

Univerzita Hradec Králové
Fakulta informatiky a managementu
Katedra informačních technologií

Hluboké učení v radiologii
se zaměřením na rentgenové snímky
rukou revmatologických pacientů

Diplomová práce

Autor: Adam Ostruszka
Studijní obor: Aplikovaná informatika
Vedoucí práce: Ing. Karel Mls, Ph.D.
Odborný konzultant: MUDr. Ondřej Živný, Radiologická klinika FNHK

Prohlašuji, že jsem diplomovou práci zpracoval samostatně a s použitím uvedených informačních zdrojů.

V Hradci Králové dne 30. 4. 2020

Adam Ostruszka

Děkuji vedoucímu diplomové práce, Ing. Karlu Mlsovi, Ph.D., za metodické vedení, cenné rady, připomínky, lidský přístup a především vůbec za zprostředkování možnosti pracovat na zajímavém a smysluplném tématu.

Můj velký dík patří MUDr. Ondřeji Živnému za obrovské nadšení pro věc, péli, nasazení a desítky hodin času věnovaného přípravě, pečlivému třídění a anonymizaci dat, sdílení poznatků a pravidelným konzultacím.

Dále děkuji Centru základního a aplikovaného výzkumu za možnost využití výkonného hardwaru financovaného z projektů prof. Ing. Ondřeje Krejčara, Ph.D., a prof. Ing. Kamila Kuči, Ph.D., a za organizační zajištění jeho využití děkuji Ing. Michalovi Dobrovolnému.

Na závěr srdečně děkuji také svým rodičům za pomoc a podporu během studia a dále kamarádům a spolužákům za jejich podněty a doporučení.

Anotace

Tato práce zkoumá současný stav a možnosti využití hlubokého učení v diagnostické radiologii, konkrétně klasifikaci dvourozměrných snímků pomocí konvolučních neuronových sítí. Práce jednak přináší shrnutí aktuálního stavu vývoje v této oblasti ve vědě i klinické praxi a jednak ověřuje možnosti nasazení světově nejběžnějších technologií na reálném a smysluplném příkladu klasifikace onemocnění na rentgenových snímcích rukou revmatologických pacientů. Vzhledem k omezené velikosti dostupného datového souboru byly vyzkoušeny metody přeneseného učení (transfer learning) a rozšíření dat (data augmentation). Zkoumaný datový soubor poskytla Fakultní nemocnice Hradec Králové a skládá se ze tří skupin snímků rukou trpících revmatickými onemocněními a rukou bez onemocnění. Řešit bylo nutné jednak velmi důkladnou analýzu a předzpracování vstupních dat a jednak samotnou klasifikaci. Nejvíce se osvědčily neuronové sítě architektur ResNet a Inception-V3. Nejúspěšnější experimenty dosáhly v klíčových metrikách hodnot okolo 90 % a taktéž další dosažené výsledky ukazují, že budoucí vývoj nastíněným směrem má smysl.

Klíčová slova

počítačové vidění, hluboké učení, přenesené učení, neuronové sítě, TensorFlow, radiologie, RTG, revmatologie, revmatoidní artritida, psoriatická artritida, osteoartróza

Annotation

Title: Deep Learning in Radiology

This thesis examines current state and possibilities of using deep learning in diagnostic radiology, specifically image classification using convolutional neural networks. The work brings a summary of the current state of development in this field in science and clinical practice and also verifies the possibility of deploying the world's most common technologies on a real and meaningful example of disease classification of X-ray images of the hands of rheumatologic patients. Due to the limited size of the available dataset, the methods of transfer learning and data augmentation were tested. The investigated dataset was provided by the Hradec Králové University Hospital (Fakultní nemocnice Hradec Králové) and contains three groups of radiographs of hands suffering from rheumatic diseases and disease-free hands. It was necessary to solve very thorough analysis and preprocessing of input data and also the classification task itself. The ResNet and Inception-V3 neural networks proved to be the most effective for solving the task. The most successful experiments reached around 90 % in all key metrics and other achieved results also show that further development in the outlined direction is promising.

Keywords

computer vision, deep learning, transfer learning, neural networks, TensorFlow, radiology, X-ray, rheumatology, rheumatoid arthritis, psoriatic arthritis, osteoarthritis

Obsah

1	Úvod	1
2	Vymezení tématu a cíle práce	3
3	Medicínské zobrazování a charakteristiky onemocnění	4
3.1	Radiologie (medicínské zobrazovací metody)	4
3.1.1	Přehled zobrazovacích metod	4
3.2	RTG vyšetření (skiografie).....	5
3.2.1	Metody snímání	6
3.2.2	Kvalita a vlastnosti RTG obrazu	7
3.2.3	Postup vyhodnocení RTG snímku radiologem.....	11
3.3	RTG ruky.....	12
3.4	Revmatická onemocnění a jejich projevy na RTG snímcích	13
3.4.1	Revmatoidní artritida.....	15
3.4.2	Psoriatická artritida.....	17
3.4.3	Osteoartróza.....	19
4	Umělá inteligence, strojové a hluboké učení.....	21
4.1	Umělá inteligence	21
4.2	Motivace – proč se zabývat umělou inteligencí v radiologii?.....	22
4.3	Strojové učení	23
4.3.1	Učení s učitelem.....	24
4.3.2	Učení bez učitele.....	24
4.4	Umělé neuronové sítě a hluboké učení	25
4.4.1	Formální neuron a neuronová síť	25
4.4.2	Matematické principy fungování neuronové sítě.....	27
4.4.3	Konvoluční neuronové sítě	29
4.4.4	Pracovní postup.....	31
4.4.5	Vyhodnocení modelu.....	35
5	Implementace neuronových sítí a související problémy	43
5.1	Architektury neuronových sítí.....	43
5.1.1	VGG16	43

5.1.2	Inception.....	43
5.1.3	ResNet.....	44
5.1.4	Inception-ResNet	44
5.1.5	MobileNet.....	44
5.2	Programovací jazyky a frameworky.....	44
5.2.1	TensorFlow.....	44
5.3	Rychlost vývoje technologií	45
5.4	Hardware	46
5.4.1	Použití GPU, CPU a TPU	46
5.4.2	Problém omezené paměti GPU.....	46
5.5	Přenesené učení.....	48
5.6	Problém nízkého rozlišení vstupních snímků	51
5.7	Problém čtvercového formátu vstupních snímků	52
5.8	Rozšíření dat	53
5.8.1	Bezpečnost a způsoby provedení.....	53
5.8.2	Přehled druhů transformací.....	55
5.8.3	Způsoby implementace rozšíření dat.....	57
6	Hluboké učení ve vědě a klinické praxi	58
6.1	Hluboké učení ve vědě	58
6.1.1	Doporučení pro výzkum a publikování.....	59
6.1.2	Medicínské datasety	61
6.1.3	Výzkumy související s tématem této práce	63
6.1.4	Detekce kloubů na ruce	66
6.1.5	RSNA Bone Age Challenge	68
6.2	Hluboké učení v praxi.....	70
6.2.1	Visana – BoneXpert.....	70
6.2.2	Vuno – Med Bone Age a další	70
6.2.3	Siemens – AI-Rad Companion, AI-Pathway Companion a Syngo.....	71
6.2.4	Další software – Brainomix a Quantib Brain	71
7	Vlastní řešení.....	72

7.1	Návrh studie – vytyčení cílů, předpoklady a očekávání	72
7.2	Analýza dat, jejich předzpracování a tvorba datasetů	74
7.2.1	Data poskytnutá FNHK	74
7.2.2	Převod z formátu DICOM	74
7.2.3	Analýza a předzpracování snímků	76
7.2.4	Dataset snímků celých rukou	78
7.2.5	Ověřovací skupina snímků rukou z externích zdrojů	79
7.2.6	Dataset snímků kloubů	80
7.2.7	Dataset snímků kloubů tříděný podle typu změn	81
7.3	Trénování modelů a jejich vyhodnocení	82
7.3.1	Použitý hardware	82
7.3.2	Použitý software	82
7.3.3	Rozšíření dat	83
7.3.4	Vytrénování modelu od základů	83
7.3.5	Přenesené učení formou extrakce rysů	84
7.3.6	Přenesené učení formou ladění	86
8	Shrnutí výsledků	97
8.1	Aktuální situace využití hlubokého učení v radiologii	97
8.2	Trénování klasifikačních modelů	98
8.3	Omezení současných technologií a další překážky	99
9	Závěr a otázky pro další výzkum	101
10	Seznam použitých zdrojů	104
11	Přílohy	111

Seznam obrázků a grafů

Obr. 1: Základní odstíny tkání pozorovatelné na RTG snímku.....	8
Obr. 2: Vliv šumu na kontrast zobrazení	9
Obr. 3: Ostrost stínu v závislosti na velikosti zdroje záření	9
Obr. 4: Vlevo úprava kontrastu a ostrosti (zřetelnosti), vpravo oprava podexponovaných míst.	10
Obr. 5: Technologie Bone Suppression (potlačení kostí) na snímku plic	10
Obr. 6: Způsoby projekce ruky	12
Obr. 7: Označení kloubů a číslování prstů ruky.....	13
Obr. 8: Deformity kloubů	14
Obr. 9: Klouby typicky postižené RA.....	15
Obr. 10: Ruce postižené RA.....	16
Obr. 11: Deformity prstů typické pro RA	16
Obr. 12: Klouby typicky postižené PsA.....	17
Obr. 13: Ruce postižené PsA.....	18
Obr. 14: Jiný příklad ruky postižené PsA	18
Obr. 15: Klouby typicky postižené OA.....	19
Obr. 16: Ruce postižené OA.....	20
Obr. 17: Vztah umělé inteligence, strojového a hlubokého učení.....	21
Obr. 18: Varianty klasifikační úlohy.....	24
Obr. 19: Obecné schéma formálního neuronu	25
Obr. 20: Obecné schéma neuronové sítě	25
Obr. 21: Tvorba map rysů pomocí konvoluce	29
Obr. 22: Posloupnost prací při vývoji řešení využívajícího strojové či hluboké učení.....	31
Obr. 23: Rozdělení datasetu a účel jednotlivých částí.....	33
Obr. 24: Možné varianty průběhu trénování modelu	35
Obr. 25: Hluboké učení jako černá skříňka	39
Obr. 26: Reprezentace rysů obrázku uvnitř hluboké konvoluční neuronové sítě	40
Obr. 27: Teplotní mapy jako součást úlohy určování kostního věku	41

Obr. 28: Schematické znázornění řešení na analýzu 3D skenu sítnice od společnosti DeepMind	41
Obr. 29: Oklamání vytrénovaných modelů pomocí protichůdných příkladů	42
Obr. 30: Využití paměti GPU při trénování modelu	47
Obr. 31: Strategie přeneseného učení	49
Obr. 32: Jak vybrat vhodnou strategii přeneseného učení	50
Obr. 33: Vliv rozlišení snímku na přesnost klasifikace diagnóz na snímcích hrudníku	51
Obr. 34: Nežádoucí podobnost obrazu vzniklá příliš velkou mírou deformace	53
Obr. 35: Druhy transformací	56
Obr. 36: Vývoj počtu publikací týkajících se radiologie a umělé inteligence	58
Obr. 37: Ukázka funkčnosti programu BoneFinder	67
Obr. 38: Způsob předzpracování datasetu v podání soutěžního týmu Michiganské univerzity....	68
Obr. 39: Rozhraní programu BoneXpert	70
Obr. 40: Chyby při převádění programem Dicom Converter (vpravo původní snímek)	75
Obr. 41: Program na označování a tvorbu výřezů kloubů	80

Seznam tabulek

Tabulka 1: Přehled druhů RTG záření	5
Tabulka 2: Přehled revmatických onemocnění řešených v této práci	14
Tabulka 3: Výzkumy úspěšně využívající přenesené učení na medicínských datech	63
Tabulka 4: Nástroje vyzkoušené k převodu z formátu DICOM do PNG	75
Tabulka 5: Počty snímků po základním předzpracování	77
Tabulka 6: Počty snímků v datasetu rukou	78
Tabulka 7: Ověřovací skupina snímků rukou z externích zdrojů	79
Tabulka 8: Dataset snímků kloubů	81
Tabulka 9: Dataset snímků kloubů tříděný podle typu změn	81
Tabulka 10: Parametry použitého hardwaru	82
Tabulka 11: Trénování na datasetech rukou a kloubů od základů	84
Tabulka 12: Přenesené učení formou extrakce rysů na datasetu rukou	85
Tabulka 13: Přenesené učení formou extrakce rysů na datasetu kloubů	86
Tabulka 14: Metriky pro model 1 na klasifikaci snímků ruky	88
Tabulka 15: Ukázka výstupů modelu 1 na vybraných snímcích z externích zdrojů	89
Tabulka 16: Metriky pro model 2 na klasifikaci snímků ruky	90
Tabulka 17: Ukázka výstupů modelu 2 na vybraných snímcích z externích zdrojů	91
Tabulka 18: Metriky pro model 3 na klasifikaci snímků ruky	92
Tabulka 19: Metriky pro model na rozlišení artritických a artrotických projevů	93
Tabulka 20: Metriky pro model na klasifikaci snímků kloubů	94
Tabulka 21: Metriky pro model 1 na rozlišení artritických a artrotických projevů	95
Tabulka 22: Metriky pro model 2 na rozlišení artritických a artrotických projevů	96

Poznámka: všechny tabulky jsou vlastním dílem autora práce.

1 Úvod

V uplynulém desetiletí došlo zejména díky nárůstu výpočetního výkonu počítačového hardwaru k významným pokrokům v hlubokém učení, což je jedna z mnoha podoblastí umělé inteligence. Internetové vyhledávače, systémy pro reklamu a doporučování obsahu na webu, emailové spam filtry, software ovládající samořiditelná auta, rozpoznávání tváří na fotografiích a záznamech z bezpečnostních kamer, rozpoznávání textu a řeči, překladače jazyků nebo chatboti... to vše dnes ke zpracování dat využívá výpočetní modely nazývané umělé neuronové sítě, potažmo hluboké učení.

Ruku v ruce s rozvojem hardwaru a algoritmů vedly pokroky v medicínských zobrazovacích technikách a postupná digitalizace této oblasti k vytvoření potenciálu pro využití hlubokého učení na různé úlohy v diagnostické radiologii. Navíc současná doba je charakteristická stárnutím populace, růstem nákladů na zdravotní péči a nedostatkem zdravotníků. Podle jednoho ze scénářů v budoucnu lidská civilizace nemusí mít dostatek financí a odborníků, aby bylo možné nadále poskytovat zdravotní péči v tradiční podobě, jakou známe dnes, pro stále početnější a starší populaci. Jedním z mnoha způsobů pomoci mohou být právě inovativní zdravotnické přístroje a software s prvky umělé inteligence.

S ohledem na technologickou, etickou, morální a právní stránku věci při současném stavu poznání nehrozí, že by radiologické odborníky nahradil počítač – namísto toho se očekává, že hluboké učení bude nástroj pomáhající lékařům lépe zvládnout narůstající počet vyšetření, vytěžit z dostupných dat maximum informací, zvýšit spolehlivost diagnostiky, ušetřit čas a zefektivnit léčbu pacientů. Předpokládá se, že následující generace radiologů již kromě medicínského vzdělání budou potřebovat i znalosti ze statistiky a informačních technologií a budou se setkávat s metodami hlubokého učení v každodenní praxi.

Již nyní je k dispozici hrstka prvních certifikovaných komerčních aplikací, které jsou dílem předních světových výrobců zdravotnické techniky či inovativních startupů, a které se o výše zmíněné přínosy pokoušejí. Vědecký výzkum se postupně tímto směrem také obrací, i když v Česku zatím velmi zvolna. Množství publikací o hlubokém učení ve zdravotnictví v posledních letech strmě roste. Open source povaha softwaru pro implementaci neuronových sítí a rostoucí výkon počítačového hardwaru budou nadcházející změny dále stimulovat. Protože jde o poměrně mladý obor na vzestupu, vše se v něm poměrně bouřlivě vyvíjí (chvílemi možná jen nekontrolovaně bují) a na rozdíl od tradičnějších oborů lidského bádání zde zatím chybí pevná půda pod nohama. Zprávy o nových technologiích, technikách a vylepšeních přicházejí téměř denně, a přestože jde většinou jen o další kapky vody v moři, několikrát do roka se objeví také „game changer“, nově zveřejněný velký projekt, který posouvá celý obor o další krůček vpřed a mění zaběhnutá pravidla. Jen během roku tvorby této práce se objevily asi tři nebo čtyři.

Aby bylo hluboké učení v radiologii dostatečně spolehlivé pro nasazení v praxi, jsou zapotřebí datové soubory natolik početné a tak dobře anotované, že dostatečně pokryjí i onemocnění méně častá, vzácná či postrádající charakteristické projevy. A zároveň soubory dostatečně rozmanité, co se týče možných zdrojů a podob snímků. Proto právě data představují velké úskalí. Volně dostupných souborů není mnoho, navíc se zaměřují jen na určité druhy snímků a diagnóz. Výzkumné týmy pak často sestavují datové soubory z vlastních zdrojů. Avšak ty často trpí nedostatečným počtem snímků nutným ke statisticky spolehlivému vytrénování neuronové sítě. Způsobem, jak se s tímto problémem vypořádat, může být například metoda přeneseného učení (transfer learning), kdy se použije již částečně předtrénovaná neuronová síť.

V roce 2019 spojily síly dvě sousedící instituce, Fakulta informatiky a managementu (FIM) UHK a Fakultní nemocnice Hradec Králové (FNHK), aby spolupracovaly na zpracování radiologických dat. Kromě ryze vědeckých záměrů má výzkum potenciál zkvalitnit péči o pacienty. Jednou z několika řešených úloh je analýza rentgenových snímků rukou revmatologických pacientů – rozpoznání, jestli jsou příznaky spíše původu zánětlivého (artritické) nebo degenerativního (artrotické) a v ideálním případě také diagnostika onemocnění v časnější fázi. Počítačová podpora této úlohy by v ideálním případě mohla zaručit, že radiologové nic nepřehlédnou, poskytnout jim oporu či alternativní názor včetně zdůvodnění, a ušetřit čas potřebný na náročnější úkoly, kde je nasazení hlubokého učení prozatím neproveditelné.

Práce je psána se záměrem obeznámit čtenáře s řešeným tématem pokud možno srozumitelně – tak, aby lékařům dokázala dát patřičný vhled do IT problematiky, a naopak datovým vědcům nebo programátorům přiblížila medicínskou stránku věci. Každé dílčí téma či oblast popsaná na následujících stranách by vydaly na samostatnou diplomovou práci, a proto se tato práce spíše než vyčerpávající míru detailů se snaží nabídnout ucelený a komplexní pohled, samozřejmě na odpovídající úrovni odbornosti. A zároveň také pohled střízlivý – o přínosech nových technologií už bylo řečeno mnoho a nepotřebují zveličovat, namísto toho je vhodné otevřeně mluvit také o omezeních a nalezených nedostatcích, protože právě to jsou největší překážky pro budoucí rozvoj a nasazení technologie.

Použitá literatura čerpá z fondů Univerzitní knihovny UHK, Lékařské knihovny Univerzity Karlovy, odborných časopisů *Radiology*, *Radiology: Artificial Intelligence*, *European Radiology*, *European Journal of Radiology* a dalších vědeckých publikací. Poznatky v oblasti hlubokého učení rychle zastarávají, a proto nejaktuálnější podklady k použitým technologiím jsou přebírány například z oficiálních webů vývojářů, odborného webu *Medium (Towards Data Science)* či blogu Andreje Karpatyho, šéfa vývoje umělé inteligence a hlubokého učení v Tesle. Popsaná teoretická východiska jsou k datu odevzdání aktuální a zahrnují i vývoj v první čtvrtině roku 2020. Praktická část využívá moderní technologie a osvědčené architektury neuronových sítí. Práce se v bodech, kde to bylo relevantní, drží teprve nedávno představené osnovy, kterou pro publikace doporučuje používat RSNA (Radiological Society of North America).

2 Vymezení tématu a cíle práce

Cílem práce je prozkoumat současné možnosti využití umělé inteligence v oboru diagnostické radiologie, konkrétně způsoby klasifikace dvourozměrných obrazových dat pomocí hlubokých konvolučních neuronových sítí – zmapovat současný stav ve vývoji technologií, vědě i praxi a vyzkoušet vybrané osvědčené druhy dostupných architektur sítí na rentgenových snímcích, které poskytla Fakultní nemocnice Hradec Králové.

Konkrétně se jedná o snímky rukou a řešenou úlohou bude určování, zda ruka na snímku má nebo nemá příznaky revmatických onemocnění – revmatoidní artritidy, psoriatické artritidy nebo osteoartrózy. Využito bude několik architektur již vyvinutých konvolučních neuronových sítí, které v době tvorby této práce patřily mezi nejlepší veřejně dostupné. Záměrem je vyzkoušení proveditelnosti na úrovni ověření koncepce (proof of concept).

Předpokladem dobrých výsledků libovolné neuronové sítě, lépe řečeno libovolného počítačového programu, je kvalita vstupních dat – proto bude odpovídající pozornost věnována právě datovým podkladům a jejich předzpracování. Vzhledem k nevelkému počtu dostupných snímků se práce zaměří na metody umožňující dosáhnout dobrých výsledků i s menším množstvím dat – přenesené učení (transfer learning) a možnosti znásobení počtu vstupních snímků technikami rozšíření dat (data augmentation).

3 Medicínské zobrazování a charakteristiky onemocnění

V této kapitole je představen obor radiologie a podrobně osvětleno RTG vyšetření s důrazem na faktory ovlivňujících kvalitu obrazu – požadavky radiologů a softwaru na hluboké učení se totiž v některých ohledech liší. Dále jsou popsána ta revmatická onemocnění, na která se tato práce vymezuje, a jejich charakteristické projevy na RTG snímcích rukou, protože právě rozdíly v projevech jsou předpokladem proveditelnosti jejich rozpoznání pomocí hlubokého učení.

3.1 Radiologie (medicínské zobrazovací metody)

Radiologie je lékařský obor využívající zobrazovací metody k diagnostice a terapii. Název je dědictvím z historie, kdy se radiologie věnovala pouze ionizujícímu záření. Dnes již pojem nevystihuje obor v plné šíři a občas se lze setkat také s přesnějším označením *medicínské zobrazovací metody (medical imaging)*. Jednoslovný pojem radiologie je ale kratší a částečně i ze setrvačnosti se nadále používá. Obor se dělí do dvou větví:

- **Diagnostická radiologie**

Získává a vyhodnocuje obraz lidského těla či jeho funkcí za účelem určení diagnózy nebo prognózy onemocnění. Dělí se na všeobecnou radiologii, pediatrickou radiologii, neuroradiologii a mamologii.

- **Intervenční / terapeutická radiologie**

Využívá zobrazovacích metod k navigaci invazivních zákroků, spadají sem např. cévní intervence, odběry vzorků tkání, diagnostické či terapeutické punkce aj.

3.1.1 Přehled zobrazovacích metod

Zobrazovací metoda (modalita) je způsob získávání obrazových dat. Následující stručný přehled radiologických zobrazovacích metod byl zpracován podle zdrojů [1, 2]. U všech uvedených metod lze použít kontrastní látky, pro každou metodu specifické.

- **Skiografie; prostý snímek; plain radiograph; x-ray; RTG**

Sumační obraz vzniklý průchodem rentgenového záření anatomickou oblastí, liší se použitým elektrickým napětím (tvrdý vs. měkký snímek) a typem projekce.

- **Skioskopie**

Dynamické sumační zobrazení taktéž využívající RTG záření. Od prostého snímku se liší tím, že umožňuje sledování dějů v čase.

- **Výpočetní tomografie; CT**

Využívá RTG ke snímání objemových dat, která převádí na řezy v určitých rovinách. Nejvíce se používají axiální řezy – kolmé na osu těla.

- **Sonografie; ultrazvuk; UZ**

Generuje obraz na základě odrazu ultrazvuku na rozhraní tkání (o různé mechanické impedanci), umožňuje zobrazení řezů, má velmi vysoké časové rozlišení.

- **Magnetická rezonance; MRI**

Využívá interakce jader atomů s magnetickým polem a generovaný signál převádí na 2D řez. Umožňuje nejen anatomické zobrazení orgánů, ale i semikvantitativní zobrazení například difuze O₂, nervových drah apod. Hlavní využití je v neuroradiologii.

- **Pozitronová emisní tomografie; PET (v kombinacích PET/CT; PET/MR)**

Využívá detekce gama záření vyvolaného následkem rozpadu pacientovi podaných radionuklidů. Umožňuje cílené zobrazení např. tkáně s vysokým metabolismem (tumor, zánět, ...), či anatomických struktur při podání specifických značených protilátek. Používá se v kombinaci s CT či MR pro určení lokality. V USA tato vyšetření spadají ještě do radiologie, v ČR již spíše do nukleární medicíny.

3.2 RTG vyšetření (skiografie)

Ze všech radiologických vyšetření je právě RTG tím nejčastějším [1, 2]. V ČR bylo v roce 2018 celkem 2338 diagnostických RTG přístrojů (mimo radiologii navíc ještě 7692 zubních), na kterých bylo provedeno 10,5 milionu výkonů [3].

Rentgenové záření je obdobně jako viditelné světlo jedním z druhů elektromagnetického záření vymezené vlnovou délkou v rozmezí 0,01 – 10 nm a jí odpovídající frekvenci 10²⁰ – 10¹⁷ Hz. RTG záření je ionizující a lidskému zdraví škodlivé [1, 4].

Zobrazení funguje na principu rozdílné míry pohlcení a rozptylu záření v různých tkáních těla. Malá část záření projde nedotčena skrz a dopadne na záznamovou vrstvu, čímž vznikne výsledný obraz [2, 5]. Na tento princip odkazuje také původní název pro RTG vyšetření – *skiografie*, do češtiny volně přeloženo jako *stínové zobrazení*, *stínová malba* nebo *stínový záznam*.

RTG záření se podle vlnové délky, respektive prostupnosti tkáněmi, dělí na měkké (vlnová délka větší než 0,1 nm) a tvrdé (menší než 0,1 nm) [1].

Tabulka 1: Přehled druhů RTG záření

označení záření	tvrdé	měkké
vlastnosti	menší vlnová délka, vyšší frekvence, vyšší energie	větší vlnová délka, nižší frekvence, nižší energie
vlnová délka	$\lambda < 0,1 \text{ nm}$	$\lambda > 0,1 \text{ nm}$
prostupnost tkáněmi	vysoká	nízká
použití na	měkké tkáně (plíce, ...), obecně tkáně tlustější nebo hutnější	skelet (ruce, žebra, ...) menší nebo jemnější měkké tkáně, mamografie, snímkování dětí

Na rozdíl od CT, které také využívá rentgenové paprsky, výstupy klasického RTG vyšetření jsou pouze dvourozměrnou reprezentací trojrozměrných objektů, takže k mnoha skiografickým vyšetřením je zapotřebí alespoň dvou projekcí – čelní (frontální) a boční (laterální), popřípadě projekce šikmé [2]. Podle toho, jakým směrem záření prošlo tělem, se rozlišují projekce předozadní (posteroanteriorní; PA) a zadopřední (anteroposteriorní; AP) [2].

3.2.1 Metody snímání

3.2.1.1 Klasické (analogové)

Podobně jako v klasické fotografii je obraz zachycen na fotografický film, který je následně vyvolán. Filmy se mohou dělit na fóliové (expozice se zesilovací fólií) a bezfóliové, které mají výbornou ostrost kresby, ale potřebují být více exponovány (větší zátěž na pacienta). Analogové metody jsou v současné době z mnoha důvodů na ústupu. [1]

3.2.1.2 Digitální

Modernější způsob, kdy se snímky získávají a zpracovávají v digitální podobě. Děje se tak dvěma způsoby: [2, 4, 5]

1. Metoda přímé digitalizace (přímé konverze) – podobně jako v digitálním fotoaparátu záření dopadá na snímač, který ho převede na elektrický signál.
2. Metoda nepřímé digitalizace (nepřímé konverze) – snímek je zachycen na fólii podobným způsobem jako u analogových metod a digitální obraz se získá až následným naskenováním fólie laserem.

Výhodami digitálních metod jsou: [4]

- rychlost – okamžitá kontrola snímku
- vysoká kvalita obrazu – omezení potřeby pořizovat další snímky
- větší dynamický rozsah – možnost dodatečně v počítači opravit pod- nebo přeexponované snímky bez nutnosti opakování vyšetření
- zkrácení expozičního času 2 až 5× podle typu vyšetření – omezení radiační zátěže pacienta a také pohybových artefaktů (způsobených chvěním, cukáním, dýcháním, ...) a snížení šance případného opakování vyšetření
- snadná distribuce po pracovištích v rámci nemocnice i mimo ni pomocí systému PACS
- možnost kopírování bez ztráty kvality
- možnost dodatečného zpracování obrazu v počítači (jas, kontrast, ostrost, zřetelnost, potlačení šumu, převod negativu na pozitiv, měření velikostí a úhlů atd.)

3.2.1.2.1 DICOM a PACS

DICOM (Digital Imaging and Communications in Medicine) je standardizovaný formát pro ukládání medicínských obrazových dat v počítači. Kromě samotných obrazových dat v rastrové 2D reprezentaci (s podporou ukládání více řezů neboli vrstev do jednoho souboru) i objemové 3D reprezentaci umožňuje ukládat i bohatá metadata s informacemi o pacientovi a parametry provedeného vyšetření, například: jméno a příjmení pacienta, pohlaví a datum narození, modalitu, název a typ přístroje, datum a čas vzniku snímku, hodnoty radiační dávky, jednoznačný identifikátor snímku (který by údajně měl být celosvětově unikátní), měřítko vzdálenosti, hodnoty

jasu a kontrastu, anotace (vyznačené oblasti na snímku) apod. Protokol DICOM také přesně definuje způsob digitální komunikace mezi přístroji. Zatím poslední verzí je specifikace 3.0, která je průběžně aktualizována [1]. DICOM soubory mohou mít přípony .dc3, .dcm, .dic a rastrová data mohou být uložena s barevnou hloubkou 8, 10, 12 nebo 16 bitů [6].

PACS (Picture Archiving and Communication System) je počítačový systém skládající se ze zabezpečené sítě, koncových stanic (počítače, terminály), úložiště dat a dat samotných. Umožňuje bezpečné skladování a snadné předávání dat ve formátu DICOM mezi lékaři a pracovišti. PACS prostřednictvím API (Application Programming Interface; rozhraní pro vzájemnou komunikaci mezi softwarovými aplikacemi) umožňuje napojení na další informační systémy a software, například i programy na analýzu radiologických snímků (klasické i využívající hlubokého učení) musí umět s tímto systémem komunikovat [1, 5, 7].

3.2.2 Kvalita a vlastnosti RTG obrazu

Kvalitou obrazu se rozumí celkový vzhled obrazu a jeho vhodnost pro dané (diagnostické) účely [5]. Kvalitu lze popsat dvěma základními parametry: [5]

1. Kontrast zobrazení – schopnost rozlišit od sebe dvě přilehlé oblasti, tedy dva různé odstíny.
2. Prostorové rozlišení – schopnost rozeznat jemné detaily. Rozlišovací schopnost je definovaná jako nejmenší vzdálenost dvou bodů, které se ještě zobrazí odděleně. Při menší vzdálenosti body splynou a zobrazí se jako jeden. Rozlišovací schopnost rentgenového zobrazení se pohybuje kolem 0,5–2 mm.

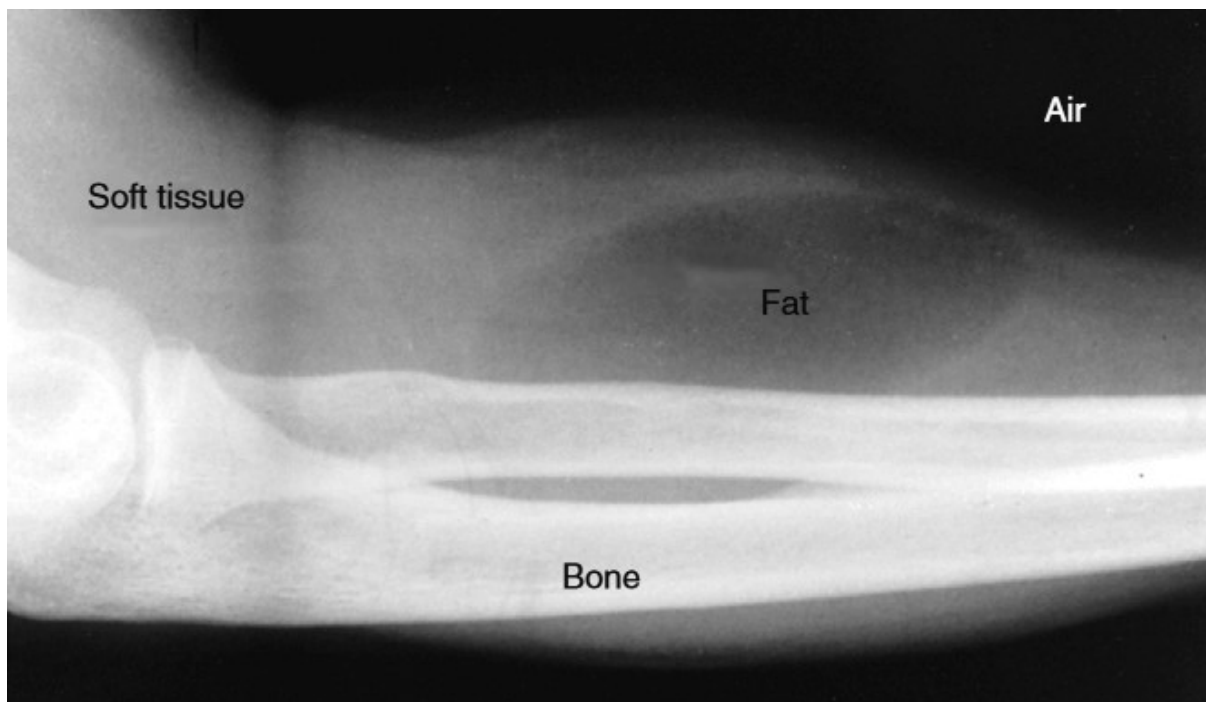
Vliv na tyto vlastnosti mají zejména charakter zobrazované tkáně, pohyby pacienta, parametry snímání a dávka záření, velikost optického ohniska, použitá metoda získání a zpracování obrazu, dodatečné úpravy a zobrazení snímku [1, 5].

3.2.2.1 Charakter zobrazované tkáně

V závislosti na míře pohlcení záření je na RTG snímcích obvykle rozlišitelných 5 úrovní stínů (tedy úrovní hustoty zobrazených tkání či objektů): [2]

1. Vzduch (resp. „prázdný“ prostor) a negativní kontrastní látky – černé nebo velmi tmavé
2. Tuk – obecně šedý, tmavší než svaly nebo krev
3. Voda (krev) a měkké tkáně – šedé
4. Kostí (vápník) – světlé, téměř bílé
5. Kov (kloubní protézy, dráty, šrouby, nesundatelné prsteny) a pozitivní kontrastní látky – bílé

RTG je účinné především na zobrazování skeletu. Měkké tkáně (jako třeba tkáň jater nebo sleziny) jsou na RTG snímku vykresleny chabě. To samé platí i pro svaly a tuk. Naopak kosti nebo vzduchové bubliny, například v žaludku, se zobrazí velmi kontrastně a zřetelně [2].



Obr. 1: Základní odstíny tkání pozorovatelné na RTG snímku

Boční projekce předloktí ukazuje 4 úrovně stínů, kov a kontrastní látky chybí. Tuk v měkké tkáni předloktí není normální stav – jde o tzv. lipom, nezhoubný nádor tvořený tukovou tkání.

Zdroj: převzato z [2]

3.2.2.2 Pohyby pacienta

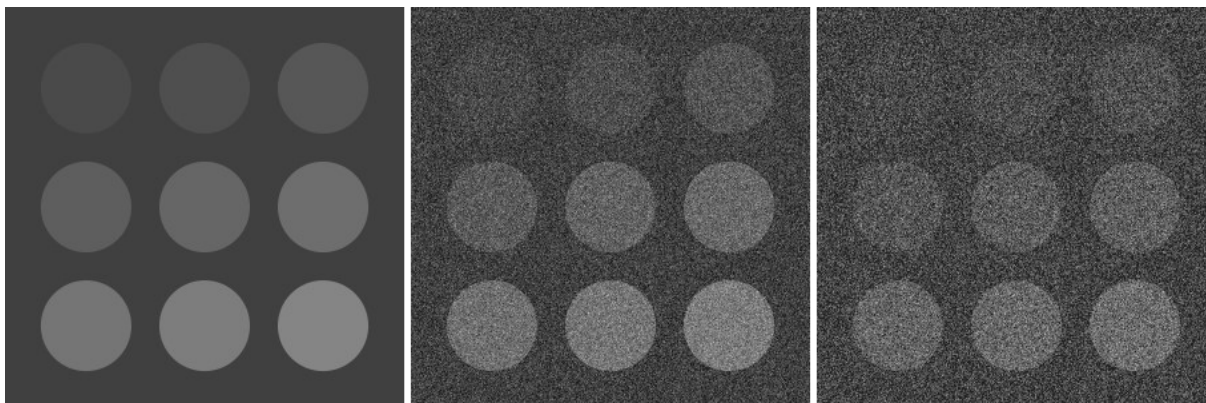
Vlivem pohybů pacienta může vzniknout pohybové rozmazání (pohybová neostrost), která snižuje prostorové rozlišení snímku. Neostrost může být způsobena zachycením některého z fyziologických pohybů (tep srdce, dýchání) a dalšími pohyby pacienta v průběhu expozice, typickými u neklidných pacientů, novorozenců a kojenců. Tento druh neostrosti lze omezit zkrácením expozičního času při současném zvýšení intenzity záření a synchronizací expozice se srdečním cyklem (tzv. *EKG gating*) nebo dýchacím cyklem (*respiratory gating*) [1, 5].

3.2.2.3 Parametry snímání a dávka ozáření

Nastavitelné parametry jsou napětí (kV), proud a čas (mAs). Těmi se ovlivňuje tvrdost či měkkost (energie) a expoziční doba. Čím je záření tvrdší, tím je snímek méně kontrastní, vyrovnanější [1]. Cílem diagnostiky pomocí ionizujícího záření je vystavit člověka pouze tak nízké dávce, jaká je postačující pro dané vyšetření. To často vede ke kompromisům, protože čím menší je dávka záření, tím horší je kvalita obrazu – kontrast zobrazení je snížen šumem.

Jde zejména o tzv. kvantový šum, který se odvíjí od počtu fotonů zachycených snímačem. Například při pediatrických RTG vyšetřeních jsou dávky záření velmi nízké a počet fotonů je poměrně malý. Zachycení fotonu snímačem je ze statistického pohledu náhodný jev, některé body snímače proto zachytí mnoho fotonů, jiné naopak málo [5].

U digitálních metod vstupuje do hry navíc elektronický šum, který má původ v elektrických obvodech snímací a zobrazovací soustavy [5].



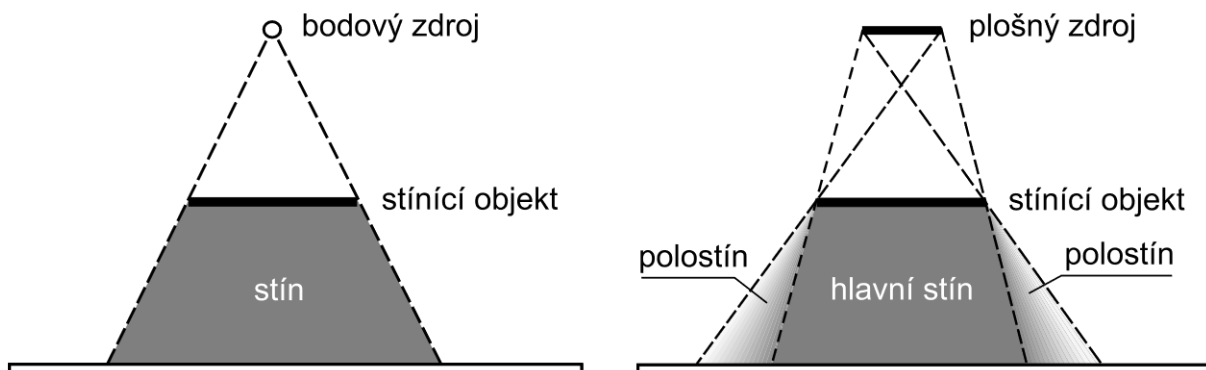
Obr. 2: Vliv šumu na kontrast zobrazení

Šum snižuje kontrast mezi jinak jasně odlišitelnými oblastmi; potlačuje jejich jinak jasně definované kontury. Jak je v této práci popsáno v kapitole Problém černé skříňky, šum také dokáže zmást algoritmy založené na neuronových sítích. V tištěné verzi této práce mohou kvůli omezenému dynamickému rozsahu papírové reprodukce splývat oblasti i v šumem nezatíženém čtverci úplně vlevo.

Zdroj: vlastní zpracování

3.2.2.4 Velikost optického ohniska

Velikost optického ohniska (zdroje záření) určuje limitní dosažitelnou prostorovou rozlišovací schopnost zobrazení, tzv. geometrickou neostrost. Z pohledu geometrické teorie platí, že v ideálním případě by stínový obraz byl dokonale ostrý, jen pokud by záření vycházelo z jednoho bodu [8]. V praxi toho nelze dosáhnout a čím je ohnisko větší, tím je obraz více neostrý, neboť vznikají polostíny. Běžná velikost ohniska RTG přístrojů je 0,5–2 mm [1, 5].



Obr. 3: Ostrost stínu v závislosti na velikosti zdroje záření

Zdroj: převzato z [8]

3.2.2.5 Použitá metoda získání a zpracování obrazu

Ostrost a rozlišovací schopnost je rovněž ovlivněna vlastnostmi obrazového detektoru a následném způsobu zpracování obrazu [5].

U analogových metod šlo o kvalitu fotografické emulze, citlivost a velikost zrna filmu, při jehož vyvolání mohlo docházet k mnoha různým vadám (závoje a skvrny, stopy nečistot, škrábance a otisky prstů, apod.) [1].

Digitální metody nepřímé konverze mají schopnost rozlišit objekty o velikosti do 100 μm , detektory s přímou konverzí objekty o velikosti 50–100 μm . U digitálních metod sice odpadají problémy doprovázející analogové metody, namísto toho se objevují nové. U nepřímé konverze

může při snímání na fólii docházet ke geometrickým deformacím vlivem ohybu fólie. Dále záleží na použitém skeneru a obslužném softwaru. U přímých metod má roli rozlišení a konstrukce snímače a obrazového procesoru převádějícího signál na obrazová data.

Oběma metodám jsou společné vlastnosti souborového formátu, do kterého je snímek uložen (zejména bitová hloubka a případná komprese) [1, 2].

3.2.2.6 Dodatečné úpravy a zobrazení snímku

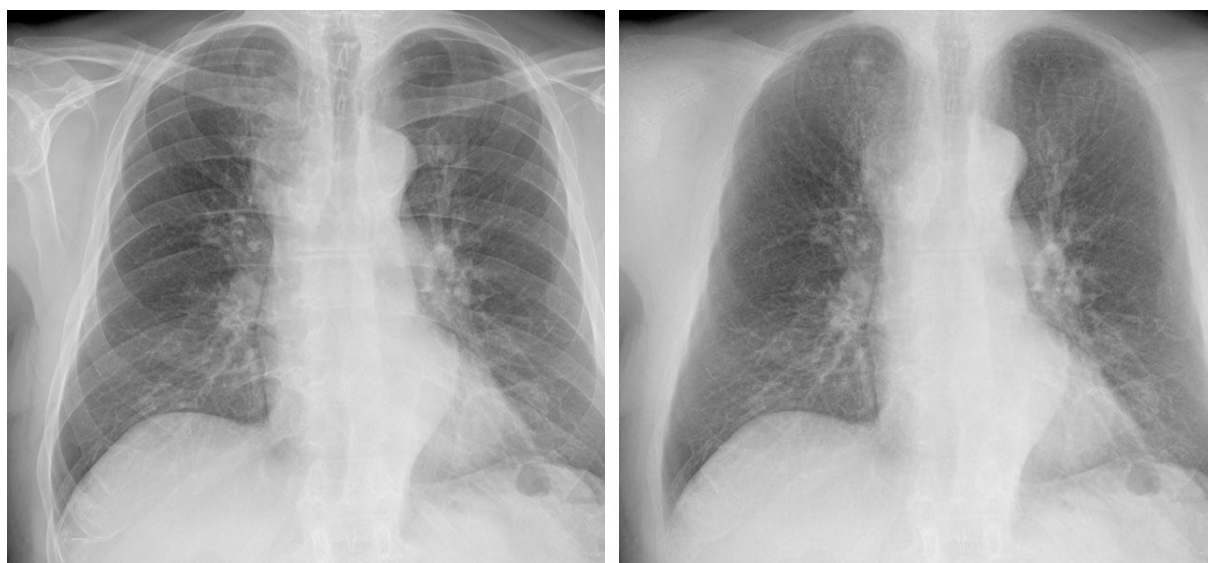
Digitální metody poskytují také řadu nových způsobů ovlivnění výsledné podoby snímku. Výrobci RTG zařízení jako jsou Siemens, Philips či Fujifilm nabízejí spolu s přístroji i software na zpracování snímků. Software všech tří výrobců dokáže optimalizovat ostrost, jas a kontrast včetně opravy míst pod- a přexponovaných. Dále nabízí funkci potlačení šumu, což umožňuje použít menší dávky záření a může se hodit zejména u pediatrických snímků.



Obr. 4: Vlevo úprava kontrastu a ostrosti (zřetelnosti), vpravo oprava podexponovaných míst
Tyto a další úpravy nabízí software DiamondView Max firmy Siemens.

