

Univerzita Karlova

1. lékařská fakulta

Lékařská biofyzika



UNIVERZITA KARLOVA
1. lékařská fakulta

Mgr. Antonín Procházka

**POČÍTAČOVÁ ANALÝZA ULTRAZVUKOVÝCH OBRAZŮ UZLŮ
VE ŠTÍTNÉ ŽLÁZE SE ZAMĚŘENÍM NA JEJICH SONOGRAFICKÉ
VLASTNOSTI A CYTOLOGICKÉ NÁLEZY**

COMPUTER ANALYSIS OF ULTRASOUND IMAGES OF THYROID NODULES, FOCUSING ON
THEIR SONOGRAPHIC FEATURES AND CYTOLOGICAL FINDINGS

Disertační práce

Školitel:

doc. Ing. MUDr. Daniel Smutek, Ph.D.

Praha, 2019

Prohlášení:

Prohlašuji, že jsem závěrečnou práci zpracoval samostatně a že jsem řádně uvedl a citoval všechny použité prameny a literaturu. Současně prohlašuji, že práce nebyla využita k získání jiného nebo stejného titulu.

Souhlasím s trvalým uložením elektronické verze mé práce v databázi systému meziuniverzitního projektu Theses.cz za účelem soustavné kontroly podobnosti kvalifikačních prací.

V Praze, 30.8.2019

Antonín Procházka

Podpis:

Identifikační záznam:

PROCHÁZKA, Antonín. *Počítačová analýza ultrazvukových obrazů uzlů ve štítné žláze se zaměřením na jejich sonografické vlastnosti a cytologické nálezy. [Computer analysis of ultrasound images of thyroid nodules, focusing on their sonographic features and cytological findings]*. Praha, 2019. 92 s., 2 příl. Disertační práce. Univerzita Karlova, 1. lékařská fakulta, Ústav biofyziky a informatiky a 3. interní klinika VFN a 1. LF UK v Praze. Vedoucí práce Smutek, Daniel.

Klíčová slova: uzly štítné žlázy, počítačem podporovaná diagnostika, rakovina, ultrazvuk, ultrasonografie, klasifikace, strojové učení

Key words: thyroid nodules, computer-aided diagnosis, cancer, ultrasound, ultrasonography, classification, machine learning

Poděkování

Děkuji svému školiteli doc. Ing. MUDr. Danielu Smutkovi, Ph.D. za odborné vedení v průběhu celého studia, dále pak kolegovi Sumeetu Gulatimu za spolupráci při psaní anglických rukopisů, ale také za jeho pozitivní přístup a nikdy neutuchající nadšení pro vědu, a v neposlední řadě všem zúčastněným zaměstnancům 3. interní kliniky 1. lékařské fakulty Univerzity Karlovy a Všeobecné fakultní nemocnice v Praze, především Ing. Štěpánu Holinkovi, za pomoc při sběru a přípravě dat.

OBSAH

1 Úvod.....	7
1.1 Uzly štítné žlázy.....	7
1.2 Sonografické vyšetření.....	13
1.2.1 Rizikovost sonografických vlastností uzlu	16
1.3 Bethesda systém	19
1.4 CAD systémy	22
1.4.1 CAD systémy založené na klinických datech.....	22
1.4.2. Klasické CAD systémy založené na obrazových datech	25
1.4.3. CAD systémy založené na hlubokém učení.....	27
1.4.4 Patch-based analýza	29
1.5 Náš přístup	31
1.6 Cíle a hypotézy práce	31
2 Metody.....	33
2.1. Data.....	33
2.2 Prediktory	34
2.2.1 Společné prediktory.....	34
2.2.2 Prediktory použité při klasifikaci celého uzlu	35
2.2.3 Prediktory použité pro patch-based klasifikaci.....	39
2.3 Klasifikace	45
2.3.1 Náhodné lesy	45
2.3.2 SVM.....	49
2.3.3 Nevyváženost tříd	50
2.3.4 Validace klasifikačních modelů	51
2.3.5 Výběr prediktorů.....	53

2.4 Technické specifikace hardwaru	53
3 Výsledky	54
3.1 Výsledky klasifikace celých uzlů	54
3.2 Výsledky patch-based metody	56
4 Diskuse	65
5 Závěr	71
Souhrn	73
Summary	74
Použitá literatura	75
Vlastní publikační aktivita	83
A) Publikace týkající se tématu dizertační práce	83
B) Ostatní publikace	84
Přílohy	85
Příloha 1: R skript – patch-based klasifikační modely	85
Příloha 2: R skript – klasifikační model pro celé uzly	91

1 ÚVOD

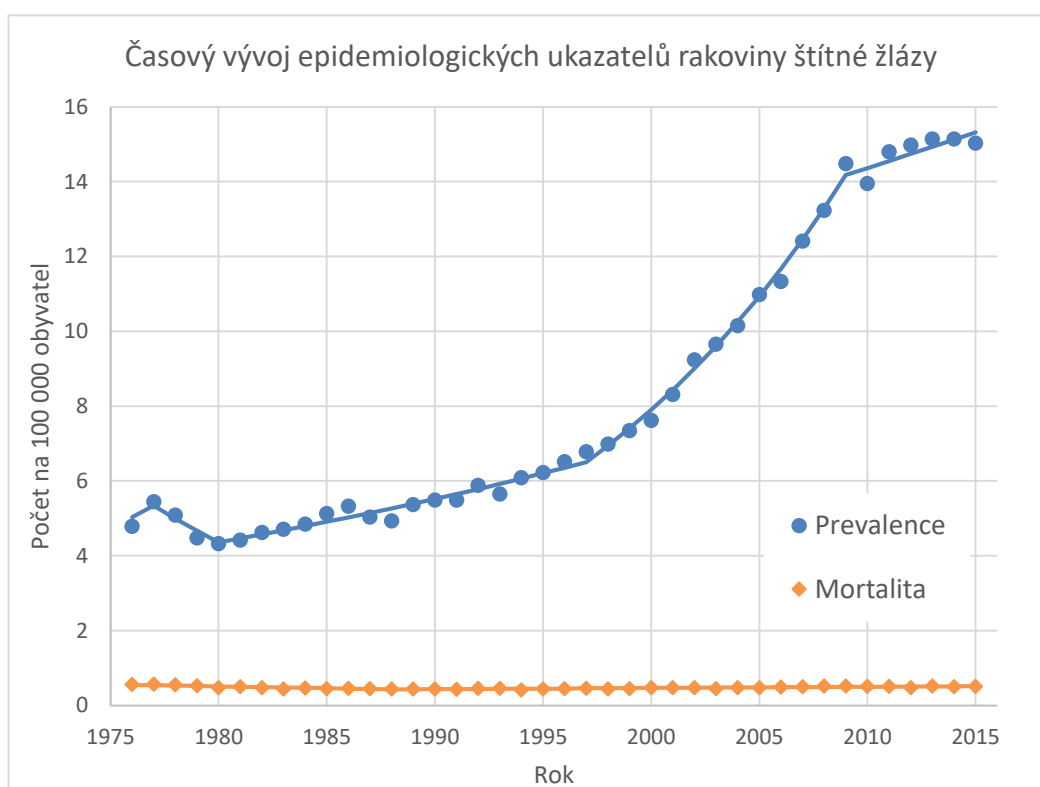
1.1 Uzly štítné žlázy

Hmatné uzly štítné žlázy jsou dlouhodobě běžným klinickým nálezem s prevalencí ve vyspělých částech světa kolem 5 % u žen a 1 % u mužů (Vander et al., 1968, Tunbridge et al., 1977). Ve skutečnosti je však prevalence uzlů ve štítné žláze mnohem vyšší, jelikož se často jedná o asymptomatické onemocnění a uzly nemusí být vždy palpovatelné. U náhodně vybraných jedinců byly uzly štítné žlázy objeveny pomocí vysokofrekvenčního ultrazvuku v 19–68 % případů s tím, že častější výskyt byl u žen a postarších jedinců (Tan and Gharib, 1997, Guth et al., 2009). Z klinického hlediska jsou uzly ve štítné žláze závažné především proto, že mohou být maligního charakteru, což nastává v 7–15 % případů v závislosti na věku, pohlaví, rodinné anamnéze, historii ozáření z přírodních či umělých zdrojů a dalších faktorech (Hegedus, 2004, Haugen et al., 2016, Papini et al., 2002).

K rozlišení maligních uzlů od benigních a ke stanovení rizika onemocnění se využívá několika metod. Kromě biochemického vyšetření je to především ultrazvukové zobrazení uzlů, které dále u podezřelých uzlů doplňuje cytologické vyšetření (označované zkratkou FNAB z angl. *fine needle aspiration biopsy* – aspirační biopsie tenkou jehlou), přičemž výsledky cytologie bývají hlavním podkladem pro rozhodování o případné operaci štítné žlázy.

Maligní nádory štítné žlázy jsou jednou z nejčastějších endokrinních malignit. Jde o autonomní růst buněk pocházejících z folikulárních nebo parafolikulárních buněk štítné žlázy, který může vést k dobře diferencovaným rakovinám, jako je papilární rakovina štítné žlázy (PTC) a folikulární rakovina štítné žlázy (FTC), dále ke špatně diferencovanému anaplastickému karcinomu štítné žlázy (ATC), nebo k medulárnímu karcinomu štítné žlázy (MTC). Diferencovaná papilární nebo folikulární forma tvoří více jak 90 % všech případů rakoviny štítné žlázy (Sherman, 2003). Celkově má rakovina štítné žlázy velmi dobrou prognózu s 98%

pravděpodobností přežití pěti let. Ačkoli se jedná o relativně častý typ rakoviny, tak od snah o zavedení screeningu se ustupuje, protože z dostupných epidemiologických dat je jednoznačně vidět (obrázek 1), že i přestože incidence rakoviny štítné žlázy roste, mortalita se rok od roku prakticky nemění a dlouhodobě se drží na prakticky stejné hodnotě kolem 5 pacientů na 1 000 000 obyvatel (Jegerlehner et al., 2017). To znamená, že ačkoli se daří objevovat stále více maligních případů, tak tento fakt nesnižuje úmrtnost.



Obrázek 1: Vývoj incidence a mortality v posledních letech, během kterých došlo ke všeobecnému rozšíření zobrazovacích metod v klinické praxi. Data z National Cancer Institute (2017).

Zajímavá data ohledně problematiky screeningu jsou k dispozici z Jižní Koreje (Ahn et al., 2014), kde začal být od roku 1999 v rámci vládního programu pro boj s rakovinou nabízen určité části populace screening pro karcinom štítné žlázy pomocí ultrasonografie. Od té doby se v Jižní Koreji dramaticky zvýšil počet nových malignit štítné žlázy ze 4 onemocnění na 100 000 obyvatel v roce 1993 na 69 onemocnění na 100 000 obyvatel v roce 2011, což je zhruba sedmnásobný nárůst. Mortalita karcinomu štítné žlázy se ale v těchto letech nezměnila a držela se

stále mezi 300 až 400 pacienty za rok, z čehož plyne, že se pomocí screeningu nepodařilo snížit počet pacientů, kteří na karcinom štítné žlázy umírají. Zavedení screeningu vedlo pouze k tomu, že se zvýšil počet diagnostikovaných pacientů s karcinomem štítné žlázy, nemělo to ale prakticky vliv na mortalitu tohoto onemocnění. Léčba rakoviny štítné žlázy je většinou chirurgická a může být zatížena komplikacemi, jako je operační hypoparatyroidismus (dočasný 19–38 %, trvalý 0–3 % operací) (Edafe et al., 2014) či poškození laryngeálního nervu (2 % operací) (Ahn et al., 2014). Z těchto důvodů také dle doporučení United States Preventive Services Task Force (USPSTF) není screening pro karcinom štítné žlázy u asymptomatických dospělých doporučen s odůvodněním, že rizika převyšují výhody (Bibbins-Domingo et al., 2017). U rakoviny štítné žlázy se tedy setkáváme s netradiční situací, kdy se zdá, že prognóza této rakoviny závisí na včasné diagnostice jen omezeně. Jinými slovy, máme zde situaci, kdy asymptomatické nediodagnostikované nádory mají výbornou prognózu.

Rostoucí incidence je většinou spojována s dlouhodobým trendem plošného rozšiřování zobrazovacích metod, který vede k tomu, že se zvyšuje počet nově diagnostikovaných maligních uzlů štítné žlázy náhodně objevených při vyšetřeních souvisejících s jinou indikací (např. při ultrazvukovém vyšetření karotid či vyšetření krku počítačovou tomografií). Avšak, jak již bylo řečeno, mortalita se dlouhodobě nemění. Zdá se tedy, že tyto náhodně objevené uzly jsou povětšinou indolentní a jejich diagnostika a následná terapie nezlepšuje celkovou statistiku léčby nádorů štítné žlázy. Z tohoto důvodu se také neuvažuje o plošném screeningu karcinomu štítné žlázy, a dokonce se zmírňují pravidla pro asymptomatické uzly, kdy se odborné společnosti přiklánějí pouze k dlouhodobému monitorování u čím dál většího množství uzlů, které byly dříve indikovány k cytologickému vyšetření či operaci (Haugen, 2017). Trend, jakým se v posledních letech mění přístup k diagnostice uzlů ve štítné žláze, je vidět v tabulce 1.

Doporučení 2009	Doporučení 2015	Potenciální klinické implikace
Indikace k FNAB především na základě velikosti uzlu (> 5 mm) a dále pak na základě sonografických vlastností uzlu	FNAB indikována na základě sonografických vlastností a zároveň na základě velikosti (musí být > 10 mm)	Méně uzlů bude indikováno k FNAB
Doporučený systém klasifikace výsledků FNAB se 4 stupni	Doporučený systém klasifikace výsledků FNAB se 6 stupni	Lépe rozliší pravděpodobnost malignity u pacientů s nejednoznačnou cytologií
Rutiní předoperační zobrazení pomocí doplňujících zobrazovacích metod (MR, CT, PET) není doporučeno	Předoperační zobrazení krku pomocí CT (s kontrastem) nebo MRI je doporučeno pro pacienty s klinickým podezřením na pokročilou malignitu ŠŽ	Zvýší cílenou operativní terapii pro pacienty s agresivním onemocněním
Pouze obecné doporučení pro použití molekulárních markerů	Velmi specifická doporučení pro použití molekulárních markerů	Sníží množství nepotřebných operací
Totální thyreoectomie je doporučena pro všechny papilární karcinomy ŠŽ > 1 cm	Doporučena totální thyreoectomie nebo lobektomie u přesně vybraných pacientů s papilárním karcinomem ŠŽ o velikosti 1–4 cm	Sníží operační hypoparathyroidismus, poškození laryngeálního nervu a potřebu hormonální terapie u některých pacientů s diferencovaným karcinomem ŠŽ

Tabulka 1: Vybrané změny v přístupu k diagnostice a terapii uzlů ve štítné žláze mezi lety 2009 a 2015. Převzato z (Haugen, 2017). Vysvětlivky: FNAB – aspirační biopsie tenkou jehlou, ŠŽ – štítná žláza, MR – magnetická rezonance, CT – počítačová tomografie, PET – pozitronová emisní tomografie.

Z tabulky 1 je vidět, že se ustupuje od toho, aby byl každý suspektní karcinom co nejdříve diagnostikován a následně chirurgicky odstraněn. Velikost uzlu už není jediným kritériem pro selekci uzlů k FNAB (Haugen et al., 2016). Tato změna zohledňuje právě poznatek trvale konstantní mortality a jinými slovy tedy říká, že pokud je uzel menší než 10 mm, tak není vhodné pacienta operovat a vystavovat ho perioperačním a pooperačním komplikacím, i kdyby se jednalo o karcinom štítné žlázy. Další zásadní změnou je, že růst uzlů ve štítné žláze nemusí indikovat malignitu (Durante et al., 2015) a že byla definována specifická doporučení, kdy molekulární markery mohou pomoci ke zpřesnění diagnostiky uzlů s nejednoznačným výsledkem cytologie a pomoci tak k rozhodnutí zda provést či neprovést thyreoectomii (Yip and Sosa, 2016).

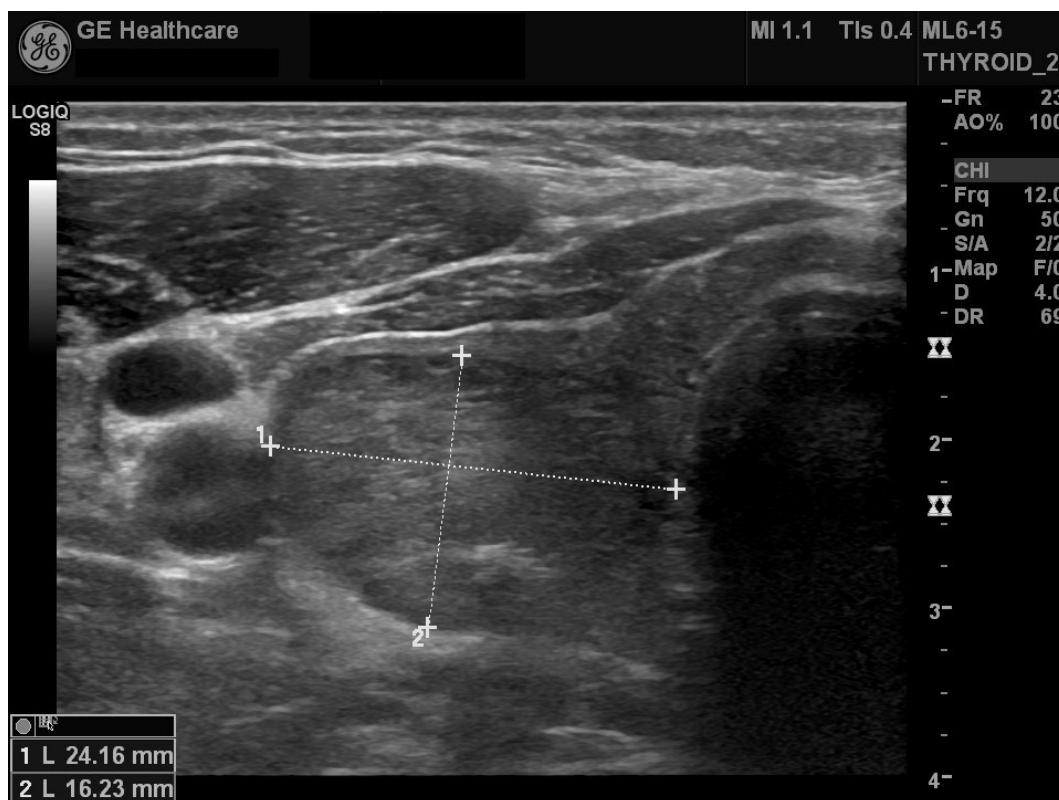
Dlouhodobý monitoring subklinických karcinomů byl zkoumán v několika studiích (Ito et al., 2010, Ito et al., 2014, Sugitani et al., 2010, Miyauchi et al., 2017), jejichž výsledky byly vždy velice podobné: u většiny pacientů se subklinickým papilárním karcinomem nedošlo k progresi onemocnění, zároveň u těch pacientů, kde k progresi (a následné operaci) došlo, nebyl pozorován signifikantně významný nárůst recidivit nebo vzdálených metastáz ve srovnání s pacienty, kteří podstoupili operaci ihned.

Přesto nelze jednoduše říct, že všechny nově diagnostikované případy jsou způsobeny nadměrnou diagnostikou (angl. *overdiagnosis*) v důsledku zvýšené dostupnosti zobrazovacích metod a že možnost včasné diagnostiky nemá žádný pozitivní efekt. Podrobnější analýza epidemiologických dat z USA (Lim et al., 2017) ukázala, že kromě vzrůstající incidence můžeme sledovat také slabý nárůst tzv. mortality založené na incidenci (*incidence-based mortality*) u pokročilejších stádií rakoviny štítné žlázy. Tento ukazatel úmrtnosti bere v potaz (oproti běžné mortalitě) závažnost onemocnění v moment diagnózy (např. histologii, stádium a velikost nádoru) a ve sledovaném období od roku 1994 do roku 2013 ukázal meziroční nárůst diagnostikovaných pokročilých stádií PTC v průměru o 2,9 % ročně, což se jeví v přímé souvislosti s rostoucí incidencí pokročilých stádií PTC (od roku 1981 v průměru o 3,5 % ročně). Tento nárůst výskytu pokročilých stádií PTC a nárůst úmrtnosti na karcinom štítné žlázy u pacientů s diagnostikovaným pokročilým stádiem PTC není v souladu s představou, že za celkový nárůst

incidence všech forem rakoviny štítné žlázy může pouze falešná pozitivita diagnostických testů. Rostoucí celková incidence rakoviny štítné žlázy může být tedy vysvětlena dvěma základními procesy: dominantní je nadměrná diagnostika (*overdiagnosis*) a druhý je malý, ale skutečný nárůst výskytu rakoviny štítné žlázy, na kterém se pravděpodobně v moderním světě podílejí také různé environmentální faktory, genetika a léčebná historie pacientů (Lim et al., 2017). Zatím jsou podle doporučení USPSTF (Bibbins-Domingo et al., 2017) většímu riziku vystaveni pacienti s genetickými syndromy asociovanými s rakovinou štítné žlázy, dále pak dospělí jedinci, kteří v minulosti prodělali nádorové onemocnění, které bylo léčeno radioterapií, což mohlo vést k nadměrnému ozáření štítné žlázy; pro obě tyto skupiny pacientů platí, že by měly být zařazeny do screeningového programu na karcinom štítné žlázy. Ostatní environmentální faktory zůstávají spekulativní (Liu et al., 2017b). Přestože mortalita rakoviny štítné žlázy je konstantně velmi nízká, nemalá část karcinomů štítné žlázy je agresivního charakteru, kdy nádorové buňky expandují do okolí uzlu, mizních uzlin a můžou tvořit vzdálené metastázy. Z těchto důvodů je tedy stále důležité, aby byl nádor objeven dříve, než se začne takovýmto způsobem projevovat.

1.2 Sonografické vyšetření

Sonografické vyšetření krku či přímo štítné žlázy by mělo být provedeno u všech pacientů s podezřelým palpovatelným uzlem ve štítné žláze, nodulární strumou nebo u pacientů s náhodně objevenou abnormalitou v oblasti štítné žlázy při jiném radiologickém vyšetření, jako je počítačová tomografie (CT), magnetická rezonance (MR) či pozitronová emisní tomografie pomocí fluorem značené deoxyglukózy (^{18}F FDG-PET) (Pellerito et al., 2018). Příklady ultrazukového snímku štítné žlázy s uzlem ve dvou běžně zobrazovaných rovinách (axiální a longitudinální) můžeme vidět na obrázku 2 a 3.



Obrázek 2: Ultrazukový snímek levého laloku štítné žlázy v axiální rovině. Lalok je téměř celý vyplněn uzlem (hranice uzlu označeny bílými křížky). Zdroj: archiv autora.



Obrázek 3: Ultrazvukový snímek štítné žlázy s uzlem v longitudinální rovině. Hranice uzlu jsou označeny bílými křížky. Zdroj: archiv autora.

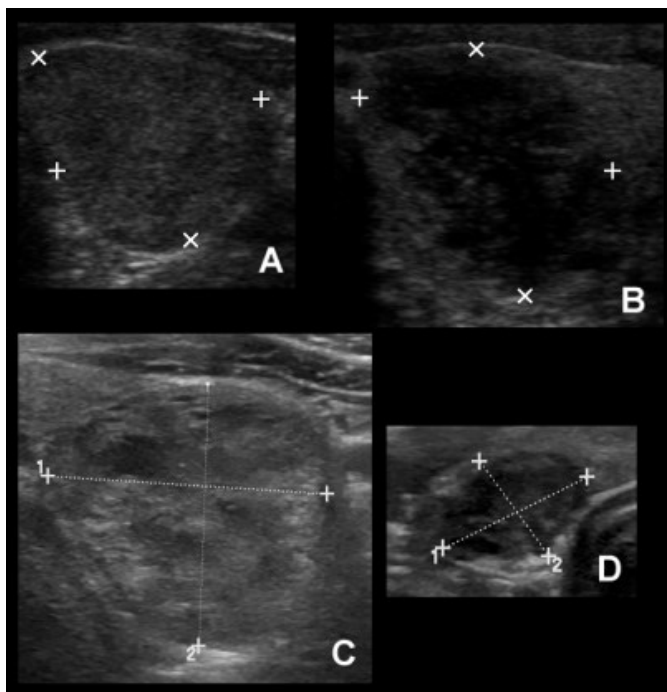
Sonografické vyšetření by mělo dát odpověď na otázky, zda se v oblasti štítné žlázy skutečně vyskytuje uzel a pokud ano, jak je velký, jaká je vnitřní struktura uzlu a jaké jsou jeho další sonografické vlastnosti (Haugen et al., 2016). Dále pak jestli je uzel z více jak z 50 % cystický a jestli je lokalizován hlouběji, což může snížit přesnost FNAB, především pokud není cílená ultrazvukem (Alexander et al., 2002, Hall et al., 1989). Následně by měl lékař pomocí ultrazvuku zhodnotit, zda je parenchym štítné žlázy homogenní či nehomogenní, jaká je velikost štítné žlázy, dále popsat přítomné uzly ve štítné žláze, případně vyloučit změnu struktury mizních uzlin.

Sonografický popis uzlů ve štítné žláze je důležitým diagnostickým podkladem pro odhad rizika malignity uzlu a pro rozhodnutí o nutnosti následného cytologického vyšetření. Sonografický popis uzlu by měl obsahovat: velikost, lokaci a sonografické vlastnosti uzlu, což jsou hlavní podklady pro rozhodování o následné FNAB (Smith-Bindman et al., 2013, Brito et al., 2014).

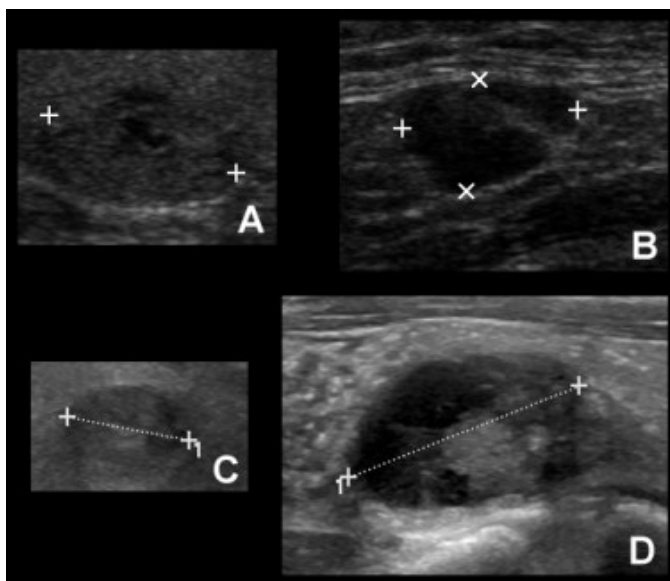
Sonografickými vlastnostmi uzlu je myšleno především:

- a) Charakter uzlu (tuhý / cystický / spongiformní).
- b) Echogenita (zvýšená / normální / snižená).
- c) Ohraničení uzlu (pravidelný / nepravidelný okraj / přítomnost tzv. *halo* okraje).
- d) Přítomnost kalcifikací (mikro- / makrokalcifikace).
- e) Tvar (širší než vyšší / vyšší než širší).

