

Západočeská univerzita v Plzni  
Fakulta aplikovaných věd  
Katedra informatiky a výpočetní techniky

**Diplomová práce**

**Automatická identifikace rakoviny  
z obrazových dat**

Plzeň, 2016

Bc. Michael Hadáček

**ZADÁNÍ DIPLOMOVÉ PRÁCE**  
(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení: **Bc. Michael HADÁČEK**  
Osobní číslo: **A14N0125P**  
Studijní program: **N3902 Inženýrská informatika**  
Studijní obor: **Softwarové inženýrství**  
Název tématu: **Automatická identifikace rakoviny z obrazových dat**  
Zadávací katedra: **Katedra informatiky a výpočetní techniky**

Z á s a d y p r o v y p r a c o v á n í :

1. Prostudujte doporučené metody v oblasti automatického zpracování obrazu.
2. Seznamte se s dodanou množinou medicínských dat. Množinu doplňte o vhodně vybraná data z volně dostupných databází.
3. Seznamte se s vybranými knihovnami pro zpracování obrazu.
4. Na základě studia literatury navrhnete dvě metody pro automatickou identifikaci rakoviny na snímcích.
5. Zvolené metody integrujte do funkčního celku.
6. Systém otestujte a zhodnoťte dosažené výsledky.
7. Na základě výsledků navrhnete další rozšíření/vylepšení.


Rozsah grafických prací: **dle potřeby**  
Rozsah kvalifikační práce: **doporuč. 50 s. původního textu**  
Forma zpracování diplomové práce: **tištěná**  
Seznam odborné literatury:  
**dodá vedoucí diplomové práce**

Vedoucí diplomové práce: **Doc. Ing. Pavel Král, Ph.D.**  
Katedra informatiky a výpočetní techniky

Datum zadání diplomové práce: **1. září 2015**  
Termín odevzdání diplomové práce: **12. května 2016**

  
Doc. RNDr. Miroslav Lávička, Ph.D.  
děkan



  
Doc. Ing. Přemysl Brada, MSc. Ph.D.  
vedoucí katedry

V Plzni dne 15. září 2015

## **Prohlášení**

Prohlašuji, že jsem diplomovou práci vypracoval samostatně a výhradně s použitím citovaných pramenů.

V Plzni dne 12. května 2016

.....  
Michael Hadáček

## **Poděkování**

Tímto bych chtěl poděkovat vedoucímu této práce doc. Ing. Pavlu Královi, Ph.D. za velmi vstřícný přístup, za čas, který mi věnoval a za odborné vedení při tvorbě této práce. Také bych chtěl poděkovat mé rodině za podporu.

## **Abstract**

This thesis explores possibilities of computer detection of breast cancer tumor. Two of detection algorithms – Density-Weighted Contrast Enhancement (DWCE) and Iris filter are examined and described in details. Afterwards, these algorithms are re-implemented. Their output is probability images where more suspicious areas are brighter. These images are further processed. There is set of regions of interest from original image as an program output.

DWCE and Iris filters are incorporated into one application. It allows user to use one of mentioned filters or their combination. Program also allows user to optimize its parameters and evaluate the implementation by annotated dataset.

## **Abstrakt**

Tato práce rozebírá možnosti počítačových systémů pro detekci nádoru rakoviny prsu z mamogramu. Následně jsou podrobněji popsány a implementovány Density-Weighted Contrast Enhancement (DWCE) a Iris filtr. Výstupem těchto algoritmů je obraz, na kterém jsou nejsvětleji označená místa možného výskytu nádoru. Tyto výstupy jsou nadále zpracovány a na konci procesu jsou označeny nejpodezřelejší oblasti na původním snímku.

Metody DWCE a Iris jsou zakomponovány do společného programu, který poskytuje prostředky pro spouštění jedné z nich nebo jejich vzájemné kombinace. Také je možné optimalizovat parametry a vyhodnotit úspěšnost detekce na základě anotovaných dat.

# Obsah

1	Úvod.....	1
2	Rakovina prsu .....	3
2.1	Rakovina .....	3
2.2	Zhoubný a nezhoubný nádor .....	3
2.3	Rakovina prsu .....	3
2.4	Včasná léčba .....	4
2.5	Statistika incidence a mortality .....	4
2.6	Metody diagnostiky.....	4
2.7	Mamografické pohledy .....	6
2.8	Artefakty na mamogramu .....	7
3	Computer-Aided Diagnosis.....	12
3.1	Druhé čtení (second reading) .....	12
3.2	Užitečnost CAD .....	12
3.3	Detekce, segmentace, klasifikace.....	13
3.4	Existující systémy .....	14
4	Metody CAD systémů.....	16
4.1	Region-based.....	16
4.2	Region growing.....	16
4.3	Metody založené na detekci kontur.....	17
4.4	Clustering .....	18
4.5	Prahování (Thresholding).....	19
4.6	Model-based.....	19
4.7	Algoritmy založené na srovnání dvou různých mamografických obrazů .....	21
4.8	Databáze.....	22
5	Návrh systému.....	25
5.1	Knihovny.....	25
5.2	Výběr metod.....	26
5.3	Předzpracování.....	26
5.4	Density-Weighted Contrast Enhancement (DWCE).....	31
5.5	Iris filter.....	37
5.6	Nalezení ROI.....	42
5.7	Hodnocení úspěšnosti algoritmů .....	45

5.8	Optimalizace parametrů .....	46
6	Vlastní implementace.....	50
6.1	DWCE.....	50
6.2	Iris .....	50
6.3	Support.....	51
6.4	ImageProcessing .....	51
6.5	Závislosti.....	52
6.6	Omezení .....	52
7	Dosažené výsledky.....	53
7.1	Porovnání Iris filtru a DWCE .....	53
7.2	Spojení filtrů pomocí lineární kombinace.....	56
7.3	Vylepšení s použitím vyrovnání histogramu.....	59
8	Budoucnost.....	61
8.1	Redukce falešně označených oblastí.....	61
8.2	Detekce dalších artefaktů .....	61
8.3	Klasifikace nálezu .....	61
8.4	Segmentace nálezu .....	61
8.5	Spojení více přístupů.....	62
8.6	Optimalizace systému .....	62
9	Závěr .....	63
	Pojmy .....	64
	Zkratky .....	64
	Reference .....	65
	Přílohy.....	73
A.	Výsledky lineární kombinace algoritmů .....	74
B.	Uživatelská dokumentace.....	80



# 1 Úvod

Rakovina – pojem, který v moderní populaci nahání strach. Jedním z nejčastějších nádorových onemocnění ve světě i u nás je karcinom prsu. Případů diagnózy rakoviny prsu každoročně přibývá. Její léčba je nejvíce účinná v situacích, kdy je nádor odhalený včas. Naopak postupem času se rizika včetně úmrtí pacienta zvyšují. Nejběžnějším prostředkem pro detekci rakoviny je mamograf, zařízení na bázi rentgenu, které zaznamenává snímky na film nebo do počítače. K interpretaci obrazu z mamografu tzv. mamogramu je zapotřebí expert – radiolog, jenž snímek prohlédne za účelem odhalení známek rakoviny a její diagnózy.

Zkoumání mamografického obrazu je náročná disciplína díky variabilitě snímků z pohledu různého kontrastu, odlišné tkáně každého pacienta a celkového množství zkoumaných případů. Proto se může stát, že radiolog udělá chybu při posuzování mamografického snímku. Pro zefektivnění interpretace mamogramů se zavedl systém nezávislého čtení dvěma experty. Tento postup má však také svá úskalí, zejména fakt, že radiologů není dostatek. Navíc s příchodem preventivního mamografického vyšetření mají k posouzení velké množství snímků.

Pomocné počítačové diagnostické systémy by takovou situaci mohly řešit. Zatím sice není možné zcela nahradit lidskou interpretaci mamogramu, ovšem počítač může posloužit jako pomocný systém. CAD (Computer Aided Detection and Diagnosis) je skupina prostředků řízených počítačem, která má za úkol pomoci radiologům při čtení mamografických snímků. Zavedení těchto systémů do běžné praxe by mohlo nahradit potřebu číst mamogram dvěma experty. Nadále by ho interpretoval pouze jeden za pomoci počítače. CAD systémy lze uplatnit ve více oblastech zpracování mamografického obrazu – úprava snímku pro efektivnější interpretaci, klasifikace nálezů, detekce známek rakoviny atd.

Na algoritmy pro vyhledávání nádorů jsem zaměřil pozornost v této práci. Na základě dostupných článků jsem vybral a implementoval dva a pokusil se zhodnotit jejich výsledky. Konkrétně jde o Density-Weighted Contrast Enhancement a Iris filtr. Tyto filtry mají za úkol zvýraznit podezřelé oblasti na mamogramech. Jako další krok při zpracování snímků jsem vytvořil funkci, jejímž úkolem je ve filtrovaných obrazech označit a vyříznout nejpodezřelější místa. Pomocí anotovaných dat z databáze MIAS jsem ověřil úspěšnost své implementace.

V další kapitole se čtenář dočte o tom, co je to vlastně rakovina a rakovina prsu, proč je nebezpečná a pomocí jakých prostředků jí lze diagnostikovat. Část nazvaná „*Computer-Aided Diagnosis*“ pojednává obecně o počítačových systémech v oblasti mamografie. Následující sekce „*Metody CAD*“ je věnována rozdělení a popisu

algoritmů, které byly použity nebo se stále používají pro práci s mamografickými daty. Na popis postupu při programování DWCE a Iris filtru se zaměřují kapitoly „*Návrh systému*“ a „*Vlastní implementace*“. V předposledních dvou sekcích („*Dosažené výsledky*“ a „*Budoucnost*“) se nachází zhodnocení úspěšnosti vytvořeného programu a návrh možného postupu v budoucnosti. V závěru je celá práce ve stručnosti shrnuta.

## **2 Rakovina prsu**

### **2.1 Rakovina**

Rakovina není jedno konkrétní onemocnění, tento termín představuje stovky různých chorob. Ty však mají, alespoň v určitých fázích, několik společných vlastností. Zpravidla je způsobují změny v buňkách a jejich kontrolních mechanismech, které vedou k jejich nekontrolovatelnému růstu. Takto rostoucí buňky se často seskupují do bulky zvané tumor nebo nádor [1][2].

### **2.2 Zhoubný a nezhoubný nádor**

Nezhoubné (benigní) i zhoubné (maligní) nádory rostou zprvu pouze na místě svého vzniku. Benigní v tomto stavu setrvávají. Buňky se shlukují pouze v místě prvotního výskytu. Jestliže se vyoperují, je velká pravděpodobnost, že se nádor znovu neobjeví.

Avšak u zhoubných nádorů v pozdějších fázích dochází k oddělování jejich částí. Volné zhoubné odnože mohou cestovat krví nebo mízními uzlinami po těle pacienta. Tak se dostávají do různých částí těla a také do orgánů, kde opět nekontrolovatelně rostou a vytváří dceřiná ložiska – metastázy. Tato ložiska pak svým neustálým zvětšováním poškozují životně důležité orgány [1][2].

### **2.3 Rakovina prsu**

Většina zhoubných nádorů je pojmenována podle místa, kde původně vznikly. Pokud se tedy rakovinné buňky z prsu rozšíří do jiných orgánů a tam vytvoří metastázy, stále jde o rakovinu prsu. Rakovina prsu je tedy nádor, který má místo vzniku v prsu pacienta. Těmi jsou většinou ženy, i když rakovina prsu mužů také existuje, je však o mnoho vzácnější. Tvoří zhruba 1 % z případů rakoviny prsu [3].

V případě karcinomu prsu neohrožuje život člověka prvotní nádor vzniklý v prsu, ale podružné metastázy, které se lymfatickým (mízním) nebo krevním systémem rozšíří do ostatních orgánů [2]. I přes intenzivní výzkum v oblasti rakoviny není stále jasné, co způsobuje maligní nádory prsu. Známé jsou pouze rizikové faktory, které zvyšují nebo snižují pravděpodobnost, že se u daného člověka rakovina rozvine. Tyto faktory jsou velmi rozmanité – např. jde o genetiku, věk, fyzickou aktivitu, obezitu, kouření, ale i to, jak dlouho žena kojí nebo kolik má za sebou porodů [1][4].

## 2.4 Včasná léčba

Nelze tedy efektivně předpovídat, u kterého člověka se objeví zhoubné bujení a u kterého ne. Dokud se tato situace nezmění, bude nejdůležitější prostředkem pro snížení úmrtnosti (kromě efektivní léčby) včasná detekce tumoru. Maligní nádor se z počátku chová stejně jako benigní. Poškozené buňky se koncentrují v místě vzniku. Včasným zásahem lze velké množství rakovinotvorných částí zničit a zastavit tak jejich rozšíření.

## 2.5 Statistika incidence a mortality

Podle posledních dostupných dat *Ústavu zdravotnických údajů a statistiky ČR* z roku 2011 [5] je rakovina prsu nejčastějším nádorovým onemocněním žen v České republice. Za zmiňovaný rok bylo diagnostikováno 6620 nových případů, což znamená přibližně 18 nových případů denně. Incidence (počet nových případů) podle historických dat rok od roku stoupá. Na druhou stranu i přes to, že bylo v roce 2011 zaznamenáno více úmrtí způsobených tímto nádorovým onemocněním, dlouhodobý trend úmrtnosti je klesající. Tyto dlouhodobé statistiky platí téměř pro celý svět s výjimkou rozvojových zemí, kde míra úmrtnosti každoročně stoupá. Globálně se rakovina prsu řadí na druhé místo (prvním je rakovina plic) nejčastějších nádorových onemocnění s přibližně 25 % případů a je klasifikována jako pátá nejčastější onkologická příčina smrti [6].

Stoupající míru incidence lze vysvětlit stálým zdokonalováním screeningových technik a zdravotnických systémů obecně. V mnoha zemích včetně České republiky bylo zavedeno bezplatné monitorování (screening) rakoviny prsu. Jde vlastně o preventivní vyšetření, na které chodí i ženy bez podezření na vznikající karcinom. Tímto krokem se zvyšuje pravděpodobnost včasné diagnózy rakoviny, a to je klíčové k její úspěšnější léčbě. Celé toto tvrzení koresponduje s dlouhodobým směrem ke snižování mortality.

## 2.6 Metody diagnostiky

### 2.6.1 Mamografie

Nejrozšířenějším prostředkem pro detekci nádorů v oblasti prsu je mamografie. Jde o speciální typ rentgenu, který je schopný zaznamenat změny v hustotě tkáně. Mamogram poskytuje komplexní obraz celého prsu. Lze na něm zpozorovat nádory, a to už od velikosti v řádu jednotek (někdy i desetin) milimetru. Lze pozorovat také shluky kalcifikací (zvápenatění), které mnohdy doprovází rakovinotvorné nálezy. Dále mohou být na snímku zachyceny i vedlejší známky rakoviny, jako např. bilaterální asymetrie (bilateral asymmetry) a deformace tkáně (architectural distortion). Dávka záření mamografu je velmi malá, a proto se tato metoda používá i jako hlavní nástroj pro preventivní vyšetření - screening.

Nevýhodami mamografie je velmi složitá interpretace snímků. Záznamy jsou často velmi málo kontrastní, zejména jde-li o snímek prsu s velmi hustou tkání. Interpretaci ovlivňuje kvalita snímku a zkušenosti experta, který ho posuzuje. Nedostatek radiologů a velké množství snímků pořízených při preventivním screeningu, může ovlivnit přesnost vyhodnocování jednotlivých záznamů. Nepřesností nebo chybou při čtení mamogramu může dojít v „lepší“ případě ke zbytečné operaci špatně označené oblasti (false positive, falešný poplach). V tom horším ovšem může dojít k přehlédnutí místa se vznikající rakovinou, a tím ke snižování úspěšnosti pozdější léčby. Takové chyby samozřejmě nejsou na denním pořádku a dochází k nim zřídka, proto byla ve výzkumech stanovena přesnost mamografie mezi 85 % až 90 % [7].

Mnohé výzkumy potvrzují, že používáním mamografie a zejména mamografického screeningu se snižuje úmrtnost na rakovinu prsu. Článek [8] rozebírá klady a zápory mamografického vyšetření. Mezi plusy se řadí včasné odhalení rakoviny při preventivních vyšetřeních, a tím i vyšší šance na úspěšnou léčbu. Na druhé straně stojí riziko vzniku rakoviny z ozáření. Avšak dávky záření používané při snímání jsou většinou tak malé, že tento zápor je naprosto převážen kladnými stránkami.

### **2.6.2 Ultrazvuk**

Ultrazvuk se používá jako doplňující vyšetření. Dokáže přesněji zobrazit postižené místo, a tím zlepšit klasifikaci daného nálezu. Jeho nevýhodou je, že nedokáže zobrazit přehledně celý prs. Navíc tumory menší než několik centimetrů mohou být při použití této metody přehlédnuty. Ultrazvukové vyšetření není schopné zobrazit mikrokalcifikace. Z těchto důvodů není tato metoda vhodná pro screeningová vyšetření. Pro užití ultrazvuku jako dodatečného vyšetření mluví jeho dostupnost, cena a hlavně fakt, že nevyužívá záření [9][10].

### **2.6.3 Tomograf, CT**

Computed Tomography či CT je vyšetření podobné mamografii. Je založeno na rentgenové technologii, ovšem místo zaznamenání na film se snímek ukládá do počítače. Při tomografii nejde o jediný snímek, ale o jejich sérii. Počítač vezme sadu získaných obrazů a dopočtem je schopný zobrazit i 3D pohled na zkoumaný objekt. Lze ho tedy prohlížet z libovolných úhlů. Nevýhodou CT je podstatně vyšší dávka záření potřebná na pořízení snímku.

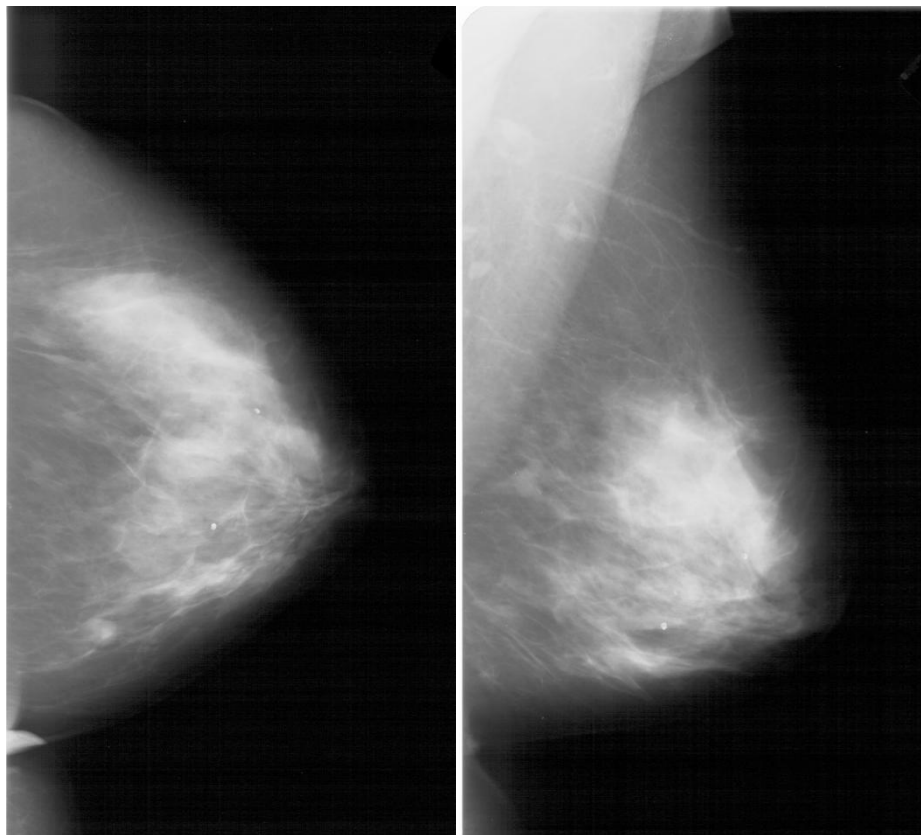
### **2.6.4 Termograf**

Termograf je zařízení založené na snímání teploty těla výkonnou termokamerou. Absolutní zdravotní nezávadnost a vysoká účinnost hovoří jasně ve prospěch termografu. Nejedná se o novou metodu v oblasti zdravotnictví. Na její konto bylo napsáno více studií. Tato metoda se nepoužívá jako screeningová metoda místo mamografie, v první řadě díky ceně zařízení a náročnému školení pro obsluhu. Další

obtížnou disciplínou je vyhodnocení termografických dat, která jsou zpracována počítačem. Při měření termografem je nutné ustálené prostředí a teplota pacienta [11][12].

## 2.7 Mamografické pohledy

Během mamografického vyšetření se často nesnímá prs pouze z jednoho úhlu, nýbrž ze dvou až čtyř. Při běžném screeningu se používají nejčastěji MLO a CC (viz Obrázek 1). Ostatní pak najdou využití při dodatečném vyšetření. Více pohledů se zavedlo z důvodu ztráty prostorové informace při snímání. Např. při pohledu shora nelze vidět detailněji tkáň okolo prsního svalu.



Obrázek 1: Mamografické pohledy - vlevo pohled shora (CC) a vpravo pohled šikmo (MLO)

### 2.7.1 Pohled shora

Craniocaudal view (CC) zaznamenává prsa při pohledu shora. Při správném umístění je prsní sval reprezentovaný úzkým pásem u kraje obrazu a bradavka se nachází naproti uprostřed [13].

### 2.7.2 Pohled šikmo

MLO – Medio-lateral oblique view ukazuje prsa při snímání šikmo. Tento pohled odhaluje i prsní sval. To je důležité, protože hodně nádorů se podle statistiky vyskytuje v horní části poblíž tohoto svalu [13].

### 2.7.3 Doplnující pohledy

Dalšími pohledy jsou medio-lateral (ML) a latero-medial (LM). Ty jsou podobné MLO. Zde se ovšem nesnímá šikmo, ale v úhlu 90° oproti CC, tedy směrem od středu hrudníku ven (ML) anebo opačně (LM).

## 2.8 Artefakty na mamogramu

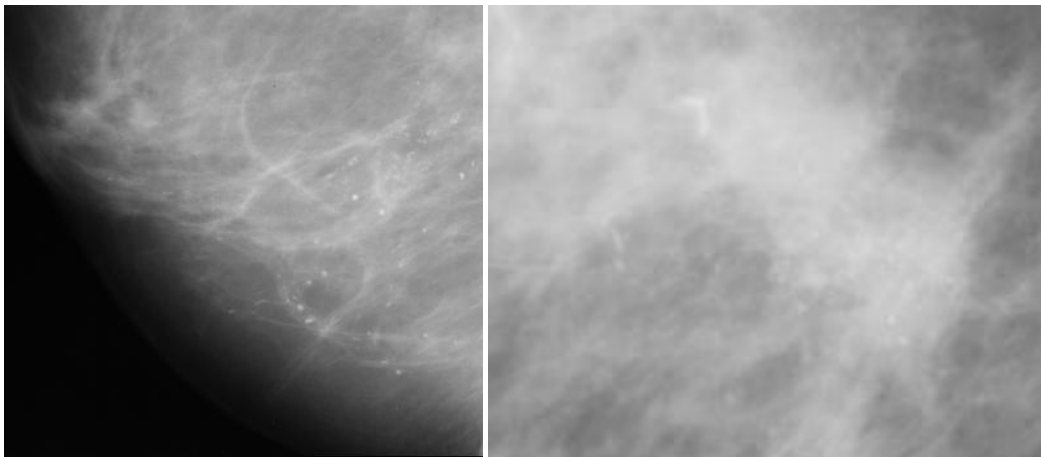
### 2.8.1 Mikrokalcifikace

Výskyt ložisek zvápenatění (microcalcification clusters) často signalizuje rakovinné bujení v brzkém stádiu. Jejich vyhledání a klasifikace je proto důležitou částí CAD programů. Kalcifikace se na mamogramu projevují jako malé světlé tečky díky tomu, že vápník více pohlcuje rentgenové záření než běžná tkáň. Avšak rozpoznání shluku vápníku je náročnější úkol, než se může zpočátku zdát.

Prvním problémem je velikost jednotlivých částí. Mikrokalcifikace zpravidla dosahují velikosti od 0,1 – 1 mm. Některé mohou být i menší, ale ty už pomocí mamografie nelze rozeznat od šumu vznikajícího při vytváření záznamu. V průměru jsou 0,3 mm velké.

Dalším úskalím pro rozpoznání pomocí počítačových algoritmů je tvar a počet kalcifikací. Shluky (clusters) se formují do všemožných tvarů a velikostí (viz Obrázek 2). Z toho důvodu je velmi náročné vyhledávat kalcifikace na základě předpokládaného modelového vzhledu nebo podle shody s případy v databázi.

I přes to, že shluky vápníku vypadají na mamografickém obraze jako světlé tečky, rozpoznat je v husté tkáni není díky minimálnímu kontrastu triviální (viz Obrázek 2 vpravo). Naopak proti tkáni s nízkou hustotou mikrokalcifikace vytváří velký kontrast [14][15].

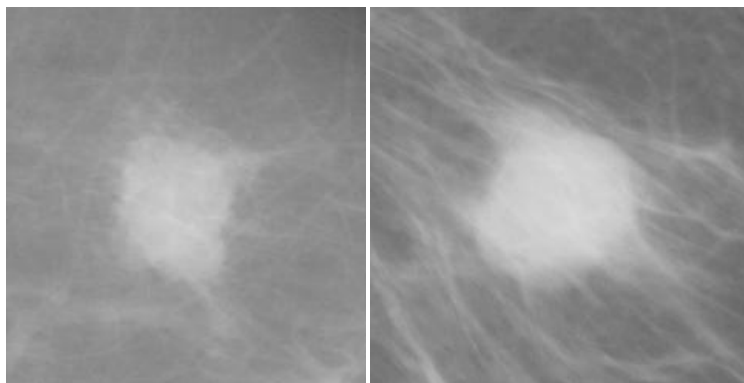


Obrázek 2: Kalcifikace - vlevo kontrastní s okolím, vpravo skryté v husté tkáni

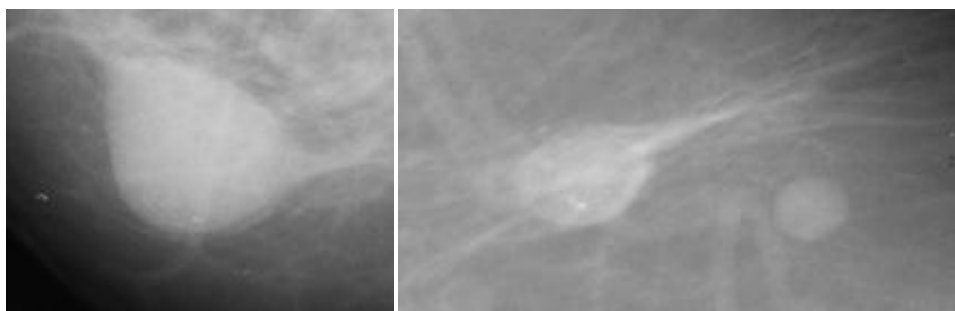
### 2.8.2 Nádor

Nádory na mamogramu reprezentuje hustší tkáň. Vypadají tedy často jako světlejší skvrna. O tvaru platí přibližně stejná pravidla jako u kalcifikací, i když u tumorů lze najít určité vzorce. Léze nejsou tak světlé a kontrastní, jak mohou být kalcifikace, proto je jejich detekce a segmentace složitější.

Nezhoubné nádory bývají zpravidla dobře ohraničené kulaté útvary. Velmi často jsou svou strukturou stejnorodé. Uvnitř nálezu se nevytváří tak vysoký kontrast jako s okolím (viz Obrázek 4). Maligní tumory naopak mívají nejasně definované ohraničení, jejich tvar mnohdy připomíná hvězdicí. Ohraničení benigních nádorů je hladké a jasné, oproti tomu u zhoubných je velmi členité a nejasné (viz Obrázek 3). Toto ovšem není všeobecné pravidlo, platí to ale u velkého množství případů. Není tak vyloučeno, že benigní nádor bude mít nejasné a členité kontury a maligní zase silně a jasně definované.



Obrázek 3: Dobře viditelné zhoubné nádory



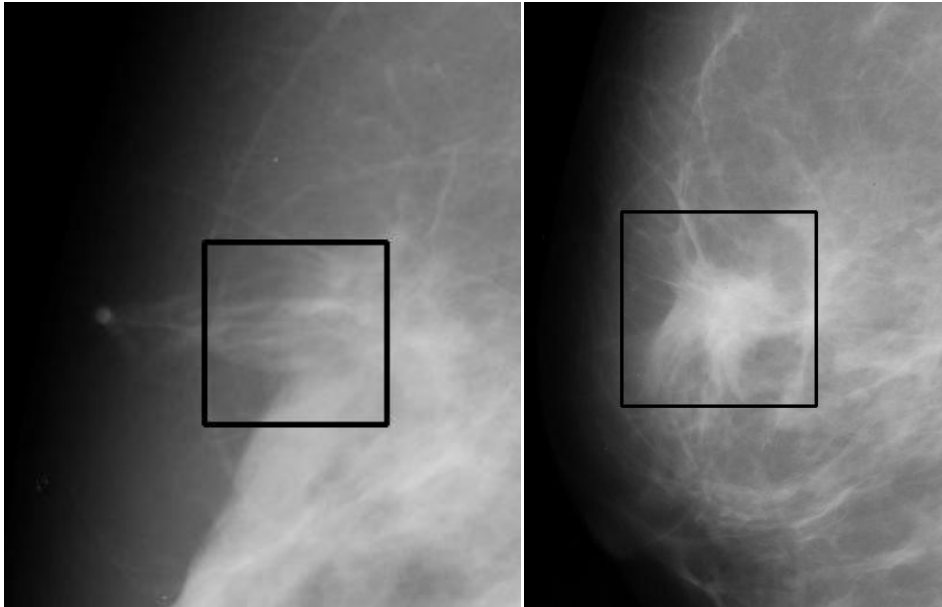
Obrázek 4: Dobře ohraničené nezhoubné nádory, na obrázku vpravo se nachází dva

### 2.8.3 Architectural Distortion

Tento termín je možno přeložit jako deformace tkáně. Jedná se o narušení „vzoru“, který vytváří tkáň prsu. Tyto deformace jdou vidět na mamografu a co víc, jsou často znakem rakoviny, a to i bez přítomného nádoru. Projevují se jako body, ze kterých směřují tenké rovné linky. Na mamogramu mohou působit dojmem místa, které je nepřírozně „zmáčknuté“ (viz Obrázek 5). Jestliže dojde ke kombinaci deformace tkáně a nádoru na jednom snímku, je velká pravděpodobnost, že tento nález bude



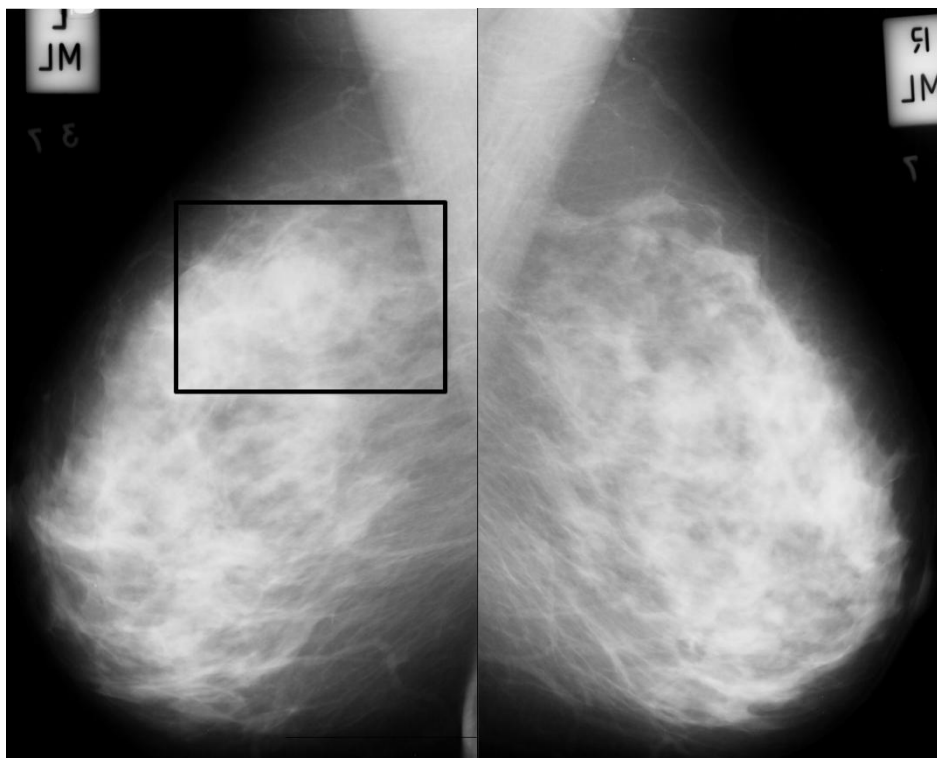
maligní. Narušení tkáně se mnohdy vyskytuje ve zpětně zkoumaných případech, tzv. false-negatives, tedy případech, kdy snímek byl hodnocen jako zdravý, ale později se u pacienta projevila rakovina. Definice tohoto jevu je relativně široká, a tak se obtížně sestavují detekční algoritmy. Vyhledávání deformace tkáně je důležité pro detekci rakoviny v raném stádiu [16][17][18].