Zdroj: převzato z [9]

Kromě těchto již běžně rozšířených funkcí každý výrobce nabízí i nějaké speciální. Fujifilm například pokročilou redukcí moaré nebo zvýraznění hran. Software společnosti Philips [10] nabízí vědecky vyvinutou úpravu Bone Suppression (potlačení kostí) [11], která dává podobné výstupy jako při odečtení dvou snímků exponovaných s různou frekvencí záření, ale zvládá to jen s jedním snímkem (tzn. menší radiční zátěž pacienta) a pokročilými algoritmy na úpravy obrazu a detekci struktur a vzorů. Zda je či není použito hluboké učení, to autoři výslovně nezmiňují [11].



Obr. 5: Technologie Bone Suppression (potlačení kostí) na snímku plic

Zdroj: převzato z [10, 11]

Softwarové úpravy mají vliv na zhodnocení snímku radiologem nebo softwarem, a to pozitivní i negativní. Například autoři softwaru BoneExpert (popsán v kapitole 6.2.1) uvádějí, že verze 1 a 2 jejich programu měly problém s přijímáním snímků s doostřenými hranami či jinak výrazněji vylepšovanými – modely hlubokého učení na takto upravených snímcích netrénovali, a tím pádem měl program nesnáze s jejich rozpoznáním [12].

Dále záleží na zobrazovacích parametrech monitoru – typu zobrazovacího panelu (TFT, IPS, OLED, ...), jeho úhlopříčce, rozlišení (pro radiologické účely typicky podobné 2K nebo 4K) a hustotě pixelů na palec, kalibraci a barevném profilu, jinými slovy schopnosti věrně reprodukovat určitou škálu odstínů. Kvalitu zobrazení mohou negativně ovlivnit také další zdroje světla v místnosti [5]. Pokud není snímek zobrazen v režimu přiblížení 1:1 (pixely obrázku odpovídají pixelům monitoru) a je nějakým způsobem zvětšen či zmenšen, znamená to, že byl upraven procesem převzorkování a interpolace, čímž mohlo dojít ke zkreslení a vzniku artefaktů.

3.2.3 Postup vyhodnocení RTG snímku radiologem

Analýza RTG snímku, respektive jakéhokoliv radiologického obrazu, je netriviální proces vyžadující notnou dávku předchozích zkušeností, trpělivosti a pečlivosti. Níže je volně parafrázován postup vyhodnocení tak, jak ho po 20 letech zkušeností v oboru v knize *Essentials of Radiology* [2] popsal Fred Mettler, profesor radiologie na Lékařské fakultě Univerzity v Novém Mexiku. Postup je zde uveden především aby ukázal, že vyhodnocení snímku v praxi neznamena jen určení diagnózy, ale detailní analýzu snímku samotného, jeho srovnání s předchozími snímky a syntézu s anamnestickými údaji o pacientovi – tedy mnohem více, než kolik toho pokrývají současná řešení využívající hluboké učení, popsaná v následujících kapitolách.

Na počátku vyhodnocení je v první řadě potřeba rozumět typu snímku, jeho orientaci a omezením použité zobrazovací metody. Dále je vhodné přihlídnout na věk a pohlaví.

Následuje hledání abnormálních nálezů. K tomu je zapotřebí znát normální anatomii zkoumané části těla, jaké může mít variace a jak běžně vypadá na výstupech dané zobrazovací metody. Následně je vhodné abnormální oblasti popsat, avšak nesnažit se hned vyvozovat diagnózu. To je správná cesta k diferenciální diagnostice (proces stanovení diagnózy na základě všech dostupných informací) – jinak totiž hrozí, že se člověk mentálně upne na předčasně vyřčenou diagnózu, která vůbec nemusí být správná.

Po popsání pozorovaných abnormalit je dobré je revidovat ve světle klinických nálezů. Některá onemocnění se projevují velice charakteristicky, většina jich ale může nabývat více podob v závislosti na fázi progresu. Zkušenosti odborníci po letech praxe vědí, na jaká místa se dívat a co hledat. Přestože z povahy věci nelze definovat jasná pravidla, obecná znalost patofyziologie onemocnění je velmi nápomocná.

Mnoho dalších informací lze získat dohledáním a porovnáním snímků z předchozích vyšetření a záznamů o pacientově stavu. To může přivést pozornost radiologa i na ty oblasti snímku, kterým by se jinak podrobněji nevěnoval.

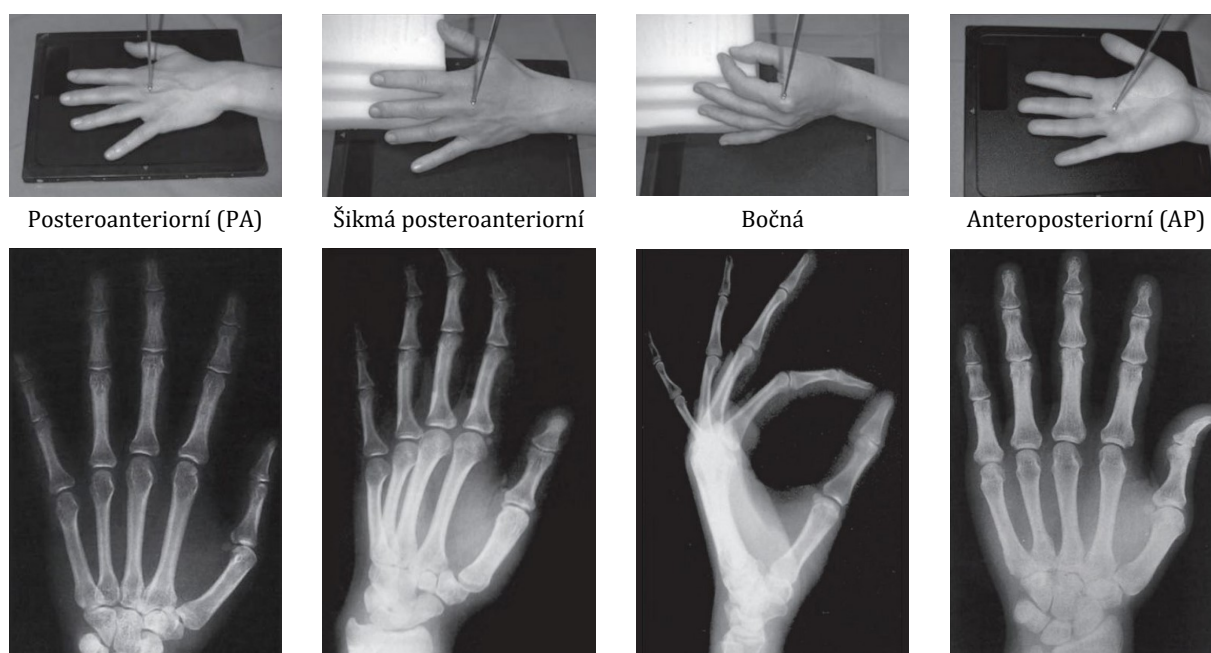
Nastává poslední krok. Pozorované nálezy mohou často být důsledkem více příčin. Aby nějaká varianta nebyla opomenuta a zároveň byly vyloučeny ty nesprávné, je vhodné mít po ruce sepsaný kontrolní seznam možných příčin/diagnóz, který si člověk na závěr projde a porovná, jaké by byly odlišnosti vůči tomu, jak se jeví nálezy na snímku.

3.3 RTG ruky

Důvody pro pořízení snímků rukou mohou být následující: [2]

- traumata (zlomeniny a jiná poranění)
- revmatická onemocnění
- změny způsobené infekcí (infekční artritida)
- kornatění tepen [13]
- tumory [2]

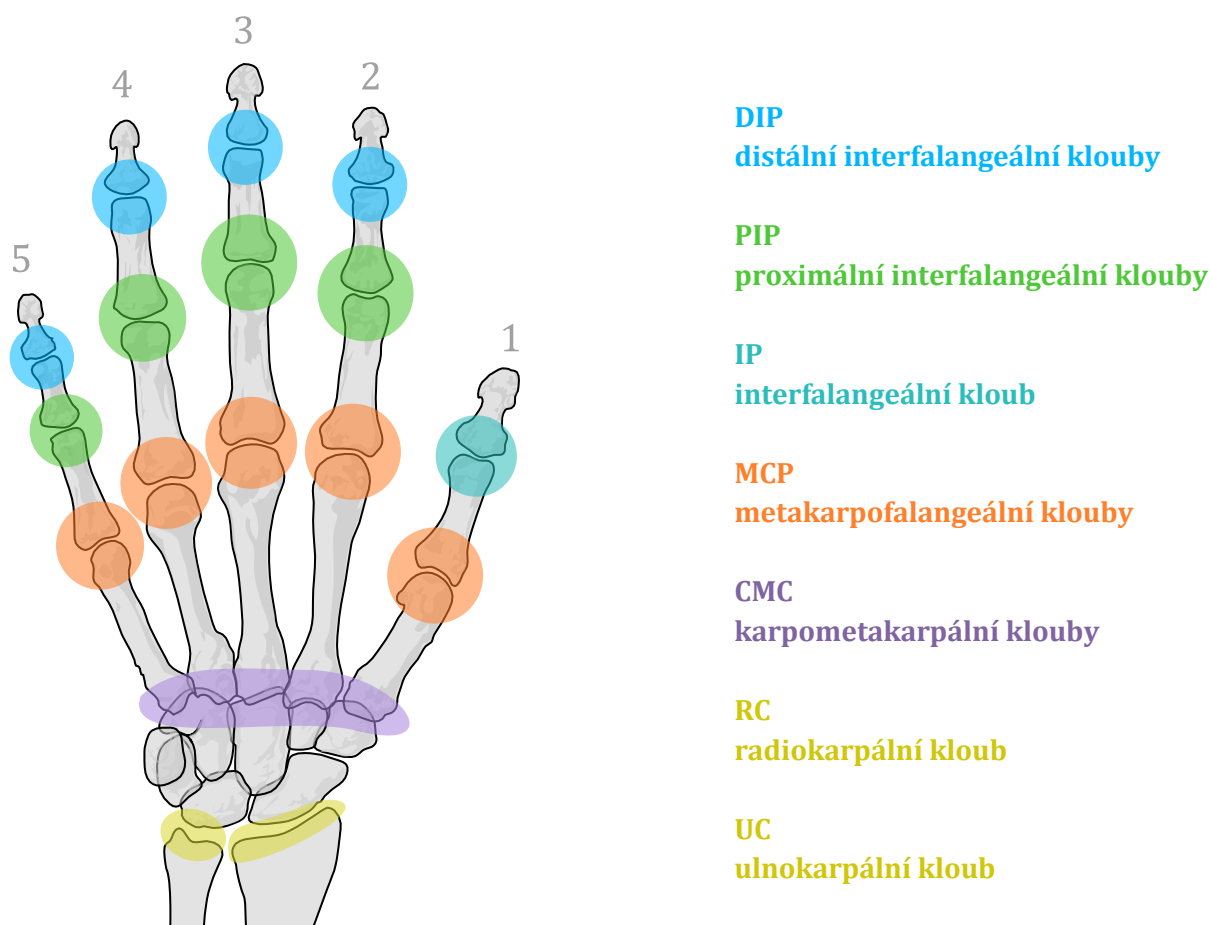
Vyšetření kvůli traumatům a revmatickým onemocněním svoji četností obecně výrazně převažují nad ostatními a nejinak je tomu i ve FNHK [2, 13]. Kornatění tepen a tumory jsou naopak méně časté a početně málo zastoupené indikace (důvody pro provedení určitého testu či výkonu). Odlišné jsou způsoby projekce. V případě traumat se snímá jen příslušná ruka a počet a typ projekcí záleží na konkrétním případě, kromě klasické PA je často zapotřebí ještě boční či šikmá projekce. Na snímcích traumatických rukou lze pozorovat zlomeniny, luxace (vykloubení) kloubů nebo cizí tělesa – kovové střepiny či špony, skleněné střepy či dřevěné třísky (poslední dva jmenované materiály nemusí být dostatečně kontrastní vůči měkkým tkáním) [2, 13].



Obr. 6: Způsoby projekce ruky

Zdroj: převzato z [14]

V následujícím textu a kapitolách popisujících vlastní řešení se zmiňují anatomické pojmy jako označení a číslování jednotlivých kloubů, tak jak jsou znázorněny na obrázku.



Obr. 7: Označení kloubů a číslování prstů ruky

Zdroj: vlastní zpracování

3.4 Revmatická onemocnění a jejich projevy na RTG snímcích

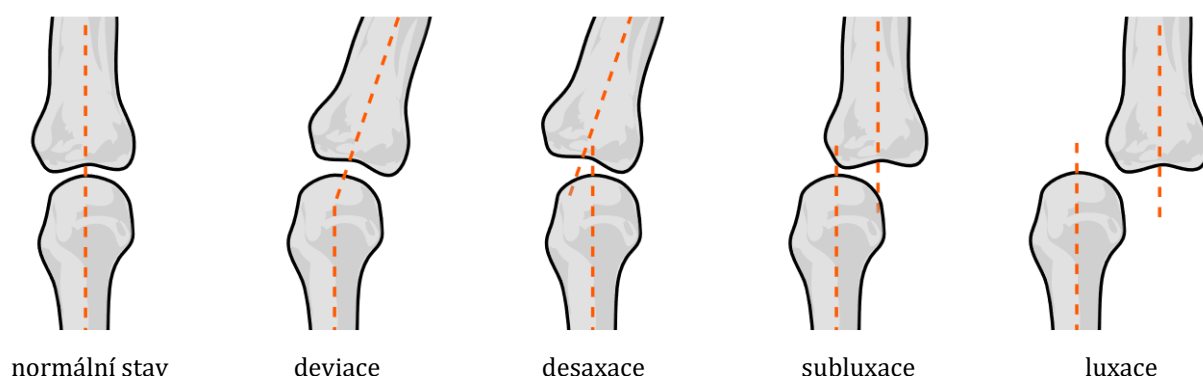
Cílem diagnostiky nejen v revmatologii je: [2, 4, 13]

- zachytit projevy nemoci co nejdříve
- správně nemoc klasifikovat (aby bylo možné určit odpovídající léčbu)
- určit progresi

RTG snímkování je i v revmatologii základní a nejčastější zobrazovací technikou. Umožňuje sledovat rozsah strukturálního poškození pohybového aparátu a jeho vývoj v dlouhodobém časovém horizontu [1, 4]. Samo o sobě je nedostačující při detekcích časných stádií onemocnění, tedy období, kdy je potřeba zahájit léčbu, aby se zabránilo nevratným poškozením kloubů [4]. Jako doplňkové vyšetření lze provést také sonografii nebo MRI, které dokáží zachytit i poškození měkkých tkání [4].

U revmatických onemocnění se zpravidla pořizuje jeden PA snímek zachycující obě ruce najednou, některá onemocnění totiž postihují obě ruce víceméně symetricky, jiná naopak asymetricky [2, 13, 15]. Pokud se nejedná o velmi ranou fázi, na RTG snímcích lze pozorovat zúžení

nebo úplný zánik kloubní štěrbiny, eroze kostí (úbytek materiálu) a změnu jejich hustoty (odvápnění), nekrózy, porušení osy kloubů a další příznaky. Porušení osy kloubů může mít formu deviace (vybočení se zachováním osy), desaxace (vybočení s odchylkou os), subluxe (neúplné vykloubení) nebo luxace (vykloubení) [15].



Obr. 8: Deformity kloubů

Zdroj: vlastní zpracování

Tato práce se z revmatických onemocnění omezuje na pouze revmatoidní artritidu, psoriatickou artritidu a osteoartrózu, a to zejména z důvodů, že tato onemocnění

- mají poměrně typické projevy, které se manifestují na rukách – dobrý předpoklad k tomu, aby úloha jejich rozpoznání byla řešitelná pomocí hlubokého učení
- jsou mezi pacienty s revmatickými onemocněními rukou zastoupena nejčastěji – lze tedy získat nejvíce obrazových dat pro využití v metodách hlubokého učení

Tabulka 2: Přehled revmatických onemocnění řešených v této práci

český název	anglický název	druh	výskyt
revmatoidní artritida	rheumatoid arthritis	zánětlivé	systémové onemocnění
psoriatická artritida	psoriatic arthritis	zánětlivé	systémové onemocnění (přidružené k psoriáze)
osteoartróza	osteoarthritis (degenerative arthritis)	degenerativní	pouze klouby

Z hlediska proveditelnosti rozpoznání těchto onemocnění pomocí hlubokého učení je důležité, že v akutních fázích by onemocnění měla být rozlišitelná, nicméně k překryvům v pozorovaných projevech docházet může [13].

Pro úplnost je potřeba uvést, že onemocnění, která zasahují ruce, existuje mnohem více. Mezi poměrně běžnými je například dna (dnavá artritida) [15, 16], reaktivní artritida, mnoho druhů infekčních (septických) artritid s původem bakteriálním, virovým či mykotickým,... Podobných chorob je ještě více, ne všechny se ale manifestují na rukou a navíc mnoho jich má poměrně různorodé projevy. Využití hlubokého učení k rozpoznání zmíněných onemocnění není předmětem této práce, ačkoliv ideální a v praxi použitelné řešení by samozřejmě mělo zahrnovat i tyto možnosti.

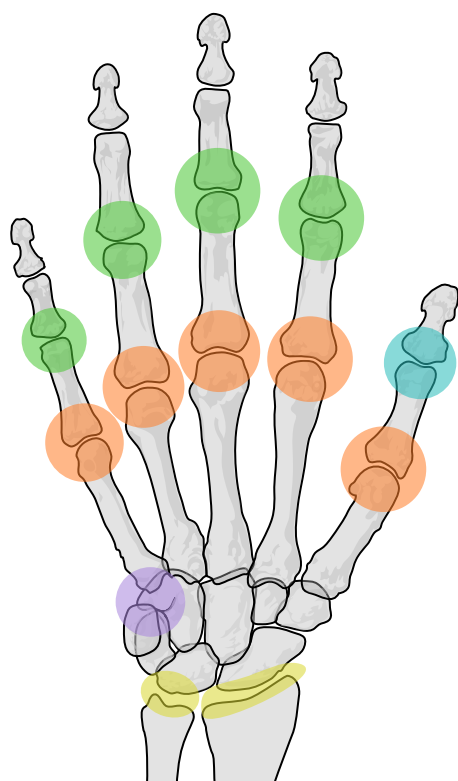
3.4.1 Revmatoidní artritida

Revmatoidní artritida (RA) je chronické systémové zánětlivé autoimunitní onemocnění, které postihuje asi 1 % populace. Častěji (v poměru až 5:1) jsou postiženy ženy [15]. Onemocnění začíná ve středním věku s maximální incidencí ve věku 50–60 let, avšak začít může v každém věku, ve velmi vzácných případech i u dětí [2, 17]. V typickém případě se RA projevuje symetrickým postižením kloubů rukou. Nejčastěji bývají zasaženy na ruce klouby RC, MCP, PIP, na nohou klouby metatarzofalangeální (MTP). Velké klouby jako kolena nebo ramena jsou v raném stádiu nemoci postiženy vzácně, jejich postižení provází spíše pokročilé stadium nemoci. Klouby DIP u RA zpravidla postiženy nebývají [15].

Plně rozvinuté onemocnění se po několikaletém průběhu bez efektivní léčby projevuje destrukcí kloubních tkání včetně šlachového aparátu a rozvojem kloubních deformit především na drobných kloubech ruky. Může se jednat o subluxace, luxace, desaxace, deviace, flekční kontraktury (vysoké napětí svalů) apod. Dále například ulnární deviace zápěstí, deformity prstů do tvaru labutí šije (hyperextenze PIP a flexe DIP), *boutonnière* (*knoflíkové dírky* nebo lépe *stříšky*) či kladívka [15, 17].

RA je systémové onemocnění, kromě kloubů postihuje také svaly, oči, plíce či srdce. U nemocných s RA se častěji vyskytují kardiovaskulární onemocnění a osteoporóza [15, 17].

3.4.1.1 Diagnostika



Obr. 9: Klouby typicky postižené RA
Zdroj: vlastní zpracování

Nemoc není možné vyléčit, ale včasná léčba dokáže oddálit či zastavit její projevy. Určení diagnózy RA je založeno na kombinaci typických klinických příznaků, specifických laboratorních nálezů a charakteristických změn na RTG snímku rukou [15, 17]. Laboratorní (krevní) testy potvrdí správnost diagnózy a umožňují porovnat stavy před a po aplikaci léčiv [17]. RTG snímek pomůže vyloučit jiné diagnózy, lokalizovat a určit stupeň poškození kloubů a kostí [17].

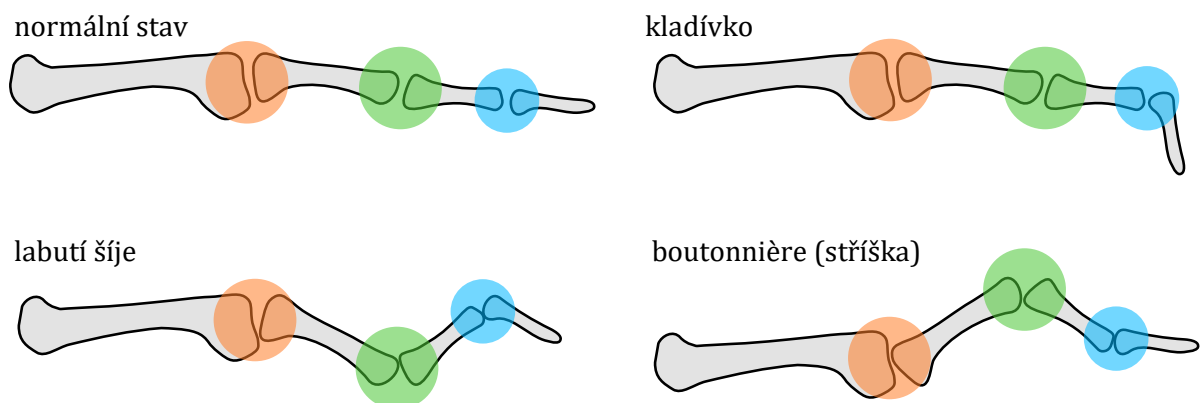
Na klasifikaci RA byla v roce 2010 sestavena nová klasifikační kritéria EULAR/ACR (European League Against Rheumatism / American College of Radiology), která umožňují diagnostikovat i raná stádia onemocnění, což je předpokladem úspěšné léčby [15, 17].



Obr. 10: Ruce postižené RA

Pokročilé stadium nemoci, lze pozorovat erozivní subluxaci a ulnární deviaci většiny MCP kloubů a dále degenerativní změny UC kloubů. AP (předozaďní) projekce.

Zdroj: převzato z [18]



Obr. 11: Deformity prstů typické pro RA

Z bočního pohledu a s barevným vyznačením kloubů MCP, PIP a DIP.

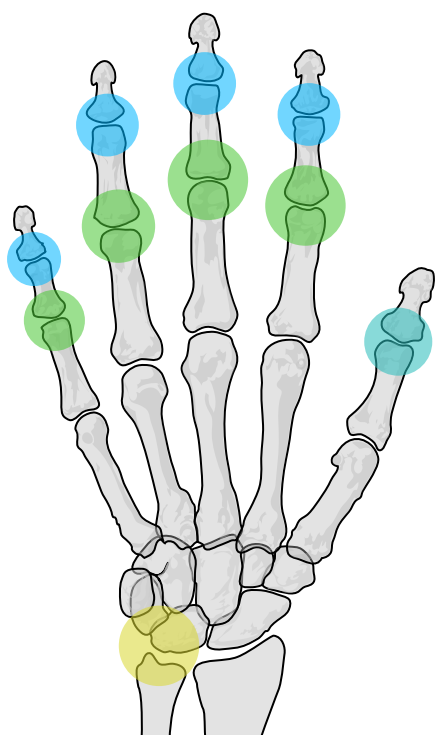
Zdroj: vlastní zpracování

3.4.2 Psoriatická artritida

Psoriatická artritida (PA; též PsA) je zánětlivé revmatické onemocnění sdružené s psoriázou. Psoriáza samotná postihuje až 3 % populace [17]. PsA se vyskytuje asi u 5–8 % pacientů s psoriázou, u závažných forem psoriázy až u 39 %. V 75 % případů se PsA vyvíjí u nemocných s kožním postižením předcházejícím o několik let. U 15 % případů se psoriáza a PsA manifestuje současně a jen zhruba u 10 % PsA předchází manifestaci psoriázy nebo je jediným projevem nemoci [15].

Rozlišuje se 5 forem nemoci: asymetrická, symetrická, distální, mutilující, axiální. Asymetrická je podle některých autorů nejběžnější formou, nejčastěji je lokalizována v drobných ručních (včetně DIP) a nožních kloubech. Distribuce postižení má radiální (paprscitý) charakter, často je současně postižena probíhající šlacha a dochází k difuznímu otoku prstu [15].

3.4.2.1 Diagnostika



Diagnóza PsA je založena na klinických a radiologických kritériích u pacienta s psoriázou – diagnóza je poměrně přímočará, problémy ale mohou nastat u lehčích forem postižení, kdy mohou pomoci RTG snímky. Laboratorní vyšetření umožňují sledování aktivity choroby [15].

Pro diagnostiku nemoci se od roku 2006 využívají klasifikační kritéria CASPAR (CLASSification criteria for Psoriatic ARthritis) [15, 17].

Obr. 12: Klouby typicky postižené PsA
Zdroj: vlastní zpracování



Obr. 13: Ruce postižené PsA
Zdroj: převzato z [19]



Obr. 14: Jiný příklad ruky postižené PsA
Zdroj: převzato z [2]

3.4.3 Osteoartróza

Osteoartróza (OA) je heterogenní skupina degenerativních chronických kloubních chorob, jež mají podobný klinický a morfologický obraz. Ženy jsou ohroženy více [17]. Jedná se o celosvětově nejčastější onemocnění kloubů, typicky postihuje kolena, kyčle, ruce, páteř a palce na nohou [15–17]. Při onemocnění dochází k degenerativní přeměně kloubní chrupavky a reaktivním změnám přilehlých měkkých tkání a subchondrální kosti [15].

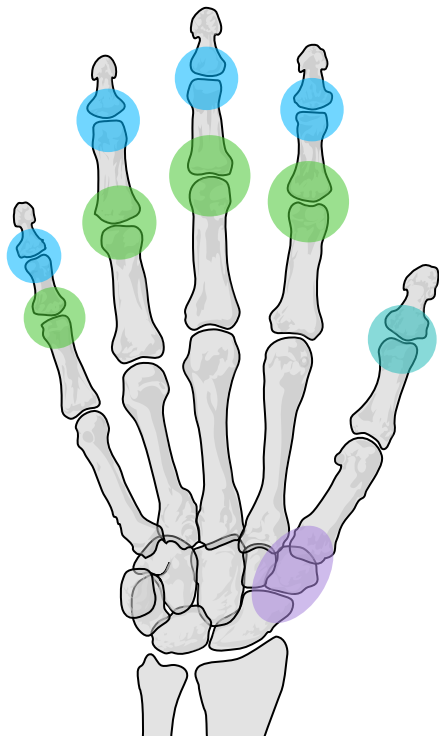
OA se začíná projevovat od středního věku. Přesné příčiny nemoci nejsou známy, předpokládá se, že vliv má dědičnost a genetické faktory, přetížení kloubů vlivem obezity nebo náročné fyzické práce. Rozlišuje se OA primární a sekundární, nasedající na předcházející patologický kloubní proces.

Příznaky jsou bolest a ztuhlost postiženého kloubu, později omezená pohyblivost a instabilita. Rozlišuje se OA rukou, nohou a páteře, případně také další vzácnější formy.

OA na ruku nejčastěji postihuje klouby DIP, PIP a kořenové klouby palců (CMC-1). Klouby MCP a zápěstí OA nepostihuje. V pokročilejším stádiu vzniká deformita a dochází k omezení pasivní i aktivní hybnosti [15].

Zvláštní formou je erozivní OA, jež se projevuje výraznějšími obtížemi, zánětlivým obrazem, rychlejší progresí a větším sklonem k deformitám [15].

3.4.3.1 Diagnostika



Obr. 15: Klouby typicky postižené OA
Zdroj: vlastní zpracování

Sestavení diagnózy OA je založeno na kombinaci typických klinických příznaků a charakteristických změn na RTG snímku rukou. Speciické laboratorní testy v tomto případě nejsou, ale jejich provedení může pomoci vyvrátit jiné diagnózy, jako například RA [16, 17].

Nálezy na RTG snímcích korelují se symptomy a klinickými nálezy chabě, mnoho starších pacientů může mít nález odpovídající OA, ačkoliv jsou bez dalších příznaků [17].

V raných stádiích lze pomocí MRI identifikovat oslabení chrupavky, edém (otok) subchondrální kosti či zánět kloubní výstelky, avšak v praxi se MRI používá jen vzácně [16, 17].



Obr. 16: Ruce postižené OA
Zdroj: převzato z [20]

4 Umělá inteligence, strojové a hluboké učení

Umělá inteligence (artificial intelligence; AI) je velmi široký vědní obor. Pojem je v současné době z marketingových důvodů velmi módní až nadužívaný [21] a často se objevuje ve spojení s dalšími výrazy jako strojové učení (machine learning; ML), umělé neuronové sítě (artificial neural networks; ANN) nebo hluboké učení (deep learning; DL). Následující kapitola vysvětlí, co tyto pojmy vlastně znamenají, proč má smysl řešit hluboké učení v radiologii, jak vypadá proces učení neuronové sítě a jak lze zkontrolovat a vyhodnotit, zda byl úspěšný.

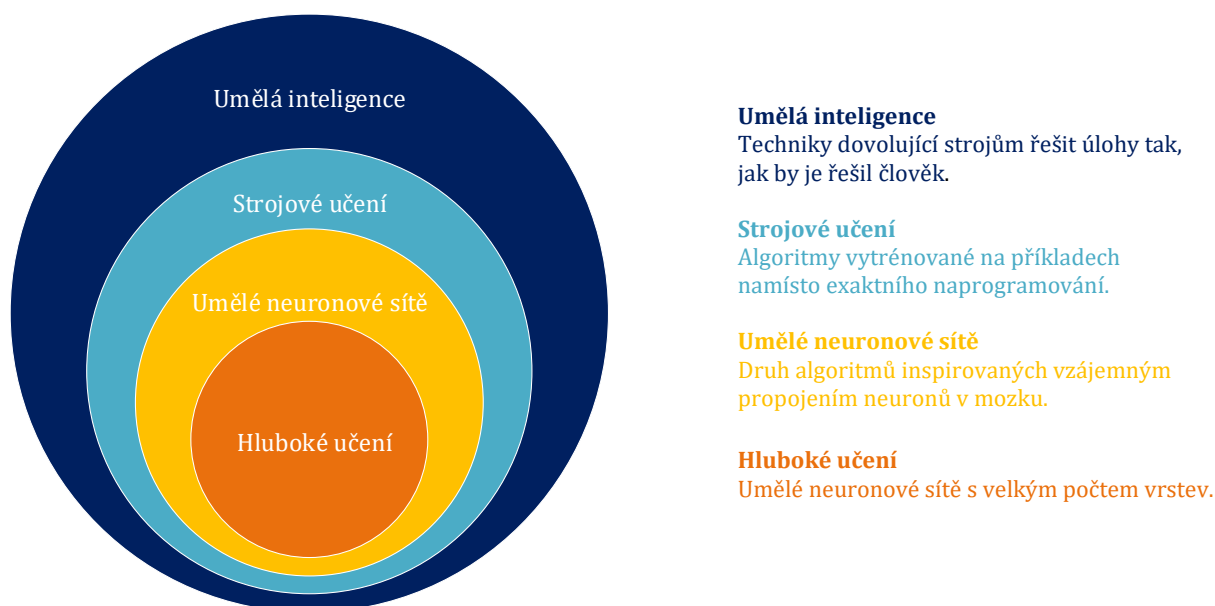
4.1 Umělá inteligence

Na inteligenci v klasickém smyslu slova existuje vícero způsobů nahlížení a vícero definic a nejinak je tomu i s inteligencí umělou [22]. Umělá inteligence jako vědní obor má mnoho různých směrů výzkumu a cílů. Zatímco strojům, které skutečně *jsou inteligentní*, je lidstvo při současném stavu poznání stále velice vzdálené, řešení do jisté míry *projevující známky inteligentního chování* [23] jsou již dnes dostupná a pomalu se začínají uplatňovat v mnoha odvětvích. V kontextu této práce, tedy vyhodnocování obrazových dat, lze umělou inteligenci vymezit takto:

„Umělá inteligence je vědní obor popisující chování strojů (počítačů) napodobující kognitivní funkce lidské mysli, jako je učení nebo řešení problémů.“ [24]

Nebo konkrétněji a s ohledem na současný stav poznání:

„Umělá inteligence je obor počítačové vědy zabývající se získáváním, rekonstrukcí, analýzou a interpretací medicínských snímků napodobováním lidského chování.“ [22]



Obr. 17: Vztah umělé inteligence, strojového a hlubokého učení

Umělé neuronové sítě se v některých materiálech již neuvádějí, protože téměř splývají s hlubokým učением. Podmnožiny ilustrují pouze vzájemné vztahy, jejich vyobrazené velikosti nemají výpovědní hodnotu.

Zdroj: vlastní zpracování podle [22]

4.2 Motivace – proč se zabývat umělou inteligencí v radiologii?

Curtis Langlotz ze Stanfordovy Univerzity v časopisu *Radiology: Artificial Intelligence* tvrdí, že na otázku „Nahradí umělá inteligence radiology?“ je správná odpověď „Radiologové používající umělou inteligenci nahradí ty, kdo ji nepoužívají.“[25] I další odborníci o umělé inteligenci v kontextu zpracování medicínských dat mluví prostě jako o *nástroji* [25, 26]. Panuje totiž všeobecná shoda na tom, že současné metody umělé inteligence, přesněji řečeno strojového a hlubokého učení, vzdělané odborníky s léty praxe v žádném případě nenahradí, zato pro ně mohou mít několik významných přínosů. Obecně se uvádějí následující:

- **Zvládnutí neustále rostoucího množství dat**

Předpokládaný objem medicínských dat se každé tři roky zdvojnásobuje [27]. Populace stárne, onemocnění přibývá, množství radiologických vyšetření roste a s tím roste i nápor na radiology. Více vyšetření znamená méně času na zpracování dat a vyšší pravděpodobnost chyb. Studie ukazují, že k chybným interpretacím dochází zhruba ve 4 % radiologických vyšetření, v závislosti na druhu vyšetření to ale může být i výrazně více [27].

Trend narůstajícího množství dat ke zpracování je vidět obecně ve všech oborech lidské činnosti a dalším cílem také je, aby znalosti z dat odvoditelné mohly být drženy aktuální a aktualizované, protože data se stále vyvíjejí [23]. Vypořádat se s rostoucím náparem dat je v silách současných IT řešení, překážky jsou typicky jinde než v technologiích [28].

- **Automatizace rutinních činností**

Na současném stupni vývoje může umělá inteligence pomoci s těmi rutinními úkoly, ke kterým je k dispozici dostatek dat, jsou víceméně přímočaré a nevyžadují kombinaci až příliš mnoha druhů vstupních údajů. Tedy úkoly, které radiologové musí provádět poměrně často. [22]

- **Podrobnější diagnostika**

Mnoho současných řešení a výzkumů se zaměřuje na extrahování a poskytování dodatečných informací. Může jít třeba o kvantitativní popisy snímků, které by umožnily porovnat odlišnosti jedince od zbytku (statistické) populace pomocí měřitelných charakteristik [22, 27].

- **Nabídnutí dalšího názoru**

Výstupy počítačových programů mohou sloužit jako druhý, nezávislý pohled na situaci. Při určení diagnózy mohou hrát roli pojistky, že nic nezůstalo opomenuto nebo přehlédnuto. Umělá inteligence má také potenciál zachytit klíčové nálezy v raných fázích onemocnění, kdy jsou pro lidské smysly na hranici rozlišitelnosti (nebo pod ní). Důležité je, že i počítačový program musí stejně jako radiolog být schopen svůj názor podložit řádnými důkazy [22, 29].

- **Omezení inter- a intraobservační variability**

Různí radiologové se mohou ve svých výstupech soustředit na různé aspekty a používat mírně odlišný způsob vyjadřování. I výstupy činnosti jednoho jedince nemusí být vždy stejnorodé a ošetřující lékaři s tím musejí počítat. Software využívající umělou inteligenci může variabilitu radiologických zpráv omezit a určitým způsobem je standardizovat [22, 27, 30].

4.3 Strojové učení

Strojové učení je jedním z podoborů umělé inteligence, zabývá se algoritmy a technikami, které dokáží z dat získávat informace a znalosti, jinak řečeno *učit se*.

V tradičním programování programátor přesně definuje pravidla, jak má algoritmus zareagovat na určitý vstup. Naproti tomu v případě strojového učení programátor pouze dodá data a vybere požadovaný algoritmus, co se pak sám snaží v datech nalézt vzorce, struktury či souvislosti. Fráze *učit se* označuje právě tento rozdílný přístup, kdy systém samočinně mění svůj vnitřní stav, aby se co nejlépe přizpůsobil vstupním datům [22, 31].

Pojmy strojové učení a umělá inteligence jsou často zmiňovány společně nebo používány zaměnitelně, ačkoliv neznamenaají totéž. Mnoho současných výtobytků, označovaných jako umělá inteligence, spadá právě do oblasti strojového (či přesněji hlubokého) učení.

Pomalou se začíná vžívat také nové dělení softwaru – programy využívající „klasické“, přesně definované algoritmy, se souhrnně označují jako Software 1.0 a programy fungující na bázi strojového učení, jež nemají chování definované exaktně, zato ale umožňují zvládat mnohem komplexnější úlohy, se označují jako Software 2.0 [32].

V širším kontextu (mimo zdravotnictví) do této nové kategorie spadají například v úvodu již zmiňované současné internetové vyhledávače, systémy pro reklamu a doporučování obsahu na webu, emailové spam filtry, software ovládající samořiditelná auta, rozpoznávání tváří na fotografiích a záznamech z bezpečnostních kamer, rozpoznávání textu a řeči, překladače jazyků nebo chatboti [32, 33].

Strojové učení se dělí do dvou základních větví: [33, 34]

1. Učení s učitelem (supervised learning)
2. Učení bez učitele (unsupervised learning)

Někteří autoři uvádějí jako třetí větev částečné učení s učitelem (semi-supervised learning) [22], jiní zpětnovazební učení (též sebeposilující učení; reinforcement learning) [33]. Tento druh učení principiálně připomíná učení s učitelem, avšak používá se pro jiné účely než zpracování dat, a proto není v této práci dále rozebírán [33].

4.3.1 Učení s učitelem

Učení s učitelem popisuje případ, kdy program pracuje s anotovaným datasetem, jinak řečeno zná vstupní hodnoty a také očekávané výstupní hodnoty. Program se *učí*, jak od daného vstupu dojít k požadovanému výstupu [23]. Algoritmy učení s učitelem lze dělit na

- **Regresní**

Cílem je schopnost pro nové vzorky vypočítat určité dodatečné nebo chybějící číselné údaje, typicky předpovědi do budoucnosti.

- **Klasifikační**

Cílem je schopnost nové vzorky zařadit, roztrždit.

Výstupem je konkrétní volba mezi třídami (kategoriemi), například:

- Zhoubný nádor / Nezhoubný nádor
- Zlomenina ano / ne
- Osteoartróza/ Revmatoidní artritida / Psoriatická artritida / Bez příznaků
- Určitá hodnota kostního věku

Pro zpracování obrazových dat existují ještě dva poddruhy klasifikační úlohy, které se nazývají detekce a segmentace [7]. Výstupem detekce jsou souřadnice či ohraničující rámečky kolem hledaných objektů, výstupem segmentace je pouze hledaný objekt vyjmutý ze zbytku obrazu.



Obr. 18: Varianty klasifikační úlohy
Vlevo klasifikace, uprostřed detekce, vpravo segmentace.
Zdroj: vlastní zpracování s využitím snímku [35]

4.3.2 Učení bez učitele

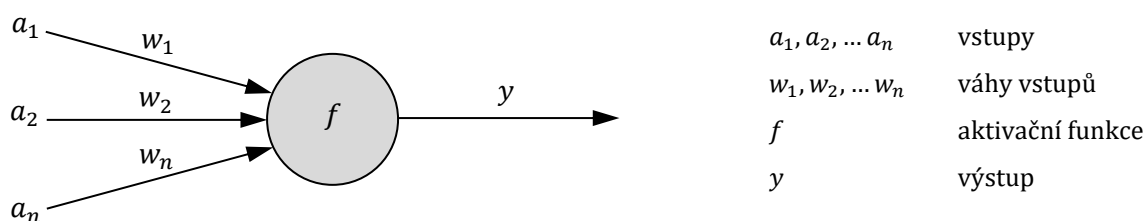
Učení bez učitele popisuje případ, kdy program pracuje s neanotovaným datasetem, tedy má k dispozici pouze vstupní hodnoty a sám se v nich snaží zorientovat, přesněji řečeno najít nějakou strukturu, vzorce, závislosti nebo korelace. Výstupem jsou data rozřazená do skupin (shluků), podle kritérií, která si definuje algoritmus sám. Pomocí učení bez učitele lze získat odlišný úhel pohledu na data a nalézt vazby a vztahy, které lidskému vnímání unikají [22, 23]. Algoritmy učení bez učitele lze dělit do dvou základních skupin:

- **Shluková analýza**
- **Analýza hlavních komponent**

4.4 Umělé neuronové sítě a hluboké učení

4.4.1 Formální neuron a neuronová síť

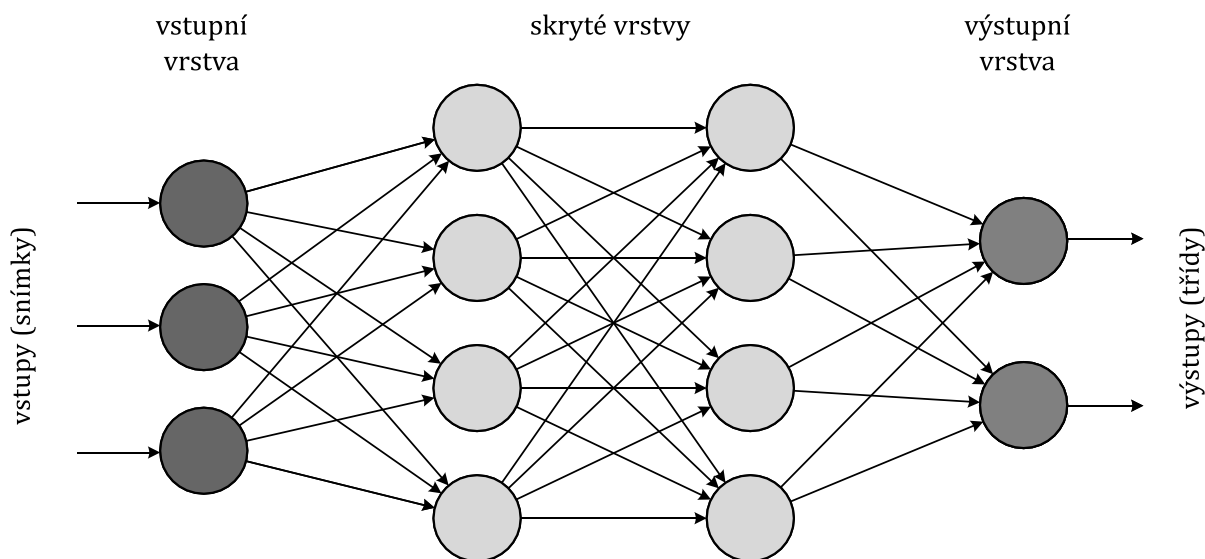
Umělé neuronové sítě jsou jedním z výpočetních modelů v rámci strojového učení. Jsou velmi volně inspirované biologickými strukturami komplexních živých organismů. Obdobně, jako je neuron základním stavebním kamenem mozku, je základním prvkem neuronové sítě tzv. formální (či umělý) neuron [33]. Zatímco v mozku existuje mnoho druhů neuronů a více způsobů přenosu a zpracování signálu, z nichž mnohé nebyly dosud objasněny, formální neurony využívají zjednodušený základní koncept, kdy neuron přijme vstupy, zpracuje je a odešle dál výstup [36].



Obr. 19: Obecné schéma formálního neuronu

Zdroj: vlastní zpracování podle [37]

Formální neurony jsou vzájemně propojeny, předávají si signály (číselné hodnoty) a transformují je prostřednictvím matematických funkcí. Každé propojení neuronů má přiřazenou určitou číselnou váhu, která reprezentuje sílu spojení [33]. Pávě tyto váhy představují paměť neuronu a během procesu *učení* (trénování) jsou nastavovány jejich hodnoty. Oproti mozku je i koncept propojení formálních neuronů značně zjednodušený. Skutečné neurony mají obecně více vstupů (dendritů) i více výstupů (axonů) a v rámci propojení neuronů mohou vznikat zpětnovazební smyčky [36]. Formální neuron má libovolný počet vstupů, ale pouze jeden výstup (aktivaci), který je vypočítán z váženého součtu všech vstupů pomocí aktivační funkce [33].



Obr. 20: Obecné schéma neuronové sítě

Zdroj: vlastní zpracování podle [37]

Formální neurony se v rámci sítě označují jako uzly, jejich propojení jako hrany. Jednotlivé uzly jsou členěny do vrstev, každá vrstva může zpracovávat vstupy odlišným způsobem. Během procesu učení putují signály z první (vstupní) vrstvy, přes vnitřní (tzv. skryté) vrstvy, do poslední (výstupní) vrstvy. Dalším zjednodušením je, že propojit lze pouze uzly mezi různými vrstvami, kdežto uzly náležící do stejné vrstvy zůstávají nepropojené [33]. Sdružování neuronů podle jejich funkce do určitých celků (vrstev, shluků,...) je známo i ze skutečného mozku [36]. Pro neuronové sítě, které mají velký počet skrytých vrstev, se vžilo označení *hluboké*. Někteří autoři z historických důvodů používají tento termín už pro sítě s jednou a více skrytými vrstvami [37]. Podle hlubokých sítí se celé této větvě strojového učení začalo říkat *hluboké učení* [37].