V souvislosti s novými metodami, které hodnotí nové vlastnosti (např. hodnocení elasticity pomocí elastografie či textury pomocí počítačové analýzy obrazu) se tyto sonografické vlastnosti někdy označují jako klasické vlastnosti uzlu. Některé z těchto vlastností jsou spojeny se zvýšeným rizikem malignity, zatímco jiné mohou poměrně přesně indikovat benignitu (Haugen et al., 2016). Mikrokalcifikace, nepravidelný okraj a vyšší než širší tvar jsou vlastnosti s největším rizikem malignity (medián specificity > 90 %), i když střední hodnoty senzitivity jsou pro každou samostatnou vlastnost uzlu výrazně nižší (Papini et al., 2002, Kwak et al., 2011, Moon et al., 2008, Gul et al., 2009). Příklady maligních a benigních uzlů použitých v této práci jsou zobrazeny na obrázku 4 a 5.



Obrázek 4: Příklad benigních a maligních uzlů v axiální rovině. Hranice uzle jsou označeny křížky. A – benigní uzel z přístroje Philips, B – maligní uzel z přístroje Philips, C – benigní uzel z přístroje GE, D – maligní uzel z přístroje GE. Zdroj: archiv autora.



Obrázek 5: Příklad benigních a maligních uzlů v longitudinální rovině. Hranice uzlů jsou označeny křížky. A – benigní uzel z přístroje Philips, B – maligní uzel z přístroje Philips, C – benigní uzel z přístroje GE, D – maligní uzel z přístroje GE. Zdroj: archiv autora.

1.2.1 Rizikovost sonografických vlastností uzlu

Remonti et al. (2015) publikovali meta-analýzu s celkovým počtem 12 786 uzlů, ve které dochází k závěru, že žádná ze sonografických vlastností uzlu není dostatečně spolehlivá na to, aby sama o sobě indikovala malignitu. Mikrokalcifikace, vyšší než širší tvar uzlu, nejasné ohraničení (nepravidelný okraj uzlu) a absence elasticity byly vyhodnoceny jako nejspecifičtější indikátory malignity, přičemž kombinace těchto vlastností riziko malignity zvyšuje. Autoři další meta-analýzy (Brito et al., 2014), která zahrnuje 31 studií a celkem 18 288 uzlů ve štítné žláze, došli k obdobnému závěru, tedy že neexistuje dostatečně spolehlivá vlastnost uzlu, která by byla schopna dostatečně přesně odhalit malignitu i přesto, že kromě výše uvedených vlastností zkoumali navíc i zvýšené prokrvení v uzlu a přítomnost/absenci tzv. *halo* okraje. V souvislosti s tím ale autoři zmiňují dvě vlastnosti, které velmi přesně predikují benignitu. Jedná se o spongioformní vzhled (průměrná specifická 99 %) a dále pak o situaci, kdy je uzel cystický (průměrná specifická 98 %). Také dle doporučení American Thyroid Association (Haugen et al., 2016), neexistuje jediná sonografická vlastnost, která by předpověděla malignitu. Za nejnebezpečnější uzly jsou dle American Thyroid

Association považovány hypoechogenní uzly nebo hypoechogenní část částečně cystických uzlů s jedním nebo více sonografickými rizikovými vlastnostmi nebo s extratyroidální expanzí (riziko malignity 70–90 %). Za středně rizikové jsou pak považovány hypoechogenní uzly s hladkým okrajem bez žádných dalších rizikových vlastností (riziko malignity 10–20 %).

V dalších studiích bylo zkoumáno, jaké přesnosti dosahují radiologové při klasifikaci uzlů podle sonografických vlastností uzlu. Autoři jedné takové studie (Kim et al., 2002) kategorizovali pomocí sonografických vlastností celkem 155 uzlů tím způsobem, že pokud byla v uzlu přítomna alespoň jedna z rizikových vlastností, byl uzel klasifikován jako maligní, a pokud se žádná riziková vlastnost nevyskytovala, pak byl uzel označen jako benigní. Tímto způsobem autoři dosáhli 74,8% klasifikační přesnosti ve srovnání s výsledky cytologie či histologie (u operovaných pacientů). Stejná kritéria (tedy pokud je přítomna jedna riziková vlastnost, pak je uzel označen jako maligní, pokud žádná, pak je uzel označen jako benigní) byla použita i v práci autorů Kim et al. (2015), kde radiologové zhodnotili celkem 633 uzlů štítné žlázy a dosáhli plochy pod křivkou ROC (*area under the curve of receiver operating characteristics*, zkráceně AUC ROC) 0,809. Autoři další studie (Trimboli et al., 2012) uplatňující stejná kritéria dosáhli senzitivity 85 % a 97 % v kombinaci s elastografií při použití 498 uzlů. Autoři další studie (Moon et al., 2007) uplatnili přísnější kritéria (pokud jsou přítomny dvě a více rizikových vlastností, pak je uzel maligní, pokud méně, je benigní) a dosáhli klasifikační přesnosti 93,7 % u 875 pacientů. Navíc u 153 operovaných pacientů byla přesnost klasifikace na základě ultrazvuku srovnatelná s přesností FNAB. Výkon radiologů byl zkoumán také v další studii (Lim et al., 2008) s jednoduchým designem, kdy byli dva radiologové požádáni, aby zařadili uzel do maligní či benigní kategorie jen na základě B-mód ultrazvuku bez přesně definovaných pravidel, která musí uzel splňovat. Plocha pod ROC křivkou byla u prvního 0,83 a u druhého 0,76. Jejich výkon byl srovnán se dvěma klasifikátory (binární logistická regrese a neuronová síť), které dosáhly signifikantně lepšího výsledku ($p < 0,05$) s plochou pod křivkou ROC 0,94 pro binární logistickou regresi a 0,90 pro neuronovou síť.

Na základě dlouhodobého klinického výzkumu byl sestaven systém pro klasifikaci uzlů na základě sonografických vlastností známý pod zkratkou TIRADS (*Thyroid Imaging and Reporting Data System*). Jedná se o poměrně rozsáhlý systém, který se v současné době vyskytuje v několika variantách. V Evropě funguje systém EU-TIRADS (Russ et al., 2017) vydávaný odbornou společností European Thyroid Association. Podle EU-TIRADS můžeme uzly ve štítné žláze na základě sonografických vlastností dělit do pěti kategorií:

1. EU-TIRADS 1: Ve štítné žláze není přítomen žádný uzel.
2. EU-TIRADS 2: Benigní třída. Zcela cystický (anechogenní) uzel nebo spongiformní vzhled. Riziko malignity blízké 0 %.
3. EU-TIRADS 3: Oválný tvar, pravidelný hladký okraj, izoechogenní nebo hyperechogenní, bez dalších rizikových vlastností. Riziko malignity 2–4 %.
4. EU-TIRADS 4: Oválný tvar, pravidelný hladký okraj, mírně hypoechogenní, bez dalších rizikových vlastností. Riziko malignity 6–17 %.
5. EU-TIRADS 5: Uzly s alespoň jednou z následujících vysoce rizikových vlastností: jiný než oválný tvar, nepravidelný okraj, mikrokalcifikace nebo výrazná hypoechogenita. Riziko malignity 26–87 %.

Jak je vidět, ultrazvukové vyšetření uzlů ve štítné žláze je naprosto zásadní při odhadování rizika malignity a pro určení dalšího diagnostického postupu. Při zvažování cytologického vyšetření se přihlíží nejen k EU-TIRADS kategorii, ale také k velikosti uzlu. U tříd EU-TIRADS 1 a 2 se FNAB zpravidla neprovádí, u třídy EU-TIRADS 3 by FNAB měla být provedena při velikosti uzlu nad 20 mm, u třídy EU-TIRADS 4 při velikosti uzlu nad 15 mm a u třídy EU-TIRADS 5 při velikosti uzlu nad 10 mm. U nejvyšší třídy navíc platí, že pokud je výsledek FNAB benigní, cytologické vzorky by měly být do tří měsíců odebrány znovu za účelem vyvrácení možné falešné negativy prvního vyšetření (Russ et al., 2017).

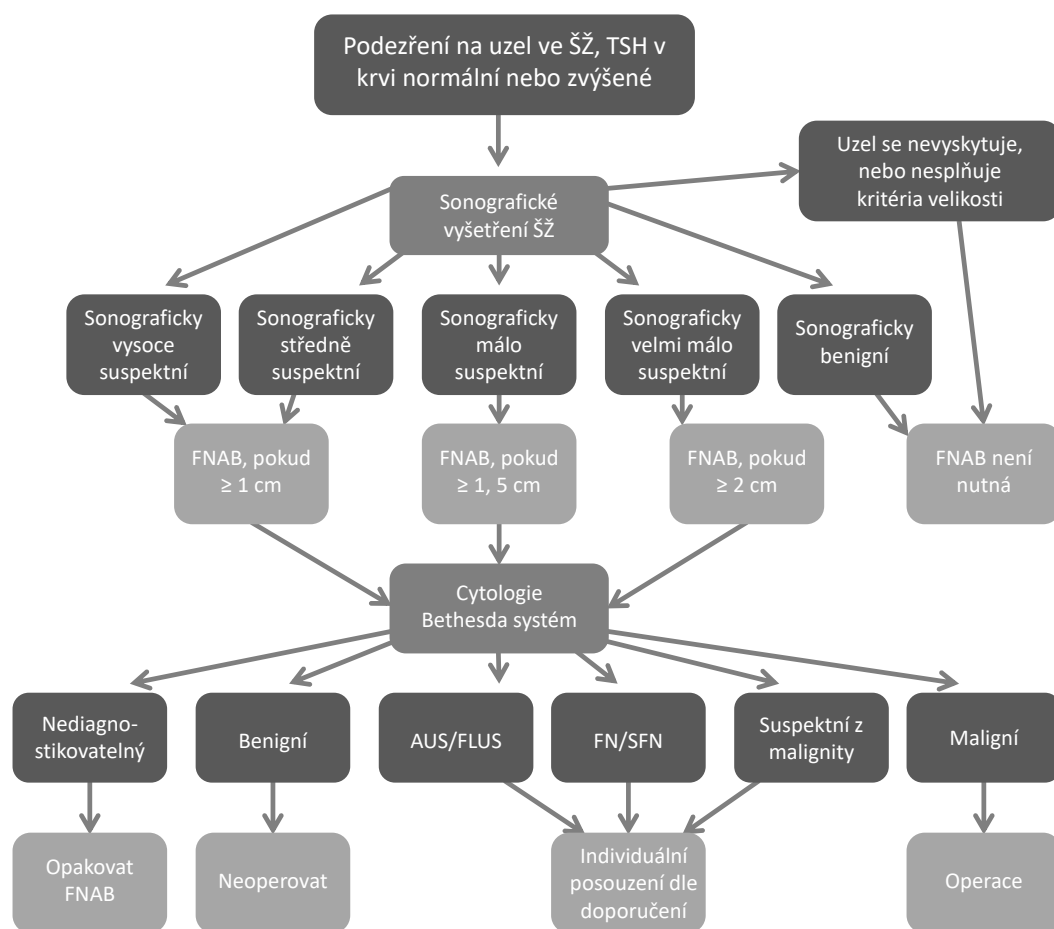
1.3 Bethesda systém

FNAB je stále jednou z nejpodstatnějších částí diagnostického procesu uzlů ve štítné žláze, kdy výsledek cytologie je zásadní při rozhodování o provedení lobectomie či thyreoectomie. V roce 2007 byla zavedena a později pak aktualizována jednotná terminologie pro hodnocení FNAB výsledků, Bethesda systém. V dnešní době Bethesda systém rozeznává 6 různých cytologických kategorií a poskytuje odhad rizika rakoviny v každé kategorii na základě přehledových článků tehdejší literatury (Baloch et al., 2008, Crippa et al., 2010). Tyto kategorie jsou:

- Bethesda 1: zcela nediodagnostikovatelný nález
- Bethesda 2: benigní uzel
- Bethesda 3: atypický uzel s nejasnou významností / folikulární léze s nejasnou významností (AUS/FLUS)
- Bethesda 4: folikulární novotvar / suspektní z folikulární neoplazie (FN/SFN)
- Bethesda 5: suspektní z malignity (SUSP)
- Bethesda 6: maligní uzel

Nedávné studie, ve kterých byl použit Bethesda systém na velkém množství pacientů (Theoharis et al., 2009, Luu et al., 2011, Bongiovanni et al., 2012, Nayar and Ivanovic, 2009), ukázaly poměrně velkou shodu v kategorizaci cytologických nálezů, kdy diagnostikovatelných bylo 89–95 % uzlů (Bethesda 2–6), z čehož 55–74 % jednoznačně benigních (Bethesda 2) a 2–5 % jednoznačně maligních (Bethesda 6). Ostatní, takzvaně cytologicky neurčité kategorie, byly zastoupeny následovně: Bethesda 3 v 2–18 % případů, Bethesda 4 ve 2–25 % případů a Bethesda 5 v 1–6 % případů. Co se týče rizika malignity v jednotlivých kategoriích, tyto studie sice vykazují poměrně velkou variabilitu, ale i tak se povětšinou shodují s odhadovaným rizikem z roku 2008 s výjimkou Bethesda 3 kategorie, kde tyto studie ukázaly signifikantně vyšší riziko malignity než původní odhadované riziko z roku 2008. Bethesda systém se ukázal jako účinné schéma pro klasifikaci uzlů štítné žlázy a odhadu rizika s těmito uzly spojenými. Přesto ale stále až 11 % všech FNAB zůstává nediodagnostikovatelných (Bethesda 1) a v průměru

kolem 23 % FNAB výsledků je cytologicky nejednoznačných (Bethesda 3, 4 a 5; dříve jedna kategorie *intermediate*), s rizikem malignity 75 % v kategorii 5, 26 % v kategorii 4 a 16 % v kategorii 3 (Bongiovanni et al., 2012). Výsledky FNAB jsou zásadní pro konečné rozhodnutí, zda štítnou žlázu operovat nebo ne. Postup diagnostiky uzlů ve štítné žláze na základě sonografických vlastností uzlu a FNAB je znázorněn na obrázku 6.



Obrázek 6: Schéma rozhodování při diagnostice uzlů ve štítné žláze podle doporučení American Thyroid Association (Haugen et al., 2016). Vysvětlivky: ŠŽ – štítná žláza, TSH – tyreotropní hormon, FNAB – aspirační biopsie tenkou jehlou, AUS/FLUS – atypický uzel s nejasnou významností / folikulární léze s nejasnou významností, FN/SFN – folikulární novotvar / suspektní z folikulární neoplazie.

Bongiovanni et al. (2012) ve své meta-analýze zpracovávají výsledky 8 studií s celkem 25 445 FNAB diagnostikovanými případy, z nichž celkem 6 362 skončilo operací. Z těchto operací bylo histologicky zhodnoceno, že 2 150 případů bylo skutečně maligních a 4 212 případů bylo benigních. Je jasné, že karcinom štítné žlázy je závažné onemocnění a obecně je důležité uplatňovat preventivní přístup, kterým pacienty chráníme před rizikem spojeným s rozvojem tohoto onemocnění, jako je především postižení dalších orgánů (expanze do okolí štítné žlázy a mízních uzlin, vzdálené metastázy), ale i přesto není optimální, že pouze 33,8 % pacientů, kteří podstoupili operaci, skutečně trpělo maligní formou onemocnění. To je velmi nízká specificita – obzvláště s přihlédnutím k tomu, že operace s sebou nese rizika (paralýza hlasivek udávána u 2 % operací (Ahn et al., 2014), dočasný hypoparatyroidismus u 19–38 %, trvalý hypoparatyroidismus u 0–3 % operací (Edafe et al., 2014)). Bongiovanniho meta-analýza je z roku 2012, takže zřejmě můžeme předpokládat, že se tato statistika mohla od té doby zlepšit díky stále specifitějším kritériím v nových doporučeních (viz tabulka 1), ale i přesto se poměr skutečně maligních případů vůči všem operovaným zdá být neuspokojivý. Není tedy divu, že vedle FNAB, která je v dnešní době zásadní metodou pro indikaci k operaci, se hledají další metody, které by byly schopny vnést do diagnostického procesu více jistoty jak v rozlišení benigních a maligních uzlů (a tím specifikovat indikaci k operaci), tak v rozpoznání potenciální agresivity novotvaru. V současné době se mluví hlavně o neinvazivních metodách, jako jsou nové metody ultrasonografie či objev dalších specifických biomarkerů (Cappola, 2017). Z nových metod ultrasonografie je to především použití kontrastu k zobrazení štítné žlázy, dále pak real-time a shear-wave elastografie a v neposlední řadě systémy počítačové podpory diagnostiky (*computer aided diagnosis*, zkratka CAD), kdy právě poslední zmiňované jsou hlavním tématem této práce. Především u pacientů s opakovaně nediodagnostikovatelými (Bethesda 1), benigními limitovanými, či suspektními uzly (Bethesda 3–5) by mohly tyto nové metody vnést nový pohled a tím poskytnout lékařům další informaci, což může zpřesnit výběr vhodných pacientů pro FNAB či operaci.

1.4 CAD systémy

Role počítačové podpory lékařské diagnostiky je vnímána jako doplněk k rutinní diagnostice. CAD pro štítnou žlázu jsou označovány jako nízkonákladové, efektivní systémy, které mohou nabízet novou diagnostickou informaci, a tím mohou zmenšit intrapersonální odchylky při rutinní diagnostice (Acharya et al., 2014). CAD systémy se nemusí zaměřovat na ty samé vlastnosti, na které se zaměřuje lékař při hodnocení ultrazvukových snímků (jako byly výše uvedené rizikové faktory, např. nepravidelnost okraje, vyšší než širší tvar, atd.). Především texturní CAD systémy využívají informace, které jsou stěží hodnotitelné pouhým okem. Texturní systémy jsou založeny na různých metodách, jako jsou morfologické charakteristiky obrazu, hodnocení opakujících se struktur (*patterns*) či metody založené na matematické a filtrační transformaci obrazu. Acharya et al. (2014) ve svém článku nabízejí přehled v minulosti publikovaných CAD systémů pro štítnou žlázu, ze kterého je zřejmé, že CAD zahrnuje mnoho přístupů a metod založených na různých statistických metodách, pomocí kterých se z ultrazvukového obrazu extrahují důležité prediktory, vlastnosti (*features*). Tyto prediktory jsou pak použity v různých typech klasifikátorů pro diferenciaci maligních a benigních uzlů, přičemž zpravidla se uplatňuje tzv. učení s učitelem (*supervised learning*). Nejběžněji používanou veličinou pro hodnocení efektivity klasifikačních modelů je buďto přesnost (počet správně klasifikovaných pozorování / počet všech pozorování) anebo plocha pod křivkou ROC (AUC ROC). Zde podáváme přehled těch nejdůležitějších prací, které byly na toto téma publikovány. Přehledová tabulka těch prací pro naši práci nejdůležitějších je pak uvedena v diskuzi (tabulka 7).

1.4.1 CAD systémy založené na klinických datech

Dříve než samotná obrazová data se k počítačové podpoře diagnostiky štítné žlázy využívala data klinická, např. koncentrace tyroxinu, trijodtyroninu či tyreotropního hormonu v krvi, dále pak výsledky fyzikálních vyšetření, popřípadě lékařský popis sonografického obrazu. Jedním z prvních CAD systémů pro štítnou žlázu byl v roce 2002 systém založený na neuronových sítích (Ozyilmaz and

Yildirim, 2002), kdy autoři testovali různé architektury neuronových sítí (*multi-layer perceptron - MLP, radial basis function – RBF* a kombinovanou architekturu), přičemž dosáhli nevyšší přesnosti přibližně 88,3 %. Autoři Polat et al. (2007) použili k diagnostice štítné žlázy klasifikační algoritmus imitující funkci lidského imunitního systému, tzv. *Artificial Immune Recognition System (AIRS)* a dosáhli 85% klasifikační přesnosti na datech z UCI Machine Learning Repository (Coomans et al., 1983), dále UCI databáze. Ačkoliv AIRS nepatří mezi nejoblíbenější klasifikační modely, na problém diagnostiky štítné žlázy byl použit ještě jednou, a to v roce 2009, kdy autoři Kodaz et al. (2009) představili *Information Gain Based Artificial Immune Recognition System (IG-AIRS)*, který dosáhl na stejných datech z UCI databáze klasifikační přesnosti 95,9 %. Autoři Keles et al. (2008) vytvořili na stejných datech z UCI databáze expertní systém, který pomocí tzv. neuro fuzzy klasifikace dosáhl klasifikační přesnosti 95,33 %. Nakonec byla data z UCI databáze použita i v práci autorů Dogantekin et al. (2010), kteří pomocí analýzy hlavních komponent a klasifikátoru SVM (*support vector machines, metoda podpurných vektorů*) dosáhli klasifikační přesnosti 97,67 %. Ač jsou všechny výše zmíněné práce zajímavé, tak ve všech se pracuje s klinickými daty a nikoli s ultrazvukovým obrazem a jedná se o práce, které se zabývaly spíše hledáním nejlepších klasifikačních metod, než „těžbou“ důležitých prediktorů. Nicméně, je z přehledu výše uvedených prací vidět, že i problém hledání nejlepšího klasifikačního modelu je zásadní, neboť přesnosti CAD systémů, které byly vytvořeny pro stejná data (data z UCI databáze), se výrazně liší.

To nás přivádí k otázce, zda potřebujeme na podobné poměrně jednoduché učení s učitelem skutečně tolik různých klasifikátorů, nebo jsou některé, které jsou výrazně přesnější? Touto otázkou se již zabývali autoři Fernandez-Delgado et al. (2014), kteří zhodnotili celkem 179 klasifikátorů založených celkem na 17 různých principech, jako jsou například bayesovské neuronové sítě, SVM, rozhodovací stromy a jiné klasifikátory založené na rozhodovacích pravidlech (*rule-based classifiers*), tzv. *boosting, bagging, stacking* klasifikátory, náhodné lesy (*random forests, RF*) a jiné souborové (*ensemble*) metody, lineární modely, kNN (*k-nearest neighbors algorithm, algoritmus k-nejbližších sousedů*), klasifikátory založené na metodě nejmenších čtverců, regresi hlavních komponent, logistické regresi či

adaptivních regresních splines funkcích. Klasifikační modely byly implementovány v jazycích a programovacích prostředích Weka (Frank et al., 2016), R (The R Foundation for Statistical Computing, Wien, Austria), Matlab (MathWorks Inc., Natick, MA, USA) a C/C++ (ISO/IEC, 2018a, ISO/IEC, 2018b) na 121 souborech dat z UCI databáze (tedy i na klinická data pro štítnou žlázu, která byla použita ve výše zmíněných studiích). Jako nejlepší klasifikační model vychází dle autorů náhodné lesy implementované v R pomocí knihovny Caret (Kuhn, 2008), který dosáhl průměrně na všech souborech dat 82% přesnosti, přičemž nejlepší průměrná přesnost (dosažená vždy nejlepším klasifikátorem pro každý soubor dat) byla 86,9 %, z čehož plyne, že v průměru RF dosáhl 94% přesnosti vůči vždy nejlepšímu klasifikátoru na daný soubor dat. Nejlépe po náhodných lesích vyšel klasifikátor SVM implementovaný pomocí LibSVM (Chang, 2011) knihovny pro C (s 92,3% průměrnou přesností vůči vždy nejlepšímu klasifikátoru na daný soubor dat). Celkem pak mezi 20 nejlepšími klasifikátory bylo šest různých implementací náhodných lesů a pět SVM. Autoři v článku bohužel neuvádí klasifikační přesnost na jednotlivých souborech dat, proto není možné porovnat jednotlivé klasifikátory na datech z UCI databáze pro štítnou žlázu, která byla použita ve výše zmíněných studiích. Práce nicméně ukazuje, že právem je SVM jednou z nejpoužívanějších klasifikačních metod na běžné klasifikační problémy, což koresponduje také s výše uvedenou nejvyšší klasifikační přesností na datech z UCI databáze pro štítnou žlázu uvedené u autorů Dogantekin et al. (2010). Ze závěrů práce autorů Fernandez-Delgado et al. (2014) je ale zřejmé, že ještě o něco lepší než SVM by měly být náhodné lesy a že jednou z nejlepších implementací je algoritmus z knihovny Caret pro R. Z tohoto důvodu jsme se rozhodli použít v této práci primárně právě náhodné lesy z knihovny Caret a sekundárně pak velmi často používaný a také velmi přesný klasifikační model SVM.

Výše zmíněné práce se sice zabývaly diagnostikou štítné žlázy, ale vzhledem k tomu, že nepoužívaly obrazová data, tak pro nás zůstávají důležité především pro demonstraci toho, že ne všechny klasifikační metody mohou být stejně vhodné a že i na stejných datech se přesnost jednotlivých metod může výrazně lišit. Pokud se ale chceme podívat na konkrétní práce, které se zabývají celým procesem CAD uzlů ve štítné žláze z obrazových dat, kdy celým procesem myslíme za prvé, „dolování“

prediktorů, za druhé selekci prediktorů a za třetí klasifikaci, pak musíme zmínit především následující práce.

1.4.2. Klasické CAD systémy založené na obrazových datech

Autoři Iakovidis et al. (2010) použili celkem 250 vzorků ultrazvukového obrazu štítné žlázy od celkem 75 pacientů k rozlišení uzlů štítné žlázy s vyšším rizikem malignity proti normální tkáni štítné žlázy. Autoři popisují, že porovnávali vzorek uzlu označený radiologem jako uzel se zvýšeným rizikem malignity a zdravou tkáň vždy od jednoho pacienta, přičemž dosáhli nejlepšího výsledku ROC AUC = 0,975.

Autoři Savelonas et al. (2007) použili ultrazvukový obraz pro klasifikaci zdravé a nodulární štítné žlázy a dále pak pro klasifikaci maligních a benigních uzlů ve štítné žláze na datech od 66 pacientů. Autoři se zaměřili na směrové texturní charakteristiky obrazu (*directional patterns*), které získali Radonovou transformací a pomocí klasifikátoru SVM dosáhli klasifikační přesnosti 89,4 %.

Ding et al. (2011) použili 125 uzlů štítné žlázy od 56 maligních a 69 benigních pacientů k extrakci statistických a texturních prediktorů z elastografického obrazu a pomocí SVM klasifikátoru dosáhli 93,6% klasifikační přesnosti u solidních uzlů (cysty a pseudocysty byly ze studie vyřazeny).