Obrázek 5: Architectural distortion (deformace tkáně) označená černým čtvercem

#### 2.8.4 Bilateral Asymmetry

Jestliže se snímek pravého a levého prsu liší v určitém místě, mluvíme o asymetrii, která může být další známkou rakoviny. Snímky samozřejmě nejsou zcela identické. U asymetrie se například v určitém místě výrazně liší hustota tkáně (viz Obrázek 6) [15][18].



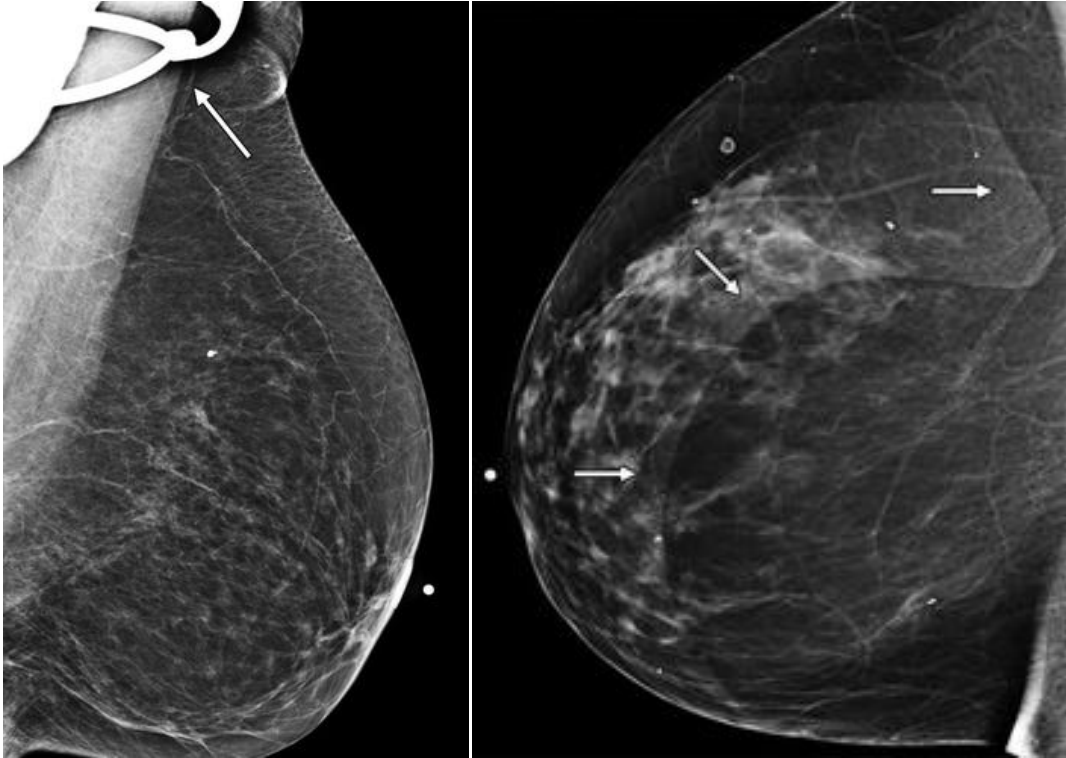
Obrázek 6: Asymetrie mezi pravým a levým prsem. U levého lze vidět zhuštěnou tkáň oproti pravému snímku

### 2.8.5 Nežádoucí artefakty na snímcích

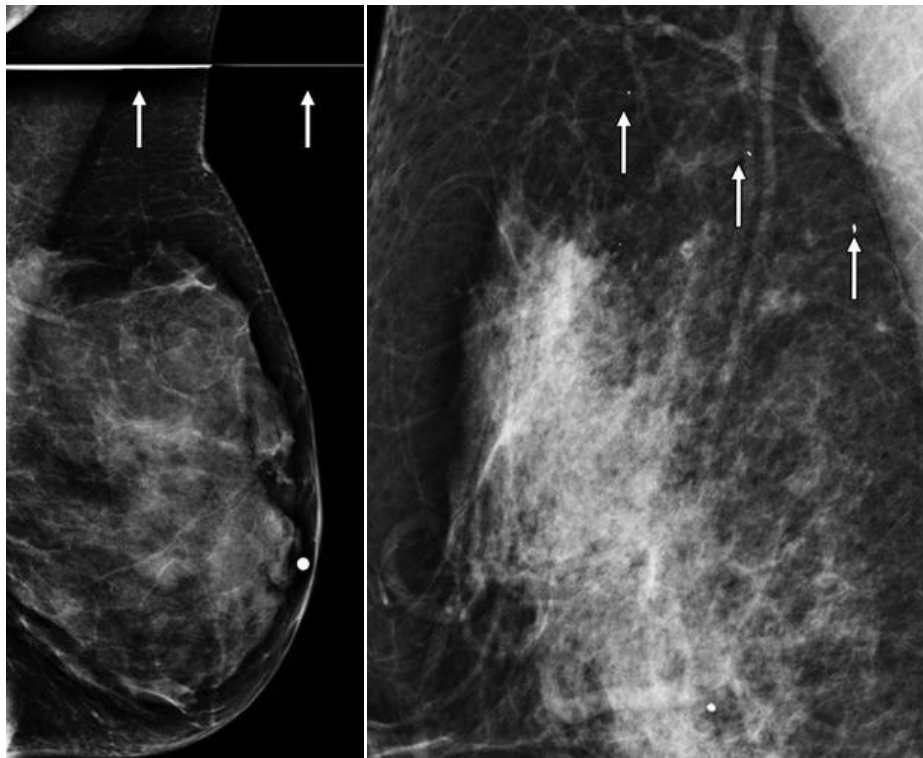
Klasifikace snímků záleží na radiologovi popř. CAD systému. Práci ale ztěžují různé artefakty a šum (noise), které se na obrázku mohou vyskytovat a občas vytvářejí dojem, že jde o nádor nebo kalcifikace, anebo naopak mohou znemožnit jejich detekci.

Tento šum může mít příčiny [19]:

- **V technice** – vypadlé pixely (viz Obrázek 8 vlevo), rozmazání nebo až nečitelnost kvůli poruše mamografu (viz Obrázek 7 vpravo).
- **Při pořizování snímku** – na kontaktních plochách se může zachytit prach (viz Obrázek 8 vpravo) nebo třeba vlas pacienta. Dalším, co zapříčiní mnohdy i úplnou nečitelnost snímku, je jeho rozmazání způsobené pohybem pacienta při zaznamenání.
- **Při digitalizaci** – čím více přístrojů zasáhne do digitalizace, tím více je prostoru pro vznik chyby. Tato kategorie částečně spadá pod první bod. Zde lze navíc uvést třeba možné horší rozlišení scannovacího zařízení, zejména pak nepřesnosti při ovládní tohoto přístroje. Tím vznikají bílá místa u nedobře zasazeného snímku na filmu, problematické části, kde jsou umístěné identifikační tabulky, atp.



Obrázek 7: Nežádoucí artefakty na mamogramu. Zachycený kus ochranné pomůcky (vlevo), mamograf zachytil "duchy" způsobené nezahřátím přístroje (vpravo) [19]



Obrázek 8: Nežádoucí artefakty na mamogramu. Vypadlý pixel (linka) z důvodu vady přístroje (vlevo). Zachycený prach při snímání (vpravo) [19]

## 3 Computer-Aided Diagnosis

Potřeba zefektivnit proces interpretace mamogramů dala podnět k vývoji tzv. CAD systémů. CAD znamená Computer Aided Detection (nebo Diagnosis) – detekce (diagnóza) pomocí počítače. Jde o souhrn počítačových algoritmů a nástrojů, které mají pomoci radiologům usnadnit nebo zpřesnit jejich interpretaci mamografických (ale i jiných) snímků.

### 3.1 Druhé čtení (second reading)

Výše jsem uvedl, že nejrozšířenějším nástrojem pro diagnostiku rakoviny prsu je mamograf. Jeho slabší stránkou jsou nároky na člověka, který má snímky vyhodnocovat. Z důvodu snížení chybovosti a hlavně zvýšení procenta úspěšnosti správné interpretace se zavádí pojem second reading. Česky ho můžeme přeložit jako „druhé čtení“.

Jde o techniku, kdy jeden mamogram prohlíží dva radiologové nezávisle. Zmenší se prostor pro vznik chyby, ovšem za cenu většího vytížení expertů. Když k tomu připočteme i jejich možný nedostatek v některých oblastech, pak taková možnost mnohdy nepřipadá v úvahu nebo je enormně náročná. V mamografii se často používají CAD systémy jako náhrada radiologa při druhém čtení.

### 3.2 Užitečnost CAD

Bylo provedeno několik výzkumů ověřujících dopad CAD systémů na detekci rakoviny. V Itálii otestovali úspěšnost klasické interpretace mamogramů dvěma experty v porovnání s jedním, kterému pomáhal počítačový systém. Ukázalo se, že oba dva přístupy jsou při detekci rakoviny stejně úspěšné [20]. Autoři jiné studie zase dlouhodobě zkoumali rozdíl mezi diagnózou rakoviny s a bez použití CAD. Jejich výsledky hovoří o téměř 20 % nárůstu odhalených případů s použitím počítačového pomocníka. Hlavně ale zaznamenali zvýšené procento zhoubné rakoviny nalezené v brzkém stádiu, a tedy s větší šancí na úspěšnou léčbu.

Na druhou stranu studie [21] ukazuje, že použití CAD zvýší úspěšnost pouze nepatrně v porovnání s druhým expertem. Obecně jsou ovšem počítačové systémy v oblasti detekce rakoviny považovány za přínos [22].

#### 3.2.1 Chybně označené oblasti (false positives)

Největší nevýhodou pomocných počítačových systémů bývá zpravidla velké množství falešně označených oblastí, které vyžadují další zkoumání – false positives [22]. Ty totiž snižují použitelnost takových programů při preventivních vyšetřeních. Radiolog musí analyzovat velké množství mamogramů, na nichž je většina případů nezasažena rakovinou. Přesto všem musí věnovat stejnou pozornost, aby snížil riziko

své chyby. CAD by mu v tom měl pomoci tím, že označí podezřelé oblasti. Avšak jestliže počítač na každém snímku vyznačí několik částí, kdy povětšinou půjde o zdravou tkáň, nelze mluvit o významné pomoci.

### 3.2.2 Kompromis mezi úspěšností a počtem falešně označených oblastí

Z výše zmíněných důvodů bývají často autoři algoritmů pro detekci rakoviny postaveni před volbu mezi úspěšností a únosným počtem falešně označených částí. Úspěchem se myslí počet nebo spíše procento správně detekovaných nádorů. Ideálním skóre pro CAD systém by byla 100 % úspěšnost a ani jeden chybně detekovaný ROI (Region of Interest, podezřelá oblast, oblast zájmu). Ovšem takový systém existuje pouze teoreticky. Zpravidla je mezi těmito dvěma hodnotami přímá úměra. Čím se algoritmus stává přesnější, tím více vykazuje i falešně pozitivních nálezů. Autoři potom musí zvolit použitelný poměr. Tento problém z části řeší klasifikační algoritmy, které třídí výřezy se zdravou tkání a nádory.

### 3.3 Detekce, segmentace, klasifikace

Zatím jsem psal výhradně o detekčních algoritmech. Computer aided diagnosis však pokrývá více odvětví v oblasti detekce rakoviny prsu. Podle výstupu a zaměření programu je můžeme rozdělit na:

- **Detekční** – má za úkol zjistit, zda se na snímku nachází rakovina, resp. znaky, které na ní mohou ukazovat. Výstupem detekčních programů bývá bod nebo jiné označení oblastí zájmu. Dalším možným výstupem může být odpověď ANO x NE. V případě kladné odpovědi se na daném snímku nachází podezřelá oblast, algoritmus už ale neumí přesně určit její místo.
- **Segmentační** – tyto metody se vyvíjí za účelem oddělit podezřelou oblast od pozadí. Na výstupu uvidíme celou oblast zájmu. Správná segmentace nádoru a jeho často velmi špatně definovaných hranic je důležitá pro některé klasifikační algoritmy. Na vstupu segmentačních algoritmů může být buď celý snímek nebo jen výřez získaný pomocí detekční metody, případně obrázek s markerem označujícím, kde se nádor nachází.
- **Klasifikační** – klasifikační algoritmy zpracovávají ve většině případů již segmentované nálezy. Jejich úkolem je v první řadě určit, zda se vůbec jedná o podezřelou oblast nebo jen o planý poplach. Poté rozlišují mezi nádory benigními a maligními. Klasifikační metody jsou často založené na učení z anotovaných dat a následném porovnávání. Nejvíce se používají jako poslední krok složených programů, kdy se v první fázi lokalizují a segmentují ROI a poté se klasifikují. Tím dojde k výraznému snížení falešně pozitivních nálezů a upřesnění diagnózy zbylých oblastí zájmu.

- **Vylepšení snímku (image enhancement)** – není zcela samostatná kategorie, ale často jako počáteční krok jedné z výše uvedených. V některých situacích počítač pouze snímek zvýrazní, a tím jeho práce končí. Ve většině případů jde o zvýraznění kontrastu, které pomáhá odhalit špatně viditelné nádory v husté tkáni.

### 3.4 Existující systémy

#### 3.4.1 R2 ImageChecker

R2 Technology je firma vyvíjející medicínské diagnostické systémy. Zaměřuje se zejména na rakovinu prsu a plic. Tato firma navrhla a vytvořila CAD systém s názvem *ImageChecker*, který má pomáhat radiologům při interpretaci mamografických snímků. Tento systém byl původně určen k redukci false-positives označených expertem. Později se ovšem rozrostl i o diagnostickou část. Jedná se nejspíše o nejrozšířenější CAD systém. V roce 2004 dostala firma R2 oprávnění americké *Food and Drug Administration* (FDA) používat *ImageChecker* ve spojení s plně digitálními snímacími zařízeními firmy Hologic [23]. Tato společnost je mimo jiné největším světovým výrobcem medicínských zařízení na zpracování obrazů.

U komerčního systému je pochopitelné, že autoři nezveřejňují použité algoritmy, které tvoří jejich CAD. Je ovšem vcelku předvídatelné, že systém používá sérii několika metod a přístupů. Při vyhodnocování snímků se řídí více pohledy a je schopný detekovat nádory i kalcifikace.

Úspěšnost tohoto komerčního CAD systému byla zdokumentována ve více studiích [24][25][26]. Podle prospektu k *ImageChecker(u)* verze 9.3 dosahuje citlivosti při detekování mikrokalcifikací až 97 % při 0,64 FP na obrázek. U nádorů je citlivost o něco nižší a to 88 % při 1 falešném ROI na snímek [27]. Výstupem *ImageChecker(u)* je snímek s vyznačenými podezřelými místy, kterým by radiolog měl věnovat zvláštní pozornost.

#### 3.4.2 The Second Look®

Druhou firmou v oboru CAD systémů je iCAD. Jejich oblast zájmu se ovšem nachází v trošku jiném spektru. iCAD si sama vyrábí a integruje fyzické snímací a digitalizační zařízení i softwarový CAD systém. Stejně jako R2 dostali oprávnění pro spolupráci s firmou Hologic a jejich digitálními snímači.

Výstupem tohoto systému je opět mamogram s označenými podezřelými oblastmi. Systém *Second Look* označuje kružnicí potenciální nádory a jejich doprovodné jevy (architectural distortion, bilateral asymmetry, asymmetry density) a obdélníkem shluky kalcifikací.

CAD systém od iCAD dosahuje úspěšnosti okolo 89 % pro nádory a 98 % pro kalcifikace [28][29], to jsou podobné výsledky jako pro *ImageChecker*.

## 4 Metody CAD systémů

Moje práce je zaměřena na detekční a částečně i segmentační algoritmy. Metod, algoritmů a heuristik se během let vyvinulo nepřehledné množství. Mnohé články uvádějí soupis těchto přístupů [15][30]. Bylo použito členění dle „*A review of automatic mass detection and segmentation in mammographic images*“ [22], na základě kterého vznikl text v této sekci. Algoritmy jsou rozděleny do kategorií podle přístupu k danému problému.

### 4.1 Region-based

Tyto metody jsou založeny na udržování homogenity v oblasti zájmu. Výstupem jsou regiony, které jsou v určitém ohledu stejnorodé. V potaz se bere možná chyba nebo lépe řečeno práh, který tvoří hranici mezi zařazením nebo nezařazením pixelu do dané oblasti. U tohoto druhu metod je ovšem těžké určit počáteční krok algoritmu, který musí ležet uvnitř toho, co se snažíme detekovat/segmentovat. V našem případě by měl seed (počáteční bod/y výpočtu) ležet uvnitř nádoru. Někdy se jako vstup tento bod nebo skupina bodů definuje manuálně. Jak bylo uvedeno v odstavci 2.8.2, nádory by měly na snímku být homogenní a tvořit určitý kontrast s okolím, proto by region-based metody měly uspět při jejich detekci a segmentaci.

### 4.2 Region growing

„Růst regionu“ je velmi stará technika pro práci s obrázky. Je založena na rozšiřování oblasti směrem od seed pixelu podle určitých kritérií. Jednoduchým příkladem může být růst regionu, jehož kritérium je intenzita šedé (jak moc je bod světlý). Začneme počátečním bodem a postupně zkoumáme všechny přilehlé pixely a ptáme se, zda se jejich intenzita neliší více než o prahovou hodnotu. Jestliže se neliší, přidáme zkoumaný bod do výsledné oblasti a prozkoumáme jeho sousedy. V opačném případě pixel nepřidáme a ani neprověřujeme jeho přilehlou oblast. Rozšiřování končí ve chvíli, kdy už se není kam dále posouvat, kdy jsme narazili na hranice segmentovaného objektu.

Postupem času se tyto algoritmy neustále vylepšovaly. Do kritérií pro přijetí nových pixelů například byly dosazeny podmínky ohledně tvaru daného regionu založené na maximalizaci radiálního gradientního indexu a na pravděpodobnostním modelu [48]. Dalším postupným zlepšením bylo rozšíření příslušnosti k regionu o další hodnotu, kde se neurčovalo pouze přijat nebo nepřiját, ale také přijat s výhradami s následným prošetřením příslušnosti [31]. Některé z variant přidávaly váhu bodům na základě výskytu hran nebo kontur nalezených v jejich okolí [49]. Jiní se vydali opačnou cestou a navrhovali metody, které segmentují pozadí a to, co jim zbude, je skutečná oblast zájmu [50].



Jestliže pro region-based metody platí, že jsou silně závislé na počátečních bodech, v případě region-growing to platí dvojnásobně. Pakliže si zvolíme špatný seed, další postup pravděpodobně nebude úspěšný. Nakonec vlastně budeme segmentovat úplně jinou oblast, než jsme původně zamýšleli. Proto je jeho volba jedním z nejdůležitějších kroků. Některé segmentační algoritmy, hlavně v dřívější době, vyžadovaly manuální vstup. To znamená, že expert označí body v podezřelé oblasti a pak spustí další proceduru.

Někteří výzkumníci se pokusili navrhnout automatickou volbu počátečního bodu. V nejjednodušším případě šlo o vybrání nejsvětějších bodů [51]. Úskalí tohoto přístupu je ovšem ve výskytu kalcifikací nebo nežádoucích artefaktů, které zpravidla dosahují vyšších intenzit šedé, než okolní tkáň. Dalším možným přístupem k tomuto problému se stala volba seed pixelů na základě upraveného obrazu. Odečtením dvou „gaussovsky“ rozostřených obrazů od sebe vznikne další obraz, na kterém jsou zvýrazněny oblasti, kde se původně měnila intenzita – byl zde přechod nebo hrana.

#### **4.2.1 Watershed metody**

Watershed transformace patří mezi morfologické operace. Je založena na topografii. Obraz představuje jakousi mapu, kde intenzita šedé značí „nadmořskou“ výšku. Transformace začíná na nejnižších bodech. Postupně od nejnižších zaplňuje segmenty mapy „vodou“. Jak se plní různé nádrže začínají se slévat dohromady. V místech, kde dochází k tomuto jevu, vytvoříme „hráze“ (předěly). Algoritmus končí v momentě, kdy se zaplaví nejvyšší bod. Výstupem je hierarchie těchto „přehrad“. Jejich filtrováním a výběrem získáme požadované regiony [32].

Nejdůležitější částí u tohoto typu algoritmů je předzpracování – úprava snímku před aplikováním. Na obraze, kde se nachází hodně šumu a nežádoucích artefaktů, vzniká velké množství regionů. Nejčastějším krokem v přípravě obrazů je jeho rozostření, a s tím redukce šumu. Watershed metody v oblasti segmentace a detekce nádorů jsou popsány v článcích [52][53][54].

#### **4.2.2 Split and merge**

Nepříliš používaný typ algoritmů v úpravě mamografických obrazů je split and merge. Na začátku obraz rozdělujeme do více a více oblastí, dokud nebudou splňovat podmínku homogenity. V další fázi regiony opět spojujeme, tentokrát však podle další podmínky [55].

### **4.3 Metody založené na detekci kontur**

Používání algoritmů pro detekci hran se objevovalo už od samých počátků analýzy medicínských dat. Myšlenkou těchto metod je detekovat přechody mezi nižší a vyšší intenzitou. Následně v těchto hranách vyhledat ty, jenž se uzavírají a ideálně tvoří

oválný tvar (potenciální nádor). Jak už bylo zmíněno výše, tento úkol může být velmi náročný díky vlastnostem mamografických snímků. Navíc hrany zhoubných nádorů jsou velmi často nejasné, rozmazané a hodně členité, tím se jejich identifikace ještě více ztěžuje.

Nevýhodou metod založených na detekci kontur je jejich nepřesnost nebo neúplnost. Tím se myslí nemožnost dopočítat chybějící kusy hran tak, jak to dělá lidské oko a mozek. Kontury mohou být narušeny různými artefakty od kalcifikací až po šum, nebo například hustější tkání, která zakryje přechod. Na výstupu algoritmů pak mohou být přerušené čáry. Dalším zatěžujícím elementem je charakter mamografických dat. Na snímku se bohužel objevuje spousta objektů, které lze klasifikovat jako hrany. S tím vším se vyhledání nádoru stává velmi obtížným. Často se metody založené na vyhledávání kontur používají pro zpřesnění dosaženého výsledku po segmentačních algoritmech.

Úprava snímku před samotným aplikováním hranových detektorů se zdá být nezbytným krokem. Příkladem může být i v této práci použitý filtr Density-Weighted Contrast Enhancement [33] pracující na bázi násobení pixelů na stejných pozicích dvou obrazů, které vznikly z původního. Jeden z nich se vytvoří aplikováním dolní propusti a hodnoty jednotlivých bodů udávají intenzitu šedé v dané oblasti. Druhý obraz naopak poskytuje informaci o změně kontrastu. Zvýrazňuje místa, kde dochází k náhlým změnám barvy (hrany, přechody). Při spojování těchto dvou verzí původního snímku dochází k násobení intenzity kontrastem. To znamená, že na výsledném filtrovaném obraze budou nejsvětější místa, která byla světlá a zároveň s kontrastem vůči okolí.

Jiný přístup byl použit v [56]. Autoři pojali opět obrázek jako mapu. Prahováním na různých topografických hladinách, se povedlo vyfiltrovat různé části prsu: sval, tkáň, ale hlavně nádory. Reprezentovat nalezené hrany jako síť bylo dalším pokusem o segmentaci mamogramu [57]. Po skončení algoritmu se s největší pravděpodobností nacházel nádor v místě s nejvyšší koncentrací uzlů. Další práce budou uvedeny pouze výčtem: použití active contour model (snakes) [58], metoda založena na dynamickém programování [59].

#### **4.4 Clustering**

Clustering neboli shlukování je metoda hodně podobná region-based algoritmům. Při svém průběhu vytváří shluky podobných pixelů tak, aby si body ve shluku byly co nejpodobnější a zároveň, aby se jednotlivé shluky neprolínaly a nepodobaly se jeden druhému.

Shlukovací algoritmy můžeme rozdělit na dva typy:

1. **Hierarchické** – vytváří strukturu nalezených clusterů; většinou dosahuje lepších výsledků, ale nepoužívají se tolik, díky jejich výpočetní náročnosti, která u velké množiny dat způsobuje dlouhý čas výpočtu.
2. **Nehierarchické** – jejich výstupem je jedna množina na sobě nezávislých shluků. Používají se daleko více kvůli menší algoritické složitosti. Za jejich hlavní nevýhodu se dá považovat fakt, že potřebují při inicializaci znát počet shluků, které se mají vytvořit. Dále neberou v potaz informace o tvaru clusterů.

## 4.5 Prahování (Thresholding)

Metody založené na prahování jsou většinou implementačně velmi jednoduché. Jsou založené na vybrání dat, která patří do určených intervalů (prahů). U mamografických snímků je nejčastější způsob založený na nastavování prahů podle intenzity šedé. Prahování lze použít na prvotní hrubé rozčlenění obrazu a odlišení nejzajímavějších částí, které se později zpracují například výše nebo níže uvedenými algoritmy. Anebo právě naopak – nejdříve je obraz zpracovaný určitým algoritmem, jehož výstupem je pravděpodobnostní „mapa“, která se následně zprahuje.

Takový přístup používá i Iris filtr [34][35] blíže popsany v 5.5. Ten nejdříve vytvoří obraz, který ukazuje stupeň konvergence gradientů k danému místu. Jeho nejsvětlejší oblasti leží pravděpodobně uprostřed oválných míst tvořících kontrast s okolím. Tento pravděpodobnostní obraz je nadále zprahován a jsou z něj vybrány pouze nejzajímavější oblasti.

Dále jsou uvedeny odkazy na různé algoritmy z této skupiny: prahování založené na minimalizaci lokální „fuzzy entropie“ [60]; určení prahu pro každý mamogram jinak na základě informací z histogramu [61]; použití prahování na více úrovních pro vylepšení hranového detektoru [62].

## 4.6 Model-based

V této kategorii se nachází metody mající nejdříve „učící“ část. Učí se prvky, které signalizují, že se na snímku vyskytuje nežádoucí artefakt. Některé pak mohou získávat znalosti i o tom, co se na obrazu nacházet může, aby dokázaly rozlišit mezi „dobrým a špatným“.

Pro učení jsou algoritmům připravena data, na kterých se nachází nádory spolu se zdravými vzorky. Ty se dají získat z anotací k dostupným databázím (např. MIAS). Druhým způsobem je přizvat k vývoji experty, kteří vyberou reprezentativní vzorky pro tuto fázi.

Algoritmy pak podle získaných informací vyhledávají na snímcích určité vzory a porovnávají je s naučenými. Podle tohoto srovnání určí výsledek. Kritické pro tyto metody je ustanovit, podle jaké míry se bude testovat podobnost jednotlivých nalezených artefaktů. Problémem však zůstává obrovská variabilita, v jaké se nádory objevují. Většinou se tyto algoritmy řadí ke klasifikačním a detekčním.

Za nejběžnějšího zástupce model-based metod se dá považovat pattern-matching (vyhledávání vzorů). Zde se algoritmus učí podle sady výřezů, na kterých se nachází nádor nebo jiný nežádoucí artefakt. Během zpracování obrazů se porovnávají výřezy z tréninkové sady s výřezy (nebo celými obrazy) zkoumaných dat. Na základě určené míry podobnosti se pak klasifikují. Touto mírou může být např. korelační koeficient [63][64]. Vzhledem k tomu, že základní myšlenkou těchto algoritmů je vyhledávání ukázkových vzorů v testovaných datech, trpí nejvíce kvůli rozmanitosti tvarů nádorů. Částečným řešením může být velká základna ukázkových ROI, která však nikdy neobsáhne všechny případy.

Obdobný přístup zvolili autoři v článku [36]. Z trénovací sady vytvořili pravděpodobnostní model nádoru, jenž vznikl sjednocením velikostí všech ROI a proložením do jednoho obrázku. V článku nezvolili pouze jednu velikost, ale několik různých, k nimž přiřazovali podobně velké výřezy. V místě, kde se kontury nejvíce překrývaly, byla v pravděpodobnostním modelu největší váha. Naopak do oblastí, kam zasahuje minimum nebo žádný ROI, byla přiřazena nejmenší. Vstupem algoritmu jsou reálné výřezy, které lze získat výše popsanými metodami. Prvním krokem bude opět přizpůsobení velikosti. Následuje pattern-matching, do kterého navíc vstupuje váha pravděpodobnostního modelu. Jako měřítko podobnosti a pomocné schéma ke zpřesnění učení posloužila Bayesova věta.

Další z řady přístupů založených na definici modelů určuje podezřelou oblast pěti klíčovými vlastnostmi [65]. Říká, že nádor na snímku by měl obsahovat:

1. Globální maximum na rozostřeném obrazu
2. Lokální maximum na původním
3. Lokální maximum na obraze, který vznikne odečtením dvou různých rozostření původního snímku.
4. Malou podezřelou oblast s nízkým kontrastem  
nebo
5. Malou podezřelou oblast s vysokým kontrastem

Do další skupiny můžeme zařadit metody založené na analýze informací získaných z mapy gradientů, která poskytuje prostorovou informaci [66].

Poslední velkou skupinu tvoří metody založené na neuronových sítích. Sítě se učí ze sady obrazů zasažených i nezasazených rakovinou. Takové algoritmy se často používají na snížení počtu chybně označených oblastí. Do neuronové sítě zadáme segmentované ROI a následně síť vyhodnotí, zda skutečně jde o podezřelé místo nebo ne. Části, které byly klasifikovány jako zdravé, vyřadíme a ponecháme pouze zasažené [67][68][69].

## **4.7 Algoritmy založené na srovnání dvou různých mamografických obrazů**

Pro přehled zde ještě uvedu metody, které hledají nádory v závislosti na porovnání dvou mamogramů. Tento typ se většinou zaměřuje na detekci podezřelé oblasti.

Může jít o porovnání:

- Mamogramů levého a pravého prsu
- Dvou různých pohledů na stejný prs
- Dvou snímků stejného prsu, které vznikly v časovém odstupu

### **4.7.1 Porovnávání snímků levého a pravého prsu**

I když prsy nikdy nebudou zcela identické, jejich struktura by do jisté míry být měla. Pro porovnání obou mamogramů se používá jejich „prosté“ odečtení. Vyniknou tak artefakty a oblasti, které jsou výraznější na jednom či druhém snímku. Největší výzvou tohoto typu algoritmů je vycentrování obou obrazů. K tomu se používá tři způsobů:

1. Najdou se jednoznačně definované prvky u obou snímků, které se překryjí. Mezi takové prvky můžeme zařadit: pozici bradavky na snímku, svalu a výrazných oblastí.
2. Z výrazných oblastí se vytvoří stromy, které se následně překryjí.
3. Lokalizace kontur prsu na snímku, nalezené kontury se dají přes sebe.

Zástupci této skupiny jsou např. [70][71].

### **4.7.2 Porovnání dvou různých pohledů na stejný prs**

Do této kategorie se řadí algoritmy, které zkoumají dva nebo více pohledů na stejný prs (viz 2.7). Používají se většinou pro určení, zda se jedná o pravý nálezný nebo o špatně označenou zdravou oblast. Nejdříve se na obou snímcích vyhledají podezřelá místa, a poté se porovnají snímky mezi sebou. Jestliže se jedna oblast nachází na obou, je zařazena do podezřelých. K detekci ROI se používá jakýkoliv detekční algoritmus. Alternativně mohou být nejdříve identifikovány oblasti na jednom snímku a na základě

získaných informací přizpůsobit vyhledávání na druhém – např. nastavení prahů, zúžení zpracovávané oblasti atp.

Po detekci na obou snímcích zbývá porovnat nalezená ROI. To ztěžuje fakt, že každý obraz je zaznamenaný z jiného úhlu. Jako orientační měřítko se používá vzdálenost od bradavky, která by u obou běžně používaných typů pohledu (MLO a CC) měla být přibližně stejná [72][73][74].

### **4.7.3 Porovnání dvou snímků jednoho prsu s časovým odstupem**

Porovnávání snímků zaznamenaných s časovým odstupem, používají radiologové, aby zjistili, jak se daný nález vyvíjí. Nejprve se najdou podezřelá místa na nejnovějším snímku a určí se jejich souřadnice stejným způsobem jako u předchozího typu. Poté se prohlíží starší snímky a zkoumá se, jak se daná podezřelá oblast měnila v minulosti [75][76].

## **4.8 Databáze**

### **4.8.1 Digitalizovaná vs. digitální**

Standardem ve vyšetření rakoviny prsu jsou screen-film (filmové) mamografy. Tedy přístroje zaznamenávající průchod rentgenového záření na film. Vyznačují se zejména vysokým rozlišením, díky kterému lze vyhledat i menší objekty, jakými jsou mikrokalcifikace. Avšak pro zpracování v počítači je nutné snímek z filmu převést do digitální podoby. O to se starají vysoce výkonné scannery. Co největší přesnost a rozlišení scannovacího přístroje je zásadní z hlediska kvality digitalizovaného obrazu.