Samotné *učení* probíhá tím způsobem, že při dopředném chodu algoritmu vstupní data postupně procházejí sítí ze vstupní vrstvy do výstupní, kde je určen výsledek. Z rozdílu skutečného (předpovězeného) a očekávaného (správného) výsledku se spočítá chyba. Na základě této chyby jsou pak zpětně upraveny váhy spojení ve snaze chybu v příštím kroku minimalizovat – tato fáze se označuje jako *zpětná propagace*. Celý proces je opakován v mnoha krocích, dokud se síť nenaučí dojít k výstupům s takovou mírou chybovosti, která je již přijatelná [37].

Neurologové věří, že umělé neuronové sítě reprezentují a uchovávají informace přibližně podobným způsobem, jako skutečný mozek [36]. Neuronové sítě napodobující kognitivní funkce, jako je rozpoznání obrazu nebo zvuku, pravděpodobně taktéž velmi zjednodušeně a vzdáleně připomínají skutečné procesy v mozku [27, 36]. Z vědeckého pohledu to znamená, že výzkum umělých neuronových sítí může přinést více vzhledu i do fungování mozku [36].

Pro pořádek je vhodné dodat, že zde popsané umělé neuronové sítě jsou realizovány jako počítačové programy – algoritmy zapsané v programovacím jazyku. Snaha vyvinout umělé neurony schopné komunikovat s biologickými patří do jiného vědního oboru. Při jakémkoliv porovnání se skutečným mozkiem je také potřeba mít na paměti, že současné neuronové sítě neumí myslet a nejsou ničím jiným, než výpočetním mechanismem a způsobem, jak od zadaných vstupů dojít k požadovaným výstupům.

Ačkoliv současný koncept neuronových sítí je oproti realitě velkým zjednodušením, na některé úlohy se již osvědčily lépe než tradiční software (Software 1.0) [36]. Za povšimnutí stojí jedno zásadní omezení současných neuronových sítí. Dokáží totiž zpracovat pouze jeden druh dat. Existují architektury sítí specializované na zpracování textu, jiné na zpracování obrazu a další na zpracování zvuku. Je-li zapotřebí použít více druhů dat najednou, vyžaduje to zkombinování výstupů z více sítí. Takovým technikám se v anglické terminologii říká *ensembles* (volně přeloženo jako *seskupení, uskupení* nebo *ansámbl*). V roce 2017 publikovali vědci z projektu Google Brain v článku *One Model to Learn Them All* [38] architekturu sítě schopnou zpracovat více druhů dat současně. Síť v žádné doméně nedala vynikající výsledky a nevyrovnala se specializovaným architektuřím, zároveň ale ve všech doménách dosáhla výsledků poměrně dobrých [38, 39].

4.4.2 Matematické principy fungování neuronové sítě

Každý uzel (neuron) v obecné neuronové síti je reprezentován poměrně jednoduchým algoritmem, který je popsán na následujících řádcích včetně běžně používané terminologie. Výpočet, který algoritmus vyjadřuje, funguje stejně pro každý neuron v síti. Popis algoritmu byl sestaven podle zdrojů [33, 37, 40].

4.4.2.1 Algoritmus výpočtu aktivací pro jeden neuron

Výstupy

Výstupem neuronu je tzv. aktivace, což je číslo vypočtené pomocí aktivační funkce z váženého součtu vstupních hodnot. Aktivace se značí písmenem a a může nabývat libovolné hodnoty z intervalu $(0,1)$.

Vstupy

Vstupy jsou tvořeny aktivacemi propojených neuronů z předchozí vrstvy a jejich váhami. Vstupy se značí hodnotami a_1, a_2, \dots, a_n a jejich váhy w_1, w_2, \dots, w_n . Kromě těchto hodnot je vstupem ještě tzv. bias (česky vychýlení), což je přidaná váha, která umožňuje přímo ovlivňovat výstup aktivační funkce.

Postup výpočtu

Ze vstupních hodnot se nejprve spočítá jejich vážený součet z , ke kterému je přičten bias b . Pro lepší čitelnost je bias v pořadí sčítanců uváděn jako první. Výpočet z probíhá následovně:

$$z = b + (w_1 a_1 + w_2 a_2 + \dots + w_n a_n)$$

Což lze schematicky zapsat jako:

$$z = b + \sum_{i=0}^n a_i w_i$$

Hodnota z se následně použije k výpočtu aktivace y pomocí aktivační funkce:

$$y = f(z)$$

Hodnota y je hledaným výsledkem a výstupem celého neuronu. Celý výše uvedený postup lze shrnout do jednoho vzorce:

$$y = f\left(b + \sum_{i=0}^n a_i w_i\right)$$

Pokud byly vstupní hodnoty zdefinovány jako vektory \vec{a} a \vec{w} , lze předchozí vzorec vyjádřit také s pomocí skalárního součinu jako:

$$y = f(b + \vec{a} \cdot \vec{w})$$

Doplňující poznámky

- Uvedená ukázka je ve všech použitých zdrojích uváděna s drobným zádrhelem v označení. Aktivace je značena buď a nebo y , podle toho, zda má v daném kontextu roli vstupní nebo výstupní. Pro úvodní seznámení to může být srozumitelnější, v podstatě jde ale o tu samou hodnotu a v praktických výpočtech je vhodnější držet se při označení stejného symbolu a kontext (aktuální vrstvu) rozlišovat pomocí indexu [40].
- Funkcí, které slouží jako aktivační, se používá celá řada, často zmiňovaná je například funkce ReLu [37]. Uplatňují se zde i funkce definované po intervalech (po částech). Jejich detailní představení, příklady užití v neuronových sítích a popis související problematiky jsou mimo rámec této práce, zároveň jde o informace vcelku obecné a snadno dohledatelné.

4.4.2.2 Výpočet aktivací v praxi

V praxi nejsou výpočty implementovány po jednotlivých neuronech, ale rovnou po celých vrstvách pomocí vektorů a matic. Z aktivací, které jsou výstupem neuronů vrstvy l , se rovnou spočítají aktivace vrstvy $l + 1$. Následující ukázka vychází stejně jako minulá ze zdrojů [33, 37, 40] a ke značení z minulé ukázky přidává navíc tuto konvenci:

$$a_{\text{pořadí}}^{(\text{vrstva})} = a_n^{(l)}$$

Např. $a_5^{(2)}$ značí v pořadí 6. neuron ve 3. vrstvě (čísluje se od nuly).

$$w_{\text{do,od}}^{(\text{vrstva})} = w_{j,k}^{(l)}$$

Např. $w_{7,5}^{(2)}$ značí váhu spojení mezi 6. neuronem ve 3. vrstvě a 8. neuronem ve vrstvě následující.

Výstupy

Vektor $a^{(l+1)}$ všech aktivací z vrstvy $l + 1$

$$a^{(l+1)} = \begin{bmatrix} a_0^{l+1} \\ a_a^{l+1} \\ \vdots \\ a_n^{l+1} \end{bmatrix}$$

Vstupy

Vektor $a^{(l)}$ všech aktivací z vrstvy l

$$a^{(l)} = \begin{bmatrix} a_0^l \\ a_a^l \\ \vdots \\ a_n^l \end{bmatrix}$$

Matrice $W^{(l)}$ s vahami všech spojení mezi vrstvami l a $l + 1$

$$W^{(l)} = \begin{bmatrix} w_{0,0} & w_{0,1} & \cdots & w_{0,k} \\ w_{1,0} & w_{1,1} & \cdots & w_{1,k} \\ \vdots & \vdots & \ddots & \vdots \\ w_{j,0} & w_{j,1} & \cdots & w_{j,k} \end{bmatrix}$$

Vektor biasů b

$$b = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_n \end{bmatrix}$$

Postup výpočtu

Vektor $a^{(l)}$ všech aktivací z vrstvy l se vynásobí s maticí W odpovídajících vah, k výsledku se přičte vektor biasů b a na závěr se použije aktivační funkce f . Výsledkem bude vektor aktivací pro vrstvu $a^{(l+1)}$.

Výpočet lze zapsat vzorcem takto:

$$a^{(l+1)} = f \left(\begin{bmatrix} w_{0,0} & w_{0,1} & \dots & w_{0,k} \\ w_{1,0} & w_{1,1} & \dots & w_{1,k} \\ \vdots & \vdots & \ddots & \vdots \\ w_{j,0} & w_{j,1} & \dots & w_{j,k} \end{bmatrix} \begin{bmatrix} a_0^0 \\ a_1^0 \\ \vdots \\ a_n^0 \end{bmatrix} + \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_n \end{bmatrix} \right)$$

Schematicky zapsáno:

$$a^{(l+1)} = f(W \cdot a^{(l)} + b)$$

Pokud mezivýsledek v kulatých závorkách stejně jako v předešlém příkladu označíme z , lze použít i zjednodušený zápis:

$$a^{(l+1)} = f(z^{(l)})$$

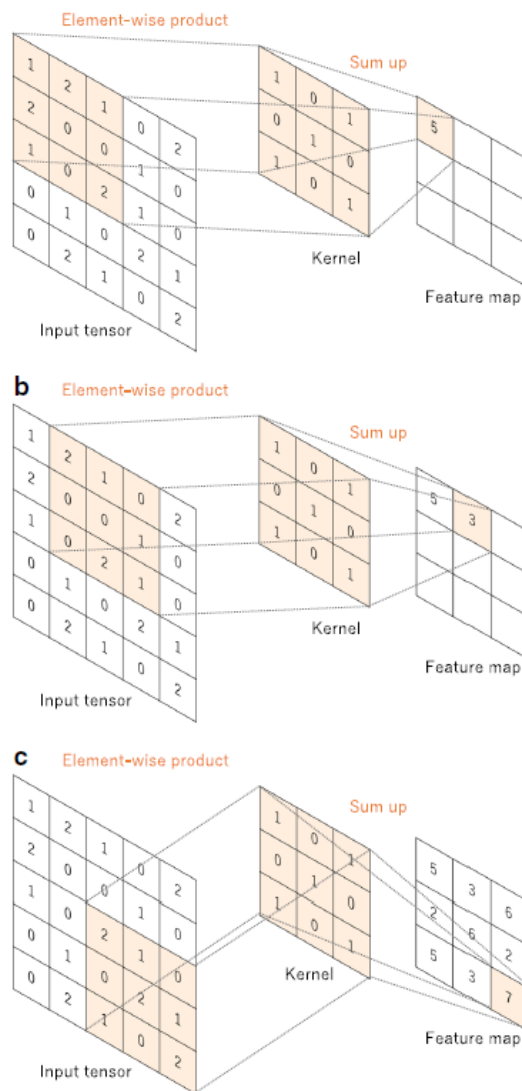
4.4.3 Konvoluční neuronové sítě

Na zpracování obrazových dat se používají tzv. konvoluční neuronové sítě (convolutional neural network; CNN; ConvNet). Tyto sítě se dokáží naučit nízko- i vysokoúrovňové rysy obrazu a existují ve variantách pro 2D i 3D data.

CNN se typicky skládá ze tří druhů vrstev: konvolučních, poolingových (sjednocujících) a plně propojených. První dva druhy ze vstupního snímku extrahují jeho rysy, poslední druh mapuje tyto rysy na výstupní třídy.

Proces konvoluce využívá čtvercovou matici vah nazývanou filtr či kernel. Princip konvoluce spočívá v tom, že se tento kernel v určitých krocích „posouvá“ po matici vstupních dat (vstupním snímku) a přitom se v každém kroku vypočítá skalární součin z hodnot v kernelu a hodnot v příslušné části matice vstupního snímku. Rozměry výstupního obrázku jsou menší a závisí na velikosti kernelu (nová délka strany se spočítá jako původní délka dělená velikostí strany kernelu) [41].

Tímto způsobem probíhá extrakce rysů obrázku. Jaké přesně rysy jsou extrahovány závisí na tom, jaké váhy jsou v kernelu nastaveny.



Obr. 21: Tvorba map rysů pomocí konvoluce
Zdroj: převzato z [41]

Nastavování vah se děje v procesu učení během zpětné propagace. Konvoluce se odehrává zpravidla v několika vrstvách, což znamená, že z původně velkého obrázku je po průchodu těmito vrstvami mnoho miniaturních obrázků zachycujících jednotlivé rysy – tzv. mapy rysů (feature maps) [41].

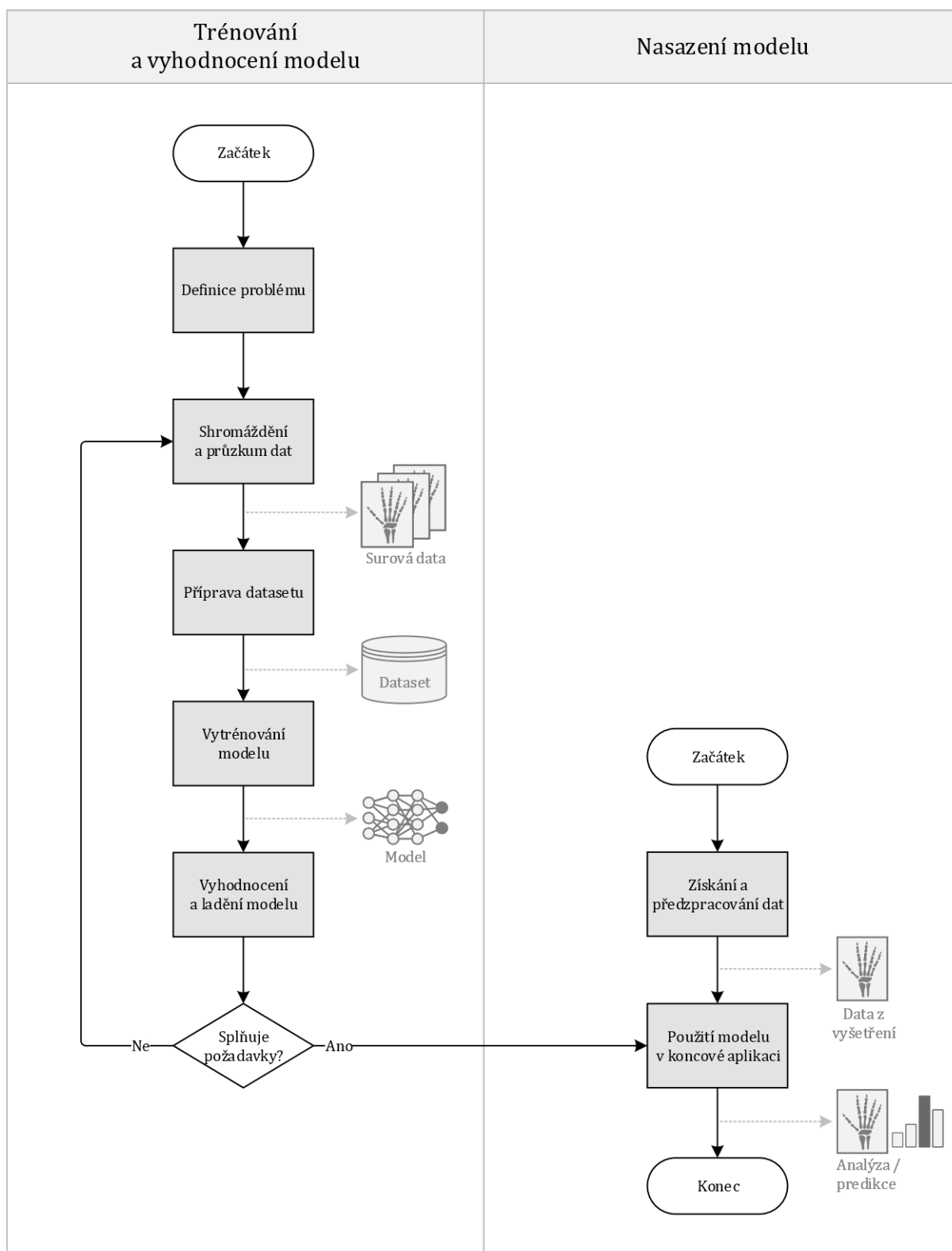
Během procesu učení se mění pouze zmíněné váhy v kernelech, naopak velikost kernelů, jejich počet a velikost kroku jsou vlastnosti, které se učením nemění a nastavují se předem (přesněji řečeno jsou pevně danou součástí architektury neuronové sítě) [41].

V poolingových vrstvách se malé bloky z konvolučních vrstev sjednocují do větších celků, tím se redukuje počet parametrů a extrahují důležité rysy. Díky konvolučním a poolingovým vrstvám funguje konvoluční neuronová síť v praxi tak, že nejnižší vrstvy reagují na hrany, vyšší vrstvy reagují na tvary a ještě vyšší vrstvy reagují na celé objekty [31]. Vědci mají za to, že zhruba podobně zpracovává zrakové vjemy i lidský mozek [27, 34].

A proč se vůbec konvoluční vrstvy používají? Bez nich by neuronová síť musela při trénování nastavovat váhy pro každý jeden pixel nezmenšeného vstupního snímku, což by bylo extrémně náročné na paměť GPU nebo RAM. Zároveň by nebyla schopná abstrakce [42].

4.4.4 Pracovní postup

U strojového či hlubokého učení je nutné rozlišovat dva rozdílné procesy. Jednak časově i výpočetně náročné *učení* (trénování), tedy přípravu a tvorbu matematického modelu, a jednak pouhé použití už hotového modelu ke zpracování (klasifikaci) nových dat.



Obr. 22: Posloupnost prací při vývoji řešení využívajícího strojové či hluboké učení

Posloupnost sama o sobě má obecnou platnost, doplňující symboly byly přizpůsobeny pro znázornění klasifikace RTG snímků rukou řešené v této práci. Zdroj: vlastní zpracování podle [31, 33, 37, 43]

4.4.4.1 Trénování modelu

Výstupem tohoto procesu je tzv. model – datová struktura obsahující neuronovou síť s naučenými váhami spojení, tedy obsahující způsob, jak od vstupních dat dojít k požadovaným výstupům. Celý proces trénování a vyhodnocení může být pojat jen jako jednorázová činnost, nebo může být podle charakteru řešené úlohy čas od času opakován, například když se po určité době nakumuluje velký počet nových dat, která je potřeba do stávajícího řešení zahrnout. Pro úspěšné vytrénování přesného a použitelného modelu jsou nejdůležitější a časově nejnáročnější fáze týkající se přípravy dat [7, 43].

A) Definice problému

Definice problému obnáší zodpovězení otázek: Co je cílem? Jaká úloha se bude řešit – regrese, klasifikace, detekce, segmentace? Je to proveditelné? Jaké budou přínosy a jaká rizika?

B) Shromáždění a průzkum dat

Otázky: Jaké zdroje dat jsou k dispozici? Bude dat dostatečné množství pro vytrénování modelu? Jak jsou početně zastoupeny jednotlivé třídy (kategorie)? V jakém formátu jsou data k dispozici? Jak řešit anonymizaci?

Následovat by měla manuální a důkladná kontrola dat. Mohou se objevit duplicity, poškozené snímky, chybně anotované snímky. Dále je potřeba zkoumat: Jakou roli hrají detaily? Jsou snímky spíše rozmanité nebo jednotvárné? Jaké mají společné vlastnosti? Kolik je okrajových případů a jak s nimi naložit?

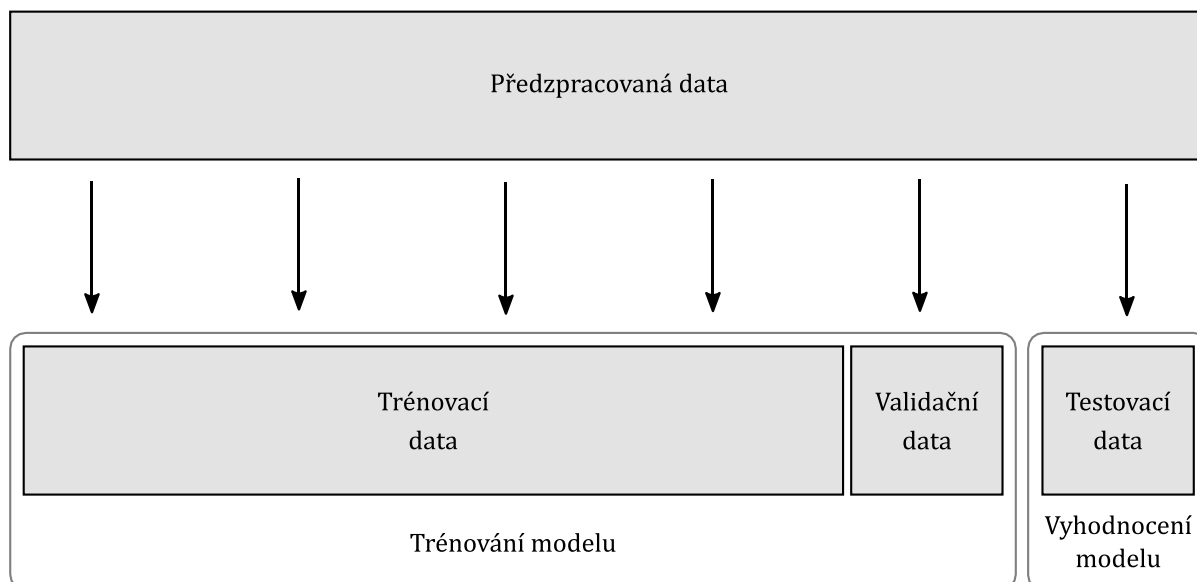
Tato fáze je klíčová, protože pouze kvalitní data jsou předpokladem kvalitních výsledků. Andrej Karpathy tuto fázi s mírnou nadsázkou pojmenovává jako *Splyňte s daty* („*Become one with the data*“). Google pro přehlednost uvádí shromáždění a průzkum jako dvě samostatné fáze [31].

C) Příprava datasetu

Po prozkoumání dat a základním odladění lze přistoupit k tvorbě datasetu. Ten se zpravidla dělí do 3 podmnožin: [7, 31]

- trénovací data (80 % z celkového množství)
- validační data (10 %)
- testovací data (10 %)

Trénovací a validační data se využívají během procesu trénování, testovací data slouží k závěrečnému vyhodnocení modelu a může být vhodné je vyčlenit ručně. V každé skupině by měly být jednotlivé třídy zastoupeny stejnou měrou, jakou jsou zastoupeny v datasetu celkově. Tedy nemělo by dojít k tomu, že například testovací skupina bude obsahovat vzorky pouze z jedné třídy. Pokud má být model univerzálně použitelný, doporučuje se, aby dataset tvořily vzorky z více rozdílných zdrojů (více nemocnic či přístrojů), což zajistí potřebnou variabilitu [29].



Obr. 23: Rozdělení datasetu a účel jednotlivých částí

Zdroj: vlastní zpracování podle [41]

Pozornost je potřeba věnovat také pořadí, data totiž mohou být ve výchozím stavu určitým způsobem seřazena, což může mít na trénování modelu vliv. K zajištění, že model nebude ovlivněn pořadím dat, se doporučuje vzorky promíchat nebo zajistit, že trénovací algoritmus je bude načítat v náhodném pořadí.

K úspěšnému vytrénování modelu jsou potřeba řádově tisíce až desítky tisíc případů (snímků). Čím více příkladů je sítí poskytnuto, tím lépe se naučí rozpoznávat obecněji platné příznaky diagnózy [44].

Pokud dat není dostatek nebo hrozí problém přeučení (kapitola 4.4.5.1), je potřeba přinejmenším implementovat techniku tzv. rozšíření dat (kapitola 5.8).

Než může být dataset použit k trénování, je nutné ho převést do formátu, se kterým neuronová síť umí pracovat. Například v této práci použitý framework TensorFlow vyžaduje data ve formátu .tfrecord v binární reprezentaci. Je také potřeba řešit, jakým způsobem jsou snímky převzorkovány a zmenšeny. Současné neuronové sítě navíc pracují se snímky pouze ve čtvercovém formátu a při převodu do něj může dojít k nechtěné deformaci snímků a jejich nesprávnému zpracování, jak bude představeno v kapitole 5.7.

D) Vytrénování modelu

Tato etapa je obecně velmi náročná na výkon hardwaru a čas, což bude osvětleno v kapitole 5.4. Hledání nejvhodnější architektury a parametrů trénování může probíhat manuálně nebo stále častěji též automatizovaně, umožňují-li to použité technologie.

Co se týče struktur sítě, většina softwarových frameworků a knihoven nabízí volně dostupné implementace špičkových struktur sítě, jako jsou VGG, Inception či ResNet. V závislosti na množství dat lze také v této fázi implementovat přenesené učení (kapitola 5.5).

E) Vyhodnocení a ladění

Po dokončení trénování, pokud se vůbec povedlo, je nutné model vyhodnotit na testovací části datasetu podle stanovených metrik. Spolu s algoritmem by měl stejná data posoudit i skutečný radiolog, aby bylo k dispozici srovnání úspěšnosti programu a člověka. Doporučuje se také, aby byl vytrénovaný model dán k dispozici na posouzení nezávislé třetí straně [29]. Více o tom v kapitole Doporučení pro výzkum a publikování.

Pokud se model neosvědčil, je potřeba vrátit se zpět do fáze průzkumu dat, pokud se osvědčil, je možné zabývat se jeho nasazením do praxe.

4.4.4.2 Nasazení modelu

Úspěšně vyhodnocený model může již být použit ve finálním softwaru. Zatímco trénování modelu je výpočetně a časově značně náročné, na jeho pouhé použití stačí průměrný hardware. Řada jednoduchých modelů v současnosti dokáže fungovat už i v internetových prohlížečích nebo na mobilech.

Výsledný software pro praktické nasazení musí načíst vstupní data, automaticky je upravit stejným způsobem, jako byla upravena trénovací data (např. oprava jasu a kontrastu), nechat snímek modelem vyhodnotit a pak prezentovat výsledky. Důležité je, že nejen pro medicínské použití by měl software být navržen tak, aby dokázal své výsledky podložit zdůvodněním (detailně rozebráno v kapitole Problém černé skříňky).

Před nasazením je potřeba vymyslet, jak bude nový software zapojen do již existující infrastruktury, tzn. jak bude komunikovat se systémem PACS a dalším softwarem. V současné době se začínají objevovat softwarové platformy, které navenek nabízejí radiologům jednotné uživatelské rozhraní a způsoby ovládání, zatímco uvnitř umožňují přes API zapojení mnoha různých modelů. Taková řešení urychlují nasazení a použitelnost vyvinutých řešení [7, 22].

Pro použití v medicíně je potřeba řešit ještě certifikační proces, kterým musí každý nový software projít před uvedením na trh. Certifikace pro umožnění prodeje v rámci EU je završena označením produktu značkou CE [45, 46]. Nasazení vytrénovaného modelu hlubokého učení, potažmo celého výsledného softwaru, může být realizováno dvěma způsoby:

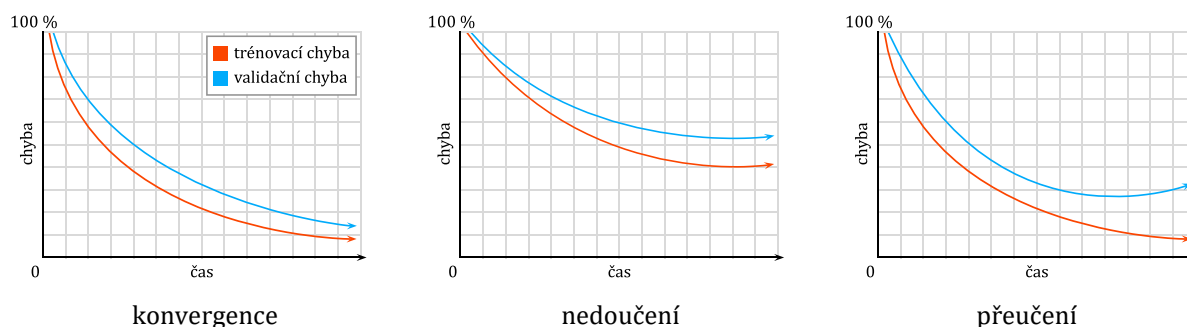
- instalace celého softwaru lokálně (s možností pravidelných občasných aktualizací)
- nasazení v cloudu (s možností častého nasazování nových verzí)

Současná softwarová řešení ukazují, že již i druhá uvedená možnost může být i v medicínském prostředí certifikována a použita [47].

4.4.5 Vyhodnocení modelu

4.4.5.1 Konvergence, nedoučení a přeučení

Ke kontrole, zda a jak úspěšně probíhá trénování modelu, slouží tzv. ztrátová funkce (loss function). Ta vyjadřuje, jak se vyvíjí celková chyba dosažená modelem, a udává se běžně v procentech. Průběh této funkce se zpravidla sleduje na grafu. Alternativně je možné tu samou skutečnost popsat nikoliv chybou, ale celkovou přesností, která se spočítá jako doplněk do 100 % k funkční hodnotě chybové funkce. Grafy jsou pak zrcadlově převrácené podle svislé osy.



Obr. 24: Možné varianty průběhu trénování modelu

Zdroj: vlastní zpracování podle [48]

Konvergence

Konvergence je ideální stav, kdy se v grafu křivky trénovací i validační chyby přibližují k ose reprezentující nulovou chybu. Trénovací chyba (chyba při klasifikaci trénovacích dat) je o něco menší než validační (chyba při klasifikaci validačních dat). Vytrénovaný model by v takovém případě neměl trpět nedoučením ani přeučením.

Nedoučení (underfitting)

K nedoučení dochází, když model selhává při zpracování nových dat, protože se nedokázal naučit komplexitu dat trénovacích. Jinak řečeno, když model až příliš generalizuje. Příčinami nedoučení mohou být:

- Trénování s nedostatečným počtem epoch nebo příliš nízkým koeficientem učení
- Málo skrytých vrstev v architektuře sítě
- Trénování na špatných rysech obrazu

Přeučení (overfitting)

Přeučení naopak vzniká, když se model až příliš specializuje na charakteristiky trénovacích dat a tím pádem ztrácí schopnost generalizovat a správně vyhodnotit data nová. Trénovací chyba sice stále klesá, ale validační chyba roste. K přeučení může dojít, pokud dataset není dostatečně obsáhlý a dostatečně rozmanitý.

4.4.5.2 Metriky pro hodnocení úspěšnosti klasifikačního modelu

Metriky v oblasti hlubokého učení zatím nejsou ustálené a standardizované [22, 49], běžně využívané metriky totiž nemusí být vždy použitelné – například u segmentace neexistuje přímočará interpretace pojmů falešně pozitivní a falešně negativní [22]. Pro praktickou část této práce byly použity níže popsané nejběžnější metriky používané pro klasifikační úlohy ve zdravotnictví i mimo něj. Obsah kapitoly je zpracován s využitím zdrojů [23, 50].

4.4.5.2.1 Binární klasifikace

Základní pojmy

Základní metriky se zpravidla uvádějí na příkladu binární klasifikace, kdy výstupem je buď zařízení do třídy 1 (pozitivní) nebo 0 (negativní). Klasifikace každého snímku může skončit jedním ze čtyř různých výsledků:

- Skutečně pozitivní (true positive)
Model předpověděl třídu 1 (pozitivní) a snímek do ní ve skutečnosti patří.
- Skutečně negativní (true negative)
Model předpověděl třídu 0 (negativní) a snímek do ní ve skutečnosti patří.
- Falešně pozitivní (false positive) – též známé jako chyba 1. druhu
Model předpověděl třídu 1, ale snímek do ní ve skutečnosti nepatří.
- Falešně negativní (false negative) – též známé jako chyba 2. druhu
Model předpověděl třídu 0, ale snímek do ní ve skutečnosti nepatří.

V závislosti na povaze řešené úlohy je nutné optimalizovat řešení tak, aby se omezil vznik buď chyby 1. druhu, nebo chyby 2. druhu.

Matice záměn (též matice chyb; confusion matrix; error matrix)

Čtyři výše uvedené varianty výsledků lze uspořádat do tzv. matice záměn, která přehledně znázorňuje, jak dobře si vytrénovaný model vede při určování jednotlivých tříd. V ideálním případě by na úhlopříčce matice (správně klasifikované případy) měly být hodnoty blížící se 100 % a ve zbytku matice (chybně klasifikované) hodnoty blížící se nule.

		Předpovědi	
		Pozitivní	Negativní
Skutečnost	Pozitivní	<i>TP</i>	<i>FN</i>
	Negativní	<i>FP</i>	<i>TN</i>

V této a následujících ukázkách je použito označení:

- TP* suma všech skutečně pozitivních případů
- TN* suma všech skutečně negativních případů
- FP* suma všech falešně pozitivních případů
- FN* suma všech falešně negativních případů

Z mace záměn lze odvodit několik metrik, z nichž nejdůležitější a nejrozšířenější jsou tyto:

Celková správnost (overall accuracy; accuracy)

Celková správnost označuje podíl všech správně klasifikovaných snímků z celkového počtu snímků. Odpovídá na jednoduchou otázku: Kolik snímků bylo správně klasifikováno?

$$A = \frac{TP + TN}{TP + TN + FP + FN}$$

čítatel: počet všech správně klasifikovaných snímků
jmenovatel: celkový počet snímků

Přesnost (precision)

Přesnost označuje podíl snímků správně klasifikovaných jako pozitivní z počtu snímků klasifikovaných jako pozitivní. Odpovídá na otázku: Kolik snímků klasifikovaných jako pozitivní je ve skutečnosti pozitivních? Vyšší přesnost znamená méně chyb 1. druhu.

$$P = \frac{TP}{TP + FP}$$

čítatel: počet snímků správně klasifikovaných jako pozitivní
jmenovatel: počet všech snímků klasifikovaných jako pozitivní

Senzitivita (sensitivity; recall; true positive rate)

Senzitivita označuje podíl snímků správně klasifikovaných jako pozitivní z počtu snímků, které skutečně pozitivní jsou. Odpovídá na otázku: Kolik snímků z těch, co měly být označeny jako pozitivní, bylo skutečně označeno? Vyšší senzitivita znamená více chyb 2. druhu.

$$R = TPR = \frac{TP}{TP + FN}$$

čítatel: počet snímků správně klasifikovaných jako pozitivní
jmenovatel: počet všech snímků ve skutečnosti patřících mezi pozitivní

Specificita (specificity; true negative rate)

Specificita označuje podíl snímků správně klasifikovaných jako negativní z počtu snímků, které skutečně negativní jsou. Odpovídá na otázku: Kolik snímků z těch, co měly být označeny jako negativní, bylo skutečně označeno?

$$S = TNR = \frac{TN}{TN + FP}$$

čítatel: počet snímků správně klasifikovaných jako negativní
jmenovatel: počet všech snímků ve skutečnosti patřících mezi negativní

F1-skóre (F1-score; F1-measure)

F1-skóre vyjadřuje úspěšnost modelu. Je definováno jako harmonický průměr přesnosti a senzitivity, a proto je vhodnou metrikou v případě, kdy jsou četnosti jednotlivých tříd nevyvážené.

$$F = 2 \cdot \frac{P \cdot R}{P + R}$$

4.4.5.2 Vícetřídní klasifikace

Následující přehled byl sestaven podle zdrojů [51, 52] V případě vícetřídní klasifikace je nutné sestavit matici záměn obsahující hodnoty pro všechny třídy. Následně se z této matice odvodí dílčí matice záměn pro binární klasifikaci každé třídy a metriky se vypočítají pro každou třídu zvlášť pomocí výše uvedených vzorců.

		Předpovědi			
		kat. A	kat. B	kat. C	kat. D
Skutečnost	kat. A	TP_A	E_{AB}	E_{AC}	E_{AD}
	kat. B	E_{BA}	TP_B	E_{BC}	E_{BD}
	kat. C	E_{CA}	E_{CD}	TP_C	E_{CD}
	kat. D	E_{DA}	E_{DB}	E_{CD}	TP_D

V této a následujících ukázkách je použito označení:

TP_s počet případů, které byly správně zařazeny do třídy s

E_{sp} počet případů, které patří do třídy s , ale byly chybně zařazeny do třídy p

Z metrik spočítaných pro jednotlivé třídy lze následně vypočítat metriky charakterizující model jako celek. Rozlišují se mikro a makro charakteristiky. Ve výpočtu makro charakteristik jsou všechny třídy zastoupeny stejným dílem. Každá makro charakteristika modelu je aritmetickým průměrem charakteristik jednotlivých tříd. Mikro charakteristiky naproti tomu zohledňují četnosti jednotlivých tříd, což je vhodné v případě, kdy jsou četnosti nevyvážené. V následujících vzorcích k označuje celkový počet tříd a i je index označující konkrétní třídu.

Celková správnost

$$A_{makro} = \frac{\sum_{i=1}^k A_i}{k}$$

$$A_{mikro} = \frac{\sum_{i=1}^k TP_i}{\sum_{i=1}^k TP_i + TN_i + FP_i + FN_i}$$

Přesnost

$$P_{makro} = \frac{\sum_{i=1}^k P_i}{k}$$

$$P_{mikro} = \frac{\sum_{i=1}^k TP_i}{\sum_{i=1}^k TP_i + FP_i}$$

Senzitivita

$$TPR_{makro} = \frac{\sum_{i=1}^k TPR_i}{k}$$

$$TPR_{mikro} = \frac{\sum_{i=1}^k TP_i}{\sum_{i=1}^k TP_i + FN_i}$$

Specificita

$$TNR_{makro} = \frac{\sum_{i=1}^k TNR_i}{k}$$

$$TNR_{mikro} = \frac{\sum_{i=1}^k TN_i}{\sum_{i=1}^k TP_i + FP_i}$$

F1-skóre

$$F1_{makro} = \frac{\sum_{i=1}^k F1_i}{k}$$

$$F1_{mikro} = 2 \cdot \frac{P_{mikro} \cdot TPR_{mikro}}{P_{mikro} + TPR_{mikro}}$$

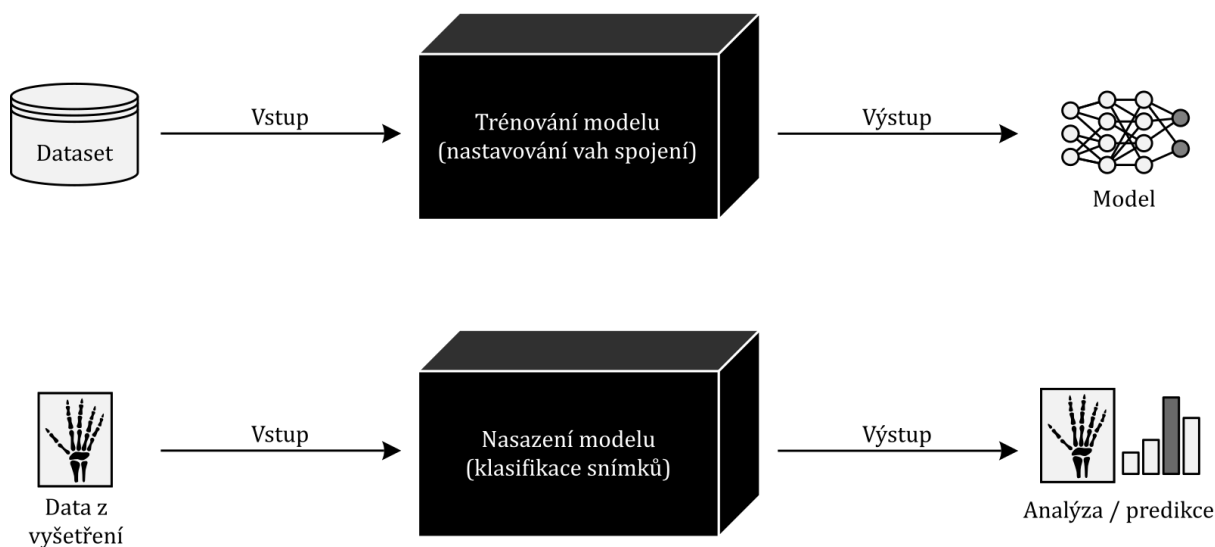
4.4.5.3 Problém černé skříňky

Pojem černá skříňka se v IT prostředí používá obecně pro libovolné zařízení nebo software, u kterého jsou zřejmé vstupy a výstupy, avšak naprosto chybí znalost o jeho vnitřním fungování. Jinak řečeno není znám způsob, jakým se od vstupů došlo k výstupům. S černou skříňkou známou z letectví nebo obecně z dopravy nemá tento pojem nic společného.

Opakem černé skříňky je bílá skříňka, někdy označovaná také jako průhledná nebo skleněná. Tedy systém, jehož fungování je zřetelné a přesně popsatelné. Systémy s pouze částečným vzhledem dovnitř se označují jako šedá skříňka. Druh skříňky záleží také na vztahu pozorovatele vůči zkoumanému systému.

Nedostatečný vzhled do činnosti algoritmu

Zatímco u klasických programů (Software 1.0) přinejmenším jejich autoři – programátoři a softwaroví architekti – mají (nebo alespoň mají šanci mít) o vnitřním fungování systému víceméně jasnou představu, u systémů na bázi hlubokého učení i samotným autorům vzhled většinou chybí. Obecně se na systémy využívající hluboké učení pohlíží jako na černou skříňku, a to jak na proces trénování, tak na nasazení již vytrénovaného modelu [22, 41].



Obr. 25: Hluboké učení jako černá skříňka

Zdroj: vlastní zpracování

Vytrénovaný model obsahuje tisíce parametrů reprezentujících nějaké statistické vlastnosti. Tyto hodnoty je obtížné interpretovat a tím pádem nelze plně porozumět způsobu, jakým se program například při určování diagnózy rozhoduje. Ve zdravotnictví jsou řešení fungující jako černá skříňka z mnoha pochopitelných důvodů nepřijatelná, a kromě toho by takový software ani neprošel certifikačním procesem [53]. Problém černé skříňky je jednou z největších překážek, kterou je pro nasazení hlubokého učení nutné překonat nejen ve zdravotnictví.

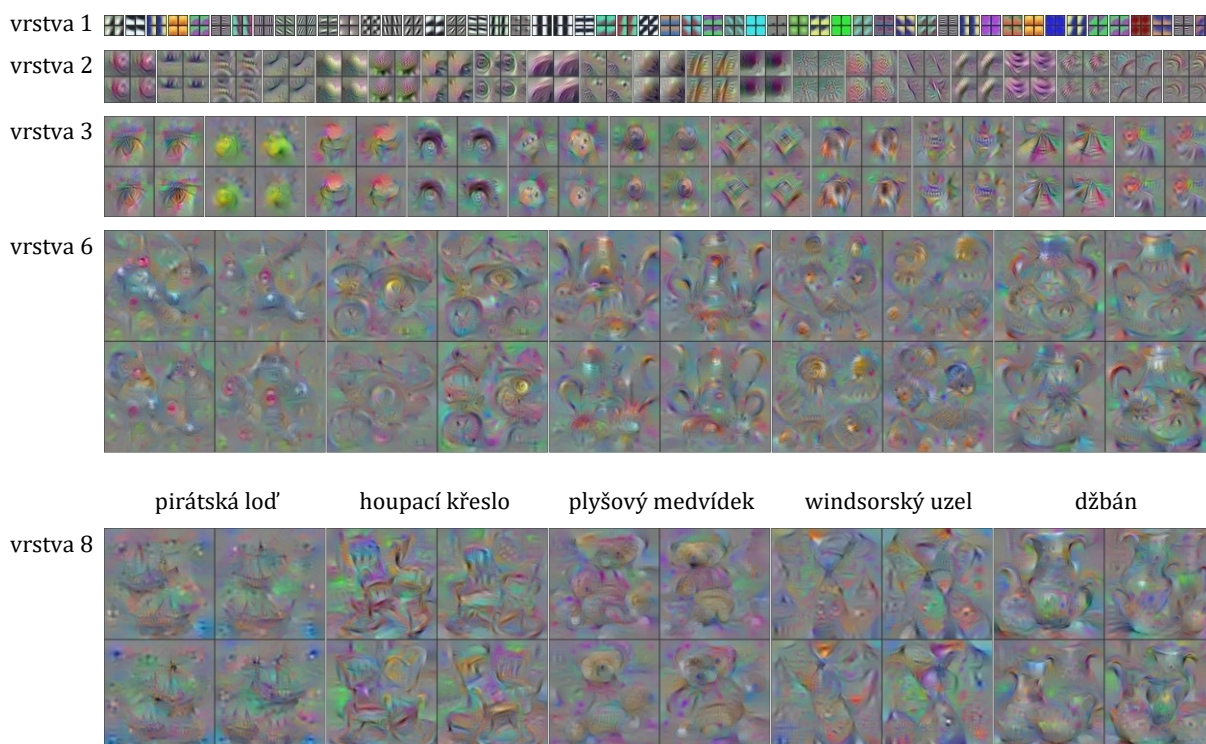
V případě neuronových sítí na zpracování obrazu bylo představeno několik způsobů, jak lze při současném stavu technologií s problémem černé skříňky alespoň částečně bojovat. Jsou to:

A) Důraz na co nejpodrobnější vyhodnocení modelu

Důkladné vyhodnocení modelu je základní způsob ke získání vhledu do jeho fungování. Často se ale opomíjí. RSNA uvádí několik způsobů vyhodnocení: doporučuje používat metriky uvedené v této práci i některé jiné, dále vyzkoušet model také na datech z jiného zdroje a také nějakým způsobem nabídnout hotový model veřejně k nezávislému prověření [49]. Zveřejnění může probíhat například po vzoru společnosti 16 Bit, která vyhrála soutěž RSNA 2017 Bone Age Challenge (popsáno v kapitole 6.1.5). Ta nabízí možnost model vyzkoušet na svém webu – veřejnost tak má možnost seznámit se s funkčností modelu, ačkoliv ten zůstává bezpečně uložen na serveru společnosti a nikdo si jej nemůže stáhnout [54].