Acharya et al. (2011) použili ultrasonografii s kontrastem (CEUS) k zobrazení uzlů ve štítné žláze u 20 pacientů (10 maligních a 10 benigních) a získané ultrazvukové obrazy byly pak dále použity k diferenciaci mezi benigními a maligními uzly na základě texturních parametrů (energie a entropie) a prediktorů získaných pomocí diskrétní waveletové transformace. Autoři použili kNN (k-nejbližší soused) klasifikátor a dosáhli přesnosti 98 %. Dále pak autoři navrhli index rizika malignity, který může být spočítán z ultrazvukového obrazu a vyjadřuje riziko malignity vyšetřovaného uzlu.

Autoři Acharya et al. publikovali také další studii (2012a), ve které použili opět prediktory založené na waveletové transformaci a některé další texturní parametry získané z B-mód ultrazvukového obrazu. Použili opět 10 maligních a 10 benigních snímků a nově klasifikátor Adaboost, přičemž pro jeho optimalizaci používali ROC AUC a dosáhli 100% klasifikační přesnosti a ROC AUC rovno 1.

V další studii autoři Acharya et al. (2012c) vypočítali velké množství prediktorů jako například Fourierovské descriptory, *local binary patterns*, fraktální dimenzi, či Lawovy texturní energie, které získali jak z CEUS obrazu, tak z klasického B-módu. Získaná data pak použili k diagnostice 10 maligních a 10 benigních pacientů pomocí GMM (Gaussian mixture model) klasifikátoru přičemž dosáhli 98,1% přesnosti.

Chang et al. (2016) použili axiální i longitudinální ultrazvukové obrazy (celkem 30 maligních a 29 benigních) pro extrakci velkého množství texturních prediktorů (celkem 74), jako jsou například parametry histogramu (průměr, směrodatná odchylka, šikmost, špičatost), rozdíly jasu v uzlu a v okolní tkáni, matice spoluvýskytu (*co-occurrence matrix*) a tzv. šedotónové *run-length* matice. Autoři dále představili vlastní segmentační algoritmus, který je schopný s velmi dobrou přesností označit hranice uzlu ve štítné žláze (u většiny snímků označeno správně více jak 90 % uzlu). Segmentované hranice byly použity k proložení elipsy uzlem a dále bylo zkoumáno, jak je elipsa natočená, čímž autoři zkoumali, jestli je uzel vyšší než širší, což je jeden z rizikových faktorů při radiologickém vyšetření uzlů ve štítné žláze. Autoři navíc kromě texturní analýzy zhodnotili 12 radiologických parametrů (6 z axiálního a 6 z longitudinálního obrazu), které odpovídaly rizikovým faktorům podle autorů Moon et al. (2008, 2011). Výsledkem studie je pak 98,3% přesnost CAD systému při použití SVM klasifikátoru vs. 98,7% přesnost hodnocení rizikových parametrů radiologem, přičemž mezi oběma výsledky nebyl pomocí t-testu sledován statisticky významný rozdíl (na hladině významnosti $p < 0,05$). Nejdůležitějšími prediktory CAD systému byly: Natočení elipsy (tzn. hodnocení, zda je uzel vyšší než širší) a maximální pravděpodobnost počítaná z matice spoluvýskytu, oboje z axiálního obrazu.

Autoři Acharya et al. v další studii (2016) použili Gaborovu transformaci pro extrakci zhruba 3 800 prediktorů, které dále redukovali pomocí lokálně citlivé diskriminační analýzy (*locality sensitive discriminant analysis*, LSDA). Autoři použili snímky od celkem 242 pacientů, kdy 211 bylo benigních a 31 maligních. Vzhledem k tomu, že u binárních klasifikátorů založených na algoritmech strojového učení může nevyváženost jednotlivých tříd (zde maligní vs. benigní třída) vést k nereprezentativním výsledkům (Branco et al., 2016), museli se autoři

s tímto problémem vypořádat. Pravděpodobně z důvodu, aby neztratili informace obsažené ve velkém množství benigních obrázků, rozhodli se nevyváženost řešit *over-samplingem*, kdy se vytvoří syntetická data minoritní třídy (maligní), a tím se zastoupení obou tříd vyváží. Jedná se o jeden z možných způsobů, kdy ale musíme mít na paměti, že informace o maligních uzlech jsou duplikovány, a proto výsledný klasifikátor nemusí dostatečně generalizovat pro maligní uzly a je otázka, jak by se klasifikátor choval na nových datech. I přesto se ale jedná o práci prezentující CAD systém pro uzly ve štítné žláze, ve které byl použit jeden z největších souborů dat. Autoři porovnali několik klasifikátorů (SVM, rozhodovací stromy, kNN, MLP neuronové sítě), přičemž dosáhli nejvyšší přesnosti 94,3 % pomocí C4.5 klasifikátoru, jehož princip je založen na rozhodovacích stromech.

Autoři Raghavendra et al. (2017) extrahovali 181 prediktorů z 242 obrázků, přičemž dosáhli nejvyšší klasifikační přesnosti 97,52 % pomocí SVM klasifikátoru. Tato studie je naší práci obzvláště důležitá, protože 120 prediktorů bylo získáno pomocí *Segmentation-based Fractal Texture Analysis* (SFTA) algoritmu. Autoři redukovali počet prediktorů pomocí marginální Fisherovy analýzy (*marginal Fisher analysis*, MFA), která funguje na takovém principu, že ze stávajících prediktorů se vytváří nové MFA prediktory. Autoři bohužel v práci neprezentují, jaké primární prediktory sloužily k vytvoření těch nejvýznamnějších MFA prediktorů, proto je obtížné hodnotit, jak velký význam na konečnou klasifikaci měly prediktory získané právě pomocí SFTA algoritmu.

1.4.3. CAD systémy založené na hlubokém učení

V roce 2017 byla publikována ještě jedna velmi zajímavá práce autorů Chi et al. (2017). Tato práce je zajímavá především proto, že ukazuje nový přístup ke klasifikaci obrazových dat a dost možná nový budoucí směr CAD nejen štítné žlázy. Doteď jsme psali o CAD systémech jako o klasifikačních modelech, kterým je potřeba dodat vhodné prediktory, které je potřeba předtím najít, „vydolovat“, přičemž způsob této „těžby“ spočívá na uvážení autora CAD systému. Autor, jako člověk, se snaží najít nejvhodnější prediktory, vlastnosti, na základě kterých půjde data vhodně rozdělit do jednotlivých kategorií. Jak je vidět na předchozím přehledu,

u uzlů ve štítné žláze má několik možností – buďto se zaměří na radiologické rizikové faktory, nebo hledá vhodnou transformaci obrazu, či sází na texturní, šedotónové nebo jiné vlastnosti obrazu. V poslední době se ale jako efektivnější ukazuje metoda, při které sám algoritmus hledá ony vhodné prediktory. Jedná se o metodu hlubokého učení (*deep learning*), kdy vstupem pro klasifikátor nemusí být vektor prediktorů vytěžených v předchozím kroku, ale jen pixely vstupního obrazu. Hledání vlastností, prediktorů, v obraze je pak ponecháno na samotném algoritmu, který pokud má dostatek tréninkových dat (data, u kterých je známa příslušná třída), tak může metodou založenou na opakovaném upravování vah a zpětné kontrole najít takové vlastnosti, které už jen těžko reprezentují nějaké konkrétní, slovy popsatelné, vlastnosti obrazu, přesto však mohou být velmi účinné pro klasifikaci obrazu do příslušných kategorií. Hluboké učení se začalo prosazovat především po roce 2012, kdy Krizhevsky et al. (2012) s metodou hlubokého učení založenou na konvolučních neuronových sítích s přehledem vyhráli nejznámější mezinárodní soutěž na téma obecného rozpoznávání objektů v obraze, ImageNet Large Scale Visual Recognition Challenge (Russakovsky et al., 2015). Od té doby probíhá mohutný rozvoj hlubokého učení v mnoha oblastech poznání. K hlubokému učení se používají především neuronové sítě, které byly použity i v práci autorů Chi et al. (2017). Aby ale neuronová síť našla relevantní prediktory, musí být trénována na obrovském množství dat, což je také hlavní problém při použití v klinickém výzkumu, kde bývají data většinou nesrovnatelně menší, než bychom k tomuto účelu potřebovali. Autoři Chi et al. (2017) tento problém obcházejí tím, že používají neuronovou síť GoogLeNet, která je „předtrénovaná“ na velkém objemu dat, kde tréninkovou úlohou je rozpoznávání obecných předmětů na běžných fotografiích. Autoři tento předem naučený GoogLeNet model „vyladili“ (učení pouze několika nových vrstev, angl. *fine-tuning*) pro získání prediktorů použitelných pro diagnostiku uzlů ve štítné žláze. Získané prediktory pak autoři klasifikovali pomocí náhodných lesů, kdy na dvou souborech dat dosáhli přesnosti 98,29 % (soubor dat s 357 obrázky) a 96,34 % (soubor dat se 164 obrázky). Výsledky ukazují, že metoda deep learningu přináší velmi slibné výsledky při klasifikaci uzlů ve štítné žláze i přes to, že byl použit velmi obecný model, který byl trénován pro obecné rozpoznávání objektů v běžných fotografiích

a pro potřeby diagnostiky štítné žlázy byl pouze vyladěn pomocí poměrně malého souboru dat. Jedná se o přístup, který je časově mnohem úspornější než trénovat algoritmus hlubokého učení od začátku, a navíc obcházíme potřebu velké tréninkové množiny dat. Zdá se tedy, že máme pro uzly ve štítné žláze nový směr, který vypadá velmi slibně. Zmíněný článek ale bohužel skýtá jednu důležitou nejasnost. Autoři se zmiňují, že řešili nevyváženost tříd při získávání prediktorů, chybí ale zmínka o tom, zda řešili nevyváženost tříd také při klasifikaci. Pokud autoři nevyváženost souboru při klasifikaci neřešili, pak výsledky těžko můžeme považovat za relevantní.

Hluboké učení bylo použito i v dalších pracích (Mei et al., 2017, Liu et al., 2017a), kdy autoři zkoušeli pomocí této metody extrahovat prediktory a pak tyto prediktory dále kombinovat s jinými, tradičními prediktory, jako jsou *local binary patterns* nebo šedotónové matice spoluvýskytu (*co-occurrence matrices*). Tento postup, kdy se klasický způsob získávání prediktorů kombinuje s hlubokým učením, je alternativou k předchozímu postupu využívajícímu *fine-tuning* předtrénovaných neuronových sítí. Autoři Liu et al. (2017a) získali pomocí konvoluční neuronové sítě s architekturou VGG-F předtrénované na datech z ImageNet databáze (Deng et al., 2009) celkem 4 096 prediktorů a tyto prediktory pak fúzovali s výše zmíněnými tradičními prediktory, čímž získali celkem 4 778 prediktorů. Autoři zkoušeli různé kombinace prediktorů, které použili ke klasifikaci a následně porovnávali výsledky, přičemž dosáhli na 651 benigních a 386 maligních snímcích nejvyšší přesnosti 93,1 % při použití kombinace prediktorů hlubokého učení s tradičními prediktory. Použití různých kombinací pouze prediktorů hlubokého učení dosahovalo o 7–8 % nižší přesnosti než nejlepší výsledek 93,1 % a použití různých kombinací pouze tradičních prediktorů mělo ve srovnání s nejlepším výsledkem přesnost nižší o 10–14 %.

1.4.4 Patch-based analýza

Počítačová podpora diagnostiky uzlů ve štítné žláze má poměrně krátkou, avšak pestrou historii, a za tu dobu bylo použito velké množství metod k těžbě prediktorů a k následné klasifikaci. Naprostá většina výše zmíněných CAD systémů pracuje

s uzlem jako celkem, pouze v práci autorů Savelonas et al. (2007) se pracuje s malým výsekem ultrazvukového obrazu o velikosti 32×32 pixelů, přičemž ale není jasné, jak byly tyto čtverečky vybírány v rámci uzle (jestli zprostřed, zkraje, náhodně, či byly vybrány radiologem jen reprezentativní části uzlu). Také konvoluční neuronové sítě používané při hlubokém učení se zaměřují na menší část obrazu (podle velikosti konvoluční masky), jinak se ale většina výše zmíněných prací zabývá uzlem jako celkem a zdá se, že metoda patch-based analýzy, tedy analýzy obrazu po malých čtverečkách, zatím není dostatečně prozkoumána. Výhodou patch-based přístupu je, že získáváme velké množství dat – v závislosti na velikosti čtverečků to v našem případě bylo několik jednotek až desítek tisíc obrazových čtverečků – a v oblastech strojového učení obecně platí, že čím více pozorování, tím lépe klasifikační modely generalizují, tedy pracují přesněji s novými daty, jak ukázaly studie z oblasti umělé inteligence na velkých datech (Alon Halevy, 2009, Hays and Efros, 2007). Rozhodli jsme se proto zaměřit se, kromě klasifikace uzlů jako celku, také na patch-based klasifikaci.

1.5 Náš přístup

Experimentální část této práce má dvě části. První část se zabývá klasifikací uzlu jako celku. Zde jsme na základě literatury vybrali několik prediktorů, které byly úspěšně použity na větším množství dat. Zároveň jsme pak hledali takové prediktory, které nebudou popisovat směrové charakteristiky, tzn. budou použitelné jak na axiální, tak longitudinální obraz. Z dříve publikovaných prediktorů tyto podmínky splňují především prediktory získané pomocí SFTA algoritmu z poslední práce autorů Raghavendra et al. (2017) a parametry histogramu, které byly dříve použity u (Kim et al., 2015, Chang et al., 2016). Druhá část této práce posouvá výzkum v oblasti CAD systémů o něco dál, jelikož se zabývá patch-based klasifikací, kdy je snímek rozdělen do malých čtverečků, které jsou použity k návrhu (trénování) klasifikátoru. Prediktory v druhé části jsme získávali podobně jako v první části, tedy pomocí parametrů histogramu a SFTA algoritmu, který je zde ale výrazně upraven a vrací se spíše k metodě, ze které SFTA vychází, k tzv. binární stack-dekompozici, BSD (více v kapitole 2.2.3.1 Binární stack dekompozice). Patch-based přístup jsme pak doplnili o 7 dalších prediktorů vhodných pro hodnocení struktur v binárním obraze.

1.6 Cíle a hypotézy práce

Cílem práce je vyvinout dva dostatečně přesné CAD systémy pro štítnou žlázu, které budou pracovat pouze se směrově nezávislými vlastnostmi obrazu a budou tedy použitelné jak v axiální, tak longitudinální zobrazovací rovině. Oba CAD systémy budou pracovat na základě *Segmentation-based Fractal Texture Analysis* (SFTA) algoritmu (případně jeho variaci), čímž bude obraz rozložen na jednotlivá šedotónová pásma. Prediktory budou zvoleny takové, aby komplexně popsaly region v binárním obraze v každém šedotónovém pásmu. Zásadní rozdíl mezi systémy bude, že první CAD systém bude pracovat s celým uzlem a druhý CAD systém bude založen na patch-based přístupu.

Hypotézy práce jsou následující:

- Předpokládáme, že SFTA algoritmus, či jeho variace, bude dostatečně silný a dovedeme pomocí něj extrahovat z ultrazvukového obrazu štítné žlázy dostatečně významné prediktory, které povedou k dostatečně přesné klasifikaci benigních a maligních uzlů štítné žlázy.
- Metoda SFTA, či její variace, bude použitelná jak na axiální, tak na longitudinální snímky.
- Předpokládáme, že nejdůležitější šedotónová pásma budou jednak ty úplně nejtmavší části obrazu (snížená echogenita maligních uzlů) a jednak ty úplně nejsvětlejší části obrazu (výskyt mikro-/makrokalcifikací).
- Patch-based klasifikace bude dostatečně přesná, i přes to, že pracuje pouze s malými kousky obrazu a nebere v potaz uzel jako celek (nepracuje ani s velikostí ani s tvarem, okrajem či jinými rizikovými vlastnostmi uzlu).
- Předpokládáme (především u patch-based analýzy), že ne všechny získané prediktory budou významné. Prediktory spolu budou korelovat, a proto bude vhodné použít metodu zpětné eliminace prediktorů (*Recursive Feature Elimination – RFE*) k výběru těch nejvýznamnějších prediktorů.
- Co se použitých klasifikačních modelů týče, předpokládáme, že náhodné lesy budou vykazovat podobné výsledky jako SVM, případně budou i o něco přesnější.

2 METODY

2.1. Data

Obrazová data byla sbírána na 3. interní klinice Všeobecné fakultní nemocnice (VFN) v Praze a 1. lékařské fakulty Univerzity Karlovy po dobu dvou let. V souladu s uděleným souhlasem Etické komise VFN v Praze byla získána data od 20 pacientů s maligními uzly (ověřeno pomocí FNAB nebo histologicky), která byla doplněna daty od 40 pacientů s benigními uzly (ověřeno FNAB), z čehož bylo celkem 11 mužů a 49 žen. Průměrný věk byl 58 let v benigní třídě a 51 let ve třídě maligní. Obrazová data byla pořízena dvěma ultrasonografickými přístroji: 10 maligních a 25 benigních uzlů pomocí přístroje Logiq S8 výrobce GE Healthcare (GE Healthcare Inc., Chicago, Illinois, United States) s lineární sondou ML6-15-D (50 mm, 4–13 MHz); a 10 maligních a 15 benigních uzlů pomocí přístroje EnVisor M2540A výrobce Philips (Philips Inc., Amsterdam, Netherlands) s lineární sondou L12-5. Obrazové analýze nepředcházelo žádné předzpracování obrazu, kromě zvýšení jasů plošně o 20 jasových úrovní pro všechny obrazy z přístroje Philips. Jinak byly obrazy použity v takovém stavu a kvalitě, jak byly pořízeny lékařem a následně uloženy v nemocničním informačním systému. V experimentální části práce byla k patch-based analýze použita data všech 60 pacientů. K analýze celých uzlů byla použita všechna data z maligní třídy, ale pouze polovina dat ze třídy benigní, přičemž rozložení benigních uzlů bylo takové, že polovina uzlů (10 uzlů) pocházela z přístroje Philips a polovina (10 uzlů) pocházela z přístroje GE (věkový průměr v takto redukované benigní třídě byl 60 let).

2.2 Prediktory

V této kapitole popíšeme veškeré prediktory, které byly použity k následné klasifikaci. Jedná se o společné prediktory, které byly použity jak ke klasifikaci uzlů jako celku, tak k patch-based klasifikaci, dále pak o prediktory získané pomocí SFTA algoritmu, publikované Costou et al. (2012), které byly použity ke klasifikaci celého uzlu, a nakonec prediktory založené na variaci SFTA algoritmu, na binární stack-dekompozici, ke kterým jsme kromě základních tří prediktorů prezentovaných Costou et al. přidali další nesměrové charakteristiky binárního obrazu, které zatím, pokud je nám známo, nebyly k analýze ultrazvukového obrazu použity.

2.2.1 Společné prediktory

Pro oba dva postupy byly použity parametry histogramu jako základní ukazatele rozložení jasu v obraze. Z histogramu byly vypočteny následující prediktory: průměr, směrodatná odchylka, šikmost a špičatost. Histogram šedotónového obrazu zobrazuje zastoupení jednotlivých jasových hodnot v obraze, tedy relativní četnosti $P(h)$ jasových hodnot h , přičemž $h \in \{1, 2, \dots, n_l\}$, kde n_l značí počet jasových úrovní šedotónového obrazu.

Průměr (značen HistMean , \bar{h}) představuje střední hodnotu jasu v histogramu, tedy průměrnou hodnotu jasu, což můžeme vyjádřit jako (Habibzadeh et al., 2013):

$$\text{HistMean} = \bar{h} = \sum_{h=1}^{n_l} h P(h) \quad . \quad (2.1)$$

Směrodatná odchylka (značena HistStdDev , σ) představuje směrodatnou odchylku jasu v obraze. Vyjadřuje tedy jak úzký či široký je histogram a lze ji vyjádřit jako (Habibzadeh et al., 2013):

$$HistStdDev = \sigma = \sqrt{\sum_{h=1}^{n_l} (h - \bar{h})^2 P(h)} \quad . \quad (2.2)$$

Šikmost (značena HistSkew) je indikátor asymetrie histogramu vzhledem k průměru. Pokud je histogram symetrický, hodnota je rovna nule. Pokud je histogram sešikmený doleva, hodnota je záporná a pokud je histogram sešikmený doprava, hodnota je kladná. Šikmost můžeme vyjádřit jako (Habibzadeh et al., 2013):

$$HistSkew = \frac{1}{\sigma^3} \sum_{h=1}^{n_l} (h - \bar{h})^3 P(h) \quad . \quad (2.3)$$

Veličina špičatost (značena HistKurt) popisuje, zda je histogram plochý nebo špičatý, což lze vyjádřit jako (Habibzadeh et al., 2013):

$$HistKurt = \frac{\sum_{h=1}^{n_l} (h - \bar{h})^4}{n_l \sigma^4} \quad . \quad (2.4)$$

Záporná hodnota špičatosti indikuje ploché rozdělení, kladná hodnota indikuje špičaté rozdělení.

2.2.2 Prediktory použité při klasifikaci celého uzlu

Pro klasifikaci celého uzlu byl použit SFTA (*Segmentation-based Fractal Texture Analysis*) algoritmus, který byl představen autory Costa et al. (2012) a byl úspěšně použit na štítnou žlázu autory Raghavendra et al. (2017). Algoritmus rozkládá šedotónový obraz na sérii binárních obrazů pomocí dvouprahové binární dekompozice (*Two-Threshold Binary Decomposition*), která vychází z dříve publikované metody nazvané binární stack-dekompozice (*Binary Stack Decomposition*) (Chen et al., 1995). SFTA algoritmus vychází z předpokladu, že vstupní šedotónový obrázek můžeme modelovat jako 2D funkci $I(x, y)$, kde (x, y)

jsou prostorové souřadnice pixelu. Pro hodnotu h obrazové funkce $I(x, y)$ platí $h \in \{1, 2, \dots, n_l\}$, kde n_l značí počet jasových úrovní šedotónového obrazu. SFTA algoritmus dělí obrázek pomocí dvouprahové binární dekompozice na sérii binárních obrazů $I_b(x, y)$, které se vytváří pomocí prahování obrazu dvěma prahy, přičemž hodnota příslušného prahu t je určena na základě distribuce jasu v obraze pomocí Otsuovy metody (Liao et al., 2001, Otsu, 1979), což je algoritmus, který rozdělí obrazové body do dvou tříd, do třídy popředí (třída C_1) a do třídy pozadí (třída C_0). Otsuova metoda k optimálnímu rozdělení používá výpočet rozptylu třídy popředí $\sigma_1^2(t)$ a třídy pozadí $\sigma_0^2(t)$, přičemž optimální práh může být nalezen minimalizací hodnoty vnitřního rozptylu $\sigma_w^2(t)$, která je dána vztahem:

$$\sigma_w^2(t) = w_0(t)\sigma_0^2(t) + w_1(t)\sigma_1^2(t) \quad , \quad (2.5)$$

kde váhy $w_0(t)$ a $w_1(t)$ vyjadřují pravděpodobnosti výskytu třídy pozadí a popředí, které jsou pro práh t počítány z n_l jasových hodnot následujícím způsobem.

$$w_0(t) = \sum_{i=1}^t P(i) \quad , \quad (2.6)$$

$$w_1(t) = \sum_{i=t+1}^{n_l} P(i) \quad . \quad (2.7)$$

Pro jednodušší výpočet se pro dvě třídy využívá toho, že minimalizace vnitřního rozptylu (rovnice 2.5) je rovna maximalizaci mezi-rozptylu, který lze spočítat následovně:

$$\sigma_B^2(t) = w_0(t)w_1(t)[\mu_0(t) - \mu_1(t)]^2 \quad , \quad (2.8)$$

kde $\mu_0(t)$, $\mu_1(t)$ jsou třídni průměry dané jako:

$$\mu_0(t) = \sum_{i=1}^t i \frac{p(i)}{w_0(t)} \quad , \quad (2.9)$$

$$\mu_1(t) = \sum_{i=t+1}^{n_l} i \frac{p(i)}{w_1(t)} \quad . \quad (2.10)$$

Otsův Algoritmus lze tedy shrnout do následujících kroků:

1. Z histogramu se určí pravděpodobnosti pro každou jasovou úroveň.
2. Nastaví se iniciační hodnoty $\mu_0(t)$ a $\mu_1(t)$.
3. Proběhne cyklus přes všechny možné prahy, tj. od $t = 1$ do $t = n_l$, kdy v každém opakování cyklu je vypočítána hodnota $\sigma_B^2(t)$ pro příslušnou jasovou úroveň.
4. Optimální práh odpovídá maximální hodnotě $\sigma_B^2(t)$.

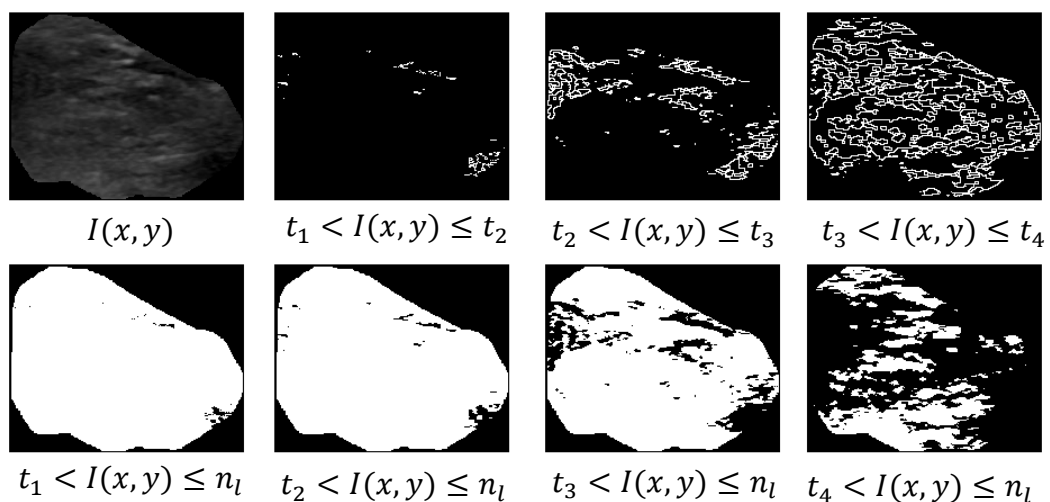
V případě SFTA algoritmu se nejedná o obyčejné prahování obrazu pomocí jednoho prahu, ale o opakované prahování pomocí celkového počtu prahů n_t . Otsuova metoda je tedy v případě SFTA algoritmu použita k určení množiny optimálních hodnot všech prahů z celkového počtu prahů n_t . Máme-li tedy předpokládanou množinu prahů $\{t_1, t_2, \dots, t_{n_t}\}$, kde n_t je celkový počet prahů, pak Otsuovou metodou získáme $n_t + 1$ tříd: C_1 pro $[1, \dots, t_1]$, C_2 pro $[t_1 + 1, \dots, t_2]$, \dots , C_i pro $[t_{i-1} + 1, \dots, t_i]$, \dots , C_{n_t+1} pro $[t_{n_t} + 1, \dots, n_l]$. Optimální prahy $\{t_1^*, t_2^*, \dots, t_{n_t}^*\}$ jsou pak určeny opět maximalizací funkce $\sigma_B^2(t)$, tedy:

$$\begin{aligned} \{t_1^*, t_2^*, \dots, t_{n_t}^*\} = \max \{ \sigma_B^2(t_1, t_2, \dots, t_{n_t}) \} \\ 1 \leq t_1 < \dots < t_{n_t} < n_l \quad , \quad (2.11) \end{aligned}$$

Takto získanou množinu všech prahů rozšíříme o maximální hodnotu jasové úrovně šedotónového obrazu n_l , tedy $\{t_1, t_2, \dots, t_{n_t}\} \cup \{n_l\}$, a použijeme k dvouprahové segmentaci, kdy se požívají vždy dva sousední prahy (spodní práh t_s a horní práh t_h , přičemž $t_s < t_h$) k získání jednoho binárního obrazu $I_b(x, y)$ tak, že platí:

$$I_b(x, y) = \begin{cases} 1 & \text{pro } t_s < I(x, y) \leq t_h \\ 0 & \text{jinak} \end{cases} \quad (2.12)$$

Tímto způsobem získáváme n_t binárních obrazů. Dalších n_t binárních obrazů pak získáváme pomocí dvouprahové segmentace (podle rovnice 2.12) pomocí všech párů prahů $\{t, n_l\}$, $t \in \{t_1, t_2, \dots, t_{n_t}\}$, čímž získáváme dalších n_t binárních obrazů. V případě páru $\{t_{n_t}, n_l\}$ se ale jedná o sousední prahy, a tak získáváme stejný binární obraz jako při dvouprahové segmentaci popsané výše se spodním prahem $t_s = t_{n_t}$ a horním prahem $t_h = n_l$. Tento duplicitní obraz tedy vypouštíme a získáváme celkem $2n_t - 1$ binárních obrazů. Příklad rozkladu obrazu pomocí dvouprahové binární dekompozice pro $n_t = 4$ můžeme vidět na obrázku 7.