S pokrokem doby a vývojem přišly plně digitální mamografické detektory, které ukládají obraz do počítače ihned při vyšetření. Jejich výhodou oproti filmovým je výrazně lepší kontrast, jenž je velmi důležitý k rozpoznávání nádorů v husté tkáni.

Nevýhodou screen-film mamografie je možnost zanesení šumu a nepříznivých artefaktů během digitalizace snímků nebo při jejich skladování. Slabší stránkou je také menší kontrast oproti digitálním snímkům, které jsou zase u rozlišení limitovány velikostí pixelů. To může vést k horší detekci opravdu malých objektů. Výzkumy však přinesly závěr, že oba přístupy jsou stejně dobré pro detekci nádorů i kalcifikací [37][38].

### **4.8.2 MIAS**

Většina vývoje a testování této práce probíhala s použitím MIAS databáze. *Mammographic Image Analysis Society* je skupina vědců a výzkumníků z Velké Británie zabývajících se zpracováním a interpretací mamogramů. Ti vytvořili volně

dostupnou databázi mamografických snímků pro vědecké účely. Důvodem vytvoření této databáze bylo ohodnocování a testování nově vznikajících CAD systémů.

Snímky do databáze byly pečlivě vybírány. Jsou zde zastoupeny různé výše popsané mamografické artefakty (viz 2.8), symbolizující zhoubné i nezhoubné nádory. Ty jsou doplněny dalšími snímky nezasaženými rakovinou. Ve výběru nechybí různé druhy a hustoty tkáně pacientek. Tyto podmínky jsou vhodně zkombinovány tak, aby se od každé varianty vyskytovalo v databázi několik případů. Tím by měla být poskytnuta dostatečně široká základna a její spektrum k objektivnímu hodnocení CAD systémů.

V databázi se nachází 322 snímků, resp. 168 párů levého a pravého prsu. Obrazy jsou zachycené pouze zešikma tedy MLO pohled (viz 2.7.2). Při tvorbě bylo jedním z omezení, že snímky budou pocházet pouze z jednoho centra *Institutu pro mamografický screening* ve Velké Británii (*UK National Breast Screening Programme*).

Jedná se o databázi digitalizovanou nikoliv digitální. Zde se snímky pořídili mamografem na film a až později byly digitalizovány výkonným scannerem. Původní obrazy byly digitalizovány na 50  $\mu\text{m}$  na pixel a s hloubkou šedé 8 bitů na pixel. Později však byla databáze redukována na 200  $\mu\text{m}$  a snímky zarovnány na rozlišení 1024 x 1024.

Mimo snímků se u databáze nachází ještě soubor s informacemi o jednotlivých případech. Data jsou anotována následujícím způsobem:

mdb012 F CIRC B 471 458 40

- mdb012 – generovaný název snímku
- F – Typ tkáně
- CIRC – typ nálezu
- B – nezhoubný x zhoubný (M)
- Další jsou souřadnice středu artefaktu a poloměr kruhu, který ho obkresluje.

Typ tkáně popisuje, jak je tkáň hustá a žlázovitá. Varianty vypadají takto:

- F - Fatty – v tkáni převažuje tuk
- G - Fatty-glandular – tučná a místy žlázovitá
- D - Dense-glandular – hustá tkáň

Výpis možných nálezů na snímcích:

- CALC – Calcification – kalcifikace
- CIRC – Well-defined/circumscribed masses – dobře ohraničený nádor

- SPIC – Spiculated masses – nádor ve tvaru hvězdice s množstvím „paprsků“ jdoucích směrem od jeho středu ven.
- MISC – Other, ill-defined masses – nádor s nejasně definovanými okraji
- ARCH – Architectural distortion
- ASYM – Asymmetry
- NORM – Normal – snímky nezasažené rakovinou

### 4.8.3 DDSM

*The Digital Database for Screening Mammography* je digitální databáze určená zejména pro vývoj počítačových algoritmů, které mají pomáhat při screeningové mamografii. Tvoří základ pro vývoj CAD systémů a také prostředků pro výuku a zdokonalování v oblasti mamografie. DDSM je projekt financovaný oddělením lékařského výzkumu a vývoje americké armády, na kterém se podílejí zejména nemocnice a univerzity.

Databáze obsahuje přibližně 2 600 případů, které se skládají z MLO a CC snímků na oba prsy pacientky. Obrazy byly digitalizovány jedním ze čtyř různých scannerů s přesností od 42  $\mu\text{m}$  do 50  $\mu\text{m}$  na pixel.

U každého případu jsou přidány informace o pacientce a datu pořízení snímků. K obrazům prsů zasažených rakovinou je přiložen soubor obsahující informace ohledně nálezu. Tyto informace jsou označeny a klasifikovány expertem a zapsány do kódu čitelného pro počítače. Mamogramy jsou uloženy ve speciálním formátu bezztrátového jpeg [39][40].

Tvůrci databáze k ní připravili software, který by měl umět stahovat a filtrovat případy, zobrazovat je a ukazovat podezřelé oblasti. Program se mi ovšem nepovedlo zprovoznit a bez něj je velmi obtížné převést snímky do formátu, jenž umí zobrazit běžně používaný software. I přes to, že lze získat tyto snímky i z jiného zdroje v čitelném formátu, vznikl další problém. Interpretace označených oblastí také není zcela triviální. Z tohoto důvodu se vývoj zvolených algoritmů odehrával především s databází MIAS.



## 5 Návrh systému

### 5.1 Knihovny

#### 5.1.1 Opencv

Opencv je volně šiřitelná knihovna napsaná v C/C++. Poskytuje však rozhraní i pro další programovací jazyky jako například pro Javu, Python, C nebo Matlab. Knihovna funguje pod operačním systémem Windows, Linux i Mac OS. Obsahuje více než 2500 algoritmů z oboru zpracování obrazu a videa a strojového učení. Záměrem autorů nebylo pouze vytvořit knihovnu, která je volně dostupná a šiřitelná, ale hlavně dobře optimalizovaná. Hlavním účelem vzniku této knihovny bylo poskytnout komplexní sadu funkcí pro podporu vývoje aplikací pro zpracování videa a obrázků a později i strojového učení, která se bude snadno používat.

Díky spolupráci se zaběhlými společnostmi jako Google, Intel, Microsoft atd. se řešení používající knihovnu OpenCV objevila ve spoustě aplikací jako jsou úprava streetview obrazů, detekování osob ve sledovacích systémech, rozpoznávání obličejů, detekce nežádoucích objektů na ranvejích atp. [41][42].

Knihovna je dostupná na [41] v aktuální verzi 3.1. Pro zpracování této práce jsem použil předchozí verzi 3.0, protože poslední release ještě nebyl dostupný.

#### 5.1.2 Intel Threading Building Blocks (TBB)

Threading Building Blocks je open source a multiplatformní knihovna pro realizaci paralelních výpočtů. TBB vyhrála několik ocenění, mezi nimi i cenu za nejlepší open source vývojářský nástroj pro paralelní programování. Jejím hlavním cílem je zjednodušit práci programátorům vyvíjejícím paralelní programy a zároveň zrychlit a optimalizovat paralelní část programu.

TBB nepoužívají přímo vlákna poskytované operačním systémem, ale pracují s tzv. „tasky“. Tedy přiřazují jednotlivým procesorovým vláknům/jádrům úkoly, na kterých daná výpočetní jednotka pracuje, dokud není úloha ukončena. Tento fakt napomáhá mimo jiné snížit dopad cache-coolingu. To je situace, která nastává, když se vlákna přepínají na procesoru např. kvůli tomu, že jich je více než výpočetních jednotek procesoru. Pak si každé vlákno musí vždy natahat svoje pracovní data do cache paměti, jenž se stále přepisuje a tím zpomaluje program. TBB ovšem poskytují daleko více optimalizačních technik, pro zefektivnění paralelních výpočtů [43].

## 5.2 Výběr metod

Při vybírání metod, které nakonec budu zkoušet implementovat, jsem se řídil třemi kritérii:

1. **Počet citací daného článku** - zajišťují alespoň nějakou relevantnost uváděné implementace. Nemělo by se mi tedy stát, že se pustím do realizování algoritmu, který funguje pouze teoreticky, ale v praxi úplně pohoří.
2. **Uváděná úspěšnost detekce/segmentace** - existuje několik článků popisujících objevené metody CAD systémů a jejich srovnání podle určitých kritérií. Rád bych se ve své práci věnoval algoritmům, jenž byly podrobeny srovnání a byly v něm úspěšné.
3. **Srozumitelnost a úplnost textu** - v mnohých člancích (nakonec i ve zvoleném) chybí spoustu podstatných detailů. To pak sice nechává prostor pro vlastní kreativitu, ovšem na druhou stranu je možné do algoritmů zanést chyby.

Na základě článku [22] jsem vybral dvě metody, které jsem později implementoval a otestoval. Autoři článku vybrali a znovu naprogramovali zástupce různých přístupů k detekci a segmentaci nádorů. Následně všechny algoritmy ohodnotili a srovnali podle stejného kritéria. Za hodnotící prostředek zvolili (F)ROC – *Free-response Receiver Operating Characteristic*, křivka sestavená jako poměr mezi dobře lokalizovanými ROI a špatně označenými (poměr true a false positives). Implementované algoritmy vyzkoušeli na dvou databázích. Jedna z nich byla výše zmíněná MIAS a druhá plně digitální Trueta.

Pro obě databáze dosáhl jedny z nejlepších výsledků **Iris filtr**. Ten dokázal překonat pouze na MIAS databázi filtr **DWCE**. Dohledal jsem více informací o těchto dvou metodách a nakonec je vybral i pro svou práci.

## 5.3 Předzpracování

Před samotným aplikováním algoritmů na zvýraznění a segmentaci nálezů je žádoucí mamografické snímky připravit. Tento krok je společný pro obě vybrané metody. Jeho účelem je co nejvíce zefektivnit detekci nádorů. Optimalizace může spočívat např. ve zmenšení obrázku z důvodu náročnosti algoritmu nebo v odfiltrování nežádoucích částí snímku.

### 5.3.1 Odstranění artefaktů a separace prsu z mamogramu

Při přípravě snímku je nutné odstranit veškeré elementy, které nejsou přímou součástí snímku prsu. V případě digitalizovaných snímků se stává, že se k nim přidávají různé popisy v podobě identifikačních tabulek (viz Obrázek 9). Jedním z dalších pro detekci nežádoucích jevů je nepřesné umístění snímku do digitalizační jednotky, kvůli kterému vznikají velmi kontrastní bílé okraje (viz Obrázek 9 vlevo nahoře).



Obrázek 9: Digitalizované mamogramy. Snímek z DDSM (vlevo) s identifikační tabulkou v pravém horním rohu a nestejným pozadím. Ukázka snímku z MIAS databáze (vpravo), na němž je zachycena kontrastní bílá tabulka.

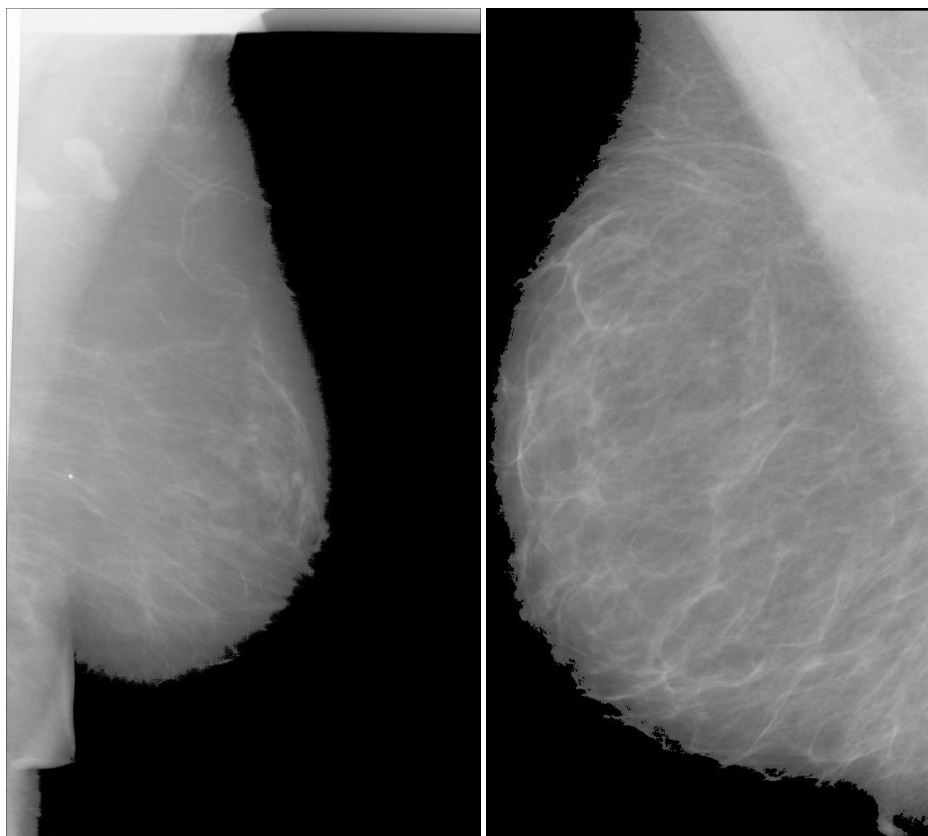
Takové části jsou často velmi světlé a leží na černém pozadí. Mají tedy vysokou intenzitu šedé a velký kontrast, proto zasahují do běhu algoritmů. Ve výsledku se v lepším případě může stát, že takové artefakty budou považovány za podezřelou oblast. To by se dalo řešit dodatečnými kroky pro redukci chybně označených částí.

Avšak v jiné situaci může dojít k vynechání opravdového nádoru na úkor věcí, které do detekce vůbec neměly být zařazeny. Například by se v průběhu metody mohly nastavovat prahy na základě informací z histogramu nebo údajů o kontrastu v různých částech snímku. Anebo by mohl být artefakt zvýrazněn daleko více než skutečný nález. Pak by vzájemné porovnání hovořilo jasně pro označení nežádoucího objektu. Odstranění nechtěných částí mamogramu je spojeno s vymezením oblasti prsu v následujícím odstavci.

Důležitým krokem při přípravě snímků se ukázala být separace prsu. Tento krok má dvě velká pozitiva. Zaprvé plocha obrázku se zmenšuje občas o více než třetinu. U metod, které počítají s každým pixelem, se tak výrazně urychlí výpočet. Druhým přínosem zvoleného postupu je nastavení zbývajícího pozadí na hodnotu nula. Pozadí se sice může zdát čistě černé, ale opak je pravdou (viz Obrázek 9). Převládá v něm sice černá, ale hodně bodů má i jinou intenzitu a vytváří tím tedy různé kontrastní útvary, které v extrému mohou vést k detekci nálezu mimo oblast prsu.

Algoritmus pro vymezení oblasti prsu jsem převzal a upravil z [33]. Jde o velmi jednoduchý ale účinný postup. Snímek se nejdříve normalizuje. Hodnoty intenzity  $\langle 0,255 \rangle$  se převedou do intervalu  $\langle 0,1 \rangle$ . Poté obraz projde prahováním – vše, co je pod stanovenou mez, se nastaví na hodnotu 0. U drtivé většiny pixelů jde o pozadí. Tento krok způsobí sjednocení pozadí na černou barvu.

Nyní je nutné nalézt, kde se nachází prs. K tomu jsem využil metodu `findContours` z knihovny OpenCV, která najde obrysy všech objektů na snímku. Díky předchozímu prahování jsou kontury zcela jasně vymezeny. Protože se na obrázku nachází kromě prsu, které nás zajímá, i výše zmíněné artefakty (viz 2.8.5), bude detekovaných kontur více. Nám stačí vybrat tu největší podle toho, jakou plochu zabírá. Naštěstí knihovna definuje k nalezeným konturám i takové prostředky, s tímto tedy není žádný problém. Do nového obrázku s nastaveným pozadím na černou zkopírujeme pouze pixely, které patří do vyplněné kontury prsu, ostatní detekované objekty ignorujeme (viz Obrázek 10). Jako hraniční hodnotu jsem stejně jako autoři článku zvolil 0,3.



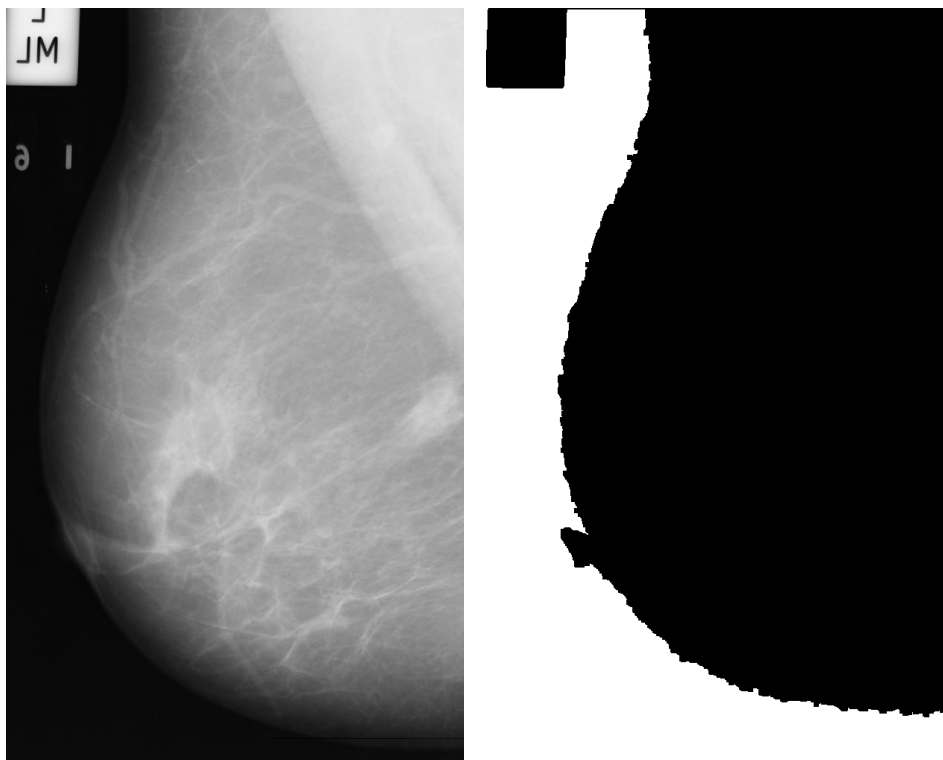
Obrázek 10: Separovaná oblast prsu ze snímků na obrázku 9

Prahování má v tomto případě jeden nežádoucí efekt. Při jeho aplikaci dochází ke ztrátě informace. Téměř vše pod normovanou hodnotu prahu bude ztraceno. Nicméně většinou se jedná o velmi tmavé body a pozadí. Je téměř vyloučeno, aby se v těchto hodnotách pohyboval nádor nebo kalcifikace.

### 5.3.2 Zero padding

Algoritmus na separaci oblasti prsu ovšem vykazoval chyby při detekci kontur. Různé objekty se do sebe spojovaly a tvořily tak jednu velkou konturu zabírající většinu snímku. Nejčastěji k takovému spojení docházelo pomocí okraje snímku nebo artefaktu u kraje. V obraze pak tyto nechtěné objekty zůstávaly (viz Obrázek 11). V krajním případě se separace vůbec nepovedla a snímek zůstal beze změny. Tento nedostatek jsem vyřešil vynulováním hraničních bodů. Snímek je tak orámovaný černou ve zvolené šířce na každé straně.

Hodnota šířky se může lišit pro jednotlivé databáze. Původně jsem podle zkušeností z databáze MIAS zvolil pouze 10ti pixelové orámování. Pro práci s DDSM databází jsem tyto hodnoty musel upravit na 50 pixelů nahoře a dole a 10 pixelů po stranách. U snímků z DDSM, které jsem testoval, mi přišlo, že obsahují daleko více artefaktů ztěžujících detekci nádorů.



Obrázek 11: Vpravo chybně detekovaná kontura (černá barva) z obrázku vlevo.

### 5.3.3 Zmenšení snímku

Pro urychlení výpočtu Iris filtru, jehož výpočetní časy jsou velmi dlouhé na velkých obrázcích, jsem se rozhodl snímky zmenšovat na úkor určité ztráty informace.

Obrazy, jejichž alespoň jeden rozměr přesáhne 4000 bodů, budou zmenšeny osmkrát, u patnácti set bodů to bude čtyřikrát a zbylé budou zmenšeny na polovinu. Všechny tyto hodnoty lze přepsat v hlavičkovém souboru `support.hpp`.

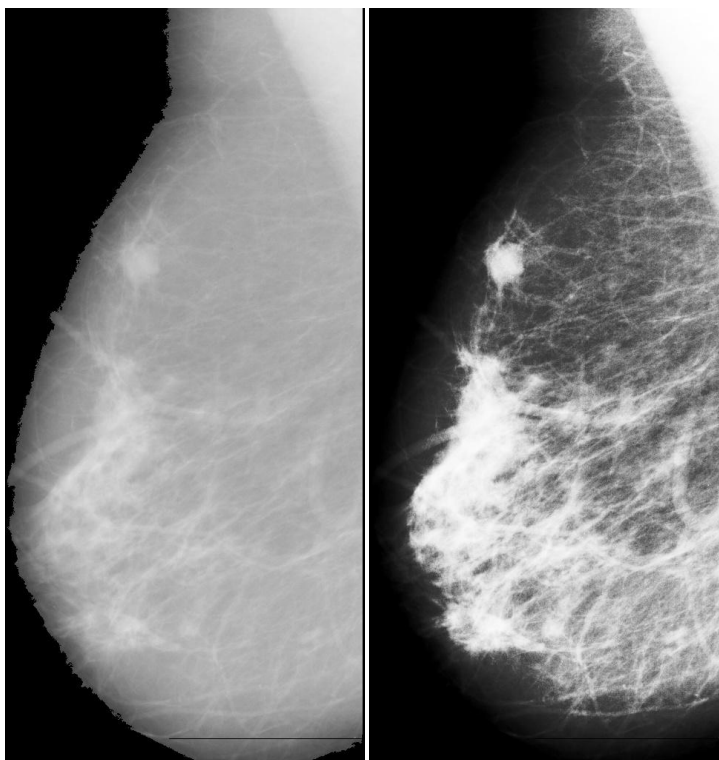
### 5.3.4 Vyrovnání histogramu (histogram equalization)

Ekvalizace (vyrovnání) histogramu je metoda, která se hojně využívá ve většině oblastech úpravy obrazu. Našla uplatnění i při zpracování medicínských snímků [77][78][79].

Jedná se o algoritmus, který je založen na práci s histogramem. Na jeho výstupu by měl být histogram rovnoměrně rozprostřený po celém intervalu. Používá se zejména ke zvýšení kontrastu. Nejlepších výsledků dosahuje, nachází-li se většina bodů v úzkém intervalu intenzit. To znamená, že je například většina bodů s hodnotou blízkou nule (hodně tmavé) a tvoří mezi sebou malý kontrast. Proto se vyrovnání histogramu dá vcelku úspěšně použít na úpravu pod nebo přeexponovaných fotografií.

Výhodou ekvalizace histogramu je fakt, že při znalosti „vyrovnávací“ funkce lze snadno změny vrátit a dostat se zpět k původnímu obrázku. Tato technika je také výpočetně nenáročná. Na druhou stranu se může stát, že dojde ke zvýraznění šumu na úkor zmenšení kontrastu a vlastně celkovému utlumení důležitých dat [44].

V této práci ekvalizaci aplikuji těsně před spuštěním algoritmů. Tedy ve fázi, kdy už mám separovanou oblast prsu a zároveň odfiltrované artefakty. Tento krok jsem zařadil, aby se ještě před spuštěním filtrů zvýšil kontrast a zvýraznila se zajímavá místa (viz Obrázek 12). U Iris filtru sice není změna tak patrná, ovšem u DWCE lze vidět jasný posun k lepšímu (viz 7.3).



Obrázek 12: Aplikace vyrovnání histogramu (vpravo) na obrázek vlevo

#### 5.4 Density-Weighted Contrast Enhancement (DWCE)

Density-Weighted Contrast Enhancement [33] – filtr sloužící ke zvýšení kontrastu v závislosti na intenzitě šedé v daném místě. Na výstupu filtru budou nejvíce zvýrazněné oblasti, které jsou nejsvětlejší a zároveň se v jejich okolí značně mění kontrast. V mnoha případech mají nádory přesně takové vlastnosti – jsou světlejší a tvoří kontrast s okolím (viz 2.8.2), proto by tento algoritmus měl být teoreticky úspěšný.

Filtr pracuje se dvěma verzemi původního snímku – obrazem hustoty (**density image**) a kontrastním obrazem (**contrast image**), které se mezi sebou násobí.

#### 5.4.1 Normalizace hodnot pixelů

Kvůli snadnější práci s různými parametry a prahy se hodnoty intenzity každého pixelu normalizují do intervalu  $\langle 0,1 \rangle$ . Potom lze se všemi snímky zacházet stejně a v průběhu algoritmu pro všechny používat totožné proměnné. Tuto techniku jsem použil už u výše zmíněné separace prsu ze snímku, abych mohl použít jeden práh pro všechny obrazy.

Pro redukci nepatřičného šumu autoři článku navrhují určit meze normalizace pomocí 5 % a 95 % kumulativní distribuční funkce hustoty intenzity. Potom nejsvětlejších 5 % z celkového počtu bodů bude mít po normalizaci hodnotu 1 a spodních 5 % nula. Tímto se zbavíme nejsvětlejších a nejtmavších bodů, které mohou na snímku značit mikrokalcifikace, prach, šum, chybu detektoru nebo porušení snímku při digitalizaci či pořizování. A navíc se mírně prohloubí rozdíly mezi jednotlivými body – zvýší se kontrast.

#### 5.4.2 Obraz hustoty (density image)

Hodnoty jednotlivých pixelů obrazu hustoty udávají informaci o intenzitě šedé barvy v daném místě. Takovou definici splňuje i původní obraz, lze tedy použít i ten. Nicméně za účelem odstranění šumu se na obraz aplikuje **dolní propust** (low-pass filter, smoothing, blurring). Ta má za úkol odstranit prudké změny intenzity, které si lze představit např. jako skupiny světlých bodů na tmavém pozadí. V nejjednodušší variantě dolní propust průměruje intenzitu zkoumaného bodu a jeho osmi sousedů.

Obraz hustoty vypadá rozostřeně nebo lze říct i uhlazeně. Vynikají na něm stejnorodé plochy a zanikají náhlé přechody, mezi jež musíme počítat i hrany, které budou na filtrovaném obraze méně výrazné (viz Obrázek 13 vpravo).

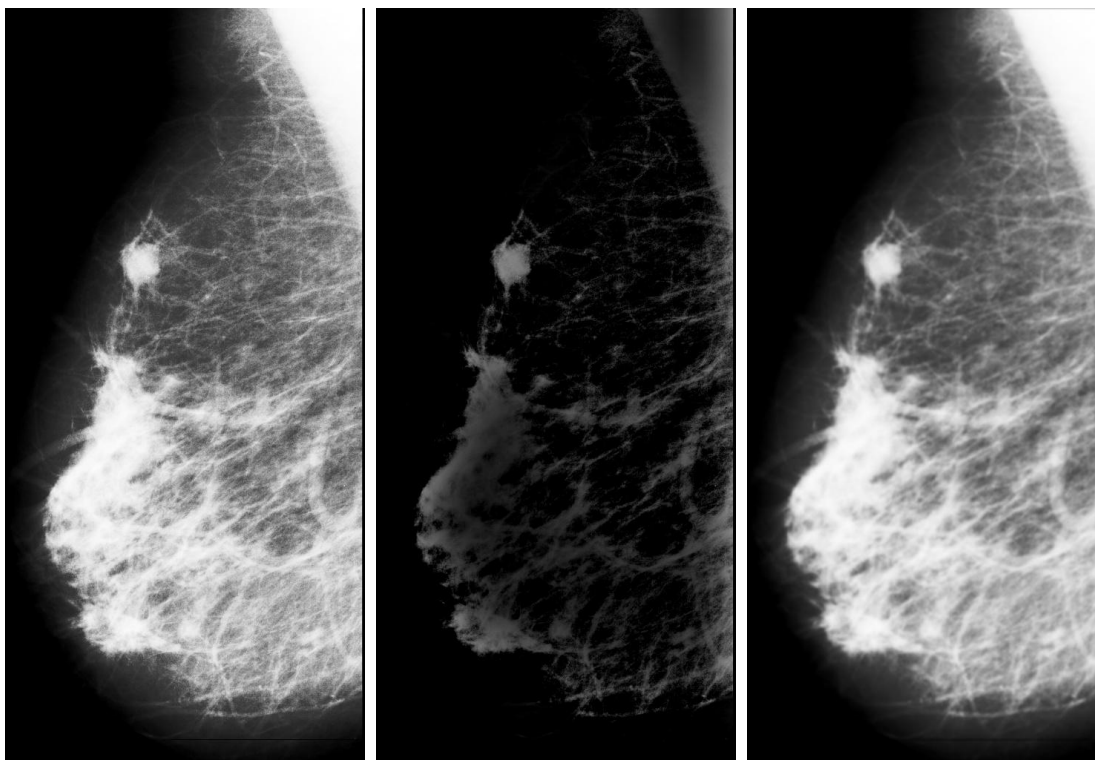
#### 5.4.3 Obraz kontrastu (contrast image)

Kontrastní obraz ukazuje, jak moc se původní snímek v daném místě mění. Hodnoty bodů udávají lokální kontrast. Nula znamená, že okolí bodu je stejnorodé. Zvětšující se hodnota říká, že v okolí je přechod z vyšší intenzity do nižší. Nejmenších hodnot tedy budou dosahovat pixely uprostřed homogenních oblastí, naopak čím více se body budou nacházet blíže hranám, tím jejich hodnota poroste (viz Obrázek 13 uprostřed).

Obraz se vytváří aplikací **horní propusti** (high-pass filter). Vlastně je to protiklad obrazu hustoty. Horní propust ponechá pouze vysoké „frekvence“ – ostré změny intenzity. Na výsledném obraze vyniknou hrany a přechody, homogenní oblasti budou tmavé. V této práci jsem zvolil postup podobný jako popisují v originálním článku. Od původního snímku odečítám jeho rozostřenou verzi.



Na ukázkových obrázcích sice nejde vidět na první pohled příliš rozdílů. Snímek vpravo reprezentuje obraz hustoty. U toho lze pozorovat, že přechody a hrany nejsou tak výrazné. Obraz působí rozmazaným dojmem. Uprostřed se nachází kontrastní obraz. Tam zase vidíme ztmavení oblasti, jež byla nejsvětlejší a stejnorodá na výchozím snímku (vlevo). Naopak zde vynikly různé vzory tkáně a žil.



Obrázek 13: Obraz hustoty (vpravo) a obraz kontrastu (uprostřed) vytvořené z originálu upraveného pomocí vyrovnání histogramu (vlevo)

#### 5.4.4 Algoritmus

Po přípravě obou verzí původního snímku už je v jednoduchosti stačí pouze vynásobit. Dosáhneme tak požadovaného výsledku, že nejsvětlejší a zároveň nejkontrastnější místa budou ve výsledku nejvýraznější.

Autoři tohoto algoritmu ovšem navrhli o něco složitější řešení, které vykazuje lepší výsledky. Postup vypadá následovně:

$$I_{dc} = K_m \cdot I_d \times I_c$$

$$I_v = K_{nl} \cdot I_{dc} \times I_{dc}$$

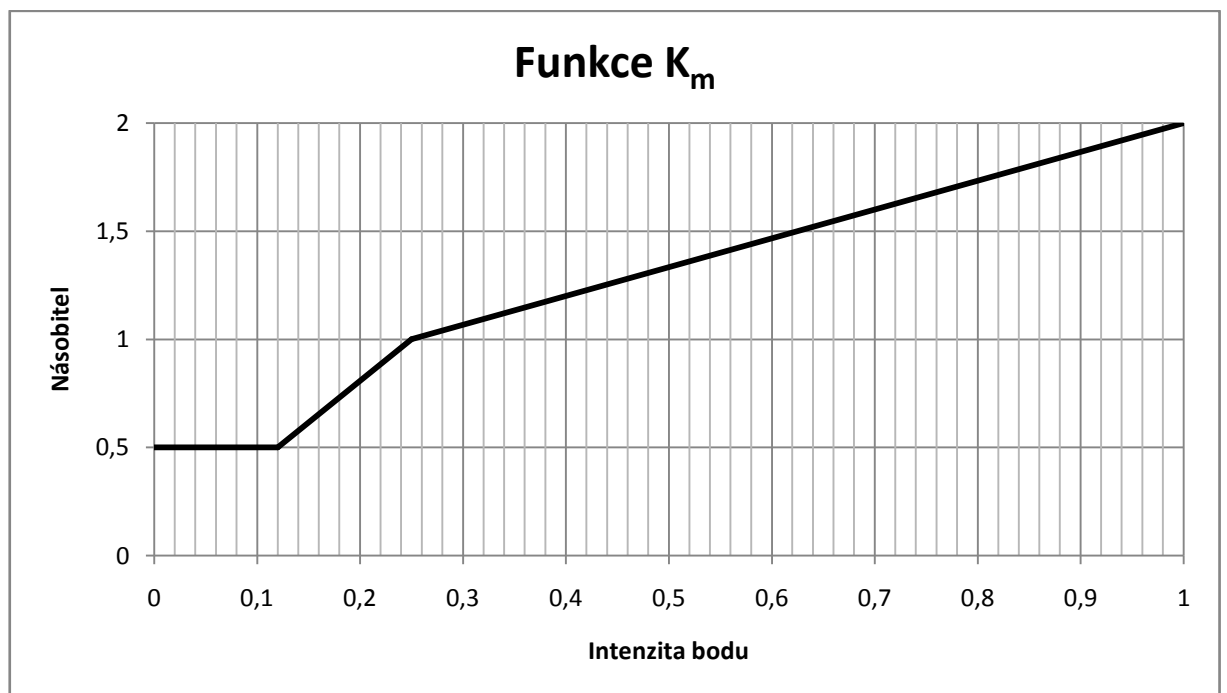
$I_d$  – obraz hustoty,  $I_c$  – obraz kontrastu,  $I_{dc}$  – obraz vzniklý násobením kontrastu a hustoty,  $I_v$  – výsledný obraz,  $K_m$  a  $K_{nl}$  nelineární funkce.

