



F3

**Fakulta elektrotechnická
Katedra kybernetiky**

Diplomová práce

Automatická segmentace ischemické léze u cévní mozkové příhody

Jakub Šmíd
Kybernetika a robotika

5. ledna 2025

Vedoucí práce: prof. MUDr. Jakub Otáhal, Ph.D., prof. Dr. Ing. Jan Kybic

I. OSOBNÍ A STUDIJNÍ ÚDAJE

Příjmení: **Šmíd** Jméno: **Jakub** Osobní číslo: **483554**
Fakulta/ústav: **Fakulta elektrotechnická**
Zadávající katedra/ústav: **Katedra kybernetiky**
Studijní program: **Kybernetika a robotika**

II. ÚDAJE K DIPLOMOVÉ PRÁCI

Název diplomové práce:

Automatická segmentace ischemické léze u cévní mozkové příhody

Název diplomové práce anglicky:

Automated Segmentation of Brain Ischemic Stroke

Pokyny pro vypracování:

Cévní mozková příhoda (mrtvice) je jedním z nejčastějších onemocnění a příčin úmrtí celosvětově. Zásadní krok při jejím vyšetření je segmentace poškozené tkáně na obrazech z magnetické rezonance. V současné klinické praxi se segmentace provádí manuálním obkreslováním a kvůli tomu je časově velmi náročná a výsledky podléhají značné subjektivitě. Cílem této práce je celý proces zrychlit přidáním prvků automatické segmentace obrazu. Práce bude probíhat v úzké spolupráci s Fakultní nemocnicí v Motole a neurovědci z EpiReC dlouhodobě se zabývajícími problematikou cévní mozkové příhody a souvisejícími onemocněními.

Postup:

1. Bude provedena teoretická analýza segmentačních metod ischemické léze mrtvice (např. z ISLES 2015).
2. Student zváží možnosti rozšíření datasetu pro účely použití deep learningových metod (datasety ISLES 2015 a ISLES 2022).
3. Student experimentálně ověří několik deep learningových metod multimodální segmentace léze (např. DeepMedic, ISLES22_SEALS, nnUNet). V rámci validace budou vyhodnoceny překryvy a objemy automatických segmentací proti expertním segmentacím lézí z Motola.
4. Bude zvolena vhodná multimodální deep learningová metoda/architektura, která bude sloužit pro potřeby automatické segmentace dat z Motola.
5. Následně bude studentem implementována optimalizace zvolené metody. Optimalizace bude provedena se zaměřením na modifikaci fúze multimodálních dat (input fusion, layer fusion, decision fusion).
6. Bude provedeno srovnání dosažených výsledků (překryvy a objemy segmentací) validace z datasetu z Motola s výsledky uváděnými v soudobé literatuře.

Seznam doporučené literatury:

- [1] Maier, Oskar, Bjoern H. Menze, Janina von der Gabeltz, Levin Häni, Matthias P. Heinrich, Matthias Liebrand, Stefan Winzeck, et al. „ISLES 2015 - A Public Evaluation Benchmark for Ischemic Stroke Lesion Segmentation from Multispectral MRI“. Medical Image Analysis. Elsevier BV, leden 2017. <https://doi.org/10.1016/j.media.2016.07.009>.
- [2] Ito, Kaori L., Hosung Kim, a Sook□Lei Liew. „A Comparison of Automated Lesion Segmentation Approaches for Chronic Stroke T1□weighted MRI Data“. Human Brain Mapping. Wiley, 26. červenec 2019. <https://doi.org/10.1002/hbm.24729>.
- [3] Subbanna, Nagesh K., Deepthi Rajashekhar, Bastian Cheng, Götz Thomalla, Jens Fiehler, Tal Arbel, a Nils D. Forkert. „Stroke Lesion Segmentation in FLAIR MRI Datasets Using Customized Markov Random Fields“. Frontiers in Neurology. Frontiers Media SA, 24. květen 2019. <https://doi.org/10.3389/fneur.2019.00541>.
- [4] Zhou, Tongxue, Su Ruan, a Stéphane Canu. „A Review: Deep Learning for Medical Image Segmentation Using Multi-Modality Fusion“. Array. Elsevier BV, září 2019. <https://doi.org/10.1016/j.array.2019.100004>.
- [5] Isensee, Fabian, Paul F. Jaeger, Simon A. A. Kohl, Jens Petersen, a Klaus H. Maier-Hein. „nnU-Net: A Self-Configuring Method for Deep Learning-Based Biomedical Image Segmentation“. Nature Methods. Springer Science and Business Media LLC, 7. prosinec 2020. <https://doi.org/10.1038/s41592-020-01008-z>.

Jméno a pracoviště vedoucí(ho) diplomové práce:

prof. MUDr. Jakub Otáhal, Ph.D. Ústav patologické fyziologie 2.LF UK Praha

Jméno a pracoviště druhé(ho) vedoucí(ho) nebo konzultanta(ky) diplomové práce:

prof. Dr. Ing. Jan Kybic algoritmy pro biomedicínské zobrazování FEL

Datum zadání diplomové práce: **21.01.2024**

Termín odevzdání diplomové práce: **07.01.2025**

Platnost zadání diplomové práce: **21.09.2025**

prof. MUDr. Jakub Otáhal, Ph.D.
podpis vedoucí(ho) práce

prof. Dr. Ing. Jan Kybic
podpis vedoucí(ho) ústavu/katedry

prof. Mgr. Petr Páta, Ph.D.
podpis děkana(ky)

III. PŘEVZETÍ ZADÁNÍ

Diplomant bere na vědomí, že je povinen vypracovat diplomovou práci samostatně, bez cizí pomoci, s výjimkou poskytnutých konzultací. Seznam použité literatury, jiných pramenů a jmen konzultantů je třeba uvést v diplomové práci.

Datum převzetí zadání

Podpis studenta

Poděkování / Prohlášení

Předně bych rád poděkoval své rodině a všem přátelům, kteří mi během vypracování závěrečné práce i během celého studia vyjadřovali svoji podporu.

Děkuji také svému vedoucímu prof. Jakubovi Otáhalovi a konzultantovi dr. Davidu Kalovi za poskytnuté konzultace a data, bez kterých by nebylo možné tuto práci vytvořit. Dále děkuji vedoucímu prof. Janu Kybicovi za umožnění práce na tomto tématu.

Grant číslo NU21-02-00289 je financován Agenturou pro zdravotnický výzkum České republiky.

Podpořeno projektem č. LX22NPO5107 (MŠMT): Financováno Evropskou unií – Next Generation EU.



**Financováno
Evropskou unií**
NextGenerationEU



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY



**NÁRODNÍ
PLÁN OBNOVY**

Prohlašuji, že jsem předloženou práci vypracoval samostatně a že jsem uvedl veškeré použité informační zdroje v souladu s Metodickým pokynem o dodržování etických principů při přípravě vysokoškolských závěrečných prací.

V Praze dne 5. ledna 2025

.....

Abstrakt / Abstract

Práce se zabývá výběrem 3D segmentačního modelu hlubokého učení pro účely 2. Lékařské fakulty UK. Cílem je nalézt kvalitní metodu, jež dokáže automaticky segmentovat subakutní léze mrtvice ze sekvencí FLAIR a DWI.

V rámci práce jsou analyzovány soutěže Ischemic Stroke Lesion Segmentation Challenge 2015 a 2022. Poznatky z nich jsou použity pro výběr ověřovaných metod v praktické části.

Sloučením veřejně dostupných datasetů ze soutěží vznikl hlavní trénovací dataset této práce a ověřované metody byly validovány na datech poskytnutých 2. LF. Ověřeny jsou metody DeepMedic, nnU-Net a vlastní implementace 3D U-Net. Zároveň je sledován vliv vstupní, výstupní, případně in-layer fúze. Dále je sledován vliv TSS, vliv koregistrace do prostoru MNI a vliv velikosti léze na přesnost segmentace.

Ukázalo se, že metoda nnU-Net funguje pro účely 2. LF nejlépe (DC 0,646). Proto byla metoda ověřena i pětinásobnou křížovou validací na datech 2. LF (DC 0,745). Metoda DeepMedic dosáhla velmi špatných výsledků a vlastní implementace 3D U-Net dosáhla nejnižšího DC.

Závěrem lze tvrdit, že metody typu U-Net dosahují kvalitních výsledků a definují stav poznání. Důležitější pro kvalitu segmentace je ale nastavení celého procesu trénování než výběr nebo vylepšení samotné architektury.

Klíčová slova: automatická segmentace; počítačové vidění; umělá inteligence; neuronové sítě; hluboké učení; mrtvice; magnetická rezonance; MRI; DWI; FLAIR; 3D U-Net; nnU-Net; DeepMedic; ISLES.

This thesis deals with the selection of a 3D deep learning segmentation model for the 2nd Faculty of Medicine CUNI (LF). The goal is to find the most appropriate method that can automatically segment sub-acute stroke lesions from FLAIR and DWI sequences.

The 2015 and 2022 Ischemic Stroke Lesion Segmentation Challenge competitions are analyzed. The insights from these are used to select the methods to be validated in the practical part.

The main training dataset is created by merging competitions training datasets and DeepMedic, nnU-Net, and custom implementation of the 3D U-Net were validated on data provided by the LF. The influence of input, decision and in-layer fusion is investigated. Furthermore, the influence of TSS, the effect of co-registration to MNI and the effect of lesion size on segmentation accuracy are investigated.

The nnU-Net method performs best for the purpose of the LF (DC 0.646). The method was also validated by 5-fold cross-validation on LF data (DC 0.745). The DeepMedic method performed very poorly and the custom 3D U-Net achieved the lowest DC.

In conclusion, U-Net methods can achieve good results and define the state of the art. But setup of the whole training process is more important for the quality of segmentation than the selection or improvement of the architecture itself.

Keywords: automatic segmentation; computer vision; artificial intelligence (AI); neural networks (NN); deep learning (DL); stroke; magnetic resonance imaging (MRI); DWI; FLAIR; 3D U-Net; nnU-Net; DeepMedic; ISLES.

Title translation: Automated Segmentation of Brain Ischemic Stroke

Obsah /

1 Úvod	1
2 Snímkování pomocí magnetické rezonance	2
2.1 Sekvence MRI	6
2.2 Formát dat	9
3 Úlohy počítačového vidění v radiologii	10
3.1 Klasifikace	10
3.2 Detekce	11
3.3 Regrese	11
3.4 Segmentace	11
4 Evaluační metriky segmentace	13
4.1 Sensitivita	14
4.2 Specificita	17
4.3 Precisnost	17
4.4 Klasifikační přesnost	17
4.5 Sørensen-Dice koeficient	18
4.6 Hausdorffova vzdálenost	18
4.7 Průměrná symetrická vzdálenost ploch	20
4.8 Cohenovo Kappa	20
4.9 ROC křivka a AUC	20
5 Metody automatické segmentace ischemické léze	22
5.1 Srovnávací soutěže ISLES	22
5.1.1 ISLES 2015	24
5.1.2 ISLES 2022	27
5.2 Segmentace chronické mrtvice z T1-w	29
5.2.1 Clusterize	30
5.2.2 Automatic Lesion Identification	30
5.2.3 Naivní Bayesovský klasifikátor	31
5.2.4 LINDA	31
5.3 Segmentace pomocí Markovských náhodných polí	31
6 Neuronové sítě	33
6.1 Vrstvy neuronových sítí	37
6.1.1 Lineární vrstvy	37
6.1.2 Konvoluční vrstvy	37
6.1.3 Transponovaná konvoluční vrstva	38
6.1.4 Pooling vrstvy	38
6.1.5 Ostatní vrstvy	38
6.2 Výzvy neuronových sítí	38
6.2.1 Mizení gradientu	38
6.2.2 Přeúčení	40
6.2.3 Nevývážený dataset	40
6.2.4 Prostorové zarovnání	40
6.3 Augmentace dat	41
6.4 Fúze dat	41
6.4.1 Vstupní fúze	41
6.4.2 In-layer fúze	42
6.4.3 Výstupní fúze	42
7 Metodika	43
8 Analýza datasetů	46
8.1 Dataset Motol	46
8.2 Dataset ISLES 2015	49
8.3 Dataset ISLES 2022	51
9 Experimentální ověření metod DeepMedic	55
9.1 Popis metody	55
9.1.1 Konfigurace ISLES 2015	55
9.1.2 Pozdější vývoj	56
9.2 Ověřované konfigurace DeepMedic	56
9.2.1 Testované modely	57
9.3 Vstupní fúze	58
9.4 Výstupní fúze	59
10 Experimentální ověření metod nnU-Net	62
10.1 Popis metody	62
10.2 Vstupní fúze	64
10.3 Výstupní fúze	66
10.4 Křížová validace	67
11 Experimenty s architekturou 3D U-Net	69
11.1 Popis metody	69
11.2 Implementace a ověření 3D U-Net	69
11.2.1 Vliv vzorkování	71
11.2.2 Vliv ztrátové funkce	72
11.2.3 Vliv velikosti modelu	72
11.2.4 Vliv augmentace	73
11.2.5 Vliv výstupní fúze	74
11.2.6 Vliv in-layer fúze	75
12 Diskuse	77
12.1 Dosažené výsledky	77

12.2 Vliv koregistrace do prostoru MNI	79
12.3 Vliv fúzí	79
12.4 Vliv velikosti léze	79
12.5 Vliv TSS	80
12.6 Srovnání s rešerší	82
12.7 Návrhy na zlepšení	83
13 Závěr	86
Literatura	87
A Matice predikcí testoválních metod	95
B Predikce metody DeepMedic	96
C Predikce metody nnU-Net	99
D Predikce metody nnU-Net (Motol)	102
E Predikce architektury 3D U-Net	105

Tabulky / Obrázky

2.1 Časy T_1 a T_2 pro různé tkáně ...	5	2.1 Nákres magnetické rezonance....	3
4.1 TP, FP, TN, FN	14	2.2 Excitace, relaxace, FID	4
5.1 Výsledky ISLES 2015	25	2.3 K-space	6
5.2 Výsledky ISLES 2022	28	2.4 Sekvence T1-w, T2-w, FLAIR a DWI.....	7
5.3 Výsledky segm. z T1-w 1	29	2.5 Porovnání CT a MRI v čase....	8
5.4 Výsledky segm. z T1-w 2	30	2.6 Souřadnicový systém	9
5.5 Výsledky MRF	32	4.1 TP, FP, TN, FN	15
8.1 Dataset Motol odstup sběrů 2-1	47	4.2 Vlastnosti evaluačních metrik .	16
8.2 Dataset Motol odstup sběrů 3-2	47	4.3 Hausdorffova vzdálenost	19
8.3 Dataset Motol nativní veli- kosti tensorů.....	48	5.1 Výsledky soutěže ISLES 2015 .	26
8.4 Dataset Motol deskriptivní statistika lézí	48	6.1 Model perceptronu	34
8.5 Dataset Motol statistika komponent.....	48	6.2 Vrstvy neuronové sítě	34
8.6 Dataset Motol analýza late- rality	48	6.3 Propagace neuronovou sítí	35
8.7 Dataset Motol analýza podle atlasu	49	6.4 Gradientní sestup	36
8.8 Dataset ISLES15 nativní ve- likosti tensorů	50	6.5 Malý krok učení	36
8.9 Dataset ISLES15 deskriptiv- ní statistika lézí	50	6.6 Velký krok učení	36
8.10 Dataset ISLES15 statistika komponent.....	50	6.7 Konvoluce	37
8.11 Dataset ISLES15 analýza la- terality	51	6.8 Aktivační funkce	39
8.12 Dataset ISLES15 analýza podle atlasu	51	7.1 Odstranění lebky	44
8.13 Dataset ISLES22 inter-rater analýza	52	7.2 Koregistrace na FLAIR	45
8.14 Dataset ISLES22 deskriptiv- ní statistika lézí	53	7.3 Koregistrace MNI	45
8.15 Dataset ISLES22 statistika komponent.....	53	8.1 Sjednocení masek	47
8.16 Dataset ISLES22 analýza la- terality	53	8.2 Dataset Motol rozložení lézí ...	49
8.17 Dataset ISLES22 analýza podle atlasu	54	8.3 Dataset ISLES15 rozložení lézí	51
9.1 DeepMedic vstupní fúze	58	8.4 Dataset ISLES22 rozložení lézí	54
9.2 DeepMedic výstupní fúze	61	9.1 Architektura DeepMedic	55
10.1 nnU-Net vstupní fúze	65	9.2 DeepMedic vstupní fúze	59
10.2 nnU-Net výstupní fúze.....	67	9.3 DeepMedic model 75FC	60
10.3 nnU-Net křížová validace	68	9.4 DeepMedic výstupní fúze	61

11.1	3D U-Net výsledky podle vzorkování	71
11.2	3D U-Net výsledky podle ztrát. funkce	73
11.3	3D U-Net výsledky podle ve- likosti modelu	73
11.4	3D U-Net výsledky podle augmentace	74
11.5	3D U-Net výstupní fúze	74
11.6	3D U-Net in-layer fúze	76
B.2	DeepMedic predikce 1	96
B.3	DeepMedic predikce 2	97
B.4	DeepMedic predikce 3	98
C.5	nnU-Net predikce 1	99
C.6	nnU-Net predikce 2	100
C.7	nnU-Net predikce 3	101
D.8	nnU-Net (Motol) predikce 1 ..	102
D.9	nnU-Net (Motol) predikce 2 ..	103
D.10	nnU-Net (Motol) predikce 3 ..	104
E.11	3D U-Net predikce 1	105
E.12	3D U-Net predikce 2	106
E.13	3D U-Net predikce 3	107

Kapitola 1

Úvod

Cévní mozková příhoda je celosvětově jedním z nejčastějších onemocnění, které vede k trvalé invaliditě nebo smrti. Její včasná diagnostika může významně zlepšit dopady tohoto onemocnění. K diagnostice cévní mozkové příhody se mj. používá snímkování magnetickou rezonancí, často několika sekvencemi. Více sekvencí poskytuje lékařům podpůrné informace pro zjištění samotného výskytu onemocnění a dále pro lokalizaci postižené mozkové tkáně nebo pro určení stáří onemocnění.