Obrázek 7: Výstup dvouprahové binární dekompozice, která je součástí SFTA algoritmu na příkladu s počtem prahů $n_t = 4$ (značeno t_1 až t_4) určených pomocí Ostuovy metody. Vlevo nahoře: Příklad benigního uzlu. Dále pak v horní řadě: Výsledné binární obrázky získané dvouprahovou segmentací pomocí dvou sousedních prahů. Spodní řada: výsledek dvouprahové segmentace pro všechny páry prahů $\{t, n_l\}$, $t \in \{t_1, t_2, \dots, t_{n_t}\}$. Celkem pomocí SFTA algoritmu získáváme $2n_t - 1$ obrazů, tedy celkem 7 binárních obrazů.

SFTA algoritmus pak v každém binárním obraze získaném výše popsanou dvouprahovou segmentací určí dva prediktory: celkový počet pixelů (Area) a fraktální dimenzi (FracDim) bílých pixelů. Fraktální dimenze FracDim je

v případě algoritmu SFTA počítána jako numerický odhad 2D Hausdorffovy dimenze D_0 , kterou pro objekt v eukleidovském prostoru E můžeme počítat jako:

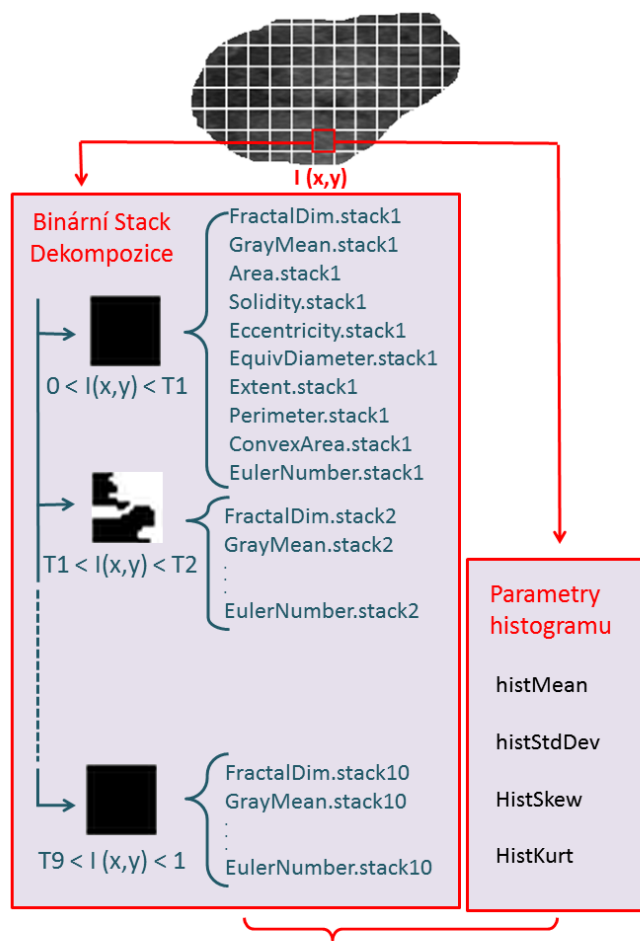
$$D_0 = \lim_{\varepsilon \rightarrow 0} \frac{\log N(\varepsilon)}{\log \varepsilon^{-1}}, \quad (2.13)$$

kde $N(\varepsilon)$ je obecně počet dílů, na který dělíme objekt a ε je měřítko. Konkrétně pak pro dvourozměrný objekt je $N(\varepsilon)$ počet čtverců o straně ε , které vyplňují tento objekt. Numerický odhad D_0 je pro binární obraz počítán pomocí box counting metody (Schroeder, 2009), která funguje následujícím způsobem: Nejprve je obraz rozdělen mřížkou na čtverečky o velikosti $\varepsilon \times \varepsilon$ a následně je spočten počet $N(\varepsilon)$ čtverečků o velikosti $\varepsilon \times \varepsilon$ obsahujících alespoň jeden pixel objektu. Tento postup se několikrát opakuje se stále menší hodnotou ε a ze spočtených hodnot je následně sestaven graf závislosti $\log N(\varepsilon)$ na $\log \varepsilon^{-1}$. Křivka je následně aproximována lineární regresí a hodnota prediktoru FracDim pak odpovídá směrnici regresní přímky.

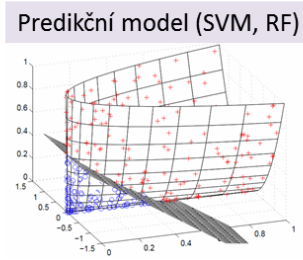
Dále pak SFTA počítá z původního obrazu třetí prediktor (GrayMean), který odpovídá průměrné jasové hodnotě těch pixelů v původním obraze, které pozičně odpovídají bílým pixelům v binárním obraze. Celkem tedy pomocí SFTA získáváme $(2n_t - 1) \cdot 3 = 6n_t - 3$ prediktorů. Například pro počet prahů $n_t = 4$, dostáváme $6 \cdot 4 - 3 = 21$ prediktorů.

2.2.3 Prediktory použité pro patch-based klasifikaci

Schéma znázorňující patch-based přístup i se získáváním prediktorů je znázorněno na obrázku 8. Celkem byly získány dva typy prediktorů: 1) výše popsané parametry histogramu 2) prediktory získané po aplikaci binární stack-dekompozice, popsané v následujícím odstavci.



Data Frame					
	FractalDim.stack1	GrayMean.stack1	...	histMean	...
1	1.468798	48.64108	...	28.67282	...
2	1.537670	78.08910	...	50.72148	...
3	1.394992	57.95667	...	40.80810	...
...



Obrázek 8: Schéma znázorňující patch-based metodu získávání prediktorů na příkladu longitudinálního obrazu benigního uzlu. Segmentovaný uzel je rozdělen do obrazových čtverečků o velikosti 17×17 px. Každý obrazový čtvereček $I(x,y)$ je rozložen pomocí binární stack-dekompozice na deset binárních obrazů, ze kterých jsou získány různé prediktory. Dále jsou pak z každého obrazového čtverečku $I(x,y)$ získány 4 prediktory histogramu. Stack1 – stack10 značí jednotlivá šedotónová pásma. Jednotlivé prediktory jsou popsány v textu.

2.2.3.1 Binární stack dekompozice

Jak bylo zmíněno v předchozím odstavci, SFTA rozkládá obraz na sérii binárních obrazů, které jsou určeny z histogramu Otsuovou metodou. SFTA bere tedy v potaz zastoupení jednotlivých jasových hodnot v obraze a na tomto základě se pak určují prahy pro binarizaci. Tento postup je aplikovatelný na ultrazvukový obraz jako celek, tj. pokud máme větší část obrazu a můžeme předpokládat, že prahy budou určeny v různých obrazech podobně. U patch-based klasifikace toto ale platit nebude. Vezmeme-li dva malé čtverečky, může být v jednom většina pixelů tmavých a ve druhém světlých, a pokud aplikujeme víceúrovňovou Otsuovu metodu, tak získáme zcela odlišné hodnoty prahů. Může tedy dojít k tomu, že u různých čtverečků budeme porovnávat různá šedotónová pásma. To, co tedy u analýzy celého uzlu bereme jako přínos – protože stanovení prahu na základě histogramu celého obrazu Otsuovou metodou funguje jako určitý typ ekvalizace různě exponovaných snímků – to u patch-based klasifikace není žádoucí. Z tohoto důvodu bylo nutné algoritmus upravit a vrátit se k předchůdci dvouprahové binární dekompozice, k metodě nazývané binární stack-dekompozice (Chen et al., 1995), při které jsou prahy stanoveny fixně u všech snímků. Platí tedy:

$$t_i = \left\lfloor \frac{n_l}{n_t + 1} i \right\rfloor, i = 1, 2, \dots, n_t, \quad (2.14)$$

kde n_t je volitelná hodnota počtu prahů a n_l je nejvyšší jasová hodnota v šedotónovém obraze. Takto získaná množina rovnoměrně rozložených prahů $\{t_1, t_2, \dots, t_{n_t}\}$ je rozšířena o minimální a maximální hodnotu jasové úrovně šedotónového obrazu, tedy $\{t_1, t_2, \dots, t_{n_t}\} \cup \{1, n_l\}$, a použita k dvouprahové segmentaci pomocí rovnice 2.12, kdy se používají vždy dva sousední prahy (spodní práh t_s a horní práh t_h , přičemž $t_s < t_h$). Tímto postupem získáváme $n_t + 1$ binárních obrazů.

Nevýhodou tohoto postupu s rovnoměrným rozložením prahů je, že nemusí fungovat, pokud jsou snímky různě exponované. Naše data jsou ale v tomto ohledu poměrně stálá (pravděpodobně z toho důvodu, že byla pořízena pouze několika lékaři a pouze na jedné klinice).

Počet prahů n_t byl fixně nastaven na $n_t = 9$, což odpovídá prahové hodnotě publikované Costou et al. (2012) při použití dvouprahové binární dekompozice a také je to již za hraniční hodnotou 6, případně 7 prahů zjištěnou při klasifikaci celých uzlů (viz 3.1 Výsledky klasifikace celých uzlů, obrázek 11).

2.2.3.2 Region popisující prediktory pro patch-based klasifikaci

Kromě tří prediktorů použitých i ke klasifikaci celých uzlů – tedy fraktální dimenze (FracDim), počtu pixelů (Area) a střední jasové hodnoty (GrayMean) – bylo navíc v každém pásmu každého binárního obrazu určeno celkem 7 dalších region popisujících prediktorů. Regionem je myšlen tvar, který v binárním obraze tvoří bílé pixely. Jednalo se o následující prediktory: ConvexArea, Solidity, Eccentricity, EquivDiameter, Extent, Perimeter a EulerNumber.

ConvexArea je skalár, který specifikuje počet pixelů v nejmenším možném konvexním polynomu, který obsáhne celý region (Hassanien and Oliva, 2017). Konvexním polynomem je myšlena podmnožina eukleidovského prostoru (v našem případě podmnožina všech pixelů binárního obrazu), která má takovou vlastnost, že úsečka spojující libovolné dva body této podmnožiny je obsažena v dané podmnožině. Pro konvexní polynom C a pro všechna $x, y \in C$ a $\lambda \in [0,1]$ tedy platí $(1 - \lambda)x + \lambda y \in C$.

Solidity je skalár, který specifikuje poměr počtu pixelů v regionu a počtu pixelů v nejmenším možném konvexním polynomu, který obsáhne celý region (Hassanien and Oliva, 2017), tedy:

$$Solidity = \frac{Area}{ConvexArea} \quad . \quad (2.15)$$

Eccentricity je skalár, který specifikuje numerickou excentricitu ε elipsy, která má stejný druhý centrální moment jako region. Excentricitu ε spočítáme jako:

$$\varepsilon = \frac{\sqrt{a^2 - b^2}}{a} \quad , \quad (2.16)$$

kde a je délka hlavní poloosy a b je délka vedlejší poloosy (Di et al., 2017). Kruhový region má Eccentricity rovno nule, pro eliptický tvar platí $0 < \text{Eccentricity} < 1$, přímka má pak Eccentricity rovno 1.

EquivDiameter je skalár, který specifikuje průměr kruhu, který má stejnou plochu jako plocha regionu (Area), tedy (Hassanien and Oliva, 2017):

$$\text{EquivDiameter} = \sqrt{\frac{4}{\pi} \text{Area}} \quad . \quad (2.17)$$

Extent je skalár, který specifikuje poměr mezi počtem pixelů v regionu (Area) a počtem pixelů v nejmenším možném obdélníku, kterým lze celý region obsáhnout, tzv. *bounding rectangle* (Russ, 2016).

$$\text{Extent} = \frac{\text{Area}}{\text{Bounding Rectangle Area}} \quad , \quad (2.18)$$

Perimeter je skalár, který specifikuje obvod celého regionu. Počítá se sečtením vzdáleností každých dvou spolu sousedících pixelů, které tvoří okraj regionu. Máme-li tedy n párů souřadnic (x, y) příslušících okrajovým pixelům regionu seřazených tak, že vždy (x_i, y_i) a (x_{i+1}, y_{i+1}) jsou spolu sousedící pixely, pak můžeme Perimeter spočítat jako součet euklidovských vzdáleností sousedních pixelů (Peters, 2017):

$$\text{Perimeter} = \sum_{i=1}^{n-1} \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2} \quad , \quad (2.19)$$

Poslední prediktor, EulerNumber, je skalár, který specifikuje celkový počet binárních objektů v obraze mínus celkový počet děr v těchto objektech (Umbaugh, 2017), tedy:

$$\text{EulerNumber} = \frac{\text{Počet objektů}}{\text{Počet děr v objektech}} \quad . \quad (2.20)$$

Obecně můžeme říct, že všechny výše popsané prediktory popisují prostorovou komplexitu tvarů (regionů), které zůstávají v jednotlivých šedotónových pásmech po dvouprahové segmentaci. Je jasné, že některé prediktory hodnotí podobné vlastnosti uzlu a budou spolu korelovat. Pro výběr těch nejvhodnějších prediktorů jsme použili metodu rekurzivní eliminace prediktorů (*recursive feature elimination*), která zpětně vyřazuje z modelu ty prediktory, které nepřispívají významně ke správné klasifikaci.

SFTA, binární stack-dekompozice stejně tak jako rozdělení obrazu do čtverečků a výpočet všech prediktorů byly provedeny v programovém prostředí Matlab (MathWorks Inc., Natick, MA, USA) s pomocí Image Processing Toolboxu.

2.3 Klasifikace

V experimentální části práce jsou používány dva klasifikační modely (klasifikátory): SVM (Cortes and Vapnik, 1995) a náhodné lesy (*random forests*) (Breiman, 2001), které byly implementovány v R 3.3.1 (The R Foundation for Statistical Computing, Wien, Austria) pomocí knihovny Caret (Kuhn, 2008). Oba modely v obecném srovnání klasifikačních modelů stojí na předních příčkách co do přesnosti, tak i do schopnosti generalizace (Fernandez-Delgado et al., 2014). SVM je jednoznačně jedním z nejčastěji používaných klasifikačních modelů v CAD systémech pro uzly ve štítné žláze, byl použit např. u autorů (Ding et al., 2011, Acharya et al., 2012b, Chang et al., 2016, Acharya et al., 2016, Tsantis et al., 2009) a náhodné lesy se obecně zdají být velmi slibným algoritmem pro klasifikaci založenou na větším počtu prediktorů, jak to v případě uzlů ve štítné žláze ukázali autoři Chi et al. (2017). Skripty klasifikačních modelů v programovacím jazyce R jsou k nahlédnutí v přílohách této práce (příloha 1 – patch-based klasifikace, příloha 2 – klasifikace celých uzlů)

2.3.1 Náhodné lesy

Náhodné lesy spadají mezi tzv. souborové klasifikační metody (*ensemble methods*), které pracují na principu budování několika jednodušších klasifikačních modelů, jejichž výsledky se nakonec průměrují (regresní problémy) nebo jsou dány většinovým hlasováním (klasifikační problémy). V případě náhodných stromů je tímto jednodušším klasifikačním modelem rozhodovací strom typu CART (*classification and regression trees*), který k predikci používá binární uzly.

2.3.1.1 Klasifikační stromy typu CART

Klasifikační strom typu CART (Breiman et al., 1984) lze popsat stromovým grafem, který je tvořen uzly a orientovanými hranami. Mějme data, kdy každé pozorování je charakterizováno vektorem hodnot $x = (x_1, x_2, \dots, x_m)$ prediktorů X_1, X_2, \dots, X_m a zároveň každé pozorování patří do jedné z tříd C_1, C_2, \dots, C_k .

V každém neterminálním uzlu klasifikačního stromu typu CART se strom větví. Větvení je založeno na hodnotě právě jednoho prediktoru X_i . U kvantitativních prediktorů X_i se testuje, zda je hodnota prediktoru menší než určitá konstanta c , tedy zda $x_i < c$. U kvalitativních prediktorů X_i se testuje, zda $x_i \in B$, kde B je neprázdná vlastní podmnožina množiny všech hodnot veličiny X_i . V každém uzlu jsou tedy právě dvě možnosti: hodnota prediktoru splňuje testovanou podmínku, nebo nesplňuje, a podle výsledku postupuje pozorování orientovanou hranou do jednoho ze dvou následujících uzlů, kde je obdobným způsobem testován další prediktor X_i , který data dále dělí. Takto postupuje pozorování celým stromem, až skončí v některém z terminálních uzlů (listů), čímž je přiřazen (klasifikován) do některé ze tříd C_1, C_2, \dots, C_k (Klaschka and Kotrč, 2004).

Při hledání správného rozdělení v uzlech se v případě klasifikace počítá kritériální statistika, která je založena na poměru tříd v uzlech. Kritériální statistikou v případě CART stromů a náhodných lesů je Gini index G , který lze spočítat jako:

$$G = \sum_{c=1}^k p_c(1 - p_c) = 1 - \sum_{c=1}^k p_c^2 \quad , \quad (2.21)$$

kde k je počet tříd C a p_c je podíl pozorování příslušících do třídy C v daném uzlu z celkového počtu pozorování v tomto uzlu, neboli pravděpodobnost kategorie C v tomto uzlu (Rebala et al., 2019).

Celková hodnota Gini indexu G_{celk} pro dané rozhodnutí je dána jako vážený součet indexů dceřiných uzlů, kdy jako váha slouží podíl počtu pozorování v daném dceřiném uzlu N_i a počtu pozorování v mateřském uzlu N . Obecně můžeme celkovou hodnota Gini indexu G_{celk} po rozdělení dat v uzlu na k dceřiných uzlů vyjádřit jako (Rebala et al., 2019):

$$G_{celk} = \sum_{i=1}^k \frac{N_i}{N} G(i) \quad . \quad (2.22)$$

Pro binární strom typu CART platí $k = 2$, celkovou hodnotu Gini indexu G_{celk} tedy můžeme spočítat jako:

$$G_{celk} = \frac{N_1}{N} G_1 + \frac{N_2}{N} G_2 \quad , \quad (2.23)$$

kde N_1 je počet pozorování v 1. dceřiném uzlu, N_2 je počet pozorování ve 2. dceřiném uzlu, N je počet pozorování v mateřském uzlu a G_1 a G_2 jsou příslušné Gini indexy 1. a 2. dceřiného uzlu (spočítané pomocí rovnice 2.21).

Celková hodnota Gini indexu G_{celk} je u jednoduchých stromů počítána pro všechny ve stromu zatím nepoužité prediktory, přičemž v uzlu se použije ten prediktor, pro který je G_{celk} nejnižší.

2.3.1.2 Náhodné lesy založené na CART

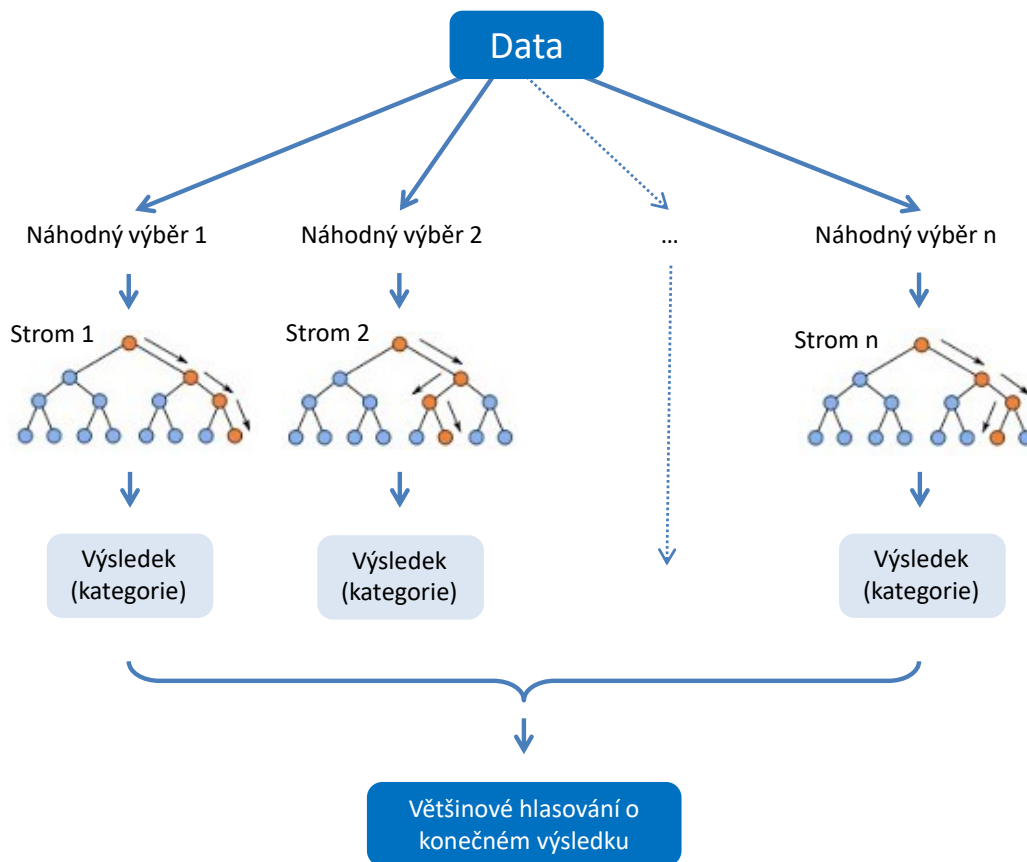
Náhodné lesy (*random forests*) jsou tvořeny velkým množstvím rozhodovacích stromů typu CART, přičemž k tomu využívají, jak název naznačuje, náhodného výběru. Konkrétně se při tvorbě stromu uplatňují dva náhodné procesy. První je, že strom není tvořen na kompletních datech, ale je tvořen vždy pouze na náhodně vybrané podskupině dat (využívá se *bootstrapping*, náhodný výběr s opakováním). Výsledný klasifikační les je pak dán většinovým hlasováním všech na dílčích datech vyrostlých stromů.

Druhý náhodný proces se pak uplatňuje při tvorbě uzlu rozhodovacího stromu, kdy je vždy vybrána pouze náhodná podskupina prediktorů, které se v daném uzlu zkouší. Z této podskupiny se pak klasicky na vytvoření uzlu použije ten prediktor, pro který vychází nejnižší hodnota G_{celk} . Tím, že nepoužíváme všechny prediktory, docílíme jednak snížení výpočetní náročnosti, ale také toho, že tentýž uzel každého rozhodovacího stromu bude s největší pravděpodobností používat jiný prediktor, a tím bude vytvořen soubor různorodých stromů, jejichž výsledky má poté cenu průměrovat.

Tyto dva náhodné procesy jsou kontrolovány dvěma základními parametry, které lze při tvorbě náhodných lesů nastavit. První, v knihovně Caret označován *n tree*, specifikuje, kolik stromů bude vytvořeno a následně zprůměrováno. Optimální hodnota *n tree* se pohybuje v řádech stovek až tisíců. Pokud není zmíněno jinak, pak

náhodné lesy publikované v této práci pracují s hodnotou $n_{tree} = 500$ v případě patch-based klasifikace a $n_{tree} = 2000$ v případě klasifikace celých uzlů. Druhý parametr, v Caret knihovně označovaný jako $mtry$, pak ovlivňuje, kolik prediktorů bude náhodně zkoušeno a porovnáváno v každém uzlu každého stromu. Optimální hodnota se může pohybovat kdekoli mezi 1 a celkovým počtem prediktorů a obvykle je potřeba tuto optimální hodnotu pro daný úkol najít.

Zjednodušené schéma náhodných lesů znázorňuje obrázek 9.



Obrázek 9: Schéma náhodných lesů s celkovým počtem stromů rovno n . Z dat je uděláno n náhodných výběrů, které se použijí k vybudování n rozhodovacích stromů. Každý strom dává dílčí výsledek klasifikace do příslušných kategorií. Výsledná klasifikace je pak modus dílčích výsledků.

Budování stromů samo o sobě dělí data v uzlech postupně pomocí těch nejdůležitějších prediktorů po ty nejméně důležité. Proto se náhodné lesy v praxi často používají k měření významnosti proměnných a zhodnocení efektu proměnných na

predikci. K tomuto účelu lze vypočítat ukazatel důležitosti I (*importance*), jehož výpočet je založený na poklesu klasifikační přesnosti při náhodné permutaci hodnoty proměnných v uzlu. Pro každý rozhodovací strom náhodného lesa máme malou část dat z celkové množiny dat, která nebyla použita při vytváření daného rozhodovacího stromu (tzv. *out-of-bag sample*). Tato část dat je použita k výpočtu ukazatele I konkrétního prediktoru. Nejprve se pro tuto *out-of-bag* část dat určí predikční přesnost na vytvořeném stromu. Poté se hodnoty zkoumaného prediktoru „zašumí“ dosazením náhodné hodnoty, přičemž hodnoty ostatních prediktorů zůstávají stejné. Tím eliminujeme vliv zkoumaného prediktoru na klasifikační proces, čímž se sníží klasifikační přesnost. Náhodné dosazování hodnot prediktoru do všech stromů pak vede k tomu, že v průměru tento prediktor ztratí s největší pravděpodobností veškerou svoji predikční sílu. Důležitost I je pak průměrný pokles přesnosti pro všechny stromy náhodného lesa normovaný tak, aby nabýval hodnot od 0 do 100 (Kuhn, 2008).