Podle původní myšlenky se obrazy násobí ještě nelineárními funkcemi  $K_m$  a  $K_{nl}$ . Můžeme si všimnout, že obraz hustoty je předně násobený funkcí  $K_m$ . Tato funkce penalizuje nízké intenzity a naopak vyšší zvýrazňuje. Následně dojde k nejdůležitějšímu násobení obraz hustoty x obraz kontrastu. Obrázky jsou sice reprezentovány jako matice, ovšem zde nejde o klasické násobení matic. V tomto algoritmu se násobí pouze pixely na odpovídajících pozicích.

Dalším krokem je násobení výsledného obrazu z předchozího kroku  $I_{dc}$  sebou samým. To ještě více zvýrazní zajímavé oblasti. Do této operace vstupuje nelineární funkce  $K_{nl}$ . Obě funkce vlastně určují násobitele v závislosti na hodnotě každého bodu v obrazu.

### Funkce $K_m$

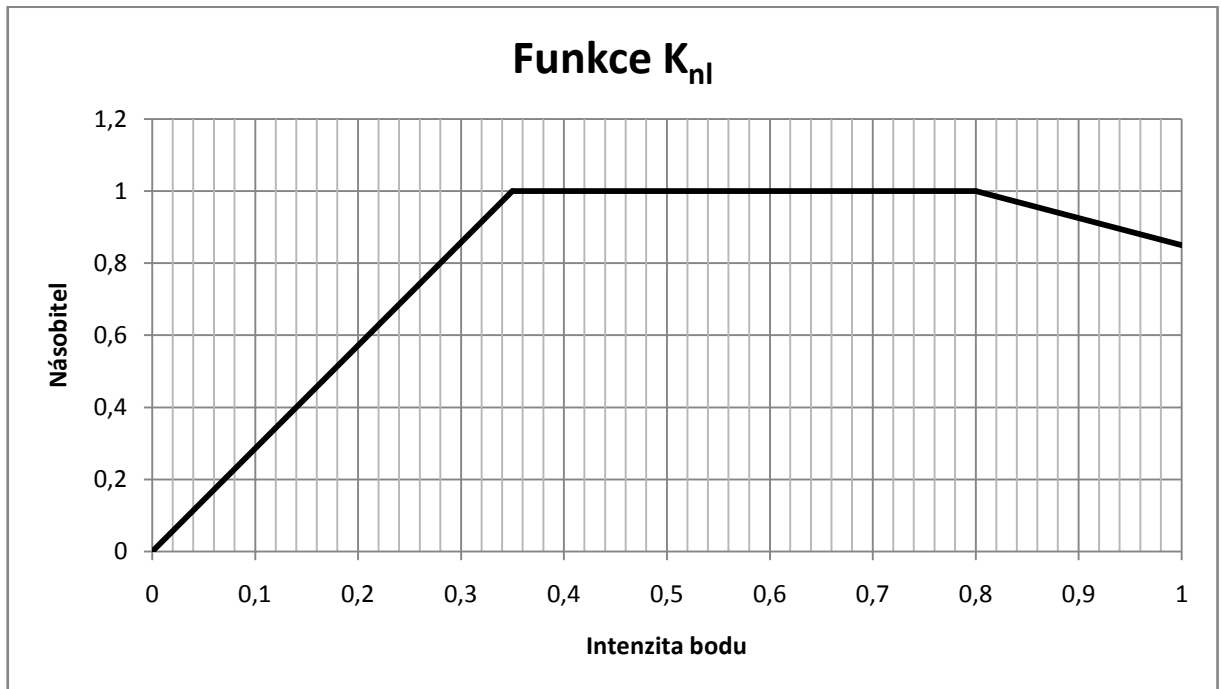
Funkce  $K_m$  (viz Graf 1) má co nejvíce penalizovat velmi tmavé hodnoty kvůli tomu, že většinou tvoří pozadí a z hlediska detekce tumorů nezajímavé oblasti. Naopak čím se daný bod blíží více k bílé, tím je násoben vyšším činitelem. Jelikož se touto funkcí násobí obraz hustoty, měly by na něm ještě více vyniknout světlé homogenní oblasti.



Graf 1: Funkce  $K_m$

## Funkce $K_{nl}$

Touto funkcí se upraví obraz vzniklý násobením hustoty a kontrastu.  $K_{nl}$  (viz Graf 2) má za úkol odfiltrout „nezajímavé“ části. Proto jsou body s nedostatečnou světlostí a kontrastem silně potlačeny. Naopak u světlejších a kontrastnějších dochází pouze k malému posunu. Za účelem řádně zformovat hrany jsou nejsvětlejší body jemně utlumeny.



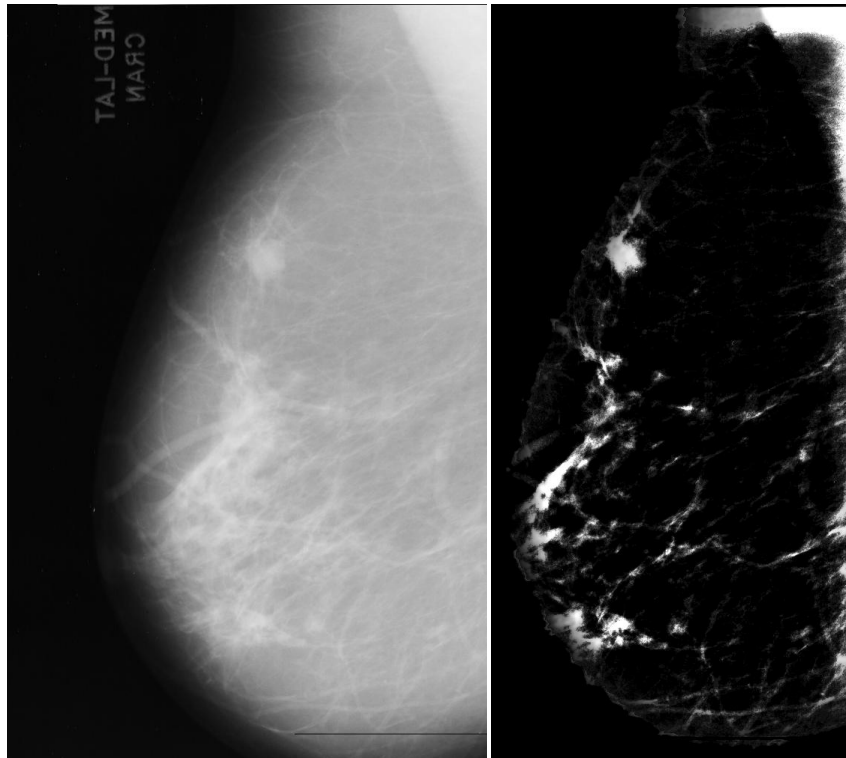
Graf 2: Funkce  $K_{nl}$

### 5.4.5 Nepřesnosti v článku a při implementaci

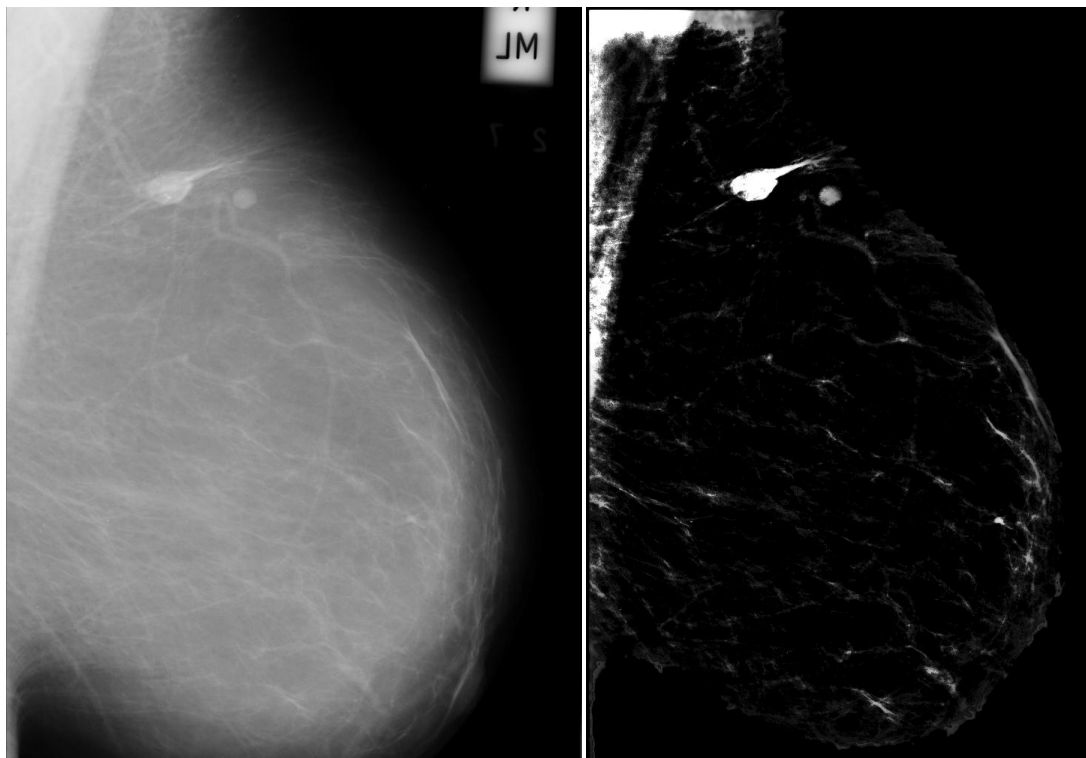
Výstup filtru je hodně závislý na realizaci obrazu kontrastu. Pro něj jsem zvolil velmi podobnou interpretaci jako autoři, tedy odečíst rozostřenou verzi od původního obrazu. Ovšem jak velké rozostření má být nebo jak ho přesně realizovat, nebylo popsáno. Můj program prošel opravdu mnoha verzemi a nakonec jsem vybral tu nejlepší. Nemůžu však tvrdit, že neexistuje lepší implementace tohoto filtru.

Článek není úplně kvalitně převeden do digitální polohy. Proto jsou v něm téměř nerozluštitelné obrázky, které by mohly napomoci zhodnotit funkčnost algoritmu. Práci, které se zabývají přímo popisem implementace DWCE filtru je sice více, ale bohužel z prostředí Západočeské univerzity nebyly dostupné [45][46].

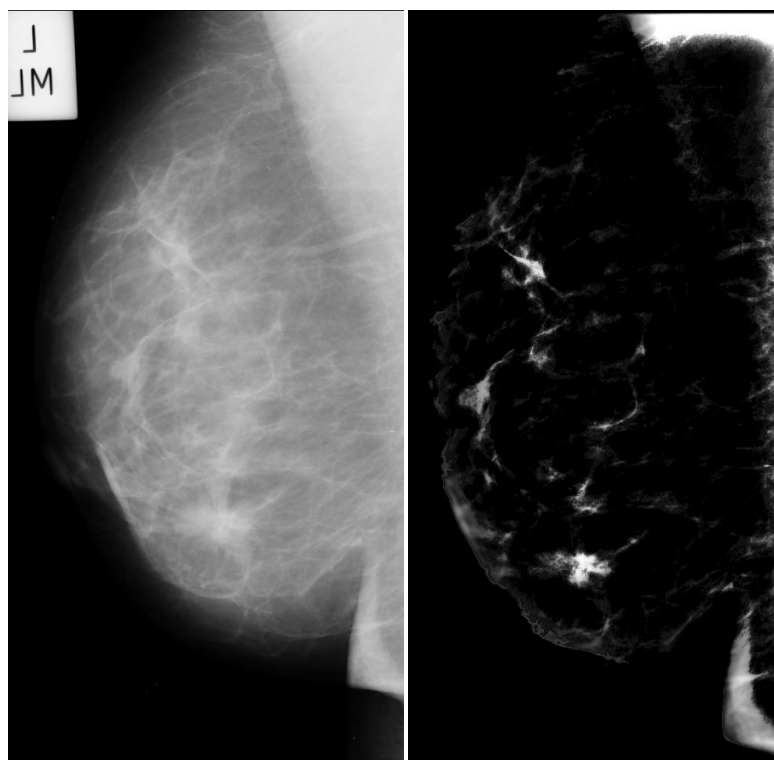
Obrázky 14-16 ukazují výstup filtru Density-Weighted Contrast Enhancement.



Obrázek 14: Výstup filtru DWCE. Originální snímek se nachází vlevo. Nejsvětlejší oblasti filtrovaných obrazů (vpravo) poukazují na podezřelá místa. Na snímcích je zachycen nádor.



Obrázek 15: Výstup filtru DWCE. Originální snímek se nachází vlevo. Nejsvětlejší oblasti filtrovaných obrazů (vpravo) poukazují na podezřelá místa. Na snímcích je zachycen nádor.



Obrázek 16: Výstup filtru DWCE. Originální snímek se nachází vlevo. Nejsvětlejší oblasti filtrovaných obrazů (vpravo) poukazují na podezřelá místa. Na snímcích je zachycen nádor.

## 5.5 Iris filter

Iris filter [34][35] slouží ke zvýraznění konvexních oblastí. Nádor se v ideálním případě jeví jako téměř homogenní oválná oblast světlejší než okolí. Vezmeme-li si mamografický snímek jako mapu, kde světlá barva značí vyšší nadmořskou výšku, pak hledaný nádor vypadá jako hora. Zabírá lokální maximum v dané oblasti nebo se k němu alespoň blíží. Gradient jako ukazatel směru největšího růstu by měl směřovat k nádoru. Při návratu k ideálnímu případu dokonce gradienty budou ukazovat do jeho středu. Na této myšlence je založený Iris filtr. Pracuje s gradientní mapou a snaží se najít místa, kam směřují gradienty z okolí.

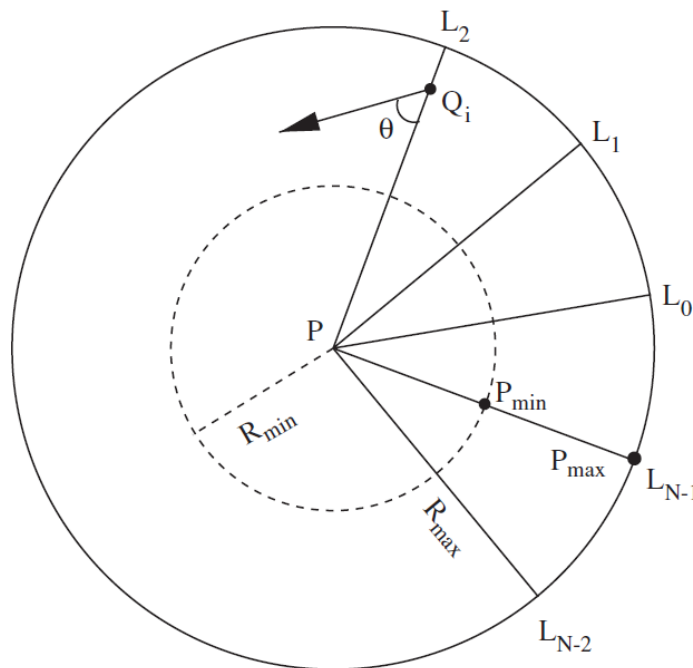
Tak to ovšem funguje pouze v ideálním případě. V praxi mají nálezy rozličné tvary a směry gradientů může ovlivňovat šum nebo místy hustší tkáň. Na rozdíl od ideálního kruhového případu se nebudou všechny gradienty sbíhat do středu a navíc se může stát, že některé dokonce budou směřovat úplně na opačnou stranu. Nelze tedy hledat přesně oblast, do které ukazují gradienty z okolí. Je zapotřebí určit nějakou míru, podle které budeme poměřovat sbíhavost gradientů k příslušnému místu – **stupeň konvergence**.

### 5.5.1 Oblast zájmu

Protože by bylo výpočetně velmi náročné zkoumat úplně všechny gradienty v okolí nějakého bodu, aproximuje se toto okolí pouze  $N$  polopřímkami. Ty jsou spuštěné ze zkoumaného bodu rovnoměrně do všech směrů. V algoritmu se vyhodnocuje stupeň konvergence pouze u bodů, které leží na těchto polopřímkách.

Aby se při hledání podezřelých oblastí zabránilo velkému vlivu malých lokálních extrémů poblíž aktuálně zkoumaného bodu, jakými jsou kalcifikace, šum nebo defekty snímku, zavádí se pojem *region of support* (oblast podpory, oblast zájmu). Ten definuje minimální a maximální vzdálenost od vyšetřovaného pixelu (viz Obrázek 17). Maximální hranice zamezuje zbytečnému zkoumání hodně vzdáleného okolí. Minimální zase eliminuje vliv malých světlých bodů.

Z obou výše uvedených odstavců plyne, že se nadále budeme zabývat pouze body na polopřímkách, jenž jsou mezi minimální a maximální stanovenou vzdáleností. Tento problém jsem ve svém programu vyřešil použitím dvou soustředných kružnic a úseček mezi nimi. Kružnice reprezentují minimální ( $R_{min}$ ) a maximální vzdálenost ( $R_{max}$ ). Oblast zájmu se tedy nachází v prostoru mezi nimi. Jestliže je  $N$  počet pomyslných polopřímek, rozdělí se obě kružnice na  $N$  stejných dílů. Vzniknou dvě kružnice s vyznačenými  $N$  body. Zkoumané pixely vyjdou ze spojení příslušných bodů v úsečky.



Obrázek 17: Znázornění důležitých termínů pro Iris filtr.  $R_{min}$  a  $R_{max}$  reprezentují minimální a maximální poloměr kružnic.  $P$  je zkoumaný pixel,  $L$  polopřímky,  $Q_i$  je právě vypočítávaný bod na polopřímce a  $\theta$  úhel mezi polopřímkou a gradientem ( $g$ ) [34].

### 5.5.2 Stupeň konvergence

Pro každou z  $N$  polopřímek určíme její stupeň konvergence podle vzorce:

$$C_i = \frac{\sum_{Q=P_{min}}^{Q_i} f(Q)}{P_{min} Q_i}$$

Kde

$$f(Q) = \begin{cases} 0 & |g| = 0 \\ \cos \theta & |g| \neq 0 \end{cases}$$

Procházíme pixely na polopřímkách náležících do oblasti zájmu a u každého zjistíme, jak moc je gradient ( $g$ ) v daném bodě  $Q_i$  odkloněn od polopřímky. Jestliže výsledek bude roven 1, gradient ukazuje přesně na zkoumaný pixel a má tedy maximální stupeň konvergence. Pokud gradient bude směřovat kolmo od polopřímky, výsledek  $f(Q)$  je rovný nule a s ním i stupeň konvergence ve zkoumaném bodě.

Jak lze vidět ze vzorce pro výpočet  $C_i$ , při výpočtu nedosazujeme pouze konvergenci v jednom bodě zkoumané polopřímky. Ale vlastně průměr všech bodů ležící mezi bodem  $P_i$  a  $Q_i$  spadajících do oblasti zájmu.

Výsledná hodnota pro danou polopřímku je maximální dosažené  $C_i$ . Takto se výsledek aproximuje pro danou polopřímku. Jedna extrémní hodnota výrazně neovlivní celkový stupeň konvergence.

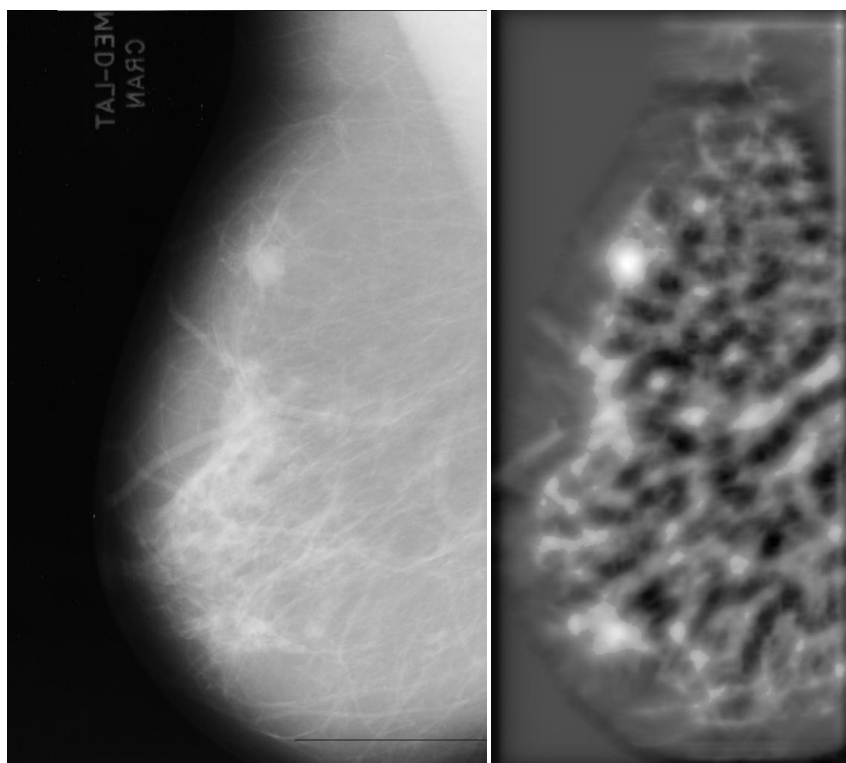
Pro zkoumaný pixel pak určíme jeho stupeň ( $C_p$ ) jako průměr  $C_i$  ze všech polopřímek. Hodnoty  $C_p$  náležejí do intervalu  $(-1,1)$ . Hodnotu blízko jedné na výstupu bude mít střed konvexní oblasti, která je světlejší než okolí.

### 5.5.3 Algoritmus

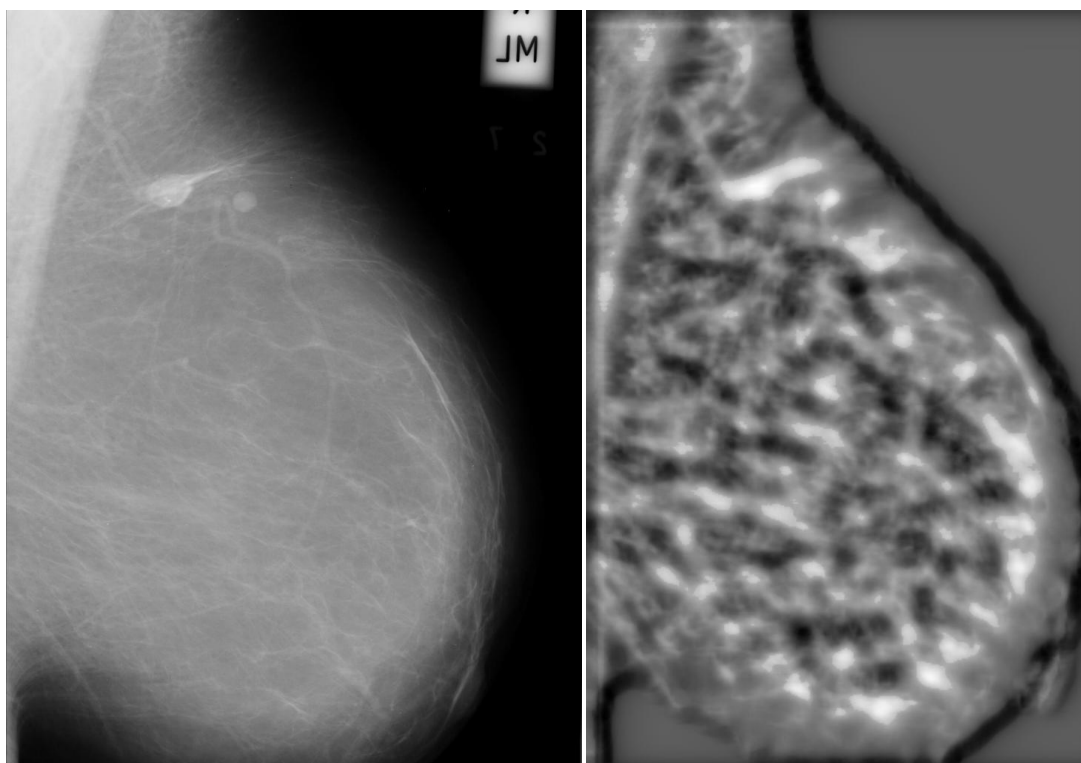
Iris filtr prochází každý pixel na snímku. U všech určí jejich stupeň konvergence a ukládá hodnoty do výsledného obrazu. Po převedení ze stupňů konvergence z intervalu  $(-1,1)$  do odstínů šedé získáváme výstup Iris filtru. Ten vlastně představuje pravděpodobnostní mapu. V místech, kde je světlejší, je větší pravděpodobnost, že se zde nachází nádor (viz Obrázek 18-20).

Kvůli zvýraznění výsledků pro další kroky ještě obraz projde lehkou úpravou. Pět procent bodů, které pokrývají nejsvětlejší oblasti snímku, budou vynásobeny konstantní hodnotou. Tím se ještě více zviditelní nejpodezřelější oblasti.

Na základě informací poskytnutých ve zdrojovém článku jsem zvolil hodnotu  $R_{min} = 2,5 \text{ mm}$ ,  $R_{max} = 17,5 \text{ mm}$ . Počet polopřímek jsem stanovil na 35. Tyto hodnoty lze ovšem velmi lehce pozměnit v konfiguračním souboru.

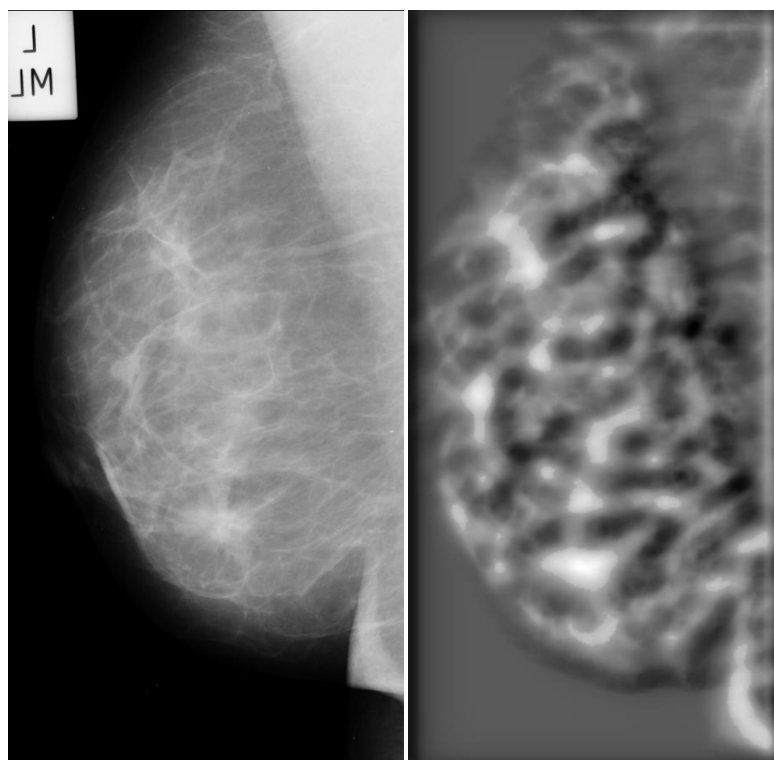


Obrázek 18: Výstup Iris filtru pro stejnou sadu snímků jako u DWCE (Obrázek 14-16)



Obrázek 19: Výstup Iris filtru pro stejnou sadu snímků jako u DWCE (Obrázek 14-16)





Obrázek 20: Výstup Iris filtru pro stejnou sadu snímků jako u DWCE (Obrázek 14-16)

#### 5.5.4 Paralelizace výpočtu

Prvotní implementace se vyznačovala časem výpočtu v řádu minut. Algoritmus totiž prochází každý bod obrazu, u něho musí spočítat stupeň konvergence pro každou polopřímku. Proto však musí projít všechny její body, které leží v oblasti podpory.

Všechno dohromady dává miliony výpočtů na snímek u MIAS databáze se snímky v rozlišení 1024 x 1024 pixelů. Např. u DDSM databáze jsou snímky i pětkrát větší. První optimalizací tedy bylo zmenšení obrazu (viz 5.3.3). To sice zrychlilo výpočet, ale na úkor kvality výsledného obrazu. Proto nelze zmenšovat snímek příliš, aby se stále zachovala důležitá data.

Jednotlivé body při výpočtu nejsou na sobě nijak závislé. Počítá se sice s jejich okolím, ale kvůli ukládání výsledků do odděleného obrazu, se okolí v průběhu nemění. Lze to tedy považovat za ideální případ pro použití výpočtu ve více vláknech. Ta budou moci pracovat každé na svém bodě a nemusejí se vůbec synchronizovat. Pro paralelizaci jsem vybral knihovnu *Intel Threading Building Blocks* (viz 5.1.2). Konkrétně jsem použil pouze konstrukci `parallel_for`. Vše ostatní již zajistila knihovna sama.

## 5.6 Nalezení ROI

Po provedení jednoho z filtrů máme zatím ve výsledku pouze jakýsi pravděpodobnostní obrázek. Už ten sám o sobě poukazuje na zajímavé oblasti na původním snímku a dal by se použít jako vodítko, kterým oblastem věnovat zvýšenou péči. My ale potřebujeme tyto oblasti nějakým způsobem označit a pokud možno pouze jen ty nejpodezřelější. Nejdřív zde popíši, jak byla tato extrakce ROI navrhována v obou článcích a poté, jak jsem jí realizoval já.

### 5.6.1 DWCE

Článků o DWCE se na internetu nachází více a v každém je uváděna trošku jiná metoda segmentace ROI [45][46]. Moje práce se zakládá na prvním návrhu tohoto algoritmu [33], protože k tomuto jedinému jsem získal přístup.

V původním návrhu byl použit detektor hran ihned po aplikaci DWCE filtru. Jako detektor byl zvolen *Laplacian-Gaussian(LG)*, který je složený ze dvou částí. První je rozostření snímku už dříve použité k snížení šumu na obraze nebo například k vytvoření obrazu hustoty u DWCE. Druhou část zastává detektor hran využívající Laplaceův operátor. Ten pracuje na základě vyhledávání nul, resp. na vyhledávání míst, kde operátor mění znaménka (zero-crossing edge detectors).

Po nalezení hran LG detektorem se najdou uzavřené oblasti. S těmi se bude pracovat dále, ostatní hrany jsou ignorovány. Tímto krokem už zbyly pouze zajímavé oblasti. Nicméně následuje ještě poslední krok této fáze algoritmu. Poněvadž po tomto kroku zůstává i spousta chybně označené normální tkáň. Občas se stane, že se sousední ROI slijí do sebe. Jejich rozdělení má na starosti krok nazvaný object-splitting (binary – splitting, rozdělování objektů). Na nalezených objektech se hledají úzká místa. Poté se zkoumají různé morfologické vlastnosti daného objektu a taky objektů, které by vznikly jeho rozdělením v nalezeném úzkém místě.

### 5.6.2 Iris filtr

Tvůrci Iris filtru pojali segmentaci ROI trošku jinak. Filtrovaný snímek převedli na binární (jen bílá a černá). Nastavili rozhraní mezi černou a bílou jako 93 % percentil. Bude tedy 93 % bodů černých a zbylých 7 % bílých představujících nadále zajímavé oblasti.

Následují morfologické operace, abychom se zbavili malých roztroušených bodů a řádně stanovili tvar potenciálních nádorů. První se aplikuje eroze za účelem odstranit osamocené body nebo jejich shluky. Následuje morfologické uzavření (closing). To zajistí jasně definované kontury podezřelých oblastí.

Všechny oblasti, jež nebyly vyloučeny během uplynulých kroků, budou podrobeny dalším testům pro snížení počtu falešně označených (false positives). Prvním z nich je test velikosti. Ten se zavádí pro odfiltrování nejmenších částí, které sice nebyly eliminovány předchozími kroky, ale jsou příliš malé, než aby mohly značit nádor. Na některých snímcích jich totiž může být opravdu velké množství a poněvadž by každá z nich značila zvýšení počtu špatně označených oblastí, úplně by znehodnotily celý algoritmus.