Pro studium mrtvice lékaři provádí segmentaci, tj. obkreslování zasažené tkáně po jednotlivých řezech 3D skenů magnetické rezonance. Segmentace precizně zachycuje prostor zasažené tkáně a poskytuje tak kvalitativní pohled na onemocnění nebo umožňuje určit přesný objem zasažené tkáně. Segmentace cévní mozkové příhody mohou být využity pro účely diagnostiky, pro výzkum mrtvice, pro predikce míry invalidity nebo pro rehabilitační účely. Provedení segmentace je ovšem časově náročné u většího počtu snímků a proto jsou snahy o automatizaci tohoto procesu. Automatická segmentace je ale jedním z nejsložitějších problémů ve zpracování medicínských obrazů.

Cílem této práce je vyvinout plně automatickou segmentační metodu pro 2. Lékařskou fakultu Univerzity Karlovy, jejímž vstupem budou modality FLAIR a DWI a výstupem metody bude segmentační maska léze. Metoda bude sloužit k výzkumu vývoje mrtvice, proto byl 2. LF poskytnut reprezentativní dataset 19 pacientů naskenovaných ve 3 časových bodech. Prostředkem pro nalezení takovéto metody je ověření několika 3D metod hlubokého učení a sledování vlivu fúze informací z jednotlivých sekvencí.

Nejprve je v práci teoretičky uveden princip získání skenů, ale především jsou popsány jednotlivé sekvence a zobrazení lézí na sekvencích v čase. Dále jsou definovány problémy počítačového vidění a segmentace. Aby bylo možné porozumět výsledkům uváděných v literatuře, jsou v textu práce nejdříve popsány evaluační metriky segmentace. Následně je provedena rešerše. Podstatná část rešerše se zaobírá soutěžemi Ischemic Stroke Lesion Segmentation Challenge. Následuje krátký úvod do neuronových sítí, jenž je důležitý pro praktickou část závěrečné práce.

V rámci praktické části byla navržena metodika přístupu k datům. V práci jsou použity tři různé datasety a tři nezávislé metody, proto je kláden důraz na konzistenci při porovnávání jednotlivých metod. Byla provedena analýza všech použitých datasetů a poté byly ověřeny metody DeepMedic a nnU-Net, jejichž implementace jsou veřejné a kód byl převzat. Nakonec je ověřena vlastní implementace architektury 3D U-Net. U všech tří ověřovaných metod byl testován vliv vstupní a výstupní fúze. Architektura 3D U-Net byla modifikována a byl testován vliv in-layer fúze. Dosažené výsledky nejlepších modelů tří metod jsou na závěr diskutovány mj. z pohledu vlivu fúzí, vlivu velikosti lézí, vlivu koregistrace do prostoru MNI a vlivu času skenování pacienta od počátku cévní mozkové příhody.

Kapitola 2

Snímkování pomocí magnetické rezonance

V této kapitole bych chtěl nejprve uvést onemocnění cévní mozkové příhody (CMP), popsat, jak vznikají a jak vypadají data, se kterými se v této práci budu dále zabývat - ať už experimentálně nebo jsou dostupná v datasetech, se kterými budu pracovat a ze kterých se ischemická léze segmentuje. Nakonec krátce popíši formát, v jakém jsou data přenášena.

Cévní mozková příhoda (známá také jako mozkový infarkt nebo mrtvice) vzniká obecně přerušením přítoku krve (blokádou cévy) nebo prasknutím cévy v mozku. V důsledku je zabráněno vyživování a okysličování části mozku a mozkové buňky v místech postižených mrtvicí po čase odumírají. Odumírání buněk se projevuje špatnou funkcí zasažené části mozku a projevuje se často špatnou pohybovou koordinací postiženého, poruchou zraku, sluchu nebo řeči v závislosti na lokalizaci léze.

Onemocnění cévní mozkovou příhodou se dělí na dvě hlavní skupiny právě podle mechanismu vzniku cévní mozkové příhody na ischemickou, která je způsobena ucpáním mozkové tepny a snížením nebo úplnou blokádou přítoku krve do postižené části mozku. Ischemická cévní mozková příhoda je nejčastějším druhem mrtvice, vyskytuje se ve více než 62 % mrtvic [1]. Druhou skupinou mrtvice je hemoragická cévní mozková příhoda, která je způsobena prasknutím mozkové tepny a následným akutním krvácením do mozku. V této závěrečné práci se budu věnovat čistě ischemické cévní mozkové příhodě.

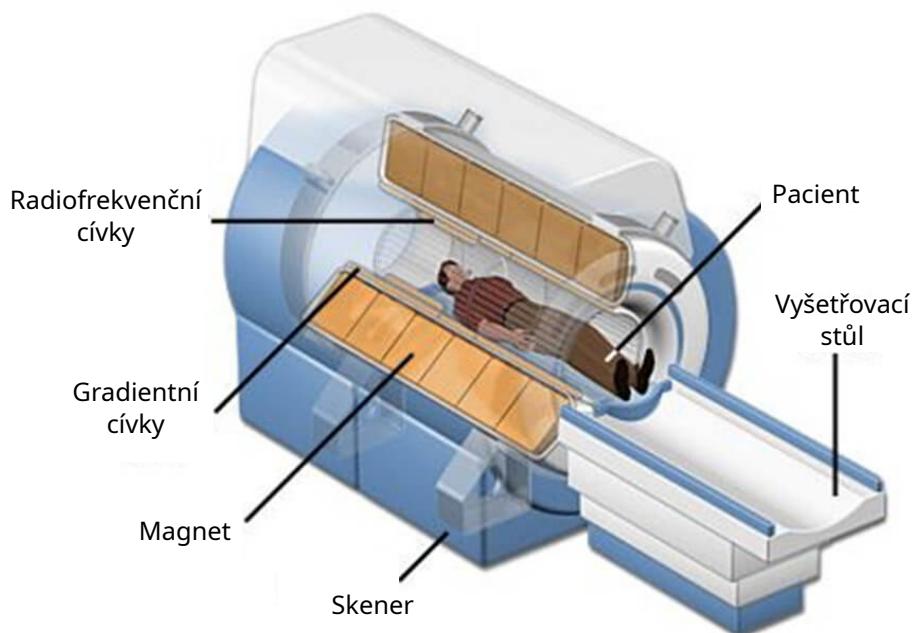
Dále se onemocnění cévní mozkové příhody dělí podle doby, která uběhla od vzniku tohoto onemocnění. Čas od počátku mrtvice je v literatuře označován zkratkou TSS (time since stroke onset). Pacienti dorazí do nemocnice v různou dobu od vzniku mrtvice a je důležité určit, kdy mrtvice vznikla, jelikož tento údaj ovlivňuje následnou léčbu lékařem. Snímkování více různými způsoby (sekvencemi) poskytuje dodatečnou informaci mj. pro určení TSS. V této práci se budu držet dělení fází na akutní cévní mozkovou příhodu (0 až 24 hodin od vzniku), subakutní (24 hodin až 2 týdny) a chronickou (více než 2 týdny od vzniku cévní mozkové příhody). [2]

Každoročně přibude celosvětově 12,2 milionu lidí zasažených první mrtvicí, z nichž 6,5 milionu v důsledku mrtvice zemře. V roce 2019 se 101 milionů lidí potýkalo s následky mrtvice. Přes 63 % mrtvic bylo u pacientů mladších 70 let a přes 16 % mrtvic se vyskytlo u pacientů mladších 49 let. Dle aktuální statistiky jeden ze čtyř lidí během svého života prodělá mrtvici (což je o polovinu větší šance než před 17 lety) [1]. Mrtvice jsou 5. nejčastější příčinou smrti v USA (hned za srdečním infarktem, rakovinou, COVID-19 a za nehodami) [3]. V ČR v roce 2023 zemřelo na následky cévní nemoci mozku 6506 lidí, což je v Česku třetí nejčastější příčina smrti s podílem 5,77 %. Před cévními nemocemi mozku jsou chronická ischemická choroba srdeční (13,32 %) a selhání srdce (6,34 %) [4].

Pro diagnostiku cévní mozkové příhody se používá výpočetní tomografie (CT - computed tomography) nebo magnetická rezonance (MRI - magnetic resonance imaging). Výpočetní tomografie je relativně levná, rychlá a dostupná radiologická metoda diagnostiky, která je založena na prozáření tkáně rentgenovými paprsky z několika úhlů. V přístroji se nachází rentgenka a detektor, které se otáčejí za postupného pohybu lůžka,

na kterém je pacient. Následnou transformací pořízených snímků se získá 3D sken, který dokáže detektovat mrtvici, avšak pokud mrtvice vznikla v krátké době od skenu, potom metoda CT nemusí být dostatečně sensitivní. Její sensitivita se pohybuje mezi 20 až 75 % mezi šesti až osmi hodinami od vzniku mrtvice [5]. CT je sensitivní na hemoragické cévní mozkové příhody [6].

Magnetická rezonance je druhým přístrojem, který slouží k diagnostice mrtvice. Oproti CT jde o komplexnější přístroj, který není tak často dostupný a kontraindikací pro vyšetření jsou kovové implantáty. Na magnetické rezonanci vynikají měkké tkáně. Přístroj a metoda snímkování je založená na rádiových signálech v silném magnetickém poli. Oproti CT a mnohým dalším metodám je tato zobrazovací technika zcela neinvazivní.

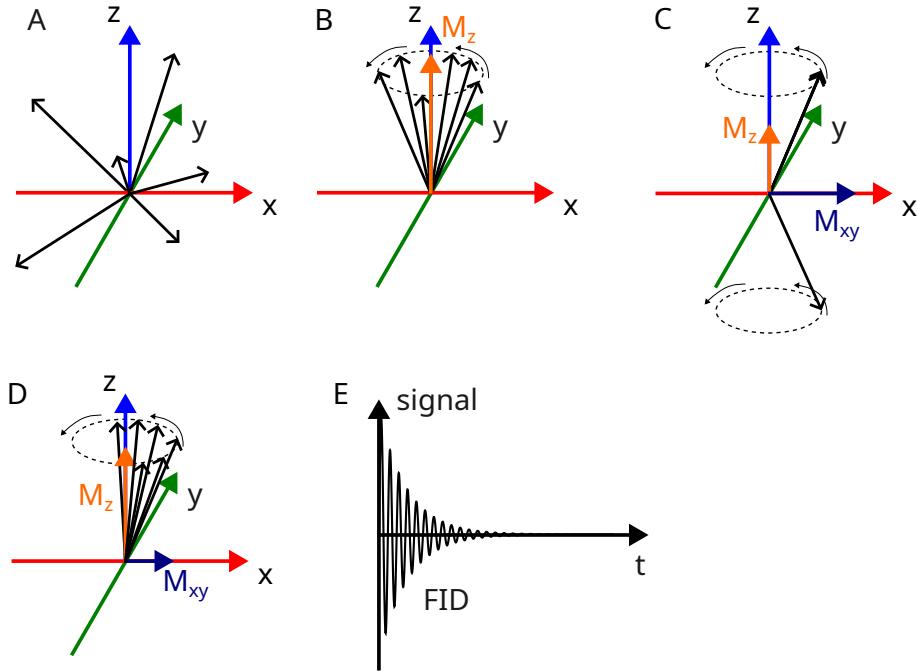


Obrázek 2.1. Zjednodušený nákres magnetické rezonance [7].

Nákres skeneru používaného pro snímkování magnetickou rezonancí je na obrázku 2.1. Skener obsahuje silný magnet, který může být buď permanentní (do zhruba 1 T) nebo supravodivý elektromagnet chlazený tekutým heliem. Supravodivé magnety jsou v nemocnicích nejčastější, poskytují vyšší magnetickou indukci (1,5 T nebo nejčastěji 3 T) a lepší rozlišení než skenery s permanentními magnety. Magnet vytváří homogenní pole uvnitř tubusu skeneru paralelně s podélnou (longitudinální) osou pacienta.

Zdrojem MRI signálu jsou primárně jádra vodíku v těle. Aby bylo možné tento signál získat a vizualizovat, je nutná série mnoha složitých kroků. Velmi zjednodušeně se je v několika následujících odstavcích pokusím popsat [8–10].

Protony kolem sebe mají v důsledku spinu (vlastnost částice, kterou si lze představit jako rotaci kolem vlastní osy - resp. moment hybnosti) slabý magnetický moment. Za běžných okolností v těle člověka jsou protony uspořádány chaoticky, tedy tyto slabé magnetické momenty se statisticky vzájemně vyruší a není možné cokoliv měřit - to odpovídá obrázku 2.2 A, kde je vyznačený prostor, jehož osa z je souběžná s podélnou osou člověka a zároveň s vnějším homogenním magnetickým polem v magnetické rezonanci. Rovina xy je poté příčnou (transversální) rovinou pacienta. Na obrázku jsou černě znázorněny orientace magnetických momentů jednotlivých protonů.



Obrázek 2.2. A: Chaotické uspořádání magnetických momentů (černých šipek) jednotlivých protonů vodíku. B: Uspořádání momentů kvůli vnějšímu magnetickému poli v ose z. Vlivem přebytku protonů uspořádaných paralelně s vnějším polem dojde ke vzniku magnetického momentu M_z , který je složený z momentů jednotlivých protonů. Magnetické momenty jednotlivých složek rotují na Larmorově frekvenci a otáčení není ve fázi. C: Přivedením RF pulzu se některé z protonů dostanou do anti-paralelního stavu, což znázorňuje černá šipka jdoucí v opačném směru osy z. Pohyb protonů se také seběhne do fáze. V důsledku těchto jevů poklesne podélná magnetizace M_z a začne se projevovat magnetizace v rovině xy označená jako M_{xy} , která také rotuje. D: Vypněme-li buzení RF cívками, dojde k postupnému rozfázování otáčení momentů a navrácení protonů z anti-paralelního stavu do stavu paralelního. Během tohoto postupného jevu můžeme měřit FID (free induction decay) signál zobrazený na obrázku E. Kvůli rotaci jader má i tento signál Larmorovu frekvenci.

Pokud jsou protony v silném vnějším magnetickém poli, potom se začnou projevit dvě věci - za prvé se magnetické momenty protonů orientují buď paralelně nebo anti-paralelně k vnějšímu magnetickému poli. Za druhé se magnetické momenty těchto protonů začnou sbíhat a rotovat po kuželu v ose vnějšího magnetického pole (precese). Anti-paralelní stav je energeticky náročnější, proto tento stav zaujmou nepatrně menší množství protonů než paralelní orientaci. Přebytek paralelně rotujících protonů je přibližně 6 na 1 milion protonů vody při teplotě lidského těla za intenzity vnějšího pole 1 T [10]. Tento přebytek způsobuje magnetizaci v longitudinální rovině, kterou je možné měřit. Magnetizace rotuje na Larmorově frekvenci (rezonanční frekvenci jader vodíku) dané vztahem (1), kde ω je úhlová rychlosť precese, γ je gyromagnetický poměr (konstantní pro dané jádro atomu) a B_0 je intenzita vnějšího magnetického pole, kde se částice nachází.

$$\omega = \gamma \cdot B_0 \quad (1)$$

V magnetickém poli Země je Larmorova frekvence vodíku zhruba 2 kHz, zatímco v magnetické rezonanci o síle magnetické indukce 1,5 T je frekvence 64 MHz [10]. Výše popsáne jevy jsou znázorněny na obrázku 2.2 B, společně s výsledným složeným vektorem magnetizace M_z .