B) Vyobrazení toho, jak je snímek reprezentován uvnitř neuronové sítě

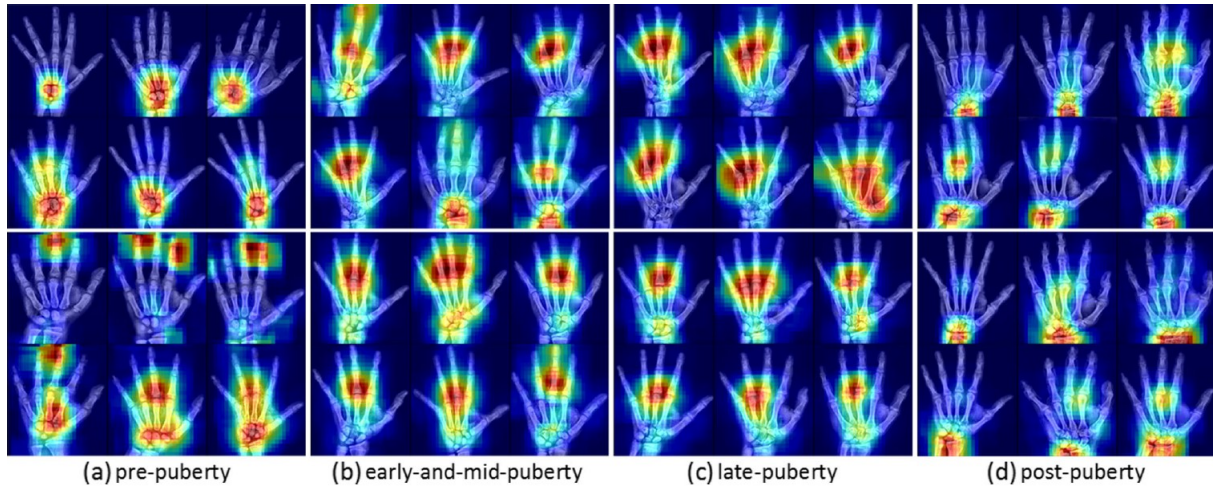
V anglické terminologii označováno pojmy feature visualization nebo též activation visualization. V obou případech jde v podstatě o znázornění map rysů (feature maps). Z vizualizovaných map je patrné, že první vrstvy sítě se zaměřují na jednotlivosti jako hrany, rohy nebo základní geometrické tvary a následující vrstvy je pak kombinují do větších a smyslupnějších celků [41, 55]. Princip ilustruje následující obrázek, relevantní příklad využívající medicínská data nebyl v době tvorby této práce k dispozici nebo nebyl dostatečně názorný [56].



Obr. 26: Reprezentace rysů obrázku uvnitř hluboké konvoluční neuronové sítě
Zdroj: převzato z [55]

C) Barevné zvýraznění oblastí snímku, podle kterých neuronová síť učinila rozhodnutí

V anglické terminologii označováno pojmy class activation maps (CAMs), saliency maps (mapy význačnosti) nebo méně odborně též jako heatmaps (teplotní mapy) – barevné zvýraznění těch částí vstupního snímku, podle kterých neuronová síť určila výstup. Zvýraznění použitou duhovou barevnou škálou připomíná obraz z termokamery [41, 57].

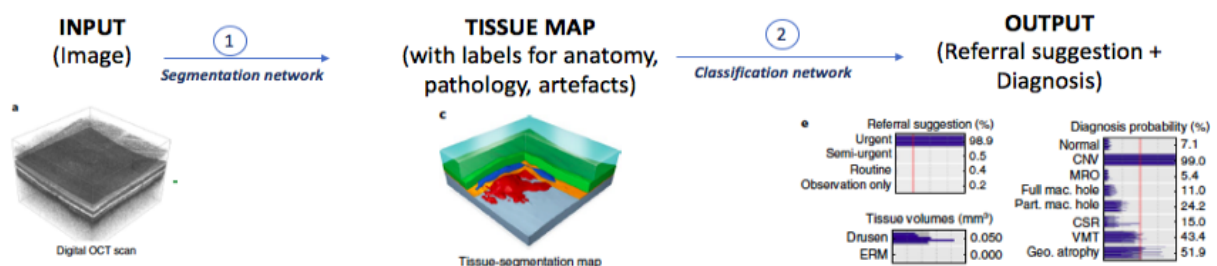


Obr. 27: Teplotní mapy jako součást úlohy určování kostního věku

Obrázek vznikl jako součást řešení RSNA Bone Age Challenge soutěžního týmu z Harvardovy univerzity. Teplotní mapy vyznačují ty oblasti, podle kterých neuronová síť snímek odlišila od zástupců dalších tříd. Horní dva řádky jsou snímky ženských rukou, dolní dva řádky mužských. Do pre-pubertální skupiny byly zařazeny snímky žen ve věku 2–7 let a mužů ve věku 3–9 let, do rané puberty ženy 7–13 let a muži 9–14 let, do pozdní puberty ženy 13–15 let a muži 14–16 let, do post pubertální skupiny ženy 15–18 let a muži 17–18 let (použitý dataset neobsahuje ruce starší 19 let). Zdroj: převzato z [58]

D) Rozdělení složitějších úloh na několik jednodušších a přidání mezivýstupů

Googlem vlastněná společnost DeepMind navrhl ve svém řešení pro diagnózu 3D skenů sítnice ještě další z možných cest. Úlohu výzkumníci rozdělili na dvě části. První neuronová síť měla za úkol pouze segmentaci, tedy rozpoznání struktur v obrazu. Teprve druhá síť rozpoznané struktury klasifikovala. Tím do celého procesu přibyl ověřitelný mezivýstup. Navíc segmentační i klasifikační síť se spouštějí v pěti instancích, takže celý proces od vstupních dat až k diagnóze proběhne pětikrát a tím pádem si uživatel může udělat představu, která místa jsou pro algoritmus jasně rozpoznatelná, a naopak se kterými má potíže [53].



Obr. 28: Schematické znázornění řešení na analýzu 3D skenu sítnice od společnosti DeepMind

Zdroj: převzato z [53]

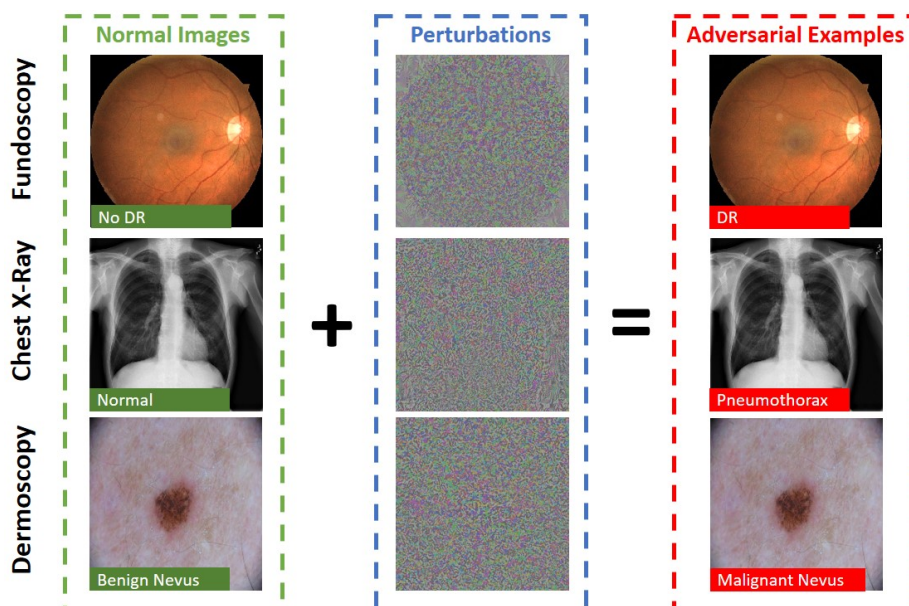
E) Zobrazení podobných případů

Další alternativu navrhli výzkumníci z Harvardovy univerzity. Jejich software pro klasifikaci krvácení na CT snímcích mozku své závěry podporuje zobrazením nejvíce podobných případů z trénovacího datasetu. Uživatel tak má možnost zjistit, na základě čeho program ke svým závěrům došel a zároveň jsou mu dány k dispozici podobné případy pro porovnání [58].

Všech pět výše uvedených variant vede k otázce, jak celý systém navrhnout tak, aby kontrola správnosti softwarového řešení nespotřebovala víc času a sil, než tradiční způsob vyšetření bez využití hlubokého učení.

Známé způsoby klamání

Problematika rozhodování neuronových sítí je zkoumána také z druhé strany, hledáním zranitelností. Mimo zdravotnickou oblastí byly popsány poměrně snadné způsoby oklamání sítě, např. pomocí tzv. protichůdných příkladů (adversarial examples) [59, 60] – vytrénované modely na bázi hlubokého učení se dají snadno zmást snímkem, do kterého se v malé míře přidá šum náhodný nebo založený na vnitřní reprezentaci obrazu uvnitř neuronové sítě. Dále například snímkem, který se se prolne či zkombinuje s jiným snímkem. Takto upravená vstupní data vedou algoritmus k jinému výsledku, přestože pro lidské vnímání se pozorovaný obraz zdá být téměř stejný [41, 60]. Vlivy protichůdných příkladů i jiných způsobů oklamání vytrénovaných modelů byly zkoumány a potvrzeny i v medicínské oblasti a to nejen na radiologických datech [61, 62].



Obr. 29: Oklamání vytrénovaných modelů pomocí protichůdných příkladů

Zdroj: převzato z [62]

Přímočaré pokusy o řešení problému prostě tím, že se do trénování zařadily i klamavé obrázky, se ukázaly jako nedostatečně účinné. Úspěšné vyřešení bude patrně vyžadovat změny v základních principech fungování současných neuronových sítí [63].

5 Implementace neuronových sítí a související problémy

Tato kapitola popisuje architektury neuronových sítí a softwarové technologie použité pro jejich implementaci v praktické části této práce. Informace jsou zasazeny do kontextu aktuálního vývoje a uvedeny jsou také známé problémy a nedostatky komplikující nasazení technologií, jako je výkon hardwaru a dostupná paměť, náročnost procesu trénování a rozlišení vstupních snímků. Možnostmi, jak některé nedostatky zmírnit, jsou metody přeneseného učení a rozšíření dat.

5.1 Architektury neuronových sítí

V oblasti konvolučních neuronových sítí je vývoj struktur úzce spojený se soutěží ILSVRC (ImageNet Large Scale Visual Recognition Challenge). Soutěž se konala v letech 2010 až 2017 v kategoriích klasifikace a detekce objektů v 2D obrazu. Účastníci soupeřili se novými strukturami sítí ve snaze dosáhnout co nevyšší přesnosti, v posledních třech ročnících dokonce vykazovaly algoritmy menší chybu než člověk. Síť byly trénovány na datasetu ImageNet s více než milionem obrázků v tisíci třídách, díky čemuž jsou výsledky z jednotlivých ročníků vzájemně porovnatelné a dataset slouží jako benchmark. Vítězné architektury sítí jsou označovány jako špičkové (*state of the art*) [41, 64].

Je běžné, že tyto architektury se následně rozšíří pro využití ve vědě i praxi, a některé níže popsané byly vyzkoušeny i pro praktickou část této. Obsah kapitoly je zpracován podle zdrojů [64–66]. Podrobnější popis je již nad rámec práce, ale zejména na webu [66] lze najít velmi přehledná schématická znázornění a detailní výčet odlišností a přínosů jednotlivých struktur.

5.1.1 VGG16

Struktura sítě VGG byla navržena výzkumníky z Oxfordu (VGG je zkratka pro Visual Geometry Group). Síť je evolucí předchozí velmi úspěšné struktury AlexNet a skončila na druhém místě v ILSVRC 2014. Ve srovnání s níže uvedenými je struktura velmi jednoduchá a přitom účinná, obsahuje pouze konvoluční vrstvy prokládané poolingovými a zakončené trojicí plně propojených. Daní za jednoduchost je pomalejší činnost a podstatně větší paměťová náročnost modelu. Objevují dvě verze sítě označované podle počtu vrstev jako VGG-16 a VGG-19. Verze VGG-16 má 138 357 544 trénovatelných parametrů, obě verze přijímají vstupní obrázky v rozlišení 224×224 .

5.1.2 Inception

Sítě typu Inception pocházejí z dílen Googlu a jsou nástupcem sítí typu NiN (Network in Network). První verze se jmenovala GoogLeNet a zvítězila v ILSVRC 2014. Inception byl původně název pro jednotlivé bloky, ze kterých byla síť složená – každý blok byl vlastně samostatnou jednoduchou konvoluční sítí. GoogLeNet měl celkem 22 vrstev a 5 milionů trénovatelných parametrů.

Další verze Inception-V2 a Inception-V3 byly představeny v roce 2015 a zvýšily velikost vstupního snímku na 299×299 px (z původních 244×244 px) a počet parametrů na 23 851 784 ve 159 vrstvách.

5.1.3 ResNet

ILSVRC 2015 vyhrál Microsoft se svou sítí ResNet. Tato síť využívá tzv. reziduální bloky, ve kterých vstup v jedné větvi projde dvojicí (v pozdějších implementacích trojicí) konvolučních vrstev a v druhé větvi projde nezměněn. Na výstupu bloku se obě větve spojí a obě hodnoty zkombinují. ResNet se běžně objevuje ve verzích s 34, 50, 101 nebo 152 vrstvami. Počet vrstev se uvádí přímo v názvu sítě. Verze Resnet-50 má 25 636 712 trénovatelných parametrů. Výchozí velikost vstupního obrázku je ve všech případech 224×224 px.

5.1.4 Inception-ResNet

Pozdějším zabudováním principů sítě ResNet do architektury Inception vznikly v roce 2017 architektury Inception-V4, Inception-Resnet-V1 a Inception-Resnet-V2. Poslední uvedená disponuje 55 873 736 trénovatelnými parametry v 572 vrstvách, výpočetní náročností odpovídá síti Inception-V4.

5.1.5 MobileNet

V roce 2017 Google představil jednoduché a nenáročné síť typu MobileNet s 53 vrstvami. Využívají podobný typ konvolucí jako Inception [67]. Existuje několik variant se vstupním rozlišením od 128 do 224 px a parametry od 0,47 milionů do 4,24 milionů.

5.2 Programovací jazyky a frameworky

Programovací jazyk Python je v oblasti datové vědy standardem a tím pádem je k němu dostupná i mnoho rozšiřujících knihoven a frameworků, které zpracování dat zjednodušují [33]. Právě pro Python vznikla také celá řada knihoven s funkcemi pro strojové a hluboké učení. Mezi ty nejrozšířenější patří TensorFlow (vyvíjeno Googlem), PyTorch (Facebook), CNTK (Microsoft), Theano, Keras, Caffé a celá řada dalších, méně známých. Oproti jiným převažují právě Tensorflow, Keras a PyTorch, protože jejich vývoj zastřešují velké týmy programátorů, mají početnou komunitu uživatelů a tím pádem lepší dokumentaci a méně bugů (chyb v programu) [33].

5.2.1 TensorFlow

Pro praktickou část této práce byl použit framework TensorFlow, který obsahuje potřebné matematické a obslužné funkcionality k implementaci a používání neuronových sítí. Byl vyvinut Googlem v oddělení Google Brain původně pro interní použití a 9. listopadu 2015 zveřejněn jako open source [33]. Výpočty jsou v TensorFlow reprezentovány jako grafy datových toků (*data flows*). Uzly grafu představují matematické operace a hrany data předávaná mezi uzly. Samotná

data jsou reprezentována pomocí tenzorů, což jsou z matematického pohledu zjednodušeně řečeno zobecnění pojmů vektor a matice (vektor je jednorozměrný tenzor, matice dvourozměrný) a z programátorského pohledu vícerozměrná pole [37]. TensorFlow dokáže být přínosný i v různých jiných odvětvích, používá se ale především pro datovou vědu a hlubokého učení [37].

TensorFlow umí podobně jako konkurenční knihovny využívat jak procesor, tak grafickou kartu, avšak pouze takovou, která je vybavená technologií Nvidia CUDA. Podpora pro OpenCL je zatím pouze experimentální [33].

Součástí TensorFlow je TensorBoard, nástroj umožňující ve webovém prohlížeči v reálném čase monitorovat graficky i číselně proces trénování neuronové sítě.

5.3 Rychlost vývoje technologií

Strojové učení a neuronové sítě existují již několik desítek let, v posledních přibližně patnácti letech postupně seskupením a zobecněním funkcionalit vznikly softwarové knihovny a frameworky zaměřené na usnadnění implementace neuronových sítí i obecně dalších matematických výpočtů, jejich efektivní reprezentaci a vykonání výpočtů na HW. Do této skupiny softwaru patří např. frameworky Torch (zveřejněn 2002), OpenNN (2003) nebo Theano (2007). Moderním přírůstkem pak bylo v roce 2015 také TensorFlow a univerzální nadstavba Keras.

Vývoj probíhá v poslední letech velmi uspěchaným tempem. Například verze TensorFlow vydaná v roce 2015 byla nedokončená a ve vývoji, verze 1.0 vyšla až roku 2017. V roce 2019 pak vyšla verze 2.0, která přinesla mnoho změn v syntaxi kódu, provázání s Kerasem a podporu nových verzí jazyka Python – to vše dohromady znamená nutnost drtivou většinu stávajícího kódu od základů přepsat, což tvorbu softwaru značně brzdí a komplikuje.

Navíc programování s pomocí výše zmíněných knihoven stále probíhá na relativně nízké úrovni a vyžaduje množství tzv. boilerplate kódu (sekce opakujícího se kódu, které je nutné v programu mít pro dodržení syntaxe a formy, ačkoliv nijak nesouvisí s primární funkcionalitou programu). Kvůli tomu je proces programování zdlouhavý a náchylný ke vzniku obtížně dohledatelných bugů [68].

Od problému boilerplate kódu se uživatelé snaží odstítnit Keras a jeho další nástavby jako Talos č AutoKeras [69]. Tato řešení ukazují současný trend vývoje – technologie označované jako automated machine learning (Auto ML) [70], které uživatele částečně nebo zcela odstiňují od zdrojového kódu i od rutinních činností, které s vývojem neuronových sítí běžně souvisejí. Odpadá například zdlouhavé hledání nejvhodnější architektury neuronové sítě, vhodných parametrů trénování a vyhodnocení a vyladění vytrénovaného modelu. Avšak ani AutoML neznamena automatizovanou datovou vědu – v této úloze hraje software a hardware stále pouze roli nástrojů [71]. Daní za pohodlí a úsporu času, kterou AutoML nabízí, je omezená možnost kontroly nad celým procesem a některými jeho aspekty [7].

5.4 Hardware

5.4.1 Použití GPU, CPU a TPU

Matematické operace, na kterých je trénování neuronových sítí založené, zahrnují práci s vektory a maticemi. Na stejných základech funguje i počítačová 3D grafika a grafické karty (GPU) jsou na výpočty s maticemi a vektory speciálně přizpůsobeny, takže oproti procesoru (CPU) zvládají tento druh operací vykonávat podstatně rychleji. Hardware i software grafických karet je přizpůsoben tak, aby na něm bylo možné provádět i negrafické výpočty, tento způsob jejich využití se označuje jako GPGPU (general-purpose computing on graphics processing units). Výpočty na GPU mají teoreticky výhodu v možnosti masivní paralelizace, protože disponují řádově více výpočetními jádry než CPU. Zatímco CPU má řádově jednotky jader, GPU jich má řádově stovky až tisíce. Tato jádra GPU se označují pojmem *shader* v terminologii výrobce Nvidia, nebo jako *stream processor* v terminologii AMD. Naopak nevýhodou je, že GPU mají značně omezenou paměť (řádově jednotky či menší desítky gigabytů), zatímco při využití CPU není technicky problém využít stovky gigabytů RAM [72]. Pro překonání paměťového omezení GPU lze trénovat paralelně na více GPU současně, k čemuž je ovšem potřeba podpora paralelismu v použitém frameworku.

Podmínkou a překážkou pro trénování neuronových sítí na GPU je, že kromě hardwaru samotného je nutná také softwarová mezivrstva, která tvoří rozhraní mezi použitým frameworkem pro hluboké učení a ovladači GPU. Výrobce Nvidia nabízí pro implementaci GPGPU vlastní technologii CUDA, konkurenční AMD naopak otevřený standard OpenCL. Avšak pokud nějaký framework pro hluboké učení podporuje běh na GPU, pak zpravidla pouze pomocí CUDA. OpenCL podporováno není, nebo je pouze výhledově v plánu do dalších let, což výrobce AMD staví mimo hru. Naopak Nvidia se hlubokému učení intenzivně věnuje a průběžně rozšiřuje svou nabídku hardwaru, softwaru i doprovodných služeb k těmto účelům. Od roku 2019 Nvidia nabízí softwarový balíček Clara specializovaný na trénování neuronových sítí pro medicínské účely. Poradí si se sběrem a anotováním dat přímo ve formátu DICOM a především pak nabízí paralelní i federativní trénování na hardwaru Nvidia [73].

Kromě CPU a GPU existuje ještě třetí možnost – TPU (tensor processing unit), hardwarová komponenta speciálně navržená pro trénování neuronových sítí. TPU bylo vyvinuto přímo Googlem pro použití s jeho frameworkem TensorFlow. Od roku 2015 byly dosud představeny již 4 generace TPU. Třetí generace TPU nabízela v součtu 8 výpočetních jader a 32 GB paměti [74]. I přesto ale aktuální vývoj v oboru působí dojmem, že hardwaru pro hluboké učení vévodí superpočítače s GPU, případně CPU [7].

5.4.2 Problém omezené paměti GPU

Trénování velkých a robustních modelů hlubokého učení je v současné době velmi časově náročné a naráží na omezení spočívající ve velikosti paměti grafických karet. Pro ilustraci: špičková Nvidia GeForce RTX 2080 Ti, která byla k tvorbě této diplomové práce použita a která

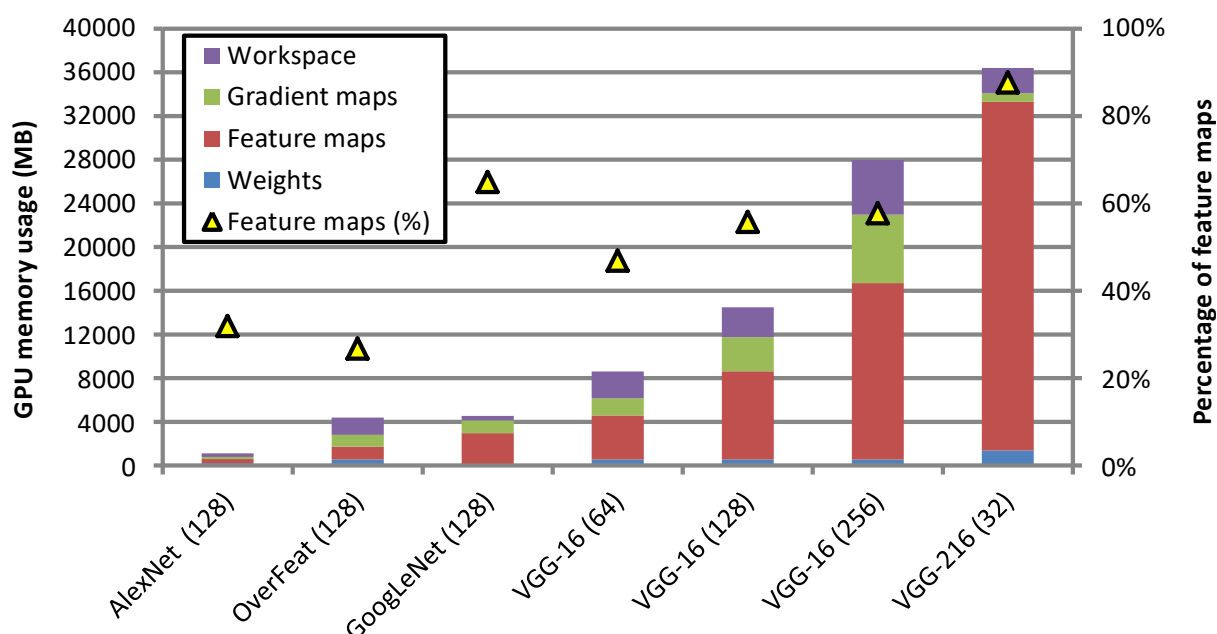
v době tvorby této práce byla nejmodernější a nejvýkonnější herní grafickou kartou v nabídce Nvidie, disponuje 11 GB paměti a prodává se za ceny okolo 1200 USD [75]. Mimo herní sektor disponovala největší pamětí Nvidia Quadro RTX 8000, která nabízí 48 GB paměti s cenou okolo 5500 USD [76]. Ačkoliv se i zmíněných 11 GB může zdát až zbytečně mnoho, pro trénování hlubokých neuronových sítí je taková paměť nezbytná a mnoha případech také čím dál více nedostačující. Velikost požadované paměti je závislá na použité architektuře sítě a na nastavené velikosti dávky (počtu snímků, na které do neuronové sítě v jednom trénovacím kroku vstoupí).

Kupříkladu síť AlexNet, tvořená 5 konvolučními vrstvami a 2 plně propojenými, vyžaduje při velikosti dávky 128 alespoň 1,1 GB paměti. Větší síť VGG-16 spotřebuje při velikosti dávky 128 už zhruba 14 GB paměti. A to je ještě poměrně málo v porovnání se sítí ResNet, která je dostupná v implementacích s 50, 101 a 152 vrstvami.

Data uložená během trénování v paměti GPU lze rozdělit do 4 kategorií: [77]

- 1) parametry modelu (váhy)
- 2) mapy rysů (feature maps) – mezivýsledky vznikající v konvolučních vrstvách při dopředné fázi běhu trénovacího algoritmu
- 3) gradientní mapy (gradient maps) – mezivýsledky vznikající při zpětném chodu
- 4) pracovní prostředí (workspace) – odkládací prostor čili mezipaměť pro dočasné proměnné (matice)

Největší část paměti zabírají právě mapy rysů. Pro velkých modelů, jako je například VGG-16, to může být i přes 50–80 % celkové dostupné paměti, jak ukazuje následující graf.



Obr. 30: Využití paměti GPU při trénování modelu

Levá osa znázorňuje využití paměti GPU na různé účely v absolutních číslech, pravá osa ukazuje podíl alokované paměti spotřebované na mapy rysů. Zdroj: převzato z [77]

Trénování modelů jako je VGG-16 nebo větších na GPU s menší pamětí, než kolik by bylo optimální, vede nutně k rozhodování mezi několika možnostmi kompromisu: [77]

- Zmenšit velikost dávky – může ovlivnit přesnost a rychlost trénování. U velkých modelů typicky dochází k situaci, že přínosy pokročilé architektury jsou znehodnoceny práce nemožností trénovat s velkými dávkami.
- Distribuovat trénování na více GPU – kromě toho, že vyžaduje investici do další GPU, je to náročný proces sám o sobě a přímo souvisí s vnitřní architekturou a principy fungování frameworku, ve kterém je neuronová síť implementována
- Zjednodušit neuronovou síť (model) odebráním vrstev
- Počkat na další generaci GPU s větší pamětí

V článku [77], z něhož informace v této kapitole pocházejí, výzkumníci pracující pro Nvidii poukázali, že současné frameworky pro hluboké učení mají vážné nedostatky a omezení co se týče alokování a správy paměti, a představili svůj vlastní, nový způsob správy paměti, který dokáže nedostatek místa na GPU obejít. V podstatě jde o softwarovou mezivrstvu, co přístup k paměti (slovy autorů) virtualizuje, takže programy na trénování modelů mohou využívat kromě paměti GPU nově také RAM, která je porovnání s pamětí GPU levná a dostatečně kapacitní. Software byl zacílen v první řadě na mapy rysů, jejichž přenesení na RAM přineslo největší úsporu paměti, řádově až 90 % v závislosti na architektuře sítě a nastavené velikosti dávky. Celé řešení autoři označili názvem *Virtualized Deep Neural Network* (vDNN). Ačkoliv byl článek publikován již v roce 2016, dosud se veřejně dostupná oficiální implementace technologie vDNN neobjevila.

5.5 Přenesené učení

Jedním ze způsobů, jak obejít výpočetní náročnost trénování, vypořádat se s nedostatkem dat a také jak předejít přeučení modelu, je metoda zvaná přenesené učení (transfer learning) [48]. V takovém případě se model už vytrénovaný na jednom datasetu upraví pro použití na jiném datasetu. Úspěšnost přeneseného učení závisí na typu řešené úlohy a podobnosti datasetů [7].

Typicky to znamená použít model náročně vytrénovaný na opravdu velkém datasetu, který díky tomu disponuje dobrou schopností generalizace a dokáže extrahovat ze snímků obecné, nízkourovňové charakteristiky. Vytrénované váhy nižších vrstev neuronové sítě se ponechají, zatímco ve vyšších vrstvách (které jsou zodpovědné za zpracování větších celků obrazu a za výsledné zařazení do třídy) se váhy odstraní a následně dotrénují znovu již na cílovém datasetu. Existují tři základní strategie, jak k trénování modelu přistoupit: [7, 41, 78]

1) Vytrénovat celý model od základu a přenesené učení neimplementovat

Tato možnost dává svobodnou volbu architektury, na druhou stranu bude vyžadovat velký a kvalitní dataset, extrémně výkonný hardware a mnoho času.

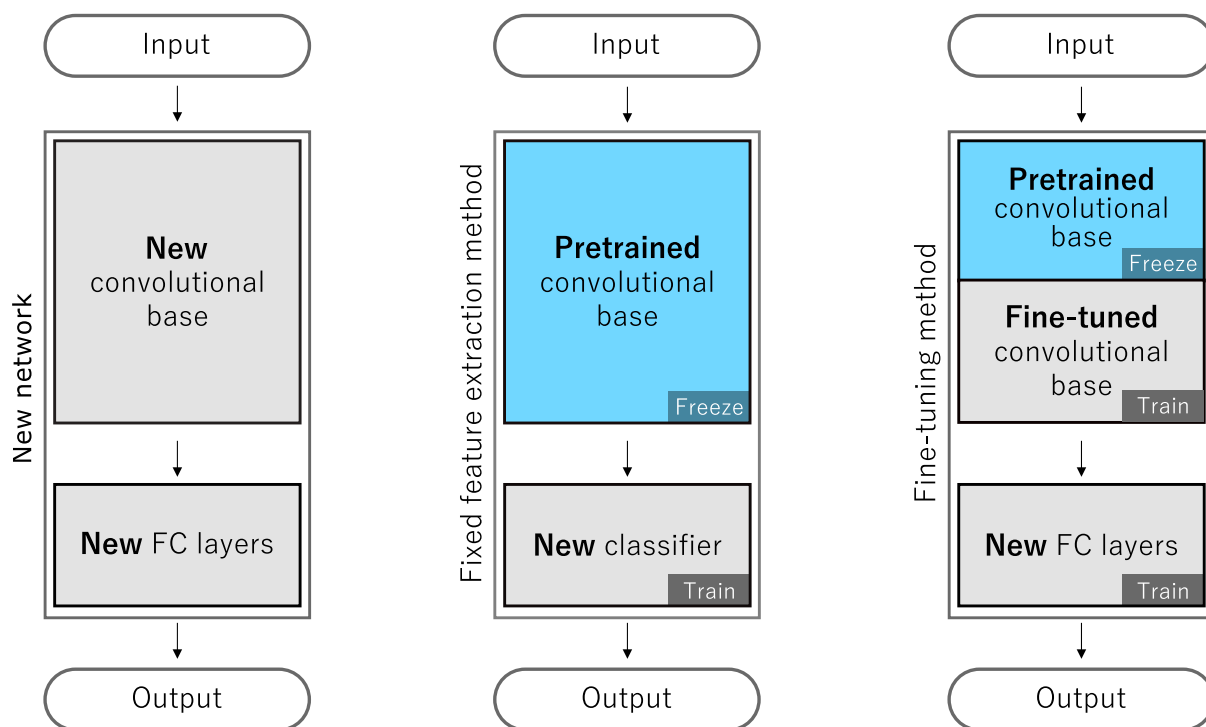
2) Přenesené učení formou extrakce rysů (feature extraction)

Váhy uložené v nižších konvolučních a poolingových vrstvách předtrénovaného modelu jsou ponechány v původním stavu (často se i v odborné terminologii používá pojem „zmražený“) – tato část modelu se označuje jako modul pro extrakci rysů obrázku (image feature extraction module). Naopak horní, plně propojené vrstvy, zodpovědné za zařazení snímku do výsledné třídy, jsou odstraněny, nahrazeny novými a vytrénovány znovu. Trénování je v tomto případě poměrně rychlé a hardwarově méně náročné.

3) Přenesené učení formou ladění (fine-tuning)

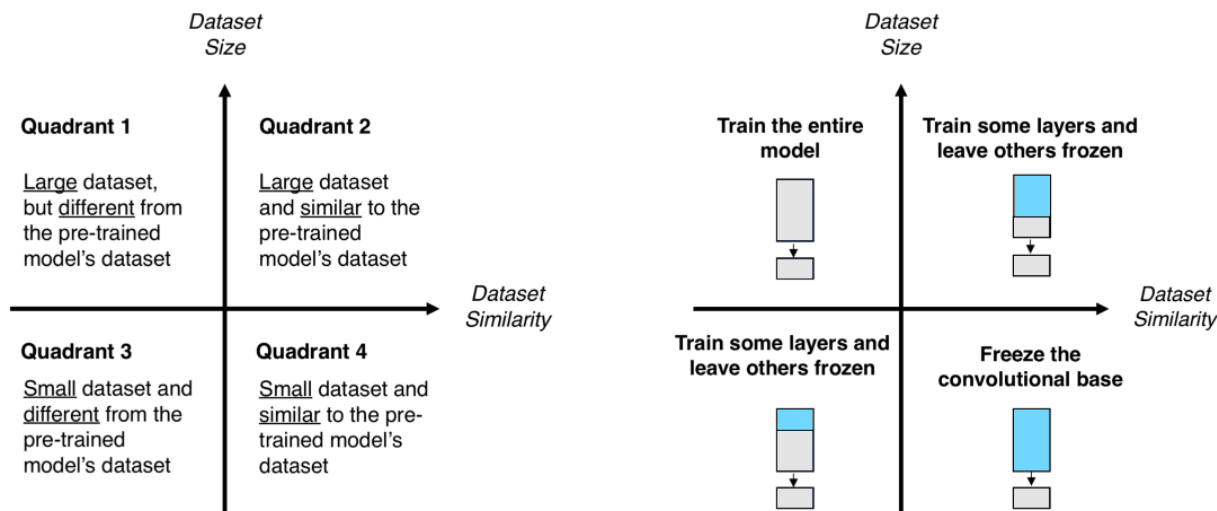
Při této strategii se zmrazí pouze nejnižší konvoluční vrstvy, další konvoluční vrstvy jsou odemčeny k tzv. ladění – předtrénované hodnoty v nich uložené se při trénování upravují, aby se přizpůsobily novým datům. Horní, plně propojené vrstvy jsou stejně jako v minulém případě odstraněny a vytrénovány znovu. Alternativně i nejnižší vrstvy mohou být odemčeny k ladění. Trénování je zpravidla delší než v předešlém případě a probíhá s velmi nízkým koeficientem učení, aby se pracně předtrénované váhy spojení neuronů ihned nevytratily, ale pouze postupně a zvolna ladily a přizpůsobovaly novým datům.

Trojici strategií z předchozí strany a důvody, proč si nějakou z nich vybrat, přehledně ilustrují následující dvě schémata.



Obr. 31: Strategie přeneseného učení

Zdroj: převzato z [41] a upraveno



Obr. 32: Jak vybrat vhodnou strategii přeneseného učení

Zdroj: převzato z [78]

Přenesené učení přináší ještě další zásadní výhodu – menší hardwarovou (i finanční) náročnost procesu trénování. Ten je totiž obecně velmi zdlouhavý, a kromě již zmiňovaného problému s pamětí GPU vyžaduje výkonný a drahý hardware. Například zmiňovaná síť VGG byla pro účely soutěže ILSVRC trénována na datasetu ImageNet na 4 grafických kartách po dobu dvou týdnů. Vítězná síť GoogleNet byla do konvergujícího stavu vytrénována během necelého týdne, omezující byla právě paměť GPU. O rok později trénování sítě ResNet trvalo tři týdny na počítači vybaveném 8 grafickými kartami [64].

A právě tyto pracně a pečlivě vytrénované špičkové modely osvědčených architektur sítí jsou pro usnadnění implementace přeneseného učení zdarma k dispozici ve většině nejrozšířenějších frameworků jako TensorFlow, Keras, PyTorch a dalších. Stačí si zvolit požadovanou architekturu a strategii přenosu učení. Čas potřebný k trénování lze v závislosti na situaci zkrátit řádově na desítky či jednotky hodin, v určitých případech i na jednotky minut [41, 78].

Přenesené učení bylo úspěšně vyzkoušeno také v medicíně, a to jak s modely předtrénovanými na medicínských datech, tak s modely využívajícími právě zmiňovaný dataset ImageNet. Konkrétní výzkumy, kterými se praktická část této práce inspirovala, jsou představeny v kapitole 6.1.3 Výzkumy související s tématem této práce.

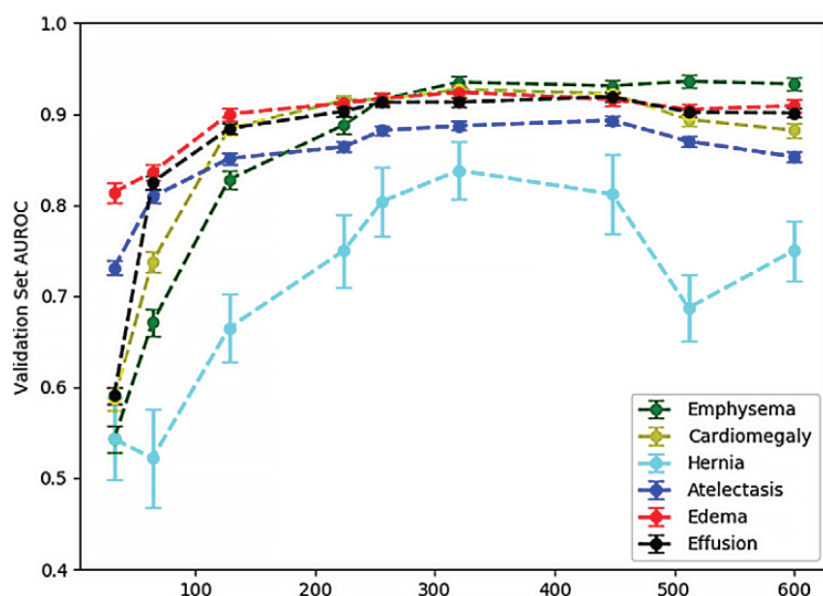
5.6 Problém nízkého rozlišení vstupních snímků

Společnou vlastností všech současných veřejně dostupných a oblíbených architektur neuronových sítí je, že jsou navrženy na poměrně nízké rozlišení vstupních snímků. Typicky může jít o hodnoty v rozmezí od 32×32 px až do 299×299 px. I dostupné modely na přenesené učení předtrénované na datasetu ImageNet využívají rozlišení snímků v uvedených rozmezích [48, 79]. Nízké rozlišení se používá z řady důvodů: [79]

- Trénování neuronových sítí s miliony parametrů je velmi náročné na výkon hardwaru a čas potřebný k vytrénování, jak už bylo popsáno v přechozím textu.
- Pro základní úlohy, jako rozpoznání objektů ve snímku, jsou nízká rozlišení naprosto dostačující a přinášejí lepší schopnost generalizace, a tedy menší riziko přeučení.
- Snímky v nižším rozlišení jsou jednodušší na skladování, manipulaci a distribuci. Mnoho skutečně velkých datasetů je k dispozici pouze ve sníženém rozlišení.

Extrémně nízká rozlišení vedou ke ztrátám informací nutných k diagnostice [48]. Pro mnoho ne-medicínských úkolů takové hodnoty rozlišení snímků mohou stačit. V případě radiologických dat výzkumy z minulých let ukazují, že i při takto nízkých rozlišeních lze dosáhnout až překvapivě zajímavých výsledků. V současné době, s narůstajícím výkonem hardwaru a novými možnostmi distribuovaných výpočtů, se postupně začínají prosazovat i snímky ve vyšším rozlišení [80].

Několik výzkumů se již zabývalo také vlivem rozlišení snímku na přesnost diagnostiky. Článek *The Effect of Image Resolution on Deep Learning in Radiography* [80] publikovaný počátkem roku 2020 představil výsledky studie provedené s datasetem NIH Chest X-ray (obsahujícím přes 100 000 snímků hrudníku) a sítí architektury ResNet. Řešenou úlohou byla pro jednoduchost binární klasifikace, tedy možnost *ano/ne* pro jednotlivé diagnózy. Autoři zkoušeli k trénování modelů použít mnoho velikostí snímku v rozmezí od 32×32 px až do 600×600 px. Nejlepších a zhruba srovnatelných výsledků dosáhli s čtverci o straně 256, 320 a 448 px.



Obr. 33: Vliv rozlišení snímku na přesnost klasifikace diagnóz na snímcích hrudníku
Délky stran čtverce jsou v pořadí: 32, 64, 128, 224, 256, 320, 448, 512 a 600 pixelů.
Zdroj: převzato z [80]

Vyšší rozlišení snímků s sebou přináší taky nesrovnatelně vyšší nároky na hardware. Nizozemští vědci ze sdružení SURF při analýze RTG snímků hrudníku použili upravenou architekturu ResNet-59 a dataset NIH Chest X-ray zpracovávali v rozlišení 1024×1024 px. Dosáhli dobrých výsledků, ale potýkali se s přeúčením. Při trénování se kvůli obrovským nárokům na paměť odchýlili od GPU a díky spolupráci s firmou Intel místo toho trénovali na superpočítačích disponujících řádově stovkami jader CPU Intel Xeon a stovkami gigabytů RAM [72, 81].

5.7 Problém čtvercového formátu vstupních snímků

Kvůli tomu, že konvoluční operace využívají čtvercové matice, jsou vstupní rozměry snímků ve všech architekturách neuronových sítí omezeny na čtvercový formát, tedy poměr stran 1:1. Framework TensorFlow ve výchozím nastavení každý vstupní snímek nejdříve automaticky převede na tento formát. Ačkoliv existuje několik možných způsobů úpravy formátu, TensorFlow jednoduše během převzorkování a interpolace na požadované vstupní rozlišení snímek na jeho kratší straně roztáhne (resp. na delší straně stlačí) tak, aby se čtverec obrazem vyplnil [82].

Takový způsob má své klady i zápory. Na jedné straně se v roztaženém rozměru může uchovat víc informací, kterých je vlivem nízkého rozlišení snímků už tak poměrně málo. Podobný princip se využíval už před desítkami let ve filmovém průmyslu. Při natáčení širokoúhlých filmů se používaly anamorfní objektivy, které obraz roztáhly tak, aby vyplnil celou plochu téměř čtvercového políčka na filmovém pásu. Díky tomu měl obraz ve svislém rozměru vyšší rozlišení [83]. V oblasti neuronových sítí je natažení jednou z běžných úprav a lze jím mimo jiné dosáhnout větší variability vstupních dat, jak detailněji popisuje následující kapitola.

Na druhé straně ovšem může natažení na čtverec způsobovat problém u medicínských dat. Například RTG snímky hrudníku zpravidla čtvercový či téměř čtvercový formát mají, snímky rukou a jiných anatomických oblastí ale ne. Při jednom z experimentů, který předcházela praktické části této práce, bylo zjištěno, že úprava na čtvercový formát natažením může v určitých případech mít nežádoucí dopad na výsledky klasifikace. Model vytrénovaný na datasetu OAI (představený v kapitole 6.1.2) měl za cíl pouze roztřídit snímky podle toho, co zobrazují – ruku, obě ruce, pánev, koleno, obě kolena nebo dolní končetiny. Snímky obou kolen mají formát nejčastěji 4:3, snímky dolních končetin mají poměr zhruba 5:18 a při převodu formátu výchozím způsobem, tedy natažením, dochází k výrazné deformaci. Výsledkem bylo, že vytrénovaný model si v mnoha případech pletl snímky kolen se snímky dolních končetin, protože vlivem až příliš velké deformace vypadaly podobně, jak dokládá obrázek na následující straně. Z toho plyne, že způsobům změny poměru stran bude vhodné v praktické části práce věnovat pozornost a zjistit, zda v případě, kdy dataset tvoří pouze snímky rukou, natažení má nebo nemá vliv.

Kromě natažení, které je v TensorFlow výchozí a jedinou možností, lze formát v závislosti na situaci upravit také ořezáním nebo vyplněním prázdného místa neutrální barvou, v případě RTG snímků například černou.



Obr. 34: Nežádoucí podobnost obrazu vzniklá příliš velkou mírou deformace

Na krajích původní snímky, uprostřed převedené na čtverec natažením. Uvedený snímek kolen byl vytrénovaným modelem nesprávně označen jako snímek dolních končetin. Použitá síť architektury Inception-V3 s využitím přeneseného učení formou extrakce rysů (stejné řešení, jaké bylo využito v praktické části práce v kapitole 7.3.3) přitom v tomto případě dosáhla celkové správnosti přes 99 %.

Zdroj: vlastní zpracování s využitím snímků z datasetu OAI [84]

5.8 Rozšíření dat

Smysluplné trénování neuronových sítí vyžaduje obrovské množství dat. Čím více různých příkladů je síti poskytnuto, tím lépe se naučí rozpoznávat obecněji platné příznaky diagnózy [44]. Navíc, pokud není dataset dostatečně rozmanitý, může při trénování docházet k přeučení. Kromě toho by dataset měl být také vyvážený, tedy zástupci všech tříd by v něm měly být zastoupeni přibližně ve stejném poměru. Modely trénované na nevyvážených datasetech totiž mají sklon při předpovědích tíhnout k nejpočetnější třídě [48].

Tato omezení datasetů jsou částečně řešitelná technikami označovanými jako *data augmentation* (počeštěně *rozšíření dat*, *zvětšení dat* či *augmentace dat*) nebo v případě obrazových dat *image augmentation*. Je to souhrnné označení pro úpravy, kterými lze zvýšit počet a rozmanitost dat [44, 48], což v případě obrázků znamená tvorbu odvozenin pomocí geometrických transformací, úprav jasu, kontrastu a barevnosti, filtrace (rozmazání, doostření), ořezání, atp. Některé výzkumné týmy prověřovaly namísto transformací i možnosti úpravy a tvorby nových snímků pomocí neuronových sítí generativního typu (generative adversarial networks; GAN) a to i v oboru medicínských dat [48].