2.3.2 SVM

Klasifikační model SVM je založen na principu prokládání dat v multidimenzionálním prostoru nadrovinami, které vedou k rozdělení dat do příslušných tříd. Lineární nadroviny jsou použity tam, kde lze prediktory rozdělit lineárně, tzv. kernelové transformace jsou použity v těch případech, kdy mezi daty nejsou lineární závislosti. Kernelové transformace jsou založeny na tom, že existují tzv. jádrové, neboli kernelové funkce dvou vektorových argumentů, $K(x, x')$, jejichž hodnoty jsou rovny skalárnímu součinu vektorů $\phi(x)$ a $\phi(x')$, které vzniknou namapováním x a x' do mnohazměrného prostoru.

Pro potřeby této práce jsme zvolili nelineární SVM model založený na základě RBF (*radial basis function*), neboli Gaussovského kernelu. Pokud máme dva případy reprezentované dvěma vektory prediktorů x a x' v daném původním prostoru, pak můžeme RBF kernel definovat jako (Jebara, 2003):

$$K(x, x') = e^{-\frac{\|x-x'\|^2}{2\sigma^2}} \quad , \quad (2.24)$$

kde σ je free-parametr, kterým lze měnit rozptyl Gaussovské funkce a $\|x - x'\|^2$ je čtverec euklidovské metriky, tedy:

$$\|x - x'\|^2 = (x_1 - x'_1)^2 + (x_2 - x'_2)^2 + \dots + (x_n - x'_n)^2 \quad , \quad (2.25)$$

kde n je velikost shodně velkých vektorů x a x' . Dále pak můžeme u SVM nastavit ladicí parametr C , který ovlivňuje, jak budou klasifikovány body, které se nacházejí v oblasti okraje nadroviny (angl. *margin*), slouží tedy k nastavení prahu mezi komplexitou rozhodovacího pravidla a četností chybně klasifikovaných případů na tréninkových datech (Cortes and Vapnik, 1995). Malá hodnota parametru C zvýší počet chyb při tréninku klasifikátoru, zatímco velká hodnota parametru C povede k menšímu množství tréninkových chyb, ale chování klasifikátoru se poté blíží tzv. *hard-margin* SVM (Joachims, 2002). Hlavní výhody SVM ve srovnání s náhodnými lesy jsou nižší nároky na výpočetní výkon a jednodušší grafická interpretace modelu. Společnou výhodou jak SVM, tak náhodných lesů je jejich dobrá generalizační schopnost (Fernandez-Delgado et al., 2014), to znamená, že nejsou náchylné k „přetrénování“ na tréninkový soubor dat a vykazují dobré výsledky i při použití na nových datech.

2.3.3 Nevyváženost tříd

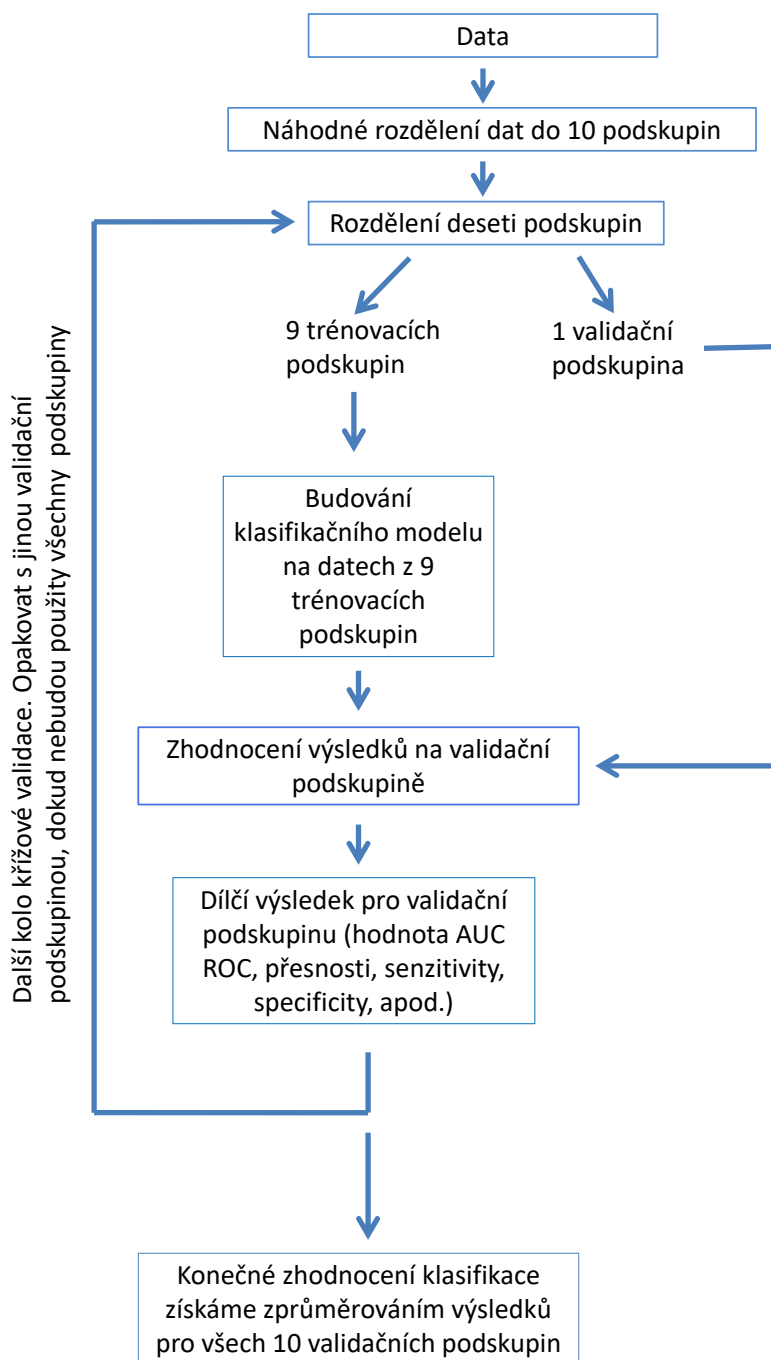
V případě patch-based klasifikace je náš datový soubor nevyvážený (větší počet benigních než maligních případů), což způsobuje vyšší apriorní pravděpodobnost benignity. Nevyváženost tříd je obecně při použití algoritmů strojového učení problém, který může vést k nereprezentativním výsledkům klasifikace (Branco et al., 2016). Abychom neztratili komplexitu dat z benigní třídy, rozhodli jsme se vyváženost tříd neřešit na úrovni obrázků (tzn. upravovat počet ultrazvukových obrazů), ale až na úrovni patchů (obrazových čtverečků). Přesně jsme postupovali tím způsobem, že jsme podzorkovali množinu čtverečků z majoritní třídy pomocí softwarového balíčku Unbalanced for R (Dal Pozzolo et al., 2013) tak, aby počet patchů v benigní a maligní třídě byl stejný. Tímto postupem došlo ke snížení celkového počtu čtverečků (pro velikost 17×17 px) z 9065 na 6864, čímž došlo k vyvážení benigní a maligní třídy.

Druhou možností, jak lze nevyváženost tříd řešit, je vytvořit syntetická data příslušící k minoritní třídě (*over-sampling*), jako v případě CAD pro štítnou žlázu použili autoři Acharya et al. (2016). Tento postup se nám ale v případě patch-based analýzy, kdy používáme malé čtverečky ultrazvukového obrazu, který sám o sobě obsahuje velké množství šumu, zdá jako méně vhodné řešení. Při podvzorkování dat sice nepracujeme s celou oblastí uzlu u maligních případů, ale alespoň nevnášíme do dat zkreslení ve formě možné syntézy dalšího šumu.

2.3.4 Validace klasifikačních modelů

K validaci klasifikačních modelů byla použita tzv. *k-fold* křížová validace (*k-fold cross-validation*), což je metoda hodnocení, při které jsou data rozdělena do *k*-podskupin, přičemž k trénování klasifikačního modelu se použije $k - 1$ podskupin a zbylá podskupina (tzv. *validation fold*, nebo *out-of-bag fold*) se ponechá stranou a použije se až k validaci vybudovaného klasifikačního modelu. Tento postup simuluje testování na nezávislých datech, což omezuje „přeučení“ klasifikačního modelu na použitá data (*overfitting*) a zvyšuje se tím generalizační schopnost klasifikačního modelu. Schéma křížové validace je znázorněno na obrázku 10.

V našem případě jsme použili pro hodnocení celých snímků metodu *leave one out cross-validation* (LOOC), což je specifický případ *k-fold* křížové validace, kdy *k* je rovno počtu případů. Při hodnocení patch-based klasifikace jsme použili skupinovou 10-fold křížovou validaci (*group 10-fold cross-validation*), kde slovo skupinová označuje, že rozdělení do 10 podskupin není zcela náhodné, ale je provedeno tak, aby bylo zaručeno, že se data ze stejné skupiny (v našem případě čtverečky z jednoho obrázku) nevyskytnou zároveň v trénovací a validační množině. Skupinová křížová validace by měla být použita v těch případech, kdy je z každého subjektu získáno několik vzorků a hrozí, že klasifikátor se může tzv. „přeučit“ na specifické, subjektivní vlastnosti. V našem případě tedy skupinová křížová validace zajišťuje, že čtverečky z jednoho obrázku vždy kompletně všechny spadají do jedné z deseti podskupin (foldů), což by mělo vést k přesnějším výsledkům při použití na nových datech.



Obrázek 10: Schéma křížové validace statistického modelu na příkladu *10-fold* křížové validace. Klasifikátor je trénován a hodnocen desetkrát, vždy na jiných datech. Výsledné hodnocení klasifikátoru (např. výsledná hodnota AUC ROC) je dána zprůměrováním všech deseti dílčích výsledků.

2.3.5 Výběr prediktorů

Při klasifikaci celých uzlů byly použity všechny získané prediktory, protože pozorování (obrazů) i prediktorů bylo relativně málo. U patch-based klasifikace bylo ale zpracovááno několik tisíc pozorování (patchů) a počet prediktorů byl výrazně vyšší (104), přičemž ne všechny musejí být nutně pro klasifikaci benigní a maligní třídy důležité. Z tohoto důvodu jsme použili metodu tzv. zpětného vyřazování prediktorů (*Recursive Feature Elimination – RFE*) pomocí stejnojmenné funkce v balíčku Caret (Kuhn, 2008), a to jak pro SVM, tak pro RF klasifikátor. Výsledky jsou znázorněny v kapitole 3.2 Výsledky patch-based metody, tabulka 2 a obrázek 14.

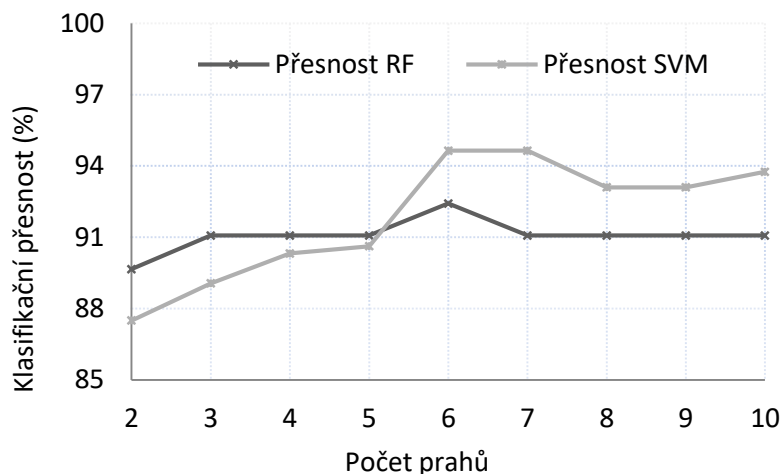
2.4 Technické specifikace hardwaru

Veškerá výpočetní část této práce (včetně získávání prediktorů a tvorby klasifikačních modelů) byla provedena na počítači s procesorem Intel (Intel Corp., Santa Clara, California, USA) i3, 2,1 GHz dual-core, s pamětí RAM o velikosti 8 GB, a operačním systémem Windows 7 Professional (Microsoft Corp., Redmond, Washington, USA), bez externí grafické karty.

3 VÝSLEDKY

3.1 Výsledky klasifikace celých uzlů

Klasifikace pomocí náhodných lesů při optimálním nastavení (počet prahů $n_t = 6$, $mtry = 12$) dosáhla přesnosti 92,5 % se směrodatnou odchylkou křížové validace $\pm 0,22$. Klasifikace uzlů pomocí metody SVM při optimálním nastavení (počet prahů $n_t = 6$, $C = 8$) dosáhla přesnosti 95 % se směrodatnou odchylkou křížové validace $\pm 0,16$, což je výsledek srovnatelný s předchozími studii publikovanými na toto téma (viz tabulka 7). Jak bylo řečeno, SFTA algoritmus rozkládá obraz na množinu binárních obrazů, jejichž počet závisí na počtu prahů n_t (viz obrázek 7). Na obrázku 11 můžeme vidět, jak závisí celková přesnost klasifikace celých uzlů na počtu prahů n_t .

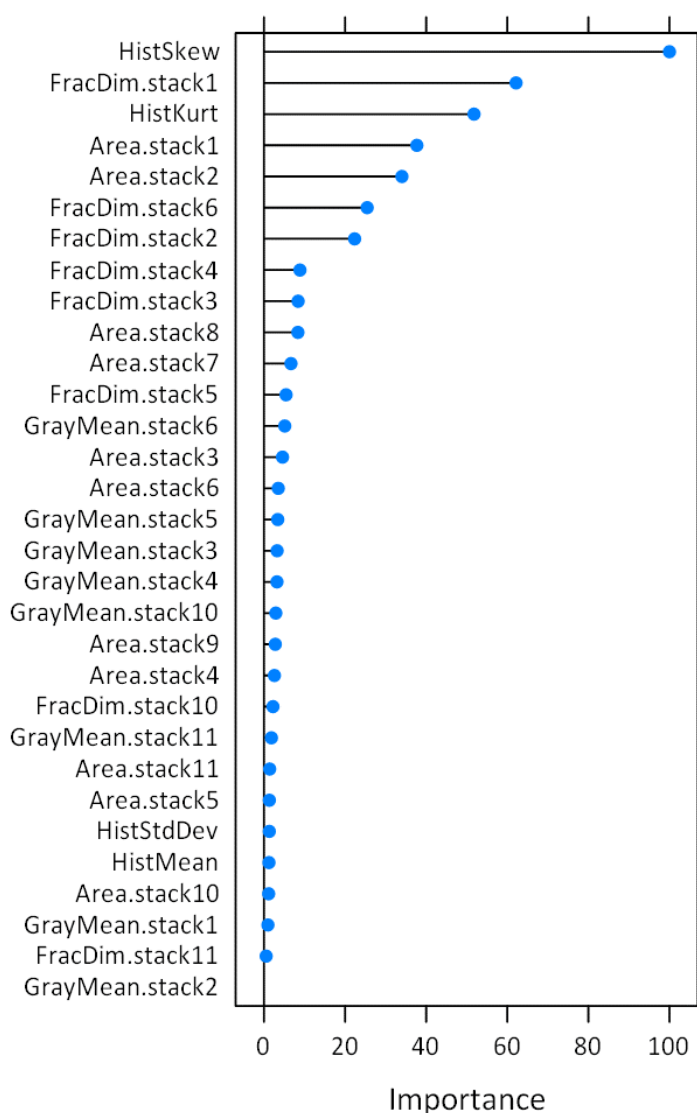


Obrázek 11: Graf znázorňující závislost celkové přesnosti klasifikace pomocí náhodných lesů (RF) a SVM na počtu prahů SFTA algoritmu.

Na grafu vidíme, že nejvyšší přesnosti – jak pro náhodné lesy, tak pro SVM – bylo dosaženo pro počet prahů 6 nebo 7, což poměrně odpovídá výsledkům prezentovaných autory SFTA algoritmu (Costa et al., 2012). Ti uvádějí, že klasifikační přesnost se na

testovacích datech zvyšovala do počtu prahů rovno 8 nebo 9 s vysvětlením, že od této hodnoty není SFTA algoritmus schopný identifikovat další texturní vlastnosti v obraze. Naše výsledky tedy poměrně odpovídají jejich závěrům, nutno ale dodat, že autoři sice používali k hodnocení algoritmu medicínský obraz, nejednalo se ale o ultrazvuk štítné žlázy, ale o vzorky CT snímků hrudníku s rozlišením 64×64 px.

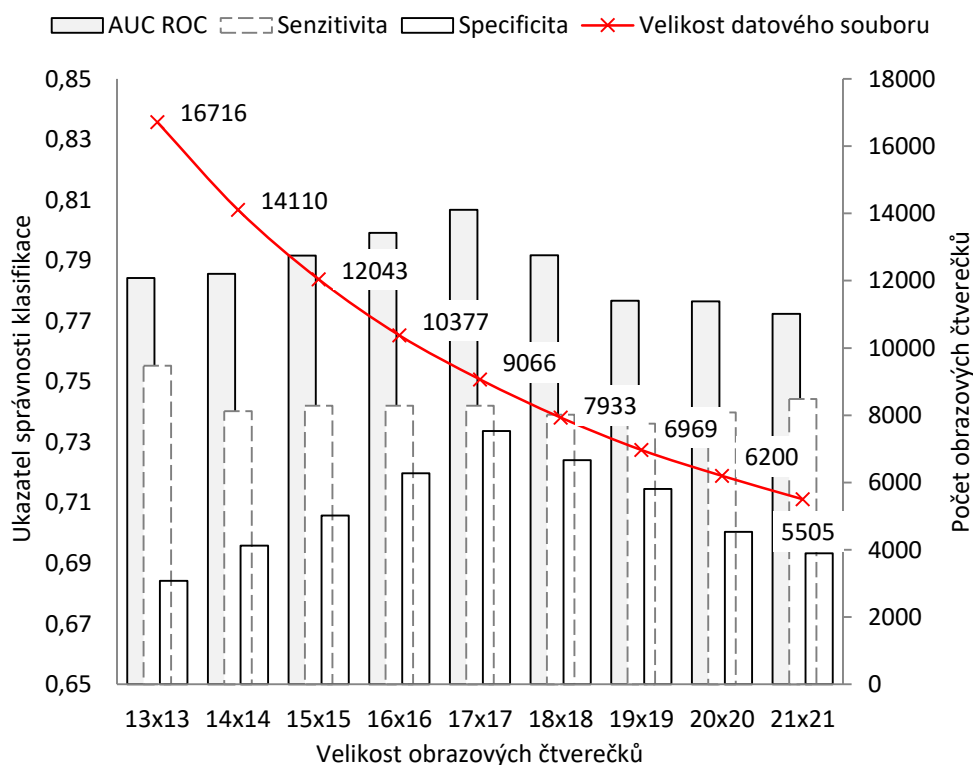
Důležitost I všech prediktorů použitých ke klasifikaci celých uzlů je znázorněna na obrázku 12. Princip určování hodnoty I je vysvětlen v metodách, obecně lze ale říct, že čím vyšší hodnota I , tím je prediktor pro správnou klasifikaci důležitější.



Obrázek 12: Hodnoty ukazatele důležitosti I (*importance*) znázorňující významnost jednotlivých prediktorů náhodného lesa (pro počet prahů SFTA algoritmu $n_t = 6$).

3.2 Výsledky patch-based metody

Velikost čtverečků může ovlivňovat přesnost klasifikace nejen tím, že se mění rozlišení analyzované části uzlu, ale také tím, že velikost čtverečků má vliv na to, kolik jich je z obrazů celkem extrahováno; ovlivňuje tedy velikost datového souboru pro klasifikaci. Proto byly testovány různé velikosti čtverečků pro určení optimální velikosti. Výsledky nejkritičtějšího úseku velikostí jsou znázorněny na obrázku 13. Na obrázku je vidět, jaký vliv má velikost obrazového čtverečku na počet extrahovaných čtverečků, kdy například pro velikost 11×11 px je počet čtverečků roven 23 750 a pro velikost 21×21 px se jedná už jen zhruba o 5 500 čtverečků. Vliv na metriky správnosti klasifikace není zas tolik velký, ale přesto je vidět, že pro velikost 17×17 px je hodnota AUC ROC nejvyšší. Proto jsme tuto velikost vyhodnotili jako optimální a všechny další výsledky publikované v této práci jsou právě pro tuto optimální velikost.



Obrázek 13: Vliv velikosti čtverečku na celkový počet extrahovaných čtverečků (pravá svislá osa) a dále pak na metriky správnosti klasifikace (levá svislá osa), konkrétně plochu pod křivkou ROC (AUC ROC), senzitivitu a specificku. Jednalo se o klasifikaci pomocí náhodných lesů při použití všech 104 extrahovaných prediktorů.

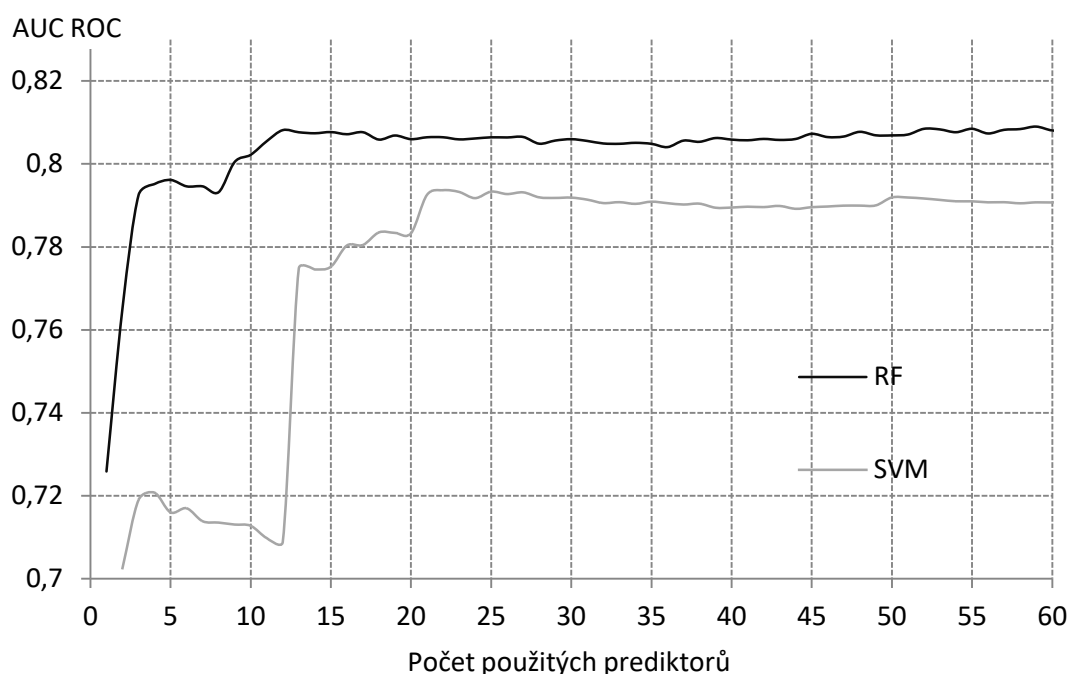
Metoda zpětné eliminace prediktorů (*Recursive Feature Elimination* – RFE) byla použita k výběru těch nejdůležitějších prediktorů pro klasifikační proces a k vyřazení těch prediktorů, které jsou pro klasifikaci nadbytečné. Tabulka 2 obsahuje seznam 30 nejdůležitějších prediktorů pro náhodné lesy a SVM určených pomocí metody RFE.

Pořadí	Prediktory náhodných lesů	Prediktory SVM
1	histMean	histMean
2	ConvexArea.stack1	ConvexArea.stack2
3	ConvexArea.stack3	Area.stack1
4	EquivDiameter.stack1	EquivDiameter.stack1
5	Area.stack1	FilledArea.stack1
6	FractalDim.stack1	FractalDim.stack1
7	FilledArea.stack1	ConvexArea.stack4
8	GrayMean.stack1	Area.stack4
9	Extent.stack1	EquivDiameter.stack4
10	EquivDiameter.stack3	FilledArea.stack4
11	Area.stack3	Extent.stack1
12	ConvexArea.stack4	FractalDim.stack4
13	GrayMean.stack3	GrayMean.stack1
14	FilledArea.stack3	Solidity.stack1
15	Eccentricity.stack1	Perimeter.stack1
16	FilledArea.stack2	GrayMean.stack3
17	Area.stack2	Eccentricity.stack1
18	Solidity.stack1	FilledArea.stack2
19	GrayMean.stack2	Area.stack2
20	FilledArea.stack4	EquivDiameter.stack2
21	EquivDiameter.stack2	ConvexArea.stack2
22	EquivDiameter.stack4	ConvexArea.stack3
23	Area.stack4	Area.stack3
24	FractalDim.stack2	EquivDiameter.stack3
25	GrayMean.stack4	Perimeter.stack4
26	ConvexArea.stack2	GrayMean.stack4
27	FractalDim.stack3	FilledArea.stack3
28	FractalDim.stack4	Extent.stack4
29	Eccentricity.stack4	Solidity.stack4
30	Solidity.stack4	Perimeter.stack2

Tabulka 2: Třicet nejdůležitějších prediktorů vyhodnocených metodou RFE.

Význam algoritmu RFE nespočívá jen v tom, že seřadí jednotlivé prediktory, ale především v tom, že určí optimální počet prediktorů, který je k přesné klasifikaci

potřeba. Závislost správnosti klasifikace na počtu prediktorů seřazených pomocí metody RFE je znázorněna na obrázku 14, ze kterého je zřejmé, že výsledek klasifikace se po dosažení určitého počtu prediktorů již s novými prediktory příliš nemění; tento počet je 14 pro náhodné lesy a 21 pro SVM. Z toho plyne, že velké množství prediktorů přispívá pouze velmi málo, až zanedbatelně, ke zlepšení klasifikace, a proto bylo vhodné tyto prediktory vyřadit. Níže publikované výsledky jsou tedy již pro redukované klasifikační modely, kdy bylo použito v případě náhodných lesů 14 předních prediktorů a v případě SVM 21 předních prediktorů tak, jak byly vyhodnoceny pomocí RFE a jak jsou seřazené v tabulce 2.



Obrázek 14: Graf znázorňující výsledky metody zpětného vyřazování prediktorů (*recursive feature elimination*), kdy jako metrika byla použita plocha pod křivkou ROC (AUC ROC). Z grafu je zřejmé, že po dosažení určitého počtu prediktorů již nedocházelo ke zlepšení klasifikace. Tento počet byl 14 pro náhodné lesy (RF) a 21 pro *support vector machines* (SVM).