Radiofrekvenční (RF) cívky se dělí na vysílací a přijímací cívky, případně může být stejná cívka použita pro oba účely. Často je přijímací cívka mimo vnitřek tubusu skeletu kvůli tomu, aby byla co nejblíže snímané části těla. RF vysílací cívky vyšlou pulz na Larmorově frekvenci, tato dodaná energie začne převracet některé protony do anti-paralelního stavu a tím se začne snižovat longitudinální magnetizace. Protony také začnou postupně rotovat ve fázi a díky tomu se objeví transversální magnetizace M_{xy} (rotující na Larmorově frekvenci). Běžně se používají RF pulzy o takové energii, aby převrátili rotující vektor magnetizace o 90° nebo o 180° . Výše popsaná část se označuje jako excitace a výsledek je ilustrován na obrázku 2.2 C.

Jakmile přestaneme vysílat RF pulz, dojde k postupnému navrácení superponovaných vektorů magnetizace do původního stavu, kvůli návratu jader vodíku z anti-paralelního do původního paralelního stavu a rozfázování precese. Postupně vymizí M_{xy} a začne se zvětšovat M_z . Tento rychlý, ale postupný proces se nazývá relaxace a je ilustrován na obrázku 2.2 D. Relaxace se dělí na tzv. longitudinální relaxaci (označovanou také jako spin-lattice relaxace), která popisuje rychlosť návratu resp. nárůstu podélné složky magnetizace. Druhý druh relaxace se nazývá transversální (označovaná spin-spin) a popisuje rychlosť návratu resp. zánik transversální složky magnetizace. Obě relaxace jsou exponenciální a časové konstanty relaxací se označují jako T_1 pro longitudinální relaxaci a T_2 pro relaxaci transversální. Tyto časové konstanty jsou různé pro různé druhy tkání, příklady některých tkání jsou uvedeny v tabulce 2.1. Díky tomu je možné zachytit kontrast mezi různými tkáněmi.

	T_1 v poli 1,5 T	T_2
Bílá hmota (WM)	783 ms	92 ms
Šedá hmota (GM)	917 ms	101 ms
Mozkomíšní mok (CSF)	3 000 ms	1 400 ms

Tabulka 2.1. Časy T_1 a T_2 pro různé druhy tkání v magnetickém poli 1,5 T. [10]

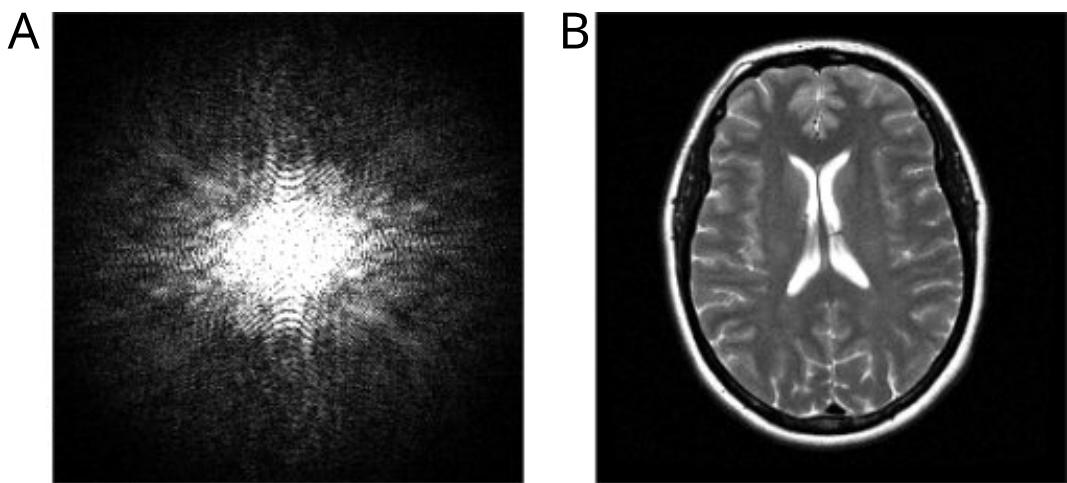
Na přijímacích resp. měřících RF cívkách se během relaxace začne indukovat FID signál (free induction decay) s průběhem na grafu 2.2 E. Tento signál osciluje na Larmorově frekvenci a je postupně tlumen tak, jak zaniká magnetizace M_{xy} .

Kontrast jednotlivých tkání se získá vhodným nastavením RF pulzů operátorem magnetické rezonance. Časování pulzů a úhel sklopení vektoru magnetizace jsou klíčové parametry pro získání kontrastu. Základní obrazy (sekvence) z magnetické rezonance jsou T_1 -vážený (T_1 -w), T_2 -vážený (T_2 -w) a protonová hustota (PD - proton density). T_1 -w obraz se získává minimalizací vlivu protonové hustoty a transversální magnetizace v měřeném signálu. Obdobně T_2 -w se získá minimalizací vlivu protonové hustoty a longitudinální magnetizace a PD obraz se získá potlačením vlivu longitudinální a transversální relaxace na měřený signál.

Poslední důležitou součástí magnetické rezonance jsou gradientní cívky, jedná se o tři nezávislé ortogonální cívky, které modulují hlavní silné magnetické pole. Slouží k zajištění prostorového rozlišení snímaného signálu. Funkce gradientních cívek lze rozdělit do tří kategorií podle toho, jakým způsobem modulují hlavní magnetické pole. První z nich je gradient výběru roviny, který moduluje hlavní silné magnetické pole a vytváří lineární gradient hlavního magnetického pole - např. souběžně v jeho směru. Tím pádem jádra vodíku v různých příčných řezech pacienta vykonávají precesi na různé Larmorově frekvenci (složením hlavního magnetického pole a magnetického pole od gradientních cívek) dojde totiž k tomu, že každá příčná plocha se nachází v různé intenzitě magnetické

indukce). RF pulz tedy excituje pouze daný příčný řez, který má stejnou Larmorovu frekvenci. Tato cívka je označována slice selection gradient G_s . Pokud nyní budeme na příčný řez pohlížet jako na 2D matici, tak další dvě cívky slouží k fázovému (cívka G_p) a frekvenčnímu kódování (cívka G_f), které zajistí, že jádra vodíku budou podél jedné osy mírně rozfázované (fázové kódování) a v ortogonálním směru budou jádra provádět precesi na mírně odlišných frekvencích (frekvenční kódování).

Opakovanou sekvencí pulzů lze změřit celou 2D matici, která obsahuje signál složený z různých frekvencí a posuvů fází. Takto získaná matice naměřeného signálu se označuje jako k-space a inverzní 2D Fourierovou transformací lze získat požadovaný obraz. Příklad změřené intenzity v k-space a výsledné intenzity obrazu po provedení 2D inverzní Fourierové transformaci jsou na obrázku 2.3.

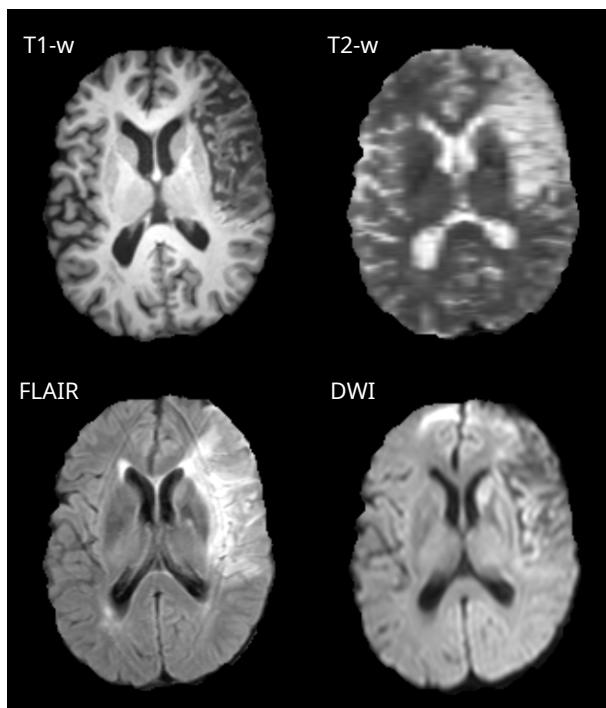


Obrázek 2.3. A: Změřené intenzity na magnetické rezonanci (k-space). B: Příslušné intenzity po provedení inverzní Fourierovy transformace. [11]

2.1 Sekvence MRI

Různé způsoby skenování pacienta, které určuje operátor, se nazývají sekvence. Operátor magnetické resonace si určí parametry skenu v závislosti na zkoumané tkáni a cílovém zobrazení, kterého chce u snímku dosáhnout. Výsledný obraz resp. 3D matice se typicky zobrazuje v odstínech šedi, kdy vysoká úroveň signálu má barvu bílou, střední intenzita signálu barvu šedou a nízká úroveň signálu je černá. Často se v literatuře používá pojem hyperintenzita, která označuje oblast s nadmerně silnějším signálem než okolí nebo porovnávaná oblast. Dále se používají obdobně pojmy hypointenzita pro oblast s významně nižším signálem než okolí nebo srovnávaná oblast. A nakonec isointenzita pro oblast, kde je stejná úroveň signálu. Různé příklady sekvencí jsou zobrazeny na obrázku 2.4.

Již zmíněné sekvence T1-w a T2-w jsou nejčastěji používané. Tyto dva snímky lze snadno rozlišit při pohledu na intenzitu signálu v místech, kde se nachází mozkomíšní mok. T1-w obraz má mozkomíšní mok a jiné tekutiny s nízkou intenzitou signálu, svaly mají střední intenzitu signálu a tuk má vysokou intenzitu signálu. Pro šedou hmotu mozkovou poté platí, že má střední intenzitu a bílá hmota má vyšší intenzitu než šedá hmota. Neboli čím kratší je čas T_1 dané tkáně - čím rychleji longitudinálně relaxuje, tím světlejší bude.



Obrázek 2.4. Porovnání obrazů T1-w, T2-w, FLAIR a DWI. Jedná se o pacienta č. 1 z datasetu ISLES 2015 [2].

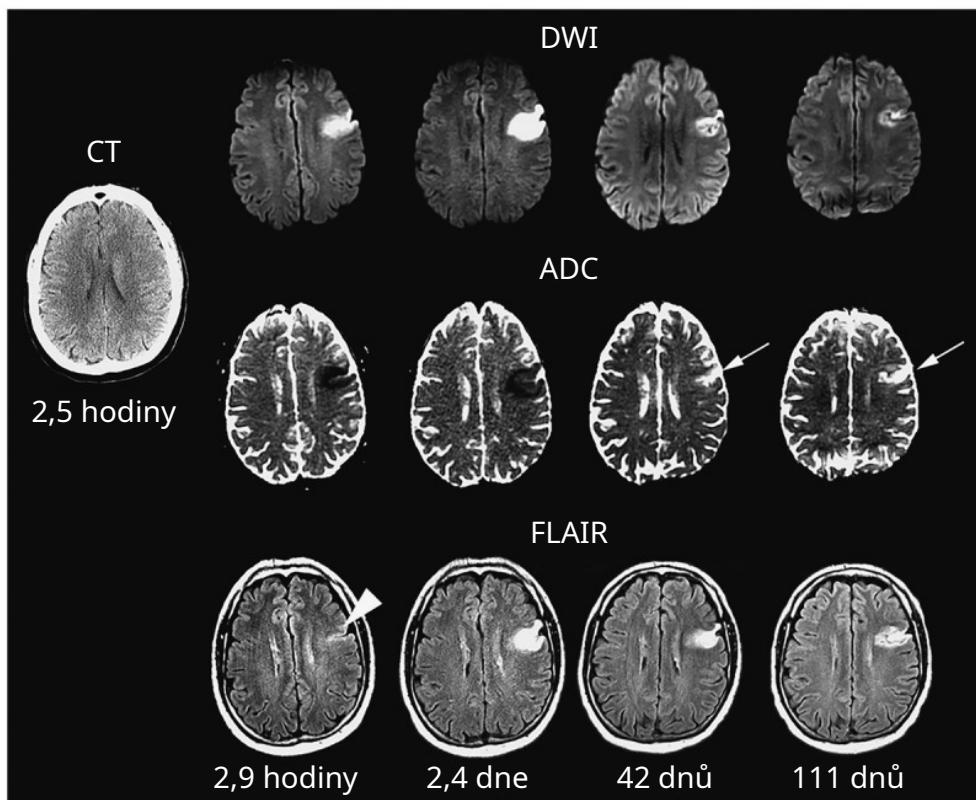
T2-w obraz má sice střední intenzitu signálu pro svaly a vysokou intenzitu signálu tuku (tedy stejně jako u T1-w), ale mozkomíšní mok má intenzitu vysokou. Pro T2-w a snímkování mozků platí, že šedá hmota je na obraze šedá, ale bílá hmota je tmavší než šedá hmota. [9] Hypointenzita v oblasti ischemické léze na T1-w se začne projevovat až po 16 hodinách od počátku cévní mozkové příhody. Signál v oblasti léze na T1-w má slabou intenzitu i během subakutní fáze, s možným výskytem hyperintenzity po 3 až 5 dnech v případě kortikální nekróz [12]. Hyperintenzita v oblasti léze na T2-w se projevuje po šesti až osmi hodinách a pokračuje v sílení signálu po následující čtyři dny. T2-w hyperintenzita v důsledku léze přetrvává až po dobu 2 až 4 měsíců. [6, 5, 13, 12] T1-w se používá pro studium chronické ischemické cévní mozkové příhody kvůli vysokému rozlišení snímků. [14]

Další známou sekvencí je FLAIR (Fluid Attenuation Inversion Recovery, inverzní zobrazení s potlačením signálu tekutiny) a je to třetí nejpoužívanější sekvence. Obraz FLAIR lze získat s mírně odlišným nastavením než je sekvence T2-w. Sekvence je velmi sensibilní na patologie - oproti T2-w je potlačen signál mozkomíšního moku, který má nyní nízkou intenzitu, ale léze zůstávají viditelné s velmi vysokou intenzitou signálu. Hyperintenzita na FLAIR v oblasti léze vzniká typicky 6 až 12 hodin (v některých případech i 0 až 2 hodiny [12]) od vzniku mrtvice a postupně se zesiluje následující dva dny [6, 5]. Vysoké intenzity na FLAIR v oblasti ischemické léze přetrvávají i v chronické fázi [12]. Ukázalo se, že sekvence FLAIR dosahuje při detekci ischemické léze lepších výsledků než při použití T2-w nebo protonové hustoty [13].

DWI (Diffusion Weighted imaging, difuzně vážené obrazy) je extrémně citlivá metoda na detekci akutní ischemické léze. Toto snímkování zachycuje se pohyb molekul vody ve směru působení magnetického pole [6]. Je to klíčová sekvence, díky které lékaři dokáží s kombinací T2-w nebo FLAIR rozlišit akutní mrtvici od subakutní nebo od chronické. [5] V oblasti ischemické léze se na obraze DWI objeví hyperintenzita v rámci minut. Tato

hyperintenzita je patrná 10 až 14 dní. [12] Nicméně ne vždy hyperintenzita na DWI znamená poškozenou tkáň akutní cévní mozkovou příhodou [5]. Pro rozlišení, zda se jedná o akutní ischemickou lézi na hyperintenzitě DWI pomůže ADC mapa (Apparent Diffusion Coefficient). V místech, kde je hyperintenzita na DWI, tak v případě, že se jedná o akutní lézi, je zasažená oblast na ADC hypointenzita. [6] Hypointenzita značící ischemickou lézi na ADC zůstává po dobu 7-10 dnů a poté přechází na hyperintenzitu [12].

Přední výhodou magnetické rezonance při diagnostice cévní mozkové příhody je vysoká sensitivita v počátečních hodinách od vzniku ischemické mrtvice (91 až 100 % během méně než 6 hodin [5]) a citlivost při snímaní měkkých tkání. Citlivost magnetické rezonance je patrná na obrázku 2.5, která zobrazuje průběh cévní mozkové příhody v čase na třech různých sekvencích magnetické rezonance. V případě, že mrtvice nastala v čase kratším než 4,5 hodiny, tak je mrtvice viditelná na sekvenci DWI, ale není patrná hyperintenzita na sekvenci FLAIR. Naopak na sekvenci ADC jsou hodnoty v zasažené oblasti během prvních 7 až 10 dnů snížené a následně se vrací k původnímu stavu. Z tohoto příkladu je patrné, proč je důležité snímat mnoha různými sekvencemi, jelikož každé z nich dávají odlišné informace o snímané tkáni.