V posledním kroku se zbylé ROI podrobí testu excentricity. Ta se stanoví jako poměr nejdelší osy zkoumaného objektu a poloměr kružnice se stejným obsahem. Určuje, jak moc je objekt oválný, resp. jestli není příliš „rozplácly“ (roztáhlý na nějakou stranu).

Protože byly často zbylé regiony zbytečně rozdělené, nastává zde opačná situace než u DWCE. Zde se naopak podobné oblasti spojují. Jestliže se oblasti dotýkají nebo překrývají, zkoumá se dále jejich průměrná intenzita šedé v původním i filtrovaném obrázku. Když se průměry příliš neliší, oblasti budou spojené do jedné velké. ROI, které prošly všemi kroky, můžeme nyní považovat za skutečné kandidáty na nádor.

### 5.6.3 Moje implementace

#### První pokus s hranovým detektorem

Nejdříve jsem implementoval DWCE filtr. V tu dobu jsem ještě neměl vybranou druhou metodu. Proto jsem segmentaci převzal z něj. Pokusil jsem se co nejpřesněji reprodukovat, co je v článku. Snímek filtrovaný DWCE jsem mírně rozostřil pomocí *Gaussian-smoothing* a následně jsem použil hranový detektor. Zde jsem na rozdíl od autorů původního článku zvolil *Canny edge detector*, protože je již implementovaný v OpenCV a navíc má mít některé lepší vlastnosti než Laplaceův operátor.

Na detekci uzavřených oblastí jsem použil metodu `findContours`. Ta nejen najde kontury objektů, ale navíc pro ně poskytuje ještě další funkce. Výstupem této první verze bylo obrovské množství ROI. Částečným zlepšením sice byla filtrace příliš malých výřezů, stejně jako popisují u filtru Iris, ale stále bylo vybraných oblastí příliš mnoho.

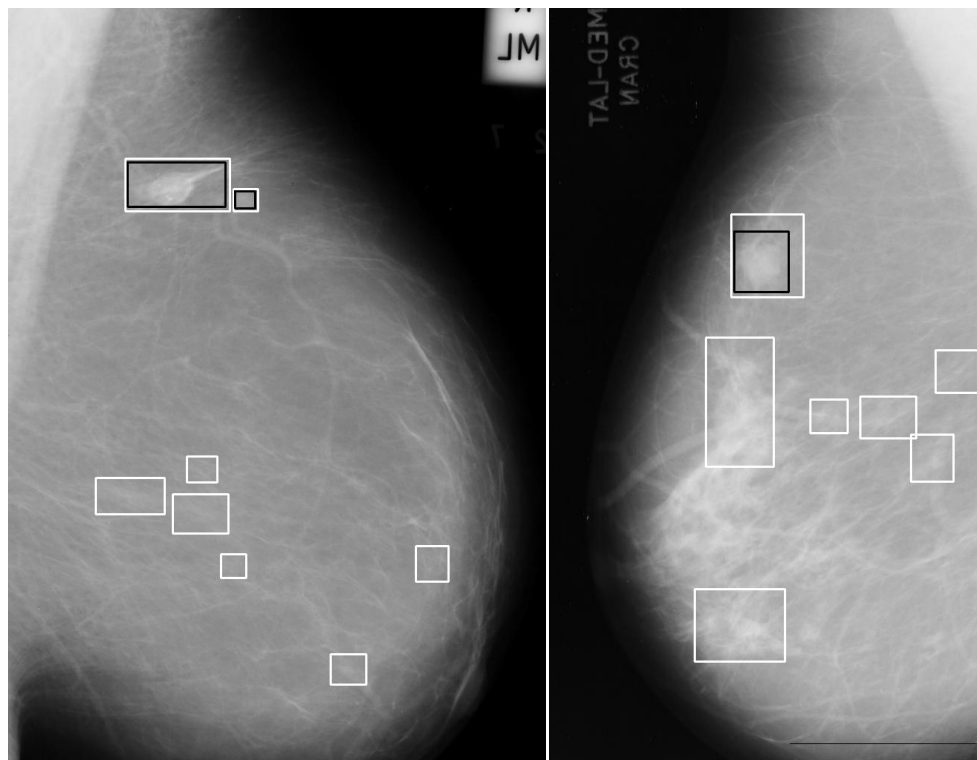
#### Přidání morfologických operací

Na základě studia dalších informací o použití morfologických operací při zpracování mamogramů, jsem je zkusil nasadit před detekci hran. Aplikace dilatace, která ztlačuje kontury, se ukázala jako dobrá volba. Dokonce dosahovala lepších výsledků, když bylo provedeno několik iterací. Nicméně s počtem iterací rostla i ztráta informace. Takže ve výsledku můj program detekoval naprosto přesně jen některé nálezy. Ostatní a zvláště menší se během morfologických operací ztratily. Potřeboval

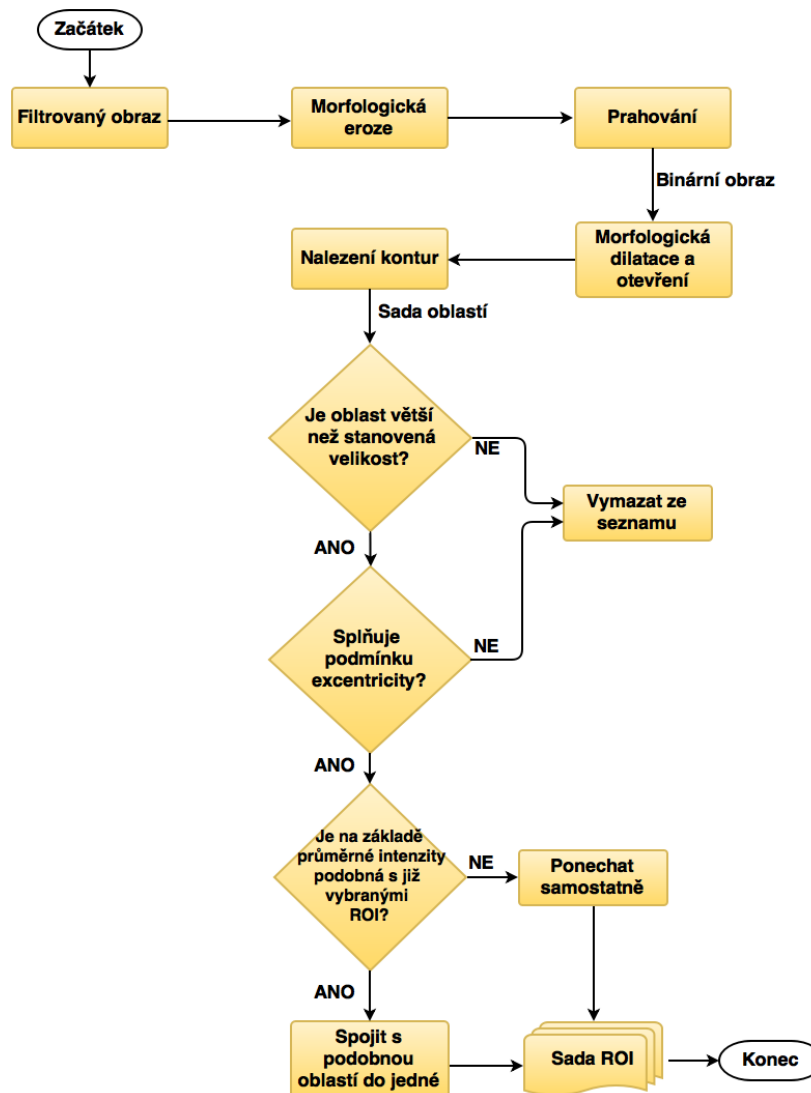
jsem detekci menších a zároveň přesnost u větších. Zkusil jsem tedy oba přístupy zkombinovat. Nejdříve jsem na původní obrázek použil morfologické operace v jedné iteraci a detekoval hrany. Poté jsem aplikoval další iterace. Znovu jsem detekoval hrany, ale tentokrát jsem je zapsal do obrazu s již nalezenými z předchozího kroku. To vedlo k doplnění necelých hran a jejich posílení při zachování menších útvarů. Ale ani tento výsledek nebyl uspokojivý, vedle velkého počtu chybně označených zdravých oblastí, bylo i několik nenalezených oblastí s nádorem (false negatives).

### Aktuální verze

S vědomím, že morfologické operace jsou primárně určené na binární obrazy, jsem zaměnil detekci hran za prahování. Vlastně jsem si zvolil stejnou techniku jako používali tvůrci Iris filtru. Určím hranici, která hodnoty podle intenzity rozdělí na bílé a černé (jedničky a nuly). Po tomto prahování použiji již připravené morfologické operace a následně už jen detekuji kontury. Postup zůstal zachovaný, jen je místo hranového detektoru provedeno prahování. Později jsem nahradil jednotný práh pro všechny percentilem hustoty distribuční funkce. To jsem převzal z návrhu filtru Iris. Odtud jsem se také inspiroval pro testování excentricity a pozdější spojování podobných ROI. Celý postup viz Obrázek 22, výstup z detekce ukazuje Obrázek 21.



Obrázek 21: Výstup z detekce ROI - bílé jsou označena všechna detekovaná podezřelá místa a černě skutečný nádor (označený podle anotací).



Obrázek 22: Diagram pro detekci ROI

## 5.7 Hodnocení úspěšnosti algoritmů

Bylo potřeba vyhodnocovat, jestli je nový směr vývoje správný. Databáze MIAS má k dispozici u snímků zasažených rakovinou anotace určující přesnou polohu a velikost nálezu.

Z různých prací jsem se dozvěděl, že pro ověřování přesnosti se používají jednoduché prostředky. Pokud jde o správnost, stačí pokrýt určitou část označeného nádoru a je to považováno za úspěšný nález. To se ukázalo jako velice nešťastné vzhledem k tomu, že se považuje za úspěch i situace, kdy je označený celý obrázek. Tak se totiž pokryje 100 % nádoru. Proto jsem musel zavést omezení velikosti. Určitě nemohla být určena jedna mez pro všechny, protože nálezy jsou v tomto ohledu velmi variabilní. Limit se musí odvíjet od udané anotace. Provedl jsem několik pokusů, kde jsem určil násobek skutečné velikosti a snímek od snímku jsem zkoumal, jak byla vyhodnocena úspěšnost. Na základě této vizuální kontroly jsem stanovil, že aby ROI úspěšně označovala nádor, musí pokrývat stanovená procenta plochy a zároveň nesmí

být více než 6 krát větší než skutečný tumor. Tato hodnota se v programu dá změnit, nicméně veškerý další postup je založen právě na ní. Jestliže je reálný nádor velký, může být o hodně větší i označená oblast, to je nevýhodou tohoto přístupu.

Počet označených oblastí, kde není nález, se také musí vzít v potaz. Jako chybně označené počítám výřezy, které pokrývají skutečný nádor z maximálně pěti procent nebo vůbec. ROI, u kterých je procento pokrytí mezi úspěšným a špatným označením, nejsou počítány ani k jedné položce.

Pro porovnání různých variant jsem zavedl jednotnou míru. Tu ovlivňuje jak úspěšnost, tak počet chybně označených částí. Protože se většinou testuje na sadě snímků, počet falešně označených oblastí je zprůměrován. Míru určuje následující vzorec:

$$m = \alpha (1 - \text{úspěšnost}) + \beta \cdot fp$$

- $(1 - \text{úspěšnost})$  – udává, kolik procent nádorů algoritmus nedetekoval správně
- $fp$  – průměrný počet false positive na obrázek.
- $\alpha$  – kvantifikátor neúspěchu – určuje, jak moc ovlivní výsledek nenalezený nádor
- $\beta$  – kvantifikátor  $fp$  – udává důležitost false positives
- $m$  – výsledná míra sloužící k porovnání dvou implementací detekce. Čím je  $m$  menší, tím je algoritmus lepší.

## 5.8 Optimalizace parametrů

V rámci vývoje metody detekce ROI jsem postupně nastavoval různé parametry, které mohou ovlivnit výsledek.

- práh pro vytvoření binárního obrazu
- počet iterací jednotlivých morfologických operací
- požadovanou excentricitu
- míru podobnosti, kdy se více oblastí může spojit do jedné.

Zkoušet zadávat tyto parametry ručně by pokaždé vedlo k jinému výstupu, ale velmi těžko bychom se dopracovali k nějakému postupu vedoucímu k lepším výsledkům. Pro tento účel jsem naprogramoval evoluční algoritmus mající za úkol všechny parametry optimalizovat. Vybral jsem si diferenciální evoluci.

### 5.8.1 Diferenciální evoluce

Diferenciální evoluce patří do skupiny evolučních algoritmů. Ty se řadí mezi globálně optimalizační algoritmy. Cílem je prohledat prostor a najít optimální řešení – tedy takové řešení, které maximalizuje nebo minimalizuje známou funkci. Jsou vhodné pro řešení obecně náročných problémů, u kterých je velké množství možností [47].

#### Populace a jedinci

U evolučních algoritmů definujeme pojem populace. Populace se skládá z jedinců a jde vlastně o skupinu několika možných řešení. Jedinec je pak jedno konkrétní. Populace se během času vyvíjí a podle Darwinovy teorie v ní přežívají pouze ti nejsilnější.

#### Fitness funkce

„Sílu“ jedince v rámci populace určuje tzv. fitness funkce. Fitness funkce je vyhodnocení určitého kritéria, u kterého hledáme maximum nebo minimum. Jde vlastně o dosažení informace z jedince do původního problému řešeného pomocí evolučního algoritmu.

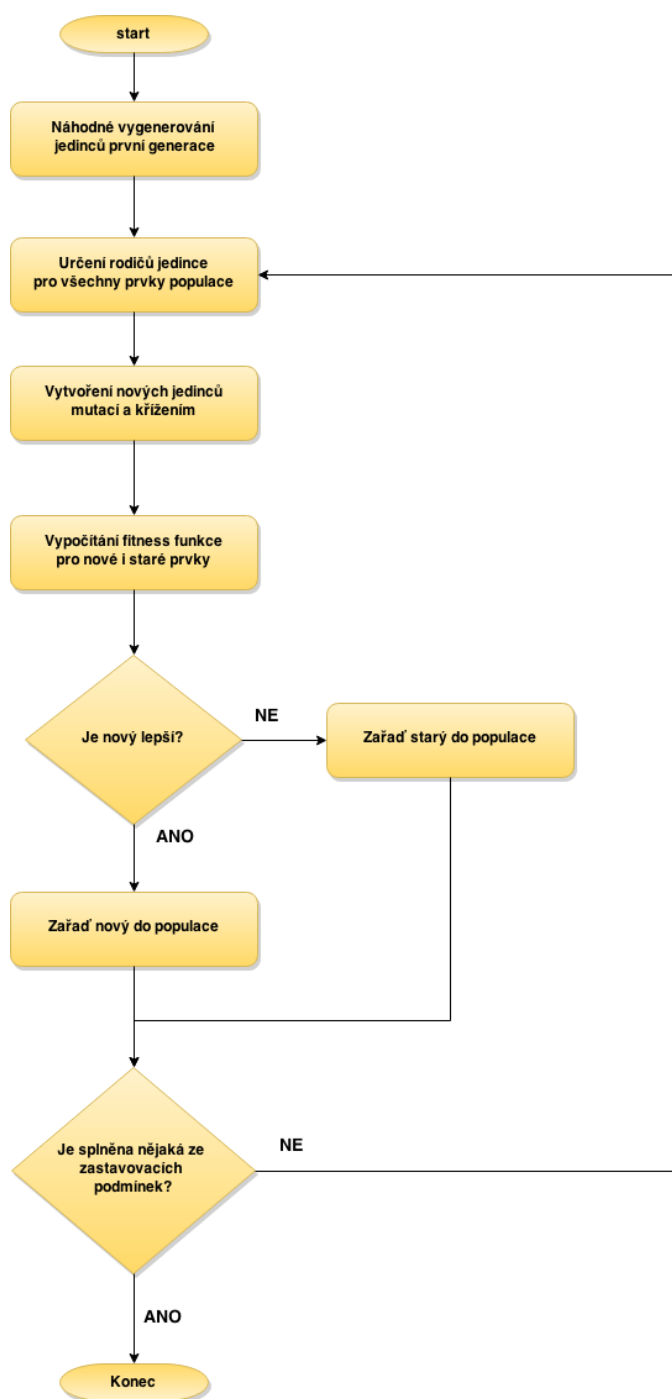
#### Evoluce

Vývoj generace způsobuje křížení vlastností jedinců. Pro křížení se definuje konstanta. Na začátku každé generace se ke všem prvkům vytvoří nový (mutovaný) jedinec. Ten je složený kombinací několika náhodně zvolených prvků dané generace (rodičů). Pro každý kus genetického kódu (zkoumané vlastnosti) vygenerujeme náhodné číslo. Je-li toto číslo větší než křížící konstanta, ponecháme prvek mutovanou vlastnost, a je-li menší, použijeme vlastnost starého prvku.

Pro nového i starého jedince vyhodnotíme fitness funkci. Do populace zařadíme prvek s lepším výsledkem. Toto zaručuje, že se postupně přibližujeme lepšímu výsledku, poněvadž vždy volíme lepší prvek (= lepší aproximaci problému).

#### Zastavovací podmínky

Pro zastavení vývoje evolucí se používá několik podmínek – hodnota fitness funkce nejlepšího prvku je lepší než zadaná hranice, hodnota fitness funkce nejlepšího prvku už se několik generací nikam neposunuje nebo konverguje příliš pomalu, anebo se spočítalo určité množství generací. Diagram, jak probíhá diferenciální evoluce viz Obrázek 23.



Obrázek 23: Diagram průběhu diferenciální evoluce

## Implementace

Při nasazení na problém optimalizace parametrů pro detekci podezřelých oblastí jedince v populaci reprezentují právě hledané parametry. Člen generace tedy bude složen z prahové hodnoty, počtu iterací morfologických operací atp. Generaci tvoří 15 jedinců. Počet je stanoven jako kompromis mezi příliš dlouhým časem výpočtu jedné generace (moc jedinců) a pomalou konvergencí (málo jedinců). Jako fitness funkce je použita míra definovaná v sekci 5.7. Během evoluce tuto míru budeme chtít minimalizovat. Jako zastavovací podmínka se využívá počet iterací.



Pro urychlení výpočtu jsem opět použil knihovnu na paralelizaci TBB. Nyní se jedinci v generaci počítají paralelně. Zde ovšem je nutná synchronizace při ukládání nejlepšího výsledku. Tu zajišťuji mutexem při kontrole a přepisování hodnoty nejsilnějšího jedince.

U tohoto algoritmu není zajištěno, že dojde ke globálnímu minimu. Přibližovat se k němu bude, avšak může se stát, že při špatně vygenerované první generaci, bude konvergovat například příliš pomalu, bude oscilovat okolo globálního minima anebo se zastaví u určitého lokálního minima.

## 6 Vlastní implementace

Kořenový adresář je rozdělený na několik podsložek. Ve složce `src` se nachází zdrojové soubory pro jednotlivé dílčí filtry a také pro hlavní modul, který je integruje do funkčního celku. `Headers` je složka obsahující hlavičkové soubory s definicemi jednotlivých funkcí v každém modulu. V kořenové složce jsou navíc přítomny prostředky pro překlad a konfigurační soubory nezbytné pro funkci programu.

Program je rozdělený do 4 částí: `DWCE`, `Iris`, `Support` a `ImageProcessing`. Zde jen ve zkratce popíši, co ve kterém zdrojovém kódu lze najít a kde to hledat v textu.

### 6.1 DWCE

V souborech `DWCE.cpp` a `DWCE.hpp` jsou umístěny prostředky, které jsou specifické pro DWCE filtr. Vedle samotného algoritmu se zde nachází metody `densityImage` a `contrastImage`. Ty vytváří obraz hustoty a kontrastu k zadanému snímku.

Další, co je potřeba pro běh DWCE filtru, jsou funkce  $K_{nl}$  a  $K_m$ , kterými se násobí různé obrazy. Ve funkcích s příslušnými názvy se všechny hodnoty v matici reprezentující snímek vynásobí příslušnými činiteli, jenž jsou do programu zaneseny natvrdo. Jejich změna je tedy možná pouze zasáhnutím do kódu.

Důležitou částí je i metoda `normalizeMat` implementovaná podle části 5.4.1.

### 6.2 Iris

Stejně jako u předchozího jsou zdrojové soubory pojmenovány `Iris.cpp` a `Iris.hpp`. V nich je programový kód k stěžejní části tohoto filtru. Metoda `performIris` dává vše dohromady. Nejdříve se vytvoří gradientní mapa použitím *Scharr* operátoru. Následně se prochází matice obrazu bod po bodu a pro každý se vypočítá konvergence.

Pro výpočet konvergence u jednotlivých bodů slouží funkce `getIrisFilterValue`. Do ní je potřeba předat hodnotu, jakou míru představují pixely. V případě MIAS databáze je to například velikost 200  $\mu\text{m}$  na pixel. Tento parametr je povinný, poněvadž se z něj dopočítají hodnoty pro minimální a maximální vzdálenost oblasti podpory (viz 5.5.1).

## 6.3 Support

V tomto modulu se nachází funkce a proměnné společné pro oba filtry (viz Tabulka 1).

Název	Funkce	Odkaz v textu
<b>resize_image</b>	slouží k zmenšení obrazu z důvodu zkrácení výpočetního času	5.3.3
<b>separateBreastRegion</b>	odfiltruje z obrazu vše ostatní kromě části, kde se nachází prs	5.3.1
<b>setPadding</b>	nastaví okraje na černou pro vylepšení separace prsu	5.3.2
<b>findROI</b>	detekuje podezřelé oblasti na snímku	5.6
<b>evaluate</b>	vyhodnotí úspěšnost algoritmu	5.7
<b>getThreshold</b>	spočítá histogram a z něj hodnotu pro zadaný percentil hustoty	
<b>optimizeParams</b>	diferenciální evoluce pro optimalizaci parametrů detekce	5.8

Tabulka 1: Důležité metody z modulu `Support.cpp`

V hlavičkovém souboru lze změnit některá nastavení programu (viz Tabulka 2). Jsou to konstanty, které jsem stanovil na základě literatury a provedených experimentů.

Konstanta	Význam	Výchozí hodnota
<b>GEN</b>	Počet jedinců v generaci	15
<b>PARAMS</b>	Počet parametrů v jedinci	6
<b>MAX_SIZE</b>	Minimální velikost pro <code>MAX_FACTOR</code>	4000
<b>MID_SIZE</b>	Minimální velikost pro <code>MID_FACTOR</code>	1500
<b>MAX_FACTOR</b>	Maximální dělitel pro zmenšení obrazu	8
<b>MID_FACTOR</b>	Střední dělitel pro zmenšení obrazu	4
<b>MIN_FACTOR</b>	Minimální dělitel pro zmenšení obrazu	2
<b>SEPARATE_THRES</b>	Prahová hodnota pro separaci region prsu	0,3
<b>CROSS_THRESHOLD</b>	Křížící konstanta pro optimalizaci	0,5
<b>DIFF_CONSTANT</b>	Konstanta pro vytvoření mutovaného jedince	0,65

Tabulka 2: Konstanty, které lze měnit ve hlavičkovém souboru `Support.hpp`

## 6.4 ImageProcessing

Tato část je vstupním bodem celého programu. Integruje všechny části. Obsahuje metodu `main`. Zpracovává parametry příkazové řádky, na jejichž základě volí, jaký modul bude použit.

## **6.5 Závislosti**

Program byl sestaven za pomoci několika knihoven. Největším podílem se o to zasloužila knihovna OpenCV pro operace s obrázky ve verzi 3.0.0. Pro paralelizaci jsem použil knihovnu Intel Threading Building Blocks v. 4.4. Práci se soubory nezávislou na platformě, generování seedu pro generátor náhodných čísel a zpracování parametrů příkazové řádky zajistila knihovna Boost v 1.55. Pro správný překlad a další vývoj, je třeba tyto knihovny mít správně nakonfigurované. Překlad je možný buď v nástrojích, které mají překladač pro C++ nebo připraveným skriptem pro Cmake verze alespoň 3.1.

## **6.6 Omezení**

Vzhledem k výzkumnému charakteru vyvíjené aplikace nejsou ošetřeny všechny vstupy a případné výjimky. Předpokládá se účast znalého uživatele.

## 7 Dosažené výsledky

Abych mohl nějakým způsobem ohodnotit úspěšnost své implementace jednotlivých filtrů stanovil jsem míru (viz 5.7). Ta se skládá ze dvou důležitých měřitelných parametrů – úspěšnosti/neúspěšnosti detekce nádoru a počtu chybně detekovaných oblastí na obrázek. Tyto dva údaje budou společně s celkovým ukazatelem předmětem zkoumání a porovnávání v této kapitole.

Testování jsem prováděl na databázi MIAS. Ta obsahuje 122 snímků, na nichž byla radiology identifikována rakovina. V této sadě jsou různé známky rakoviny od kalcifikací, přes architectural distortion až po nádory. Protože jsem se na začátku práce rozhodl, že moje zaměření bude detekce nádorů, musel jsem vybraný soubor ještě zúžit. Potom mi zbylo 53 snímků se zhoubnými a nezhojnými nádory označenými v MIAS terminologii jako CIRC, SPIC, MISC.

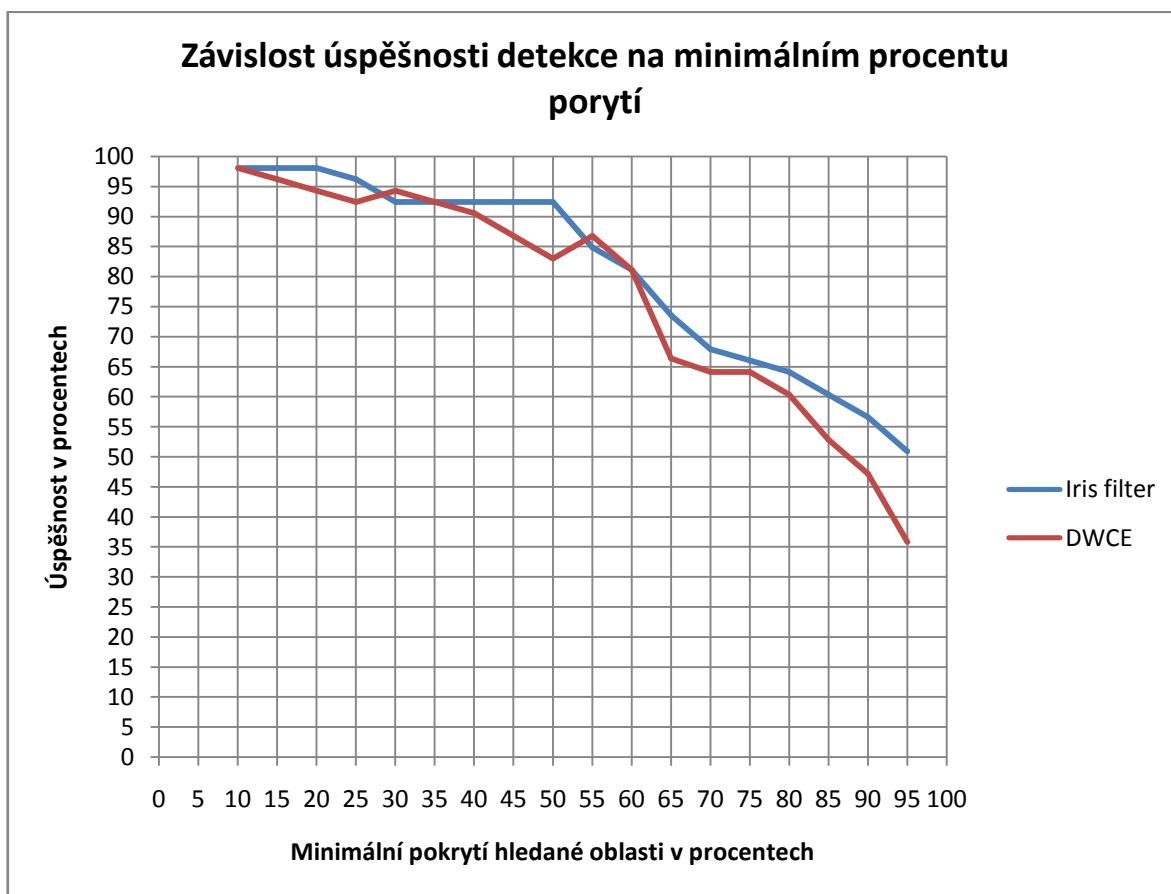
### 7.1 Porovnání Iris filtru a DWCE

Na tomto relativně malém datasetu jsem vyzkoušel mojí interpretaci DWCE a Iris filtru. V odstavci 5.7 uvádím, že nález je správně identifikovaný, jestliže ho počítačem určený ROI pokrývá z určitého procenta. Protože je pro různé aplikace vhodná jiná přesnost, sestavil jsem tabulku, kde postupně spouštím algoritmus pro různá hraniční pokrytí. Před každým testem ještě spustím optimalizaci parametrů pro daný minimální limit, abych dosáhl co nejpřesnějších výsledků. Výsledky se nachází v tabulce (Tabulka 3) a na jejím základě jsem vytvořil graf porovnání filtrů (Graf 3 a Graf 4).

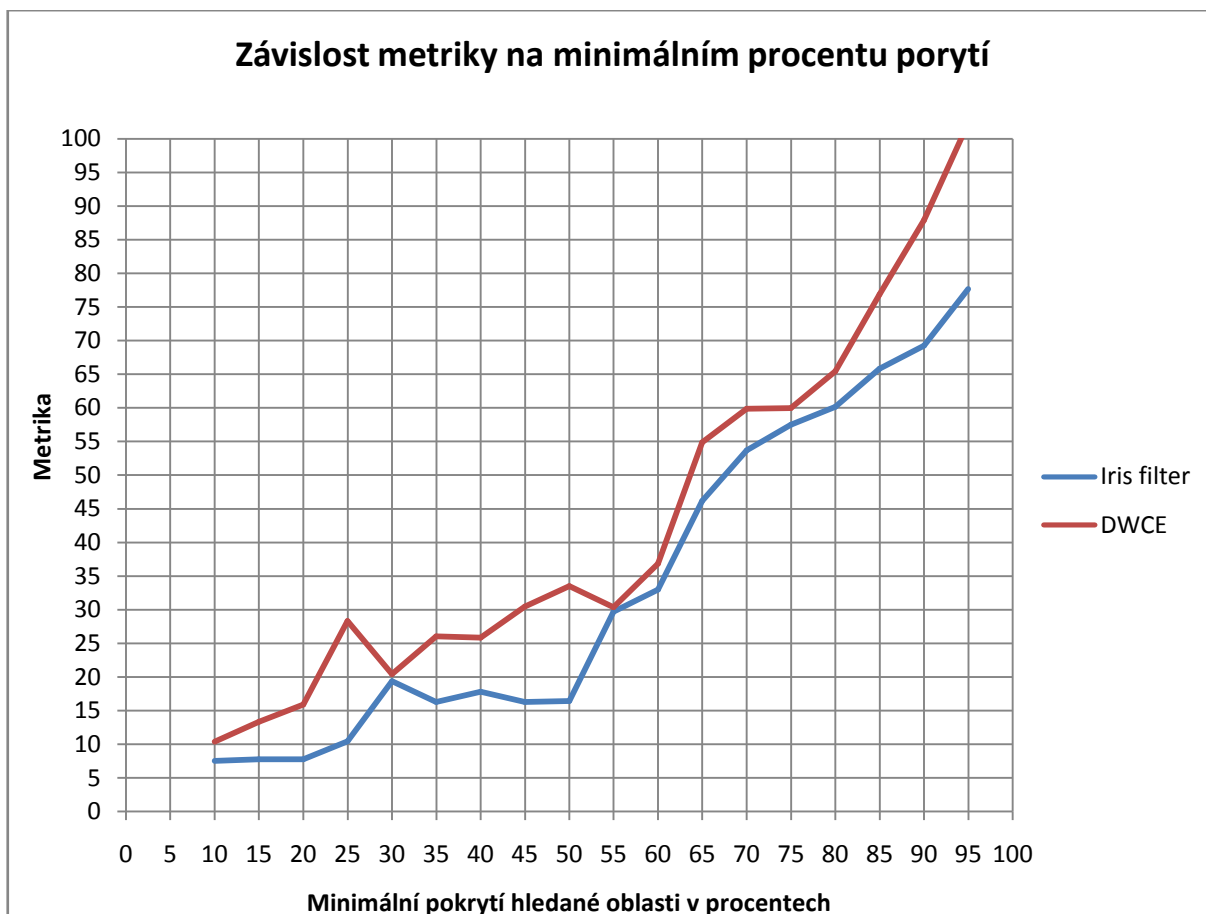
Metrika ve výsledcích je uvedena v odstavci 5.7. Parametry pro ní byly nastaveny  $\alpha = 150$  a  $\beta = 1$ .