Klasifikace všech 6 864 obrazových čtverečků (počet po vyrovnání nevyváženosti tříd z celkového počtu 9 066 extrahovaných čtverečků) pomocí náhodných lesů při optimálním nastavení dosáhla hodnot AUC ROC = 0,81, senzitivity = 73 % a specificity = 75 %. Podobně klasifikace všech obrazových čtverečků pomocí SVM při optimálním nastavení dosáhla hodnoty AUC ROC = 0,79, senzitivita byla rovna 73 % a specificita 75 %. Výsledky klasifikace jednotlivých čtverečků pomocí náhodných lesů v závislosti na ladícím parametru *mtry* a výsledky klasifikace jednotlivých čtverečků pomocí SVM v závislosti na ladícím parametru *C* jsou vidět v tabulce 3. Pro náhodné lesy vychází optimální hodnota *mtry* = 2 a pro SVM vychází optimální hodnota *C* = 0,25.

Náhodné lesy			
<i>mtry</i>	ROC	Senzitivita	Specificita
1	0,8042	72,49 %	76,30 %
2	0,8087	72,98 %	75,08 %
3	0,8060	73,77 %	73,96 %
4	0,8035	73,91 %	73,52 %
5	0,8023	73,44 %	73,50 %
:	:	:	:
Support Vector Machines			
<i>C</i>	ROC	Senzitivita	Specificita
0,25	0,7900	73,46 %	75,00 %
0,50	0,7862	73,59 %	74,87 %
1,00	0,7816	73,50 %	74,66 %
2,00	0,7803	73,58 %	74,25 %
4,00	0,7786	73,51 %	73,76 %

Tabulka 3: Výsledky klasifikace jednotlivých obrazových čtverečků pro různé nastavení náhodných lesů a SVM. 1) Náhodné lesy: Tabulka ukazuje pět výsledků různého nastavení ladícího parametru *mtry*, který představuje počet prediktorů, které jsou náhodně vybírány pro každý uzel rozhodovacího stromu. Hodnota *mtry* = 2 byla vyhodnocena jako optimální nastavení náhodných lesů. 2) SVM: Tabulka ukazuje prvních 5 testovaných hodnot ladícího parametru *C*, kdy *C* = 0,25 bylo vyhodnocena jako optimální nastavení.

Zásadní otázkou u patch-based klasifikace uzlů ve štítné žláze je, jestli klasifikátor špatně klasifikuje celé uzly, nebo jestli klasifikuje špatně pouze některé obrazové čtverečky z daného uzlu (například ty čtverečky, které jsou na okraji uzlu; čtverečky v té části uzlu diagnostikovaného jako benigní, které mohou být zasaženy počínajícími maligními změnami; nebo naopak části maligního uzlu, kde stále převládají buňky benigního charakteru). Proto i u patch-based analýzy zůstává hlavním záměrem dosáhnout co nejlepších výsledků při diagnostice uzlu jako celku a ne se zaměřovat pouze na co nejpřesnější klasifikaci obrazových čtverečků.

Pro klasifikaci uzlů jako celku můžeme vzít v úvahu procento správně diagnostikovaných čtverečků ze všech čtverečků v daném uzlu. Toto procento vyjadřuje svým způsobem riziko malignity pro daný uzel a pro účely této práce jsme ho nazvali Index rizika malignity (*IRM*). Výsledky procentuálního zastoupení správně klasifikovaných čtverečků i s indexem *IRM* u deseti vybraných benigních a deseti vybraných maligních uzlů je vidět v tabulce 4.

Po výpočtu *IRM* byla sestavena ROC křivka (viz obrázek 15), pomocí které byly hodnoceny konečné výsledky (senzitivita, specificita, přesnost a AUC ROC) hodnotící klasifikaci na celém souboru dat (viz tabulka 5). Analýza ROC křivky ukázala, že optimální práh pro *IRM* je kolem hodnoty 0,5, což koresponduje s předpokladem, že maligní uzly by měly být označeny za maligní, pokud je více jak 50 % obrazových čtverečků maligních; a podle stejného klíče by měly být označeny benigní uzly za benigní, pokud je více jak 50 % čtverečků hodnoceno jako benigní.

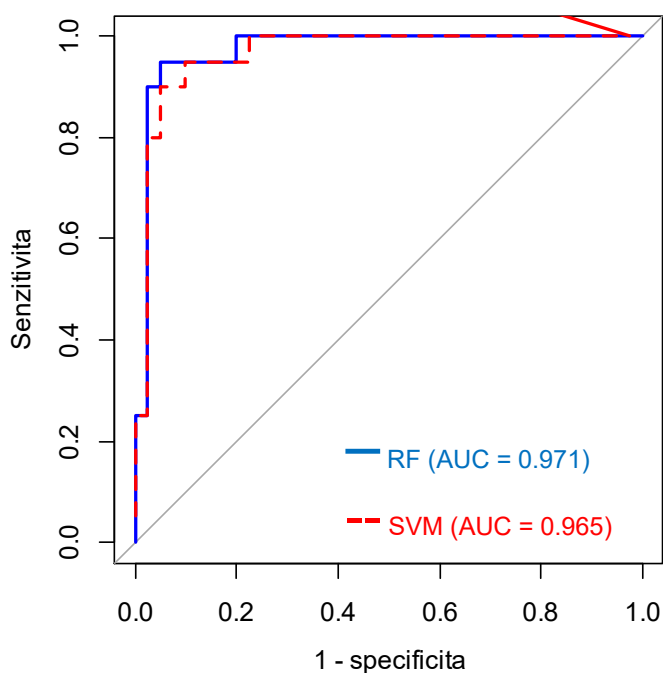
Uzel	Cytologie	Pohlaví	Přístroj	Správně klasifikováno	Nesprávně klasifikováno	IRM
1	Benigní	Ž	GE	13	3	0,19
2	Benigní	Ž	Philips	44	23	0,34
3	Benigní	M	GE	19	0	0,00
4	Benigní	Ž	GE	123	30	0,20
5	Benigní	Ž	GE	64	45	0,41
6	Benigní	Ž	GE	151	12	0,07
7	Benigní	Ž	Philips	81	9	0,10
8	<i>Benigní</i>	Ž	GE	9	48	0,84
9	Benigní	Ž	Philips	58	12	0,17
10	Benigní	Ž	Philips	50	20	0,29
:	:	:	:	:	:	:
40	Maligní	Ž	GE	35	1	0,97
41	Maligní	Ž	GE	196	153	0,56
42	Maligní	Ž	Philips	33	2	0,94
43	Maligní	Ž	GE	41	18	0,69
44	Maligní	M	Philips	153	70	0,67
45	Maligní	Ž	GE	88	29	0,75
46	Maligní	Ž	GE	122	42	0,74
47	Maligní	Ž	Philips	153	70	0,67
48	Maligní	Ž	GE	49	1	0,98
49	Maligní	M	Philips	97	24	0,80
50	Maligní	Ž	GE	36	9	0,80

Tabulka 4: Výsledky klasifikace na příkladu 10 benigních a 10 maligních uzlů ukazující počet správně a nesprávně klasifikovaných obrazových čtverečků pro každý uzel. Index rizika malignity (*IRM*) odpovídá procentu čtverečků v uzlu, které byly klasifikovány jako maligní. Nesprávně klasifikovaný uzel při prahové hodnotě *IRM* rovno 0,5 je označen kurzívou (uzel 8).

Při použití všech obrazových čtverečků bylo z celkového počtu 60 uzlů špatně diagnostikováno pouze několik případů, a to jak pomocí náhodných lesů, tak pomocí SVM. Celkové výsledky jsou znázorněny v tabulce 5 a výkon klasifikátoru je také vidět na ROC křivkách na obrázku 15.

Přesná segmentace – všechny obrazové čtverečky					
Klasifikátor	Přesnost	Senzitivita	Specifická	ROC AUC	Chybně klasifikováno
RF	95,0 %	95,0 %	95,0 %	0,9712	2 Philips, 1 GE
SVM	91,6 %	95,0 %	90,0 %	0,965	3 Philips, 2 GE

Tabulka 5: Výsledky klasifikace všech 60 uzlů při použití všech obrazových čtverečků. Poslední sloupec ukazuje, kolik uzlů bylo chybně klasifikováno a kterým přístrojem byly tyto chybně klasifikované obrázky uzlů pořízeny.



Obrázek 15: Graf znázorňující ROC křivky klasifikace 60 uzlů po zprůměrování výsledků patch-based klasifikace. RF značí náhodné lesy a SVM značí *support vector machines*, AUC značí plochu pod křivkou ROC.

Velkou výhodou patch-based klasifikace je, že z podstaty věci nemusí být tolik závislá na přesné segmentaci uzlu, jelikož zde platí předpoklad, že i když je uzel špatně ohraničen – a nemáme všechny obrazové čtverečky – tak i přes to může být Index rizika malignity (*IRM*) vypovídající. Z tohoto důvodu jsme se rozhodli zhodnotit také výkon klasifikátoru při použití pouze 75 % obrazových čtverečků, kdy výsledky jsou vidět v tabulce 6. Předpoklad, že si klasifikace při použití menšího počtu čtverečků

udrží svoji přesnost, se potvrdil; dokonce předčil očekávání, jelikož došlo k nepatrnému zlepšení jak pro SVM, tak pro náhodné lesy.

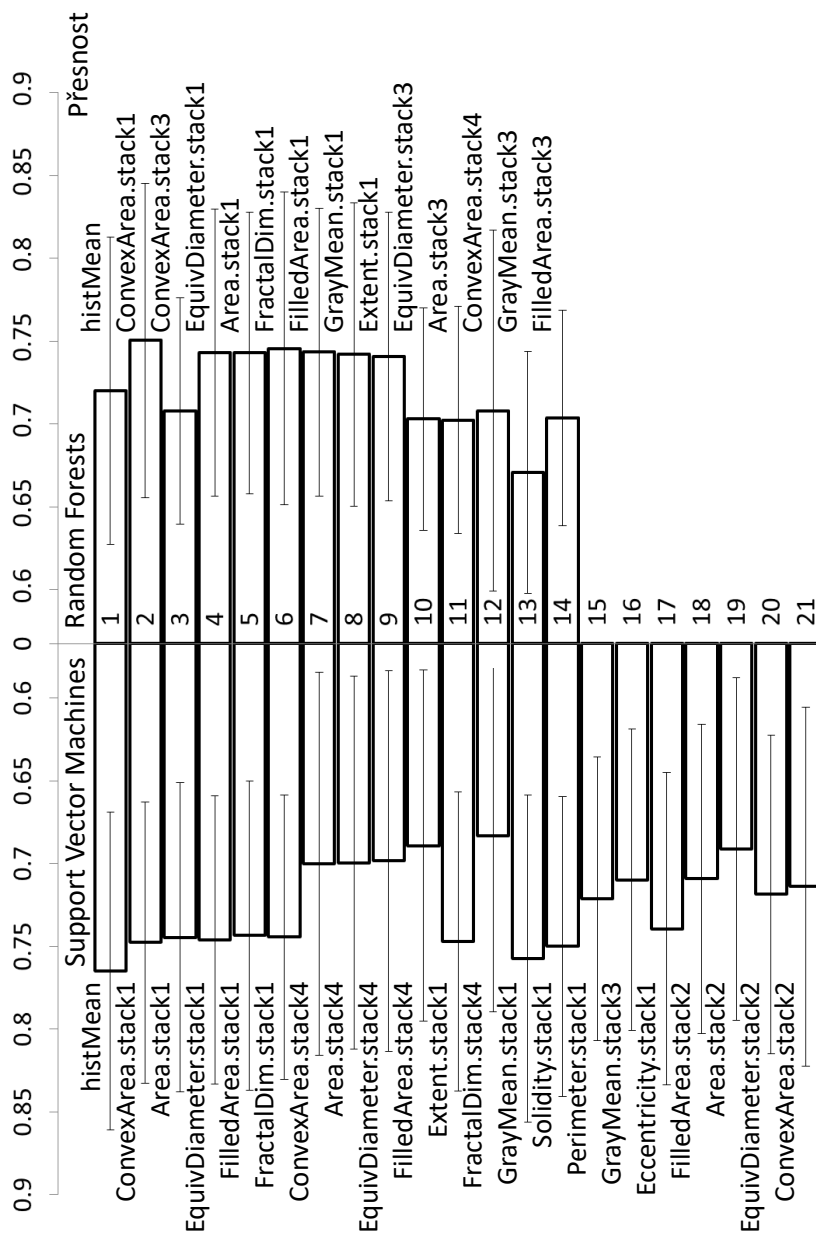
Hrubá segmentace – 75% obrazových čtverečků					
Klasifikátor	Přesnost	Senzitivita	Specifická	ROC AUC	Chybně klasifikováno
RF	96,7 %	97,5 %	95	0,999	1 Philips, 1 GE
SVM	95 %	95 %	95	0,998	2 Philips, 1 GE

Tabulka 6: Výsledky klasifikace všech 60 uzlů pro hrubě segmentované uzly (75 % celkové plochy uzlu). Poslední sloupec ukazuje, kolik uzlů bylo chybně klasifikováno a kterým přístrojem byly tyto chybně diagnostikované uzly pořízeny.

Jak již bylo řečeno, ne všechny prediktory přispívají stejnou měrou ke klasifikačnímu procesu. Prediktory můžeme řadit podle významnosti a následně z klasifikačního modelu vyřadit ty, které jsou nejméně důležité. Tabulka 2 nám ukazuje pořadí jednotlivých prediktorů seřazených pomocí metody RFE a obrázek 14 nám ukazuje, jak se mění klasifikační metrika AUC ROC v závislosti na počtu použitých prediktorů. Výsledky této metody nám ale neřeknou, jak důležité jsou jednotlivé prediktory samy o sobě.

Jednou z metod, jak kvantifikovat významnost jednotlivých prediktorů, je analýza s jednou proměnnou (*univariate analysis*). Jedná se o jednoduchou metodu, kdy je v klasifikačním modelu použita jen jedna proměnná (prediktor) a vyhodnocuje se její schopnost rozdělit případy do jednotlivých tříd. Na obrázku 16 vidíme výsledky analýzy s jednou proměnnou pro ty prediktory, které byly metodou RFE vyhodnoceny jako nejvýznamnější, tedy pro 14 prediktorů klasifikátoru RF a pro 21 prediktorů klasifikátoru SVM. Mohli bychom předpokládat, že přesnost jednotlivých prediktorů na obrázku 16 bude růst s rostoucím pořadovým číslem podle výsledků analýzy RFE, tedy že na grafu uvidíme klesající hodnoty přesnosti s klesajícím pořadovým číslem daným výsledky RFE. Nemusí tomu ale tak být, protože metoda RFE pracuje na základě zpětné eliminace těch nejméně významných prediktorů daného modelu, což může být i prediktor, který koreluje s jiným prediktorem. Takže i kdyby mohl sám o sobě být významný a při analýze s jednou proměnnou by dosáhl vysoké přesnosti, nemusí se mezi nejdůležitější prediktory pomocí RFE dostat. Cestou při výběru prediktorů pro stavbu co nejefektivnějšího modelu tedy není výběr prediktorů na

základě analýzy s jednou proměnnou, ale spíše použití nějakého nástroje přímo k tomu určeného, například právě metody RFE. Výsledky analýzy s jednou proměnnou na obrázku 16 slouží spíše pro zhodnocení potenciálu jednotlivých prediktorů a tím, že prediktory vždy něco reprezentují, nám mohou říci, jaká charakteristika textury uzlu je nejdůležitější, čehož můžeme využít v dalším výzkumu.



Obrázek 16: Graf znázorňující výsledky analýzy s jednou proměnnou pro náhodné lesy a SVM. Graf ukazuje nejdůležitější prediktory podle metody zpětné eliminace prediktorů (RFE) seřazených odshora dolů, kdy 21 nejlépe hodnocených prediktorů SVM je zobrazeno nalevo a 14 nejlepších prediktorů náhodných lesů je zobrazeno napravo. Délka horizontálně natočených sloupců ukazuje průměrnou přesnost příslušného prediktoru a chybové úsečky zobrazují směrodatnou odchylku skupinové 10-fold křížové validace. Stack1 – stack4 značí pořadí šedotónových pásem. Nepřítomnost vyšších pásem značí nevýznamnost světlých částí obrazových čtverečků.

4 DISKUSE

Výzkum počítačem podporované diagnostiky uzlů ve štítné žláze se v posledním desetiletí zaměřoval především na hledání nových, dostatečně přesných metod, které by byly schopny buď zcela automatické diagnostiky, nebo by poskytly nový vhled do diagnostiky tím, že by představily nový rizikový faktor malignity uzlů ve štítné žláze. Přehled těch nejvýznamnějších prací na téma počítačové analýzy ultrazvukových obrazů uzlů ve štítné žláze vidíme v tabulce 7, kde se v posledních dvou řádcích nachází přehled a výsledky našich dvou v této práci předkládaných postupů, tzn. jak klasifikace celých uzlů, tak klasifikace pomocí patch-based metody. Výsledky naší práce jsou srovnatelné s výsledky jiných prací publikovaných na toto téma v minulých letech. Z výsledků je zřejmé, že jak klasifikace uzlů jako celku, tak patch-based klasifikace jsou schopny vyhodnotit z použité množiny ultrazvukových obrazů podstatné informace o probíhajícím patologickém procesu v uzlech štítné žlázy.

Jak je vidět v tabulce 2 a na obrázku 16, průměrná hodnota histogramu (HistMean) byla u patch-based klasifikace vyhodnocena pomocí RFE jako nejvýznamnější prediktor jak u SVM, tak u náhodných lesů. U SVM to byl dokonce i nejučinnější prediktor nejen podle výsledků RFE, ale také podle výsledků analýzy s jednou proměnnou (obrázek 16). Vysoká významnost tohoto prediktoru odpovídá dříve publikovaným výsledkům jak studií zabývajících se texturní analýzou (Kim et al., 2015, Chang et al., 2016), tak studií, ve kterých lézi hodnotili radiologové (Haugen et al., 2016, Remonti et al., 2015, Kim et al., 2002, Moon et al., 2007, Kim et al., 2015, Trimboli et al., 2012), kde snížené hodnotě průměrné hodnoty histogramu odpovídá hypoechogenita maligního uzlu.

Dále pak na obrázku 16 vidíme, že hlavní přínos binární stack-dekompozice leží ve 4 spodních šedotónových pásmech odpovídajících nejtmavším částem obrazu. Z prediktorů popisujících region v daném šedotónovém pásmu vycházejí jako nejvýznamnější ConvexArea, EquivDiameter, FilledArea, a dále pak tři původní prediktory publikované spolu s algoritmem SFTA (Costa et al., 2012), tedy FractalDim, GrayMean a Area. Všechny tyto prediktory nabírají na důležitosti

především v 1., 3. a 4. šedotónovém pásmu. Jednoznačně nejdůležitějším pásmem je ale stack1, tedy úplně nejtmavší část obrazu. Jak ukazují výsledky analýzy s jednou proměnou (obrázek 16), tak v tomto pásmu dosahuje celkem 8 prediktorů při použití SVM a 7 prediktorů při použití náhodných lesů přesnosti nad 74 %. Navíc tyto prediktory prošly přes RFE do výběru nejdůležitějších prediktorů, z čehož lze usuzovat, že spolu korelují méně než s ostatními prediktory. Když to tedy shrneme, při patch-based texturní analýze se nejvíce uplatňují nejtmavší části obrazu, přičemž nestačí jen výskyt tmavých pixelů v uzlu štítné žlázy (tzn. nestačí hodnotit pouze hypoechogenitu), ale důležitá je také složitost tvaru regionu tvořeného těmito tmavými pixely.

Parametry histogramu hrály důležitou roli také u klasifikace celých uzlů. Jak je vidět na obrázku 12, HistSkew a HistKurt patří mezi nejdůležitější parametry, což je opět ve shodě s dříve publikovanými pracemi (Chang et al., 2016, Kim et al., 2015) a stejně jako u patch-based analýzy i zde největší síla spočívala v analýze tmavých částí obrazu, což dokazuje vysoká významnost prediktorů v prvním a druhém šedotónovém pásmu (stack1, stack2), kdy nejdůležitějšími vlastnostmi uzlu byla jednak celková velikost tvarů v těchto tmavých částech (Area) a jednak komplexita tvarů, které se v těchto tmavých pásmech vyskytují, popsána fraktální dimenzí (FracDim).

Na druhou stranu, výsledky ukazují malou důležitost informace vyskytující se ve světlých pásmech obrazu, a to jak v případě analýzy celých uzlů, tak v případě patch-based analýzy, což neodpovídá předpokladům. Zde by se totiž měly vyskytnout mikro-/makrokalcifikace, které jsou považovány za rizikový faktor malignity (Haugen et al., 2016, Remonti et al., 2015, Brito et al., 2014). Výsledky sice na žádnou souvislost mezi světlými pásmy a malignitou neukazují, je to však pravděpodobně způsobeno malým množstvím dat a nízkou relativní četností calcifikací v uzlech štítné žlázy.

Práce	Prediktory	Klasifikátor	Počet uzlů	Výsledky
(Iakovidis et al., 2010)	Fuzzy LBP, fuzzy intensity histograms	SVM	75 (uzly byly porovnávány s fyziologickou tkání ŠŽ)	AUC ROC: 97,5 %
(Ding et al., 2011)	Statické a texturní prediktory	SVM	125 (69 benigních, 56 maligních)	Přesn. 93,6 %
(Acharya et al., 2011)	Texturní a DWT prediktory získané z CEUS obrazů	kNN	20 (10 benigních, 10 maligních)	Přesn. 98,9 % Senz. 98 % Spec. 99,8 %
(Acharya et al., 2012a)	Texturní a DWT prediktory	Ada-Boost	20 (10 benigních, 10 maligních)	Přesn. 100 %
(Acharya et al., 2012c)	Fraktální dimenze, LBP, Fourier spectrum deskriptor, LTE; získané z 3D HRUS a CEUS	GMM, SVM	20 (10 benigních, 10 maligních)	Přesn. 98,1–100 %
(Chang et al., 2016)	Texturní prediktory	SVM	59 (29 benigních, 30 maligních)	Přesn. 98,3 %
(Acharya et al., 2016)	Různé typy texturní entropie: Fuzzy, Shannon, Renyi, HOS, Kapur, Vajda	C4.5	242 (211 benigních, 31 maligních), nevyváženost tříd řešena oversamplingem	Přesn. 94,3 %
(Raghavendra et al., 2017)	Prediktory prostorové závislosti šedotónových struktur, fraktální prediktory, SFTA	SVM	242 (211 benigních, 31 maligních) – nevyváženost tříd řešena oversamplingem	Přesn. 97,52 % AUC ROC: 0,9445

(Chi et al., 2017)	Prediktory hlubokého učení získané pomocí předtrénované GoogLeNet neuronové sítě	RF	Dva datové soubory čítající 357 a 164 obrazů	Přesn. 98,29 % a 96,34 %
(Liu et al., 2017a)	Předtrénovaná VGG-F neuronová síť, LBP, šedotónové matice spoluvýskytu	SVM	651 benigních a 386 maligních snímků	Přesn. 93,1 %
(Prochazka et al., 2019a)	Prediktory získané z histogramu, SFTA prediktory	SVM RF	40 (20 maligních, 20 benigních)	Přesn. RF: 92,5 % SVM: 95 %
(Prochazka et al., 2019b)	Patch-based klasifikace pomocí prediktorů histogramu a prediktorů získaných pomocí BSD	RF SVM	60 (20 maligních, 40 benigních) - nevyváženost tříd řešena pomocí undersamplingu na úrovni patchů	Přesn. RF: 95 % SVM: 91,6 %
Práce	Prediktory	Klasifikátor	Počet uzlů	Výsledky

Tabulka 7. Přehledová tabulka ukazující srovnání všech zajímavých v poslední době publikovaných prací na téma počítačové podpory diagnostiky uzlů ve štítné žláze. Vysvětlivky: Přesn. – přesnost, Senz. – senzitivita, Spec. – specificita, AUC ROC – plocha pod křivkou ROC, dále pak BSD – binární stack-dekompozice, CEUS – kontrastem obohacená ultrasonografie, DWT – diskretní waveletová transformace, GMM - *Gaussian mixture models*, HOS - *higher order spektra*, kNN – algoritmus k-nejbližších sousedů, LBP – *local binary patterns*, LTE - Lawsova texturní energie, RF – náhodné lesy, SFTA – *segmentation-based fractal texture analysis*, SVM – *support vector machines*, 3D HRUS – trojdimenzionální ultrazvuk s vysokým rozlišením.

Velkým nedostatkem dosud publikovaných CAD systémů pro štítnou žlázu je fakt, že dosud neexistuje dostatečně přesná segmentační metoda, pomocí které by bylo možné automaticky označit uzel v ultrazvukovém obraze. V současné době, pokud je nám známo, existují pouze dva segmentační algoritmy pro uzly ve štítné

žláze (Tsantis et al., 2006, Chang et al., 2016), které jsou oba schopny označit oblast uzlu s přibližnou 90% přesností. Nicméně tyto algoritmy byly použity pouze v rámci jednotlivých studií na omezeném množství dat, a proto zůstává otázkou, zda jsou dostatečně evaluovány. Ostatní studie (včetně této práce) spoléhají na manuální segmentaci lékařem, která ovšem, pokud se lékař snaží o co největší přesnost, vyžaduje poměrně dost času a trpělivosti. Zde se rýsuje další možná výhoda patch-based metody, a to ta, že tato metoda nepotřebuje mít k přesné klasifikaci přesně ohraničený uzel, protože sama o sobě tvar uzlu nehodnotí. Z výsledků je zřejmé, že pokud nevezmeme všechny obrazové čtverečky z uzlu, ale pouze 75 % z nich, přesnost klasifikace se nesnižuje, dokonce se lehce zvyšuje. Je tedy zjevné, že tím, že patch-based metoda zkoumá jemnou texturu uzlu na malém měřítku, tak není závislá na přesnosti segmentace uzlu. Vstupem pro patch-based systém tedy může být i nedokonale označená oblast, což je výhodná vlastnost při použití v klinické praxi, kdy by lékaři stačilo označit uzel jen zběžně, např. pomocí zanesení několika bodů, kterými by byl proložen polynom.