Obrázek 2.5. Zobrazení pacienta nasnímaného pomocí CT a pomocí magnetické rezonance (DWI, ADC a FLAIR) v několika různých časech. Pacient byl 54 letý muž s ischemickou lézí nasnímaný nekontrastním CT v čase 2,5 hodiny od vzniku mrtvice, na tomto snímku nejsou viditelné žádné změny. Oproti tomu na modalitě DWI se patrná zasažená oblast s vysokou intenzitou již od počátku, na ADC je opět od počátku viditelná zasažená oblast a její intenzita se v čase zvětšuje. Na FLAIR je patrná abnormalita během 2,9 hodin od počátku mrtvice, ale maximum je patrné po 2,4 dnech od počátku mrtvice. [5]

Nesoulad mezi DWI a FLAIR v prvních 4,5 hodinách se označuje jako „DWI-FLAIR mismatch“ [15] a je velmi důležitý z hlediska diagnostiky, protože poskytuje lékařům

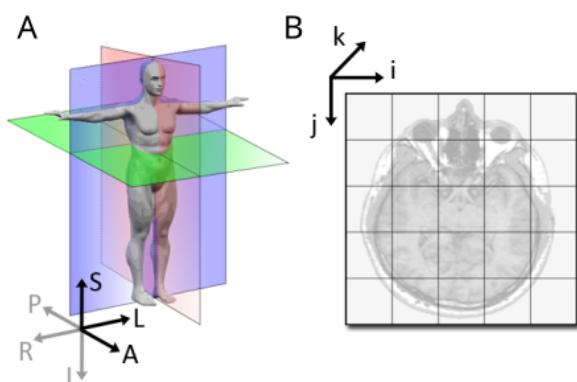
informace pro zhodnocení rizika následující léčby. Nejefektivnější léčbou mrtvice je trombolýza, která spočívá v podání látek rozpouštějících krevní sraženinu, která blokuje tok krve. Trombolýzu lze provést mezi 3 a 4,5 hodinami od počátku mrtvice. V případě takové léčby mimo toto okno převyšují rizika nad úspěšností této léčby [16]. Nesoulad mezi FLAIR a DWI pomůže lékaři zjistit kolik času již uběhlo od vzniku a zda může takovou léčbu zahájit. Stejně jako tato práce řeší automatickou segmentaci cévní mozkové příhody, tak existuje mnoho studií a experimentů, které řeší úlohu FLAIR-DWI mismatch a odhadují čas od počátku mrtvice - např. studie z roku 2023 [17] uvádí srovnatelné výsledky jako expertní odhady lidí. Stejně tak srovnávací studie [18] článků z let 2017-2022 popisuje přístup a výsledky pěti různých autorů, kteří řeší klasifikaci počátku mrtvice (více či méně než 4,5 hodiny od počátku) na základě metod umělé inteligence.

2.2 Formát dat

Skenery nejčastěji produkují data ve formátu DICOM (Digital Imaging and Communications in Medicine), který je však příliš rozsáhlý pro účely automatického zpracování obrazu. Naskenované modality, které v této práci používám, mají výhradně formát NIfTI (Neuroimaging Informatics Technology Initiative). Tento formát umožňuje uložení jak skenů, tak segmentací nebo 4D snímků. Celkem formát podporuje uložení dat až o 7 dimenzích. Pro vizualizace, analýzu, úpravu, segmentace a další zpracování biomedicínských snímků se často používá Pythonový program 3D Slicer [19].

NIfTI soubor s příponou *.nii* je složen z hlavičky a nekomprimovaných dat. Často se však pro snížení objemu dat používá komprese a poté má soubor příponu *.nii.gz*. Nejdůležitější součást hlavičky souboru je afinní matice o rozměru (4, 4), která pomocí 12 parametrů mapuje data do referenčního souřadnicového systému. Hlavička dále obsahuje například rozměry voxelu (jedné položky 3D matice, analogie pixelu u 2D obrázku) a definici jednotky těchto rozměrů (typicky mm), případně definici časové jednotky. [20]

Affinní matice slouží ke transformaci dat v souboru ze souřadnicového systému datové matice (označováno jako ijk systém) do anatomického souřadnicového systému (reálné souřadnice světa). NIfTI předpokládá, že affinní matice je transformace z ijk do systému RAS (Right-Anterior-Superior) - viz obrázek 2.6, tato orientace se označuje jako neurologická konvence. Radiologická konvence má převrácenou levo-pravou osu, proto má označení LAS. 3D Slicer defaultně zobrazuje snímky v LAS, tedy pravá strana je zobrazena vlevo.



Obrázek 2.6. A: Anatomická souřadnicová soustava. B: Souřadnicová soustava (i, j, k) s urových dat. [19]

Kapitola 3

Úlohy počítačového vidění v radiologii

V posledních letech se rychle rozvíjí možnosti a uplatnění automatického počítačového zpracování obrazu. Výhodou oboru radiologie je, že téměř všechna hlavní data používaná v radiologii jsou již v digitální podobě. Počítačové metody zpracování dat lze tedy aplikovat na rozsáhlou škálu problémů, které se v radiologii řeší. Jednou z nejsložitějších úloh je interpretace obrazu resp. detekce nemoci. V některých případech je totiž složité vyhodnotit, zda ložisko v obrazu je abnormální či se jedná o běžný nález - takovým příkladem může být snímek z mamografu, kde se vyskytují tisíce individuálních hyperintenzit, které je třeba jednotlivě rozlišit, zda se jedná o abnormální nález. V momentě, kdy je detekována abnormalita, přichází další komplexní úloha, a tou je určení diagnózy a léčba. Často platí, že pro určení diagnózy je potřeba sesbírat mnoho různých příznaků, kterými mohou být velikost léze, její poloha, intenzita signálu léze nebo její tvar a změny v čase. Všechny tyto příznaky lze zpracovat počítačově. Někdy lze dokonce zpracovat příznaky, které mají vliv na určení diagnózy, ale tyto příznaky radiolog dříve nebral při svém zhodnocení v úvahu. Třeba z důvodu, že příznaků bylo mnoho nebo jejich vliv na určení diagnózy byl latentní. Úkoly, které nejsou rutinní prací radiologa mohou být další oblasti, kde lze aplikovat počítačové vidění. Příkladem může být vytváření individuální prognózy pacienta, tvorba predikcí, jak se bude léze dále vyvíjet, predikce stupně invalidity a další podobné úlohy. Takové predikce mají potenciál sloužit k cílené léčbě a zmírnění dopadů nemoci v budoucnu po rehabilitaci. Přesto se v klinické praxi tyto predikce příliš nevyužívají. [21]

Problémy počítačového zpracování obrazu v oboru radiologie lze rozdělit do čtyř hlavních skupin podle formulace problému, který řeší. [21]

3.1 Klasifikace

Prvním problémem je klasifikace obrazu, při které vstup (obraz nebo jiné příznaky) přiřazujeme do jedné či více tříd nebo kategorií (labelů). Principem klasifikace je určení, zda se ve vstupních datech (v obrazu) nachází nějaký objekt, jev či stav. Resp. jaký objekt je v datech zakódován. Opticky objekt nemusí být na první pohled patrný nebo je člověku v obrazu úplně skrytý. Problém klasifikace je jedním z nejstarších a nejčastěji řešených problémů v počítačovém vidění. Řešením problému je nalezení pravděpodobnostního rozdělení každé ze tříd a následné přiřazení vstupního obrazu ke třídě, která má nejvyšší pravděpodobnost pro daný vstup.

Běžně se klasifikace používá na popisování fotek nebo rozpoznávání jednotlivých lidí podle jejich obličejů nebo k moderování škodlivého a ilegálního obsahu na sociálních sítích.

V radiologii lze aplikovat klasifikaci na určení, zda se v obrazu nachází či nenachází abnormalita, klasifikaci, zda je v obrazu zhoubný či nezhoubný nádor, klasifikovat lze i vstupní obraz na nějakou prognózu (zde platí, že musí být dopředu definované tyto prognózy - resp. třídy) nebo lze klasifikovat, zda uplynulý čas od vzniku cévní mozkové příhody již překonal určitou dobu či ne. [21]

Klasifikaci můžeme rozdělit na klasifikaci do jedné nebo do více tříd. Jestliže například rozpoznáváme kamerou na dálnici vozidla a chceme počítat, jaké typy vozidel projely, tak problém odpovídá klasifikaci do jedné ze tříd. Vozidlo, které projelo, tak může spadat do nákladních vozidel nebo do osobních vozidel nebo do motocyklů, ale nemůže spadat současně do více než jedné z těchto tříd (jednoduše - vozidlo nemohlo být osobní i motocykl zároveň). Klasifikaci do více tříd lze použít například při vytváření automatických popisků fotografií v domácím fotoalbu, kdy se na jedné fotce může vyskytovat hora, strom, člověk i batoh současně. Proto lze více tříd přidělit jediné fotce.

3.2 Detekce

Druhá skupina problémů je detekce. Cílem detekce je identifikování objektů zájmu a lokalizace těchto objektů ve vstupní obrazu formou parametrického minimálního ohraňčení těchto objektů. Podobně jako klasifikace, i u detekce je možné rozlišit detekci objektů jedné třídy a více tříd. Typicky se detekce používá v bezpečnostních kamérách nebo při mapování pomocí satelitních snímků. Detekce umožňuje nalezení objektu zájmu v obrazu, jeho identifikaci a případné sledování jeho pohybu v obrazu. Lze tedy pomocí bezpečnostních kamer odhalit pohyb osob v prostoru, kde nemají být, případně detektovat podezřelé chování osob. V případě mapování lze detekcí vyhledávat body zájmu (různé urbanistické změny, změny ve vodstvu nebo např. lesní požáry) v datech ze satelitu a následně tyto body zájmu zakreslovat do mapy a zpřesňovat tím mapové podklady nebo pomocí detekce získat nové informace, které bez ní nebyly dostupné. Detekci lze opět použít i k automatické katalogizaci fotoalb, kdy v jedné fotce je lokalizováno několik obličejů lidí a ke každému obličeji je detekčním algoritmem přiřazeno jméno osoby.

V radiologii se detekce používá k nalezení abnormalit, orgánů nebo anatomických struktur, které mohou být použity pro další zpracování. Detekce se dělí na dvě podúlohy - za prvé nalezení podezřelých oblastí a parametrizace oblastí pomocí ohraňčujícího boxu, tato část má typicky mnoho falešně pozitivních nálezů. Teprve druhá podúloha, kterou je klasifikace těchto oblastí, vyfiltruje falešně pozitivní nálezy. [21]

3.3 Regrese

Třetí úlohou, která se v radiologii řeší, tak je regrese [22]. Výstupem regrese je nějaká číselná hodnota, která je závislá na vstupu. Regrese se používá například pro odhad různých vzdáleností, při predikci cen na trhu, při modelování počasí, předpovídání pravděpodobnosti, že nastane nějaký jev a podobně. V lékařství se regrese používá pro odhad různých biomarkerů z obrazu jako je hustota minerálů v kostech [23], k predikci dožití pacientů s rakovinou [24] nebo k určení a predikci závažnosti ischemické cévní mozkové příhody, pomocí hodnotícího skóre [25].

3.4 Segmentace

Čtvrtou kategorií úkolů počítačového zpracování obrazu v radiologii je segmentace. Segmentace je úloha, během které dojde k preciznímu rozdělení vstupního obrazu na různé objekty nebo na různé jeho části. Segmentací lze také pojmenovat jako klasifikaci každého pixelu nebo voxelu v obrazu. Výsledkem segmentace je tedy zařazení každého pixelu do nějaké ze tříd - v praxi jde o masku obrázku, která obsahuje celá čísla v

závislosti na kategorii daného pixelu nebo voxelu. Rozdíl segmentace a detekce spočívá právě v preciznosti ohraničení segmentovaného objektu. Segmentace se používá jak v oboru zpracování medicínských dat, tak například v oboru autonomního řízení aut, kdy vozidlo segmentuje scénu kolem sebe. Díky tomu pozná, kde se nachází například ostatní účastníci provozu nebo překážky. I na základě těchto znalostí ovlivňuje svoji autonomní jízdu.

Segmentaci lze rozdělit stejně jako v předchozích příkladech na segmentaci do jedné či více tříd a dále na segmentaci sémantickou, kdy instance jednotlivých tříd ve vstupním obrazu nemusíme rozlišovat. Například segmentace ischemických lézí řešená v této práci je sémantická segmentace do jedné třídy, jelikož je důležité pro každý voxel určit, zda jde o ischemickou lézi nebo ne. V případě, kdy záleží na jednotlivých instancích segmentovaných objektů, jde o instanční segmentaci - chceme-li rozlišit více objektů stejné třídy v jedné scéně. Obecným příkladem sémantické segmentace může být segmentace lidí, kdy výstupem je maska, která jednou třídou označuje pixely, kde se na fotografii tito lidé nachází. V případě instanční segmentace bychom navíc získali informaci nejen o tom, kde tito lidé jsou, ale i jednotlivé osoby by byly v rámci segmentace rozlišeny.

Segmentace se v radiologii používá v mnoha případech, kterými mohou být segmentace tumorů, segmentace anatomických částí lebky a mozku, segmentace ložisek roztroušené sklerózy v bílé hmotě a tak dále [21]. Segmentace ischemické léze dosahuje globálně mnohem horší přesnosti než segmentace tumorů [14]. Segmentace představuje důležitý nástroj pro poskytnutí kvantitativní a prostorové informace o abnormalitách způsobených cévní mozkovou příhodou. [26]

Segmentaci lze rozdělit na manuální, semi-automatickou a automatickou.

- Manuální segmentaci provádí radiologové nebo jiní experti. Je to tradiční způsob, jakým lékaři popisují a hodnotí mrtvici za účelem zlepšení péče o pacienta. Segmentace spočívá v procházení 3D prostoru snímku pacienta řez po řezu a vyhledávání a trasování lézí na těchto snímcích. Takovéto vytváření manuální segmentace je závislé na zkušenosti experta a je jak časově náročné, tak není přesné v rámci segmentací několika různých expertů, jelikož závisí na úsudku jednotlivých expertů a často se tyto úsudky napříč experty různí. To vede k mírně odlišným segmentacím jednoho pacienta v případě segmentování více experty - typicky k rozdílným segmentacím na hranici léze. Výhodou manuální segmentace je to, že expert může využít znalosti anatomie (např. kde se hledaná léze nemůže vyskytovat) a znalosti typických tvarů a struktur léze (např. její hladkost, polohu, vzdálenosti jednotlivých ucelených ložisek). Tyto expertní znalosti mohou pomoci v rozhodovacím procesu a zlepšit tak výsledky manuální segmentace, přičemž tyto znalosti jsou špatně přenositelné do metod automatické segmentace. [27]
- Semi-automatické metody pro vytvoření segmentace potřebují zásah člověka. Typicky takové metody automaticky osegmentují více oblastí, než jaká je korektní expertní segmentace. Expert následně vybere chytrým nástrojem jen jednu z těchto oblastí.
- Plně automatickou segmentaci lze dále rozdělit do dvou hlavních kategorií:
 - Učení s učitelem, kdy segmentační algoritmus potřebuje již nějaká data, která jsou expertem anotována.
 - Učení bez učitele, kdy k segmentaci není třeba anotovaných dat. Motivací k tomuto druhu automatické segmentace může být nedostatek veřejně dostupných datasetů a zároveň intenzivní a časově náročná práce při expertní manuální segmentaci. Tento druh učení využívá matematického modelování k vytvoření charakteristik pro rozpoznání ischemické tkáně od jiných druhů tkání a následně vyhledává oblasti s touto charakteristikou. [28]

Kapitola 4

Evaluacní metriky segmentace

Abychom mohli metody z rešerše vzájemně porovnávat a abychom dokázali zjistit výkon jednotlivých segmentační metod, je nutné nejdříve definovat metriky, podle kterých to budeme dělat. Díky uvedeným metrikám je možné srovnat kvalitu dvou segmentací stejného vzoru - typicky segmentaci experta (radiologa) a automatickou segmentaci. Tyto metriky jsou zvlášt důležité při velkém množství dat, protože dovolují kvantitativní srovnání výsledků, aniž bychom museli prohlížet všechna data a vizualizovat si výsledky.

V mnoha studiích v posledních letech byla uváděna úspěšnost segmentačních algoritmů blížící se 100 %. Po bližším zkoumání těchto článků vědci došli k závěrům, že tyto velmi kvalitní výsledky byly dosaženy nějakou z metodologických chyb jako je zvolení špatné metriky nebo tzv. cherry picking - resp. evaluační metrika byla počítaná na validačním setu nereprezentující reálná data nebo byla vysoká úspěšnost dosažena špatným oddělením trénovacích a validačních dat [29]. Takovéto výzkumy poté způsobily potíže týmům, které se snažily zreplikovat výsledky například na vlastních datech nebo na datech z jiných zdravotnických center, než jaké byly použity v původní studii, protože výkonnost algoritmů v jiných prostředích se značně zhoršila.