Minimální porytí [%]	Úspěšnost [%]		FP/snímek		Metrika	
	Iris	DWCE	Iris	DWCE	Iris	DWCE
10	98,1132	98,1132	4,71698	7,56604	7,54718	10,39624
15	98,1132	96,2264	4,92453	7,67925	7,75473	13,33965
20	98,1132	94,3396	4,92453	7,41509	7,75473	15,90569
25	96,2264	92,4528	4,75472	17	10,41512	28,3208
30	92,4528	94,3396	8,03774	11,9434	19,35854	20,434
35	92,4528	92,4528	4,96226	14,717	16,28306	26,0378
40	92,4528	90,566	6,4717	11,69811	17,7925	25,84911
45	92,4528	86,7925	4,98113	10,6414	16,30193	30,45265
50	92,4528	83,0189	5,11321	8,0566	16,43401	33,52825
55	84,9057	86,7925	7,0566	10,5472	29,69805	30,35845
60	81,1321	81,1321	4,67925	8,5283	32,9811	36,83015
65	73,5849	66,3774	6,54717	4,43396	46,16982	54,86786
70	67,9245	64,1509	5,56604	6,09434	53,67929	59,86799
75	66,0377	64,1509	6,5283	6,18868	57,47175	59,96233
80	64,1509	60,3774	6,39623	6,01887	60,16988	65,45277
85	60,3774	52,8302	6,39623	6,11321	65,83013	76,86791
90	56,6038	47,1698	4,15094	8,64151	69,24524	87,88681
95	50,9434	35,8491	4,09434	6,07547	77,67924	102,3018

Tabulka 3: Výsledky Iris a DWCE filtru v závislosti na minimálním pokrytí ROI



Graf 3: Závislost úspěšnosti detekce na minimálním procentu porytí



Graf 4: Závislost výsledné metriky Iris a DWCE na minimálním pokrytí ROI

Z dosažených výsledků lze vyčíst, že implementace Iris filtru byla o něco úspěšnější. A to jak v přesnosti při detekování nálezů, tak i v menším počtu zbytečně označených oblastí na snímek. Okolo 50 % pokrytí dosáhl Iris slušné úspěšnosti v průměru 89,9 % při 5,7 FP oproti 85,5 % a 9,74 FP u DWCE. Toto je pravděpodobně způsobeno sofistikovanějším návrhem Iris filtru.

V původním článku o DWCE [33] provedli testy na 25 snímcích obsahující nádor. Jejich algoritmus dosáhl s první fází úspěšnosti okolo 90 % při 10 FP. První fáze se týkala pouze popsané techniky detekce a segmentace (viz 5.4). Dále ještě autoři vytvořili krok s cílem snížit počet chybně označených oblastí. Nakonec jejich implementace dosáhla úspěšnosti 96 % s 4,5 FP. Šlo pouze o předběžné výsledky a prvotní testování daného přístupu na velmi malém souboru obrázků. Nicméně i srovnání [22] uvádí výsledky 91 % 8,129 FP, které jsou o něco lepší než v mém případě.

Autoři Iris filtru uvádí 93 % správně nalezených ROI při 1 FP na snímek. S těmito výsledky se může mnou implementovaný Iris filtr měřit pouze při porovnání úspěšnosti. Ovšem ve stejném srovnávacím testu jako DWCE dosáhli „pouze“ 89,5 % s 7,295 FP tedy podobných údajů jako moje interpretace.

Musíme také vzít v úvahu, že jsem neimplementoval krok k redukci falešně označených oblastí. Ten obvykle bývá zařazen za detekční část programu a zpravidla je realizován pomocí klasifikačních algoritmů. Myslím, že obě moje interpretace dosáhly velmi dobré úspěšnosti.

## 7.2 Spojení filtrů pomocí lineární kombinace

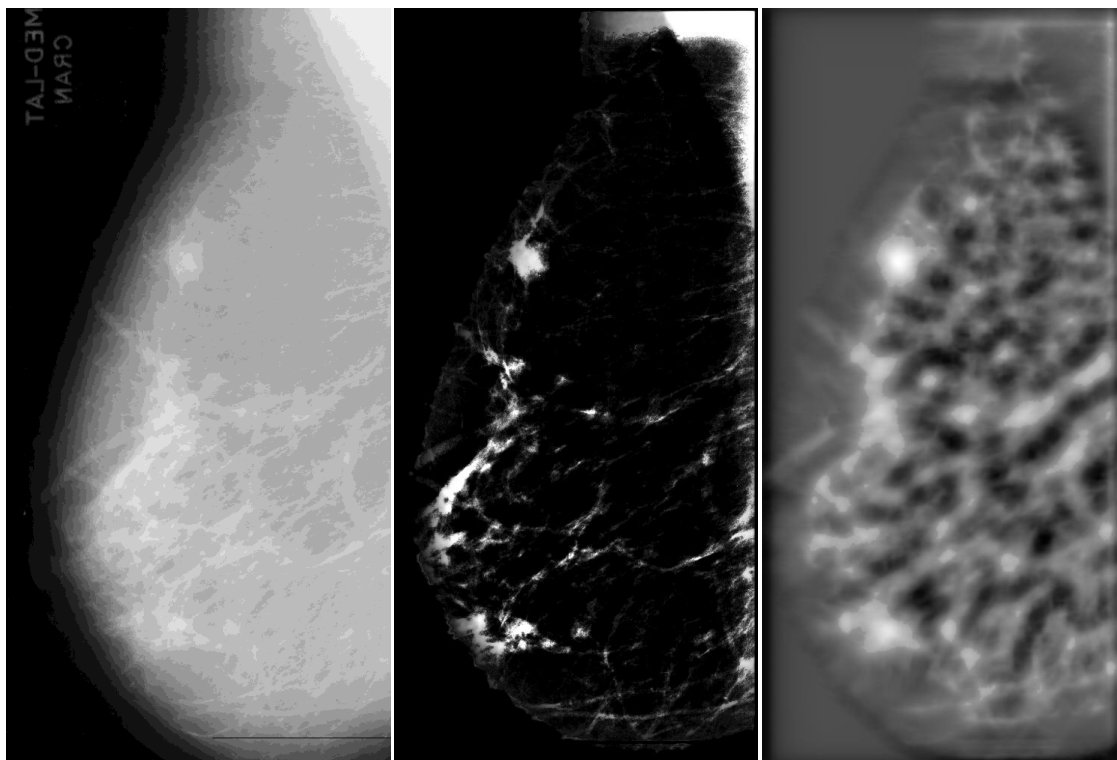
Za účelem zlepšení vlastností jsem zkusil oba filtry spojit pomocí lineární kombinace. Na to používám filtrované snímky (po průběhu filtrů, před zahájením detekce). Výsledný obraz vzniká sečtením hodnot pixelů na stejných souřadnicích podle vzorce:

$$I_v = \gamma \cdot I_{iris} + (1 - \gamma) \cdot I_{dwce}$$

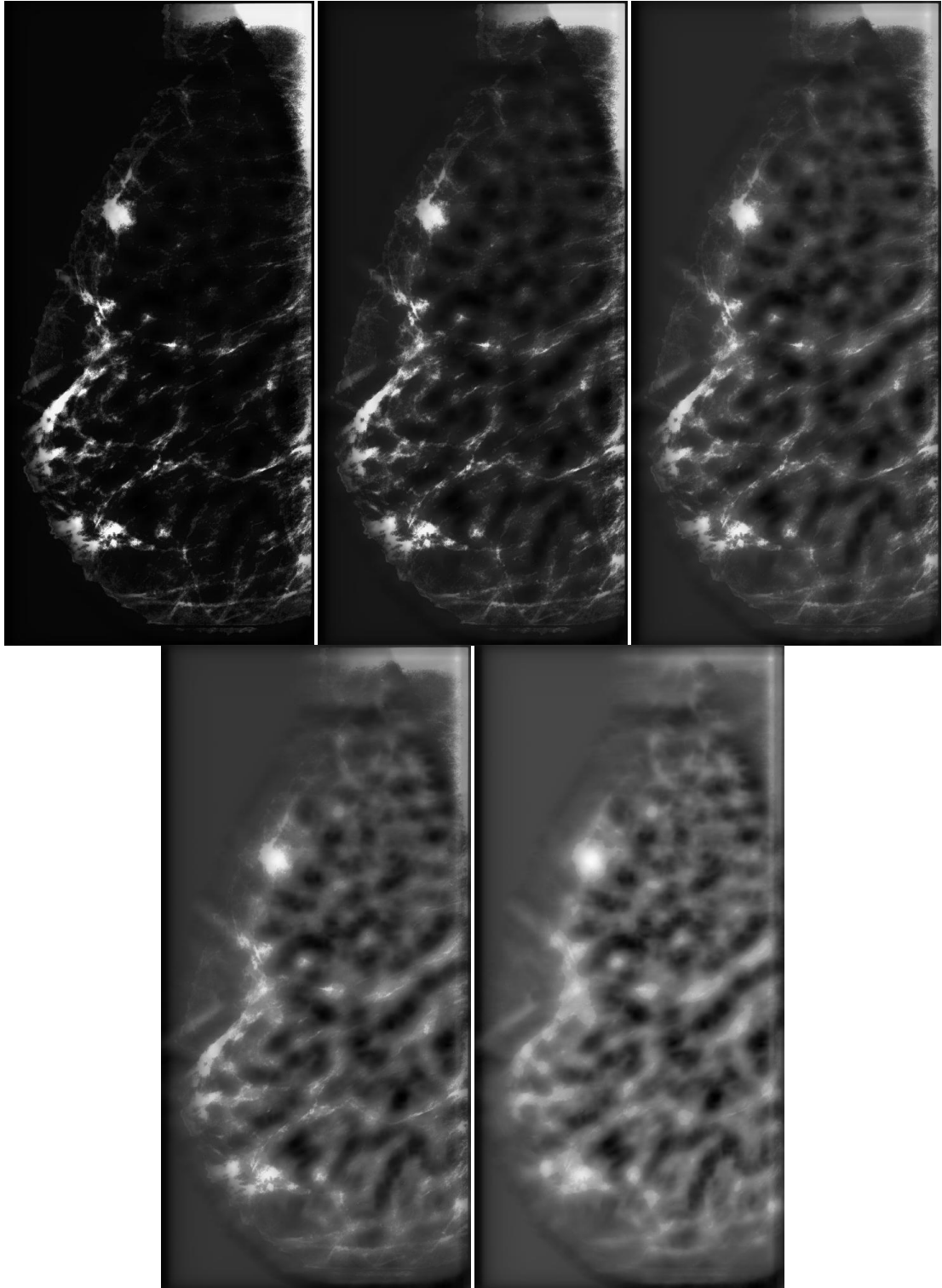
Kde  $I_v$  je výsledný obraz,  $I_{dwce}$  a  $I_{iris}$  jsou filtrované obrazy a  $\gamma$  určuje procentuální váhu, s jakou přispějí jednotlivé snímky do výsledného. Jestliže  $\gamma = 0,5$ , pak se obě složky podílejí stejně. Po vytvoření tohoto hybridního obrázku (viz Obrázek 25) následuje už známá fáze detekce ROI. Abych otestoval úspěšnost pro různé kombinace, zvolil jsem 5 zástupných hodnot poměrů. Pro  $\gamma = 0,1; 0,3; 0,5; 0,7$  a  $0,9$  jsem spustil stejný experiment s různými prahy pro pokrytí jako pro DWCE a Iris. Vytvořil jsem několik grafů, v nichž porovnávám dosažené výsledky s jednotlivými filtry a mezi sebou (viz příloha A).

Potvrdilo se, co bylo vidět už u srovnání obou původních algoritmů. Čím je  $\gamma$  větší, čím více přispívá Iris, tím je detekce úspěšnější. Zajímavostí se ovšem stal výsledek pro hodnotu  $\gamma = 0,7$ . Tato kombinace dosahuje místy lepších výsledků než úspěšnější z původních filtrů. Ukazuje tedy, že přeci jen nakonec spojení obou filtrů může přinést zlepšení.





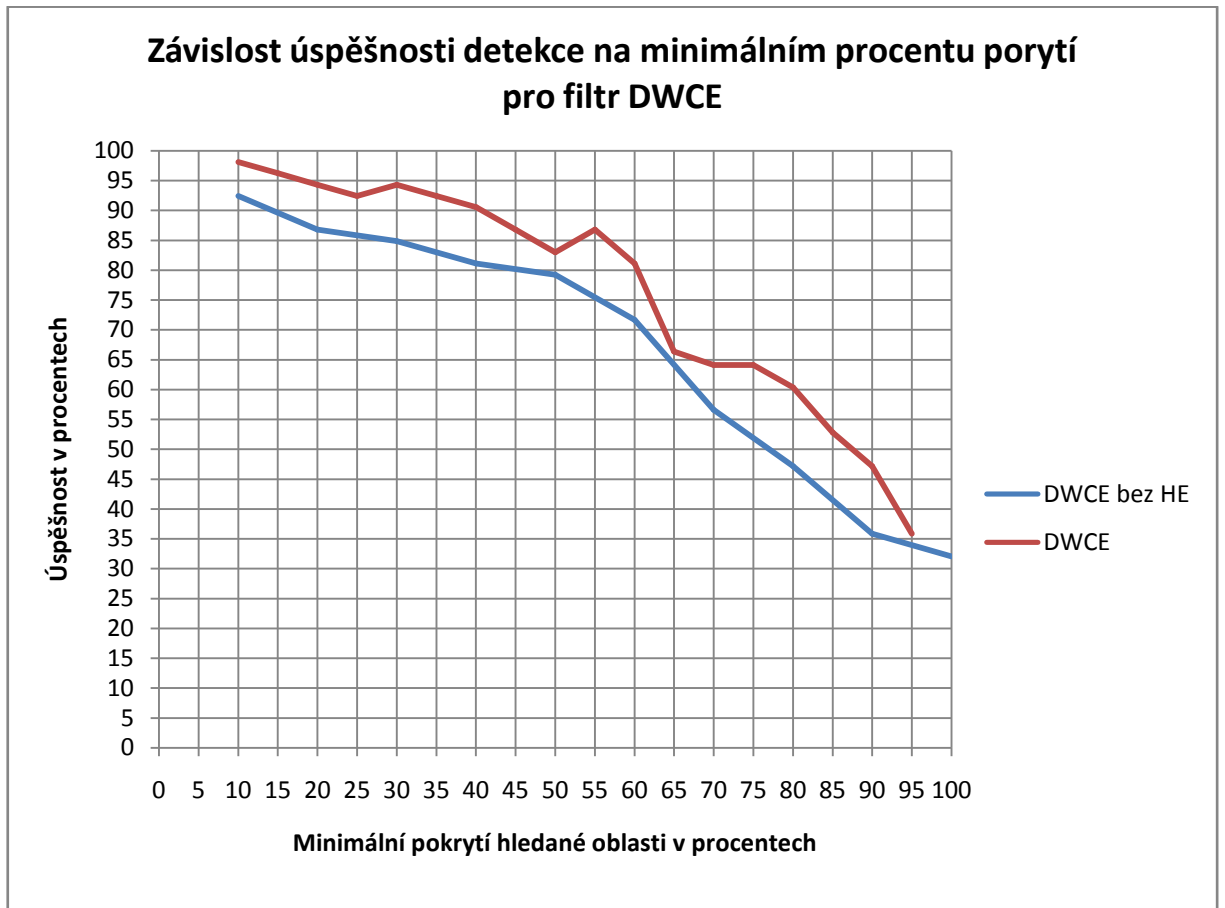
Obrázek 24: Zleva – originální obraz, výstup pro DWCE filter a výstup Iris fitru



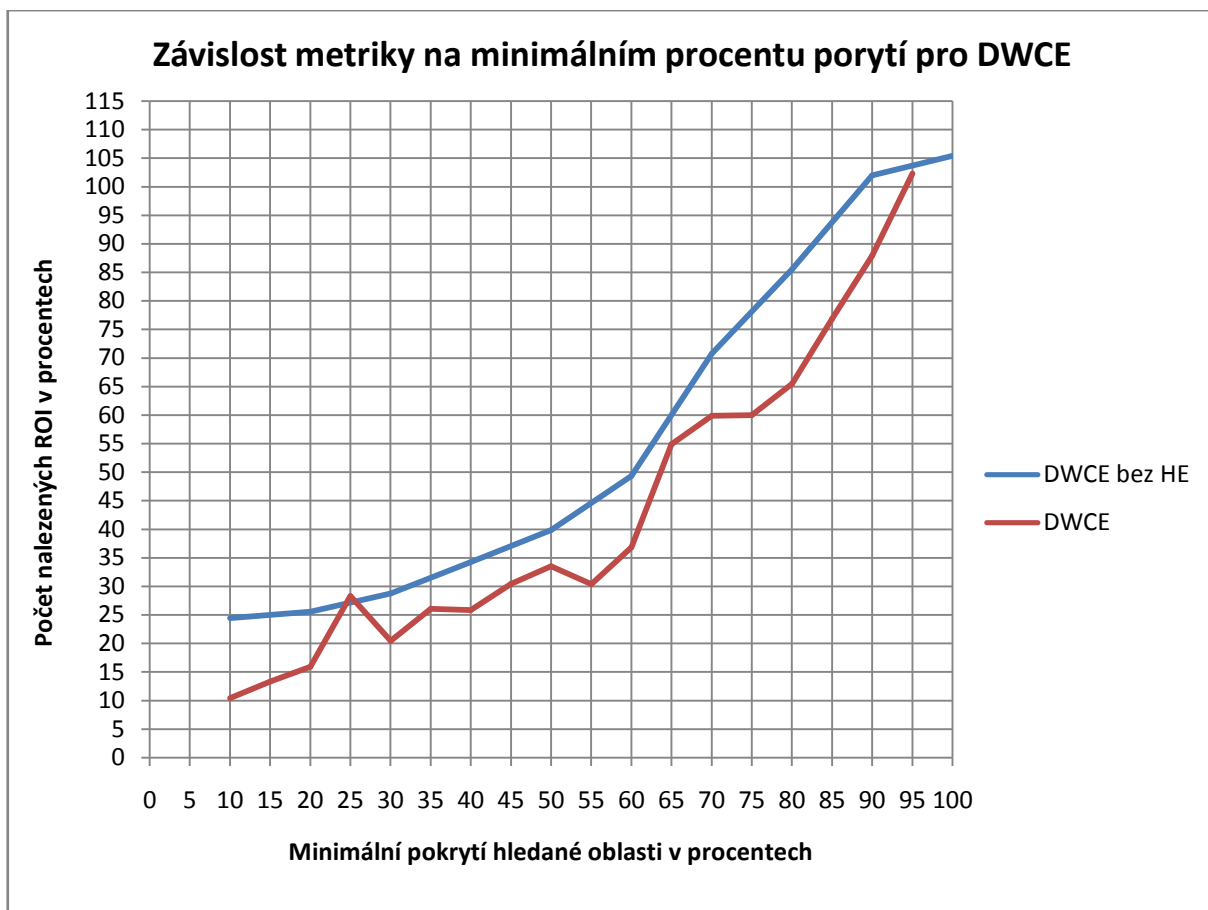
Obrázek 25: Obraz vzniklý spojením filtrů. Zleva pro  $\gamma = 0, 1; 0, 3; 0, 5; 0, 7$  a  $0, 9$

### 7.3 Vylepšení s použitím vyrovnání histogramu

Jak jsem psal v části 5.3.4, přidal jsem před samotnou aplikaci filtrů ještě vylepšení kontrastu pomocí vyrovnání histogramu. Podle následující grafů jde vidět, že tento krok pomohl k lepším výsledkům. U DWCE se dosáhlo velmi znatelného posunu oproti implementaci bez použití ekvalizace histogramu (viz Graf 5 a Graf 6). U Iris filtru jsou výsledky víceméně podobné s a bez použití této techniky, proto ani neuvádím jejich srovnání.



Graf 5: Úspěšnost DWCE s a bez použití vyrovnání histogramu (HE)



Graf 6: Porovnání míry u DWCE s a bez použití vyrovnání histogramu (HE)

## **8 Budoucnost**

Věcí, které lze vylepšit anebo přidat do implementovaného programu, je opravdu velké množství. Vývoj komplexního CAD systému rozhodně není jednoduchá záležitost. Zde uvedu několik částí, z nichž by se měl ideální systém skládat.

### **8.1 Redukce falešně označených oblastí**

Důležitou částí, kterou by bylo vhodné zařadit za mnou naprogramované filtry a detekci, je redukce špatně označených oblastí. Ideálním stavem by samozřejmě bylo, kdyby počet chybně označených oblastí na snímek byl rovný nule, nebo se alespoň k tomu přibližoval. Algoritmus pro snížení počtu FP má místo ve velkém počtu článků o CAD systémech. Jejich vývoj a úspěšné použití je kritickým krokem pro používání počítačových systémů ve screeningové mamografii. Velmi často jsou založeny na bázi učení (viz 4.6).

### **8.2 Detekce dalších artefaktů**

V této práci probírané detekční filtry lze použít téměř výhradně pro zvýraznění nádorů. Jenomže jak uvádím v sekci 2.8, doprovodných jevů u rakoviny je více. Navrhoval bych přidat další moduly, které budou detekovat kalcifikace, architectural distortion, bilateral asymmetry, případně další.

Protože se každý jev ukazující rakovinu chová a vypadá jinak, bude nutné pro každý z nich přidat vlastní programovou část. Problémem u takového rozmanitého systému ovšem bude, jak už bylo vícekrát zmíněno, počet FP. Ten by mohly řešit metody přidané z kroku redukce špatně označených oblastí.

### **8.3 Klasifikace nálezu**

V mamografii může být hodně zrádné určování, zda jde o zhoubný nebo nezhooubný nádor. To vede k většímu počtu operací, které nemusí být nutné. Pro tento účel se hodí klasifikátory, jenž určují typ nálezu. Jde vlastně o rozšíření kroku redukce chybně označených částí, kdy se algoritmy nebudou učit rozlišovat pouze zdravou a zasaženou tkáň, ale budou určovat i o jaký nález jde a jestli je benigní nebo maligní.

### **8.4 Segmentace nálezu**

Pro kvalitní klasifikaci nálezu je důležitým krokem řádná segmentace nálezu. V případě nádoru totiž o tom, že je zhoubný, často rozhodují jeho kontury a jejich nejbližší okolí.

## **8.5 Spojení více přístupů**

Ke zdokonalení kompletního systému by bylo vhodné zdvojnásobení či ztrojnásobení počtu implementovaných metod a následná kombinace jejich výsledků. Kdyby šlo o různé přístupy, mohlo by to přispět k stále vyšší úspěšnosti detekce rakoviny. Ve výsledku by bylo ale nutné, tyto algoritmy spojit do sebe tak, aby nedocházelo k nežádoucím jevům (zvýšení FP, snížení úspěšnosti).

## **8.6 Optimalizace systému**

Po složení komplexního systému ze všech navrhovaných vylepšení, by muselo následovat ještě důsledné otestování a vyladění tohoto CAD. S takovou složitou implementací se může stát, že by systém nemusel být použitelný třeba už jen z důvodu dlouhého výpočetního času nebo velkého množství chybně označených oblastí.

## 9 Závěr

Účelem této práce bylo prozkoumat možnosti detekce rakoviny z obrazových dat. Postupným získáváním znalostí a opatřením dat jsme se spolu s vedoucím této práce rozhodli pro zaměření na detekci rakoviny prsu z mamografických snímků.

V teoretickém úvodu seznamuji čtenáře s tím, co je to vlastně rakovina a rakovina prsu a jaká jsou možná vyšetření s důrazem na mamografii. V dalších částech se lze dozvědět o počítačových systémech a algoritmech pro napomáhání ke správné diagnóze.

Na základě referované úspěšnosti algoritmů v odborných člancích a podle uváděného srovnání jsem vybral dvě vhodné metody pro detekci karcinomu prsu z mamogramu: Density-Weighted Contrast Enhancement a Iris filtr. Ty jsem po jejich širším prozkoumání naprogramoval v jazyce C++ za využití knihovny pro podporu zpracování obrazových dat OpenCV.

Následně jsem tyto programy otestoval nad databází MIAS a ověřil jejich úspěšnost v porovnání s udávanou hodnotou, ať už ve zdrojovém článku nebo ve srovnávacím testu. Nakonec moje implementace Iris a DWCE dosáhly 89,9 %, resp. 85,5 % správně detekovaných tumorů při 5,7, resp. 9,74 FP. Tyto výsledky jsou porovnatelné s článkem se srovnáním, ze kterého jsem vycházel.

V rámci pokusu o vylepšení výsledků jsem vyzkoušel i různé přístupy pro kombinování těchto dvou filtrů.

Vytvořil jsem také program, který obě implementace spojuje do jednoho celku. Pomocí něj lze zpracovávat jednotlivé mamogramy, jejich celé dávky a vyhodnotit úspěšnost nastavených parametrů. Další částí programu jsou i prostředky pro optimalizaci parametrů filtrů s cílem dosáhnout co nejvyšší přesnosti. Přidal jsem i volbu pro zopakování uváděných experimentů.

V sekci nazvané „*Budoucnost*“ jsem v bodech navrhl možné zaměření na rozšíření a celkové zformování komplexního pomocného počítačového systému pro mamografii. Byl bych rád, kdyby se moje implementace někdy stala částí takového programu, který bude nasazen v praxi a napomůže zlepšení v oblasti diagnostiky rakoviny prsu.

## Pojmy

Pojem	Význam
<b>Tumor</b>	Nádor, rakovinné bujení
<b>Benigní, Maligní</b>	Nezhoubný, zhoubný nádor
<b>Screening</b>	V mamografii tento termín symbolizuje pravidelné preventivní vyšetření
<b>False positive(s)</b>	Chybně označená oblast se zdravou tkání
<b>False negative(s)</b>	Oblast neoznačená jako podezřelá, ale nachází se na ní nádor nebo jeho doprovodné jevy
<b>True positive(s)</b>	Úspěšně označená oblast zachycující nádor
<b>(Mikro)kalcifikace</b>	Shluk vápníku viditelný jako světlé tečky na mamogramu
<b>Second reading</b>	Systém nezávislého čtení jednoho snímku dvěma radiology pro zvýšení přesnosti
<b>Seed</b>	Počáteční bod výpočtu
<b>Intenzita</b>	V textu použito pro intenzitu šedé barvy na černobílém obrázku. Většinou v intervalu $\langle 0;255 \rangle$ , kde nula je černá a 255 bílá.
<b>Násobitel</b>	Vznikl z anglického multiplier, ovšem správný český termín je činitel
<b>Mutex</b>	Synchronizační prostředek pro vzájemné vyloučení

## Zkratky

Zkratka	Význam
<b>CAD</b>	Computer-Aided Detection (Diagnosis), pomocný počítačový systém pro mamografii popř. jiné odvětví medicíny
<b>DWCE</b>	Density-Weighted Contrast Enhancement – filtr pro zvýraznění podezřelých oblastí na mamogramu
<b>FP</b>	False positive(s), chybně označená oblast se zdravou tkání někdy i jako průměrná hodnota těchto chybných oblastí na snímek.
<b>CC</b>	Craniocaudal view – mamografický pohled shora
<b>MLO</b>	Medio-lateral oblique– mamografický pohled ze šikma
<b>ROI</b>	Region of Interest – oblast zájmu, podezřelá oblast
<b>DDSM</b>	<i>The Digital Database for Screening Mammography</i> – mamografická databáze
<b>MIAS</b>	<i>Mammographic Image Analysis Society</i> – skupina, která se stará o vývoj a údržbu mamografické databáze se stejným názvem



## Reference

- [1] ABRAHÁMOVÁ, Jitka. *Co byste měli vědět o rakovině prsu*. Praha: Grada, 2009. Doktor radí. ISBN 978-80-247-3063-9.
- [2] PETRÁKOVÁ, Katarína a Rostislav VYZULA. *O nádorech prsu* [online]. [cit. 2016-05-02]. Dostupné z: <http://www.linkos.cz/nadory-prsu-c50/o-nadorech-prsu>
- [3] *Breastcancer.org* [online]. [cit. 2016-05-02]. Dostupné z: [http://www.breastcancer.org/symptoms/types/male\\_bc](http://www.breastcancer.org/symptoms/types/male_bc)
- [4] Rizikové faktory. *Rakovinaprsu.cz* [online]. [cit. 2016-05-02]. Dostupné z: <http://www.rakovinaprsu.cz/o-rakovine-prsu/rizikove-faktory>
- [5] *Novotvary ČR* [online]. Praha: Ústav zdravotnických informací a statistiky České republiky, 1989- [cit. 2016-05-02]. ISBN 1210-857X.
- [6] *Globocan 2012: Estimated Cancer Incidence Mortality and Prevalence Worldwide in 2012* [online]. [cit. 2016-05-02]. Dostupné z: [http://globocan.iarc.fr/Pages/fact\\_sheets\\_cancer.aspx](http://globocan.iarc.fr/Pages/fact_sheets_cancer.aspx)
- [7] *National Cancer Institute: Breast Cancer Surveillance Consortium* [online]. [cit. 2016-05-02]. Dostupné z: <http://breastscreening.cancer.gov/statistics/performance/screening/2009/performance.html>
- [8] METTLER, Fred A., et al. Benefits versus risks from mammography: A critical reassessment. *Cancer*, 1996, 77.5: 903-909.
- [9] JOCHELSON, Maxine. *Pokročilé zobrazovací techniky detekce karcinomu prsu* [online]. [cit. 2016-05-02]. Dostupné z: <http://www.linkos.cz/asco/chapter/pokrocile-zobrazovaci-techniky-detekce-karcinomu-prsu/>
- [10] VIČAROVÁ, Vítězslava. *Problematiká nádoru prsu*. Zlín, 2007. Dostupné také z: [http://digilib.k.utb.cz/bitstream/handle/10563/3464/vi%20C4%8Darov%C3%A1\\_2007\\_bp.pdf?sequence=1](http://digilib.k.utb.cz/bitstream/handle/10563/3464/vi%20C4%8Darov%C3%A1_2007_bp.pdf?sequence=1). Bakalářská práce. Univerzita Tomáše Bati.
- [11] NG, E.Y.-K. A review of thermography as promising non-invasive detection modality for breast tumor. *International Journal of Thermal Sciences* [online]. 2009, 48(5), 849-859 [cit. 2016-05-02]. DOI: 10.1016/j.ijthermalsci.2008.06.015. ISSN 12900729. Dostupné z: <http://linkinghub.elsevier.com/retrieve/pii/S129007290800149X>
- [12] *Medicinální termografie* [online]. [cit. 2016-05-02]. Dostupné z: <http://meditermo.cz/homepage.html>
- [13] *Radiopaedia.org: Mammography views* [online]. [cit. 2016-05-02]. Dostupné z: <http://radiopaedia.org/articles/mammography-views>

- [14] CHENG, H.D., Xiaopeng CAI, Xiaowei CHEN, Liming HU a Xueling LOU. Computer-aided detection and classification of microcalcifications in mammograms: a survey. *Pattern Recognition* [online]. 2003, **36**(12), 2967-2991 [cit. 2016-05-02]. DOI: 10.1016/S0031-3203(03)00192-4. ISSN 00313203. Dostupné z: <http://linkinghub.elsevier.com/retrieve/pii/S0031320303001924>
- [15] RANGAYYAN, Rangaraj M., Fábio J. AYRES a J.E. LEO DESAUTELS. A review of computer-aided diagnosis of breast cancer: Toward the detection of subtle signs. *Journal of the Franklin Institute* [online]. 2007, **344**(3-4), 312-348 [cit. 2016-05-02]. DOI: 10.1016/j.jfranklin.2006.09.003. ISSN 00160032. Dostupné z: <http://linkinghub.elsevier.com/retrieve/pii/S001600320600127X>
- [16] GAUR, Shantanu, Vandana DIALANI, Priscilla J. SLANETZ a Ronald L. EISENBERG. Architectural Distortion of the Breast. *American Journal of Roentgenology* [online]. 2013, **201**(5), W662-W670 [cit. 2016-05-02]. DOI: 10.2214/AJR.12.10153. ISSN 0361-803x. Dostupné z: <http://www.ajronline.org/doi/abs/10.2214/AJR.12.10153>
- [17] *The Radiology Assistant* [online]. [cit. 2016-05-02]. Dostupné z: <http://www.radiologyassistant.nl/en/p53b4082c92130/bi-rads-for-mammography-and-ultrasound-2013.html#in53b4082cb1634>
- [18] *American Journal of Roentgenology* [online]. 1984, **143**(3) [cit. 2016-05-02]. ISSN 0361-803x. Dostupné z: <http://www.ajronline.org/doi/abs/10.2214/ajr.143.3.461>
- [19] GEISER, W. R., T. M. HAYGOOD, L. SANTIAGO, T. STEPHENS, D. THAMES a G. J. WHITMAN. Challenges in Mammography: Part 1, Artifacts in Digital Mammography. *American Journal of Roentgenology* [online]. 2011, **197**(6), W1023-W1030 [cit. 2016-05-02]. DOI: 10.2214/AJR.10.7246. ISSN 0361-803x. Dostupné z: <http://www.ajronline.org/cgi/doi/10.2214/AJR.10.7246>
- [20] CIATTO, Stefano, Marco Rosselli DEL TURCO, Gabriella RISSO, et al. Comparison of standard reading and computer aided detection (CAD) on a national proficiency test of screening mammography. *European Journal of Radiology* [online]. 2003, **45**(2), 135-138 [cit. 2016-05-02]. DOI: 10.1016/S0720-048X(02)00011-6. ISSN 0720048x. Dostupné z: <http://linkinghub.elsevier.com/retrieve/pii/S0720048X02000116>
- [21] KHOO, Lisanne A. L., Paul TAYLOR a Rosalind M. GIVEN-WILSON. Computer-aided Detection in the United Kingdom National Breast Screening Programme: Prospective Study 1. *Radiology* [online]. 2005, **237**(2), 444-449 [cit. 2016-05-02]. DOI: 10.1148/radiol.2372041362. ISSN 0033-8419. Dostupné z: <http://pubs.rsna.org/doi/abs/10.1148/radiol.2372041362>