Podobně jako se princip fungování neuronových sítí přirovnává k lidskému mozku, přirovnává se k němu i proces rozšíření dat [44, 48]. Tak, jako si lidská představitost dokáže na základě podobností a předešlých zkušeností vyobrazit mnoho situací nových, snaží se tyto techniky vyobrazit nové variace dat. Neuronová síť by se pak měla lépe naučit obecné charakteristiky zobrazovaných tkání a struktur, potažmo obecné charakteristiky diagnóz, a neupínat se tolik na rysy konkrétních snímků [44, 48].

Ačkoliv toto řešení dokáže uměle zvýšit četnost a rozmanitost datasetu, nic se nevyrovná dostatečnému množství skutečných a dobře anotovaných snímků [44].

5.8.1 Bezpečnost a způsoby provedení

Veškeré druhy úprav je potřeba hodnotit z pohledu „bezpečnosti“ jejich použití. Bezpečností je v tomto případě míněno, zda a do jaké míry z obrázku bude i po transformaci rozpoznatelné to,

co zobrazoval původně. Jinak řečeno, zda i nadále bude odpovídat svému původnímu popisku (anotaci) [48]. Bezpečnost transformace se odvíjí od daného kontextu a záleží na kombinaci:

- druhu transformace
- míry transformace (míry zkreslení)
- charakteru dat
- řešené úlohy

Shorten a Khoshgoftaar [48] jako jednoduchý příklad transformace, která nezachovává popisek, a tedy není bezpečná, uvádějí otočení a překlopení u datasetů zaměřených na rozpoznání písmen a číslic, kdy velká míra transformací může vést k zaměnitelnosti číslic 6 a 9.

Obecně pak platí:

- nedoporučuje se transformacemi vytvářet stavy, které by nemohly nastat v realitě [44]
- při vyšší míře zkreslení je nebezpečná každá transformace [48]
- bezpečnější jsou ty transformace, které jsou v matematické terminologii označovány jako rigidní, zachovávají totiž klíčové vlastnosti obrazu [44]
- všechny odvozeniny by měly být zařazeny v té části datasetu (trénovací, validační nebo testovací), ve které se nachází původní snímek – jako prevence před přeučení [48]

Rigidní transformace je kombinací otočení a posunutí a patří mezi afinní transformace (spolu se škálováním, zkosením, zrcadlením a transformacemi vzniklými jejich složením) [8]. Technicky vzato ovšem původní obrazovou informaci zachovávají pouze ty transformace, které mění jenom souřadnice pixelů, tedy otočení o násobky 90 stupňů, posunutí a ořezání. U všech ostatních transformací dochází k převzorkování obrazu, přesněji k dopočítávání nových pixelů pomocí interpolačních algoritmů. V závislosti na rozlišení snímku, míře transformace a druhu interpolačního algoritmu dochází ke zkreslení původní obrazové informace a mohou vznikat nové, nežádoucí obrazce a struktury. Kromě toho může docházet i k částečné ztrátě ostrosti [8].

Pokud je na jeden obrázek použito více transformací, roli hraje také způsob, jakým je jejich složení implementováno. Zda se transformace provedou odděleně jedna po druhé, nebo najednou v rámci tzv. jednotné transformace (*unified transformation*). Při provádění odděleně interpolace zkreslí obraz během každé transformace, zatímco při použití jednotné transformace dojde pouze k jediné interpolaci a výsledek bude tím pádem přesnější [8]. V praxi ale často záleží na tom, jaké možnosti poskytuje grafický software a programátorské knihovny a frameworky, které jsou k práci použity [48]. V profesionálních grafických editorech nechybí možnost vybrat si výchozí interpolační algoritmus pro každý druh transformace zvlášť. Editory ale nejsou uzpůsobené pro dávkové automatizované zpracování celých datasetů. V programovacích knihovnách a frameworkcích pro jsou naproti tomu zase jiná omezení – mají zpravidla úzkou nabídku dostupných druhů transformací i interpolačních metod a trpí také dalšími problémy, které nastiňuje praktická část této práce.

5.8.2 Přehled druhů transformací

Zpracován podle zdrojů [44, 48].

- **Posunutí (translace)**

Posunutí je implementačně i výpočetně jednou z nejjednodušších transformací. Jeho využitím se neuronová síť může lépe zaměřit na obecné rysy zobrazených tkání.

- **Otočení (rotace)**

Otočení může napodobit pacienty, kteří byli při snímání o malý úhel natočeni. Zpravidla se obraz otáčí pouze náhodný úhel v předem zadaném rozmezí [44].

- **Škálování (změna měřítka; zoom)**

Stejně anatomické struktury se na různých snímcích různých pacientů prezentují ve více velikostech, což lze (do určité míry) nasimulovat touto jednoduchou úpravou.

- **Překlopení; zrcadlení (flip; mirror)**

Je možné využít u anatomických struktur, které jsou párové nebo symetrické. Překlopením se předchází situacím, kdy by neuronová síť upřednostňovala rysy nalezené na jedné straně orgánu, přestože postižená by byla druhá strana orgánu. Překlápět obrázek lze v závislosti na okolnostech podle vodorovné i svislé osy [44].

- **Natáhnutí (změna poměru stran)**

V určitých případech a bezpečně míře natáhnutí obrázku přidává potřebou variaci. V určitých případech ovšem může být deformace netáhnutím nežádoucí, jak ukázala předchozí kapitola.

- **Naklonění**

Funguje podobně jako natáhnutí, ale s tím rozdílem, že horní část obrázku se natahuje jedním směrem a spodní část obrázku směrem opačným.

- **Pružná (elastická) deformace**

Předchozích 5 technik jsou příklady tzv. lineárních transformací. Pružná deformace pokrývá obraz tak, jako by byl z gumy nebo látky. Avšak je potřeba ji používat jen v rozumné míře se zvláštním ohledem na bezpečnost.

- **Ořezání**

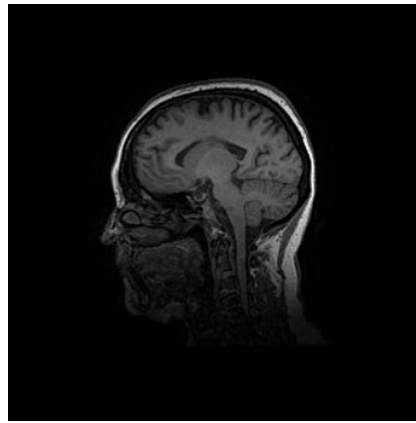
I prostým ubráním volného místa například po stranách se oblast zájmu ve snímku posune na jiné souřadnice. Pro člověka se snímek nemění, pro neuronovou síť jde o nový příklad.

- **Filtry (doostření, rozostření)**

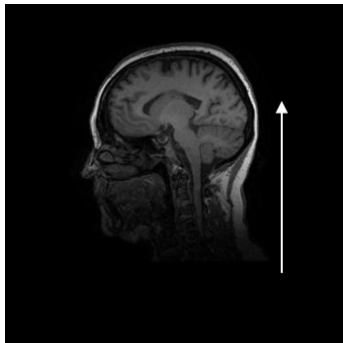
Úpravy ostrosti jsou dobrý způsob, jak ze snímků vytáhnout více detailů, případně jak trénovaný model udělat odolnější vůči pohybové neostrosti.

- **Úpravy barevnosti**

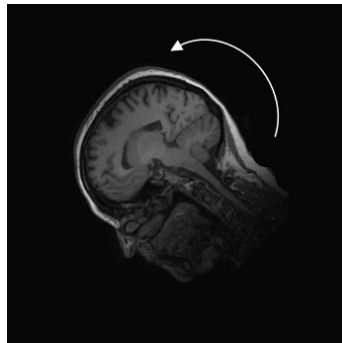
V nejjednodušším případě obnáší náhodné úpravy jasu a kontrastu v předem definovaném rozmezí (např. $\pm 15\%$), u medicínských snímků tyto hodnoty už ve výchozí podobě snímků mohou různit a je vhodné zajistit, že si s nimi neuronová síť dokáže poradit. U barevných snímků jsou možnosti úprav větší.



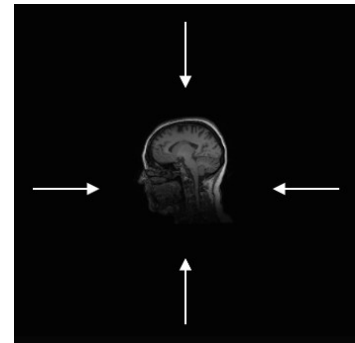
původní snímek



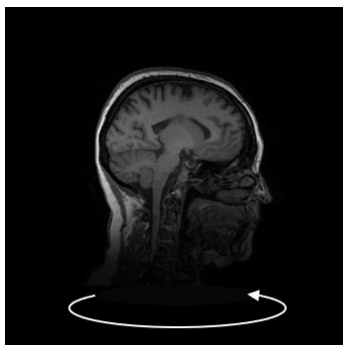
posunutí



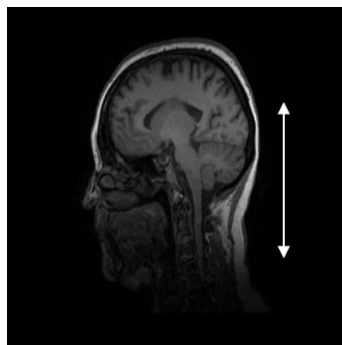
otočení



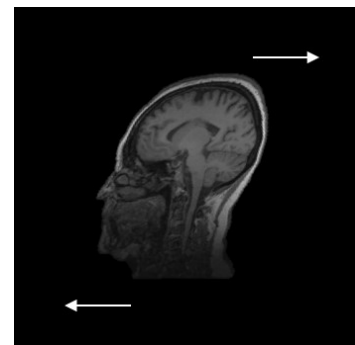
škálování



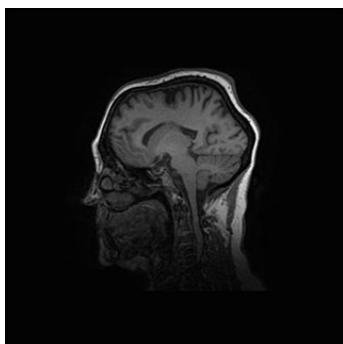
překlopení



natáhnutí



naklonění



pružná deformace



rozostření



úprava jasu a kontrastu

Obr. 35: Druhy transformací

Pro názornost jsou transformace předvedeny ve větší, a tedy nebezpečné míře. Z obrázku je patrné, že zejména natáhnutí, naklonění, pružná deformace a rozostření vyžadují velmi opatrné zacházení, protože jinak obraz deformují až příliš. Zdroj: převzato z [44]

5.8.3 Způsoby implementace rozšíření dat

Studii porovnávajících rozdílné způsoby rozšíření dat a jejich přínosy zatím není mnoho. Zvážit je každopádně potřeba cílovou velikost datasetu. Například i tvorba jediné odvozeniny každého snímku (třeba zrcadlové překlopení) zvýší počet snímků v datasetu na dvojnásobek. Na výběr jsou dvě možnosti, jak rozšíření dat implementovat [44, 48]:

1) Za běhu (online; on the fly)

Algoritmus, který za běhu tvoří odvozeniny dat, se nazývá *generátor*. Než se snímkem začne pracovat neuronová síť, projde úpravou generátorem, který s ním provede předem definované transformace v náhodně stanovené míře. Tento způsob šetří místo na disku, ale zpomaluje proces trénování. Používá se zpravidla na trénovací část datasetu, ale občas je zmiňován i pozitivní dopad při použití na validačních datech, protože může vést k vytrénování robustnějšího modelu [48].

2) Předem (offline)

Náročnější na místo na disku, ale neovlivňuje rychlost trénování modelu, a především dává možnost si zkontrolovat, zda transformace byly použity bezpečně.

Některé výzkumy došly i k závěrům, že trénování je nejlepší začít s původními daty a odvozeniny přidat až v pozdější fázi trénování. Obecná shoda na optimálním postupu ale zatím nepanuje [48].

6 Hluboké učení ve vědě a klinické praxi

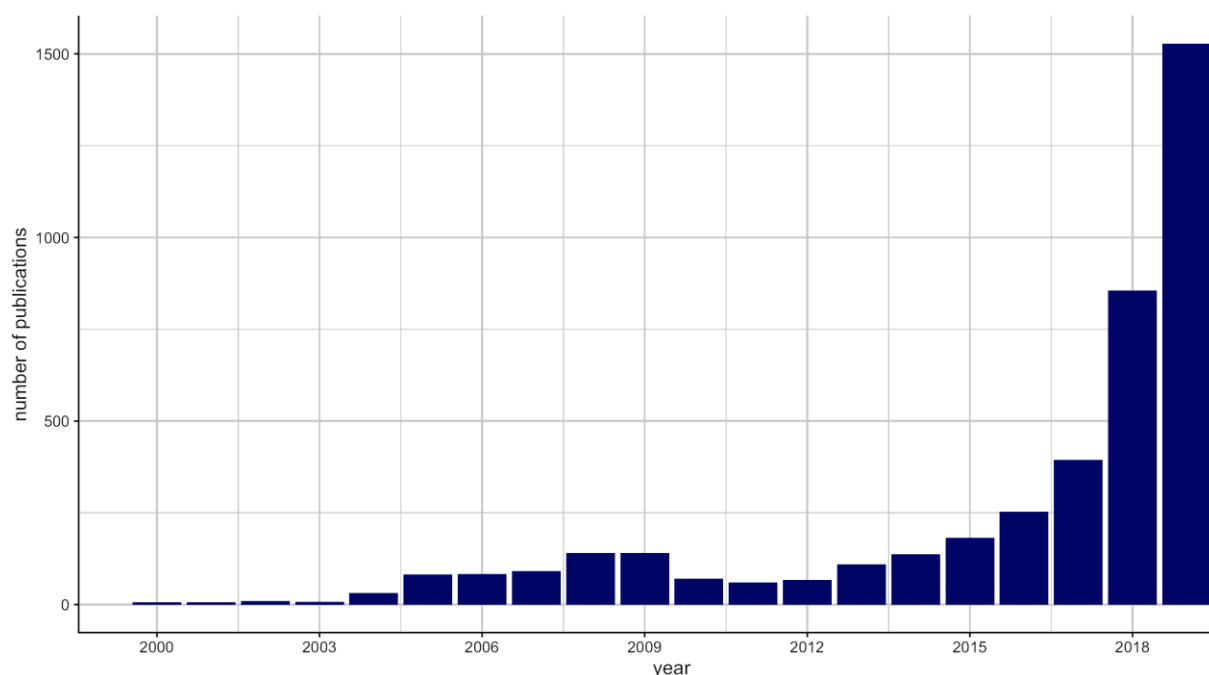
Tato kapitola shrnuje současný stav využití hlubokého učení v radiologii. Stejně jako ostatní kapitoly se vymezuje především na ty výzkumy a technologie, které mají nějaký vztah k tématu této diplomové práce, tedy konvoluční neuronové sítě a RTG snímky zejména rukou – uváděny jsou ty výzkumy, které ovlivnily směřování praktické části této práce. Prozkoumána byla také současná nabídka komerčně dostupného softwaru, který hluboké učení využívá.

6.1 Hluboké učení ve vědě

Množství publikací na téma hlubokého učení v radiologii významně roste, jak ilustruje graf níže. Z organizací zastřešujících odborníky z praxe a vědy se k tématu velmi aktivně staví například již mnohokrát zmiňovaná RSNA (Radiological Society of North America).

Společnost od roku 2017 pořádá každoroční soutěž o nejúspěšnější řešení určitého problému pomocí hlubokého učení. První ročník byl na téma analýzy kostního věku na základě RTG snímku (obsahem má blízko k tématu této práce a je podrobněji rozepsána v kapitole 6.1.5). Výsledkům soutěže je věnována i každoroční konference RSNA. Od ledna 2019 společnost začala vydávat nový odborný časopis *Radiology: Artificial Intelligence* kvůli neustále rostoucímu podílu článků na toto téma v jejich původním časopisu *Radiology* [49]. RSNA tvrdí, že jde prozatím o jediný odborný časopis věnující se současně radiologii a umělé inteligenci [49].

Aktuální stav v ČR byl také zjišťován, avšak Radiologická společnost České lékařské společnosti J. E. Purkyně ani Česká revmatologická společnost České lékařské společnosti J. E. Purkyně na dotazy nereagovaly a z informací dostupných na jejich webech [85, 86] se nezdá, že by v oblasti hlubokého učení podnikaly nějaké významnější aktivity.



Obr. 36: Vývoj počtu publikací týkajících se radiologie a umělé inteligence

Výsledky pochází z databáze PubMed pro hledaný výraz: *(strojové učení NEBO hluboké učení NEBO umělá inteligence) A radiologie*. Platné k 12. únoru 2020. Zdroj: převzato z [29]

6.1.1 Doporučení pro výzkum a publikování

S rostoucím počtem publikací vyplovají na povrch často nedostatečně popsané či záměrně zamlžované aspekty jednotlivých výzkumů. Tyto nejasnosti jsou patrné i z některých výzkumů, které jsou uvedeny na následujících stranách. Na situaci reagovala redakční rada časopisu *Radiology* iniciativou snažící se o systematictější a transparentnější přístup k vyhodnocování výsledků výzkumu. Problematiku rada koncem roku 2019 shrnula v článku *Assessing Radiology Research on Artificial Intelligence* [29, 49], z nějž pochází i následující stručně shrnuté devatero zásad dobré praxe, kterých by se měly výzkumné týmy v publikacích držet.

1. Jasně definovat všechny tři datasety (trénovací, validační a testovací)

Výzkum by měl využívat jeden dataset na trénování, druhý na ladění a validaci a třetí na závěrečné otestování. Datasety by měly být vzájemně nezávislé – bez překryvů. Kromě toho by měla být zveřejněna a vysvětlena kritéria, podle kterých byly některé snímky zahrnuty a jiné naopak vyřazeny.

2. Pro závěrečné vyhodnocení použít externí (cizí) dataset

Neuronové sítě jsou náchylné k přeučení, a proto by na ověření schopnosti dostatečné generalizace měl být využit dataset od někoho jiného, například jiné instituce.

3. Každý z trojice datasetů by měl obsahovat snímky z více různých zdrojů

Snímky z určitého zdroje (např. jednoho typu přístroje či jedné nemocnice) zpravidla vypadají i pro oči radiologů trochu jinak, než snímky od jiného zdroje. Aby se předešlo nechtěnému zaměření algoritmu na data pouze jediného zdroje, měly by trénovací, validační i testovací datasety být tvořeny snímky z více různých zdrojů.

4. Množství snímků v jednotlivých datasetech by mělo být zdůvodněno

Určení přiměřeného množství snímků k vytrénování modelu může být složité a na počátku výzkumu nebývá známo. Proto by mělo být provedeno vyhodnocení úspěšnosti trénovaného modelu v závislosti na množství snímků. Jako příklad může posloužit výzkum, který začal s 200 000 snímků hrudníku, avšak algoritmus už po využití 20 000 snímků vykazoval jen velmi malý přírůstek v přesnosti.

5. Lékařské výstupy použité k trénování by měly mít jednotný a uznávaný formát

Anotování snímků by mělo probíhat způsobem standardizovaným napříč pracovišti. Samotná radiologická zpráva nemusí být pro tyto účely vždy optimální – na stanovení diagnózy mohou být zapotřebí ještě výsledky z jiných vyšetření. Klinické zprávy vznikají kvůli péči o pacienta a obecně nemají kvalitu na to, aby byly použity pro vědecké účely. Jejich použití se výzkumným tým často pravděpodobně nevyhne a v takovém případě je přinejmenším potřeba uvažovat nad způsoby, jak se s nižší kvalitou či chybovostí zpráv zmírnit.

6. Příprava dat a jakákoliv další práce s nimi by měla být zdokumentována

Jakékoliv ořezání snímků, jejich anotace, ruční dotřídění apod. by měly být popsány, protože pomáhají odhadnout použitelnost algoritmu v praxi.

7. Výsledky algoritmu by měly být porovnány s výsledky radiologického odborníka

Výzkumníci často porovnávají výsledky vzájemně mezi sebou a opomínají, že pro léčbu pacientů je důležitější srovnání algoritmu s radiologem. Navíc by se mělo by se jednat o odborníka (včetně uvedení počtu let a v oboru a zaměření). Je pěkné, když algoritmus překoná studenty či nováčky, jenže to ani zdaleka nestačí.

Avšak ani překonání odborníka ještě neznamena, že umělá inteligence je schopná radiology nahradit [25].

8. Způsob, jakým se algoritmus rozhoduje, by měl být jasný a transparentní

V medicíně se nelze spokojit s tím, že na algoritmy umělé inteligence budeme nahlížet jako na černou skříňku. Navíc mnoho výzkumníků shrnuje své výsledky do jednoho číselného údaje (ve stylu např.: *přesnost = 0,95*), který nemá pro praktické využití velkou výpovědní hodnotu. Namísto jednoho čísla by měly být vždy uváděny metriky jako senzitivita, specificita apod. Dále by na každém snímku měly být vyobrazeny oblasti, podle kterých se algoritmus rozhodoval – tedy již zmiňované mapy význačností či teplotní mapy.

9. Algoritmus by měl být veřejně zpřístupněn, aby mohla být jeho úspěšnost nezávisle vyhodnocena

I tato zásada se snaží změnit nahlížení na algoritmy umělé inteligence jako na černou skříňku. Vyvinutý algoritmus by měl být veřejně dostupný k vyzkoušení. Neznamena to, že algoritmus musí být celý volně dostupný ke stažení, pouze to, že zájemcům z řad dalších výzkumníků, recenzentů i čtenářů by mělo být umožněno si ho vyzkoušet. Zveřejněné algoritmy by měly mít standardizovanou formu, kterou se redakce časopisu *Radiology* teprve bude zabývat. Komerční řešení jsou považována za veřejně dostupná.

Koncem března 2020 na tuto problematiku další autoři navázali a v časopise *Radiology: Artificial Intelligence* představili v článku *Checklist for Artificial Intelligence in Medical Imaging (CLAIM)* [87] jasnou osnovu, jíž by se měli výzkumníci držet, a vyjmenovali 42 bodů, co by neměly být při výzkumu opomenuty a v publikacích dostatečně ozřejměny. Osnova je k nalezení v přílohách této práce.

Uvedené dva seznamy doporučení představují v současnosti v podstatě jediná vodítka, jak k výzkumu a publikacím v oblasti hlubokého učení v medicíně přistupovat. Ačkoliv byla vydány téměř v závěru tvorby této práce, byla snaha je v maximální možné míře využít.

6.1.2 Medicínské datasey

Postup přípravy dat a tvorby datasetu již byl popsán, následující kapitola se věnuje detailněji oblasti medicínských dat a náročnosti jejich sběru, zpracování a veřejně dostupným datovým souborům, z nichž některé byly v praktické části této práce využity.

Medicínská data jsou svázána mnoha omezeními – ochranou soukromí a dalšími regulacemi a předpisy, a proto nejsou jednoduše shromažďovatelná, přenositelná mezi institucemi, natož veřejně přístupná [44, 88, 89]. Kvůli těmto legislativním překážkám je nejen velmi obtížné vytvořit rozsáhlé a kvalitní soubory dat, ale jak podotknul profesor Jaroslav Král z Matematicko-fyzikální fakulty Univerzity Karlovy, vůbec efektivně využívat přínosy, které lidstvu výpočetní technologie nabízejí [88, 89].

Další potíž je s anotováním dat. Ruční anotování dat je časově velmi náročné a tím pádem odborníci většinou skončí u malých datových souborů [44, 48]. Ve světovém měřítku se navíc, ať už v závislosti na státu (resp. zdravotnickém systému), ale i na jednotlivých odbornících v rámci jednoho systému, mohou lišit výsledné diagnózy i způsoby jejich určování. Odlišné jsou také formy elektronických záznamů, způsob jejich vyplňování, kvalita a přesnost zaznamenaných údajů. Nepřesnosti se mohou objevit v odborné terminologii, mnoho pojmů je často považováno za synonyma, přestože jejich pravé významy nejsou totožné, ale pouze se z části překrývají [30].

Zádrhelem jsou i velmi vzácná onemocnění – při výskytu několika případů ročně může trvat dlouhá léta, než se podaří sestavit nashromáždít dostatečný počet případů, aby byla data použitelná pro hluboké učení [44, 48].

Snímky, které jsou výstupem medicínských zobrazovacích technik, navíc oproti snímkům z ne-medicínského prostředí využívají k uložení informací zpravidla jen jeden barevný kanál namísto tří, což je pro neuronové sítě obtížnější, neboť ty dokáží dobře zužitkovat i barevnou informaci – použití pouze černobílých snímků podle některých výzkumů může vést k propadu přesnosti vytrénovaného modelu o několik jednotek procent [44].

Následkem všech uvedených překážek je, že pro využití ve strojovém učení musí výzkumníci často rozličné datové soubory složitě spojovat a sjednocovat, aby dosáhli alespoň trochu smysluplného množství dat [44]. Tím pádem typický soubor medicínských snímků zřídka přesahuje 500–1000 pacientů, přičemž od každého pochází řádově jednotky snímků [44]. Tvorbě datových souborů se často věnují univerzity ve spolupráci s nemocnicemi nebo v případě obrazových dat také spolky zastřešující odborníky z oboru diagnostické radiologie.

Pro porovnání situace mimo zdravotnictví: například v praktické části této práce použité předtrénované modely pro přenesené učení využívají již zmíněný dataset ImageNet obsahující přes milionů barevných obrázků [44]. Tento a další tzv. akademické datasey jako CIFAR, MNIST se využívají v pravidelných soutěžích jako standard pro porovnání úspěšnosti nových algoritmů a technik hlubokého učení [48]. Ve zdravotnictví se o porovnání formou soutěže ve větším měřítku snaží zatím pouze společnost RSNA.

Pro praktickou část této práce bylo zkoumáno, zda kromě dat z FNHK lze použít ještě nějaké další snímky z veřejně dostupných zdrojů, aby byl výsledný dataset početnější. Proto byly zmapovány datasey medicínských RTG snímků, které byly v době tvorby této práce veřejně dostupné. Asi nejpočetnější a nejvíce používané datasey vznikly v USA a spravují je buď příslušné univerzity nebo organizace National Institutes of Health (NIH). Evropská radiologická společnost (ESTR) na webu nabízí pouze databázi jednotlivých okomentovaných případů pro účely výuky a vzdělávání, nikoliv pro trénování neuronových sítí [90].

Usnadnit přístup k datasetům se snaží vyhledávač Kaggle (www.kaggle.com/datasets), který nabízí také hodnocení jejich použitelnosti na základě formátu souborů, popisků, licenčních podmínek a četnosti aktualizací. Od ledna 2020 přišel s vlastním vyhledávačem datasetů také Google, služba je dostupná na adrese datasetsearch.research.google.com [91].

Ukázalo se, že většina datasetů (NIH Chest X-ray, MIMIC, CheXpert, ...) se zaměřuje hlavně na snímky hrudníku, dataset RTG snímků pro výzkum revmatoidní či psoriatické artritidy nebyl v době tvorby této práce nalezen žádný. Tématu práce nejvíce podobné byly datasey RSNA Bone Age a OAI, ze kterých byla hrstka náhodně vybraných snímků rukou v praktické části práce využita k ověření funkčnosti vytrénovaných modelů.

RSNA Bone Age

Sestavený pro účely stejnojmenné soutěže. Obsahuje 14 036 snímků převážně levých rukou žen a mužů do 19 let, pořízených za účelem určení kostního věku. Snímky jsou k dispozici v původním rozlišení a formátu PNG s bezztrátovou kompresí, byly dodány dětskými nemocnicemi přidruženými ke Stanfordově univerzitě (2983 snímků) a Univerzitě v Coloradu (11 053 snímků) [92]. Přestože dataset vznikl k jinému účelu, bylo poučné sledovat, jak se jednotlivé soutěžní týmy postavily k otázce předzpracování dat – stejné otázky bylo totiž potřeba řešit i v praktické části této práce.

OAI (Osteoarthritis Initiative)

Sestavený pro výzkum osteoartrózy organizací National Institutes of Health. Obsahuje výstupy z RTG a MRI vyšetření od několika tisíc pacientů. Dataset je dostupný v několika variantách [84]. Zkoumaná varianta obsahuje 43 737 snímků obou kolen, jednoho kolene, pánví, dolních končetin, obou rukou a jedné ruky. Avšak snímky rukou mají velmi nízké rozlišení a kvalitu silně degradovanou vysokou mírou ztrátové komprese formátu JPG.

6.1.3 Výzkumy související s tématem této práce

Níže jsou stručně představeny výzkumy nějakým souvisejícím s problematikou řešenou v této práci. Jde o výzkumy podobné buď řešenou úlohou, tedy zpracováním snímků revmatologických pacientů nebo podobné po technické stránce, mnoho výzkumů poměrně úspěšně na medicínských obrazových datech využilo přenesené učení a podobné či stejné architektury sítí, a to nejen v radiologii, jak ukazuje tabulka.

Přestože mnoho výzkumů prokázalo vhodnost využití hlubokého i přeneseného učení s využitím ne-medicínského datasetu ImageNet, žádný z uvedených výzkumů se nesoustředil na rozlišení jednotlivých revmatických onemocnění – pokud byly využity RTG snímky rukou, bylo to pouze pro analýzu jediného onemocnění nebo k úplně jiným účelům.

V kontextu uvedených výzkumů je tedy úloha řešená v praktické části této práce nová a smysluplná i pro další zkoumání i po dokončení této diplomové práce.

Tabulka 3: Výzkumy úspěšně využívající přenesené učení na medicínských datech

řešená úloha	architektura sítě	původní dataset	metoda přenosu učení	zdroj
Klasifikace eroze kostí při revmatoidní artritidě	AlexNet	ImageNet	extrakce rysů	[93]
Klasifikace CT snímků ledvin s podezřením na rakovinu	Inception-V3	ImageNet	extrakce rysů	[94]
Detekce diabetické retinopatie na snímcích sítnice	Inception-V3	ImageNet	extrakce rysů	[95]
Klasifikace RTG snímků hrudníku	Inception-ResNet-V2	ImageNet	ladění	[96]
Detekce zlomenin na RTG snímcích zápěstí	Inception-ResNet-V2	COCO	neuvedeno, ladění (?)	[97]
Klasifikace RTG mamografických snímků	CheXNet	NIH Chest X-ray	vyzkoušeno více metod	[56]

6.1.3.1 Detekce a kvantitativní popis osteoartrózy rukou a kolen

Podobnost: řešená problematika, RTG rukou

Zdroj: [98]

Výzkumníci z Curtinovy univerzity v Austrálii se v roce 2016 pokoušeli vyvinout software, který by uměl rozpoznat, kvantitativně popsat a předpovědět progresi osteoartrózy kolen a rukou.

Využili dataset OAI. Vstupní snímky software pomocí segmentačních algoritmů rozřeže na předem definované oblasti zájmu, ve kterých se následně provedou měření kloubních štěrbin. Strukturu kostí se snažili popsat pomocí číselných charakteristik vypočítaných, k těmto datům přidali údaje o věku, pohlaví a BMI a použili algoritmy strojového (nikoliv hlubokého) učení k vytvoření klasifikačního modelu. Funkčnost řešení je možné si vyzkoušet na webu readmyxray.curtin.edu.au, výstupem softwaru je zpráva s kvantitativním hodnocením snímku.

Podle autorů je jejich řešení schopné i předpovídat progresi onemocnění a plánují jen do budoucna rozšířit i o další onemocnění, zejména revmatoidní artritidu.

6.1.3.2 Klasifikace eroze kostí při revmatoidní artritidě

Podobnost: RTG rukou, přenesené učení z datasetu ImageNet, metoda extrakce rysů

Zdroj: [93]

Úlohu blízkou tématu této práce řešil kolektiv japonských vědců a své výsledky představil v roce 2017. Jejich řešení spočívalo v tom, že ze vstupního RTG snímku se nejdříve softwarově odstraní (potlačí) měkká tkáň, následně se pomocí segmentace vyřezou snímky jednotlivých prstů a ty se v konečné fázi pomocí neuronové sítě klasifikují podle míry eroze kostí.

Vlastní dataset tvořilo celkem 129 snímků pro trénování a validaci (90 případů RA a 39 případů bez příznaků) a dalších 30 snímků pro testování a vyhodnocení. Z metod rozšíření dat využili překlopení a otočení o násobky devadesáti stupňů.

K přenesenému učení formou extrakce rysů využili síť architektury AlexNet vytrénovanou na datasetu ImageNet. Dosáhli senzitivity 80,5 % a specifity 0,84 %

Autoři konstatovali, že jejich metoda nedosáhla úrovně přijatelné k nasazení do praxe a v budoucnu chtějí zlepšit přesnost klasifikace vytrénováním na větším datasetu.

6.1.3.3 Klasifikace RTG snímků hrudníku z datasetu NIH Chest X-ray

Podobnost: přenesené učení z datasetu ImageNet, metoda ladění

Zdroj: [96]

V dubnu 2019 byl publikován článek výzkumníků z IBM Research, kteří přenesené učení využili při klasifikaci RTG snímků hrudníku. Řešenou úlohou byla binární klasifikace – rozlišit zdravé pacienty od nemocných.

Vycházeli z datasetu NIH Chest X-ray, který ale obsahuje nepřesné anotace vygenerované pomocí strojového učení. Pro část snímků z datasetu vyjmuli a znovu anotovali ručně, čímž získali 1300 snímků bez jakéhokoliv nálezu a 1917 snímků s jedním nebo více nálezy. Celkem 3217 snímků. Snímky upravili na rozměr pouze 128 × 128 px a pomocí metod rozšíření dat přidali otočení, posunutí a škálování.

Ve frameworku Keras použili síť architektury Inception-ResNet-v2 předtrénovanou na datasetu ImageNet. Přenesené učení mělo formu ladění a zmrazeny byly pouze vrstvy extrahující nízkourovňové charakteristiky obrazu.

Pokud třídu zdravých snímků označíme jako negativní, své řešení autoři optimalizovali tak, aby nedocházelo k chybě 2. druhu, tedy aby žádný snímek s příznaky onemocnění nebyl označen jako zdravý. Takto optimalizovaný vytrénovaný model byl schopen do třídy *bez nálezu* zařadit polovinu všech snímků, které tam doopravdy patří. Naopak ani jeden snímek s nálezem chybně neoznačil jako zdravý.

6.1.3.4 Detekce zlomenin na RTG snímcích zápěstí

Podobnost: RTG rukou, přenesené učení

Zdroj: [97]

Tým výzkumníků z Národní univerzity v Singapuru publikoval počátkem roku 2019 článek popisující využití neuronových sítí na detekci a lokalizaci zlomenin na RTG snímcích zápěstí.

Sestavili vlastní dataset 7356 snímků použitých k trénování a validaci a dalších 524 k závěrečnému otestování a vyhodnocení. Z metod rozšíření dat využili pouze překlopení podle svíslé osy.

K trénování použili síť architektury Inception-ResNet-V2 předtrénovanou na datasetu COCO (Common Objects in COntext). Formu přeneseného učení neuvádějí, ale vzhledem k délce a parametrům trénování mohlo jít o ladění. Ve všech sledovaných metrikách model dosáhl výsledků v rozmezí 82 až 97 %.

6.1.3.5 Klasifikace RTG mamografických snímků

Podobnost: přenesené učení, RTG snímky

Zdroj: [56]

Ze všech uvedených příkladů v této kapitole je tento jediný, který popisuje přenesené učení mezi medicínskými datasety. Výzkumníci z Jakarty využili dataset DDSM, původní čtyři třídy sloučili do dvou a úlohu zjednodušili na klasifikaci pozitivních a negativních nálezů (binární klasifikaci). Pro trénování jim Pekingská univerzita poskytla síť CheXNet předtrénovanou na datasetu NIH Chest X-ray a experimentálně vyzkoušeli několik strategií přeneseného učení s cílem zjistit, které vrstvy původní sítě je lepší ponechat a které upravit. Bylo dosaženo celkové správnosti 90,38 %, podrobnější vyhodnocení autoři neuvádějí.

6.1.3.6 Klasifikace CT snímků ledvin s podezřením na rakovinu

Podobnost: přenesené učení, Inception, dataset ImageNet, extrakce rysů

Zdroj: [94]

V roce 2019 kolektiv výzkumníků z Kalifornské univerzity publikoval své výsledky z hledání softwarového řešení klasifikace nádorů ledvin na snímcích z magnetické rezonance.

Dataset tvořily snímky od celkem 179 pacientů. K binární klasifikaci použili formou extrakce rysů síť Inception-V3 předtrénovanou na datasetu ImageNet. Ve sledovaných metrikách dosáhli hodnot od 75 do 90 % a výsledek své studie hodnotí jako slibný.

6.1.4 Detekce kloubů na ruce

Jednou z úloh řešenou v praktické části této práce je i klasifikace snímků jednotlivých kloubů. K tvorbě výřezů kloubů byla použit na míru vyvinutý program, jehož tvorbě předcházely průzkum již hotových (níže popsaných řešení), z nichž ale žádné nebylo možné využít.

6.1.4.1 FingerNet

Zdroj: [99]

Tým vědců z Národní univerzity v Soulu vyvinul v roce 2015 software s názvem FingerNet pro detekci kloubů prstů z rentgenových snímků rukou. Podle autorů údajně dosáhl průměrné celkové správnosti 98,02 %. FingerNet kombinuje dva způsoby – konvoluční neuronové sítě vlastní architektury vycházející z typu LeNet-5 a hledání lokálních maxim (kloubních štěrbin) v obrázku převedeném na signál, tedy na matematickou funkci. Oba výstupy jsou následně složeny rozhodovacím algoritmem podle pevně daných pravidel.

Autoři do datasetu použili pouze snímky od osob ve věku 80–180 měsíců. Dětské klouby jsou na snímcích jasně definované a mají široké kloubní štěrbin. Tím se autoři vyhnuli problémům se snímky starších osob, na kterých kvůli stárnutí a onemocnění některé jsou klouby být závažně poškozené nebo se vzájemně překrývají kvůli nemožnosti během snímání natáhnout prsty.

Trénovací dataset tvořilo 3000 výřezů prostředních tří prstů z 1000 snímků rukou a k tomu dále 10 000 výřezů neobsahujících klouby. Z metod rozšíření dat využili pouze pružnou deformaci k vygenerování 100 dalších odvozenin z každého snímku. Celkem tedy bylo k dispozici 1,3 milionů snímků pro trénování a dalších 130 (zachycujících 1950 kloubů) pro testování a závěrečné vyhodnocení.

Pro předcházení omylům je vhodné dodat, že pod názvem FingerNet se ve vědecké sféře lze setkat jednak s touto korejskou aplikací na detekci kloubů a jednak také s čínskou aplikací na rozpoznání otisků prstů, kterou v roce 2017 vyvinuli výzkumníci na Pekingské univerzitě [100]. [Zdrojový kód FingerNetu, který je k dispozici na GitHubu](#), patří právě této čínské aplikaci.

6.1.4.2 BoneFinder

Manchesterská univerzita představila v roce 2013 software BoneFinder, jehož hlavní autorkou je Claudia Lindner, která ho vytvořila v rámci PhD studia [101]. BoneFinder automaticky ohraničí a segmentuje kosti z rentgenových snímků a umožňuje uložit souřadnice nalezených klíčových bodů do samostatného souboru [102]. Výsledné body je možné použít (už mimo rámec funkcionality BoneFinderu) na statistickou analýzu tvaru kostí nebo na automatizovaný geometrický popis (např. velikost kloubních štěrbin apod.) [101].

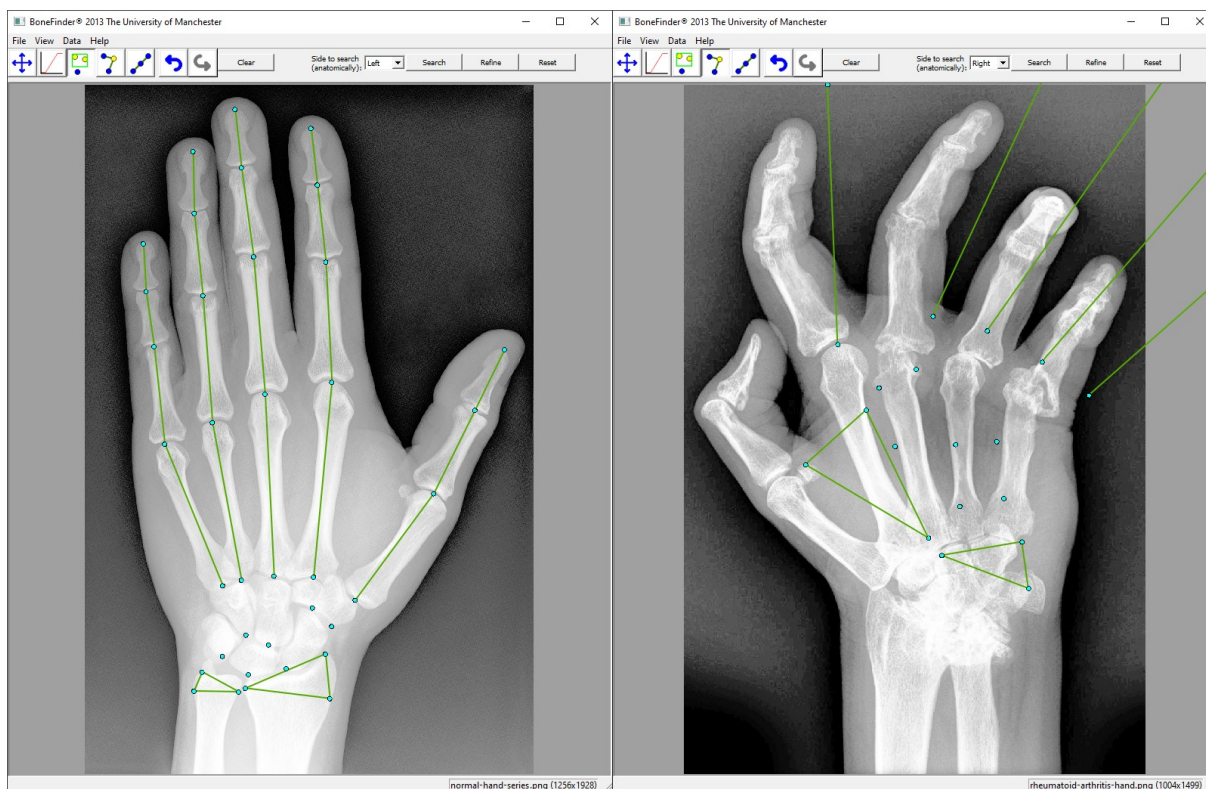
BoneFinder využívá na míru připravené algoritmy strojového učení, pomocí kterých v prvním kroku ve snímku najde oblast zájmu a hrubé pozice bodů a následujícím kroku pozici začistí a upřesní. Technologie byla patentována a patentový záznam je k dohledání pod označením EP 2893491 (země EU) nebo US9928443 (Spojené státy) [101, 102].

Vlastní zkušenost

Pro nekomerční výzkumné účely je BoneFinder k dispozici zdarma na oficiálním eshopu www.click2go.umip.com/i/software/Bonefinder.html. Důležité je, že pro práci s programem je potřeba zažádat vždy také o příslušný model. Ve výchozím stavu se program dodává s modelem vytrénovaným na analýzu snímků kyčelních kloubů. Model vytrénovaný na snímky rukou je v eshopu nabízen jako samostatný produkt [103].

Model rukou byl vytrénován na 564 snímcích rukou v obvyklé AP projekci s cílem vyznačit klouby rukou s pomocí 37 klíčových bodů. Ve výchozím stavu se počítá, že vstupem bude levá ruka. Program ve výchozím nastavení poměrně drasticky u každého snímku vyrovná jas a kontrast úpravou histogramu. Nalezené body se ukládají ve strojově čitelném formátu do souboru s příponou .pts (zkratka od slova points). Dávkové zpracování snímků není možné, pouze každý jednotlivě.

BoneFinder byl vyzkoušen na několika náhodně vybraných snímcích poskytnutých pro tuto práci Fakultní nemocnicí Hradec Králové. V případě rukou bez příznaků revmatických onemocnění si BoneFinder vedl skvěle a potvrdilo se, co se o něm píše na oficiálním webu – že způsob umisťování bodů je opravdu robustní a velmi přesný [101]. Nicméně na revmatických rukou BoneFinder označil klouby jen v případech, kde nebyla nemocí výrazněji zasažena oblast zápěstí. Jinak program nedokázal lokalizovat referenční pozici pro lokalizaci dalších kloubů. Na označování kloubů pro potřeby této diplomové práce se tím pádem BoneFinder neosvědčil.



Obr. 37: Ukázka funkčnosti programu BoneFinder

Vlevo správné označení neartritické ruky, vpravo neúspěšný pokus s rukou v pokročilém stadiu RA.
Zdroj: Vlastní zpracování s využitím programu BoneFinder a snímků [18, 35]

6.1.5 RSNA Bone Age Challenge

Roku 2017 uspořádala společnost RSNA soutěž o nejpřesnější odhadnutí kostního věku z rentgenových snímků rukou dětí a mladistvých. Kromě věku bylo součástí i určení pohlaví. Kostní věk se běžně určuje pomocí porovnávacích atlasů a pro svou rutinní povahu, zdlouhavost a zastaralost metod se zdál být úlohou pro řešení pomocí hlubokého učení [54].