Nejenom z výše uvedených důvodů je důležité zvolit správné evaluační metriky. V případě zvolení správné metriky, při její správné implementaci a při dodržení správných postupů při vytváření závěrů lze dosáhnout spolehlivějšího porovnání modelů napříč různými studiemi a lepší opakovatelnosti experimentů z těchto studií. [29]

Prostor medicínských dat je reprezentovaný maticí $X = x_1, \dots, x_n$, kde rozměry této matice jsou udávány jako shape (w, h, d) , kde w, h a d jsou šířka, výška a hloubka matice. Počet prvků matice je pak $|X| = w \times h \times d = n$. Formálně lze expertní segmentaci (Ground Truth) reprezentovat jako množinu $S_g = \{S_g^0, S_g^1\}$, vytvořenou funkcí $f_g^i(x)$, která přiřazuje hodnoty podmnožinám S_g^0 nebo S_g^1 podle příslušnosti prvku $x \in X$ k dané podmnožině. $f_g^i(x) = 1$, pokud platí, že $x \in S_g^i$ a $f_g^i(x) = 0$, pokud platí, že $x \notin S_g^i$. V případě, že je segmentace reprezentována pravděpodobnostní mapou, tedy fuzzy množinou, poté $f_g^i(x) \in (0, 1)$ reprezentuje pravděpodobnost příslušnosti prvku $x \in X$ do podmnožiny S_g^i . Analogicky je definována množina $S_t = \{S_t^0, S_t^1\}$, která reprezentuje testovanou segmentaci (např. predikci automatického algoritmu) a hodnoty jsou přiřazeny funkci $f_t^i(x)$. Jelikož v této práci je úkolem jasně rozlišit lézi cévní mozkové příhody a pozadí (tkáň nezasaženou cévní mozkovou příhodou), pracujeme pouze se dvěma třídami a každá z nich je jasně definována. Množiny S_g a S_t jsou ostré množiny a prvky množin nabývají jené z hodnot $\{0, 1\}$, protože i funkce $f_g^i(x)$ a $f_t^i(x)$ nabývají hodnot $\{0, 1\}$. Předpokládá se, že (S_g^1, S_t^1) jsou množiny reprezentující lézi a (S_g^0, S_t^0) jsou množiny reprezentující pozadí. Platí tedy, že $f_g^0(x) + f_g^1(x) = 1$ a $f_t^0(x) + f_t^1(x) = 1$ pro všechna $x \in X$.

Většina používaných metrik je definována pomocí poměrů následujících základních parametrů:

- Pravdivě pozitivních (TP) - míňeno počtu voxelů, které byly osegmentovány jak expertem, tak predikováný algoritmem.

- Falešně pozitivních (FP) - tedy počtu voxelů, které expert neosegmentoval, ale automatický algoritmus ano.
- Pravdivě negativních (TN) - počtu voxelů, které neosegmentoval ani expert, ani segmentační algoritmus.
- Falešně negativních (FN) - počtu voxelů, které osegmentoval expert, ale tyto voxely nejsou zahrnuty v predikci automatického algoritmu.

Tyto parametry vypovídají o překryvech dvou ostrých množin S_g a S_t . Každý z parametrů lze spočítat jako sumu shody segmentace prvků (i, j) pro každý pár prvků $i \in S_g$ a $j \in S_t$ [30]. Formálně lze tedy definovat vztah

$$m_{ij} = \sum_{r=1}^{|X|} f_g^i(x_r) f_t^j(x_r) \quad (1)$$

kde v případě porovnávání predikce proti expertní segmentaci (tedy S_t proti S_g) lze psát $TP = m_{11}$, $FP = m_{01}$, $TN = m_{00}$ a $FN = m_{10}$. Platí, že $TP + FP + TN + FN = |X|$.

Množina	S_t^0	$S_t^1 (= \overline{S_t^0})$
S_g^0	TN	FP
$S_g^1 (= \overline{S_g^0})$	FN	TP

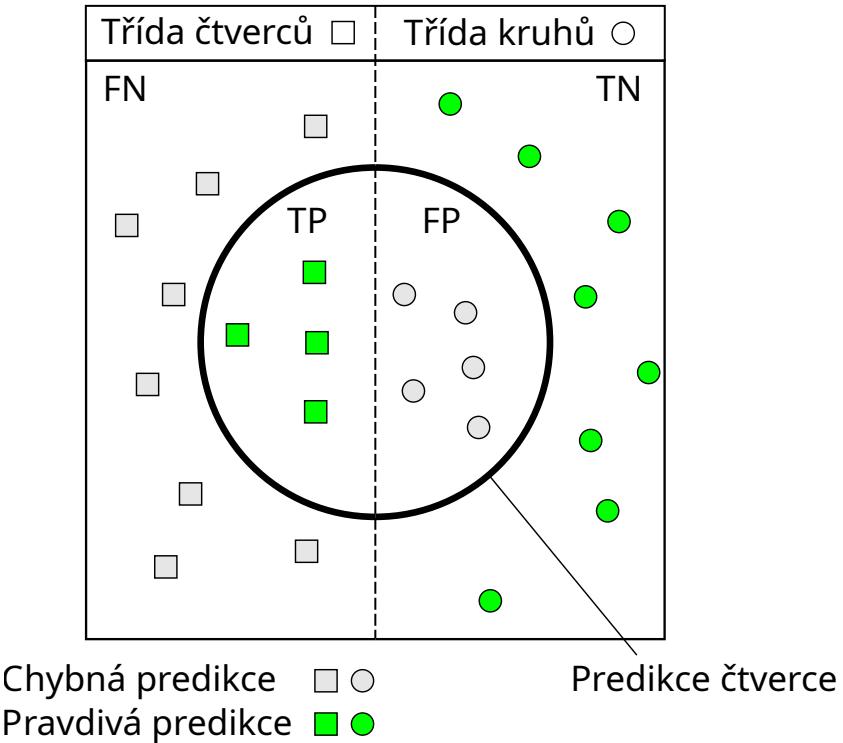
Tabulka 4.1. Rozdělení pravdivě pozitivních (TP), falešně pozitivních (FP), pravdivě negativních (TN) a falešně negativních (FN) kategorií podle příslušnosti prvku x matici dat X do některé z množin S_g^i nebo S_t^i , kde (S_g^1, S_t^1) jsou množiny reprezentující lézi a (S_g^0, S_t^0) jsou množiny reprezentující pozadí. Předpokládáme porovnání S_t vůči S_g .

Tyto vztahy jsou popsány v tabulce 4.1 a také jsou graficky zobrazeny v horní části obrázku 4.1. V levé části obrázku se nachází vzorky spadající do třídy čtverců (v reálném případě tato třída může zastupovat voxelů s lézí) a v pravé části grafiky se nachází vzorky spadající do třídy kruhů (v reálném případě může tato třída reprezentovat voxelů s tkáním, která není postižena mrtvicí). Cílem klasifikátoru resp. segmentačního algoritmu je klasifikovat vzorky do třídy čtverců. Klasifikátor z populace udělal výběr vzorků, který je uvnitř vyznačeného kruhu, tento výběr dle klasifikátoru označuje třídu čtverců. Je evidentní, že úspěšnost není stoprocentní. V grafice jsou tedy zeleně označeny vzorky, které klasifikátor úspěšně klasifikoval - tyto vzorky spadají do skupiny TP a TN. Ostatní vzorky jsou algoritmem chybně klasifikovány.

Pro vyhodnocení přesegmentování (tedy míry přesahu predikce a expertní segmentace nebo predikce v oblastech, kde expert neosegmentoval) a podsegmentování (nedostatečně automaticky osegmentovaná oblast vůči oblasti, kterou osegmentoval expert) se nejčastěji používají metriky označované jako precisnost nebo sensitivita (výtěžnost, recall). Pro porovnání kvality dvou segmentací jsou používány následující míry: Dice koeficient nebo Jaccardův index. V následujících podkapitolách rozeberu jednotlivé míry a jejich definice, protože jsou použité pro srovnávání segmentačních metod uvedených později. Aby bylo možné porozumět výsledkům prezentovaným ve studiích i výsledkům této práce, je nutné znát, jak se jednotlivé míry chovají na různých datech. K tomu dobře poslouží obrázek 4.2.

4.1 Sensitivita

Sensitivita známá také pod pojmem recall (česky výtěžnost), označována jako Sens nebo TPR (True Positive Rate), je ve zdravotnictví spolu se specificitou nejznámějším



$$\text{Sens (výtěžnost)} = \frac{\text{TP}}{\text{Třída čtverců } \square}$$

$$\text{Spec} = \frac{\text{TN}}{\text{Třída kruhů } \circ}$$

$$\text{Precisnost (PPV)} = \frac{\text{TP}}{\text{Predikce čtverce } \square \text{ } \circ \text{ } \text{TP} \text{ } \text{FP}}$$

Obrázek 4.1. V horní části obrázku je znázornění vzorků spadajících do dvou klasifikovaných tříd, je vyobrazena predikce algoritmu a rozdělení vzorků do skupin FN - falešně negativní predikce, TP - pravdivě pozitivní predikce, FP - falešně pozitivní predikce a TN - pravdivě negativní predikce. V dolní části obrázku jsou grafické vztahy pro často používané metriky sensitivita neboli výtěžnost (Sens), specificita (Spec) a precisnost.

standardem pro vyhodnocování testů nebo modelů. Sensitivita je metrika udávající poměr pravdivě pozitivní segmentace vůči všem segmentacím, které expert označil za pozitivní. Vztah pro výpočet sensitivity je definován vztahem (2). Sensitivita nabývá

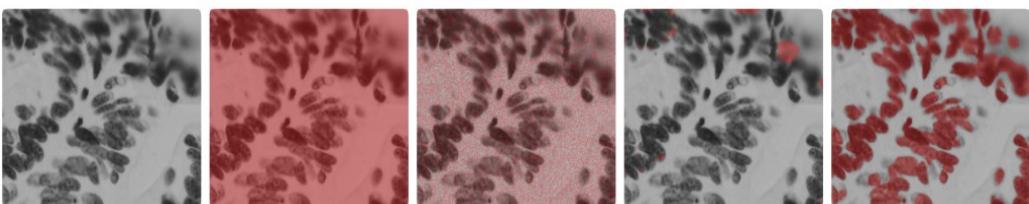
Nádor mozku - malá segmentovaná oblast (0,3 % plochy)

Prázdná predikce	Predikce všech pixelů	Náhodná predikce	Nenatrénovaný model	Natrénovaný model
DC: 0,0	DC: 0,006	DC: 0,006	DC: 0,026	DC: 0,880
IoU: 0,0	IoU: 0,003	IoU: 0,003	IoU: 0,013	IoU: 0,785
Sens: 0,0	Sens: 1,0	Sens: 0,506	Sens: 0,404	Sens: 0,923
Spec: 1,0	Spec: 0,0	Spec: 0,502	Spec: 0,905	Spec: 0,999
Acc: 0,997	Acc: 0,003	Acc: 0,502	Acc: 0,903	Acc: 0,999
KAP: 0,0	KAP: 0,0	KAP: 0,0	KAP: 0,020	KAP: 0,879
AUC: 0,5	AUC: 0,5	AUC: 0,504	AUC: 0,785	AUC: 0,992
AHD: 297	AHD: 437	AHD: 437	AHD: 326	AHD: 5



Histopatologie - velká segmentovaná oblast (35,8 % plochy)

Prázdná predikce	Predikce všech pixelů	Náhodná predikce	Nenatrénovaný model	Natrénovaný model
DC: 0,0	DC: 0,527	DC: 0,418	DC: 0,031	DC: 0,865
IoU: 0,0	IoU: 0,358	IoU: 0,264	IoU: 0,016	IoU: 0,761
Sens: 0,0	Sens: 1,0	Sens: 0,502	Sens: 0,017	Sens: 0,936
Spec: 1,0	Spec: 0,0	Spec: 0,497	Spec: 0,973	Spec: 0,873
Acc: 0,642	Acc: 0,358	Acc: 0,499	Acc: 0,631	Acc: 0,895
KAP: 0,0	KAP: 0,0	KAP: -0,001	KAP: -0,013	KAP: 0,780
AUC: 0,5	AUC: 0,5	AUC: 0,5	AUC: 0,524	AUC: 0,913
AHD: 402	AHD: 67	AHD: 67	AHD: 129	AHD: 29



Obrázek 4.2. Demonstrace chování jednotlivých metrik v závislosti na velikosti expertní segmentace a na kvalitě predikce. Červeně je na obrázcích zobrazena predikce. V horní části se nachází segmentace nádoru mozku, jež má malý obsah vůči celému obrázku. V dolní části jsou pak snímky histopatologie, kde expertní segmentace zabírá více než třetinu plochy obrázku. [29]

hodnot od 0 po 1 od nejhorší po nejlepší segmentaci.

$$Sens = \frac{TP}{TP + FN} \quad (2)$$

Grafické vyjádření sensitivity je prezentováno v již uvedeném obrázku 4.1.

Nicméně sensitivita se příliš nehodí pro porovnávání dvou segmentací a její použití jako evaluační metriky není úplně běžné v oblasti segmentace lékařských dat, jelikož je metrika velmi citlivá na velikost vyhodnocovaného segmentu (chyby v malých vy-

hodnocovaných segmentech jsou penalizovány více než stejné chyby vyhodnocované ve větších segmentech) [30].

4.2 Specificita

Specificita označována také zkratkou Spec nebo TNR (True Negative Rate) měří správnost negativní segmentace (tedy správnost segmentace pozadí) a je tedy analogií k sensitivitě. Specificita také nabývá hodnot mezi 0 a 1, kde 0 je nejhorší a 1 je nejlepší výsledek. Specificita je definována následujícím vztahem:

$$Spec = \frac{TN}{TN + FP} \quad (3)$$

Grafické vyjádření specificity je prezentováno v již uvedeném obrázku 4.1.

Tato metrika však v případě segmentování lézí cévní mozkové příhody, a obecně u segmentací medicínských dat, je nepříliš vypovídající [29]. Konkrétně v případu segmentování lézí cévní mozkové příhody jsou léze objemově velmi malé (typicky v řádu ml až desítek ml) proti objemu nepostižené mozkové tkáně (typicky kolem 1300 ml). Za negativní segmentaci se považuje tkáň nepostižená cévní mozkovou příhodou, tedy i kdyby algoritmus automatické segmentace neosegmentoval žádný voxel jako lézi, dosáhne model specificity blížící se k 1, jelikož počet TN bude velmi velký proti počtu FP. Tato metrika se hodí spíše na ověření, zda daná automatická metoda vůbec funguje. Nehodí se už ale na měření výkonnosti a porovnávání jednotlivých modelů mezi sebou.

Stejně jako u sensitivity je specificita návyklá na velikost segmentu, z něhož je metrika počítána. Tedy i zde platí, že chyby na malých segmentech, ze kterých metriku počítáme, jsou penalizovány více než stejné chyby vyhodnocované z větších navzorkovaných segmentů [30].

4.3 Precisnost

Další velmi známou metrikou je precisnost, označována také jako PPV (Positive Predictive Value). Tato metrika se samostatně běžně nepoužívá pro validaci modelů určených pro segmentaci lékařských dat, ale používá se pro výpočet dalších metrik uvedených v této práci [30]. Precisnost nabývá hodnot mezi 0 a 1 od nejhoršího po nejlepší segmentaci a vypovídá o poměru pravdivě pozitivní segmentace vůči všem automaticky segmentovaným voxelům. Precisnost je definována vztahem (4).

$$Precision = PPV = \frac{TP}{TP + FP} \quad (4)$$

Grafické vyjádření precisnosti je prezentováno v již uvedeném obrázku 4.1.

4.4 Klasifikační přesnost

Klasifikační přesnost, známá také jako Randův index (RI) [31] nebo overall accuracy (Acc), je definována jako poměr korektních (tedy pravdivě pozitivních a pravdivě negativních) segmentací a počet všech prvků v matici dat, formálně vyjádřen vztahem (5) a nabývá opět hodnot mezi 0 a 1 od nejhorší po nejlepší segmentaci.

Tato metrika je citlivá na poměry tříd stejně jako některé již dříve uvedené metriky (primárně specificita). Jelikož léze cévní mozkové příhody jsou typicky malé proti

objemu nezasažené tkáně, jsou data silně nevyvážená a tato metrika by se neměla používat [29], poněvadž při prázdné predikci bude tato predikce vyhodnocena s klasifikační přesností blížící se 1.

$$\text{Accuracy} = \text{Acc} = \frac{TP + TN}{TP + TN + FN + FP} \quad (5)$$

4.5 Sørensen-Dice koeficient

Sørensen-Dice koeficient [32] je označovaný také jako Dice koeficient, DC nebo DSC (Dice similarity coefficient) nebo koeficient překryvu. Jedná se o nejčastěji používanou metriku při vyhodnocení kvality segmentačních algoritmů, která je uváděna ve velké většině studií zpracovávajících segmentaci medicínských dat [30, 29]. Dice koeficient je vhodný na validaci modelů, tzn. je vhodný pro statistickou evaluaci výkonnosti modelu a může být použit jako metrika určující opakovatelnost experimentu, resp. opakovatelnost výsledku modelu na nových datech [30]. Koeficient nabývá hodnot od 0 (reprezentující nekvalitní segmentaci) po 1 (reprezentující excelentní segmentaci).