- [22] OLIVER, Arnau, Jordi FREIXENET, Joan MARTÍ, Elsa PÉREZ, Josep PONT, Erika R.E. DENTON a Reyer ZWIGGELAAR. A review of automatic mass detection and segmentation in mammographic images. *Medical Image Analysis* [online]. 2010, **14**(2), 87-110 [cit. 2016-05-02]. DOI: 10.1016/j.media.2009.12.005. ISSN 13618415. Dostupné z: <http://linkinghub.elsevier.com/retrieve/pii/S1361841509001492>
- [23] *Hologic: The Science of Sure* [online]. [cit. 2016-05-02]. Dostupné z: [http://investors.hologic.com/news\\_releases?item=233](http://investors.hologic.com/news_releases?item=233)
- [24] GARVICAN, L. a S. FIELD. A Pilot Evaluation of the R2 Image Checker System and Users' Response in the Detection of Interval Breast Cancers on Previous Screening Films. *Clinical Radiology* [online]. 2001, **56**(10), 833-837 [cit. 2016-05-02]. DOI: 10.1053/crad.2001.0776. ISSN 00099260. Dostupné z: <http://linkinghub.elsevier.com/retrieve/pii/S0009926001907767>
- [25] BAUM, F, U FISCHER, S OBENAUER a E GRABBE. *Computer-aided detection in direct digital full-field mammography: initial results* [online]. [cit. 2016-05-02]. DOI: 10.1007/s00330-002-1393-3. ISSN 1432-1084. Dostupné z: <http://link.springer.com/article/10.1007/s00330-002-1393-3>
- [26] BAKER, Jay A., Eric L. ROSEN, Joseph Y. LO, Edgardo I. GIMENEZ, Ruth WALSH a Mary Scott SOO. Computer-Aided Detection (CAD) in Screening Mammography: Sensitivity of Commercial CAD Systems for Detecting Architectural Distortion. *American Journal of Roentgenology* [online]. 2003, **181**(4), 1083-1088 [cit. 2016-05-02]. DOI: 10.2214/ajr.181.4.1811083. ISSN 0361-803x. Dostupné z: <http://www.ajronline.org/doi/abs/10.2214/ajr.181.4.1811083>
- [27] *R2 ImageChcecker* [online]. [cit. 2016-05-02]. Dostupné z: <http://www.alpha-imaging.com/files/document%20library/hologic/hologic%20brochures/hologic%20cad%20image%20checker.pdf>
- [28] BREM, Rachel F., Jocelyn A. RAPELYEA, Gilat ZISMAN, Jeffrey W. HOFFMEISTER a Martin P. DESIMIO. Evaluation of breast cancer with a computer-aided detection system by mammographic appearance and histopathology. *Cancer* [online]. 2005, **104**(5), 931-935 [cit. 2016-05-02]. DOI: 10.1002/cncr.21255. ISSN 0008-543x. Dostupné z: <http://doi.wiley.com/10.1002/cncr.21255>
- [29] MALICH, Ansgar, Christiane MARX, Mirjam FACIUS, Thomas BOEHM, Marlies FLECK a Werner A. KAISER. Tumour detection rate of a new commercially available computer-aided detection system. *European Radiology* [online]. 2001, **11**(12), 2454-2459 [cit. 2016-05-02]. DOI: 10.1007/s003300101079. ISSN 0938-7994. Dostupné z: <http://link.springer.com/10.1007/s003300101079>

- [30] TANG, Jinshan, et al. Computer-aided detection and diagnosis of breast cancer with mammography: recent advances. *Information Technology in Biomedicine, IEEE Transactions on*, 2009, 13.2: 236-251.
- [31] GULIATO, D., R.M. RANGAYYAN, W.A. CARNIELLI, J.A. ZUFFO a J.E.L. DESAUTELS. Segmentation of breast tumors in mammograms by fuzzy region growing. *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Vol.20 Biomedical Engineering Towards the Year 2000 and Beyond (Cat. No.98CH36286)* [online]. IEEE, 1998, , 1002-1005 [cit. 2016-05-02]. DOI: 10.1109/IEMBS.1998.745618. ISBN 0-7803-5164-9. Dostupné z: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=745618>
- [32] ŠPANĚL, Michal. *Obrazové segmentační techniky* [online]. [cit. 2016-05-02]. Dostupné z: <http://www.fit.vutbr.cz/~spanel/segmentace/>
- [33] PETRICK, Nicholas, et al. An adaptive density-weighted contrast enhancement filter for mammographic breast mass detection. *Medical Imaging, IEEE Transactions on*, 1996, 15.1: 59-67.
- [34] VARELA, Celia, et al. Computerized detection of breast masses in digitized mammograms. *Computers in Biology and Medicine*, 2007, 37.2: 214-226.
- [35] KOBATAKE, Hidefumi, et al. Computerized detection of malignant tumors on digital mammograms. *Medical Imaging, IEEE Transactions on*, 1999, 18.5: 369-378.
- [36] OLIVERA, Arnau, et al. Mammographic mass eigendetection.
- [37] FARIDAH, Y. Digital versus screen film mammography: a clinical comparison. *Biomedical Imaging and Intervention Journal* [online]. 2008, 4(4), - [cit. 2016-05-02]. DOI: 10.2349/bij.4.4.e31. ISSN 1823-5530. Dostupné z: <http://www.bij.org/2008/4/e31/e31.pdf>
- [38] LEWIN, John M., Carl J. D'ORSI, R. Edward HENDRICK, Lawrence J. MOSS, Pamela K. ISAACS, Andrew KARELLAS a Gary R. CUTTER. Clinical Comparison of Full-Field Digital Mammography and Screen-Film Mammography for Detection of Breast Cancer. *American Journal of Roentgenology* [online]. 2002, 179(3), 671-677 [cit. 2016-05-02]. DOI: 10.2214/ajr.179.3.1790671. ISSN 0361-803x. Dostupné z: <http://www.ajronline.org/doi/abs/10.2214/ajr.179.3.1790671>
- [39] *University of South Florida Digital Mammography Home Page: DDSM: Digital Database for Screening Mammography* [online]. [cit. 2016-05-02]. Dostupné z: <http://marathon.csee.usf.edu/Mammography/Database.html>

- [40] HEATH, M, K BOWYER, D KOPANS, R MOORE a P KEGELMEYER. *The Digital Database for Screening Mammography* [online]. [cit. 2016-05-02]. Dostupné z: [http://marathon.csee.usf.edu/Mammography/software/HeathEtAlIWDM\\_2000.pdf](http://marathon.csee.usf.edu/Mammography/software/HeathEtAlIWDM_2000.pdf)
- [41] *OpenCV: Open Source Computer Vision* [online]. [cit. 2016-05-02]. Dostupné z: <http://opencv.org/>
- [42] BRADSKI, Gary R. *Learning OpenCV* [online]. Sebastopol: O'Reilly, c2008 [cit. 2016-05-02]. ISBN 978-0-596-51613-0.
- [43] *Intel® Threading Building Blocks* [online]. [cit. 2016-05-02]. Dostupné z: <https://www.threadingbuildingblocks.org/>
- [44] Histogram equalization. In: *Wikipedia: the free encyclopedia* [online]. San Francisco (CA): Wikimedia Foundation, 2001- [cit. 2016-05-02]. Dostupné z: [https://en.wikipedia.org/wiki/Histogram\\_equalization](https://en.wikipedia.org/wiki/Histogram_equalization)
- [45] PETRICK, Nicholas, et al. Combined adaptive enhancement and region-growing segmentation of breast masses on digitized mammograms. *Medical physics*, 1999, 26.8: 1642-1654.
- [46] PETRICK, Nicholas, et al. Automated detection of breast masses on mammograms using adaptive contrast enhancement and texture classification. *Medical physics*, 1996, 23.10: 1685-1696.
- [47] *Biologické algoritmy - evoluční algoritmy* [online]. [cit. 2016-05-02]. Dostupné z: <http://www.root.cz/clanky/biologicke-algoritmy-1-evolucni-algoritmy>

## Převzaté reference

- [48] KUPINSKI, M.A. a M.L. GIGER. Automated seeded lesion segmentation on digital mammograms. *IEEE Transactions on Medical Imaging* [online]. **17**(4), 510-517 [cit. 2016-05-03]. DOI: 10.1109/42.730396. ISSN 02780062. Dostupné z: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=730396>
- [49] BÅRMAN, H.; GRANLUND, G. H. Computer aided diagnosis of mammograms using a hierarchical framework. In: *EXCERPTA MEDICA INTERNATIONAL CONGRESS SERIES*. Elsevier ltd, 1994. p. 271-280.
- [50] POHLMAN, Scott, et al. Quantitative classification of breast tumors in digitized mammograms. *Medical Physics*, 1996, 23.8: 1337-1345.
- [51] ZHANG, Hun, et al. Automated breast masses segmentation in digitized mammograms. In: *Biomedical Circuits and Systems, 2004 IEEE International Workshop on*. IEEE, 2004. p. S2/2-S1-4.
- [52] HERREDSVELA, Jostein; GULSRUD, Thor Ole; ENGAN, Kjersti. Detection of circumscribed masses in mammograms using morphological segmentation. In: *Medical Imaging*. International Society for Optics and Photonics, 2005. p. 902-913.

- [53] SHESHADRI, H. S.; KANDASWAMY, A. Detection of breast cancer tumor based on morphological watershed algorithm. *ICGST-GVIP Journal*, 2005, 5.5.
- [54] GULSRUD, Thor Ole; ENGAN, Kjersti; HANSTVEIT, Thomas. Watershed segmentation of detected masses in digital mammograms. In: *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*. IEEE, 2006. p. 3304-3307.
- [55] RANGAYYAN, Rangaraj M., et al. Measures of acutance and shape for classification of breast tumors. *Medical Imaging, IEEE Transactions on*, 1997, 16.6: 799-810.
- [56] HONG, Byung-Woo; BRADY, Michael. A topographic representation for mammogram segmentation. In: *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2003*. Springer Berlin Heidelberg, 2003. p. 730-737.
- [57] YIN, Lijun; DESHPANDE, Sandeep; CHANG, Ja Kwei. Automatic lesion/tumor detection using intelligent mesh-based active contour. In: *Tools with Artificial Intelligence, 2003. Proceedings. 15th IEEE International Conference on*. IEEE, 2003. p. 390-397.
- [58] SAHINER, Berkman, et al. Improvement of mammographic mass characterization using spiculation measures and morphological features. *Medical Physics*, 2001, 28.7: 1455-1465.
- [59] TIMP, Sheila; KARSSEMELJER, Nico. A new 2D segmentation method based on dynamic programming applied to computer aided detection in mammography. *Medical physics*, 2004, 31.5: 958-971.
- [60] ABDEL-DAYEM, Amr R.; EL-SAKKA, Mahmoud R. Fuzzy entropy based detection of suspicious masses in digital mammogram images. In: *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*. IEEE, 2005. p. 4017-4022.
- [61] MATSUBARA, Tomoko, et al. Development of a new algorithm for detection of mammographic masses. In: *Digital Mammography*. Springer Netherlands, 1998. p. 139-142.
- [62] ROCHA, Alberto; TONG, Fu; YAN, Zhuangzhi. A logic filter for tumor detection on mammograms. *Journal of Computer Science and Technology*, 2000, 15.6: 629-632.
- [63] NG, Shun Leung; BISCHOF, Walter F. Automated detection and classification of breast tumors. *Computers and Biomedical Research*, 1992, 25.3: 218-237.
- [64] CONSTANTINIDIS, A. S.; FAIRHURST, M. C.; RAHMAN, Ahmad Fuad Rezaur. Detection of circumscribed masses in digital mammograms using behaviour-knowledge space method. *Electronics Letters*, 2000, 36.4: 1.

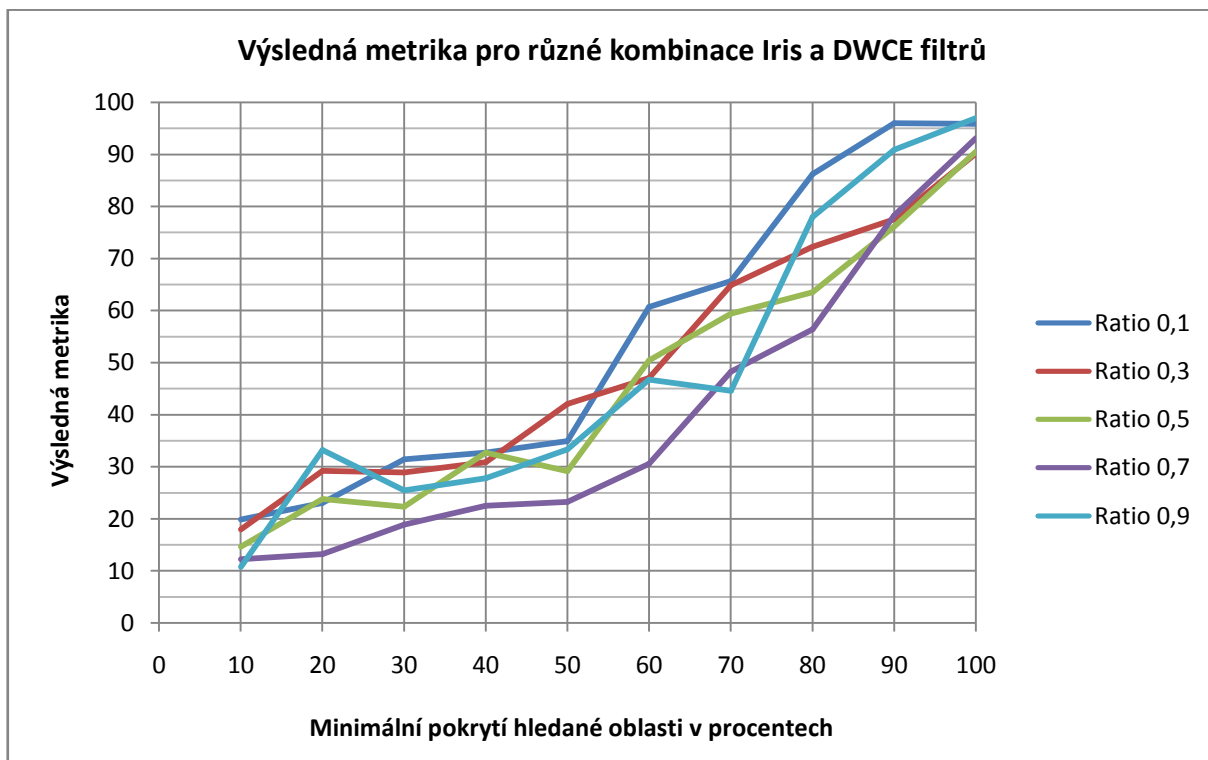
- [65] CHANG, Yuan-Hsiang; ZHENG, Bin; GUR, David. Computerized identification of suspicious regions for masses in digitized mammograms. *Investigative radiology*, 1996, 31.3: 146-153.
- [66] KARSSEMEIJER, Nico; TE BRAKE, Guido M. Detection of stellate distortions in mammograms. *Medical Imaging, IEEE Transactions on*, 1996, 15.5: 611-619.
- [67] STATHAKI, Tania; CONSTANTINIDES, A. G. Neural networks and higher order spectra for breast cancer detection. In: *Neural Networks for Signal Processing [1994] IV. Proceedings of the 1994 IEEE Workshop*. IEEE, 1994. p. 473-481.
- [68] WEI, Datong, et al. Detection of masses on mammograms using a convolution neural network. In: *Acoustics, Speech, and Signal Processing, 1995. ICASSP-95., 1995 International Conference on*. IEEE, 1995. p. 3483-3486.
- [69] ÖKTEM, V.; JOUNY, I. Automatic detection of malignant tumors in mammograms. In: *Engineering in Medicine and Biology Society, 2004. IEMBS'04. 26th Annual International Conference of the IEEE*. IEEE, 2004. p. 1770-1773.
- [70] YIN, Fang-Fang, et al. Computerized detection of masses in digital mammograms: Analysis of bilateral subtraction images. *Medical Physics*, 1991, 18.5: 955-963.
- [71] RICHARD, Frédéric JP; COHEN, Laurent D. A new image registration technique with free boundary constraints: application to mammography. *Computer Vision and Image Understanding*, 2003, 89.2: 166-196.
- [72] PAQUERAULT, Sophie, et al. Improvement of computerized mass detection on mammograms: Fusion of two-view information. *Medical Physics*, 2002, 29.2: 238-247.
- [73] ZHENG, Bin, et al. Multiview-based computer-aided detection scheme for breast masses. *Medical physics*, 2006, 33.9: 3135-3143.
- [74] VAN ENGELAND, Saskia; KARSSEMEIJER, Nico. Exploitation of correspondence between CC and MLO views in computer aided mass detection. In: *Digital Mammography*. Springer Berlin Heidelberg, 2006. p. 237-242.
- [75] SANJAY-GOPAL, S., et al. A regional registration technique for automated interval change analysis of breast lesions on mammograms. *Medical Physics*, 1999, 26.12: 2669-2679.
- [76] VUJOVIC, Nenad; BRZAKOVIC, Dragana. Establishing the correspondence between control points in pairs of mammographic images. *Image Processing, IEEE Transactions on*, 1997, 6.10: 1388-1399.

- [77] ZHU, Hui, Francis H.Y. CHAN a F.K. LAM. Image Contrast Enhancement by Constrained Local Histogram Equalization. *Computer Vision and Image Understanding* [online]. 1999, **73**(2), 281-290 [cit. 2016-05-02]. DOI: 10.1006/cviu.1998.0723. ISSN 10773142. Dostupné z: <http://linkinghub.elsevier.com/retrieve/pii/S1077314298907238>
- [78] *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* [online]. IEEE, 2008 [cit. 2016-05-02]. ISBN 978-1-4244-1814-5. Dostupné z: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4649636>
- [79] SARAGE, Govind N a Sagar JAMBHORKAR. *Enhancement of Mammography Images for Breast Cancer Detection using Histogram Processing Techniques* [online]. [cit. 2016-05-02]. ISSN 0976-8491. Dostupné z: <http://www.ijcst.com/vol24/1/govind.pdf>

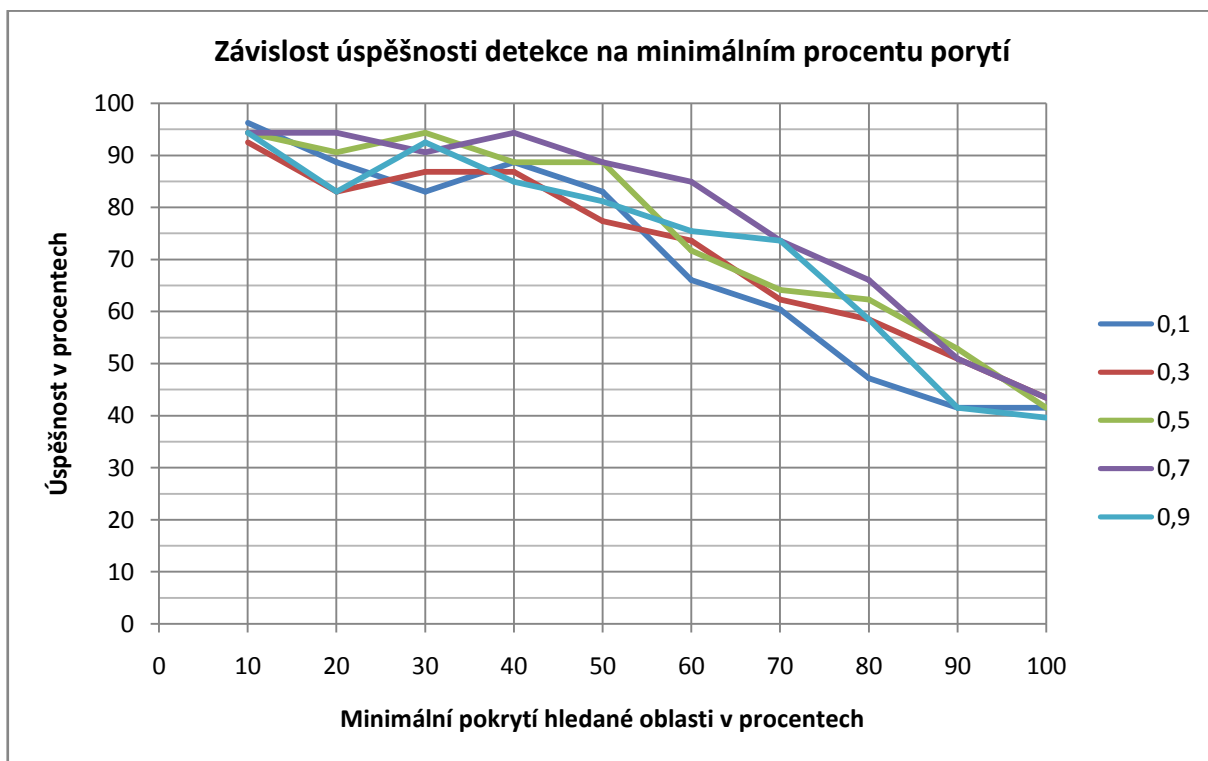


## **Přílohy**

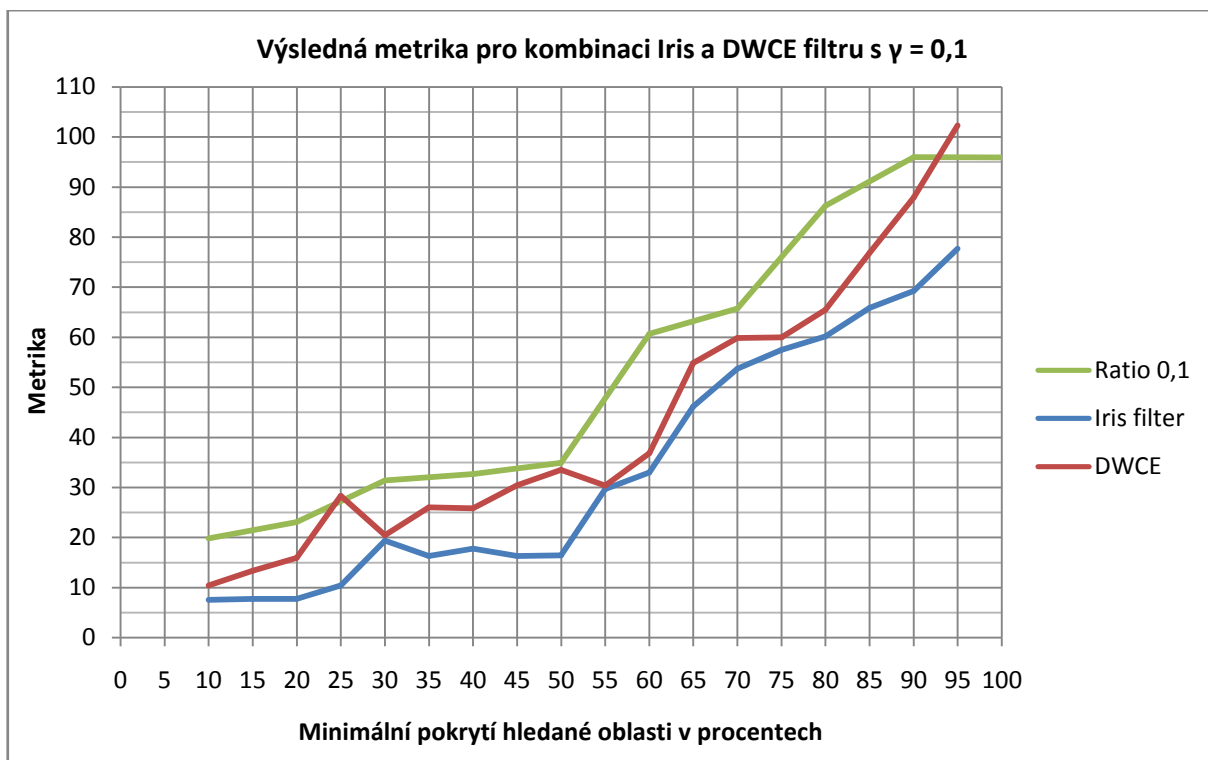
## A. Výsledky lineární kombinace algoritmů



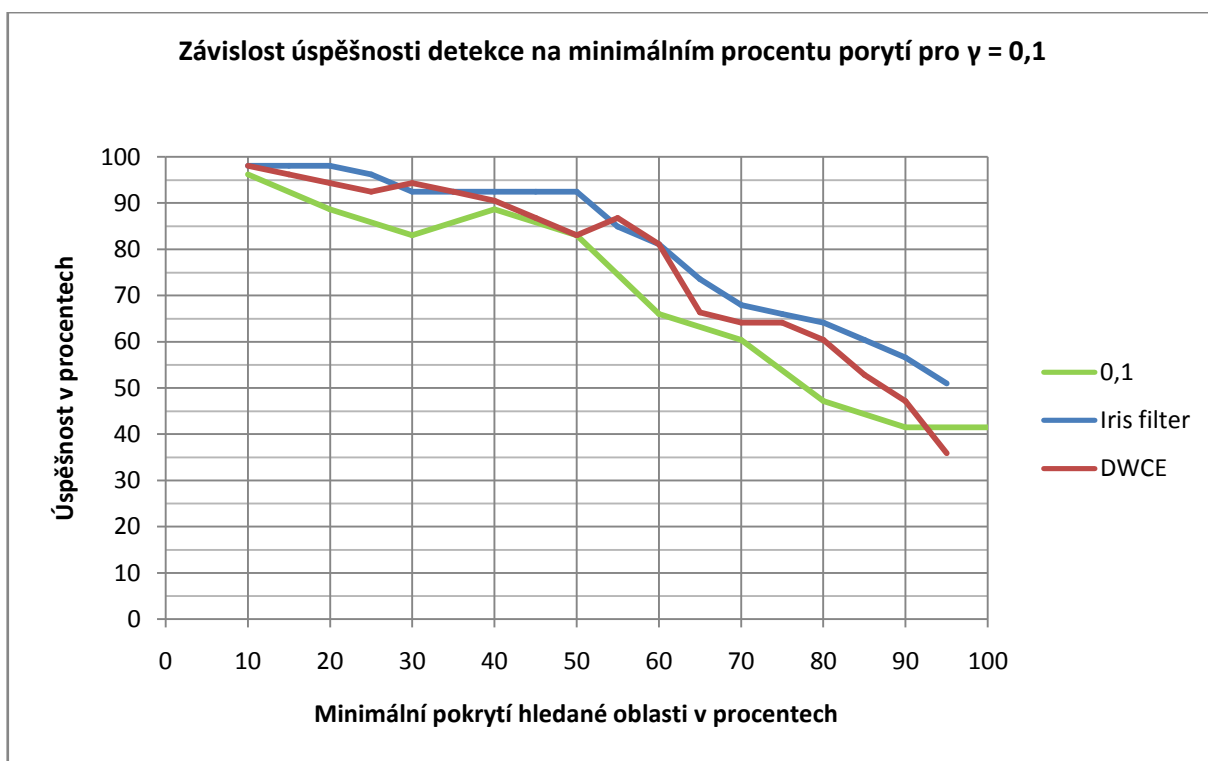
Graf 7: Přehled metrik pro různé hodnoty  $\gamma$



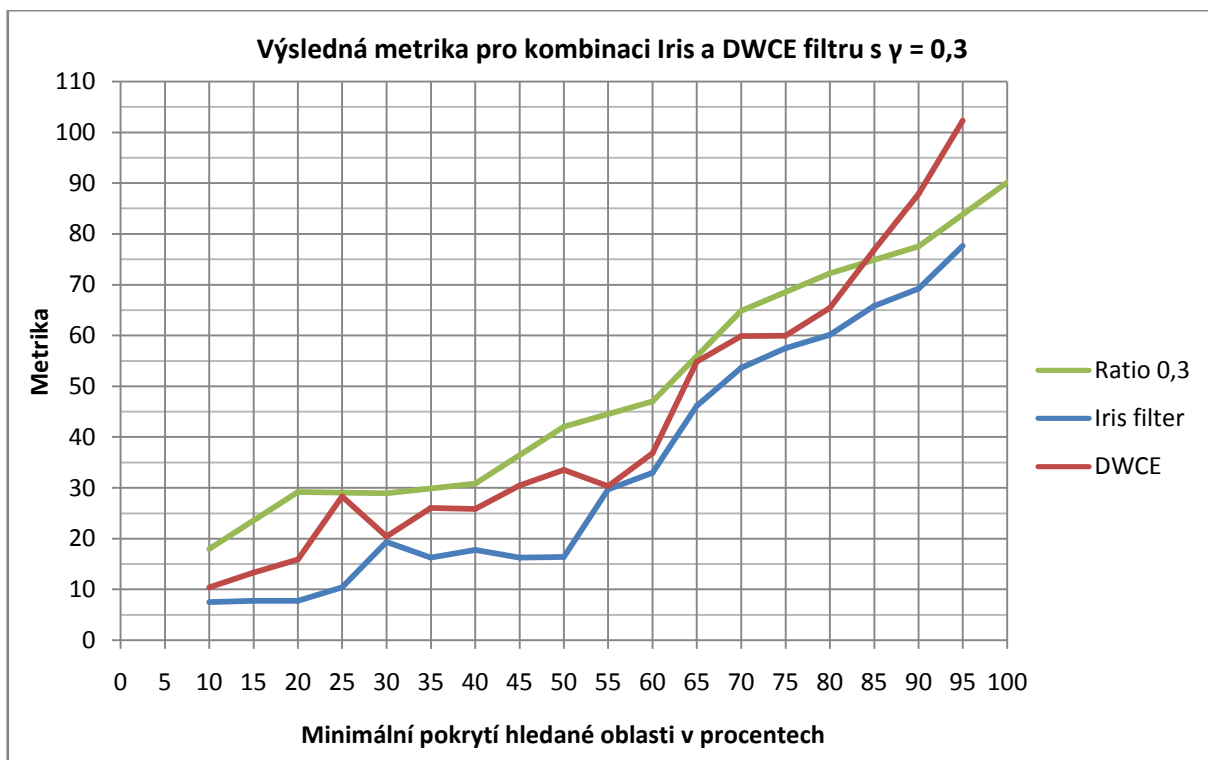
Graf 8: Přehled úspěšnosti pro různé hodnoty  $\gamma$



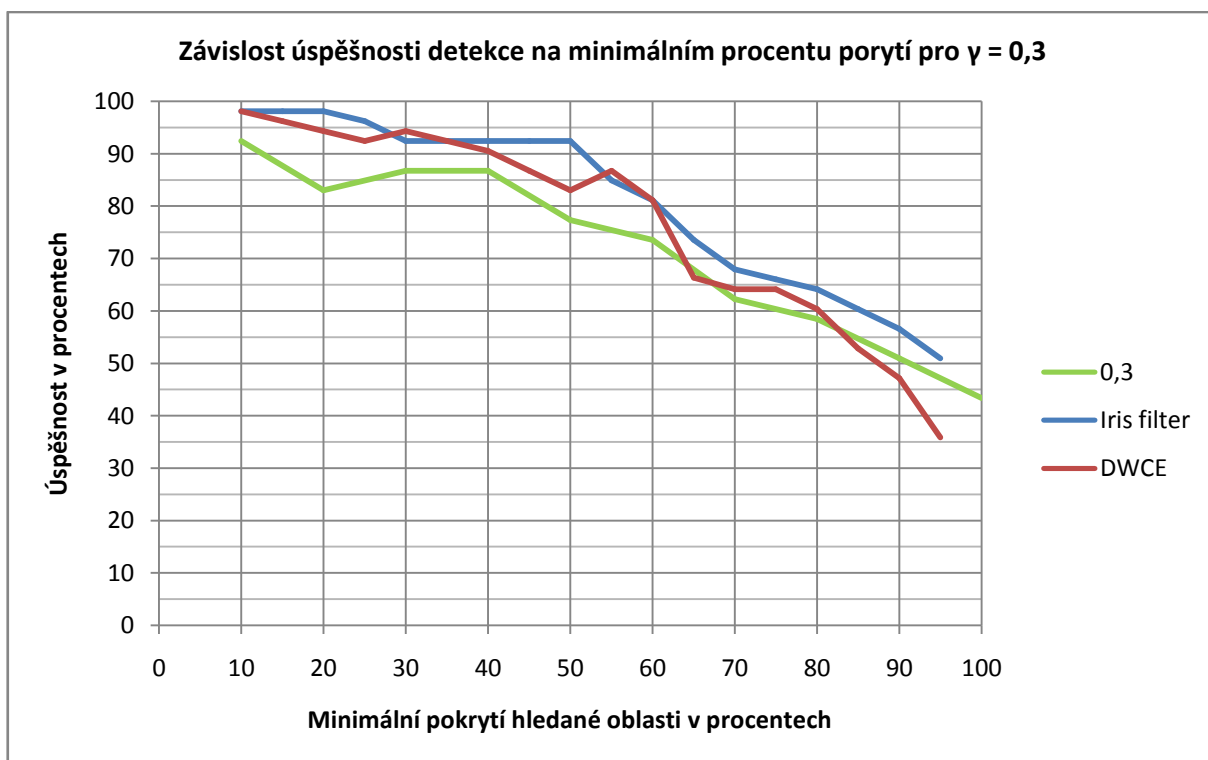
Graf 9: Metrika v závislosti na pokrytí pro  $\gamma = 0,1$