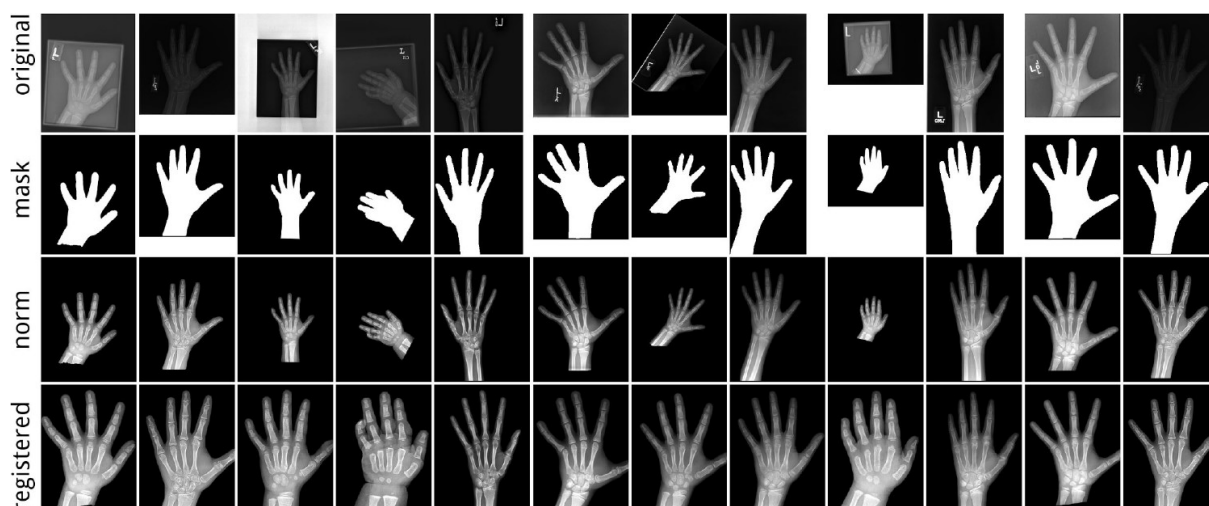
I když bylo zadání specificky zaměřené, soutěž stejně jako tato práce řešila úlohu klasifikace RTG snímku rukou a z výstupů účastníků vzešly poznatky cenné i pro tuto práci, zejména co se týče způsobu předzpracování vstupních dat (dataset popsán v kapitole 6.1.2). Soutěži a rozdílným přístupům některých týmů je proto věnováno více prostoru na následujících stranách.

Soutěže se zúčastnilo přes 37 týmů čítajících dohromady přes 250 členů – radiologů a jiných lékařů, počítačových vědců z akademického prostředí a odborníků z technologických firem.

6.1.5.1 Řešení týmu Michiganské univerzity

Tento tým nevyhrál, jeho řešení je zajímavé především způsobem předzpracování dat. Protože snímky v datasetu měly velmi rozdílnou kvalitu, co se týče jas a kontrastu, orientace snímku a nadbytečných okrajů, rozhodli se autoři pro velmi důslednou úpravu všech snímků do standardizovaného formátu. Použili segmentační síť U-Net na odstranění pozadí a orientačních značek, vyrovnali jas a kontrast a pomocí sítě architektury podobné VGG otočili a případně stranově převrátili všechny snímky do jednotné podoby [104]. Zmíněné úpravy znázorňuje obrázek na následující straně. Obdobným způsobem dataset upravil i tým z Harvardovy univerzity, který taktéž nevyhrál [105].

Ke klasifikaci věku autoři použili vlastní síť opět na bázi architektury VGG a dosáhli poměrně dobrých výsledků [104]. K zamyšlení je, zda se až příliš důkladnou snahou o sjednocení podoby snímku nepřipravili o důležité informace – například malé ruce batolat a velmi mladých dětí (na obrázku 4. zleva a 4. zprava) byly původně na snímcích výrazně menší a vyobrazeny méně kontrastně kvůli tomu, že tkáně byly pro rentgenové paprsky menší překážkou.



Obr. 38: Způsob předzpracování datasetu v podání soutěžního týmu Michiganské univerzity

Zdroj: převzato z [104]

6.1.5.2 Řešení týmu 16 Bit

Vítězné řešení pochází od startupu 16 Bit, který v Torontu založili Mark Cicero a Alexander Bilbily, dva radiologové zajímající se o počítačové vědy [106]. Vítězné řešení dali autoři veřejně zdarma k vyzkoušení také na svůj web: <https://www.16bit.ai/bone-age> [54]. (Ke smysluplnému otestování je ale zapotřebí mít RTG snímek ruky ve věku do 19 let, a to ideálně takový, který nebyl součástí trénovacího datasetu.)

Rozlišení snímků

Obvyklé architektury neuronových sítí přijímají vstupní snímky ve čtvercovém formátu s délkou strany v rozmezí 200–300 px, což se autorům zdálo být až příliš málo. Na druhou stranu plné rozlišení snímků se jim zdálo až zbytečně velké, protože ani radiolog, ani počítač nepotřebují k rozhodování všechny zhruba 3 milionu pixelů, které průměrný snímek má. Proto experimentovali s různými rozměry až do 750 × 750 pixelů a nakonec se rozhodli použít velikost 500 × 500 px, což se jim zdálo jako rozumný kompromis vzhledem k řešené úloze, poskytnutému datasetu a dostupné paměti na grafické kartě [21].

Počet snímků a jejich předzpracování

Autoři sestavili pouze trénovací a validační skupinu, čím obohatili dataset o snímky, které by jinak padly na vytvoření testovací skupiny (soutěž totiž měla stranou vyčleněných dalších 1425 snímků k závěrečnému porovnání jednotlivých týmů). Trénovací data autoři údajně nijak nepředzpracovávali (alespoň to nezmiňují), implementovali pouze metody rozšíření dat ve formě otočení v rozmezí 20 stupňů, vodorovný a svislý posun v rozmezí 20 %, změnu měřítko v rozmezí 20 % a vodorovné překlopení. Tyto hodnoty vybrali na základě svého úsudku tak, aby zhruba odpovídaly variacím, které by se mohly objevit i přirozeně [21].

Architektura neuronové sítě

Autoři zkoušeli několik architektur neuronových sítí, např. Inception-V4, ResNet-152, DenseNet a jiné a nakonec se rozhodli pro Inception-V3. Volbu nevyužít předtrénovaný model (přenesené učení), zdůvodnili tím, že [54]:

- 1) pracovali s jinou velikostí snímků
- 2) množství snímků v datasetu jim přišlo dostatečné k vytrénování modelu od základů

Výsledné řešení bylo tvořeno *seskupením (ensemble)* více neuronových sítí, protože kromě věku na základě snímku bylo nutné určovat i pohlaví na základě přiložené textové informace. Na trénování měli k dispozici jeden počítač vybavený grafickými kartami Nvidia P40 a dvojicí Titan X a druhý počítač s jedinou kartou Titan X. Všechny grafické karty firmě věnovala Nvidia v rámci členství v startup akcelérátoru Nvidia Inception Program [21].

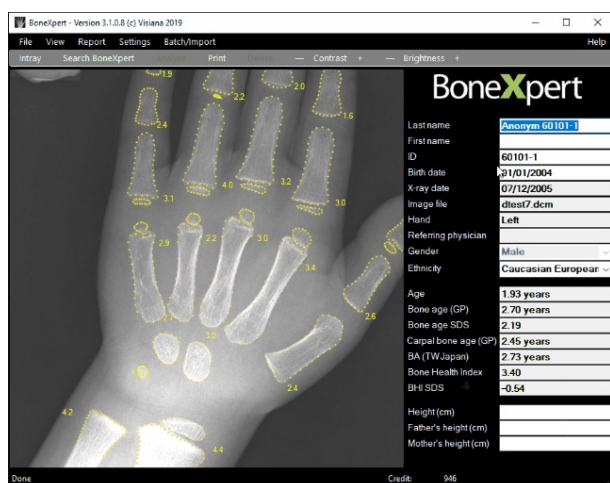
6.2 Hluboké učení v praxi

Komerční software využívající hluboké učení se v současné době zaměřuje především na co nejlepší kvantitativní popis snímku a určité sjednocení výstupního formátu zpráv z radiologických vyšetření. To je předpokladem k tomu, aby v budoucích letech mohla další generace softwaru asistovat při stanovování diagnóz.

V následujícím textu jsou některá softwarová řešení představena. Přehled slouží k ilustrování aktuálního stavu a zdaleka není úplný. Společnosti Siemens, Fujifilm a Philips byly osloveny s prosbou o detailnější informace, avšak k datu odevzdání této práce žádná z nich oficiálně neodpověděla.

K tematice této práce má nejbližší BoneXpert, ovšem software specificky zaměřený na revmatická onemocnění patrně zatím neexistuje a tím pádem zde existuje určitý potenciál.

6.2.1 Visana – BoneXpert



Obr. 39: Rozhraní programu BoneXpert
Zdroj: převzato z [12]

Snahy využít hluboké učení v praxi pro určování kostního věku tu byly už dávno před RSNA Bone Age Challenge. Už v roce 2009 představila dánská společnost Visana na k těmto účelům software BoneXpert. Podle jeho autorů je údajně jedním z prvních příkladů využití metod umělé inteligence v radiologii. I tento software se zmiňované soutěže zúčastnil a byl následně velmi citelně vylepšen [12].

K říjnu 2018 mělo licenci na BoneXpert zakoupeno 135 nemocnic z celého světa, mezi nimi také Fakultní nemocnice v Motole [107].

6.2.2 Vuno – Med Bone Age a další

Počátkem roku 2019 obdržel certifikaci pro prodej na evropském trhu také software Med Bone Age korejské firmy Vuno [45]. Funkčnost programu byla prověřena klinickými zkouškami. Software je navržen tak, že trénování nových modelů a jejich nasazování probíhá průběžně [45].

Kromě Bone Age nabízí společnost Vuno také obdobné programy Deep Brain na segmentaci a kvantitativní analýzu neurodegenerativních postižení mozku, Chest X-ray na diagnózu hlavních abnormalit na snímcích hrudníku, Fundus AI na detekci lézí na snímcích sítnice a dále Lung CT AI na detekci a analýzu plicních uzlíků [108].

6.2.3 Siemens – AI-Rad Companion, AI-Pathway Companion a Syngo

Siemens vyvíjí softwarovou platformu AI-Rad Companion, do které patří několik produktů. První z této řady byl Chest CT (představen na konferenci RSNA 2018), který automaticky segmentuje jednotlivé anatomické struktury a označí změny vyvolané onemocněním. Z vyšetření software generuje standardizované a reprodukovatelné zprávy kvantitativního vyhodnocení snímku. Samotné zpracování snímků probíhá v cloudu. Aplikace byla certifikována pro USA i EU a dokáže zpracovávat data všech výrobců CT systémů. Dalšími produkty v platformě jsou Brain MR, Chest X-ray, Organs RT a Prostate MR. Ve všech případech software cílí na reprodukovatelné kvantitativní výstupy a vyznačení potenciálních nálezů [47].

Dalším druhem softwaru je pak například AI-Pathway Companion, do kterého se mají soustředit uchovávaná data o pacientech – ze všech vyšetření (pracovišť) současně. Na základě anamnézy, výstupů z radiologie, laboratorních zkoušek, genetiky a patologie dokáže program s pomocí strojového učení analyzovat stav pacienta a navrhnout další kroky [47].

Dále Siemens vyvíjí software Syngo pro vyhodnocování výstupů mnoha zobrazovacích metod. I tento software se postupně posouvá od popisu snímku k nabízení diagnóz s pomocí hlubokého učení [47].

Jen pro srovnání – zatímco výzkumy univerzit a nemocnic se typicky potýkají s nedostatkem dat, Siemens uvádí, že má pro trénování algoritmů hlubokého učení k dispozici databázi s 325 miliony záznamů: anotovanými snímky, lékařskými zprávami a dalšími daty [109].

6.2.4 Další software – Brainomix a Quantib Brain

Pro zpracování snímků z vyšetření mozku kromě řešení od Siemensu existují ještě dva produkty taktéž zaměřené na tuto anatomickou oblast:

- Brainomix, který se kromě kvantitativního popisu soustředí na rozpoznání a vyhodnocení cévní mozkové příhody [110]
- Quantib Brain taktéž nabízející kvantitativní popis a zaměření pro změnu na neurodegenerativní choroby [46]

Uvedený software nastiňuje hlavní směry vývoje hlubokého učení v radiologii, avšak s tématem práce bezprostředně nesouvisí, a proto delší představení není potřeba.

7 Vlastní řešení

Následující část práce popisuje praktické nasazení konvolučních neuronových sítí ke klasifikaci RTG snímků rukou s příznaky revmatických onemocnění. Byly aplikovány poznatky popsané v předchozích částech práce, zužitkovány závěry jiných výzkumů a také zmapováno několik nových úskalí a problémů. Všude, kde to bylo relevantní, se řešení snaží rámcově držet doporučení pro publikování definovaných redakční radou časopisu *Radiology* a související osnovy CLAIM. Na základě nich oproti jiným výzkumům detailně popisuje způsoby zpracování a třídění vstupních dat a vyhodnocení vytrénovaných modelů pomocí běžných metrik. Cílem je ověření proveditelnosti v podobě tzv. ověření koncepce (proof of concept).

Je potřeba dodat, že ačkoliv proces přípravy dat a trénování modelů sice vycházel ze schématu uvedeného v kapitole 4.4.4, vzhledem k experimentální povaze celého vědního oboru vůbec nebyl lineární a ve schématu nakreslená smyčka byla mnohokrát využita. Kvůli lineární podstatě textu byla snaha výstupy práce určitým způsobem seřadit a popsány jsou pouze ty cesty, které vedly k úspěšným či jinak zajímavým výsledkům.

Kvůli právním omezením není možné uvádět příklady skutečných snímků, namísto toho byly alespoň v několika málo případech použity volně dostupné snímky ilustrační.

7.1 Návrh studie – vytyčení cílů, předpoklady a očekávání

Obecná vize – kontext, ve kterém se studie pohybuje

Počáteční velice obecná vize byla vyvinutí softwaru, který by byl nasazen na radiologických pracovištích a dokázal identifikovat jednotlivá revmatická onemocnění, podrobně popsat pozorované projevy (změny na kloubech), navrhl diagnózu či alespoň určil, zda jsou změny zánětlivého původu (artritické) nebo způsobené stářím (artrotické). V duchu teoretických východisek a již dostupných řešení popsaných v předchozích kapitolách by role softwaru rozhodně neměla být taková, že radiology nahradí, pouze by při rozhodování mohl nabídnout další názor, umět ho patřičně zdůvodnit a snížit riziko, že lidské oko něco přehlédne.

Klíčovou částí celého systému by bylo několik vytrénovaných neuronových sítí (modelů) – jeden na klasifikaci celých snímků rukou, další na detekci kloubů a poslední na klasifikaci způsobu a míry postižení jednotlivých kloubů. Výsledné řešení by mohlo nabídnout radiologům dva nezávislé názory, obdobně jako například software Googlu pro diagnostiku skenu sítnice zmiňovaný v kapitole 4.4.5.3. Šlo by jednak o výsledek klasifikace snímku celé ruky, kde výstupem by byly pravděpodobnosti jednotlivých diagnóz. Druhým výstupem by byla míra postižení jednotlivých kloubů. Na základě toho, které klouby a jak jsou postižené, by pak mohl například i obyčejný rozhodovací systém nabídnout druhý pohled na celkovou diagnózu. Využití snímků kloubů je lákavé i z toho důvodu, že pokud by se ukázal počet snímků celých rukou jako nedostačující pro vytrénování modelu, snímků kloubů by mohlo být až 13× více.

Užší vymezení

Následující kapitoly této práce se budou soustředit právě na zmíněné klasifikační modely. Konkrétně budou hledány cesty k jednotlivým modelům, které by dokázaly:

- 1) V lepším případě rozlišení tříd:
 - bez příznaků revmatických onemocnění (dále označováno zkratkou NEART)
 - revmatoidní artritida (RA)
 - psoriatická artritida (PsA)
 - osteoartróza (OA)
- 2) V nejjednodušším případě rozlišení rukou artritických a artrotických – tedy zda se příznaky zdají být spíše zánětlivého původu nebo způsobené stářím. Toto je zjednodušený případ, pokud by se ukázalo množství snímků z FNHK jako nedostačující pro výše uvedenou úlohu.
- 3) Určení způsobu postižení jednotlivých kloubů

Kvalita tří výše zmíněných modelů je limitována daty, která jsou k dispozici. Jak už bylo uvedeno v kapitole 6.1.2 Medicínské datasey, kromě osteoartrózy nejsou zatím k dispozici žádné volně dostupné velké datasey rukou s revmatickými onemocněními a zmíněný dataset OAI zaměřující se na osteoartrózu navíc trpí nízkým rozlišením a kompresí snímků. Proto je nutné vycházet pouze ze snímků, které poskytla FNHK a hledat cesty, jak z nich získat maximum – vyzkoušet metody přeneseného učení a rozšíření dat. V souladu s doporučeními publikovanými redakční radou časopisu *Radiology* [49] budou snímky z externích zdrojů vyzkoušeny alespoň k závěrečnému vyhodnocení vytrénovaných modelů.

Třetí úloze by měla předcházet samotná detekce kloubů na snímcích rukou – ta však řešena nebude a na místo toho budou klouby vyznačeny ručně. Důvody jsou následující:

- Není záruka, že úloha klasifikace jednotlivých kloubů přinese nějaké výsledky. Proto se zdá být smysluplnější přednostně vyzkoušet právě klasifikaci a cestami, jak jednotlivé klouby ze snímků získat, se zabývat až v případě, že výsledky klasifikace budou příznivé.
- Pro nalezení souřadnic kloubů na snímcích zdravých nebo nemocí málo zasažených rukou již nástroje existují a dosahují vynikajících výsledků. Jako příklad byl v kapitole 6.1.4 vyzkoušen software BoneFinder a popsána aplikace FingerNet.
- Pro detekci souřadnic kloubů na nemocí výrazně postižených rukou by bylo zapotřebí mnoho snímků tak jako tak označit ručně, aby bylo možné na čem vytrénovat. Vzhledem k tomu, jak drastické v těžkých případech deformace kloubů jsou, není záruka, že by tato cesta byla úspěšná – pokud by se klouby ručně neoznačily už pro fázi trénování, velice pravděpodobně by stejně musely být označeny ručně, protože by si model s jejich detekcí nemusel poradit.

7.2 Analýza dat, jejich předzpracování a tvorba datasetů

7.2.1 Data poskytnutá FNHK

Ve FNHK bylo v době přípravy dat v létě 2019 vedeno 494 pacientů s diagnózou revmatoidní artritidy [13]. Dále bylo k dispozici poměrně velké množství snímků pacientů trpících psoriatickou artritidou a osteoartrózou (přesné počty nebyly upřesněny). V kapitole 3.2.3 Postup vyhodnocení RTG snímku radiologem je uvedeno, že radiolog by měl mít znalost normálního anatomického stavu, aby byl schopen posoudit odchylky. Tento postup se zdálo smysluplné napodobit i při použití neuronové sítě, a tak byly k uvedeným třem skupinám přidány ještě snímky rukou z revmatologického pohledu v pořádku, které byly pořízeny kvůli podezření na trauma nebo jiným důvodům.

Snímky uchovává FNHK ve formátu DICOM (s příponou .dcm) v systému PACS. Snímky použité pro tuto práci byly ještě na straně nemocnice ze systému exportovány v anonymizované podobě, tzn. bez jakýchkoliv metadat o pacientech. Z podstaty řešené úlohy byl každý snímek opatřen pouze jediným údajem – diagnózou. Podle té byly snímky dodány již seřazené do čtyř základních skupin: osteoartróza (OA), psoriatická artritida (PsA), revmatoidní artritida (RA) a ruce bez příznaků revmatických onemocnění (NEART). Diagnózy se nezakládají pouze na radiologických výstupech, jsou skutečně potvrzené revmatology s využitím dalších vyšetření a informací o pacientech. Celkový počet snímků byl v této fázi lehce přes 2000.

7.2.2 Převod z formátu DICOM

Prvním úkolem řešeným v rámci předzpracování dat byl převod obrazového formátu ze zdravotnického standardu DICOM do formátů běžných v oblasti počítačové grafiky a hlubokého učení. Důvody byly následující:

1. Běžně dostupné frameworky pro hluboké učení formát DICOM nepodporují.
2. Grafické editory a softwarové knihovny, které jsou potřeba na předzpracování snímků, disponují pouze velmi omezenou podporou formátu DICOM – jde pouze o určité kombinace verzí softwaru a specifikací DICOM.
3. Ne-zdravotnické formáty dat mohou nabídnout srovnatelnou kvalitu obrazu a na rozdíl od formátu DICOM jdou zobrazit v běžných prohlížečích obrázků, což přináší možnost vizuální kontroly a snadnějšího a rychlejšího dodatečného třídění snímků.

Cílem bylo snímky převést na rastrový formát, který kompresi nepoužívá vůbec, nebo pouze bezztrátovou. Volba padla na formát PNG, protože umožňuje bezztrátovou kompresi. K převodu byla vyzkoušena celkem tři volně a bezplatně dostupná řešení, z nichž pouze jedno se osvědčilo. Přehled použitých nástrojů je shrnut v tabulce na následující straně.

Tabulka 4: Nástroje vyzkoušené k převodu z formátu DICOM do PNG

	operační systém	výstupní formát PNG	počet chyb	dostupnost a cena
Morgify	Linux	libovolný	asi 2000	zdarma
Dicom Converter	Windows	8 bit stupně šedi	239	zkušební verze na 5 dnů, cena 199 USD
MicroDicom	Windows	24 bit RGB	7	zdarma pro nekomerční využití, jinak cena 99 USD

Morgify

Morgify je zdarma dostupný, podporuje pouze operační systém Linux a ovládá se výhradně pomocí příkazové řádky. Teoreticky umožňuje převod mezi zhruba 200 vstupními a výstupními formáty, změnu velikosti a základní úpravy jako jas, kontrast, doostření, rozostření apod. [111]

Při dávkovém převádění snímků z FNHK program naprosto selhal. Úspěšně dokázal převést pouze několik desítek snímků a zbylé tisíce se mu převést nepodařilo. Na vině může být špatně implementovaná podpora novějších specifikací formátu DICOM.

Dicom Converter

Dicom Converter je zdarma k vyzkoušení na 5 dnů bez omezení funkčnosti, poté je nutné ho odinstalovat nebo zakoupit licenci za 199 USD. Má jednoduché a přehledné grafické rozhraní. Podporuje formáty JPG, PNG, BMP, TIF a při převádění umožňuje zpracovat celé složky včetně zachování struktury podsložek. Výstupní snímky se ukládají v odstínech šedi s hloubkou 8 bitů.

Při dávkovém převádění snímků z FNHK program pracoval pomalu, občas úplně přestal reagovat a bylo potřeba ho vypnout a znovu spustit. Celkem 239 snímků převedl chybně.

**Obr. 40: Chyby při převádění programem Dicom Converter (vpravo původní snímek)**

Zdroj: vlastní zpracování s využitím snímku [19]

MicroDicom

MicroDicom je pro nekomerční použití zdarma dostupný na www.microdicom.com. Cena za jednu komerční licenci činí 99 USD [112]. Má přehledné grafické rozhraní ve stylu starších verzí Microsoft Office. Umožňuje prohlížení a základní úpravy DICOM snímků včetně metadat. V režimu prohlížení nabízí úpravu jasu a kontrastu, měření úhlů, vzdáleností, vyznačování oblastí zájmu a vkládání pomocných textů a šipek – víceméně ty funkce, na které jsou radiologové zvyklí [112].

Do nového formátu lze převádět každý snímek samostatně nebo po celých složkách. Na výběr je formáty JPG s nastavitelnou kvalitou 0–100 % a dále PNG, BMP, TIF, GIF. Výstupní snímky se

ukládají v RGB s hloubkou 8 bitů na kanál, tedy celkem 24 bitů na zakódování barevné informace jednoho pixelu.

Při dávkovém převádění snímků z FNHK program pracoval rychle a stabilně, chyba nastala pouze u 7 snímků, které program uložil jako „prázdné“ – vyplněné pouze bílou barvou.

Nakonec se ukázalo, že v průniku množin snímků, které nešly převést ani jedním nástrojem, se nacházejí tři snímky. Na vině je nejspíš poškozený výchozí DICOM soubor. Snímky byly z dalšího zpracování vynechány bez pokusu o přesné hledání příčiny.

7.2.3 Analýza a předzpracování snímků

7.2.3.1 Okrajové případy

Příprava datasetu začala základní vizuální kontrolou všech snímků. V této fázi bylo úplně vyřazeno 12 snímků, které způsobem polohy ruky a prstů na ní neodpovídaly klasické PA projekci. Šlo o případy, kde vlivem omezené pohyblivosti kvůli pokročilému stádiu revmatického onemocnění docházelo k tak extrémním překryvům prstů, že snímky ztratily diagnostickou kvalitu i pro radiology. Některé z těchto snímků byly navíc pořízeny v provizorních podmínkách u ležícího pacienta a zachycují ruku v nestandardní projekci. Tyto snímky byly vyřazeny a do výsledného datasetu nezahrnuty. Pokud by se jich časem sešlo více, řádově alespoň desítky až stovky, mohly by být zařazeny v samostatné třídě. Vytrénovaný model by v takovém případě dokázal ohlásit, že snímek nedokáže klasifikovat a musí být posouzen lidským odborníkem.

Ponechány byly snímky s pouze částečnými překryvy, tedy nejružnější případy luxace, subluxace, deviace, desaxace, deformity typu labutí šíje, kladívka či boutonnière – tedy typické projevy onemocnění popsané v kapitole 3.4. Protože nešlo o extrémní případy, měla by se s nimi neuronová síť být schopná vypořádat.

Dále byl vyřazen jeden snímek vzniklý spojením dvou snímků jedné ruky do jednoho. Polovina snímku, zachycující ruku v nepoužitelné projekci, byla odstraněna. Druhá polovina s rukou v běžné PA projekci byla zařazena zpátky do vznikajícího datasetu.

Na snímcích artritických i artrotických rukou se objevují také cizí předměty, jako nesundatelné prsteny, náramkové hodinky, u hospitalizovaných pacientů navíc identifikační jmenovky na zápěstí nebo zavedená infuze. Ve třech případech se objevily také kloubní protézy a zadrátování kostí. Celkem jsou snímků s cizími předměty řádově desítky, a proto byly ponechány.

7.2.3.2 Obrazová kvalita a provedené úpravy

Z pohledu vlastností a kvality RTG obrazu, které byly uvedené v kapitole 3.2.2, je možné dostupné snímky popsat následovně:

- Svým charakterem jsou data spíše homogenní, tvrdost použitého RTG záření je pro všechny snímky stejná. Snímky zachycují buď obě, levou nebo pravou ruku v PA projekci.

Ne vždy jsou ale ruce orientovány ideálně, v některých případech dochází k natočení a překryvům, především v pokročilých stádiích onemocnění.

- Veškeré snímky již byly pořízeny moderními digitálními metodami. Zhruba 95 % snímků je pořízeno digitalizací nepřímou, zbývajících 5 % už přímou. Diagnostická kvalita výstupů obou metod je v tomto případě téměř totožná [13].
- Prostorové rozlišení snímků se liší v závislosti na použité metodě digitalizace a reálné velikosti ruky, resp. počtu pixelů, na které se její obraz promítne. Některé snímky jsou celé vyplněny rukou, na některých ruka zabírá jen část celkové plochy. Snímky obou rukou mají různá rozlišení v rozmezí od 1976 × 1576 (3,1 MP) až do 2660 × 2041 (5,4 MP).
- Žádný snímek netrpí nadměrným šumem, naopak všechny snímky mají velmi nízkou a konzistentní míru šumu, která nepřekáží diagnostice.
- Neostrosti způsobené pohyby pacienta nejsou přítomny na žádném snímku.
- Téměř všechny snímky jsou opatřeny značkou určující, kde je na snímku pravá (písmeno P nebo R) a kde levá (L) ruka. Na snímcích obou rukou bývá občas tato značka jenom na jedné straně. Značky mají mnoho různých podob softwarových i fyzických.
- Celkem 3 snímky byly ve výchozím stavu otočené o 90 stupňů doprava.
- U snímků se různí jas, kontrast a míra artefaktů vzniklých softwarovým doostřením. Některé snímky mají obecně velmi nízký kontrast zobrazení v celé ploše snímku, řada snímků trpí přexponováním typicky v oblasti zápěstí, které je diagnosticky důležité.

Ze všech čtyř tříd (diagnóz) celkem zhruba 600 snímků prošlo ruční softwarovou úpravou zahrnující doostření, ořezání nebo vyrovnání jasu a kontrastu. U řady snímků byla opravena přexpozice zápěstní oblasti. Ořez byl nutný u snímků, na kterých ruka v poměru k rozměrům snímku zabírala zbytečně málo místa, a naopak velkou část plochy vyplňovalo tmavé pozadí bez obrazové informace. Při zmenšení snímků na vstupní velikost potřebnou k trénování neuronové sítě bude už tak docházet k obrovským ztrátám obrazové informace a bez oříznutí by docházelo k ještě větším. Taktéž byly oříznuty snímky zachycující zbytečně velkou část předloktí. Popsané softwarové úpravy se snažily poučit z výstupů soutěže RSNA Bone Age Challenge, a proto nebylo cílem podobu snímků úplně sjednotit (jak to udělal tým Michiganské univerzity v kapitole 6.1.5), ale pouze je k sobě vzhledově přiblížit (což více odpovídá vítěznému řešení týmu 16 Bit).

Tabulka 5: Počty snímků po základním předzpracování

třída	počet snímků obou rukou	počet snímků jedné ruky	počet snímků celkem
NEART	0	646	646
OA	251	15	266
PsA	213	74	287
RA	717	6	723

7.2.4 Dataset snímků celých rukou

Pro tvorbu datasetu bylo potřeba překonat ještě jednu zásadní překážku – sjednotit, co snímky vyobrazují. U revmatologických pacientů převažovaly snímky zobrazující obě ruce, zbytek snímků a všechny snímky ve třídě NEART zachycovaly buďto pravou nebo levou ruku. I když pro radiology toto nehraje roli, z pohledu neuronové sítě to jsou rozdílné druhy snímků. Formát proto bylo potřeba sjednotit – už jen kvůli tomu, aby bylo možné analyzovat každou ruku nezávisle. Opět však po vzoru výstupů z RSNA Bone Age Challenge nebylo cílem dosáhnout naprosto jednotné podoby, ale ponechat určitou variabilitu.

Protože ruční zpracování by bylo zbytečně zdlouhavé, byla i na rozřezání snímků použita neuronová síť. Ručně byla setříděna pouze přibližně třetina celkového počtu snímků. S využitím přeneseného učení formou extrakce rysů byl s architekturou Inception-V3 předtrénovaný na datasetu ImageNet vytrénován jednoduchý model na klasifikaci, zda snímek zobrazuje levou, pravou nebo obě ruce. Stejně softwarové řešení [113] bylo použito i k následné klasifikaci diagnóz a je detailněji popsáno v kapitole 7.3.2. V jazyce Python byl implementován program, který s využitím vytrénovaného modelu zpracoval zbytek snímků. Konkrétně program pracoval takto:

- Snímky pravých rukou nechal beze změny.
- Snímky levých rukou zrcadlově převrátil na pravé.
- Snímky obou rukou rozřezal na pravé a levé, následně levé zrcadlově převrátil.
- Pro lepší orientaci do názvu snímku před uložením doplnil jednoslovné označení, jakou ruku snímek původně zachycoval.

Vytrénovaný model klasifikoval správně všechny snímky obou rukou. V jednom případě nesprávně zařadil do této třídy i snímek pravé ruky, ve kterém našel známky symetrie – ruka měla osu prostředníčku téměř přesně svislou a umístěnou v geometrickém středu snímku. U snímků pravých / levých rukou bylo nesprávně klasifikovaných snímků zhruba okolo 20 %, tyto případy byly dotříděny ručně. Použití neuronové sítě celý proces třídění výrazně urychlilo.

V této fázi, po sjednocení formátu všech snímků, byla provedena ještě závěrečná kontrola a dotřídění. Ruce s diagnózou RA byly pro interní potřeby ještě ručně rozděleny do dvou skupin podle počtu pozorovatelných znaků charakteristických pro toto onemocnění. Konečné počty snímků ukazuje následující tabulka.

Tabulka 6: Počty snímků v datasetu rukou

třída	počet snímků celkem	trénovací	validační	testovací
NEART	633	505	64	64
OA	525	419	53	53
PsA	500	400	50	50
RA	1433	1147	143	143

V součtu tedy dataset tvoří 3091 snímků nerovnoměrně rozdělených do čtyř tříd. Počet snímků se ještě zvyšoval při nasazení metod rozšíření dat v závislosti na konkrétním experimentu. Tyto následné úpravy datasetu jsou shrnuty vždy v popisu příslušného experimentu.

7.2.5 Ověřovací skupina snímků rukou z externích zdrojů

V souladu se zmiňovanými doporučeními RSNA (kapitola 6.1.1) byl připraven ještě nezávislý dataset snímků rukou z jiných zdrojů, který pomůže při závěrečném vyhodnocení. Vytrénované modely by měly být vyzkoušeny na testovací části výše uvedeného datasetu a poté ještě na tomto ověřovacím externím datasetu. Snímky z jiných zdrojů pořízené na jiných přístrojích, s jinými parametry a na jiném vzorku pacientů podle RSNA pomáhají otestovat, zda je vytrénovaný model dostatečně robustní a zda má schopnost generalizace [49].

Naneštěstí snímků rukou s revmatickými onemocněními není ve volně dostupných zdrojích tolik, aby mělo smysl provádět statistické vyhodnocení modelu. Proto bude dataset z externích zdrojů v tomto případě sloužit čistě jen k demonstraci funkčnosti modelu. Příléhavější označení je proto spíše ověřovací skupina než dataset.

Skupinu tvoří 15 ukázkových případů [18–20, 35, 114, 115] z Radiopedie – odborné internetové radiologické příručky tvořené radiology a jinými odborníky. Stejně snímky byly použity na mnoha místech v textu této práce jako ilustrační. Zastoupeny jsou 2 ruce s psoriatickou artritidou, 2 s osteoartrózou, 6 s revmatoidní artritidou a 5 bez příznaků. Onemocnění jsou spíše v pokročilejší fázi a mají tedy poměrně výrazné a zřetelně rozpoznatelné projevy.

Dále bylo přidáno 10 náhodně vybraných snímků rukou trpících osteoartrózou z datasetu OAI. Vzhledem k nízké obrazové kvalitě datasetu se výběr omezil na snímky spíše ve vyšším rozlišení a méně degradované ztrátovou kompresí.

Poslední přírůstek tvoří 10 snímků zdravých rukou z datasetu RSNA Bone Age. Snímky byly náhodně vybrány z podmnožiny nejstarších pacientů (18–19 let) s již plně vyvinutými kostmi.

Snímky ze všech tří uvedených zdrojů se obrazovou i technickou kvalitou vzájemně velmi odlišovaly a před klasifikací pomocí vytrénovaných modelů byly předzpracovány totožným způsobem, jako všechny ostatní snímky rukou z datasetu FNHK, tedy rozřezáním na pravé a levé, převrácením na pravé, ořezem přebytečného místa a vyrovnáním jasu a kontrastu.

Tabulka 7: Ověřovací skupina snímků rukou z externích zdrojů

třída	počet snímků celkem	RSNA Bone Age	OAI	Radiopaedia
NEART	16	10	-	6
OA	12	-	10	2
PsA	4	-	-	2
RA	6	-	-	6

7.2.6 Dataset snímků kloubů

Druhý dataset vznikl s cílem umožnit klasifikaci každého kloubu zvlášť. Jeho sestavení obnáší z každého snímku celé ruky vyříznout samostatné snímky jednotlivých kloubů. Klouby palce nejsou brány v potaz [13], protože palec na PA projekci ruky je oproti ostatním prstům zobrazen v natočené poloze. Přínosné naopak jsou zápěstní klouby. Celkem lze tedy ze snímku jedné ruky získat až 13 dalších snímků: 4 klouby DIP, 4 klouby PIP, 4 klouby MCP a jeden snímek zachycující klouby UC a RC.

Dataset kloubů byl odvozen z výše uvedeného datasetu snímků celých rukou, tedy ze snímků již předzpracovaných, rozřezaných na pravé a levé a převrácených na stejnou stranu. Důvody, proč nebyla k detekci kloubů použita neuronová síť již byly uvedeny. Nástroje pro pouhou anotaci snímků (tedy vyznačení kloubů) jako například MicroDicom neměly ovládání přizpůsobené pro zpracování velkého množství snímků, takže by práce byla až příliš zdlouhavá.

Proto byl v jazyce Python vytvořen program umožňující snadnou tvorbu výřezů kloubů. Program byl navržen s důrazem na jednoduchost ovládání a co nejrychlejší a nejintuitivnější zpracování snímků. Kliknutím myši uživatel označí geometrický střed kloubu, kolečkem myši lze zvětšit či zmenšit velikost výřezu. Výchozí velikosti jsou 299×299 px pro klouby MCP, PIP, DIP a 500×500 px pro radiokarpální kloub. Každý výřez je uložen jako samostatný soubor ve formátu .png a souřadnice kloubu ve strojově čitelné podobě v samostatném souboru, takže pro případné využití v mimo rámec této diplomové práce již nebude potřeba procházet všechny snímky znovu.

Tvorba datasetu byla časově značně náročná. Zpracování jednoho snímku neartritické ruky, tedy ruky s jasně definovanými klouby, zabralo v průměru okolo 40 vteřin, tempo bylo zhruba 80 snímků za hodinu. U rukou s deformovanými klouby zabralo zpracování jednoho snímku zhruba minutu, protože nebylo snadné střed kloubu lokalizovat. Celkem zabrala pouhá tvorba výřezů kloubů zhruba týden čistého času a další tři dny programování a úpravy aplikace na tvorbu výřezů.



Obr. 41: Program na označování a tvorbu výřezů kloubů

Zdroj: vlastní zpracování s využitím snímku [18]

Počet snímků ve výsledném datasetu neodpovídá třináctinásobku počtu snímků rukou. Mnoho DIP a PIP kloubů vlivem ohybu prstů splývalo do jednoho nebo se překrývaly se sousedními prsty, což lze vidět i na obrázku výše, kde byl kvůli překryvům vynechán kloub DIP-4. Vynechány byly také tři případy kloubních protéz a zadrátování kostí. U kloubů zápěstí byly vynechány případy, kdy došlo k překrytí kloubu kapačkami nebo nemocničními jmenovkami. Aby byl dataset vyváženější (a také kvůli nesmírné časové náročnosti tvorby výřezů) bylo z rukou trpících revmatoidní artritidou zahrnuto pouze 1805 snímků rukou se třemi znaky onemocnění. Zbývajících 535 snímků s jedním až dvěma znaky použito nebylo. Celkové počty snímků v datasetu ukazuje následující tabulka.

Tabulka 8: Dataset snímků kloubů

třída	počet snímků celkem	trénovací	validační	testovací
NEART	8407	7575	416	416
OA	6824	6148	338	338
PsA	6840	5838	321	321
RA	11671	10511	580	580

7.2.7 Dataset snímků kloubů tříděný podle typu změn

Při trénování modelů s výše uvedeným datasetem kloubů se ukázala jedna zásadní nevýhoda. Průměrný snímek ruky s určitou diagnózou má každý kloub v jiném stupni poškození, některé mohou být poškozeny výrazně, jiné mírně nebo téměř vůbec – takové klouby jsou samy o sobě těžko klasifikovatelné. Proto byla s poměrně velkou časovou investicí konzultanta práce část datasetu kloubů přetříděna podle typu pozorovaných změn. Opět s použitím přeneseného učení formou extrakce rysů a následným ručním dotříděním bylo celkem 7348 snímků jednotlivých kloubů rozříděno do čtyř hlavních skupin podle projevů artritických (trpících revmatoidní či psoriatickou artritidou), artrotických (osteoartróza), s mírnými (těžko specifikovatelnými) projevy a kloubů bez projevů revmatických změn. Nevyvážené počty kategorií jsou dány různou četností poškození kloubů ve výchozích snímcích rukou.

Tabulka 9: Dataset snímků kloubů tříděný podle typu změn

třída	počet snímků celkem	trénovací	validační	testovací
ARTRITICKÉ	1080	864	108	108
ARTROTICKÉ	2158	1726	216	216
NEGATIVNÍ	2378	1902	238	238
NEJISTÉ	1448	1156	146	146

7.3 Trénování modelů a jejich vyhodnocení

7.3.1 Použitý hardware

Centrum základního a aplikovaného výzkumu Fakulty informatiky a managementu ochotně poskytlo pro tvorbu této práce přístup k počítači vyhrazenému pro hluboké učení a obecně náročnější výpočty. Počítač disponoval dvojicí výkonných grafických karet s platformou Nvidia CUDA, a ačkoliv ani zdaleka nedosahoval parametrů, na kterých se trénovaly modely pro ILSVRC (kapitola 5.5), výrazně převyšoval možnosti jiných počítačů na fakultě a pro účely této práce víceméně dostačoval. Trénování modelů probíhalo na jedné nebo na obou grafických kartách a případně i na procesoru.

Tabulka 10: Parametry použitého hardwaru

CPU	Intel (R) Core™ i7-8700 3.20 GHz (6×CORE)	
RAM	16 GB X 4 (2666 Mhz) CL13	
GPU	GeForce RTX 2080 TI	GeForce GTX 1080 TI
paměť [GB]	11,019	11,176
počet CUDA jader (shaderů)	4352	3584
frekvence [MHz]	1545	1607

7.3.2 Použitý software

Software využitý pro předzpracování dat a trénování modelů hlubokého učení byl implementován v jazyku Python 3.6 s využitím frameworků TensorFlow 1.12 pro neuronové sítě, OpenCV 4.1 a Pillow 5.3 pro úpravy obrazu a implementaci metod rozšíření, PyCM 2.6 pro výpočet metrik na základě matice záměn, XlsxWriter 1.2 pro ukládání výstupů do Excelu. Většina napsaného kódu se týkala práce s daty – šlo o programy na kopírování a přejmenování souborů, export popisků snímků v datasetu do formátu CSV, geometrické transformace a úpravu barev nebo již zmíněné vyznačení souřadnic kloubů ve snímcích rukou a následnou automatickou tvorbu jejich výřezů.

Pro přenesené učení formou extrakce rysů bylo základem řešení poskytované Googlem [113], které umožňuje využít předtrénovaných modelů a obsahuje obslužný boilerplate kód k jejich dotrénování a nasazení. Toto řešení bylo pro potřeby práce upraveno a rozšířeno.

Pro trénování modelu od základů a pro přenesené učení formou ladění byla využita větev TensorFlow nazývaná TF-Slim [116], která pracuje na vyšší úrovni abstrakce než běžné TensorFlow a v době svého vzniku byla jakýmsi mezistupněm mezi TensorFlow a Kerasem.

Všechny programy byly spouštěny v Docker kontejnerech (verze Nvidia-Docker, která oproti normálnímu Dockeru umožňuje využívat i GPU).

7.3.3 Rozšíření dat

Metody rozšíření dat byly s ohledem na množství snímků nezbytné. Rozšíření bylo implementováno ve variantě offline. Rozšiřování za běhu (online) v tomto případě nedávalo smysl, protože provedených experimentů bylo mnoho, avšak hlavní datasety zůstávaly stejné – bylo tedy výhodnější věnovat strojový čas na předzpracování snímků jednorázově, než jím plýtvat na opakované rozšiřování dat při každém trénování. Navíc implementace offline dává plnou kontrolu nad upravenými snímky, umožňuje vyzkoušet libovolné kombinace úprav a například také porovnat vliv různých způsobů rozšíření snímků na čtvercový formát.

S využitím knihoven OpenCV a Pillow byl v jazyce Python napsán program, který z každého snímku umožňuje vytvořit několik odvozenin:

- Až 4 odvozeniny vzniklé otočením o náhodný úhel v rozmezí nejvýše ± 20 stupňů. Po otočení zároveň dochází k náhodnému ořezu v rozsahu nejvýše 3 % z každé strany.
- Až 2 odvozeniny vzniklé posunutím o 20 % z šířky snímku doleva či doprava.
- 1 odvozenina vzniklá zrcadlovým překlopením podél svislé osy.
- 1 odvozenina vzniklá pružnou deformací.

Uvedené 4 transformace bylo možné doplnit o posun jasu v rozmezí ± 20 %. Stejný program sloužil také k převodu snímků na čtvercový formát doplněním prázdného (černého) prostoru po stranách a k převzorkování snímků na rozměry, které na vstupu neuronové sítě vyžadují. Ukázalo se, že podpora interpolace je v OpenCV zakompilována chybně, a na Windows tyto funkce nelze využít. OpenCV na Linuxu Ubuntu pro změnu špatně podporovalo zobrazení obrázku v okně a využívalo jiné mapování kláves na klávesnici. Proto byla pro interpolaci a převzorkování nakonec použita knihovna Pillow. Byla použita nejkvalitnější dostupná interpolační metoda, v dokumentaci nazývaná *antialias* [117].

7.3.4 Vytrénování modelu od základů

Nejpřímočařejší cestou k vytrénování klasifikačního modelu je trénování od základů bez využití přeneseného učení. Tato cesta byla prověřena, ačkoliv vzhledem k počtu snímků neměla velkou naději na úspěch, což se také potvrdilo.

Byl využit již zmíněný balíček TF-Slim [116], který nabízel výběr z mnoha druhů architektur sítí a uměl pro trénování využít obě grafické karty i procesor současně.

Provedeno bylo mnoho experimentů s různými kombinacemi rozšíření dat a parametrů. Nejúspěšnější trénování probíhala s parametry: velikost dávky 64 (větší nebyla kvůli omezené paměti GPU možná), počet paralelních vláken 2, optimizér Adam, koeficient učení 0,001. V každém experimentu bylo provedeno alespoň 30 000 kroků, v nadějnějších případech i 100 000. Přes všechny snahy však žádné trénování nekonvergovalo k výsledku a končilo nedoučením a oscilací celkové správnosti kolem určité a poměrně nízké hodnoty. Výsledky nejúspěšnějších pokusů ukazuje tabulka na následující straně.