Dice koeficient lze chápat jako procento překryvu mezi expertní segmentací a segmentací automatického modelu. Koeficient je definován jako harmonický průměr sensitivity a precnosti, formálně zapsán ve vztahu (6).

$$DC = \frac{2}{\frac{1}{Sens} + \frac{1}{PPV}} = \frac{2TP}{2TP + FP + FN} = \frac{2|S_g^1 \cap S_t^1|}{|S_g^1| + |S_t^1|} \quad (6)$$

Alternativně k Dice koeficientu lze použít Jaccardův index (JAC) [33], známý také jako Jaccardův podobnostní koeficient nebo Intersection over Union (IoU). Jaccardův index je definován jako průnik dvou segmentací vydelený jejich sjednocením, formálně vyjádřeno vztahem (7).

$$JAC = \frac{|S_g^1 \cap S_t^1|}{|S_g^1 \cup S_t^1|} = \frac{TP}{TP + FP + FN} \quad (7)$$

Jaccardův index je vyjma extrémů $\{0, 1\}$ vždy menší než Dice koeficient. V bodech extrémů jsou si koeficienty rovny. Vztah mezi Jaccardovým indexem a Dice koeficientem je vyjádřen v rovnici (8).

$$JAC = \frac{|S_g^1 \cap S_t^1|}{|S_g^1 \cup S_t^1|} = \frac{2|S_g^1 \cap S_t^1|}{2(|S_g^1| + |S_t^1| - |S_g^1 \cap S_t^1|)} = \frac{DC}{2 - DC} \quad (8)$$

Vzhledem k tomu, že Jaccardův index lze vyjádřit pomocí Dice koeficientu a opačně, nemá smysl používat Dice koeficient a Jaccardův koeficient společně, jelikož druhá z použitých metrik neposkytne žádnou přidanou informaci. Obě metriky zohledňují stejné příznaky a poskytují tedy stejné hodnocení.

4.6 Hausdorffova vzdálenost

Hausdorffova vzdálenost (Hausdorff Distance - HD) [34] je metrika, jejíž jednotka je prostorová vzdálenost, tudíž Hausdorffovu vzdálenost lze spočítat pouze na datech, které se vyskytují v metrickém prostoru. Její hodnoty nabývají od 0 k nekonečnu v závislosti na rozložení dat v prostoru.

Hausdorffova vzdálenost je symetrická a je definována jako

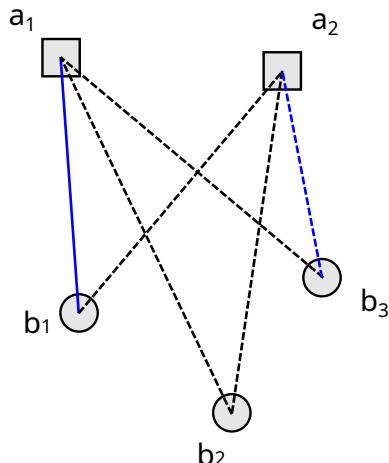
$$HD = \max(h(S_g^1, S_t^1), h(S_t^1, S_g^1)) \quad (9)$$

kde S_g^1 a S_t^1 jsou expertně a automaticky osegmentovaná data, resp. obecně to mohou být dvě jakékoli množiny. Funkci h lze označit za asymetrickou Hausdorffovu vzdálenost, tu pak definujeme následovně

$$h(A, B) = \max_{a \in A} \min_{b \in B} \|a - b\| \quad (10)$$

kde A a B jsou měřené množiny bodů a $\|a - b\|$ je norma mezi těmito body. V případě reprezentace dat použité v této práci jde o Euklidovskou normu. Typicky se HD uvádí v mm při vyhodnocení segmentací lézí cévní mozkové příhody.

Výsledek asymetrické Hausdorffovy vzdálenosti lze vágně popsat jako maximální vzdálenost mezi nějakým bodem jedené množiny nebo segmentace a jejím nejbližším bodem v druhé množině nebo segmentaci. Což je zpracováno na obrázku 4.3.



Obrázek 4.3. Znázornění asymetrické Hausdorffovy vzdálenosti $h(A, B)$. Na obrázku se nachází množina čtverců A a množina kruhů B . Mezi prvky množiny A jsou vyobrazeny vzdálenosti k prvkům množiny B (spojnice mezi prvky). Modré spojnice pak vyznačují nejkratší vzdálenosti od prvků množiny A k prvkům množiny B . Maximální (Hausdorffova) vzdálenost je vyznačena plnou čarou. Všechny prvky množiny A tedy leží nejvýše v Hausdorffově vzdálenosti od množiny B . Z příkladu je evidentní, že výsledek se pro $h(B, A)$ od prezentovaného liší a jde o asymetrickou Hausdorffovu vzdálenost.

Hausdorffova vzdálenost je velmi citlivá na odlehlé hodnoty (tzv. outliers). Pro představu stačí, pokud automatický algoritmus špatně osegmentuje byť jediný voxel, který leží na opačné straně segmentovaného prostoru, než kde je korektní segmentace experta. V takovém případě bude Hausdorffova vzdálenost extrémně velká, jelikož se v predikci vyskytuje falešné ložisko, které leží velmi daleko. Tyto příklady jsou poměrně časté a jsou tedy zdrojem outlierů, kvůli kterým Hausdorffova vzdálenost neúměrně vzroste.

Z popsaných důvodů se mnohem častěji používá průměrná Hausdorffova vzdálenost (AHD), uváděna také jako Average Distance (AVD) [35–36] nebo ASD. AVD lze spočítat pomocí vztahu (9) záměnou funkce h za funkci d definovanou vztahem

$$d(A, B) = \frac{1}{N} \sum_{a \in A} \min_{b \in B} \|a - b\| \quad (11)$$

kde N značí počet měřených prvků množiny A .

4.7 Průměrná symetrická vzdálenost ploch

Průměrná symetrická vzdálenost ploch uváděná pod zkratkou ASSD z anglického Average Symmetric Surface Distance [2, 28] je opět míra, jejíž jednotka je typicky v mm a měří globálně vzdálenost všech bodů první množiny od množiny druhé. Tato metrika nabývá hodnot od nuly k nekončenou v závislosti na velikosti prostoru. Čím menší je hodnota metriky, tím jde o lepší překryv.

Abychom mohli definovat ASSD, je potřeba nejprve definovat průměrnou vzdálenost ploch (ASD). Tato funkce je již uvedena v předchozí sekci jako funkce $d(A, B)$ ve vztahu (11).

Symetrická varianta ASD - tedy ASSD se poté vypočítá podle vztahu

$$ASSD(S_g^1, S_t^1) = \frac{d(S_g^1, S_t^1) + d(S_t^1, S_g^1)}{2} \quad (12)$$

4.8 Cohenovo Kappa

Cohenovo Kappa (KAP) [37] je metrika shody mezi dvěma výběry. Tato metrika byla uvedena původně pro obor psychologie. Výhodou této metriky je, že v rámci hodnocení zohledňuje i náhodu, při níž došlo ke shodě v obou výběrech (segmentacích) - tím se zvyšuje spolehlivost této metriky. Tato metrika dosahuje hodnot mezi -1 a 1 pro nejhorší, resp. nejlepší segmentaci. V případě, že Cohenovo Kappa má hodnotu 0, jde o náhodný klasifikátor.

Cohenovo Kappa je definováno

$$KAP = \frac{P_a - P_c}{1 - P_c} \quad (13)$$

kde P_a je míra shody mezi vzorky a P_c je pravděpodobnost, že vzorek byl shodně označen náhodou. V případě mnoha měření lze vztah (13) přepsat ve formě frekvencí jednotlivých pozorování

$$KAP = \frac{f_a - f_c}{N - f_c} \quad (14)$$

kde N je počet jednotlivých vzorků nebo měření.

Vztah (14) lze již použít spolu se základními parametry [30, 29].

$$f_a = TP + TN \quad (15)$$

$$f_c = \frac{(TN + FN)(TN + FP) + (FP + TP)(FN + TP)}{N} \quad (16)$$

Dle článku [29] Cohenovo Kappa neumožňuje porovnání výsledků na různých datasetech ani neinterpretuje přesnost predikce.

4.9 ROC křivka a AUC

ROC z anglického Receiver Operating Characteristic je křivka grafu, sloužící pro vizualizaci výkonu klasifikátoru při různých nastaveních prahu. Prahování uvažujeme v případě, kdy na výstupu algoritmu je pravděpodobnostní mapa a je nutné zvolit jistý práh pravděpodobnosti, při jehož překročení daný voxel přiřadíme do třídy lézí, případně naopak.

Nejprve je nutné definovat tzv. fallout neboli poměr falešně pozitivních (FPR) vztahem (17), který je doplňkem ke specificitě.

$$\text{Fallout} = \text{FPR} = \frac{FP}{FP + TN} = 1 - \text{Spec} \quad (17)$$

Graf ROC křivky má na svislé ose sensitivitu (True Positive Rate) a na vodorovné ose je vynesen fallout. Pro vynesení křivky do grafu je tedy nutné volit několik různých prahů a provést několik měření.

Z ROC lze spočítat AUC (Area Under the ROC Curve), což je skalární veličina charakterizující ROC křivku, která byla představena J. A. Hanleyem a B. J. McNeilem [38]. V obecném případě, kdy klasifikátor přiřazuje pravděpodobnosti, lze AUC slovy popsat na následující situaci: vybereme náhodně pozitivní a nezávisle na něm náhodně negativní vzorek, AUC poté reprezentuje pravděpodobnost, s jakou klasifikátor ocení pozitivní vzorek vyšší pravděpodobností než vzorek negativní.

V případě, že AUC nabývá hodnoty 0,5, jde o náhodný klasifikátor. V případě nižší hodnoty (jdoucí až k 0) bude klasifikátor predikovat horší výsledky než náhodný klasifikátor. Ideální klasifikátor má AUC 1. AUC je pro současné články poměrně relevantní a často udávaná metrika [29].

V případě pouze jednoho měření (aplikováním pouze jednoho prahu), kdy porovnáváme testovanou segmentaci vůči expertní, lze uvažovat definici AUC podle [39], která vychází z plochy lichoběžníku.

Nejprve definujeme poměr falešně negativních (False Negative Rate) vztahem (18), který je doplňkem k sensitivitě.

$$\text{FNR} = \frac{FN}{FN + TP} = 1 - \text{Sens} \quad (18)$$

AUC je poté definována vztahem (19).

$$\text{AUC} = 1 - \frac{\text{FPR} + \text{FNR}}{2} = 1 - \frac{1}{2} \left(\frac{FP}{FP + TN} + \frac{FN}{FN + TP} \right) \quad (19)$$

Kapitola 5

Metody automatické segmentace ischemické léze

V této kapitole popíší souhrn současného stavu a rozeberu několik článků řešící segmentaci cévních mozkových příhod. Kapitola má tři podsekce, ve kterých se zabývám populárními soutěžemi ISLES [2, 40–41] segmentující subakutní léze; metodami z přehledového článku [42] řešící segmentaci chronické léze; a na závěr popisuji unimodální metodu pro segmentaci subakutních lézí [43]. Metody lze srovnat na základě uváděných výsledků, jimiž jsou metriky, které jsou definovány v předchozí kapitole.

5.1 Srovnávací soutěže ISLES

Soutěže Ischemic Stroke Lesion Segmentation Challenge (dále uváděné jako ISLES) [2, 40–41] konané ve spojení s konferencí MICCAI (International Conference on Medical Image Computing and Computer Assisted Intervention) vznikly v roce 2015 jako reakce na rostoucí zájem o automatickou segmentaci ischemických lézí, na rychlý rozvoj metod a na rostoucí množství publikovaných článků v oblasti segmentace ischemických lézí cévní mozkové příhody.

Přestože se dle autorů všechny tyto dřívější články snažily nějakým způsobem kvantifikovat své výsledky, často chyběl detailní popis datasetu, na kterém byly tyto výsledky získány. Tento popis (jako je čas od vzniku cévní mozkové příhody) je velmi důležitá informace, jelikož velikost, tvar nebo intenzita léze se v čase na různých sekvenčních dramaticky mění, jak již bylo popsáno v této práci. Výsledky mohou kromě času ovlivnit různé modality vstupující do automatického algoritmu, velikost samotného datasetu, na kterém byly výsledky dosaženy nebo metriky, které byly použity. Popsaná nekonzistence a netransparentnost v důsledku znamenala, že články prezentují velmi široké spektrum výsledků, jejichž interpretace není snadná a jejich srovnání je prakticky nemožné.

Před první ISLES soutěží existovala již řada soutěží nebo srovnávacích studií, které řešily např. segmentaci nádorů mozku (BRATS) [44] z MRI, roztroušené sklerózy z MRI [45], rakoviny prostaty z MRI [46], segmentaci mozkové tkáně z MRI [47], plicních uzlů z CT [48] nebo plicních cév z CT [49]. Některé z uvedených soutěží se staly v podstatě standardem pro ověřování jak tehdy moderních algoritmů, tak algoritmů nových. Tyto soutěže mají relevantní a pečlivě vybrané snímky v datasezech, se kterými pracují. Tento výběr dat a velikost datasetu zároveň dostatečně pokrývá danou oblast segmentací medicínských obrazů. Výsledky dosažené v těchto studiích by tedy měly spolehlivě odpovídat reálnému výkonu testovaných metod. Dále tyto soutěže mají jasně definovaná pravidla, podle kterých expert segmentoval oblasti zájmu. Srovnávací metriky nebo hodnotící skóre zúčastněných metod zahrnuje všechny aspekty relevantní pro daný problém. A také tyto soutěže mají často k dispozici zveřejněný dataset a jsou otevřené, aby mohly být srovnány i novodobé algoritmy s algoritmy zúčastněnými v dané soutěži.

Pro segmentaci ischemických lézí cévní mozkové příhody nebyla k dispozici podobná soutěž, jaké jsou uvedené v předchozím odstavci. V důsledku výše popsaného byla na jaře roku 2015 zveřejněna a na podzim vyhodnocena první soutěž ISLES 2015 [2], jako doprovodný program MICCAI. Cílem této soutěže bylo srovnat nejmodernější metody segmentace ischemických lézí cévní mozkové příhody.

Soutěž ISLES 2015 [2] byla tedy zaměřena na srovnání mnoha automatických segmentačních metod ischemické cévní mozkové příhody z multimodálních MRI dat. Byly zveřejněny dva rozmanité datasety pokrývající dvě úlohy, na které byla soutěž ISLES 2015 [2] rozdělena: SISS (sub-acute ischemic stroke lesion segmentation) a SPES (acute stroke pneumbra estimation). Jelikož je tato srovnávací studie relevantní k problému, který řeší tato práce, tak bude ISLES 2015 [2] podrobněji rozebrán v následující podkapitole 5.1.1.

Po úspěchu, v roce 2016, proběhla další soutěž ISLES 2016 [41]. Motivace pro uspořádání tohoto ročníku byla stejná jako motivace u ročníku předchozího, avšak zásadně se lišily úkoly, které participanti řešili. V tomto ročníku byl zveřejněn multimodální dataset MRI skenů akutní cévní mozkové příhody a patřičné expertní segmentace - ty však byly anotovány na skenech, které byly provedeny až po třech až devíti měsících. Kromě segmentací bylo ke každému skenu k dispozici ještě klinické mRS skóre (modifikovaná Rankinova škála), získané ze škály, podle které se hodnotí míra tělesného postižení po prodělání cévní mozkové příhody. Nebo byly v rámci datasetu poskytnuty i další klinické informace (jako je např. čas od vzniku CMP). Celá soutěž tedy opět byla rozdělena na dvě různé podsoutěže - první z nich byla zaměřena na predikci a segmentaci finální léze z MRI obrazu a druhá byla zaměřena na odhad mRS skóre. ISLES 2016 obsahoval 35 trénovacích subjektů a 19 testovacích, na kterých byla soutěž vyhodnocena.

V následující soutěži ISLES 2017 [41] zůstaly úkoly stejné jako v ISLES 2016 a v zásadě došlo pouze k rozšíření datasetu o 8 pacientů v trénovací fázi (tedy veřejné části datasetu) a o 13 pacientů v testovací fázi.

Soutěž ISLES 2018 byla zaměřena na segmentaci akutních ischemických lézí z dat CT perfúze. Byl vytvořen nový dataset čítající snímky 103 pacientů postižených mrtvicí spolu s příslušnými expertními segmentacemi k nim. Z datasetu byla uveřejněna trénovací množina o velikosti 63 pacientů. Dataset obsahoval snímky CT perfúze, DWI z magnetické rezonance a manuální expertní segmentaci léze na DWI. Snímky DWI byly pořízeny během 3 hodin od snímkování CT. Úkolem účastníků bylo vytvořit algoritmy, které ze vstupní dat CT perfúze vytvoří odpovídající segmentaci léze, kterou by expert vytvořil, kdyby provedl MRI a segmentoval lézi na sekvenci DWI.

Poslední ISLES soutěž, která byla dokončena, proběhla v roce 2022 [40]. Autoři zhodnocují, že ISLES 2015 [2] a ISLES 2018 hrály zásadní roli v nalezení metod, které opravdu dobře fungují a dokáží řešit předložené problémy. Proto lze ISLES 2022 [40] chápat jako rozšíření soutěže ISLES 2015 [2], kdy autoři ISLES 2022 [40] připravili soutěž na základě již získaných zkušeností. Jelikož zadání ISLES 2022 [40] odpovídá i zadání této práce, rozeberu soutěž podrobněji v následující podkapitole 5.1.2.