Hlavní limitací této práce zůstává malý počet obrázků, které, i když byla zkoumána jejich textura na velmi podrobné úrovni, stále nemohou reprezentovat vzhledovou diverzitu veškerých uzlů, které se mohou ve štítné žláze vyskytovat. Předkládané výsledky mohou spíše ukázat novou cestu, jakou lze při tvorbě CAD pro uzly ve štítné žláze jít. Práce ukazuje, že k hledání optimálního klasifikačního modelu lze použít obrazové čtverečky a následně lze uzel hodnotit jako celek zprůměrováním výsledků, a tím odhadnout riziko malignity. Teoreticky by se patch-based klasifikace mohla kombinovat s jinými postupy a riziko malignity vyjádřeno hodnotou *IRM* by mohlo sloužit jako samostatný prediktor jiného klasifikačního modelu. To vše jsou ale spíše vyhlídky do budoucna, kdy by bylo možné zkoumat klasifikační modely na větším množství dat. Sběr většího množství dat je do budoucna klíčový a v našem případě to zřejmě bude vyžadovat spolupráci s dalším klinickým pracovištěm či dalšími klinickými pracovišti, jelikož četnost maligních uzlů v populaci je nízká.

Z přehledu v tabulce 7 je znát, že jen za dobu posledních deseti let se automatické diagnostice uzlů ve štítné žláze věnovalo poměrně úspěšně několik týmů. Přístupy většiny autorských kolektivů využívaly metodu strojového učení,

kteřou bychom mohli nazvat jako klasickou. Tato metoda spočívá v tom, že autor sám se snaží najít ty nejvhodnější prediktory, o kterých je předem přesvědčen, že z nějakého důvodu mohou fungovat (například popisují nějakou konkrétní, i viditelnou, vlastnost uzlu), nebo jen do určité míry náhodně zkusí použít metody, které by mohly pro daný problém extrahovat z obrazu nějakou užitečnou informaci. Naproti tomuto klasickému postupu ale stojí hluboké učení, které je sice známo již několik desetiletí, ale masivně se rozšířilo teprve v posledních několika letech. Principem hlubokého učení je, že nikoli člověk, ale sám algoritmus pomocí vzájemně propojených vrstev extrahuje ty vlastnosti, které jsou pro daný klasifikační problém důležité. V případě klasifikace obrazu jsou tedy vstupem přímo pixely. Kenji Suzuki (2017) podává ve své publikaci přehled použití hlubokého učení v oblasti zpracování medicínského obrazu a z jeho práce plyne, že zatím nejde říct, jestli na tomto poli hluboké učení přebírá roli klasických klasifikačních modelů nebo ne. V současné chvíli jsme v situaci, kdy metody hlubokého učení dosahují podobných výsledků jako klasické klasifikátory. Suzuki ale uzavírá, že se dá předpokládat, že v následujících několika dekádách bude hluboké učení ve zpracování medicínského obrazu hlavním proudem. Důvod, proč zatím ve zpracování medicínského obrazu hluboké učení nepřevládá, spočívá pravděpodobně v tom, že v medicíně je obecně složitější získat dostatek relevantních dat než je tomu v jiných oblastech, kde se již hluboké učení prosadilo. U problémů rozpoznávání medicínského obrazu, kde se podařilo nasbírat velké množství dat, hluboké učení jednoznačně vede, jak bylo v poslední době publikováno například u počítačem podporované diagnostiky karcinomu kůže (Esteva et al., 2017) nebo diabetické retinopatie (Gulshan et al., 2016). Vedle dalšího zdokonalování klasických klasifikačních modelů, je hluboké učení jednoznačně cesta, kterou stojí za to se na poli CAD systémů pro štítnou žlázu v budoucnu vydat.

5 ZÁVĚR

Tato práce se zabývá počítačem podporovanou diagnostikou uzlů ve štítné žláze. V úvodu byla popsána specifika tohoto onemocnění a jeho diagnostiky v souladu se současnými doporučenými postupy. Především na studiích zabývajících se screeningem a na meta-analýzách zabývajících se odhadováním rizika v jednotlivých rizikových kategoriích je vidět, že současné metody diagnostiky stále nejsou dostatečně přesné. Ani odhad rizika na základě cytologického vyšetření není uspokojivě přesný a není tedy divu, že se stále hledají metody, které by do diagnostiky vnesly více jistoty. Výzkum v této oblasti si klade za cíl snížit množství nežádoucích operací falešně pozitivních nálezů, a to především proto, že operace s sebou nese rizika, jako jsou paralýza hlasivek či hypoparatyroidismus. V současné době se výzkum zaměřuje hlavně na neinvazivní metody, jako jsou nové metody ultrasonografie nebo specifické biomarkery. Z nových metod ultrasonografie se ale nejedná pouze o nové „fyzikální“ metody jako je například použití kontrastu či real-time a shear-wave elastografie, ale také o „softwarové“ metody počítačem podporované diagnostiky. Především u pacientů s opakovaně nediodagnostikovatelými (Bethesda 1), benigními limitovanými, či suspektními uzly (Bethesda 3–5) by mohly tyto nové metody přinést nový pohled na věc, a tím poskytnout lékařům další informaci, což může zpřesnit výběr vhodných pacientů pro FNAB či operaci.

V poslední dekádě bylo vyvinuto několik velmi přesných systémů počítačové podpory diagnostiky uzlů ve štítné žláze, které byly diskutovány v úvodu. Hlavní limitací většiny dosud publikovaných systémů je především to, že pracují se směrovými charakteristikami ultrazvukového obrazu a jsou tedy použitelné (resp. verifikované) pouze ve statickém obraze pořízeném v jedné běžně zobrazované rovině (axiální či longitudinální). Další vlastností, kterou lze považovat za limitující, je, že tyto systémy často duplikují práci lékaře sonografisty a hodnotí uzel jako celek pomocí všech dostupných rizikových faktorů, které běžně pomocí ultrazvuku hodnotí lékař, čímž jsou sice schopny dostatečně přesně diagnostikovat uzel, ale nedávají lékařům do ruky novou metodu, kterou by mohl zařadit po bok těch, které k diagnostice běžně používá. Z těchto důvodů jsme se zaměřili na nesměrové

charakteristiky obrazu a dále pak na patch-based analýzu, která hodnotí jemnou texturu, což je vlastnost uzlu, která je při běžném ultrazvukovém vyšetření jen obtížně kvantifikovatelná.

V praktické části této práce byly představeny dva CAD přístupy sloužící k diagnostice uzlů ve štítné žláze z ultrazvukového obrazu. Oba přístupy vykazují podobně dobré výsledky (při optimálním nastavení 95% přesnost), přičemž patch-based přístup lze považovat za robustnější, protože pracuje s větším množstvím ultrazvukových snímků. Navíc tím, že se patch-based přístup nezabývá primárně uzlem jako celkem, ale zaměřuje se na analýzu jemné textury po malých čtverečcích ultrazvukového obrazu, tak přináší do klinické praxe nový diagnostický faktor hodnotící vnitřní strukturu uzlu, která se při běžném sonografickém vyšetření takto podrobně nehodnotí. Proto by doporučení těchto systémů mohlo sloužit lékařům jako přídavný diagnostický faktor, který by ze své podstaty neměl korelovat s ostatními rizikovými vlastnostmi uzlu, na které se lékař při sonografickém vyšetření zaměřuje (velikost, charakter, tvar, ohraničení, apod.). Tím odhad rizika na základě textury (třeba ve formě předloženého Indexu rizika malignity, *IRM*) přináší do diagnostiky novou, dosud v klinické praxi nehodnocenou informaci, která může být zařazena po bok klasické ultrasonografie. Tím může naše metoda vést ke zpřesnění výběru pacientů pro následnou biopsii tenkou jehlou či operaci.

SOUHRN

Ultrazvukové zobrazování patří mezi základní vyšetření uzlů ve štítné žláze, na jejichž základě se rozhoduje, zda pacient podstoupí cytologické vyšetření, které je hlavním podkladem pro rozhodování o případném chirurgickém odstranění štítné žlázy. Cytologické vyšetření má ale bohužel omezenou specificitu a případná operace s sebou nese rizika. Proto jsou hledány další metody, které by byly schopny vnést do diagnostiky více jistoty. Jednou z nových metod je počítačová podpora diagnostiky (CAD), která pomocí analýzy obrazu a strojového učení vykazuje poměrně slibné výsledky. V této práci představujeme dva do určité míry podobné, přesto však odlišné, CAD přístupy. První přístup spočívá v analýze celých uzlů pomocí *Segmentation Based Fractal Texture Analysis* (SFTA) algoritmu, který rozkládá obraz na jednotlivá šedotónová pásma pomocí metody binární stack-dekompozice. Pomocí tohoto přístupu bylo na datovém souboru 40 snímků hodnocených metodou křížové validace dosaženo přesnosti 92,5 % při použití náhodných lesů a 95 % při použití *support vector machines* (SVM). Druhý CAD přístup vychází také z metody vícenásobného prahování obrazu, ale s tím rozdílem, že z jednotlivých šedotónových pásem je extrahováno větší množství prediktorů popisujících binární texturu a dále pak, že analýza neprobíhá na uzlu jako celku, ale na malých čtverečcích (patch-based přístup) o velikosti 17×17 px. Tímto postupem bylo dosaženo klasifikační přesnosti jednotlivých čtverečků hodnocené metodou křížové validace 74,4 %. Po zprůměrování výsledků klasifikace všech čtverečků v jednotlivých uzlech bylo na 60 snímcích dosaženo přesnosti 91,6 % při použití SVM a 95,0 % při použití náhodných lesů. Vzhledem k tomu, že se patch-based přístup nezabývá primárně uzlem jako celkem, ale zaměřuje se na analýzu jemné textury, která je pro lékaře těžko kvantifikovatelná, tak přináší do klinické praxe novou informaci, která může pomoci k přesnější diagnostice uzlů ve štítné žláze pomocí ultrazvuku, a tím může částečně snížit počet falešně pozitivních nálezů indikovaných k cytologickému vyšetření či operaci.

SUMMARY

Ultrasound imaging is one of the fundamental examinations of thyroid nodules, determining whether a patient undergoes a cytological examination, which is essential for the decision on a possible thyroid surgery. Unfortunately, the cytological examination has limited specificity and potential surgery carries risks. Therefore, other diagnostic methods are being sought with hope that they will be able to bring more certainty into diagnostics. One of the new methods is computer-aided diagnosis (CAD), which exhibits promising results using image analysis and machine learning. In this study, we present two somewhat similar, yet different, CAD approaches. The first approach is based on analysing entire nodules using a Segmentation Based Fractal Texture Analysis (SFTA) algorithm that splits the image into individual grayscale bands. Using this approach, we have achieved an accuracy of 92.4% using random forests (RF) and 95% using support vector machines (SVM) on a data set of 40 images evaluated by the cross-validation method. The second CAD approach is also based on the method of multiple image thresholding, but the difference is, that a larger number of predictors describing the binary texture are extracted from the individual grayscale bands. Furthermore, the analysis did not take place on whole nodules, but on patches of a size 17×17 px. Using this approach, we have achieved an accuracy of 74.4% when classifying individual patches. After averaging the results of classification of all patches in each nodule, an accuracy of 91.6% was achieved using an SVM classifier and 95.0% using an RF classifier on a dataset of 60 images. Since the patch-based approach does not deal primarily with the nodule as a whole, but focuses on the fine texture analysis in small patches, it brings new, clinically relevant information into clinical practice, which may contribute to increase an accuracy of thyroid ultrasound, and thus can partly reduce false positive findings indicated for cytology or surgery.

POUŽITÁ LITERATURA

- National Cancer Institute. *Cancer Stat Facts: Thyroid Cancer* [Online]. Available: <https://seer.cancer.gov/statfacts/html/thyro.html> [Accessed 23. 6. 2017].
- UCI Repository of Machine Learning Databases, University of Columbia at Irvine, Department of Computer Science [Online]. Available: <ftp://ftp.ics.uci.edu/pub/machine-learning-databases> [Accessed 30.6.2017].
- AHN, H. S., KIM, H. J. & WELCH, H. G. 2014. Korea's thyroid-cancer "epidemic"--screening and overdiagnosis. *N Engl J Med*, 371, 1765-7.
- ACHARYA, U. R., FAUST, O., SREE, S. V., MOLINARI, F., GARBEROGLIO, R. & SURI, J. S. 2011. Cost-effective and non-invasive automated benign and malignant thyroid lesion classification in 3D contrast-enhanced ultrasound using combination of wavelets and textures: a class of ThyroScan algorithms. *Technol Cancer Res Treat*, 10, 371-80.
- ACHARYA, U. R., FAUST, O., SREE, S. V., MOLINARI, F. & SURI, J. S. 2012a. ThyroScreen system: high resolution ultrasound thyroid image characterization into benign and malignant classes using novel combination of texture and discrete wavelet transform. *Comput Methods Programs Biomed*, 107, 233-41.
- ACHARYA, U. R., CHOWRIAPPA, P., FUJITA, H., BHAT, S., DUA, S., KOH, J. E. W., EUGENE, L. W. J., KONGMEBHOL, P. & NG, K. H. 2016. Thyroid lesion classification in 242 patient population using Gabor transform features from high resolution ultrasound images. *Knowledge-Based Systems*, 107, 235-245.
- ACHARYA, U. R., SREE, S. V., KRISHNAN, M. M. R., MOLINARI, F., GARBEROGLIO, R. & SURI, J. S. 2012b. Non-invasive automated 3D thyroid lesion classification in ultrasound: A class of ThyroScan (TM) systems. *Ultrasonics*, 52, 508-520.
- ACHARYA, U. R., SWAPNA, G., SREE, S. V., MOLINARI, F., GUPTA, S., BARDALES, R. H., WITKOWSKA, A. & SURI, J. S. 2014. A review on ultrasound-based thyroid cancer tissue characterization and automated classification. *Technol Cancer Res Treat*, 13, 289-301.
- ACHARYA, U. R., VINITHA SREE, S., KRISHNAN, M. M., MOLINARI, F., GARBEROGLIO, R. & SURI, J. S. 2012c. Non-invasive automated 3D thyroid lesion classification in ultrasound: a class of ThyroScan systems. *Ultrasonics*, 52, 508-20.
- ALEXANDER, E. K., HEERING, J. P., BENSON, C. B., FRATES, M. C., DOUBILET, P. M., CIBAS, E. S. & MARQUSEE, E. 2002. Assessment of nondiagnostic ultrasound-guided fine needle aspirations of thyroid nodules. *J Clin Endocrinol Metab*, 87, 4924-7.
- ALON HALEVY, P. N., AND FERNANDO PEREIRA 2009. The Unreasonable Effectiveness of Data. *IEEE Intelligent Systems*, 24, 8-12.
- BALOGH, Z. W., LIVOLSI, V. A., ASA, S. L., ROSAI, J., MERINO, M. J., RANDOLPH, G., VIELH, P., DEMAY, R. M., SIDAWY, M. K. & FRABLE, W. J. 2008. Diagnostic terminology and morphologic criteria for cytologic diagnosis of thyroid lesions: a synopsis of the National Cancer Institute Thyroid Fine-Needle Aspiration State of the Science Conference. *Diagn Cytopathol*, 36, 425-37.

- BIBBINS-DOMINGO, K., GROSSMAN, D. C., CURRY, S. J., BARRY, M. J., DAVIDSON, K. W., DOUBENI, C. A., EPLING, J. W., JR., KEMPER, A. R., KRIST, A. H., KURTH, A. E., LANDEFELD, C. S., MANGIONE, C. M., PHIPPS, M. G., SILVERSTEIN, M., SIMON, M. A., SIU, A. L. & TSENG, C. W. 2017. Screening for Thyroid Cancer: US Preventive Services Task Force Recommendation Statement. *JAMA*, 317, 1882-1887.
- BONGIOVANNI, M., SPITALE, A., FAQUIN, W. C., MAZZUCHELLI, L. & BALOCH, Z. W. 2012. The Bethesda System for Reporting Thyroid Cytopathology: a meta-analysis. *Acta Cytol*, 56, 333-9.
- BRANCO, P., TORGO, L. & RIBEIRO, R. 2016. A Survey of Predictive Modelling under Imbalanced Distributions. *ACM Computing Surveys*, 49, 31-48.
- BREIMAN, L. 2001. Random forests. *Machine Learning*, 45, 5-32.
- BREIMAN, L., FRIEDMAN, J. H., OLSHEN, R. A. & STONE, C. J. 1984. *Classification and Regression Trees*, Monterey, CA, USA, Wadsworth and Brooks.
- BRITO, J. P., GIONFRIDDO, M. R., AL NOFAL, A., BOEHMER, K. R., LEPPIN, A. L., READING, C., CALLSTROM, M., ELRAIYAH, T. A., PROKOP, L. J., STAN, M. N., MURAD, M. H., MORRIS, J. C. & MONTORI, V. M. 2014. The accuracy of thyroid nodule ultrasound to predict thyroid cancer: systematic review and meta-analysis. *J Clin Endocrinol Metab*, 99, 1253-63.
- CAPPOLA, A. R. 2017. How to Look for Thyroid Cancer. *JAMA*, 317, 1840-1841.
- COOMANS, D., BROECKAERT, I., JONCKHEER, M. & MASSART, D. L. 1983. Comparison of multivariate discrimination techniques for clinical data--application to the thyroid functional state. *Methods Inf Med*, 22, 93-101.
- CORTES, C. & VAPNIK, V. 1995. Support-Vector Networks. *Machine Learning*, 20, 273-297.
- COSTA, A. F., HUMPIRE-MAMANI, G. T. & TRAINA, A. J. M. 2012. An Efficient Algorithm for Fractal Analysis of Textures. *Proceedings of the 2012 25th SIBGRAPI Conference on Graphics, Patterns and Images*, 39-46.
- CRIPPA, S., MAZZUCHELLI, L., CIBAS, E. S. & ALI, S. Z. 2010. The Bethesda System for reporting thyroid fine-needle aspiration specimens. *Am J Clin Pathol*, 134, 343-4; author reply 345.
- DAL POZZOLO, A., CAELEN, O., WATERSCHOOT, S. & BONTEMPI, G. 2013. Racing for Unbalanced Methods Selection. In: YIN, H., TANG, K., GAO, Y., KLAWONN, F., LEE, M., LI, B., WEISE, T. & YAO, X. (eds.) *Intelligent Data Engineering and Automated Learning - Ideal 2013*. New York: Springer.
- DENG, J., DONG, W., SOCHER, R., LI, L., LI, K. & LI, F. F. 2009. ImageNet: a Large-Scale Hierarchical Image Database. *IEEE Conference on Computer Vision and Pattern Recognition*. 248-255.
- DI, Y., LI, M. Y., QIAO, T. & LU, N. 2017. Edge detection and mathematic fitting for corneal surface with Matlab software. *International Journal of Ophthalmology*, 10, 336-342.
- DING, J. R., CHENG, H. D., NING, C. P., HUANG, J. H. & ZHANG, Y. T. 2011. Quantitative Measurement for Thyroid Cancer Characterization Based on Elastography. *Journal of Ultrasound in Medicine*, 30, 1259-1266.

- DOGANTEKIN, E., DOGANTEKIN, A. & AVCI, D. 2010. An automatic diagnosis system based on thyroid gland: ADSTG. *Expert Systems with Applications*, 37, 6368-6372.
- DURANTE, C., COSTANTE, G., LUCISANO, G., BRUNO, R., MERINGOLO, D., PACIARONI, A., PUXEDDU, E., TORLONTANO, M., TUMINO, S., ATTARD, M., LAMARTINA, L., NICOLUCCI, A. & FILETTI, S. 2015. The natural history of benign thyroid nodules. *JAMA*, 313, 926-35.
- EDAFE, O., ANTAKIA, R., LASKAR, N., UTTLEY, L. & BALASUBRAMANIAN, S. P. 2014. Systematic review and meta-analysis of predictors of post-thyroidectomy hypocalcaemia. *British Journal of Surgery*, 101, 307-320.
- ESTEVA, A., KUPREL, B., NOVOA, R. A., KO, J., SWETTER, S. M., BLAU, H. M. & THRUN, S. 2017. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542, 115-+.
- FERNANDEZ-DELGADO, M., CERNADAS, E., BARRO, S. & AMORIM, D. 2014. Do we Need Hundreds of Classifiers to Solve Real World Classification Problems? *Journal of Machine Learning Research*, 15, 3133-3181.
- FRANK, E., HALL, M. A. & WITTEN, I. H. 2016. *The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques"*, Morgan Kaufmann.
- GUL, K., ERSOY, R., DIRIKOC, A., KORUKLUOGLU, B., ERSOY, P. E., AYDIN, R., UGRAS, S. N., BELEMLI, O. K. & CAKIR, B. 2009. Ultrasonographic evaluation of thyroid nodules: comparison of ultrasonographic, cytological, and histopathological findings. *Endocrine*, 36, 464-72.
- GULSHAN, V., PENG, L., CORAM, M., STUMPE, M. C., WU, D., NARAYANASWAMY, A., VENUGOPALAN, S., WIDNER, K., MADAMS, T., CUADROS, J., KIM, R., RAMAN, R., NELSON, P. C., MEGA, J. L. & WEBSTER, R. 2016. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *Jama-Journal of the American Medical Association*, 316, 2402-2410.
- GUTH, S., THEUNE, U., ABERLE, J., GALACH, A. & BAMBERGER, C. M. 2009. Very high prevalence of thyroid nodules detected by high frequency (13 MHz) ultrasound examination. *Eur J Clin Invest*, 39, 699-706.
- HABIBZADEH, M., KRZYŻAK, A. & FEVENS, T. White Blood Cell Differential Counts Using Convolutional Neural Networks for Low Resolution Images. 2013 Berlin, Heidelberg. Springer Berlin Heidelberg, 263-274.
- HALL, T. L., LAYFIELD, L. J., PHILIPPE, A. & ROSENTHAL, D. L. 1989. Sources of diagnostic error in fine needle aspiration of the thyroid. *Cancer*, 63, 718-25.
- HASSANIEN, A. E. & OLIVA, D. A. 2017. *Advances in Soft Computing and Machine Learning in Image Processing*, Springer International Publishing.
- HAUGEN, B. R. 2017. 2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer: What is new and what has changed? *Cancer*, 123, 372-381.
- HAUGEN, B. R., ALEXANDER, E. K., BIBLE, K. C., DOHERTY, G. M., MANDEL, S. J., NIKIFOROV, Y. E., PACINI, F., RANDOLPH, G. W., SAWKA, A. M., SCHLUMBERGER, M., SCHUFF, K. G., SHERMAN, S. I., SOSA, J. A., STEWARD, D. L., TUTTLE, R. M. & WARTOFSKY, L. 2016. 2015 American Thyroid Association Management Guidelines for Adult

- Patients with Thyroid Nodules and Differentiated Thyroid Cancer: The American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer. *Thyroid*, 26, 1-133.
- HAYS, J. & EFROS, A. A. 2007. Scene Completion Using Millions of Photographs. *ACM Transactions on Graphics*, 26, 4-2 - 4-7.
- HEGEDUS, L. 2004. Clinical practice. The thyroid nodule. *N Engl J Med*, 351, 1764-71.
- CHANG, C.-C. A. L., CHIH-JEN 2011. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2, 27:1--27:27.
- CHANG, Y., PAUL, A. K., KIM, N., BAEK, J. H., CHOI, Y. J., HA, E. J., LEE, K. D., LEE, H. S., SHIN, D. & KIM, N. 2016. Computer-aided diagnosis for classifying benign versus malignant thyroid nodules based on ultrasound images: A comparison with radiologist-based assessments. *Medical Physics*, 43, 554-567.
- CHEN, Y. Q., NIXON, M. S. & THOMAS, D. W. 1995. Statistical Geometrical Features for Texture Classification. *Pattern Recognition*, 28, 537-552.
- CHI, J., WALIA, E., BABYN, P., WANG, J., GROOT, G. & ERAMIAN, M. 2017. Thyroid Nodule Classification in Ultrasound Images by Fine-Tuning Deep Convolutional Neural Network. *J Digit Imaging*, 30, 477-486.
- IAKOVIDIS, D. K., KERAMIDAS, E. G. & MAROULIS, D. 2010. Fusion of fuzzy statistical distributions for classification of thyroid ultrasound patterns. *Artif Intell Med*, 50, 33-41.
- ISO/IEC 2018a. ISO International Standard ISO/IEC 9899:2018. Programming Language C.: Geneva, Switzerland: International Organization for Standardization (ISO).
- ISO/IEC 2018b. ISO International Standard ISO/IEC 14882:2017. Programming Language C++. Geneva, Switzerland: International Organization for Standardization (ISO).
- ITO, Y., MIYAUCHI, A., INOUE, H., FUKUSHIMA, M., KIHARA, M., HIGASHIYAMA, T., TOMODA, C., TAKAMURA, Y., KOBAYASHI, K. & MIYA, A. 2010. An observational trial for papillary thyroid microcarcinoma in Japanese patients. *World J Surg*, 34, 28-35.
- ITO, Y., MIYAUCHI, A., KIHARA, M., HIGASHIYAMA, T., KOBAYASHI, K. & MIYA, A. 2014. Patient age is significantly related to the progression of papillary microcarcinoma of the thyroid under observation. *Thyroid*, 24, 27-34.
- JEBARA, T. 2003. *Machine Learning: Discriminative and Generative*, Springer US.
- JEGERLEHNER, S., BULLIARD, J. L., AUJESKY, D., RODONDI, N., GERMANN, S., KONZELMANN, I., CHIOLERO, A. & GROUP, N. W. 2017. Overdiagnosis and overtreatment of thyroid cancer: A population-based temporal trend study. *PLoS One*, 12, e0179387.
- JOACHIMS, T. 2002. *Learning to Classify Text Using Support Vector Machines*, Kluwer.
- KELES, A. & KELES, A. 2008. ESTDD: Expert system for thyroid diseases diagnosis. *Expert Systems with Applications*, 34, 242-246.
- KIM, E. K., PARK, C. S., CHUNG, W. Y., OH, K. K., KIM, D. I., LEE, J. T. & YOO, H. S. 2002. New sonographic criteria for recommending fine-needle aspiration biopsy of

- nonpalpable solid nodules of the thyroid. *American Journal of Roentgenology*, 178, 687-691.
- KIM, S. Y., KIM, E. K., MOON, H. J., YOON, J. H. & KWAK, J. Y. 2015. Application of Texture Analysis in the Differential Diagnosis of Benign and Malignant Thyroid Nodules: Comparison With Gray-Scale Ultrasound and Elastography. *American Journal of Roentgenology*, 205, W343-W351.
- KLASCHKA, J. & KOTRČ, E. 2004. Klasifikační a regresní lesy. In: ANTOCH, J. & DOHNAL, G. (eds.) *ROBUST*. Třešt: JČMF.
- KODAZ, H., OZSEN, S., ARSLAN, A. & GUNES, S. 2009. Medical application of information gain based artificial immune recognition system (AIRS): Diagnosis of thyroid disease. *Expert Systems with Applications*, 36, 3086-3092.
- KRIZHEVSKY A, S. I., E. HINTON G. 2012. ImageNet Classification with Deep Convolutional Neural Networks. *Neural Information Processing Systems*, 25.
- KUHN, M. 2008. Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*, 28, 1-26.
- KWAK, J. Y., HAN, K. H., YOON, J. H., MOON, H. J., SON, E. J., PARK, S. H., JUNG, H. K., CHOI, J. S., KIM, B. M. & KIM, E. K. 2011. Thyroid imaging reporting and data system for US features of nodules: a step in establishing better stratification of cancer risk. *Radiology*, 260, 892-9.
- LIAO, P.-S., CHEN, T.-S. & CHUNG, P.-C. 2001. A Fast Algorithm for Multilevel Thresholding. *Journal of Information Science and Engineering*, 17, 713-727.
- LIM, H., DEVESA, S. S., SOSA, J. A., CHECK, D. & KITAHARA, C. M. 2017. Trends in Thyroid Cancer Incidence and Mortality in the United States, 1974-2013. *Jama-Journal of the American Medical Association*, 317, 1338-1348.
- LIM, K. J., CHOI, C. S., YOON, D. Y., CHANG, S. K., KIM, K. K., HAN, H., KIM, S. S., LEE, J. & JEON, Y. H. 2008. Computer-aided diagnosis for the differentiation of malignant from benign thyroid nodules on ultrasonography. *Acad Radiol*, 15, 853-8.
- LIU, T., XIE, S., YU, J., NIU, L. & SUN, W. 2017a. Classification of thyroid nodules in ultrasound images using deep model based transfer learning and hybrid features. *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA*, pp. 919-923.
- LIU, Y., SU, L. & XIAO, H. 2017b. Review of Factors Related to the Thyroid Cancer Epidemic. *Int J Endocrinol*, 2017, 5308635.
- LUU, M. H., FISCHER, A. H., PISHARODI, L. & OWENS, C. L. 2011. Improved preoperative definitive diagnosis of papillary thyroid carcinoma in FNAs prepared with both ThinPrep and conventional smears compared with FNAs prepared with ThinPrep alone. *Cancer Cytopathol*, 119, 68-73.
- MEI, X., DONG, X., DEYER, T., ZENG, J., TRAFALIS, T. & FANG, Y. 2017. Thyroid Nodule Benignity Prediction by Deep Feature Extraction. *2017 IEEE 17th International Conference on Bioinformatics and Bioengineering (BIBE), Washington, DC*, pp. 241-245.
- MIYAUCHI, A., ITO, Y. & ODA, H. 2017. Insights into the management of papillary microcarcinoma of the thyroid. *Thyroid*.