Tabulka 11: Trénování na datasetech rukou a kloubů od základů

dataset	třídy	rozšíření dat	architektura sítě	vstupní rozměr	celková správnost
ruce	NEART, OA, PsA, RA	otočení, posun, zrcadlení	Inception-ResNet-V2	299 px	34,5 %
ruce	NEART, OA, PsA, RA	otočení, posun, zrcadlení	Inception-V3	299 px	49,3 %
ruce	NEART, OA, PsA, RA	otočení, posun, zrcadlení	Inception-V4	299 px	31,1 %
klouby	NEART, OA, PsA, RA	-	Inception-ResNet-V2	299 px	32,7 %
klouby	NEART, OA, PsA, RA	-	Inception-V3	299 px	37,6 %
klouby	NEART, OA, PsA, RA	-	Inception-V4	299 px	29,3 %

Trénování od základů ukázalo, že neuronové sítě mají problém se v datech zorientovat, a i přes veškerou snahu nevedlo k úspěchu. Z tabulky je však patrné, že architektura Inception-V3 si vedle lépe než její pokročilejší a složitější nástupci Inception-V4 a Inception-ResNet-V2. Varianta trénování od základů byla opuštěna.

7.3.5 Přenesené učení formou extrakce rysů

Nejjednodušší cestou k vytrénování modelu se zdálo být přenesené učení, a to zejména kvůli nevelkému množství dostupných snímků. Google pro přenos formou extrakce rysů nabízí sítě architektury Inception-V3 a MobileNet předtrénované na datasetu ImageNet, a to jako téměř hotové řešení včetně základního obslužného boilerplate kódu [113], který dokáže spustit trénování modelu a následně model nasadit ke klasifikaci snímků z testovací části datasetu. Pro účely této práce byl kód obstarávající trénování rozšířen o metody pro automatické ukládání parametrů trénování do souboru a kód pro nasazení modelu byl upraven pro provádění opakovaných experimentů a zpracování všech souborů ve složce s automatickým zápisem výstupů do tabulky ve formátu Excel.

Uvedené řešení funguje tak, že z původního modelu odebere poslední, plně propojenou vrstvu neuronů, v terminologii TensorFlow nazývanou *bottleneck*. Zbyde tzv. modul na extrakci rysů obrázku, ke kterému je následně přidán a dotrénován nový bottleneck. Při trénování se nejvíce osvědčily tyto parametry: počet kroků 4000, koeficient učení 0,01, velikost dávky 128, velikost validační dávky –1 (využije najednou všechny snímky z validační podmnožiny).

Bylo provedeno mnoho experimentů s mírně odlišným nastavením, různými datasety a různými způsoby rozšíření dat, ale celková správnost vytrénovaného modelu se pohybovala většinou v rozmezí 70–80 % a hodnotu 90 % přesáhla pouze ve výjimečných a velmi zjednodušených případech, jak přibližuje text na následující straně.

7.3.5.1 Rozlišení diagnóz na datasetu rukou

Přenos učení formou extrakce rysů na datasetu rukou nebyl příliš úspěšný. Jak dokládá tabulka, při klasifikaci všech čtyř tříd dosáhl model celkové správnosti pouze 75,8 %. Při sloučení nebo vynechání některých tříd bylo dosaženo vyšší správnosti, ovšem vzhledem k výraznému zjednodušení úlohy nemá smysl zabývat se podrobným vyhodnocováním pomocí metrik. Za neúspěchem stojí patrně velmi malá podobnost původního datasetu ImageNet s datasetem RTG snímků. Výsledky však ukazují, že neuronová síť příznaky revmatických onemocnění dokáže mezi snímky bez příznaků rozlišit – tedy že úloha by mohla být s pokročilejšími technikami řešitelná.

Tabulka 12: Přenesené učení formou extrakce rysů na datasetu rukou

dataset	třídy	rozšíření dat	architektura sítě	vstupní rozměr	celková správnost
ruce	NEART, OA, PsA, RA	otočení, posun, zrcadlení	Inception-V3	299 px	75,8 %
ruce	NEART, (OA+PsA+RA)	otočení, posun, zrcadlení	Inception-V3	299 px	89 %
ruce	NEART, RA	-	Inception-V3	299 px	91,9 %

7.3.5.2 Rozlišení diagnóz na datasetu kloubů

Přenos učení formou extrakce rysů na datasetu kloubů vykazoval podobnou celkovou správnost jako v případě výše zmíněného datasetu rukou. Jak ukazuje tabulka níže, nejvyšší dosažená správnost byla 72,4 %.

Jedním z důvodů může být ztráta informace vlivem interpolace a převzorkování, výřezy kloubů prstů mají totiž velmi různá rozlišení nejčastěji od zhruba 145 × 145 px až 260 × 260 px (klouby zápěstí okolo 400 × 400 px), což znamená, že před vstupem do neuronové sítě dochází ke zvětšení snímků na vstupní rozměr a tím pádem k nežádoucímu zkreslení a vzniku artefaktů. Míra zkreslení je pro každý snímek jiný. Z toho důvodu byla vyzkoušena síť jednodušší architektury MobileNet, která přijímá snímky v nižším rozlišení. Síť MobileNet vykazovala nejvyšší celkovou přesnost při rozlišení 128 × 128 px – v tomto případě u žádného snímku nedošlo ke zkreslujícímu zvětšení, všechny byly pouze zmenšeny. Experimenty se sítí MobileNet ukázaly, že u datasetu kloubů hraje zvětšení snímků důležitou a negativní roli.

Posledním provedeným experimentem bylo ještě otestování architektury Inception-V3 na datasetu kloubů se sloučenými třídami PsA a RA, úlohou tedy bylo rozlišení změn artritických a artrotických. Experiment skončil s již lepší, ale stále velmi nízkou celkovou správností 77,0 %.

Z důvodu nízké úspěšnosti nebylo provedeno podrobné vyhodnocení modelů a dosažené výsledky byly pouze shrnuty do následující tabulky. Každý experiment proběhl několikrát s drobnými úpravami v nastavení a uváděn je vždy nejlepší dosažený výsledek.

Tabulka 13: Přenesené učení formou extrakce rysů na datasetu kloubů

dataset	třídy	rozšíření dat	architektura sítě	vstupní rozměr	celková správnost
klouby	NEART, OA, PsA, RA	zrcadlení	Inception-V3	299 px	72,4 %
klouby	NEART, OA, PsA, RA	-	Inception-V3	299 px	71,6 %
klouby	NEART, OA, PsA, RA	-	MobileNet 1.0	160 px	52,3 %
klouby	NEART, OA, PsA, RA	-	MobileNet 1.0	224 px	57,5 %
klouby	NEART, OA, PsA, RA	-	MobileNet 1.0	128 px	73,5 %
klouby	NEART, OA, (PsA+RA)	-	Inception-V3	299 px	77,0 %

Závěrem z této fáze je, že přenesené učení formou extrakce rysů obrázků se pro řešené úlohy neosvědčilo, na vině je v případě rukou i kloubů patrně jednak malá podobnost datasetu ImageNet s použitými daty a v případě kloubů pak navíc ještě problematické převzorkování snímků s nízkým rozlišením na větší rozměry.

7.3.6 Přenesené učení formou ladění

Přenesené učení formou ladění na rozdíl od extrakce rysů umožňuje přizpůsobit novým datům i jiné vrstvy než jen horní plně propojené. Vzhledem k tomu, že trénování od základů selhalo a přenos učení formou extrakce rysů dal výsledky neuspokojivé, ale nadějně, měl přenos formou ladění největší šanci na úspěch. Přenesené učení formou ladění bylo implementováno stejně jako trénování od základů byly pomocí balíčku TF-Slim [116] s rozdílem, že v tomto případě namísto vytváření nových modelů byly využity modely již předtrénované na datasetu ImageNet. Stejně jako v případě trénování od základů byly vyzkoušeny architektury Inception-V3, Inception-V4, Inception-ResNet-V2, dále síť VGG a různé varianty sítě ResNet. Oproti předchozím dvěma způsobům učení tato metoda již byla úspěšnější.

7.3.6.1 Shrnutí provedených experimentů

Rozlišení diagnóz na datasetu rukou

Nejlepších výsledků bylo dosaženo s kombinací modelu využívajícího architekturu ResNet-50, velikostí snímků 224×224 px a rozšířením dat formou otočení, úprav jasu a kontrastu a zrcadlovým převrácením. Celkem byl tímto způsobem počet trénovacích snímků rozšířen na čtyřnásobek. Váhy všech plně propojených vrstev byly znovu vytrénovány, váhy v ostatních vrstvách byly odemčeny k úpravám. Nejlepších výsledků bylo dosaženo s nastavením koeficientu učení na 0,001, velikostí dávky na 64 a optimizéru na SGD.

Po nalezení schůdné cesty byl model na rozlišení jednotlivých diagnóz vytrénován celkem třikrát. Dvakrát naprosto stejným způsobem, potřetí se úpravou vstupních snímků na čtvercový formát – zatímco pro první dva modely byla vstupní data připravena se zachováním původního

poměru stran, data pro třetí model byla dodáním černých okrajů rozšířena na poměr 1:1, aby při vstupu dat do neuronové sítě nedocházelo k deformaci obrazu natažením. Ukázalo se, že tato úprava má na většinu metrik minimální vliv, dokonce si tento model vedl hůře než předchozí modely s nataženými snímky, a to jak na testovacích datech, tak na ověřovací skupině z externích zdrojů. Modely byly vyhodnoceny na testovacích datech a výsledky byly shrnuty do matic záměn, ze kterých byly následně vypočítány metriky uvedené v tabulkách.

Výsledky vyhodnocení trojice modelů na klasifikaci jednotlivých diagnóz jsou pro přehlednost uvedeny na [následujících stranách](#) 88 až 92 pro každý model zvlášť.

Rozlišení pouze artritických a artrotických změn na datasetu rukou

Protože se modely na klasifikaci všech čtyř tříd osvědčily, posledním experimentem se snímky rukou bylo ještě trénování se sloučenými třídami RA a PsA – tedy schopnost rozlišit obecně pouze projevy artritické a artrotické (a bez příznaků). Teoreticky by mohl model dosáhnout ze všech experimentů nejvyšší přesnosti, což se také potvrdilo. Výsledky jsou pro přehlednost opět na [samostatné straně č. 93](#).

Rozlišení diagnóz a rozlišení artritických a artrotických změn na datasetu kloubů

Obdobným způsobem jako na datasetu rukou bylo ladění použito i na dataset kloubů. I zde se osvědčila síť ResNet. Jako jediná forma rozšíření dat bylo použito zrcadlení podél svislé osy. Výsledky byl opět shrnuty na [samostatné straně č. 94](#) a oproti očekávání nedopadly tak dobře, což bylo způsobené pravděpodobně tím, že (jak již bylo uvedeno v kapitole zabývající se tvorbou datasetu kloubů) průměrný snímek ruky s určitou diagnózou má každý kloub v jiném stupni poškození, některé mohou být poškozeny výrazně, jiné mírně nebo téměř vůbec – takové klouby jsou samy o sobě těžko klasifikovatelné. Proto byl vytvořen [dataset snímků kloubů tříděný podle typu změn](#) a s ním byl učiněn další experiment. Ten však vzhledem k velkému počtu kloubů s mírnými či nejistými projevy taktéž nedopadl příliš úspěšně, jak [dokládají výsledky na stranách 95 a 96](#). V posledním experimentu byla tedy třída kloubů s nejasnými projevy vyřazena, čímž se výsledky zlepšily. Všechny experimenty s klouby vedou k závěru, že ve výzkumu navazujícím na tuto práci by bylo vhodné věnovat pozornost ještě dodatečnému pečlivému rozčlenění snímků kloubů na více vzájemně odlišitelných podskupin a návrh nových experimentů.

Poznámka k problému černé skříňky

Pro lepší vyhodnocení modelů by byla vhodná implementace teplotních map. Ty lze bez větších překážek realizovat ve frameworku Keras nebo novějších verzích TensorFlow (s pomocí rozšiřujících balíčků), ne však v použité verzi Tensorflow 1.12. Upgrade na novou verzi by spolu s nutností upgradovat také verzi jazyka Python znamenal kompletní přepsání téměř všech zdrojových kódů. Proto byl problém černé skříňky řešen především podrobným vyhodnocením modelů pomocí metrik a vyzkoušením na ověřovací skupině snímků z externích zdrojů.

7.3.6.2 Rozlišení diagnóz na datasetu rukou

Vyhodnocení modelu 1

Model 1		Předpovědi			
		RA	OA	PsA	NEART
Skutečnost	RA	139	0	4	0
	OA	8	35	10	0
	PsA	7	3	40	0
	NEART	4	1	2	57

RA		Poz.	Neg.
Poz.	139	4	
Neg.	19	148	

OA		Poz.	Neg.
Poz.	35	18	
Neg.	4	253	

PsA		Poz.	Neg.
Poz.	40	10	
Neg.	16	244	

NEART		Poz.	Neg.
Poz.	57	7	
Neg.	0	246	

- Model 1 ze všech 310 snímků v testovacím datasetu správně klasifikoval revmatoidní artritidu u 139 snímků, 4 snímky do této třídy nezařadil, i když do ní zařazeny být měly. Naopak předpověděl revmatoidní artritidu u 19 snímků, které ve skutečnosti patří do jiných tříd.
- Model správně klasifikoval osteoartrózu u 35 snímků, chybně do této třídy 18 snímků nezařadil. Naopak chybně zařadil 4 snímky, které ve skutečnosti patří do jiných tříd.
- Model správně klasifikoval psoriatickou artritidu u 40 snímků, chybně do této třídy 10 snímků nezařadil. Naopak chybně předpověděl psoriatickou artritidu u 16 snímků.
- Model správně klasifikoval 57 snímků bez příznaků revmatoidních onemocnění, chybně do této třídy 7 snímků nezařadil. Žádný snímek ruky bez příznaků nezařadil do nesprávné třídy.









Tabulka 14: Metriky pro model 1 na klasifikaci snímků ruky

	charakteristiky pro jednotlivé třídy				charakteristiky pro celý model	
	RA	OA	PsA	NEART	mikro	makro
celková správnost	0,926	0,929	0,916	0,977	0,874	0,937
přesnost	0,880	0,897	0,714	1,000	0,874	0,873
senzitivita	0,972	0,660	0,800	0,891	0,874	0,831
specifická	0,886	0,984	0,938	1,000	0,958	0,952
F1-skóre	0,924	0,761	0,755	0,942	0,874	0,851

Metriky odvozené z matic ukazují, že z jednotlivých tříd si model nejhůře poradil s psoriatickou artritidou (celková správnost 91,6 %, F1-skóre 75,5 %). Toto onemocnění má ze všech tří nejvíce variabilní projevy, takže nejhorší výsledek je v souladu s očekáváním. Dále je vidět, že model měl nesnáze také s klasifikací osteoartrózy, kde dosáhl senzitivity pouze 66 %. Model dosáhl celkové správnosti přes 90 % pro všechny čtyři třídy. Celková správnost celého modelu se zohledněním četností jednotlivých tříd je 87,4 %, bez zohlednění četností (aritmetický průměr) pak 93,7 %.

Z ověřovací skupiny snímků z externích zdrojů model správně klasifikoval 32 z 35 případů (91,4 %). Chybně klasifikoval 2 případy rukou ve skutečnosti trpících osteoartrózou a jeden případ ruky ve skutečnosti trpící psoriatickou artritidou. Osm vybraných výstupů modelu bylo na ukázkou sestaveno do následující tabulky.

Tabulka 15: Ukázka výstupů modelu 1 na vybraných snímcích z externích zdrojů

	<p>Skutečnost: OA</p> <p>Předpovědi: PA 0,4156 RA 0,2705 OA 0,1803 NEART 0,1372</p> <p>Zdroj snímku: dataset OAI [84]</p>		<p>Skutečnost: OA</p> <p>Předpovědi: OA 0,7450 RA 0,1176 NEART 0,0823 PA 0,0549</p> <p>Zdroj snímku: Radiopaedia [20]</p>
	<p>Skutečnost: PA</p> <p>Předpovědi: PA 0,4784 RA 0,3725 NEART 0,0901 OA 0,0627</p> <p>Zdroj snímku: Radiopaedia [19]</p>		<p>Skutečnost: PA</p> <p>Předpovědi: RA 0,6156 PA 0,2078 NEART 0,1137 OA 0,0705</p> <p>Zdroj snímku: Radiopaedia [19]</p>
	<p>Skutečnost: RA</p> <p>Předpovědi: RA 0,8941 NEART 0,0509 PA 0,0352 OA 0,0274</p> <p>Zdroj snímku: Radiopaedia [114]</p>		<p>Skutečnost: RA</p> <p>Předpovědi: RA 0,9176 NEART 0,0274 PA 0,0274 OA 0,0274</p> <p>Zdroj snímku: Radiopaedia [18]</p>
	<p>Skutečnost: bez příznaků</p> <p>Předpovědi: NEART 0,5960 OA 0,2352 RA 0,1019 PA 0,0666</p> <p>Zdroj snímku: Radiopaedia [115]</p>		<p>Skutečnost: bez příznaků</p> <p>Předpovědi: NEART 0,4000 RA 0,3411 PA 0,1882 OA 0,0745</p> <p>Zdroj snímku: RSNA Bone Age [92]</p>

Vyhodnocení modelu 2 (obdoba předchozího modelu)

Tento model je uváděn z důvodu, že jeho výsledky byly velmi blízké modelu 1 a nebylo možné jednoznačně určit, který z nich je lepší.

Model 2		Předpovědi			
		RA	OA	PsA	NEART
Skutečnost	RA	134	1	6	2
	OA	6	39	5	3
	PsA	2	4	41	3
	NEART	2	1	1	60

RA	Poz.	Neg.
Poz.	134	9
Neg.	10	157

OA	Poz.	Neg.
Poz.	39	14
Neg.	6	251

PsA	Poz.	Neg.
Poz.	41	9
Neg.	12	248

NEART	Poz.	Neg.
Poz.	60	4
Neg.	8	238

- Model 1 ze všech 310 snímků v testovacím datasetu správně klasifikoval revmatoidní artritidu u 134 snímků, 9 snímků do této třídy nezařadil, i když do ní zařazeny být měly. Naopak předpověděl revmatoidní artritidu u 10 snímků, které ve skutečnosti patří do jiných tříd.
- Model správně klasifikoval osteoartrózu u 39 snímků, chybně do této třídy 14 snímků nezařadil. Naopak chybně zařadil 6 snímků, které ve skutečnosti patří do jiných tříd.
- Model správně klasifikoval psoriatickou artritidu u 41 snímků, chybně do této třídy 9 snímků nezařadil. Naopak chybně zařadil 12 snímků, které ve skutečnosti patří do jiných tříd.
- Model správně klasifikoval 60 snímků bez příznaků revmatoidních onemocnění, chybně do této třídy 4 snímky nezařadil. Chybně 8 snímků nemocných rukou označil jako bez příznaků.









Tabulka 16: Metriky pro model 2 na klasifikaci snímků ruky

	charakteristiky pro jednotlivé třídy				charakteristiky pro celý model	
	RA	OA	PsA	NEART	mikro	makro
celková správnost	0,939	0,935	0,932	0,961	0,884	0,942
přesnost	0,931	0,867	0,774	0,882	0,884	0,863
senzitivita	0,937	0,736	0,820	0,938	0,884	0,858
specifická	0,940	0,977	0,954	0,967	0,961	0,960
F1-skóre	0,934	0,796	0,796	0,909	0,884	0,860

Metriky odvozené z matic ukazují, že z jednotlivých tříd si model nejhůře poradil s psoriatickou artritidou (celková správnost 93,2 %, F1 skóre 79,6 %). Stejně jako v případě modelu 1 jde o očekávaný výsledek. Nejnižší senzitivita je opět vidět u osteoartrózy (73,6 %). Model dosáhl celkové správnosti přes 90 % pro všechny čtyři třídy. Celková správnost celého modelu se zohledněním četností jednotlivých tříd je 88,4 %, bez zohlednění četností (aritmetický průměr) pak 94,2 %. Naopak v některých charakteristikách, například pro ruce bez příznaků revmatických onemocnění, je tento model v porovnání s předchozím horší.

Z ověřovací skupiny snímků z externích zdrojů model správně klasifikoval 33 z 35 případů (94,2 %). Chybně klasifikoval obě ruce trpící psoriatickou artritidou. Osm vybraných výstupů modelu bylo na ukázkou sestaveno do následující tabulky.

Tabulka 17: Ukázka výstupů modelu 2 na vybraných snímcích z externích zdrojů

	Skutečnost: OA		Skutečnost: OA
	Předpovědi: OA 0,6235 PA 0,2549 NEART 0,0705 RA 0,0509		Předpovědi: OA 0,8156 RA 0,0705 PA 0,0627 NEART 0,0509
	Zdroj snímku: dataset OAI [84]		Zdroj snímku: Radiopaedia [20]
	Skutečnost: PA		Skutečnost: PA
	Předpovědi: RA 0,5529 PA 0,3529 NEART 0,0588 OA 0,0392		Předpovědi: RA 0,4156 PA 0,3019 OA 0,1921 NEART 0,0901
	Zdroj snímku: Radiopaedia [19]		Zdroj snímku: Radiopaedia [19]
	Skutečnost: RA		Skutečnost: RA
	Předpovědi: RA 0,8901 PA 0,0431 OA 0,0392 NEART 0,0313		Předpovědi: RA 0,8352 OA 0,0823 PA 0,0470 NEART 0,0392
	Zdroj snímku: Radiopaedia [114]		Zdroj snímku: Radiopaedia [18]
	Skutečnost: bez příznaků		Skutečnost: bez příznaků
	Předpovědi: NEART 0,6705 RA 0,2431 PA 0,0509 OA 0,0392		Předpovědi: NEART 0,6745 RA 0,1882 PA 0,0705 OA 0,0666
	Zdroj snímku: Radiopaedia [115]		Zdroj snímku: RSNA Bone Age [92]

Vyhodnocení modelu 3 (vstupní data rozšířena na čtverec doplněním místa po stranách)

Model 3		Předpovědi			
		RA	OA	PsA	NEART
Skutečnost	RA	132	5	5	1
	OA	2	43	5	3
	PsA	1	14	33	2
	NEART	1	0	2	61

RA		Poz.	Neg.
Poz.	132	11	
Neg.	4	163	

OA		Poz.	Neg.
Poz.	43	10	
Neg.	19	238	

PsA		Poz.	Neg.
Poz.	33	17	
Neg.	12	248	

NEART		Poz.	Neg.
Poz.	61	3	
Neg.	6	240	

- Model 1 ze všech 310 snímků v testovacím datasetu správně klasifikoval revmatoidní artritidu u 132 snímků, 11 snímků do této třídy nezařadil, i když do ní zařazeny být měly. Naopak předpověděl revmatoidní artritidu u 4 snímků, které ve skutečnosti patří do jiných tříd.
- Model správně klasifikoval osteoartrózu u 43 snímků, chybně do této třídy 10 snímků nezařadil. Naopak chybně předpověděl osteoartrózu u 19 snímků, které ve skutečnosti patří do jiných tříd.
- Model správně klasifikoval psoriatickou artritidu u 33 snímků, chybně do této třídy 17 snímků nezařadil. Naopak chybně předpověděl psoriatickou artritidu u 12 snímků, které ve skutečnosti patří do jiných tříd.
- Model správně klasifikoval 61 snímků bez příznaků revmatoidních onemocnění, chybně do této třídy 3 snímky nezařadil. Chybně 6 snímků nemocných rukou označil jako bez příznaků.

Tabulka 18: Metriky pro model 3 na klasifikaci snímků ruky

	charakteristiky pro jednotlivé třídy				charakteristiky pro celý model	
	RA	OA	PsA	NEART	mikro	makro
celková správnost	0,952	0,906	0,906	0,971	0,868	0,934
přesnost	0,971	0,694	0,733	0,910	0,868	0,827
senzitivita	0,923	0,811	0,660	0,953	0,868	0,837
specifická	0,976	0,926	0,954	0,976	0,956	0,958
F1-skóre	0,946	0,748	0,695	0,931	0,868	0,832

Metriky odvozené z matic ukazují, že z jednotlivých tříd si model nejhůře poradil s psoriatickou artritidou a osteoartrózou. Nejnižší senzitivita je opět vidět u psoriatické artritidy (66 %). Model dosáhl celkové správnosti přes 90 % pro všechny čtyři třídy. Celková správnost celého modelu se zohledněním četností jednotlivých tříd je 86,8 %, bez zohlednění četností (aritmetický průměr) pak 93,4 %.

Z ověřovací skupiny snímků z externích zdrojů model správně klasifikoval překvapivě pouze 23 z 35 případů (65,7 %).

7.3.6.3 Rozlišení pouze artritických a artrotických změn na datasetu rukou

		Předpovědi		
		ARTRI	ARTRO	NEART
Skutečnost	ARTRI	180	12	2
	ARTRO	3	50	0
	NEART	2	0	61

ARTRI		Poz.	Neg.
Poz.	180	14	
Neg.	5	111	

ARTRO		Poz.	Neg.
Poz.	50	3	
Neg.	12	245	

NEART		Poz.	Neg.
Poz.	61	2	
Neg.	2	245	

- Model ze všech 310 snímků v testovacím datasetu správně klasifikoval 180 artritických snímků, chybně do této třídy 14 snímků nezařadil. Naopak sem zařadil 5 snímků, které ve skutečnosti patří do jiných tříd.
- Model správně klasifikoval artrotické projevy u 50 snímků, 3 snímky do této třídy nezařadil, i když do ní zařazeny být měly. Naopak zařadil mezi artrotické 12 snímků, které ve skutečnosti patří do jiných tříd.
- Model správně klasifikoval 61 snímků bez příznaků revmatoidních onemocnění, chybně do této třídy 2 snímků nezařadil. Chybně 2 snímky nemocných rukou označil jako bez příznaků.

Tabulka 19: Metriky pro model na rozlišení artritických a artrotických projevů

	charakteristiky pro jednotlivé třídy			charakteristiky pro celý model	
	ARTRI	ARTRO	NEART	mikro	makro
celková správnost	0,939	0,952	0,987	0,939	0,959
přesnost	0,973	0,806	0,968	0,939	0,916
senzitivita	0,928	0,943	0,968	0,939	0,946
specifická	0,957	0,953	0,992	0,969	0,967
F1-skóre	0,950	0,870	0,968	0,939	0,931

Metriky odvozené z matic ukazují, že z jednotlivých tříd nejnižší přesnosti (80,6 %) model dosáhnul u artrotických rukou. Většina charakteristik popisujících chování modelu jako celku je vzácně vyrovnaných, celková správnost celého modelu se zohledněním četností jednotlivých tříd je 93,9 %, bez zohlednění četností (aritmetický průměr) pak 95,9 %.

7.3.6.4 Rozlišení diagnóz na datasetu kloubů

		Předpovědi			
		RA	OA	PsA	NEART
Skutečnost	RA	996	16	27	12
	OA	53	451	78	33
	PsA	39	25	510	10
	NEART	19	29	64	646

RA	Poz.	Neg.
Poz.	996	55
Neg.	111	1846

OA	Poz.	Neg.
Poz.	451	164
Neg.	70	2323

PsA	Poz.	Neg.
Poz.	510	74
Neg.	169	2255

NEART	Poz.	Neg.
Poz.	646	112
Neg.	55	2195

- Model ze všech 3008 snímků v testovacím datasetu správně klasifikoval revmatoidní artritidu u 996 snímků, 55 snímků do této třídy nezařadil, i když do ní zařazeny být měly. Chybně předpověděl revmatoidní artritidu u 111 snímků, které ve skutečnosti patří do jiných tříd.
- Model správně klasifikoval osteoartrózu u 451 snímků, chybně do této třídy 164 snímků nezařadil. Naopak chybně předpověděl osteoartrózu u 70 snímků.
- Model správně klasifikoval psoriatickou artritidu u 510 snímků, chybně do této třídy 74 snímků nezařadil. Naopak chybně předpověděl psoriatickou artritidu u 169 snímků.
- Model správně klasifikoval 646 snímků bez příznaků revmatických onemocnění, chybně do této třídy 112 snímků nezařadil. Chybně 55 nemocných kloubů označil jako bez příznaků.

Tabulka 20: Metriky pro model na klasifikaci snímků kloubů

	charakteristiky pro jednotlivé třídy				charakteristiky pro celý model	
	RA	OA	PsA	NEART	mikro	makro
celková správnost	0,945	0,922	0,919	0,944	0,865	0,933
přesnost	0,900	0,866	0,751	0,922	0,865	0,860
senzitivita	0,948	0,733	0,873	0,852	0,865	0,852
specifická	0,943	0,971	0,930	0,976	0,955	0,955
F1-skóre	0,923	0,794	0,808	0,886	0,865	0,856

Metriky odvozené z matic ukazují, že z jednotlivých tříd si model nejhůře poradil s psoriatickou artritidou a osteoartrózou. Jak je z vyhodnocení vidět, mnoho snímku skončilo zařazených špatně. Může za to pravděpodobně fakt, že mnoho kloubů má projevy onemocnění pouze počínající, mírné nebo téměř žádné a rozeznat onemocnění pouze z jednoho takového snímku kloubu je velmi obtížné i pro člověka. Výsledek proto dopadl podle očekávání. Celková správnost celého modelu se zohledněním četností jednotlivých tříd je 86,5 %, bez zohlednění četností (aritmetický průměr) pak 93,3 %.

7.3.6.5 Rozlišení pouze artritických a artrotických změn na datasetu kloubů

Model 1 (s třídou kloubů s nejasnými projevy)

Model 1		Předpovědi			
		ARTRI	ARTRO	NEGAT	NEJ.
Skutečnost	ARTRI	94	4	5	5
	ARTRO	10	170	23	13
	NEGAT	9	12	191	26
	NEJ	13	31	52	50

ARTRI		Poz.	Neg.
Poz.	94	14	
Neg.	32	568	

ARTRO		Poz.	Neg.
Poz.	170	46	
Neg.	47	445	

NEGAT		Poz.	Neg.
Poz.	191	47	
Neg.	80	390	

NEJ.		Poz.	Neg.
Poz.	50	96	
Neg.	44	518	

- Model 1 ze všech 708 snímků v testovacím datasetu správně klasifikoval artritické změny u 94 snímků, 14 snímků do této třídy nezařadil, i když do ní zařazeny být měly. Naopak předpověděl artritické změny u 32 snímků, které ve skutečnosti patří do jiných tříd.
- Model správně klasifikoval artrotické změny u 170 snímků, chybně do této třídy 46 snímků nezařadil. Naopak chybně předpověděl osteoartrózu u 47 snímků.
- Model správně klasifikoval klouby bez projevů onemocnění u 191 snímků, chybně do této třídy 47 snímků nezařadil. Naopak chybně předpověděl psoriatickou artritidu u 80 snímků.
- Model správně klasifikoval 50 snímků jako nejasných, chybně do této třídy 96 snímků nezařadil. Chybně 44 snímků kloubů označil jako nejasné.

Tabulka 21: Metriky pro model 1 na rozlišení artritických a artrotických projevů

	charakteristiky pro jednotlivé třídy				charakteristiky pro celý model	
	ARTRI	ARTRO	NEGAT	NEJ.	mikro	makro
celková správnost	0,935	0,869	0,821	0,802	0,713	0,857
přesnost	0,746	0,783	0,705	0,532	0,713	0,692
senzitivita	0,870	0,787	0,803	0,342	0,713	0,701
specifická	0,947	0,904	0,830	0,922	0,904	0,901
F1-skóre	0,803	0,785	0,750	0,417	0,713	0,696

Metriky odvozené z matic ukazují, že model měl největší potíže s třídou kloubů s nejasnými či nevýraznými projevy nemoci. Celková správnost celého modelu se zohledněním četností jednotlivých tříd je proto pouze 71,3 %, bez zohlednění četností (aritmetický průměr) pak 85,7 %.

Model 2 (bez třídy s nejasnými projevy)

Model 2		Předpovědi		
		ARTRI	ARTRO	NEGAT
Skutečnost	ARTRI	90	12	6
	ARTRO	3	182	32
	NEGAT	5	18	215

ARTRI		Poz.	Neg.
Poz.	90	18	
Neg.	8	447	

ARTRO		Poz.	Neg.
Poz.	182	35	
Neg.	30	316	

NEGAT		Poz.	Neg.
Poz.	215	23	
Neg.	38	287	

- Model 2 ze všech 563 snímků v testovacím datasetu správně klasifikoval 90 artritických snímků, chybně do této třídy 18 snímků nezařadil. Naopak sem chybně zařadil 8 snímků.
- Model správně klasifikoval artrotické projevy u 182 snímků, 35 snímků do této třídy nezařadil. Naopak chybně 30 snímků jako artrotické.
- Model správně klasifikoval 215 snímků bez příznaků revmatických onemocnění, chybně do této třídy 23 snímků nezařadil. Chybně 38 nemocných kloubů označil jako bez příznaků.

Tabulka 22: Metriky pro model 2 na rozlišení artritických a artrotických projevů

	charakteristiky pro jednotlivé třídy			charakteristiky pro celý model	
	ARTRI	ARTRO	NEGAT	mikro	makro
celková správnost	0,954	0,885	0,892	0,865	0,910
přesnost	0,918	0,858	0,850	0,865	0,876
senzitivita	0,833	0,839	0,903	0,865	0,858
specifická	0,982	0,913	0,883	0,933	0,926
F1-skóre	0,874	0,848	0,876	0,865	0,867

Metriky odvozené z matic ukazují, že z jednotlivých tříd nejnižší přesnosti (80,6 %) model dosáhl u artrotických rukou. Vynechání třídy nejasných projevů pomohlo a oproti předchozímu modelu si tento poradil s rozlišením tříd lépe. Celková správnost celého modelu se zohledněním četností jednotlivých tříd je 86,5 %, bez zohlednění četností (aritmetický průměr) pak 91 %.

8 Shrnutí výsledků

8.1 Aktuální situace využití hlubokého učení v radiologii

- Ve vědecké sféře se začínají objevovat snahy oddělené výzkumy zastřešit, dát jim směr a mimo jiné stanovit jednotná kritéria pro publikace. V ČR je vývoj v oblasti hlubokého učení pomalý, avšak ve světě je této oblasti věnována velká pozornost. Velkou měrou se ve výzkumu angažuje zejména RSNA (Radiological Society of North America) a některé univerzity a instituce v USA. Tématu hlubokého učení v radiologii se dostává stále větší pozornosti i v odborných časopisech a vědeckých konferencích. Díky tomu vznikají specializované datasety a jsou pořádány soutěže v jejich klasifikaci, které vývoj skokově posouvají vpřed.
- Výzkum se soustředí především na hledání cest k vysvětlitelnosti a interpretovatelnosti výstupů neuronových sítí, což je pro nasazení technologie ve zdravotnictví klíčové. Současné jsou zkoumány slabiny a zranitelnosti, možnosti zvýšení přesnosti využitím snímků ve vyšším rozlišení a implementací přeneseného učení, které poměrně dobře funguje i při přenosu z ne-medicínské oblasti do medicínské.
- V komerční sféře se vývojem softwaru využívajícího hluboké učení zabývají především výrobci zdravotnických přístrojů, kteří díky tomu mají k dispozici dostatek dat na trénování a vyhodnocování. V době tvorby této práce již bylo na trhu dostupných několik druhů softwaru od různých dodavatelů. Jde především o programy snažící se o automatizaci rutinních a zdoluhavých úloh jako je určování kostního věku – například Med Bone Age nebo BoneXpert, používaný i v České republice. Většina ostatních programů zpracovává data z jiných zobrazovacích metod než RTG a snaží se asistovat radiologům u náročnějších vyšetření mozku, plic nebo prostaty. Veškerý takový software nabízí především zvýraznění jednotlivých oblastí ve snímku pomocí segmentace a možnost vygenerování radiologické výstupní zprávy s podrobným kvantitativním popisem snímků, což je cesta k vzájemné porovnatelnosti jednotlivých případů napříč nemocnicemi i státy a v budoucnu by právě tato standardizace formátů a doplňující číselné údaje mohly vést k vytrénování podrobných modelů, které budou využívat tyto a další údaje o pacientovi a budou schopny asistovat při stanovování diagnóz.
- Software pro pomoc s vyhodnocováním snímků rukou trpících revmatickými onemocněními zatím neexistuje.
- Příklady komerčního softwaru také ukázaly, že je již možné získat certifikaci i na software, který má modely hlubokého učení nasazené v cloudu a pravidelně je aktualizuje.

8.2 Trénování klasifikačních modelů

- Trénování klasifikačních modelů od základů, tedy bez použití přeneseného učení, nekonvergovalo k cíli a vždy končilo nedoučením. Na vině je patrně nedostatečný počet snímků, které ani po znásobení jejich počtu metodami rozšíření dat nestačí v kombinaci s poměrně nízkým vstupním rozlišením k tomu, aby neuronová síť dokázala zaměřit pozornost na podstatné rysy obrazu. Přesto je možné konstatovat, že si v těchto úlohách vedla lépe síť architektury Inception-V3 a na druhé straně architektury Inception-V4, Inception-ResNet-V2 a ResNet dosahovaly výsledků nepatrně horších.
- Přenesené učení formou extrakce rysů se v řešených úlohách klasifikace jednotlivých diagnóz ukázalo jako úspěšnější než trénování modelu od základů, přesto ale bylo nedostatečně přesné. Klasifikace všech čtyř tříd snímků na datasetech rukou i kloubů obvykle při použití architektury Inception-V3 dosahovala celkové správnosti pouze okolo 70 %. I přesto byla extrakce rysů velmi nápomocná při tvorbě datasetu, kdy si dokázala dobře poradit s rozpoznáním, zda snímek zachycuje levou, pravou či obě ruce. Dosažená přesnost v tomto případě stačila k tomu, aby nasazení vytrénovaného modelu ušetřilo velké množství času ve srovnání s tříděním snímků manuálně. Stejným způsobem bylo toto řešení nápomocné také při třídění snímků kloubů podle míry a druhu poškození.
- Přenesené učení formou ladění se ze všech tří metod osvědčilo nejlépe a umožnilo dosáhnout velmi zajímavých výsledků. V úloze klasifikace všech čtyř tříd na snímcích rukou dosáhly vytrénované modely s architekturou ResNet-50 na testovacích datech ve většině sledovaných metrik výsledků přibližně v rozmezí 85–95 % a to v makro i mikro charakteristikách. Podrobné vyhodnocení modelů dokazuje, že jejich vlastnosti jsou v mnoha ohledech velmi dobré a při klasifikaci jednotlivých tříd si modely počínaly zhruba očekávaným způsobem, kdy případy mnohem početněji zastoupené třídy revmatoidní artritidy poznaly velmi spolehlivě, a naopak třídu psoriatické artritidy, která má mnohem variabilnější projevy, poznaly s přesností nižší. Modely byly dále prověřeny na skupině snímků z externích zdrojů, čímž byly jejich vlastnosti potvrzeny. Experimentálně bylo také vyzkoušeno, že lze poměrně přesně od sebe odlišit také projevy obecně artritické a artrotické, vytrénovaný model dosáhl ve všech charakteristikách vyjma dvou výsledků přes 90 %.
- Přenesené učení formou ladění umožnilo dosáhnout zajímavých výsledků i v experimentech s datasety kloubů, opět se nejlépe osvědčila síť ResNet. Při klasifikaci jednotlivých onemocnění celkové výsledky pravděpodobně zhoršovaly snímky kloubů s nejasnými či velmi mírnými projevy. Při rozdělení na klouby artritické, artrotické, beze změn a s nejasnými projevy se naplno potvrdilo, že se snímky s nejasnými projevy má neuronová síť potíže a když byla tato třída vynechána, výsledky se zlepšily.

- Klasifikace snímků kloubů oproti snímkům rukou nedosáhla tak dobrých výsledků. Na vině mohou být následující důvody:
 - Na snímcích celých rukou patrně neuronová síť dokáže při rozhodování vzít v potaz více informací, které z výřezů samostatných kloubů získat nelze.
 - Snímky kloubů by patrně potřebovaly roztrždit do více podtříd podle diagnóz, druhů a míry poškození – roztrždit více snímků, než bylo k dispozici nyní, a následně provést celou řadu experimentů. Vše je časově velmi náročné a nad rámec této práce, která měla za cíl pouze prozkoumat potenciál řešené úlohy.
 - Roli mohlo hrát i rozlišení výřezů kloubů – mnoho snímků muselo být zvětšeno a převzorkováno, protože jejich původní rozměry byly menší, než kolik vyžadovala neuronová síť jako vstup. Uspokojivé objasnění této otázky je opět nad rámec.
- Tato práce se přidává k dalším výzkumům s potvrzením, že přenositelnost naučených vah neuronových spojení mezi ne-medicínskými a medicínskými daty je možná a že lze takto dosáhnout až překvapivě příznivých výsledků.

8.3 Omezení současných technologií a další překážky

- V oblasti zpracování medicínských dat je překážkou, že všechny současné veřejně dostupné a rozšířené architektury neuronových sítí zpracovávají snímky ve velmi nízkém rozlišení, typicky nejvýše do 224×244 nebo 299×299 px. Navzdory tomuto omezení mnoho výzkumů dosáhlo pozoruhodných úspěchů, a stejně tak výsledky dosažené v této práci dokládají, že sofistikované a citlivé algoritmy dokážou i z takto malého počtu pixelů vytěžit překvapivě množství informací potřebných k poměrně úspěšné klasifikaci. Přesto je však rozlišení snímků nevyhovující. Dalším úskalím je převod snímků na čtvercový formát. Řešení se zachováním původního poměru stran a vyplněním prázdné plochy neutrální barvou sice teoreticky dává neuronové síti možnost přesněji se zaměřit i na tvar objektů v obrazu, na druhou stranu v praxi nedosáhlo tak dobrých výsledků. Nepatrně lepších výsledků překvapivě dosáhlo výchozí řešení, kdy dochází k automatickému roztažení kratšího rozměru tak, aby výsledný snímek vyplnil čtvercový formát. I přesto, že toto řešení může způsobovat chyby při klasifikaci těch objektů, u kterých hraje tvar roli, při klasifikaci revmatických onemocnění patrně převažuje skutečnost, že všechny snímky v datasetu zobrazují to samé a že upravený snímek v nataženém rozměru pojme více informací, než kdyby byl zmenšen se zachováním původního poměru stran.
- V oblasti hardwaru lze při současném stavu vývoje technologií využít tři cesty k trénování modelů: GPU, CPU a TPU. Každé z těchto zařízení má své nedostatky.
 - Použitelné grafické karty se omezují na monopol Nvidie a její technologie CUDA, neboť technologie OpenCL používaná výrobcem AMD nemá softwarovou podporu ze strany frameworků pro implementaci neuronových sítí.

- Grafické karty mají často paměť nedostačující pro trénování velkých modelů, což lze řešit finančně náročným pořízením více karet, ovšem opět je problém se softwarovou podporou paralelních výpočtů ze strany frameworků. Při trénování na CPU problém s pamětí odpadá, trénování je ale obecně pomalejší a vyžaduje velký počet jader a počítače speciálně sestavené jen za účelem hlubokého učení.
- Současné architektury neuronových sítí jsou omezené pouze na jeden druh dat. Existují architektury sítí specializované na zpracování textu, jiné na zpracování obrazu a další na zpracování zvuku. Při nutnosti použít více druhů dat najednou je nutné zkombinovat výstupy z více sítí pomocí technik zvaných *ensembles (seskupení)*. Vývoj naznačuje, že se časem mohou rozšířit i architektury schopné zpracovat více druhů dat současně.
- Hluboké učení je v současnosti středem pozornosti, technologie procházejí bouřlivým vývojem a vše se mění doslova pod rukama. Například použitá literatura [23, 33, 37] popisující práci s TensorFlow nebyla aktuální již v roce 2017, kdy byla vydána, a v letech 2019 a 2020, kdy vznikala tato práce, byla již značně zastaralá a nedostačující. Stejně tak zastarala i použitá verze frameworku TensorFlow a objevily se nové verze a nové frameworky, které do budoucna dávají příslib lepších výsledků. S publikováním nových verzí z webu mizela dokumentace ke starším verzím, která i tak měla ne vždy dostatečnou kvalitu a pokrytí tématu. Nelze se také ubránit dojmu, že společnosti stojící za použitými technologiemi si část know-how nechávají pro sebe a některé překážky jsou uměle vytvořené, aby technologie nebyly přístupné pro až příliš velkou masu uživatelů.
- Nesnáze přináší také software přímo nesouvisející s neuronovými sítěmi. Ve frameworku OpenCV některé funkce nelze použít na Windows, jiné naopak na Linuxu. Obecně se ukázalo, že zatímco jeden potřebný software lépe funguje na Windows, jiný funguje lépe na Linuxu, a během tvorby práce bylo potřeba rozdělit úkony mezi oba systémy. Systém Windows se při osvědčil v práci s obrazovými daty, Linux naopak v tvorbě a spouštění programů využívajících Python, TensorFlow a Docker. Pro vývoj softwaru využívajícího hluboké učení na obrazových datech je tedy vhodné mít základní znalost práce v obou systémech.

9 Závěr a otázky pro další výzkum

Tato práce popsala aktuální stav využití hlubokého učení v radiologii a výsledky vlastního řešení představeného v předešlých kapitolách ukazují, že současné technologie lze i přes mnohá omezení poměrně úspěšně použít také na klasifikaci revmatických onemocnění na snímcích celých rukou i jednotlivých kloubů. Potvrdilo se, že jednotlivá onemocnění jsou pomocí neuronových sítí rozlišitelná a stejně tak jsou rozlišitelné obecně i projevy artritické či artrotické.

Uvedené výsledky jsou na úrovni tzv. ověření koncepce (proof of concept). Celé problematika je velmi rozsáhlá do šířky i do hloubky a také řešená úloha byla už jen svým vymezením a dostupnými daty oproti realitě značně zjednodušena. Jak bylo v textu práce popsáno, výstupem radiologů není jen pouhé vyslovení diagnózy, jejich činnost je mnohem komplexnější a při dalším výzkumu a vývoje technologie je nutné to zohlednit.