A pro úplnost doplním, že aktuálně probíhá soutěž ISLES 2024, která je nejspíše rozšířením nebo alternativou k ISLES 2018, jelikož i zde jde o segmentaci léze ze snímků CT perfúze. Cílem je vytvořit predikci segmentace po zaléčení mrtvice ze snímků pořízených před léčbou. Soutěž však zatím není vyhodnocená, datasety nejsou veřejné a zatím k této soutěži nebyl autory publikován vědecký článek.

5.1.1 ISLES 2015

Jak jsem již uvedl, soutěž ISLES 2015 [2] byla zaměřena na férové porovnání plně automatických segmentačních metod ischemických lézí, avšak mimo to bylo cílem odpovědět na otázky, jako jsou: jaká je reálná úspěšnost moderních segmentačních metod; jaké přístupy resp. metody jsou vhodné a dosahují zajímavých výsledků; jaké problémy dokáží moderní metody překonat a jaké problémy zůstávají stále výzvou; jaké závěry a doporučení předat dalším řešitelům stejných úloh.

Aby bylo možno dosáhnout stanovených cílů, bylo nutné zveřejnit relevantní trénovací datasety, na kterých účastníci soutěže nachystané metody natrénují. Zároveň bylo třeba vytvořit testovací datasety, které nebudou uveřejněny, aby byla zaručena férovorost soutěže. Autory byl také stanovený striktní postup pro expertní segmentaci (mj. aby zde byla konzistence) a také postup pro srovnávání jednotlivých metod.

Do soutěží se registrovalo celkem 16 výzkumných týmů, ve kterých bylo registrováno celkem 120 lidí. Týmy (účastníci) se mohli registrovat do SISS i SPES s tím, že do SISS se registrovalo 14 týmů a do SPES 7 týmů. Každý tým pro každou soutěž obdržel anotovaný trénovací dataset s daty příslušnými pro konkrétní soutěž a týden před termínem odevzdání byl týmům rozdistribuován testovací dataset. Testovací dataset však neobsahoval expertní segmentace, obsahoval pouze MRI snímky. Účastníci mohli během týdne opakovat nahrání svého algoritmu do systému pro evaluaci, započítalo se jim ale pouze poslední dosažené skóre.

Tyto obecné informace platí pro obě podsoutěže SISS a SPES. Protože dataset po-skytnutý Motolem, pro který byl v rámci této diplomové práce doručen automatický segmentační systém, obsahuje data odpovídající převážně subakutní fázi, je odpovídajícím zdrojem pouze podsoutěž SISS, a proto se dále budu zabývat pouze tou.

Pro soutěž SISS obsahoval trénovací dataset pouze 28 snímků, které byly pořízeny v jediném zdravotnickém centru a expertní segmentaci provedl pouze jeden člověk. Testovací dataset obsahoval 36 snímků, které byly pořízeny ve dvou zdravotnických centrech a každého pacienta segmentovali dva experti. Veřejně jsou k dispozici sekvence FLAIR, T2 TSE, T1 TFE/TSE, DWI a expertní segmentace k trénovacímu datasetu. Trénovací dataset ISLES 2015 [2] je použit v praktické části této diplomové práce, proto je jeho podrobnější analýza uvedena později v samostatné kapitole o použitých datasetech v sekci 8.2.

Jako evaluační metriky byly použity Dice koeficient (DC), Hausdorffova vzdálenost (HD) a průměrná symetrická vzdálenost ploch (ASSD). Následně pro každého pacienta a každou evaluační metriku bylo vytvořeno pořadí úspěšnosti metod. Toto pořadí bylo u každého pacienta pro všechny tři metriky zprůměrováno a nakonec se výsledné umístění získalo jako průměr pořadí ze všech pacientů.

Výsledky soutěže jsou uvedeny v tabulce 5.1. Nejprve je důležité zmínit Dice koeficient, který vznikl porovnáním segmentací dvou expertů, kteří anotovali testovací dataset: 0,70. Znamená to tedy, že i segmentace dvou odborníků se poměrně zásadně liší, resp. překryvy segmentací nejsou dokonalé. Sami autoři ISLES uvádí příklad, že při segmentaci nádorů mozku [44] dosahuje inter-rater Dice koeficient až $0,85 \pm 0,02$, což je významný rozdíl proti ischemickým lézím.

Dále je zajímavé, že každý z algoritmů alespoň u jednoho pacienta z testovacího datasetu nedokázal odhalit lézi (překryv automatické a expertní segmentace byl prázdný neboli Dice koeficient je nulový). A dále je v Dice koeficientech značná variabilita.

Ukázalo se, že Dice koeficienty, které jsou uvedeny v tabulce 5.1, dosáhly poměrně horšího skóre než výsledky tehdy moderních metod publikovaných individuálně ve vě-

Metoda (tým)	Úspěšnost	ASSD [mm]	DC	HD [mm]
UK-Imp2	34/36	05,96 ± 09,38	0,59 ± 0,31	37,88 ± 30,06
CN-Neu	32/36	03,27 ± 03,62	0,55 ± 0,30	19,78 ± 15,65
FI-Hus	31/36	08,05 ± 09,57	0,47 ± 0,32	40,23 ± 33,17
US-Odu	33/36	06,24 ± 05,21	0,43 ± 0,27	41,76 ± 25,11
BE-Kul2	33/36	11,27 ± 10,17	0,43 ± 0,30	60,79 ± 31,14
DE-UzL	31/36	10,21 ± 09,44	0,42 ± 0,33	49,17 ± 29,60
US-Jhu	33/36	11,54 ± 11,14	0,42 ± 0,32	62,43 ± 28,64
UK-Imp1	34/36	11,71 ± 10,12	0,44 ± 0,30	70,61 ± 24,59
CA-USher	27/36	09,25 ± 09,79	0,35 ± 0,32	44,91 ± 32,53
BE-Kul1	30/36	12,24 ± 13,49	0,37 ± 0,33	58,65 ± 29,99
CA-McGill	31/36	11,04 ± 13,68	0,32 ± 0,26	40,42 ± 26,98
SE-Cth	30/36	10,00 ± 06,61	0,38 ± 0,28	72,16 ± 17,32
DE-Dkfz	35/36	14,20 ± 10,41	0,33 ± 0,28	77,95 ± 22,13
TW-Ntust	15/36	07,59 ± 06,24	0,16 ± 0,26	38,54 ± 20,36
inter-rater	36/36	02,02 ± 02,17	0,70 ± 0,20	15,46 ± 13,56

Tabulka 5.1. Výsledky soutěže ISLES 2015 (průměr ± směr. odchylka) v pořadí odpovídajícím umístění daných metod. Poslední řádek „inter-observer“ reprezentuje výsledky dosažené porovnáním dvou nezávislých segmentací od dvou expertů, kteří anotovali testovací dataset ISLES 2015, jehož masky nejsou veřejné. Sloupec s úspěšností v tomto případě značí, u kolika pacientů byl neprázdný překryv expertní segmentace a segmentace automatického algoritmu. ASSD: Průměrná symetrická vzdálenost ploch. DC: Dice koeficient. HD: Hausdorffova vzdálenost. [2]

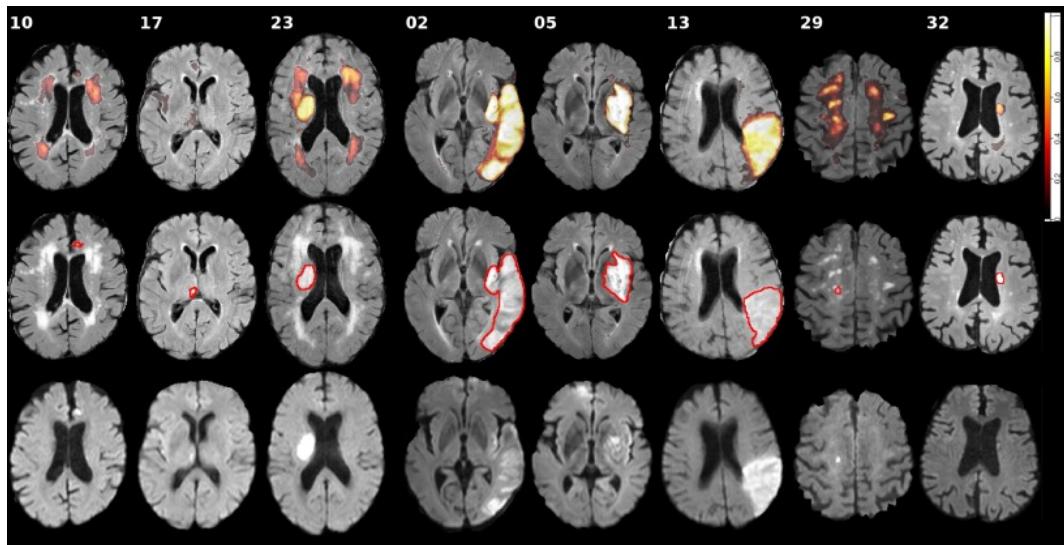
deckých článcích, často s neveřejnými datasety. Dice koeficienty dosažené v této soutěži dokazují obtížnost segmentace subakutních ischemických lézí.

Vybrané příklady z testovacího datasetu jsou uvedeny na obrázku 5.1. Obecně měly všechny algoritmy problém se segmentací malých lézí se slabým FLAIR signálem a většina algoritmů měla problém s odlišením subakutní ischemické léze od tzv. hyperintenzity v bílé hmotě (WMH), které se typicky vyskytují v okolí mozkových komor, jsou časté u starší populace a nejsou objektem zájmu při segmentaci subakutní ischemické CMP. WMH jsou patrné u pacientů 10 a 23 na obrázku 5.1. Na druhou stranu všechny algoritmy dosahly vysokých Dice koeficientů u velkých lézí s jasnou hyperintenzitou na FLAIR.

Statistickou analýzou autoři zjistili, že první dvě metody nedosahují signifikantně odlišných výsledků (na základě DC) a žádný z ostatních algoritmů nedosahuje signifikantně lepších výsledků než tyto první dvě metody.

Velmi důležitým výstupem soutěže však bylo zjištění, že každý z testovaných algoritmů dosáhl extrémně špatných výsledků na testovacích datech, pocházejících z druhého zdravotnického centra než na testovacích datech ze zdravotnického centra, ze kterého pocházela i trénovací data. Jinými slovy algoritmy dobře zobecňují pro nová data, která byla pořízena podobnými parametry skenu, ale nedokáží se adaptovat na variace způsobené jiným nezávislým zdrojem dat. Tento jev však mohl vzniknout tím, že byl vzorek dat (8 pacientů) z druhého datacentra malý a léze v něm obsažené byly pouze léze malého objemu.

Dalším výsledkem srovnávací studie je, že při validaci stejných pacientů proti segmentaci nového (druhého) experta, nedošlo k žádným výrazným změnám v Dice koeficientu (přestože algoritmy nebyly natrénovány na žádných snímcích segmentovaných druhým



Obrázek 5.1. Výsledky ISLES 2015. První řádek zobrazuje distribuci segmentace testovaných metod na pozadí FLAIR. Druhý řádek obsahuje vyznačenou expertní segmentaci léze na pozadí FLAIR. Třetí řádek zobrazuje příslušné sekvence DWI. Pacienti 2, 5 a 13 mají velké, jasně ohraničené léze, na nichž všechny algoritmy vykázaly vynikající výsledky. Se segmentací pacientů 10, 17 a 23 naopak mají algoritmy problém - jedná se buď o malé léze nebo pacienty s WMH. Snímky pacientů 29 a 32 pochází z druhého datacentra. [2]

expertem). Pouze v jednom případě nejlepší metoda týmu UK-Imp2 vykázala signifikantní rozdíl, který byl ale ve prospěch onoho druhého experts. Autoři to vysvětlují tím, že vlivem relativně vysoké variability segmentací expertů nedošlo k přeúčení strojového učení pouze na jednoho experts. Dalším možným vysvětlením je, že kritická místa, kde nedošlo k překryvům segmentací obou experts, nenesou jednoznačnou nebo dostatečnou informaci o výskytu léze.

Signifikantně koreloval i Dice koeficient s velikostí léze, stejně tak Dice koeficient s lateralitou. Lateralita však také signifikantně korelovala s velikostí léze. Z korelací proto nelze vyvodit jednoznačné závěry, jelikož samotný Dice koeficient je ovlivněný velikostí léze.

V zásadě autoři ISLES 2015 [2] nemohou rozhodnout, jaký z druhů algoritmů (hluboké učení, náhodné lesy nebo fuzzy c-means), které účastníci používali, se hodí pro segmentaci nejvíce. Například algoritmus založený na náhodných lesích se umístil jak na třetím místě, tak na předposledním místě. Autori proto zmiňují, že klíčem je ladění hyperparametrů modelu a jeho úzká adaptace na předložený problém.

Autoři srovnávací studie doporučují, aby se nevyhodnocovaly nové segmentační metody na uzavřených datasetech, jelikož ty bývají příliš malé na to, aby rozmanité pokrývaly skeny obsahující např. různé velikosti lézí, WHM, chronické léze, ...

Závěrem autorů upozorňují, že v blízké budoucnosti nejspíše nebude dosaženo dostačující spolehlivého plně automatického řešení problému. Segmentační metody mohou sloužit jako pomoc při segmentaci, ale úloha je příliš složitá na to, aby algoritmy mohly pracovat plně automaticky, zvláště u malých lézí.

Jestliže se podíváme na jednotlivé účastníky, tak tým UK-Imp2 (Konstantinos Kamnitsas et al.) použil 11 vrstev konvoluční sítě se dvěma paralelními vstupy, které se lišily v rozlišení vstupního obrazu. Tato metoda byla následně zveřejněna a je dále rozvíjena pod názvem DeepMedic [50–51]. DeepMedic [50–51] je blíže popsán v kapitole 9, protože byl experimentálně ověřen v rámci této práce.

Druhá metoda v pořadí CN-Neu od autora Chaolu Feng je založená na fuzzy c-means a není tedy závislá na trénovacích datech. Nejprve jsou provedeny segmentace bílé hmoty, šedé hmoty, mozkomíšního moku a ischemické léze na vstupních modalitách. Segmentace jsou následně fúzovány pomocí většinového hlasování a následně jsou hranice lézí opraveny pomocí Level Set metody.

Na třetím místě se umístil tým Hanna-Leena Halme et al. s metodou FI-Hus založenou na náhodných lesích. Nejprve byla data koregistrována na společnou šablonu a následně byl pro každý voxel a každou modalitu vypočítán průměr intenzity a směrodatná odchylka intenzity. Náhodný les byl natrénován na příznacích, jako jsou normalizovaná intenzita voxelu, odchylka od průměrné intenzity šablony, asymetrie v intenzitách napříč hemisférami, atd.

5.1.2 ISLES 2022

ISLES 2022 [40, 52] značně rozšířil SISS ISLES 2015 [2] - konkrétně autoři vytvořili dataset o 400 pacientech. Dataset je mnohem rozmanitější než u soutěže ISLES 2015 [2], avšak obsahuje odlišné modality (DWI, ADC a FLAIR) a obsahuje i akutní léze. Úkol soutěžících byl stejný jako v soutěži SISS ISLES 2015 [2]. ISLES 2022 [40] představuje unikátní výzvu obsahující široké spektrum mrtvic, v nichž jsou zahrnutы léze různých velikostí v rozličných lokacích mozku, mnohočetné léze a léze s komplexními tvary, atd. Bližší popis datasetu se nachází v kapitole 8.3.

ISLES 2022 [40] byl navržen pro nalezení algoritmu nebo strategie, která dokáže co nejlépe segmentovat ischemické léze na klinických datech, obsahujících mnoho variací. Takový algoritmus by měl být schopen dostatečného zobecňování i na datech pocházejících mimo soutěž a měl by být tudíž vhodný pro nasazení v reálném světě.

Soutěžícím tedy byla rozdistribuována data 250 pacientů, která byla pouze anonymizována a nedošlo k jakémukoliv dalšímu zpracování. Tím byly simulovány podmínky reálného světa a výstupem týmů by měl být opravdu obecný algoritmus, který zahrnuje výběr modality, registraci nebo normalizaci dat. Testovací množina obsahovala 150 pacientů, z toho 50 snímků pacientů z datacentra, které není obsaženo v trénovací množině.

Do soutěže se registrovalo celkem 476 lidí a celkem 20 týmů nahrálo své řešení do testovací fáze, která předcházela přijetí algoritmu autory soutěže. Z toho 8 týmů bylo vyřazeno autory pro nedodržení nějakého z pravidel soutěže.

Pro vyhodnocení soutěže byl použit stejný postup jako při vyhodnocení ISLES 2015 [2] - tedy nejprve seřazením výsledků týmů pro jednotlivé metriky a následným postupným průměrováním pořadí. Lišily se ale použité metriky. Nyní to byl DC, absolutní rozdíl predikovaného a expertního objemu (AVD), F1 skóre a absolutní rozdíl počtu predikovaných ložisek a počtu expertem odhalených ložisek (ALD).