- MOON, H. G., JUNG, E. J., PARK, S. T., HA, W. S., CHOI, S. K., HONG, S. C., LEE, Y. J., JOO, Y. T., JEONG, C. Y., CHOI, D. S. & RYOO, J. W. 2007. Role of ultrasonography in predicting malignancy in patients with thyroid nodules. *World Journal of Surgery*, 31, 1410-1416.
- MOON, W. J., BAEK, J. H., JUNG, S. L., KIM, D. W., KIM, E. K., KIM, J. Y., KWAK, J. Y., LEE, J. H., LEE, J. H., LEE, Y. H., NA, D. G., PARK, J. S., PARK, S. W., KOREAN SOCIETY OF THYROID, R. & KOREAN SOCIETY OF, R. 2011. Ultrasonography and the ultrasound-based management of thyroid nodules: consensus statement and recommendations. *Korean J Radiol*, 12, 1-14.
- MOON, W. J., JUNG, S. L., LEE, J. H., NA, D. G., BAEK, J. H., LEE, Y. H., KIM, J., KIM, H. S., BYUN, J. S., LEE, D. H., THYROID STUDY GROUP, K. S. O. N., HEAD & NECK, R. 2008. Benign and malignant thyroid nodules: US differentiation--multicenter retrospective study. *Radiology*, 247, 762-70.
- NAYAR, R. & IVANOVIC, M. 2009. The indeterminate thyroid fine-needle aspiration: experience from an academic center using terminology similar to that proposed in the 2007 National Cancer Institute Thyroid Fine Needle Aspiration State of the Science Conference. *Cancer*, 117, 195-202.
- OTSU, N. 1979. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9, 62-66.
- OZYILMAZ, L. & YILDIRIM, T. 2002. Diagnosis of thyroid disease using artificial neural network methods. *Proceeding of the 9th International Conference on Neural Information Processing*, 2033-2036.
- PAPINI, E., GUGLIELMI, R., BIANCHINI, A., CRESCENZI, A., TACCOGNA, S., NARDI, F., PANUNZI, C., RINALDI, R., TOSCANO, V. & PACELLA, C. M. 2002. Risk of malignancy in nonpalpable thyroid nodules: predictive value of ultrasound and color-Doppler features. *J Clin Endocrinol Metab*, 87, 1941-6.
- PELLERITO, J., BROMLEY, B., ALLISON, S., CHAUHAN, A., DESTOUNIS, S., DICKMAN, E., KLINE-FATH, B., MASTROBATTISTA, J., NEUMYER, M., RUNDEK, T., SAKHEL, K., SHWAYDER, J., TOI, A., WAX, J., WILKINS, I. & COMM, A. C. S. 2018. AIUM-ACR-SPR-SRU Practice Parameter for the Performance and Interpretation of a Diagnostic Ultrasound Examination of the Extracranial Head and Neck. *Journal of Ultrasound in Medicine*, 37, E6-E12.
- PETERS, J. F. 2017. *Foundations of Computer Vision: Computational Geometry, Visual Image Structures and Object Shape Detection*, Springer International Publishing.
- POLAT, K., SAHAN, S. & GUNES, S. 2007. A novel hybrid method based on artificial immune recognition system (AIRS) with fuzzy weighted pre-processing for thyroid disease diagnosis. *Expert Systems with Applications*, 32, 1141-1147.
- PROCHAZKA, A., GULATI, S., HOLINKA, S. & SMUTEK, D. 2019a. Classification of Thyroid Nodules in Ultrasound Images Using Direction-Independent Features Extracted by Two-Threshold Binary Decomposition. *Technology in Cancer Research & Treatment*, 18.
- PROCHAZKA, A., GULATI, S., HOLINKA, S. & SMUTEK, D. 2019b. Patch-based classification of thyroid nodules in ultrasound images using direction independent features extracted by two-threshold binary decomposition. *Computerized Medical Imaging and Graphics*, 71, 9-18.

- RAGHAVENDRA, U., ACHARYA, U. R., GUDIGAR, A., TAN, J. H., FUJITA, H., HAGIWARA, Y., MOLINARI, F., KONGMEBHOL, P. & NG, K. H. 2017. Fusion of spatial gray level dependency and fractal texture features for the characterization of thyroid lesions. *Ultrasonics*, 77, 110-120.
- REBALA, G., RAVI, A. & CHURIWALA, S. 2019. *An Introduction to Machine Learning*.
- REMONTI, L. R., KRAMER, C. K., LEITAO, C. B., PINTO, L. C. & GROSS, J. L. 2015. Thyroid ultrasound features and risk of carcinoma: a systematic review and meta-analysis of observational studies. *Thyroid*, 25, 538-50.
- RUSS, G., BONNEMA, S. J., ERDOGAN, M. F., DURANTE, C., NGU, R. & LEENHARDT, L. 2017. European Thyroid Association Guidelines for Ultrasound Malignancy Risk Stratification of Thyroid Nodules in Adults: The EU-TIRADS. *European Thyroid Journal*, 6, 225-237.
- RUSS, J. C. 2016. *The Image Processing Handbook*, CRC Press.
- RUSSAKOVSKY, O., DENG, J., SU, H., KRAUSE, J., SATHEESH, S., MA, S., HUANG, Z., KARPATHY, A., KHOSLA, A., BERNSTEIN, M., BERG, A. & LI, F. F. 2015. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*.
- SAVELONAS, M. A., IAKOVIDIS, D. K. & DIMITROPOULOS, N. 2007. Computational Characterization of Thyroid Tissue in the Radon Domain. *Twentieth IEEE International Symposium on Computer-Based Medical Systems, 2007. CBMS '07.*, 189-192.
- SHERMAN, S. I. 2003. Thyroid carcinoma. *Lancet*, 361, 501-11.
- SCHROEDER, M. 2009. *Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise*. Dover Publications.
- SMITH-BINDMAN, R., LEBDA, P., FELDSTEIN, V. A., SELLAMI, D., GOLDSTEIN, R. B., BRASIC, N., JIN, C. & KORNAK, J. 2013. Risk of thyroid cancer based on thyroid ultrasound imaging characteristics: results of a population-based study. *JAMA Intern Med*, 173, 1788-96.
- SUGITANI, I., TODA, K., YAMADA, K., YAMAMOTO, N., IKENAGA, M. & FUJIMOTO, Y. 2010. Three distinctly different kinds of papillary thyroid microcarcinoma should be recognized: our treatment strategies and outcomes. *World J Surg*, 34, 1222-31.
- SUZUKI, K. 2017. Overview of deep learning in medical imaging. *Radiol Phys Technol*, 10, 257-273.
- TAN, G. H. & GHARIB, H. 1997. Thyroid incidentalomas: management approaches to nonpalpable nodules discovered incidentally on thyroid imaging. *Ann Intern Med*, 126, 226-31.
- THEOHARIS, C. G., SCHOFIELD, K. M., HAMMERS, L., UDELSMAN, R. & CHHIENG, D. C. 2009. The Bethesda thyroid fine-needle aspiration classification system: year 1 at an academic institution. *Thyroid*, 19, 1215-23.
- TRIMBOLI, P., GUGLIELMI, R., MONTI, S., MISISCHI, I., GRAZIANO, F., NASROLLAH, N., AMENDOLA, S., MORGANTE, S. N., DEIANA, M. G., VALABREGA, S., TOSCANO, V. & PAPINI, E. 2012. Ultrasound Sensitivity for Thyroid Malignancy Is Increased by

- Real-Time Elastography: A Prospective Multicenter Study. *Journal of Clinical Endocrinology & Metabolism*, 97, 4524-4530.
- TSANTIS, S., DIMITROPOULOS, N., CAUOURAS, D. & NIKIFORIDIS, G. 2006. A hybrid multi-scale model for thyroid nodule boundary detection on ultrasound images. *Computer Methods and Programs in Biomedicine*, 84, 86-98.
- TSANTIS, S., DIMITROPOULOS, N., CAVOURAS, D. & NIKIFORIDIS, G. 2009. Morphological and wavelet features towards sonographic thyroid nodules evaluation. *Computerized Medical Imaging and Graphics*, 33, 91-99.
- TUNBRIDGE, W. M., EVERED, D. C., HALL, R., APPLETON, D., BREWIS, M., CLARK, F., EVANS, J. G., YOUNG, E., BIRD, T. & SMITH, P. A. 1977. The spectrum of thyroid disease in a community: the Whickham survey. *Clin Endocrinol (Oxf)*, 7, 481-93.
- UMBAUGH, S. E. 2017. *Digital Image Processing and Analysis: Applications with MATLAB and CVIPtools*, CRC Press.
- VANDER, J. B., GASTON, E. A. & DAWBER, T. R. 1968. The significance of nontoxic thyroid nodules. Final report of a 15-year study of the incidence of thyroid malignancy. *Ann Intern Med*, 69, 537-40.
- YIP, L. & SOSA, J. A. 2016. Molecular-Directed Treatment of Differentiated Thyroid Cancer: Advances in Diagnosis and Treatment. *JAMA Surg*, 151, 663-70.

VLASTNÍ PUBLIKAČNÍ AKTIVITA

A) Publikace týkající se tématu dizertační práce

A-1) Články s impact factorem

PROCHÁZKA, Antonin, Sumeet GULATI, Stepan HOLINKA a Daniel SMUTEK. Patch-based classification of thyroid nodules in ultrasound images using direction independent features extracted by two-threshold binary decomposition. *Computerized Medical Imaging and Graphics*. 2019, 71, 9-18. DOI:10.1016/j.compmedimag.2018.10.001. ISSN 08956111. IF: 3,30.

PROCHÁZKA, Antonin, Sumeet GULATI, Stepan HOLINKA a Daniel SMUTEK. Classification of Thyroid Nodules in Ultrasound Images Using Direction-Independent Features Extracted by Two-Threshold Binary Decomposition. *Technology in Cancer Research & Treatment*. 2019. DOI: 10.1177/1533033819830748. ISSN 1533-0346. IF: 1,48.

A-2) Aktivní účast na konferencích se sborníkem abstrakt

PROCHÁZKA, Antonín a Daniel SMUTEK. Texturní analýza uzlů ve štítné žláze pomocí binární stack-dekompozice a patch-based analýzy. In: *XLII. Dny lékařské biofyziky: Sborník abstrakt*. Praha: Ústav biofyziky 2. lékařské fakulty Univerzity Karlovy ve spolupráci s Českou společností lékařské fyziky České lékařské společnosti J. Ev. Purkyně, 2019, s. 36. *Příspěvek byl v rámci konference oceněn 3. místem v soutěži doktorandů.*

PROCHÁZKA, Antonín a Daniel SMUTEK. Počítačová diagnostika uzlů ve štítné žláze z ultrazvukového obrazu pomocí klasifikační metody random forests. In: *XLII. Dny lékařské biofyziky: Sborník abstrakt*. Praha: Ústav lékařské biofyziky a lékařské informatiky 3. lékařské fakulty Univerzity Karlovy ve spolupráci s Českou společností lékařské fyziky České lékařské společnosti J. Ev. Purkyně, 2017, s. 46. *Příspěvek byl v rámci konference oceněn 1. místem v soutěži doktorandů.*

PROCHÁZKA, Antonín, Daniel SMUTEK, Štěpán HOLINKA. Užití matematické morfologie při segmentaci ultrazvukového obrazu. In: *XXXVI. Dny lékařské biofyziky: Sborník abstrakt*. Hradec Králové: Univerzita Karlova v Praze, Lékařská fakulta v Hradci Králové, Ústav lékařské biofyziky ve spolupráci s Českou společností lékařské fyziky České lékařské společnosti J. Ev. Purkyně, 2013, s. 64.

B) Ostatní publikace

B-1) Články s impact factorem

WEISZENSTEIN, Martin, Martina MUSUTOVA, Andrea PLIHALOVA, Katerina WESTLAKE, Moustafa ELKALAF, Michal KOC, Antonin PROCHAZKA, Jan PALA, Sumeet GULATI, Jan TRNKA, Jan POLAK. Adipogenesis, lipogenesis and lipolysis is stimulated by mild but not severe hypoxia in 3T3-L1 cells. *Biochemical and Biophysical Research Communications*. 2016, 478(2), 727-732. DOI: 10.1016/j.bbrc.2016.08.015. ISSN 0006291X. IF: 2,47.

PROCHAZKA, A., J. DAMMER, F. WEYDA, V. SOPKO, J. BENES, J. ZEMAN a I. JANDEJSEK. Biological object recognition in μ -radiography images. *Journal of Instrumentation*. 2015, 10(03). DOI: 10.1088/1748-0221/10/03/C03023. ISSN 1748-0221. IF: 1,31.

PROCHÁZKOVÁ, Marie, Jaroslav TINTĚRA, Antonín PROCHÁZKA, Patrícia MARTINKOVÁ a Kamila ŘASOVÁ. Changes of Effective Connectivity after Facilitation Physiotherapy in Multiple Sclerosis. *Česká a slovenská neurologie a neurochirurgie*. 2015, 78/111(4), 423-429. DOI: 10.14735/amcsnn2015423. ISSN 12107859. IF: 0,21.

PROCHAZKA, Antonin a David KORPAS. Monitoring biventricular pacing parameters depending on the left ventricle lead configuration. *Experimental and Clinical Cardiology*. Pulsus Group. 2013, 18(2), 85-88. IF: 0,76.

B-2) Aktivní účast na konferencích se sborníkem abstrakt

PROCHÁZKA, Antonín, Sumeet GULATI, Jan POLÁK. Určení počtu a velikosti lipidových kapiček v mikroskopickém obraze silně a mírně hypoxických adipocytů. In: XXXIX. *Dni lékařské biofyziky: Sborník abstrakt*. Bratislava: Ústav lékařské fyziky, biofyziky, informatiky a telemedicíny Lékařské fakulty Univerzity Komenského v Bratislave, 2016, s. 59.

PROCHÁZKOVÁ, Marie, Kamila ŘASOVÁ, Jaroslav TINTĚRA, Patrícia MARTÍNKOVÁ, Antonín PROCHÁZKA. Monitoring the effect of facilitation physiotherapy in multiple sclerosis – using functional magnetic resonance imaging. In: *RIMS 2014 – Abstracts*. Multiple Sclerosis Journal. 2014, 20(7): S1003-1004.

PROCHÁZKA, Antonín. Vliv konfigurace biventrikulárních kardiostimulátorů na vývoj parametrů kardiostimulace. In: XXXV. *Dny lékařské biofyziky: Sborník abstrakt*. Olomouc: Ústav lékařské biofyziky Lékařské fakulty Univerzity palackého v Olomouci ve spolupráci Českou společností lékařské fyziky České lékařské společnosti J. Ev. Purkyně, 2012, s 63.

PŘÍLOHY

Příloha 1: R skript – patch-based klasifikační modely

```
##### Data loading and preprocessing part #####

#load data
Data_raw = read.csv("17_17_px_patches.csv", header = TRUE, sep =
";" );

#dealing with class imbalance - under-sampling the majority class
library(unbalanced)
set.seed(91906)
n<-ncol(Data_raw)
output<-Data_raw[ ,n]
input<-Data_raw[ ,-n]
Data_raw2 <-ubUnder(X=input, Y= output, perc = 50, method =
"percPos", w = NULL)
newData<-cbind(Data_raw2$X, Class=Data_raw2$Y)
Data_raw <- newData
detach(package:unbalanced, unload=TRUE)

#loading caret and preprocessing of data for caret,
#e.g. dealing with NA values and assigning text labels to class
#instead of 0 and 1
library(caret);
Data_raw <- Data_raw[complete.cases(Data_raw),]
Data_raw <- Data_raw[, colSums(is.na(Data_raw)) == 0]
Data_raw <- Data_raw[, colSums(Data_raw != 0) > 0]
Data_raw$class[Data_raw$class == 0] <- 'nben'
Data_raw$class[Data_raw$class == 1] <- 'mal'
Data_raw$class <- factor(Data_raw$class)

# Creating of 10 folds grouped by Image (for group 10 fold cross-
# validation)
group_cv <- function(x, k = length(unique(x))) {
  dat <- data.frame(index = seq(along = x), group = x)
  groups <- data.frame(group = unique(dat$group))
  group_folds <- createFolds(groups$group, returnTrain =
TRUE, k = k)
  group_folds <- lapply(group_folds, function(x, y)
y[x,,drop = FALSE], y = groups)
```

```

        dat_folds <- lapply(group_folds, function(x, y)
        merge(x, y), y = dat)
        lapply(dat_folds, function(x) sort(x$index))
    }
groups <- Data_raw$image
folds_1 <- group_cv(groups, k =10)
for_mod <- lapply(folds_1, function(x, y) y[x], y = groups)
holdout <- lapply(folds_1, function(x, y) y[-unique(x)], y =
    groups)
for(i in seq(along = folds_1)) {
    if(any(unique(for_mod[[i]]) %in% unique(holdout[[i]])))
        cat("Not working correctly")
}

#setting control parameters and group folds (variable folds_1) and
#dropping identifying columns
args(trainControl)
ctrl <-trainControl(method = "CV", index = folds_1, returnResamp
    = "all", savePredictions = TRUE, summaryFunction =
    twoClassSummary, classProbs = TRUE)
drops <- c("Nodule", "Image")
DataSet <- Data_raw[ , !(names(Data_raw) %in% drops)]

##### End of Data loading and preprocessing part #####

##### Random Forest part #####

#Recursive feature elimination method (RFE)
control <- rfeControl(functions=caretFuncs, method="cv",index =
    folds_1) # Setting control parameters for RFE

# run the RFE algorithm
features <- rfe(DataSet[,1:104], DataSet[,105], sizes=c(2: 104),
    rfeControl=control, method = "rf", ntree = 500,
    trControl =ctrl,
    preProc = c("center", "scale"),
    metric = "ROC")

# printing results of RFE
print(features)
predictors(features) # list ranked predictors
plot(features, type=c("g", "o")) # plot the features accuracy

# Training different models using different number of
# predictors (from 2 to 104 predictors)
# ROC AUC results are saved in the variable named values
values <- data.frame()
for (i in c(2:104)){

```

```

columnsToKeep <- predictors(features)[1:i]
DataSet_reduced <- Data_raw[ , (names(Data_raw) %in%
columnsToKeep)]
DataSet_reduced$class <- DataSet[,105]
tunegrid <- expand.grid(.mtry=c(2))

      rffit <- train(Class ~ ., data = DataSet_reduced,
                    method = "rf",
                    ntree = 500,
                    trControl = ctrl,
                    preProc = c("center", "scale"),
                    tuneGrid=tunegrid,
                    metric = "ROC",
                    prox=TRUE)
values[i,1]<- rffit$features$ROC
values[i,2]<- rffit$features$ROCS
}

# plot the graph showing how ROC AUC depend on number of features
plot(values$V1)

# Based on the previous graph, keep only first 14 predictors
# (other predictors, when added in the model, do not increase
# overall value of ROC AUC)
columnsToKeep <- predictors(features)[1:14]
DataSet_reduced <- DataSet[ , (names(DataSet) %in% columnsToKeep)]
DataSet_reduced$class <- DataSet[,105]

# Building the final model using only the best 14 predictors
tunegrid <- expand.grid(.mtry=c(2))
rffit <- train(Class~., data=DataSet_reduced,
              method="rf",
              ntree = 500,
              preProcess=c("center", "scale", "zv"),
              tuneGrid=tunegrid,
              trControl = ctrl,
              metric = "ROC",
              prox=TRUE)

rffit

#Outputting the results of cross-validation
data_train_resample <-rffit$resample
data_train_resample

#Outputting the results for whole nodules - number of correctly
#classified and misclassified
predictions_CV <-rffit$pred
results <-DataSet
results$Nodule <- Data_raw$Nodule

```

```

counts <- setNames(data.frame(table(results$Nodule)),
                  c("Category", "count"))
predictions_CV_best <- predictions_CV[
  which(predictions_CV$mtry==2),]
results$predRight <-
predictions_CV_best$obs==predictions_CV_best$pred

# final results showing number of correctly and wrongly classified
# patches in each nodule
count_of_corrclass_and_missclass_nodules <- with(results,
  tapply(predRight, list(Nodule, predRight), FUN =
    function(x) table(x)))
count_of_corrclass_and_missclass_nodules

##### End of Random Forests part #####

##### Support Vector Machines part #####

#Recursive feature elimination method (RFE)
control <- rfeControl(functions=caretFuncs, method="cv", index =
  folds_1)

# run the RFE algorithm
features <- rfe(DataSet[,1:104], DataSet[,105], sizes=c(2:104),
  rfeControl=control,
  method = "svmRadial",
  trControl =ctrl,
  preProc = c("center", "scale"),
  metric = "ROC")

# print the results of RFE
print(features)
predictors(features) # list of ranked predictors
plot(features, type=c("g", "o")) # plot features accuracy

# Training the different models using different number of
# predictors (from 2 to 104 predictors)
# ROC AUC results are saved in the variable named values
values <- data.frame()
for (i in c(2:104)){
  columnsToKeep <- predictors(features)[1:i]
  DataSet_reduced <- DataSet[ , (names(DataSet) %in%
  columnsToKeep)]
  DataSet_reduced$Class <- DataSet[,105]
  svmFit <- train(Class ~ ., data = DataSet_reduced,
    method = "svmRadial",
    trControl =ctrl,

```



```

preProc = c("center", "scale"),
tuneLength = 1,
metric = "ROC")

values[i-1,1]<- svmFit$features$ROC
values[i-1,2]<- svmFit$features$ROCS
assign(paste("run", i, sep=""), svmFit)
}

# plot the graph showing how ROC AUC depend on number of features
plot(values$V1)
plot( seq(1, length(values$V1), 6), values$V1[seq(1,
length(values$V1),6 )], type='l' ) #smoother graph

# Based on the previous graph, keep only first 21 predictors
#(other predictors, when added in the model, do not increase
# overall value of ROC AUC)
columnsToKeep <- predictors(features)[1:21]
DataSet_reduced <- DataSet[ , (names(DataSet) %in% columnsToKeep)]
DataSet_reduced$Class <- DataSet[,105]

# Building the final model using only the best 21 predictors
svmFit <- train(Class ~ ., data = DataSet_reduced,
method = "svmRadial",
trControl =ctrl,
preProc = c("center", "scale"),
tuneLength = 5,
metric = "ROC")

# showing main results
svmFit

#Outputting the results of cross-validation
data_train_resample <-svmFit$resample
data_train_resample

#Outputting the results for whole nodules
predictions_CV <-svmFit$pred
results <-DataSet
results$Nodule <- Data_raw$Nodule
counts <- setNames(data.frame(table(results$Nodule)),
c("Category", "count"))
predictions_CV_best <- predictions_CV[
which(predictions_CV$C==0.25),]
results$predRight <-
predictions_CV_best$obs==predictions_CV_best$pred

```

```
# final results showing number of correctly and wrongly classified
# patches in each nodule
count_of_corrclass_and_missclass <- with(results,
      tapply(predRight, list(Nodule, predRight), FUN =
            function(x) table(x)))
count_of_corrclass_and_missclass

##### End of Support Vector Machines part #####
```

Příloha 2: R skript – klasifikační model pro celé uzly

```
##### Data loading and preprocessing part #####
#load data
Dataset = read.csv("vysledky_6_prahu.csv", header = TRUE, sep =
";" );
#load ggplot2
library(ggplot2);

# Loading caret and preprocessing of data for caret,
# e.g. dealing with NA values and assigning text labels to class
# instead of 0 and 1
library(caret);
Dataset$class[Dataset$class == 0] <- 'nben'
Dataset$class[Dataset$class == 1] <- 'mal'
Dataset$class <- factor(Dataset$class)
Dataset <- Dataset[, colSums(is.na(Dataset)) == 0]

# Setting control parameters for training and dropping identifying
# column
ctrl <- trainControl(method = "repeatedcv", number= 39, classProbs
= TRUE)
tuneGrid <- expand.grid(.mtry=c(2:31))
drops <- c("Image")
Dataset <- Dataset[ , !(names(Dataset) %in% drops)]
dim(Dataset)
summary(Dataset)
##### End of Data loading and preprocessing part #####

#Training of Random Forests
rfFit <- train(Class~., data=Dataset,
method="rf",
ntree = 2000,
preProcess=c("center", "scale", "zv"),
tuneGrid=tuneGrid,
trControl =ctrl,
metric = "Accuracy",
prox=TRUE)

#Evaluation
rfFit$results
rfFit$finalModel

#Variables importance
varImportance <- varImp(rfFit)
plot(varImportance)
```

```
#Fitting of SVM model
svmFit <- train(Class ~ ., data = Dataset,
               method = "svmRadial",
               trControl =ctrl,
               preProc = c("center", "scale"),
               tuneLength = 20,
               metric = "Accuracy")

#Evaluation of SVM model
svmFit
svmFit$results
svmFit$finalModel
```