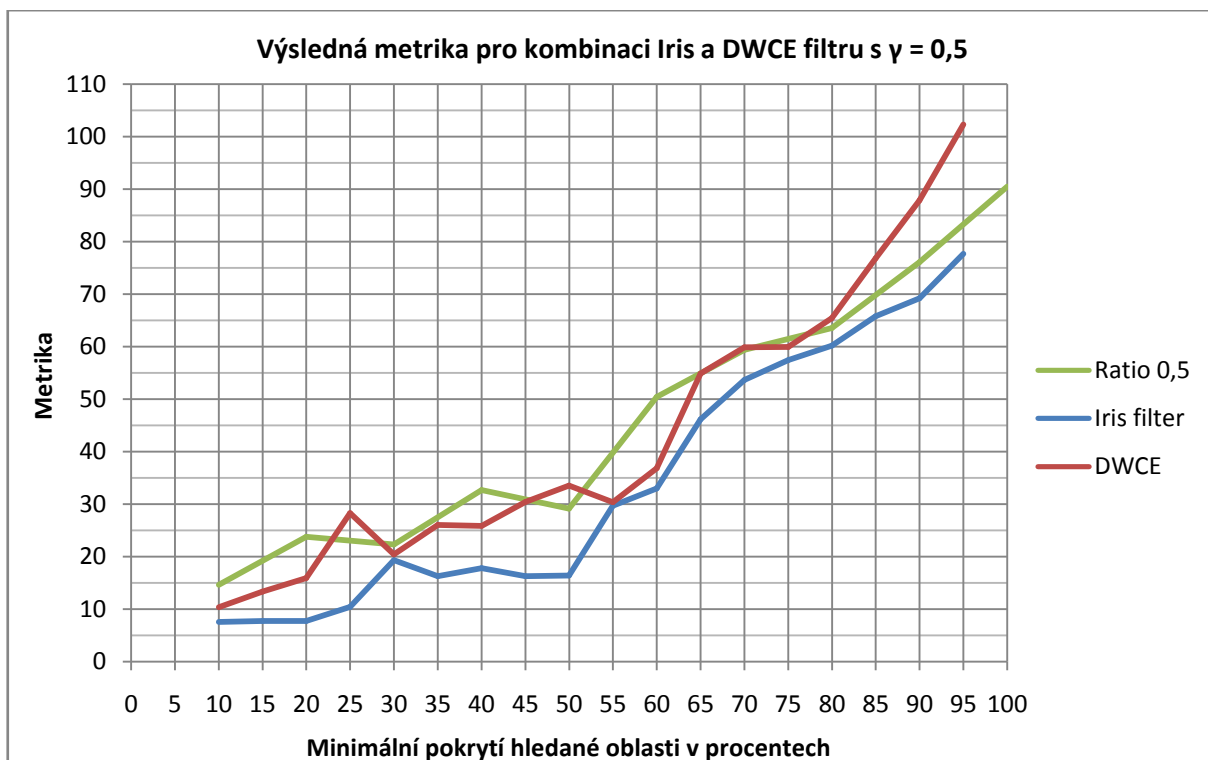
Graf 10: Úspěšnost v závislosti na pokrytí  $\gamma = 0,1$



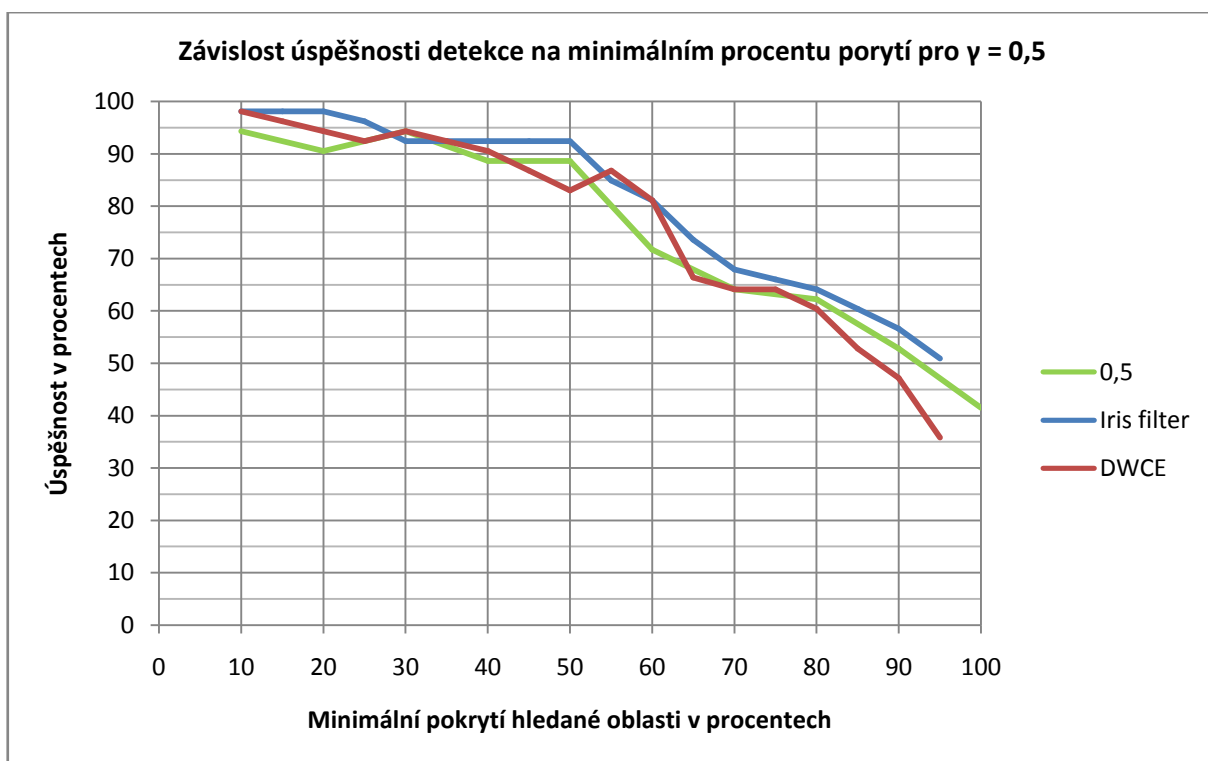
Graf 11: Metrika v závislosti na pokrytí pro  $\gamma=0,3$



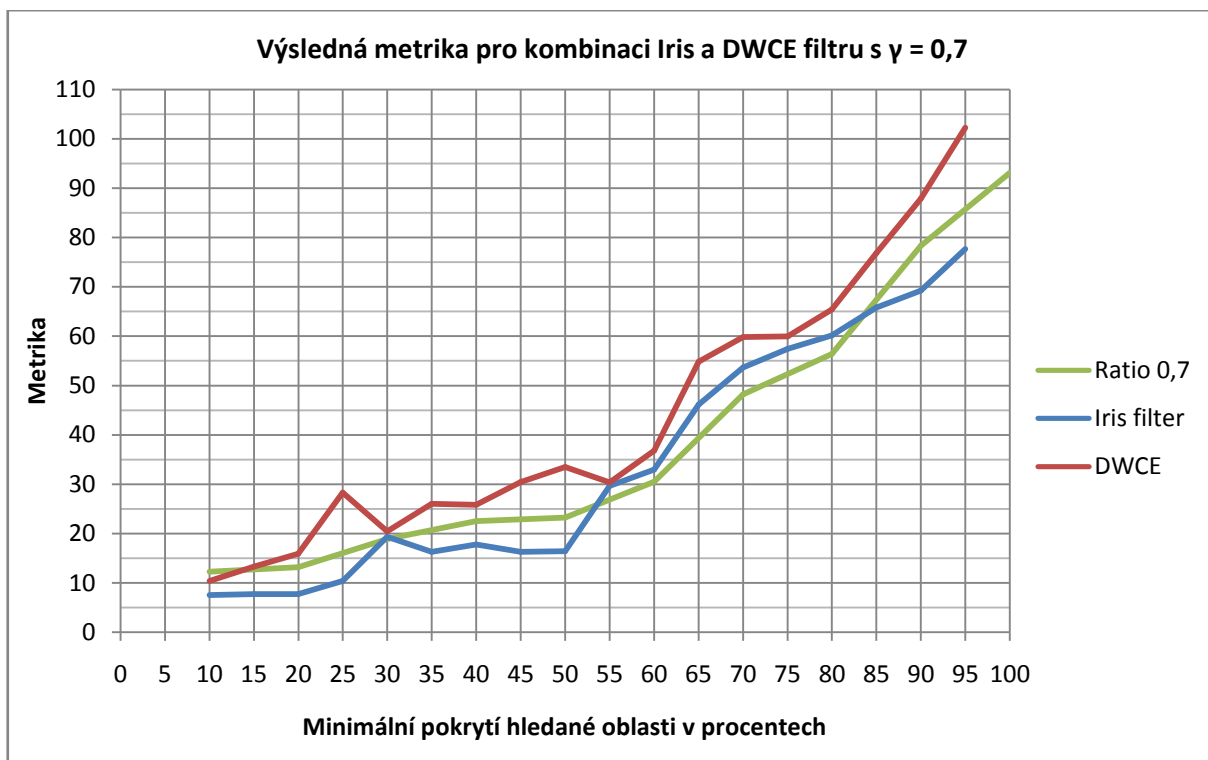
Graf 12: Úspěšnost v závislosti na pokrytí  $\gamma = 0,3$



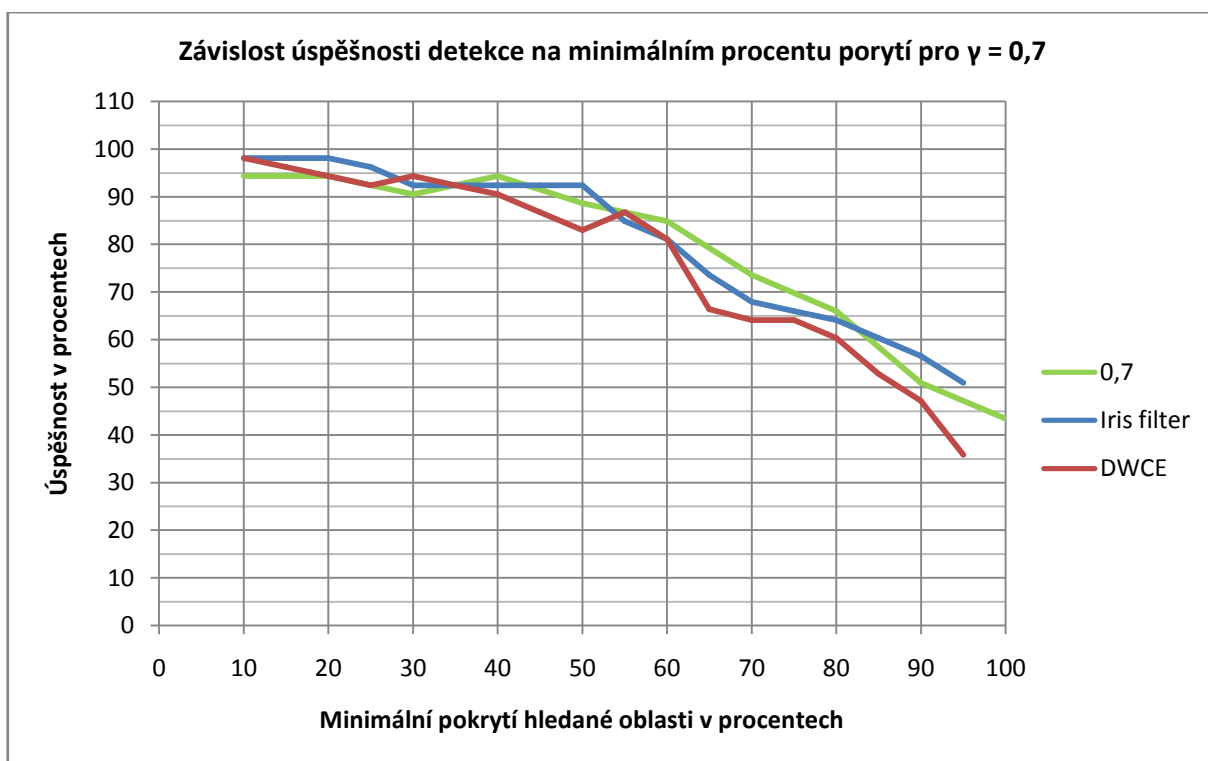
Graf 13: Metrika v závislosti na pokrytí pro  $\gamma = 0,5$



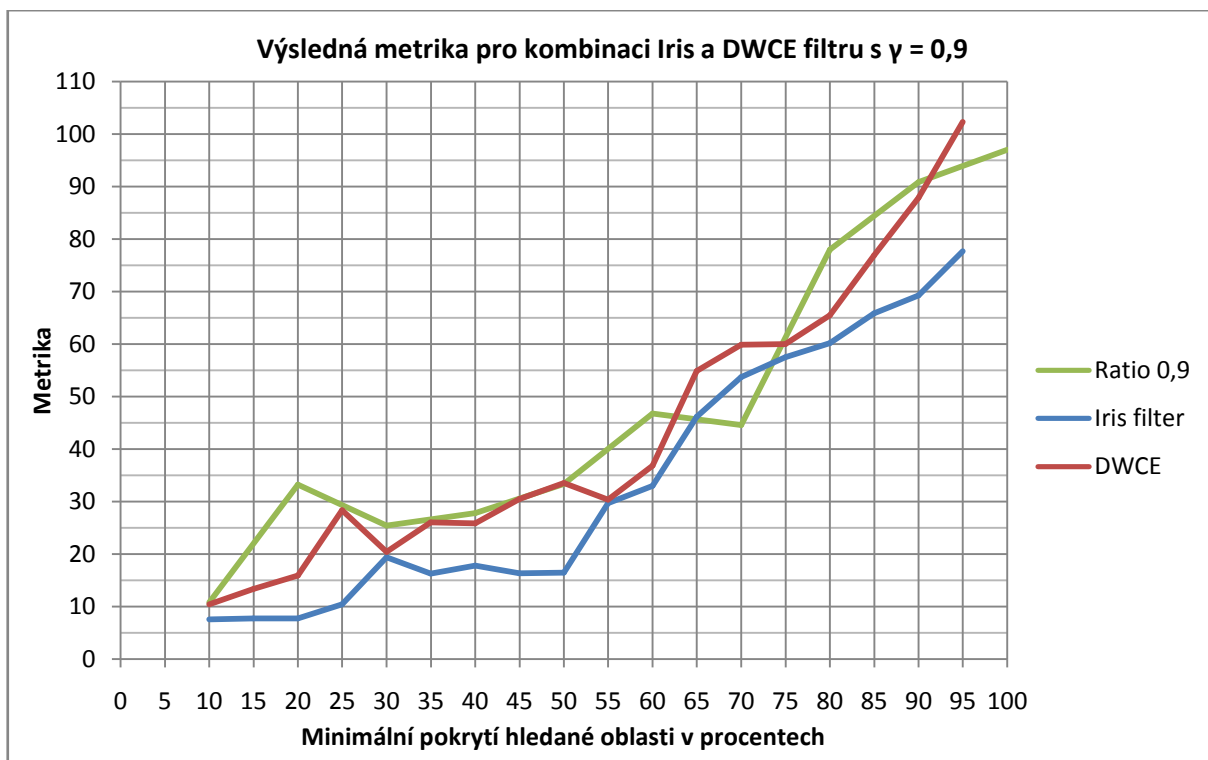
Graf 14: Úspěšnost v závislosti na pokrytí  $\gamma = 0,5$



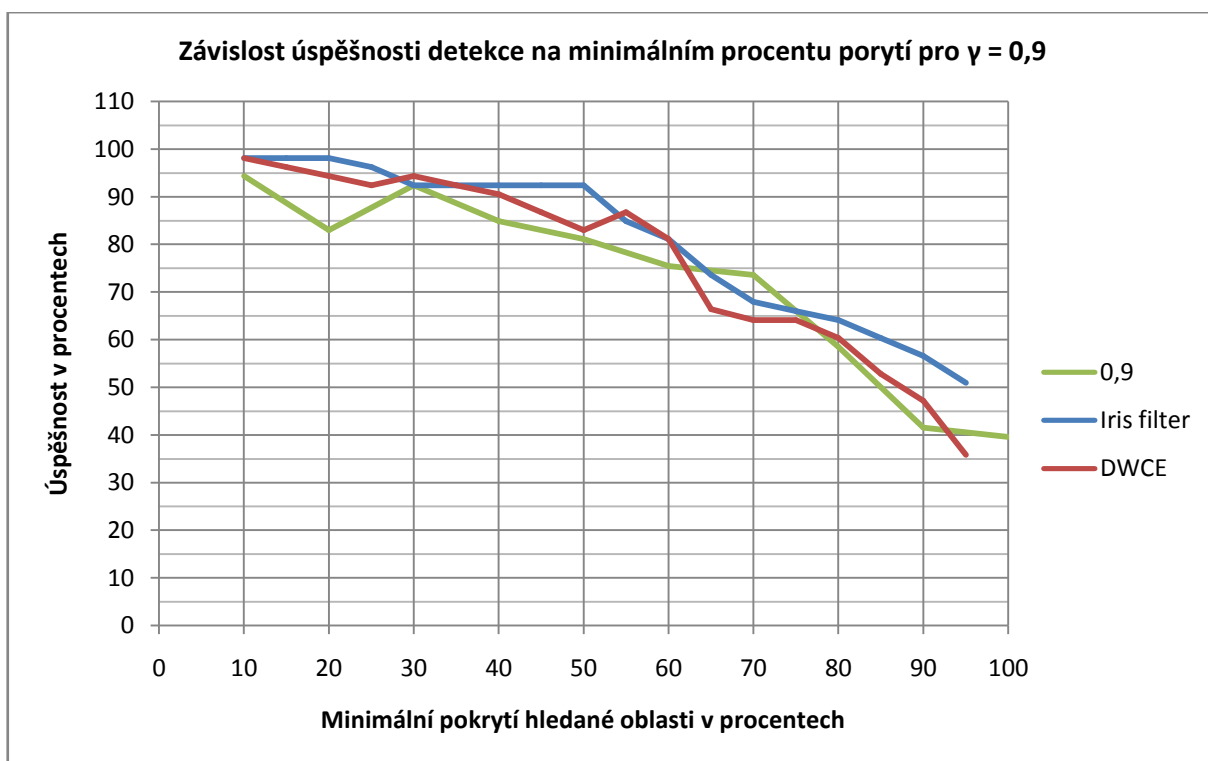
Graf 15: Metrika v závislosti na pokrytí pro  $\gamma=0,7$



Graf 16: Úspěšnost v závislosti na pokrytí  $\gamma = 0,7$



Graf 17: Metrika v závislosti na pokrytí pro  $\gamma = 0,9$



Graf 18: Úspěšnost v závislosti na pokrytí  $\gamma = 0,9$

## B. Uživatelská dokumentace

### Programové vybavení

Před překladem nebo spuštěním aplikace je nutné připravit veškeré knihovny. Jedná se o *OpenCV* v 3.0.0, Intel *Threading Building Blocks* (TBB) v 4.4, *Boost* v 1.55 a *CMake* verze minimálně 3.1 nebo jejich kompatibilní verze. Jejich výčet a další podrobnosti jsou uvedeny v kapitolách 6.5 a 6.6. Jestliže vám některá knihovna chybí obraťte se na manuálové stránky autorů.

Knihovna	Stránky
<b>Opencv</b>	<a href="http://opencv.org/">http://opencv.org/</a>
<b>TBB</b>	<a href="https://www.threadingbuildingblocks.org/">https://www.threadingbuildingblocks.org/</a>
<b>Boost</b>	<a href="http://www.boost.org/">http://www.boost.org/</a>
<b>CMake</b>	<a href="https://cmake.org/">https://cmake.org/</a>

### Překlad

Překlad je samozřejmě možný manuálně pomocí C++ překladače s přílinkováním veškerých závislostí, nebo v nějakém vývojovém prostředí podporujícím C++ s využitím složek `src` a `headers`.

### CMake

Pro automatizovaný překlad je přiložen skript `CMakeLists.txt` pro multiplatformní nástroj CMake. Ten je schopný vygenerovat prostředky pro překlad na různých platformách od Makefile pro Linux až po projekty ve Visual Studiu.

### Spuštění

Pro spuštění generátoru CMake je nutné být v kořenové složce, kde se nachází `CMakeLists.txt` a `FindTBB.cmake` a složka se zdrojovými soubory `src` a hlavičkovými `headers`. `FindTBB.cmake` je soubor pro vyhledání knihovny TBB, protože ta nemá standardní zastoupení v CMake.

Generování se spouští příkazem:

```
cmake .
```

### Překlad

Zde se budou jednotlivé verze různit a záležít, na jaké platformě proběhlo generování. Na Linux-like systémech se defaultně jedná o *Makefile*, u Windows bývá jako výchozí nástroj zvoleno *Visual Studio*. Seznam dostupných cílových platform lze dohledat pomocí `cmake -help`. Zde je možno se dozvědět i potřebná nastavení pro vygenerování souborů pro jinou platformu a sadu nástrojů.



Nyní už stačí projekt přeložit obvyklým způsobem pro vygenerovaný nástroj: `make` pro *Makefile*, překlad pomocí prostředí s *Visual Studiem*.

## **Spuštění**

V případě úspěšného překladu se vytvořila spustitelná aplikace s výchozím názvem `CancerDetection`. Pro spuštění programu je nutné, aby se ve stejné složce nacházel konfigurační soubor. Aplikace se spouští pomocí příkazu:

```
[./]CancerDetection [snímek] [přepínače]
```

- `[]` – volitelné položky
- `./` - příkaz pro spuštění na dané platformě
- `snímek` - cesta ke snímku, který se má zpracovat

### **Přepínače**

#### **Nápověda**

**Přepínač:** `-h` nebo `--help`

**Funkce:** Zobrazí krátkou nápovědu k ovládání programu

#### **Dávka**

**Přepínač:** `-b` nebo `--batch`

**Parametry:** cesta k souboru se jmény snímků, které se mají zpracovat. Snímky se musí nacházet ve stejné složce jako soubor s názvy.

**Funkce:** Zpracuje všechny snímky uvedené v zadaném textovém souboru s názvy. Pro každý vytvoří složku s příslušným jménem. V této složce se budou nacházet vyříznuté podezřelé oblasti (ROI) a originální obrázek s poznačenými ROI.

#### **Výběr metody**

**Přepínač:** `-m` nebo `--method`

**Parametry:** Číslo metody – 0 = DWCE (výchozí), 1 = Iris, 2 = lineární kombinace DWCE a Iris.

**Funkce:** Nastaví, jaká metoda se má používat při aktuálním spuštění programu.

#### **Optimalizace**

**Přepínač:** `-o` nebo `--optim`

**Parametry:** Cesta k souboru se jmény a anotacemi ke snímkům, pro které se mají optimalizovat parametry. Snímky se musí nacházet ve stejné složce jako soubor s anotacemi.

**Funkce:** Spustí diferenciální evoluci s cílem optimalizovat parametry detekce ROI pro danou sadu obrázků.

## **Konfigurační soubor**

**Přepínač:** -c nebo --config

**Parametry:** Cesta k nově zvolenému konfiguračnímu souboru

**Funkce:** Změní pro aktuální spuštění konfigurační soubor. Výchozí konfigurační soubor je nastaven na `default.ini`.

## **Načítání z disku**

**Přepínač:** -p nebo --prepared

**Funkce:** Filtrované obrázky s příslušným jménem se načtou z disku. Používá se v případech, že už jednou byly filtry aplikovány a jeho výstupy jsou uloženy. DWCE a Iris nemusí tedy být znovu aplikovány z důvodu úspory času. Zde je akorát výjimka u kombinace filtrů. U té se z disku načítají filtrované obrázky z DWCE a Iris a pak až se spojují.

## **Vyhodnocení**

**Přepínač:** -e nebo --evaluate

**Parametry:** Soubor s anotovanými daty

**Funkce:** Vyhodnotí úspěšnost vybraného filtru pro danou sadu anotovaných snímků.

## **Experiment**

**Přepínač:** -t nebo --test

**Parametry:** Soubor s názvy pro optimalizaci, soubor s anotovanými daty pro vyhodnocení

**Funkce:** Spustí test závislosti úspěšnosti na minimálním pokrytí nalezeného ROI (viz 7). Výsledky zapíše do souboru s vygenerovaným názvem (viz Vstupy a výstupy)

## **Vstupy a výstupy**

Veškeré výstupní soubory se budou ukládat do místa, odkud se program spouští. Vstupy lze uvedením cesty nalézt kdekoliv. Po zpracování jednoho snímku nebo dávky se vždy zapisují jejich upravené verze s příslušným prefixem (DWCE\_ ,IRIS\_ a DI\_ pro kombinaci)

## **Soubor s názvy**

Soubor obsahuje názvy snímků, které se nachází ve stejné složce jako tento soubor. Jako parametr lze uvést cestu k tomuto souboru.

Obrázky jsou uvedeny v seznamu bez uvedených koncovek. Tyto koncovky se doplní na základě nastavení v konfiguračním souboru (viz Konfigurační soubor).

mdb001

mdb003

snimek8

### **Soubor s anotovanými daty**

Podobný předchozímu, ovšem zde se za názvy uvádí ještě informace o nálezů na snímku. Tyto anotace jsou stejného formátu jako u databáze MIAS (viz 4.8.2)

### **Výstupní soubor s informacemi**

Výstupem z optimalizace parametrů a experimentů jsou soubory s názvem podle klíče

Datum\_optim\_metoda.txt

- Datum se vyplní automaticky podle toho, kdy byl program spuštěn. Je ve formátu YYYYMMDDHHmmSS. Když se optimalizace spustí v 14. 1. 2012 v 21:45 bude vypadat datum v názvu 20120114214500.
- Metoda udává jaký filtr byl použit (0 = DWCE, 1 = Iris, 2 = Kombinace)

Po optimalizaci bude výstup vypadat následovně:

```
precision = 0.913043 fp/pic = 4.04348
threshold = 0.99551
dilations = 24
opens = 13
eccentricity = 0.094833
growth = 50
match = 13.7384
```

Kde první řádka udává dosažené procento úspěšnosti v desetinném čísle (1 = 100%) a průměrný počet false positives na snímek.

Zbytek řádků je nastavení pro vyhledávání ROI (viz Nastavení pro vyhledávání)

Výstupním soubor po provedení experimentu bude mít tvar:

```
coverage 0.1 percentage 0.924528 fp 13.0943 0.872795 0.357013 5.62793
0.307586 26.8365 20
coverage 0.2 percentage 0.867925 fp 5.75472 0.829153 4.99516 15.1309
0.343888 50 20
coverage 0.3 percentage 0.849057 fp 6.09434 0.806479 3.58599 14.9786
0.369516 50 20
coverage 0.4 percentage 0.245283 fp 1.22642 1 32.1556 11.3592 0.601376
50 14.8341
coverage 0.5 percentage 0.792453 fp 8.73585 0.856972 9.23388 18.2579
0.408989 12.9379 0
coverage 0.6 percentage 0.716981 fp 6.88679 0.684468 4.05525 17.0999
0.325082 50 6.06719
coverage 0.7 percentage 0.566038 fp 5.62264 0.769022 3.53978 16.802
0.434372 50 15.2166
coverage 0.8 percentage 0.471698 fp 6.30189 0.838841 4.04194 15.9411
0.300858 43.0471 15.2038
coverage 0.9 percentage 0.358491 fp 5.77358 0.725375 3.14949 18.3993
0.432683 47.8129 7.06062
coverage 1 percentage 0.320755 fp 3.49057 0.792333 6.16566 19.9131
0.406018 45.6412 19.5076
```

Zde první údaj značí minimální pokrytí ROI a další dva jsou opět úspěšnost pro danou hladinu pokrytí a počet chybně označených částí na snímek. Zbytek už jsou parametry pro vyhledávání ROI, které byly vypočítány při optimalizaci.

### Výstup ze zpracování snímku

Po filtrování snímku vybranou metodou a nalezení podezřelých oblastí se vytvoří složka s názvem podle zpracovávaného snímku. V této složce se budou nacházet výřezy ROI a také snímek s klíčovým slovem `overall` v názvu. Na tom budou vyznačeny jednotlivé oblasti.

### Konfigurační soubor

Konfigurační soubor musí obsahovat všechny následující informace:

Název	Význam	Výchozí hodnota
<code>suffix</code>	Přípona vstupních a výstupních souborů	<code>.pgm</code>
<code>optimizing_loops</code>	Počet generací diferenciální evoluce při optimalizování parametrů (viz 5.8.1)	200
<code>coverage</code>	Minimální pokrytí ROI	0,5 (50%)
<code>padding-top</code>	Počet pixelů nahoře a dole, které budou černé (viz 5.3.2)	10
<code>padding-sides</code>	Počet pixelů vpravo a vlevo, které budou černé (viz 5.3.2)	10
<code>evaluation_alfa</code>	Parametr $\alpha$ pro výpočet míry (viz 5.7)	150
<code>evaluation_beta</code>	Parametr $\beta$ pro výpočet míry (viz 5.7)	1
<code>max_roi_factor</code>	Maximum, kolikrát může být ROI větší než skutečný nádor	6
<code>fp_max_cover</code>	Maximální pokrytí pro označení jako <code>false positive</code>	0,05
<code>min_roi_area</code>	Minimální velikost ROI	100

Dále musí konfigurační soubor obsahovat nastavení pro jednotlivé filtry. Ukázka celého souboru:

```
suffix = .pgm
optimizing_loops = 200
coverage = 0.5
padding-top = 10
padding-sides = 10
evaluation_alfa = 150
evaluation_beta = 1
max_roi_factor = 6
fp_max_cover = 0.05
min_roi_area = 100
```

```
[DWCE]
blurring = 65
threshold = 0.497875
```

```
dilations = 2
opens = 14
eccentricity = 0.498437
growth = 50
match = 2.74539
```

```
[IRIS]
microns = 200
rmin = 2500
rmax = 17500
threshold = 0.93499
dilations = 10
opens = 22
eccentricity = 0.384391
growth = 50
match = 0.367546
halflines = 35
```

```
[BOTH]
ratio = 0.7
threshold = 0.1
dilations = 20
opens = 20
eccentricity = 0.6
growth = 50
match = 1
```

## Nastavení pro vyhledávání

Název	Význam
threshold	Práh pro vytvoření binárního obrazu
dilations	Počet iterací morfologické dilatace
opens	Počet iterací morfologického otevření
eccentricity	Prahová hodnota excentricity
growth	Hodnota v pixelech, o kterou se mohou podobné oblasti při spojování vzdalovat
match	Maximální intenzita, o kterou se mohou podobná ROI lišit při spojování

## Ukázkové spuštění programu

```
CancerDetection -m 1 obrazky/mdb001.pgm
```

Zpracuje snímek mdb001.pgm ze složky obrazky pomocí Iris filtru. Jako výstup se vytvoří složka mdb001, která bude obsahovat všechny nalezené ROI.

```
CancerDetection -b cancer.txt
```

Pomocí DWCE upraví snímky uvedené v souboru cancer.txt. U tohoto popisného souboru se musí nacházet uvedené obrázky. Pro všechny zpracované snímky vytvoří příslušné složky.

```
CancerDetection --batch cancer.txt -m 2 -p
```

Podobné jako předchozí. Tentokrát načte výstupy s prefixem IRIS a DWCE z disku a spojí je podle nastavení v konfiguračním souboru. A dále pokračuje stejně jako u výše zmíněného – najde zajímavé oblasti na všech snímcích z cancer.txt a zapíše je do příslušných složek.

```
CancerDetection -o optimalizace/iris/vyber.txt --method 1 --prepared
```

Načte z disku snímky filtrované Iris filtrem a provede pro ně optimalizaci. Výstup se zapíše do souboru např. 20160505151512\_optim\_1.txt (program se spouštěl v 15:15 5. 5. 2016). V tom se budou nacházet hodnoty pro jednotlivé části detekce, jenž vyšli jako nejoptimálnější.

```
CancerDetection -e anotovana.txt -m 1 -c new_config.ini
```

Nastaví konfigurační soubor na new\_config.ini a provede vyhodnocení úspěšnosti nad anotovanými daty ze souboru anotovana.txt. Jako zvolená metoda bude Iris filtr.

```
CancerDetection --test vyber.txt vyhodnoceni.txt
```

Pomocí filtru DWCE upraví obrazy z obou vstupních souborů. Následně spustí test pro různá minimální pokrytí ROI (viz 7). Pro každou úroveň optimalizuje parametry na výběru snímků ze souboru vyber.txt a po tomto kroku vyhodnotí úspěšnost na vyhodnoceni.txt a zapíše do příslušného souboru (viz Vstupy a výstupy).