Na cestě k řešení robustnímu a spolehlivému natolik, aby mohlo být nasazeno do praxe, je nutné věnovat pozornost ještě celé řadě dalších aspektů, jejichž uspokojivé řešení by výrazně převyšovalo rámec této práce a autor se jimi zamýšlí zabývat v navazujícím studiu. Konkrétně jde zejména o následující otázky:

Množství dat

Dostupný počet snímků stačil na dokázání, že na rozpoznávání revmatických onemocnění lze využít informační technologie. Pro spolehlivější výsledky či použití v praxi by počet snímků musel být řádově větší, zároveň by data měla pocházet z více různých přístrojů, více nemocnic a především více pacientů. Vše je organizačně i časově poměrně náročné. Může trvat roky, než se nakumuluje dostatečný počet nových pacientů – zejména těch co trpí jinými onemocněními než revmatoidní artritidou, která je v současném datasetu zastoupená nejpočetněji. Kromě toho by mezi stávající třídy měla patřit také speciální třída snímků, které kvůli špatné projekci, například extrémním překryvům prstů, nemají diagnostickou kvalitu a nelze je vyhodnotit.

Velkou pozornost by si zasloužily také snímky samostatných kloubů – jejich množství je pro trénování neuronových sítí vhodné, na druhé straně jejich ruční anotování bude vyžadovat obrovské množství času a ideálně také naprogramování samostatné aplikace určené jen k efektivnímu a rychlému označování a třídění.

Informace o pacientech

Kvůli technickým omezením se v popsaném řešení mohly snímky jednoho pacienta (v různých stadiích onemocnění) objevit v trénovacím, validačním i testovacím datasetu, což může teoreticky mít vliv na přeučení. S automatizovaným řešením stahování dat ze systému PACS a jejich anonymizace by bylo vhodné při dostatečném množství pacientů neprovádět dělení do datasetů podle procent z počtu snímků, ale z počtu pacientů.

Automatizované řešení by taktéž při zachování anonymity mohlo ze systému PACS stáhnout i další data o pacientovi, které by mohly pomoci při rozhodování – například věk, pohlaví apod. Místo jedné neuronové sítě na zpracování obrazu by mohlo být vyzkoušeno použití *seskupení sítí (ensemble)* na zpracování obrazových i doplňujících číselných či textových dat.

Kvantitativní popis snímků a vzhled do fungování systému

Řešení navržené v této práci navrhuje diagnózu přímočaře pouze na základě vstupního snímku. Pro přesnější a spolehlivější diagnózu a také pro řešení problému černé skříňky by při dostatečném množství dat mohla být prozkoumána cesta rozložení úlohy na dvě části – první neuronovou sítí, která by zajistila kvantitativní popis snímku, tedy které klouby a jak jsou zasažené, zda a jak jsou klouby a prsty deformované apod. V podstatě by umožnila vytvoření automatizované a standardizované radiologické zprávy. Teprve následně by se aplikovala druhá neuronová síť, která by měla k dispozici tyto a další informace o pacientovi, a na základě nich by nabídla diagnózu a případně také srovnání s podobnými případy.

Míra automatizace procesu

Celý proces vedoucí k vytrénování a vyhodnocení modelu by potřeboval automatizovat. Kromě průzkumu a dotřídění dat, které už z principu musí zůstat manuální činností, by šel zautomatizovat téměř celý proces počínaje získáním dat a konče vyhodnocením vytrénovaného modelu. Především části získávání dat a část vytrénování a vyhodnocení modelu by výrazně ušetřily čas, umožnily vyzkoušet více variant a potenciálně dosáhnout lepších výsledků. Bylo by vhodné implementovat zpracování snímků přímo ve formátu DICOM (včetně anonymizace) a nějakou formu AutoML.

Pro zaručení dostatečné kvality výstupů by celý proces by měl být od základu tvořen s přihlédnutím na v textu zmiňovaná doporučení RSNA, případně i jiný publikační standard, vznikne-li ještě nějaký.

Technologie a pevná půda pod nohama

Programovací jazyky a frameworky procházejí poměrně rychlým vývojem. Stejně tak grafické karty. Technologie zvolené na počátku tvorby práce se postupem času ukázaly jako nedostatečně stabilní a nedostatečně robustní a v době odevzdání práce byly již výrazně zastaralé.

Během zhruba roku a půl přípravy této práce vyšla nová verze TensorFlow (vyžadující upgrade na novou verzi jazyku Python a přepsání většiny kódu). Rozšířily se nové technologie pro paralelní i federativní trénování modelů, které mají potenciál využít výkon stávajícího hardwaru mnohem lépe, než tomu bylo dosud a než tomu bylo v této práci. Rozšířily se techniky AutoML pro trénování lokálně i v cloudu. Nvidia představila framework Clara specializovaný na medicínská obrazová data, který pokrývá předzpracování a anotaci dat, trénování i nasazení modelů včetně komunikace s nemocničními systémy PACS.

Překotný vývoj technologií přináší problém hlavně v nedostačující technické dokumentaci, bez které je efektivní implementace nových verzí frameworků téměř nemožná.

V souvislosti s novými technologiemi by potřebovaly revidovat také možnosti vzdáleného připojení na školní výpočetní servery, protože se v některých ohledech ukázaly být svazující až nedostačující.

Při dalším pokračování v této práci by bylo vhodné vybrat takovou kombinaci technologií, které dokáží pokrýt současné nároky na jednoduchost používání, paralelní výpočty, vizualizaci alespoň teplotních map, detailní vyhodnocení modelu – a především kvalitně zdokumentované a nepodléhající příliš častým změnám.

Rozlišení snímků

Co se naopak v posledních letech nezměnilo, jsou veřejně dostupné architektury sítí. Stále se většinou jedná o architektury představené zhruba do roku 2016 a v nejlepším případě pracující se snímky ve čtvercovém formátu o hraně maximálně 224 nebo 299 pixelů. I řešení představené v této práci využívá pouze první jmenovaný rozměr (a přesto dosahuje poměrně zajímavých výsledků). Úprava architektury na zpracování větších rozměrů je pro medicínské účely žádoucí a jak ukazují některé zmiňované výzkumy, je také technicky proveditelná. Přináší s sebou ale výrazně vyšší nároky na hardware a software. Také by bylo zapotřebí mnohem více snímků, aby trénování konvergovalo k řešení, nebo jinak řečeno, aby se neuronová síť v množství přidávaných parametrů dokázala zorientovat. S ohledem na probíhající neustálý vývoj technologií je zároveň je velmi pravděpodobné, že se brzy (v nejbližších letech) objeví inovované či nové architektury, které budou navrženy a optimalizované přímo na snímky ve vyšším rozlišení.

Přenesené učení

V této práci bylo dosaženo poměrně dobrých výsledků právě díky využití přeneseného učení formou ladění, ačkoliv předtrénovaný model využíval ne-medicínská data. Nabízí se proto otázka, zda by model vytrénovaný čistě na medicínských datech nebyl pro přenesené učení ještě vhodnější. Některé takové modely nabízí Nvidia v rámci svého softwarového frameworku Clara. Velmi obsáhlý a rozmanitý dataset by šlo také získat sloučením všech veřejně dostupných medicínských datasetů dohromady – mohlo by jít řádově o desetitisíce až jednotky statisíců snímků při použití pouze dat z RTG, při rozšíření o příbuzné zobrazovací metody pak možná i půl milionu až milion snímků. Model předtrénovaný na takovém množství medicínských snímků by mohl v přeneseném učení umožnit dosažení lepších výsledků, na druhé straně trénování takového modelu od základů klade velmi vysoké nároky na hardware.

10 Seznam použitých zdrojů

- [1] VOMÁČKA, Jaroslav, Josef NEKULA a Jiří KOZÁK. *Zobrazovací metody pro radiologické asistenty*. 1. vydání. Olomouc: Univerzita Palackého v Olomouci, 2012. ISBN 978-80-244-3126-0.
- [2] METTLER, Fred A. *Essentials of Radiology*. Third edition. Philadelphia, PA: Saunders, 2014. ISBN 978-1-4557-4225-7.
- [3] *Přístrojové vybavení zdravotnických zařízení ČR v roce 2018* [online]. [cit. 2020-03-09]. Praha: Ústav zdravotnických informací a statistiky České republiky, 2019. Dostupné z: <https://www.uzis.cz/res/f/008275/ai-2019-06-t1-pristrojove-vybaveni-zz-2018.pdf>
- [4] PAVELKA, Karel, Jiří VENCOVSKÝ, Pavel HORÁK, Ladislav ŠENOLT, Heřman MANN a Jan ŠTĚPÁN. *Revmatologie*. 2. aktualizované a rozšířené vydání. Praha: Maxdorf, 2018. ISBN 978-80-7345-583-5.
- [5] ALLISY-ROBERTS, Penelope a Jerry WILLIAMS. *Farr's Physics for Medical Imaging*. Second edition. B.m.: Elsevier, 2008. ISBN 978-0-7020-2844-1.
- [6] Soubory DICOM. *Adobe Help* [online]. [cit. 2020-04-09]. Dostupné z: <https://helpx.adobe.com/cz/photoshop/using/dicom-files.html>
- [7] MONTAGNON, Emmanuel, Milena CERNY, Alexandre CADRIN-CHÊNEVERT, Vincent HAMILTON, Thomas DERENNES, André ILINCA, Franck VANDENBROUCKE-MENU, Simon TURCOTTE, Samuel KADOURY a An TANG. Deep learning workflow in radiology: a primer. *Insights into Imaging* [online]. 2020, **11**(1), 22. ISSN 1869-4101. Dostupné z: doi:10.1186/s13244-019-0832-5
- [8] ŽÁRA, Jiří, Bedřich BENEŠ, Jiří SOCHOR a Petr FELKEL. *Moderní počítačová grafika*. 2. přepracované a rozšířené vydání. Praha: Computer Press, 2004. ISBN 978-80-251-0454-5.
- [9] *Siemens Luminos dRF Max* [online]. [cit. 2020-02-25]. Dostupné z: https://static.healthcare.siemens.com/siemens_hwem-hwem_sxxa_websites-context-root/wcm/idc/groups/public/@global/@imaging/@fluoro/documents/download/mda4/mti3/~edisp/xp-fluoroscopy-luminos-drf-max-brochure-05210418.pdf
- [10] Philips DigitalDiagnost C90. *Philips* [online]. [cit. 2020-02-25]. Dostupné z: <https://www.philips.cz/healthcare/product/HC712034/digitaldiagnost-c90>
- [11] FREEDMAN, Matthew Thomas, Shih-Chung Benedict LO, John C. SEIBEL a Christina M. BROMLEY. Lung nodules: improved detection with software that suppresses the rib and clavicle on chest radiographs. *Radiology* [online]. 2011, **260**(1), 265–273. ISSN 1527-1315. Dostupné z: doi:10.1148/radiol.11100153
- [12] September 2019: BoneXpert version 3.0 released. *BoneXpert* [online]. 17. září 2019 [cit. 2020-02-29]. Dostupné z: <https://bonexpert.com/september-2019-bonexpert-version-3-0-released/>
- [13] ŽIVNÝ, Ondřej. Rozhovor s radiologem FNHK MUDr. Ondřejem Živným o metodách diagnostické radiologie. 3. prosinec 2018
- [14] JAROSLAV, Pilný, Slodička ROMAN a KOLEKTIV. *Chirurgie ruky*. 2. aktualizované a doplněné vydání. Praha: Grada Publishing, 2017. ISBN 978-80-271-0180-1.
- [15] OLEJÁROVÁ, Marta. *Revmatologie v obrazech*. 1. vydání. Praha: Mladá fronta, 2016. ISBN 978-80-204-4249-9.
- [16] DOHERTY, Michael, Hans BIJLSMA, Nigel ARDEN, David HUNTER a Nicola DALBETH, ed. *Oxford Textbook of Osteoarthritis and Crystal Arthropathy*. Third edition. Oxford: Oxford University Press, 2016. ISBN 978-0-19-966884-7.
- [17] CLUNIE, Gavin, Nick WILKINSON, Elena NIKIPHOROU a Deepak JADON, ed. *Oxford Handbook of Rheumatology*. Fourth edition. Oxford: Oxford University Press, 2018. ISBN 978-0-19-872825-2.
- [18] GAILLARD, Frank. Rheumatoid arthritis – hands, case 3. *Radiopaedia.org* [online]. [cit. 2020-01-11]. Dostupné z: <https://radiopaedia.org/cases/7245>

- [19] TOWNSVILLE RADIOLOGY TRAINING. Psoriatic arthritis – hands. *Radiopaedia.org* [online]. [cit. 2020-01-12]. Dostupné z: <https://radiopaedia.org/cases/18381>
- [20] GAILLARD, Frank. Osteoarthritis – hands. *Radiopaedia.org* [online]. [cit. 2020-01-15]. Dostupné z: <https://radiopaedia.org/cases/7546>
- [21] CICERO, Mark a Alexander BILBILY. Machine Learning and the Future of Radiology: How we won the 2017 RSNA ML Challenge. *16 Bit* [online]. 23. listopad 2017 [cit. 2019-11-30]. Dostupné z: <https://www.16bit.ai/blog/ml-and-future-of-radiology>
- [22] SIX, Ory. The ultimate guide to AI in radiology. *Quantib B.V.* [online]. [cit. 2020-02-18]. Dostupné z: <https://www.quantib.com/the-ultimate-guide-to-ai-in-radiology>
- [23] JOSHI, Prateek. *Artificial Intelligence with Python*. First edition. Birmingham: Packt Publishing, 2017. ISBN 978-1-78646-439-2.
- [24] RUSSELL, Stuart a Peter NORVIG. *Artificial Intelligence: A Modern Approach*. Third edition. Upper Saddle River: Pearson, 2009. ISBN 978-0-13-604259-4.
- [25] LANGLLOTZ, Curtis P. Will Artificial Intelligence Replace Radiologists? *Radiology: Artificial Intelligence* [online]. 2019, **1**(3), e190058. Dostupné z: doi:10.1148/ryai.2019190058
- [26] ROUGER, Melisande. AI in radiology: What patients really want – the best possible diagnosis with the highest possible precision. *European Society of Radiology AI Blog* [online]. 10. března 2020 [cit. 2020-03-19]. Dostupné z: <https://ai.myesr.org/healthcare/ai-in-radiology-what-patients-really-want-the-best-possible-diagnosis-with-the-highest-possible-precision/>
- [27] *White Paper: Medical Imaging in the Age of Artificial Intelligence* [online]. B.m.: Siemens Healthcare GmbH. 2017. Dostupné z: https://static.healthcare.siemens.com/siemens_hwem-hwem_sxxa_websites-context-root/wcm/idc/groups/public/@global/documents/download/mda3/mzy4/~edisp/white-paper_artificial-intelligence-in-medical-imaging-04384573.pdf
- [28] *Clinical radiology UK workforce census report 2018* [online]. B.m.: The Royal College of Radiologists. 2019 [cit. 2020-04-04]. Dostupné z: <https://www.rcr.ac.uk/publication/clinical-radiology-uk-workforce-census-report-2018>
- [29] PINTO DOS SANTOS, Daniel. Assessing Radiology Research on Artificial Intelligence: A Brief Guide for Authors, Reviewers and Readers. *European Society of Radiology AI Blog* [online]. 27. února 2020 [cit. 2020-02-29]. Dostupné z: <https://ai.myesr.org/publications/assessing-radiology-research-on-artificial-intelligence-a-brief-guide-for-authors-reviewers-and-readers/>
- [30] BRINK, James A., Ronald L. ARENSON, Thomas M. GRIST, Jonathan S. LEWIN a Dieter ENZMANN. Bits and bytes: the future of radiology lies in informatics and information technology. *European Radiology* [online]. 2017, **27**(9), 3647–3651. ISSN 1432-1084. Dostupné z: doi:10.1007/s00330-016-4688-5
- [31] *Why TensorFlow* [online]. [cit. 2020-03-19]. Dostupné z: <https://www.tensorflow.org/about>
- [32] KARPATY, Andrej. Software 2.0. *Medium* [online]. 13. června 2018 [cit. 2020-03-19]. Dostupné z: <https://medium.com/@karpaty/software-2-0-a64152b37c35>
- [33] RASCHKA, Sebastian a Vahid MIRJALILI. *Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow*. Second edition. Birmingham: Packt Publishing, 2017. ISBN 978-1-78712-593-3.
- [34] SABA, Luca, Mainak BISWAS, Venkatanareshbabu KUPPILI, Elisa CUADRADO GODIA, Harman S. SURI, Damodar Reddy EDLA, Tomáš OMERZU, John R. LAIRD, Narendra N. KHANNA, Sophie MAVROGENI, Athanasios PROTOGEROU, Petros P. SFIKAKIS, Vijay VISWANATHAN, George D. KITAS, Andrew NICOLAIDES, Ajay GUPTA a Jasjit S. SURI. The present and future of deep learning in radiology. *European Journal of Radiology* [online]. 2019, **114**, 14–24. ISSN 0720-048X. Dostupné z: doi:10.1016/j.ejrad.2019.02.038
- [35] HACKING, Craig. Normal hand series. *Radiopaedia.org* [online]. [cit. 2020-04-04]. Dostupné z: <https://radiopaedia.org/cases/37486>

- [36] BAARS, Bernard J. a Nicole M. GAGE. *Cognition, Brain, and Consciousness: Introduction to Cognitive Neuroscience*. Second edition. B.m.: Elsevier, 2010. ISBN 978-0-12-375070-9.
- [37] BUDUMA, Nikhil a Nicholas LOCASCIO. *Fundamentals of Deep Learning: Designing Next-Generation Machine Intelligence Algorithms*. First edition. Sebastopol, CA: O'Reilly Media, 2017. ISBN 978-1-4919-2561-4.
- [38] KAISER, Lukasz, Aidan N. GOMEZ, Noam SHAZEER, Ashish VASWANI, Niki PARMAR, Llion JONES a Jakob USZKOREIT. One Model To Learn Them All. *arXiv:1706.05137 [cs, stat]* [online]. 2017 [cit. 2020-02-07]. Dostupné z: <http://arxiv.org/abs/1706.05137>
- [39] ELFOULY, Sharif. One model to learn them all. *Medium* [online]. 16. červenec 2019 [cit. 2020-02-07]. Dostupné z: <https://towardsdatascience.com/one-model-to-learn-them-all-3e0b0ebfb01f>
- [40] HANSEN, Casper. Neural Networks: Feedforward and Backpropagation Explained. *Machine Learning From Scratch* [online]. 5. srpen 2019 [cit. 2020-04-12]. Dostupné z: <https://mlfromscratch.com/neural-networks-explained/>
- [41] YAMASHITA, Rikiya, Mizuho NISHIO, Richard Kinh Gian DO a Kaori TOGASHI. Convolutional neural networks: an overview and application in radiology. *Insights into Imaging* [online]. 2018, **9**(4), 611–629. ISSN 1869-4101. Dostupné z: [doi:10.1007/s13244-018-0639-9](https://doi.org/10.1007/s13244-018-0639-9)
- [42] ALOM, Md Zahangir, Tarek M. TAHA, Christopher YAKOPCIC, Stefan WESTBERG, Paheding SIDIKE, Mst Shamima NASRIN, Brian C. VAN ESESN, Abdul A. S. AWWAL a Vijayan K. ASARI. The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches. *arXiv:1803.01164 [cs]* [online]. 2018 [cit. 2020-02-24]. Dostupné z: <http://arxiv.org/abs/1803.01164>
- [43] KARPATY, Andrej. A Recipe for Training Neural Networks. *Andrej Karpathy's Blog* [online]. zima 2019 [cit. 2019-10-10]. Dostupné z: <http://karpathy.github.io/2019/04/25/recipe/>
- [44] TOL, Jim. *Image Augmentation: How to overcome small radiology datasets?* [online]. [cit. 2020-01-27]. Dostupné z: <https://www.quantib.com/blog/image-augmentation-how-to-overcome-small-radiology-datasets>
- [45] LEE, Han-soo. Vuno wins European certification for AI-based medical device. *Korea Biomedical Review* [online]. [cit. 2020-04-18]. Dostupné z: <http://www.koreabiomed.com/news/articleView.html?idxno=5034>
- [46] Quantib Brain Radiology Software. *Quantib B.V.* [online]. [cit. 2020-03-10]. Dostupné z: <https://www.quantib.com/solutions/partnerships/quantib-brain>
- [47] Siemens Healthineers na RSNA 2018: Efektivní řešení pro vysoce kvalitní péči. *Trend – magazín pro partnery a zákazníky společnosti Siemens Healthcare* [online]. 2018, (3). Dostupné z: https://static.healthcare.siemens.com/siemens_hwem-hwem_sxxa_websites-context-root/wcm/idc/groups/public/@at/documents/download/mda4/nzgy/~edis/trend-3-18-web-05957037.pdf
- [48] SHORTEN, Connor a Taghi M. KHOSHGOFTAAR. A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data* [online]. 2019, **6**(1), 60. ISSN 2196-1115. Dostupné z: [doi:10.1186/s40537-019-0197-0](https://doi.org/10.1186/s40537-019-0197-0)
- [49] BLUEMKE, David A., Linda MOY, Miriam A. BREDELLA, Birgit B. ERTL-WAGNER, Kathryn J. FOWLER, Vicky J. GOH, Elkan F. HALPERN, Christopher P. HESS, Mark L. SCHIEBLER a Clifford R. WEISS. Assessing Radiology Research on Artificial Intelligence: A Brief Guide for Authors, Reviewers, and Readers—From the Radiology Editorial Board. *Radiology* [online]. 2019, **294**(3), 487–489. ISSN 0033-8419. Dostupné z: [doi:10.1148/radiol.2019192515](https://doi.org/10.1148/radiol.2019192515)
- [50] GHONEIM, Salma. Accuracy, Recall, Precision, F-Score & Specificity, which to optimize on? *Medium* [online]. 8. duben 2019 [cit. 2020-04-22]. Dostupné z: <https://towardsdatascience.com/accuracy-recall-precision-f-score-specificity-which-to-optimize-on-867d3f11124>
- [51] SOKOLOVA, Marina a Guy LAPALME. A systematic analysis of performance measures for classification tasks. *Information Processing & Management* [online]. 2009, **45**(4), 427–437. ISSN 03064573. Dostupné z: [doi:10.1016/j.ipm.2009.03.002](https://doi.org/10.1016/j.ipm.2009.03.002)

- [52] THARWAT, Alaa. Classification assessment methods. *Applied Computing and Informatics* [online]. 2018 [cit. 2020-03-01]. ISSN 2210-8327. Dostupné z: doi:10.1016/j.aci.2018.08.003
- [53] QI, Susan Ruyu. Google DeepMind might have just solved the “Black Box” problem in medical AI. *Medium* [online]. 25. srpen 2018 [cit. 2020-03-20]. Dostupné z: <https://medium.com/health-ai/google-deepmind-might-have-just-solved-the-black-box-problem-in-medical-ai-3ed8bc21f636>
- [54] CICERO, Mark a Alexander BILBILY. Predicting Skeletal Age – Analyze Your Image. *16 Bit* [online]. [cit. 2019-11-30]. Dostupné z: <https://www.16bit.ai/bone-age>
- [55] YOSINSKI, Jason, Jeff CLUNE, Anh NGUYEN, Thomas FUCHS a Hod LIPSON. Understanding Neural Networks Through Deep Visualization. *arXiv:1506.06579 [cs]* [online]. 2015 [cit. 2020-03-20]. Dostupné z: <http://arxiv.org/abs/1506.06579>
- [56] PARDAMEAN, Bens, Tjeng Wawan CENGGORO, Reza RAHUTOMO, Arif BUDIARTO a Ettikan Kandasamy KARUPPIAH. Transfer Learning from Chest X-Ray Pre-trained Convolutional Neural Network for Learning Mammogram Data. *Procedia Computer Science* [online]. 2018, **135**, The 3rd International Conference on Computer Science and Computational Intelligence (ICCSICI 2018) : Empowering Smart Technology in Digital Era for a Better Life, 400–407. ISSN 1877-0509. Dostupné z: doi:10.1016/j.procs.2018.08.190
- [57] TF-Explain. *Interpretability Methods for tf.keras Models with Tensorflow 2.0* [online]. 20. březen 2020 [cit. 2020-03-20]. Dostupné z: <https://github.com/sicara/tf-explain>
- [58] LEE, Hyunkwang, Sehyo YUNE, Mohammad MANSOURI, Myeongchan KIM, Shahein H. TAJMIR, Claude E. GUERRIER, Sarah A. EBERT, Stuart R. POMERANTZ, Javier M. ROMERO, Shahmir KAMALIAN, Ramon G. GONZALEZ, Michael H. LEV a Synho DO. An explainable deep-learning algorithm for the detection of acute intracranial haemorrhage from small datasets. *Nature Biomedical Engineering* [online]. 2019, **3**(3), 173–182. ISSN 2157-846X. Dostupné z: doi:10.1038/s41551-018-0324-9
- [59] GOODFELLOW, Ian J., Jonathon SHLENS a Christian SZEGEDY. Explaining and Harnessing Adversarial Examples [online]. 2015 [cit. 2020-03-01]. Dostupné z: <http://arxiv.org/abs/1412.6572>
- [60] SZEGEDY, Christian, Wojciech ZAREMBA, Ilya SUTSKEVER, Joan BRUNA, Dumitru ERHAN, Ian GOODFELLOW a Rob FERGUS. Intriguing properties of neural networks [online]. 2014 [cit. 2020-03-01]. Dostupné z: <http://arxiv.org/abs/1312.6199>
- [61] FINLAYSON, Samuel G., Hyung Won CHUNG, Isaac S. KOHANE a Andrew L. BEAM. Adversarial Attacks Against Medical Deep Learning Systems. *arXiv:1804.05296* [online]. 2019 [cit. 2020-04-01]. Dostupné z: <http://arxiv.org/abs/1804.05296>
- [62] MA, Xingjun, Yuhao NIU, Lin GU, Yisen WANG, Yitian ZHAO, James BAILEY a Feng LU. Understanding Adversarial Attacks on Deep Learning Based Medical Image Analysis Systems [online]. 2020 [cit. 2020-03-01]. Dostupné z: <http://arxiv.org/abs/1907.10456>
- [63] KARPATY, Andrej. Breaking Linear Classifiers on ImageNet. *Andrej Karpathy's Blog* [online]. [cit. 2020-03-01]. Dostupné z: <http://karpathy.github.io/2015/03/30/breaking-convnets/>
- [64] SINGH, Rajat Vikram. ImageNet Winning CNN Architectures – A Review. *Rajat Vikram Singh – Institute of Software Research at Carnegie Mellon University* [online]. Dostupné z: http://rajatvikramsingh.github.io/media/DeepLearning_ImageNetWinners.pdf
- [65] SHAHIN, Ahmed H., Ahmed KAMAL a Mustafa A. ELATTAR. Deep Ensemble Learning for Skin Lesion Classification from Dermoscopic Images. In: *2018 9th Cairo International Biomedical Engineering Conference (CIBEC): 2018 9th Cairo International Biomedical Engineering Conference (CIBEC)* [online]. 2018, s. 150–153. ISSN 2156-6097. Dostupné z: doi:10.1109/CIBEC.2018.8641815
- [66] KARIM, Raimi. Illustrated: 10 CNN Architectures. *Medium* [online]. 17. říjen 2019 [cit. 2020-04-11]. Dostupné z: <https://towardsdatascience.com/illustrated-10-cnn-architectures-95d78ace614d>
- [67] HOWARD, Andrew G., Menglong ZHU, Bo CHEN, Dmitry KALENICHENKO, Weijun WANG, Tobias WEYAND, Marco ANDREETTO a Hartwig ADAM. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv:1704.04861 [cs]* [online]. 2017 [cit. 2020-04-12]. Dostupné z: <http://arxiv.org/abs/1704.04861>

- [68] *Deep Learning Frameworks Comparison – Tensorflow, PyTorch, Keras, MXNet, The Microsoft Cognitive Toolkit, Caffe, Deeplearning4j, Chainer* [online]. [cit. 2020-02-29]. Dostupné z: <https://www.netguru.com/blog/deep-learning-frameworks-comparison>
- [69] *Talos* [online]. 29. únor 2020 [cit. 2020-02-29]. Dostupné z: <https://github.com/autonomio/talos>
- [70] *AutoML* [online]. [cit. 2020-02-29]. Dostupné z: <https://www.automl.org/automl/>
- [71] VÁZQUEZ, Favio. Auto-Keras, or How You can Create a Deep Learning Model in 4 Lines of Code. *Medium* [online]. 4. prosinec 2018 [cit. 2020-01-17]. Dostupné z: <https://towardsdatascience.com/auto-keras-or-how-you-can-create-a-deep-learning-model-in-4-lines-of-code-b2ba448ccf5e>
- [72] SURFsara: Switch to CPUs Improves X-Ray Analysis. *Intel* [online]. [cit. 2020-04-08]. Dostupné z: <https://www.intel.com/content/www/us/en/customer-spotlight/stories/surfsara-improves-xray-analysis-customer-story.html>
- [73] NVIDIA Clara Imaging. *NVIDIA Developer* [online]. 7. září 2018 [cit. 2020-03-11]. Dostupné z: <https://developer.nvidia.com/clara-medical-imaging>
- [74] System Architecture – Cloud TPU. *Google Cloud* [online]. [cit. 2020-04-09]. Dostupné z: <https://cloud.google.com/tpu/docs/system-architecture?hl=cs>
- [75] Graphics Reinvented: NVIDIA GeForce RTX 2080 Ti Graphics Card. *NVIDIA* [online]. [cit. 2020-02-16]. Dostupné z: <https://www.nvidia.com/en-us/geforce/graphics-cards/rtx-2080-ti/>
- [76] NVIDIA Quadro RTX 8000 Graphics Card. *NVIDIA* [online]. [cit. 2020-02-16]. Dostupné z: <https://www.nvidia.com/en-us/design-visualization/quadro/rtx-8000/>
- [77] RHU, Minsoo, Natalia GIMELSHEIN, Jason CLEMONS, Arslan ZULFIQAR a Stephen W. KECKLER. vDNN: Virtualized Deep Neural Networks for Scalable, Memory-Efficient Neural Network Design. *arXiv:1602.08124 [cs]* [online]. 2016 [cit. 2020-02-16]. Dostupné z: <http://arxiv.org/abs/1602.08124>
- [78] MARCELINO, Pedro. Transfer learning from pre-trained models. *Medium* [online]. 23. říjen 2018 [cit. 2020-04-08]. Dostupné z: <https://towardsdatascience.com/transfer-learning-from-pre-trained-models-f2393f124751>
- [79] LAKHANI, Paras. The Importance of Image Resolution in Building Deep Learning Models for Medical Imaging. *Radiology: Artificial Intelligence* [online]. 2020, **2**(1), e190177. Dostupné z: doi:10.1148/ryai.2019190177
- [80] SABOTTKE, Carl F. a Bradley M. SPIELER. The Effect of Image Resolution on Deep Learning in Radiography. *Radiology: Artificial Intelligence* [online]. 2020, **2**(1), e190015. Dostupné z: doi:10.1148/ryai.2019190015
- [81] CODREANU, Valeriu, Damian PODAREANU a Vikram SALETORÉ. Large Minibatch Training on Supercomputers with Improved Accuracy and Reduced Time to Train. In: *2018 IEEE/ACM Machine Learning in HPC Environments (MLHPC): 2018 IEEE/ACM Machine Learning in HPC Environments (MLHPC)* [online]. 2018, s. 67–76. Dostupné z: doi:10.1109/MLHPC.2018.8638634
- [82] TensorFlow Docs – Core r1.15. *TensorFlow* [online]. [cit. 2020-02-11]. Dostupné z: https://www.tensorflow.org/versions/r1.15/api_docs/python/tf/
- [83] HART, Martin. Wide Screen Apertures and Aspect Ratios. *WideScreen Museum – The Internet's Largest Film Technology Resource* [online]. [cit. 2020-03-12]. Dostupné z: <http://www.widescreenmuseum.com/widescreen/apertures.htm>
- [84] OAI – permanent archive of the clinical data, patient reported outcomes, biospecimen analyses, quantitative image analyses, radiographs and magnetic resonance images. *Osteoarthritis Initiative (OAI) – National Institutes of Health* [online]. [cit. 2020-03-12]. Dostupné z: <https://nda.nih.gov/oai/>
- [85] *Radiologická společnost ČLS JEP* [online]. [cit. 2020-04-18]. Dostupné z: <http://www.crs.cz/>
- [86] *Česká revmatologická společnost ČLS JEP* [online]. [cit. 2020-04-18]. Dostupné z: <https://www.revmatologicka-spolecnost.cz/>

- [87] MONGAN, John, Linda MOY a Charles E. KAHN. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): A Guide for Authors and Reviewers. *Radiology: Artificial Intelligence* [online]. 2020, **2**(2) [cit. 2020-04-06]. Dostupné z: doi:10.1148/ryai.2020200029
- [88] SAVADJIEV, Peter, Jaron CHONG, Anthony DOHAN, Maria VAKALOPOULOU, Caroline REINHOLD, Nikos PARAGIOS a Benoit GALLIX. Demystification of AI-driven medical image interpretation: past, present and future. *European Radiology* [online]. 2019, **29**(3), 1616–1624. ISSN 1432-1084. Dostupné z: doi:10.1007/s00330-018-5674-x
- [89] KRÁL, Jaroslav. Výzvy, hrozby a úzká místa informatiky. In: *Hovory s informatiky* [online]. Praha: MFF UK a Ústav informatiky AV ČR, 2010. Dostupné z: <http://www.cs.cas.cz/hsi1/clanky/kral.pdf>
- [90] Eurorad.org. *Eurorad – Brought to you by the ESR* [online]. [cit. 2020-01-27]. Dostupné z: <https://www.eurorad.org/homepage>
- [91] WATERMAN, Tom. Google just published 25 million free datasets. *Medium* [online]. 24. leden 2020 [cit. 2020-01-27]. Dostupné z: <https://towardsdatascience.com/google-just-published-25-million-free-datasets-d83940e24284>
- [92] LARSON, David B., Matthew C. CHEN, Matthew P. LUNGREN, Safwan S. HALABI, Nicholas V. STENCE a Curtis P. LANGLOTZ. Performance of a Deep-Learning Neural Network Model in Assessing Skeletal Maturity on Pediatric Hand Radiographs. *Radiology* [online]. 2017, **287**(1), 313–322. ISSN 0033-8419. Dostupné z: doi:10.1148/radiol.2017170236
- [93] MURAKAMI, Seiichi, Kazuhiro HATANO, JooKooi TAN, Hyoungseop KIM a Takatoshi AOKI. Automatic identification of bone erosions in rheumatoid arthritis from hand radiographs based on deep convolutional neural network. *Multimedia Tools and Applications* [online]. 2018, **77**(9), 10921–10937. ISSN 1573-7721. Dostupné z: doi:10.1007/s11042-017-5449-4
- [94] COY, Heidi, Kevin HSIEH, Willie WU, Mahesh B. NAGARAJAN, Jonathan R. YOUNG, Michael L. DOUEK, Matthew S. BROWN, Fabien SCALZO a Steven S. RAMAN. Deep learning and radiomics: the utility of Google TensorFlow™ Inception in classifying clear cell renal cell carcinoma and oncocytoma on multiphasic CT. *Abdominal Radiology* [online]. 2019, **44**(6), 2009–2020. ISSN 2366-0058. Dostupné z: doi:10.1007/s00261-019-01929-0
- [95] HAGOS, Misgina Tsighe a Shri KANT. Transfer Learning based Detection of Diabetic Retinopathy from Small Dataset. *arXiv:1905.07203 [cs]* [online]. 2019 [cit. 2019-11-25]. Dostupné z: <http://arxiv.org/abs/1905.07203>
- [96] WONG, Ken C. L., Mehdi MORADI, Joy WU a Tanveer SYEDA-MAHMOOD. Identifying disease-free chest X-ray images with deep transfer learning. *Medical Imaging 2019: Computer-Aided Diagnosis* [online]. 2019, 24. Dostupné z: doi:10.1117/12.2513164
- [97] THIAN, Yee Liang, Yiting LI, Pooja JAGMOHAN, David SIA, Vincent Ern Yao CHAN a Robby T. TAN. Convolutional Neural Networks for Automated Fracture Detection and Localization on Wrist Radiographs. *Radiology: Artificial Intelligence* [online]. 2019, **1**(1), e180001. Dostupné z: doi:10.1148/ryai.2019180001
- [98] STACHOWIAK, G. W., M. WOLSKI, T. WOLOSZYNSKI a P. PODSIADLO. Detection and prediction of osteoarthritis in knee and hand joints based on the X-ray image analysis. *Biosurface and Biotribology* [online]. 2016, **2**(4), 162–172. ISSN 2405-4518. Dostupné z: doi:10.1016/j.bsbt.2016.11.004
- [99] LEE, S., M. CHOI, H. CHOI, M. S. PARK a S. YOON. FingerNet: Deep learning-based robust finger joint detection from radiographs. In: *2015 IEEE Biomedical Circuits and Systems Conference (BioCAS): 2015 IEEE Biomedical Circuits and Systems Conference (BioCAS)* [online]. 2015, s. 1–4. Dostupné z: doi:10.1109/BioCAS.2015.7348440
- [100] TANG, Yao, Fei GAO, Jufu FENG a Yuhang LIU. FingerNet: An Unified Deep Network for Fingerprint Minutiae Extraction. *arXiv:1709.02228 [cs]* [online]. 2017 [cit. 2019-08-19]. Dostupné z: <http://arxiv.org/abs/1709.02228>
- [101] *BoneFinder* [online]. [cit. 2020-01-25]. Dostupné z: <http://bone-finder.com/>

- [102] LINDNER, C., S. THIAGARAJAH, J. M. WILKINSON, THE ARCOGEN CONSORTIUM, G. A. WALLIS a T. F. COOTES. Fully Automatic Segmentation of the Proximal Femur Using Random Forest Regression Voting. *IEEE Transactions on Medical Imaging* [online]. 2013, **32**(8), 1462–1472. ISSN 1558-254X. Dostupné z: doi:10.1109/TMI.2013.2258030
- [103] BoneFinder Hand Model. *Click2go, online licensing portal* [online]. [cit. 2020-03-20]. Dostupné z: https://www.click2go.umip.com/i/technology_marketplace/Bonefinder_hand_model.html
- [104] KALININ, Alexandr. *RSNA 2017 Pediatric Bone Age Challenge* [online]. 17. prosinec 2017. Dostupné z: https://alxndrkalinin.github.io/pdf/2017-12_CFT_BoneAge.pdf
- [105] LEE, Hyunkwang, Shahein TAJMIR, Jenny LEE, Maurice ZISSEN, Bethel Ayele YESHIWAS, Tarik K. ALKASAB, Garry CHOY a Synho DO. Fully Automated Deep Learning System for Bone Age Assessment. *Journal of Digital Imaging* [online]. 2017, **30**(4), 427–441. ISSN 1618-727X. Dostupné z: doi:10.1007/s10278-017-9955-8
- [106] Canadian company wins competition at RSNA. *Canadian Healthcare Technology* [online]. [cit. 2020-02-07]. Dostupné z: <https://www.canhealth.com/2018/02/21/canadian-company-wins-competition-at-rsna/>
- [107] October 2018: BoneXpert in 135 hospitals. *BoneXpert* [online]. 21. říjen 2018 [cit. 2020-04-18]. Dostupné z: <https://bonexpert.com/october-2018-bonexpert-in-135-hospitals/>
- [108] VUNO Inc. *VUNO – View the Invisible, Know the Unknown* [online]. [cit. 2020-03-10]. Dostupné z: <https://www.vuno.co/products>
- [109] Artificial intelligence – Our Expertise. *Siemens Healthineers* [online]. [cit. 2020-03-10]. Dostupné z: <https://www.siemens-healthineers.com/en-us/infrastructure-it/artificial-intelligence/our-expertise>
- [110] About Brainomix. *Brainomix* [online]. [cit. 2020-03-10]. Dostupné z: <https://www.brainomix.com/company>
- [111] LLC, ImageMagick Studio. ImageMagick. *ImageMagick* [online]. [cit. 2020-02-24]. Dostupné z: <https://imagemagick.org/>
- [112] *MicroDicom – free DICOM viewer* [online]. [cit. 2019-02-24]. Dostupné z: <http://www.microdicom.com/>
- [113] Transfer Learning with TensorFlow. *GitHub* [online]. [cit. 2018-11-12]. Dostupné z: https://github.com/tensorflow/tensorflow/blob/master/tensorflow/examples/image_retraining/retrain.py
- [114] GAILLARD, Frank. Rheumatoid arthritis – hands, case 2. *Radiopaedia.org* [online]. [cit. 2020-01-09]. Dostupné z: <https://radiopaedia.org/cases/3031>
- [115] KNIPE, Henry. Both hands series – normal. *Radiopaedia.org* [online]. [cit. 2020-03-09]. Dostupné z: <https://radiopaedia.org/cases/76539>
- [116] GUADARRAMA, Sergio a Nathan SILBERMAN. *TensorFlow Slim* [online]. [cit. 2019-10-17]. Dostupné z: <https://github.com/tensorflow/tensorflow/tree/master/tensorflow/contrib/slim>
- [117] *Image Module — Pillow (PIL Fork) 2.6.1 documentation* [online]. [cit. 2020-02-12]. Dostupné z: <https://pillow.readthedocs.io/en/3.0.0/reference/Image.html>

11 Přílohy

Stručný přehled jednotlivých bodů osnovy CLAIM (*Checklist for Artificial Intelligence in Medical Imaging*). Převzato z [87].

	No.	Item
TITLE or ABSTRACT		
	1	Identification as a study of AI methodology, specifying the category of technology used (eg, deep learning)
ABSTRACT		
	2	Structured summary of study design, methods, results, and conclusions
INTRODUCTION		
	3	Scientific and clinical background, including the intended use and clinical role of the AI approach
	4	Study objectives and hypotheses
METHODS		
Study Design	5	Prospective or retrospective study
	6	Study goal, such as model creation, exploratory study, feasibility study, noninferiority trial
Data	7	Data sources
	8	Eligibility criteria: how, where, and when potentially eligible participants or studies were identified (eg, symptoms, results from previous tests, inclusion in registry, patient-care setting, location, dates)
	9	Data preprocessing steps
	10	Selection of data subsets, if applicable
	11	Definitions of data elements, with references to common data elements
	12	De-identification methods
	13	How missing data were handled
Ground Truth	14	Definition of ground truth reference standard, in sufficient detail to allow replication
	15	Rationale for choosing the reference standard (if alternatives exist)
	16	Source of ground truth annotations; qualifications and preparation of annotators
	17	Annotation tools
	18	Measurement of inter- and intrarater variability; methods to mitigate variability and/or resolve discrepancies
Data Partitions	19	Intended sample size and how it was determined
	20	How data were assigned to partitions; specify proportions
	21	Level at which partitions are disjoint (eg, image, study, patient, institution)
Model	22	Detailed description of model, including inputs, outputs, all intermediate layers and connections
	23	Software libraries, frameworks, and packages
	24	Initialization of model parameters (eg, randomization, transfer learning)
Training	25	Details of training approach, including data augmentation, hyperparameters, number of models trained
	26	Method of selecting the final model
	27	Ensembling techniques, if applicable
Evaluation	28	Metrics of model performance
	29	Statistical measures of significance and uncertainty (eg, confidence intervals)
	30	Robustness or sensitivity analysis
	31	Methods for explainability or interpretability (eg, saliency maps) and how they were validated
	32	Validation or testing on external data
RESULTS		
Data	33	Flow of participants or cases, using a diagram to indicate inclusion and exclusion
	34	Demographic and clinical characteristics of cases in each partition
Model performance	35	Performance metrics for optimal model(s) on all data partitions
	36	Estimates of diagnostic accuracy and their precision (such as 95 % confidence intervals)
	37	Failure analysis of incorrectly classified cases
DISCUSSION		
	38	Study limitations, including potential bias, statistical uncertainty, and generalizability
	39	Implications for practice, including the intended use and/or clinical role
OTHER INFORMATION		
	40	Registration number and name of registry
	41	Where the full study protocol can be accessed
	42	Sources of funding and other support; role of funders



Zadání diplomové práce

Autor:	Bc. Adam Ostruszka
Studium:	I1700006
Studijní program:	N1802 Aplikovaná informatika
Studijní obor:	Aplikovaná informatika
Název diplomové práce:	Hluboké učení v radiologii
Název diplomové práce AJ:	Deep Learning in Radiology

Cíl, metody, literatura, předpoklady:

Cílem práce je prozkoumat současné možnosti využití umělé inteligence v oboru diagnostické radiologie, konkrétně způsoby analýzy 2D obrazových dat pomocí hlubokých neuronových sítí.

Osnova:

1. Úvod
2. Umělá inteligence, strojové a hluboké učení
3. Popis zkoumané problematiky, existující řešení, dostupná data
4. Návrh nového řešení
5. Testování a zhodnocení
6. Závěry

BUDUMA, Nikhil. *Fundamentals of deep learning : designing next-generation machine intelligence algorithms*. First edition. Boston : O'Reilly, 2017, 283 stran. ISBN 978-1-491-92561-4.

RASCHKA, Sebastian a MIRAJALILI, Vahid. *Python machine learning*. Birmingham : Packt, 2017. xviii, 595 stran. ISBN 978-1-78712-593-3.

CLUNIE, Gavin P. R., Nick WILKINSON, Elena NIKIPHOROU a Deepak JADON. *Oxford handbook of rheumatology*. Fourth edition. Oxford: Oxford University Press, 2018. ISBN 978-0-19-872825-2.

HRBA, Jaroslav, Jiří ŠTOLFA a Karel PAVELKA. *Artritidy a artróza v ordinaci praktického lékaře*. Praha: Apotex ČR, 1999.

PAVELKA, Karel. *Revmatologie. 2., přeprac. vyd.* Praha: Galén, 2010. ISBN 978-80-7262-688-5.

Garantující pracoviště: Katedra informačních technologií,
Fakulta informatiky a managementu

Vedoucí práce: Ing. Karel Mls, Ph.D.

Datum zadání závěrečné práce: 21. 10. 2019