Výsledky soutěže jsou uvedeny v tabulce 5.2. Všechny zúčastněné metody byly založené na deep-learningu. Celkem 8 metod vycházelo z architektur konvolučních sítí nnU-Net [53] nebo U-Net. Ve dvou případech byla implementována metoda ResU-Net, ale tyto týmy se umístily na 10. a na posledním místě. Ve dvou zbylých případech byla použita síť Factorizer a SegResNet.

Jako ztrátovou funkci účastníci nejčastěji využívali součet Dice koeficientu s kategorickou křížovou entropií (v 7 případech). Tato kombinace zároveň obecně dosahovala nejlepších výsledků. Ve dvou případech účastníci použili pouze Dice koeficient jako ztrátovou funkci, ale umístili se až na 8. a 9. místě. Tým, který použil pouze kategorickou křížovou entropii se umístil na předposledním místě. Na druhém a posledním místě se umístil tým využívající součet Dice koeficientu a fokální ztrátové funkce.

Metoda (tým)	DC	AVD [ml]	F1	ALD
SEALS	$0,82 \pm 0,12$	$1,63 \pm 5,43$	$0,86 \pm 0,19$	$1,00 \pm 3,00$
NVAUTO	$0,82 \pm 0,12$	$1,63 \pm 4,27$	$0,80 \pm 0,23$	$2,00 \pm 2,75$
SWAN	$0,81 \pm 0,15$	$1,94 \pm 4,40$	$0,80 \pm 0,24$	$2,00 \pm 2,75$
PAT	$0,82 \pm 0,15$	$1,95 \pm 4,71$	$0,82 \pm 0,30$	$1,00 \pm 3,00$
CTRL	$0,80 \pm 0,14$	$2,15 \pm 5,32$	$0,79 \pm 0,23$	$1,50 \pm 3,00$
VICOROB	$0,78 \pm 0,18$	$2,04 \pm 5,29$	$0,74 \pm 0,31$	$2,00 \pm 3,00$
PLORAS	$0,76 \pm 0,16$	$1,69 \pm 4,71$	$0,73 \pm 0,33$	$2,00 \pm 3,00$
MIRC	$0,76 \pm 0,21$	$2,78 \pm 6,12$	$0,67 \pm 0,27$	$2,00 \pm 4,00$
CREATIS	$0,71 \pm 0,18$	$3,30 \pm 7,16$	$0,67 \pm 0,25$	$3,00 \pm 5,00$
Dolphins	$0,74 \pm 0,21$	$2,44 \pm 6,15$	$0,60 \pm 0,38$	$3,00 \pm 5,75$
A55972	$0,06 \pm 0,22$	$10,49 \pm 29,17$	$0,44 \pm 0,44$	$4,00 \pm 6,78$
AICON	$0,40 \pm 0,50$	$18,86 \pm 31,59$	$0,26 \pm 0,19$	$5,00 \pm 6,00$

Tabulka 5.2. Výsledky soutěže ISLES 2022 (medián \pm IQR) v pořadí odpovídajícím umístění daných metod. DC: Dice koeficient. AVD: Průměrný absolutní rozdíl predikovaného objemu léze a objemu léze nalezené expertem. F1: F1 skóre. ALD: Průměrný absolutní rozdíl počtu nalezených lézí automaticky a expertem. [52]

Účastníci na prvních dvou místech a na posledním místě používali pouze DWI a ADC sekvence. Nejčastěji týmy používaly všechny tři modality, které byly k dispozici (v 8 případech). Na 10. místě skončil tým používající pouze sekvenci DWI.

Tým SEALS [54] snímky transformoval na isotropní velikost voxelu (1, 1, 1) mm a normalizoval snímky. Autoři použili augmentaci dat pomocí převracení obrazu podél os a aplikací Gaussovského šumu. Tým natrénoval nnU-Net [53], konkrétně jeho konfiguraci 3D full-resolution U-Net. Trénovací dataset účastníci rozdělili v poměru 1/7 za účelem vytvoření vlastního evaluačního datasetu. Většina takto rozdělených dat sloužila pro natrénování pěti modelů pomocí křížové validace.

Tým firmy Nvidia NVAUTO [55] převzorkoval vstupní obrazy opět na isotropní velikost voxelu a normalizoval snímky. Byla použita augmentace dat převracením podél os, rotací, posunem, změnou intenzity, změnou velikosti, vyhlazením, ... Následně tým natrénoval SegResNet architekturu pro sémantickou segmentaci implementovanou ve frameworku MONAI. Aby se neuronová síť vešla do paměti grafiky, autoři mohli použít vstupní matici o rozměrech maximálně (192, 192, 128). Z toho důvodu autoři oříznou vstupní obraz tak, aby v něm bylo obsaženo co nejvíce tkáně. Jako optimalizátor byl použit AdamW. Bylo natrénováno celkem 15 modelů v rámci tří pětinásobných křížových validací. Každý z modelů byl nejprve předtrénován na datasetu BRATS 2021, určenému pro segmentaci nádorů mozku.

V této soutěži, ve srovnání s ISLES 2015 [2], došlo k poměrně výraznému zlepšení úspěšnosti modelů napříč různými metrikami. Úspěšné jsou týmy s implementací poměrně širokých a komplexních metod, které fúzují data z mnoha natrénovaných modelů. Otázkou je, zda pokrok např. ve zlepšení Dice koeficientů proti roku 2015 je dán dokonalejším a rychlejším hardwarem - hlavně grafickými kartami (tým NVAUTO totiž uvádí, že trénování probíhalo na 8 grafických kartách NVIDIA V100) či zda jsou opravdu dokonalejší samotné metody. Jedním z obecných, ale důležitých rozdílů je, že v tomto ročníku se nacházely pouze metody založené na neuronových konvolučních sítích.

5.2 Segmentace chronické mrtvice z T1-w

Další přehledová studie [42] z roku 2019 se zaobírá segmentováním chronické CMP (segmentace lézí starých několik týdnů až let) ze sekvence T1-w. Motivací pro použití T1-w bylo, že se nejčastěji používá pro výzkum a účely rehabilitace, jelikož tato sekvence má vysoké rozlišení a je poměrně časově i finančně levná. Zároveň T1-w ukazuje dobrý kontrast zasažené tkáně od 2 týdnů od počátku mrtvice. Dle autorů doposud chyběla taková studie z důvodu chybějícího dostatečně velkého datasetu s T1-w sekvencemi a z důvodu velké časové náročnosti manuální segmentace.

Cíl studie byl podobný jako cíl u ISLES, tedy nalezení nejlépe fungujících metod a případná identifikace nedostatků těchto metod. Autoři v článku [42] přímo zmiňují soutěže ISLES a berou je jako dobrý příklad, který však není cílený na T1-w unimodální segmentace chronické CMP. Oproti ISLES zde však nejde o soutěž, ale autoři studie provedli rešerši metod souvisejících se segmentací mrtvice z let 2007 až 2017. Identifikovali celkem 4 články vhodné pro tuto srovnávací studii. Zajímavé je, že žádny z vybraných článků neobsahuje metody hlubokého učení. Principy těchto metod jsou popsány dále v podkapitolách.

Srovnávané metody byly buď již předučené nebo byly natrénovány na datech od jejich původních autorů. Metody poté byly testovány na části datasetu ATLAS [56], konkrétně na 181 skenech pocházejících z 8 různých skenerů. Část datasetu byla použita z důvodu vyřazení nevyhovujících snímků a expertní segmentace byly autory dále revidovány a upraveny tak, aby každý snímek obsahoval právě jedno ohraničené ložisko, jelikož i testované algoritmy byly testovány pouze na jediném ložisku v jednom skenu.

Z výsledných statistik automatických metod byly odstraněny snímky, u kterých metoda ALI nebo LINDA měla prázdnou segmentaci. Takové případy jsou velmi časté. Všechny tři automatické metody selhaly u 32 pacientů (převážně šlo o léze malé, 30 pacientů spadalo do 33 % percentilu objemů), což autoři přičítají velmi omezeným trénovacím datasetům. Výsledky jsou zobrazené v tabulkách 5.3 a 5.4. Pro automatické metody ALI, lesionGnb a LINDA jsou proto vypočítané na 132 pacientech. Výsledky semi-automatické metody Clusterize jsou na všech 152 pacientech.

Metoda	F	ASSD [mm]	DC	HD [mm]
Clusterize	-	12,64 ± 7,68	0,18 ± 0,31	80,89 ± 36,60
ALI	52	9,58 ± 13,17	0,40 ± 0,44	62,79 ± 48,49
lesionGnb	39	8,75 ± 7,89	0,42 ± 0,37	58,19 ± 25,22
LINDA	68	4,97 ± 13,98	0,50 ± 0,61	36,34 ± 42,48

Tabulka 5.3. Výsledky segmentace chronické léze z T1-w (medián ± IQR). F: Počet prázdných segmentací nebo segmentací bez překryvu s expertem. ASSD: Průměrná symetrická vzdálenost ploch. DC: Dice koeficient. HD: Hausdorffova vzdálenost. [42]

Dle vyhodnocení fungovala nejlépe metoda LINDA, ale pouze v případě, že dokázala lézi odhalit. Nicméně všechny metody fungovaly nejhůře při segmentaci malých lézí, při segmentaci lézí v mozečku a v mozkovém kmeni (léze v těchto oblastech jsou však menší). Nejlépe byly segmentovány léze nacházející se v kortikální oblasti, hůře v subkortikální. Vzhledem k výsledkům zvolených metrik, které jsou rozlišné, se ukázalo, že každá z metod má své výhody i nevýhody. Např. metoda lesionGnb odhalila lézi u všech pacientů, ale ostatní výsledky nemá příliš dobré. Naopak metoda LINDA dosáhla velmi dobrého Dice koeficientu a ostatních metrik, ale jinak u velkého počtu pacientů selhala.

Metoda	Precisnost	Sensitivita
Clusterize	$0,11 \pm 0,22$	$0,89 \pm 0,26$
ALI	$0,31 \pm 0,45$	$0,61 \pm 0,51$
lesionGnb	$0,29 \pm 0,33$	$0,80 \pm 0,44$
LINDA	$0,60 \pm 0,63$	$0,59 \pm 0,63$

Tabulka 5.4. Precisnost a sensitivita segmentace chronické léze z T1-w (medián \pm IQR). [42]

Výsledky dále ukazují zajímavě velký rozdíl precisnosti a sensitivity u metod ALI a lesionGnb. Ukázalo se, že obě metody používající učení s učitelem dosahují vyšších Dice koeficientů jak metoda ALI, která nepotřebuje trénovací data.

Co se týče semi-automatické metody Clusterize, její autoři doporučují manuální korekci po manuálním výběru segmentovaného klastru. Korekce ve srovnávací studii nebyla prováděna, protože by pak šlo prakticky o manuální segmentaci. Metoda tak dosáhla extrémně nízkého Dice koeficientu. Díky expertovi provádějícímu manuální krok Clusterize dosáhla velmi vysoké sensitivity.

Na výsledky mohly mít vliv i použité nástroje pro práci s daty (ANTs, SPM, ...), jelikož jednotlivé metody používaly odlišné nástroje.

Zajímavé je, že dosažené Dice koeficienty v rámci srovnávací studie byly o 0,2 až 0,24 nižší než koeficienty udávané v původních článcích testovaných metod. To dle autorů mohlo vzniknout tak, že původní autoři použili data z jednoho zdrav. centra nebo data z jediného skeneru. Další příčinou mohlo být, že volitelné parametry testovaných metod byly převzaty od původních autorů a tudíž nebyly optimalizovány na testovací dataset srovnávací studie.

Autoři doporučují trénovat budoucí algoritmy na rozmanitých datasetech z mnoha skenerů a více zdravotnických center. Zároveň znalost polohy a velikosti testovaných lézí může vést k optimalizaci parametrů algoritmů. Dále doporučují používat výhradně plně automatické metody s učitelem, ale po vytvoření predikce je nutná visuální inspekce experta.

V následujících podkapitolách popíšu jednotlivé metody ze srovnávací studie [42].

5.2.1 Clusterize

První testovanou metodou byla Clusterize [57]. Jedná se o semi-automatickou metodu vyvinutou v roce 2012 pro sledování ztráty myelinu během leukodystrofie ze sekvence T2-w. Clusterize však lze použít i pro segmentace chronické CMP z T1-w. První krok této metody spočívá ve vyhledání lokálních maxim ve 2D řezech snímku a následného rozšiřování okolí od téhoto maxim v závislosti na prahování intenzity voxelů vůči lokálnímu maximu. Druhý krok metody spočíval v tom, že expert manuálně vybral klastry, které co nejlépe překrývaly segmentovanou lézi.

5.2.2 Automatic Lesion Identification

Zbylé tři testované metody jsou plně automatické. První takovou je metoda Automatic Lesion Identification (ALI) [58], jedná se o učení bez učitele a je založená na fuzzy klastrování. Při klastrování se používá kontrolní vzorek zdravé tkáně, díky kterému je možné najít odlehle hodnoty představující lézi na testovaném snímku.

Porovnávání se děje voxel po voxelu. Autoři nejprve provedou segmentaci tkání bílé a šedé hmoty za pomoci pravděpodobnostních map tkání v toolboxu SPM, s možností iterativního zlepšování téhoto segmentací. Následně se segmentace téhoto dvou tkání

porovnají se segmentacemi kontrolních zdravých tkání a pomocí fuzzy klastrování se vyhledají se odlehle hodnoty. Výstupem je fuzzy množina odpovídající sjednocení nalezených odlehlych hodnot pro obě tkáně. Při implementaci této metody je třeba mít vzorek zdravých pacientů, ze kterých se nejdříve vygenerují kontrolní data.

5.2.3 Naivní Bayesovský klasifikátor

Druhá plně automatická metoda v porovnávací studii označená jako lesionGnb [59] je založená na naučení naivního Bayesovského klasifikátoru z pravděpodobnostních map bílé a šedé hmoty a mozkomíšního moku.

Nejprve je nutné vygenerovat pravděpodobnostní mapy pomocí SPM, mapy jsou následně společně s originálním T1-w obrazem transformovány do prostoru MNI a vyhlazeny Gaussovským kernelem. Poté jsou z pravděpodobnostních map vygenerovány dvě mapy příznaků, pro jejichž vygenerování je potřebná znalost, v jaké hemisféře se léze nachází. První příznaková mapa obsahuje informaci o chybějící tkáni, která odpovídá jádru chronické léze. Vztah pro výpočet mapy je odvozen ze znalosti, že SPM segmentuje chronickou lézi spíše jako mozkomíšní mok než jako bílou nebo šedou hmotu. Druhá příznaková mapa obsahuje informaci o abnormální tkáni a vztah pro vygenerování mapy vychází z faktu, že tato tkáň má podobné intenzity na T1-w jako šedá hmota. Mapy příznaků jsou následně použity pro natrénování klasifikátoru. Pro získání segmentace byl výstup prahován.

5.2.4 LINDA

Poslední testovaná metoda LINDA [60] implementuje náhodné lesy, které vytváří predikce jak na základě voxelu, tak jeho okolí.

LINDA nejprve získá 12 příznaků z T1-w snímků - zvolené příznaky zohledňují levoprávou asymetrii nebo rozdíl intenzit proti kontrolním datům bez léze nebo rozdíl intenzit proti atlasu. Extrahované příznaky tvoří trénovací matici, kde jsou všichni pacienti a sloupce matice odpovídají příznakům jednoho voxelu a příznakům jeho bezprostředního okolí. Autoři převzorkují snímky pacientů tak, aby získali různé rozlišení. Do prvního náhodného lesa vstupuje nejhrubší rozlišení a výstupní pravděpodobnostní mapa je zahrnuta jako další příznak do druhého náhodného lesa, který je natrénován na jemnějším rozlišení. Autoři zvolili celkem 3 různá rozlišení, na kterých natrénují náhodné lesy. Pravděpodobnostní mapa tvořící predikci třetího náhodného lesa je přivedena na segmentaci léze.

Pro správné generování příznaků je nutné, aby léze byla vždy pouze v levé hemisféře. Autoři srovnávací studie proto převrátili snímky, které měly lézi v pravé hemisféře, podél sagitální roviny.

5.3 Segmentace pomocí Markovovských náhodných polí

Článek [43] z roku 2019 navrhoje postup segmentování subakutních lézí z jediné modality FLAIR pomocí Markovovských náhodných polí (MRF). Zaměřuje se pouze na modalitu FLAIR, jelikož ta je typickou sekvencí prováděnou prakticky vždy při diagnostice CMP. Výhoda unimodální automatické metody tedy spočívá i při segmentaci historických dat.

Autoři vytvořili dataset subakutní CMP sloučením 23 snímků z SFB databáze, 108 snímků z I-KNOW studie a 28 snímků z ISLES 2015 [2]. Celkem bylo k dispozici 159

snímků pořízených z mnoha zdrav. středisek, pořízených různými skenery s indukcí 1,5 T i 3 T a mající různé rozlišení. Autoři 8 snímků z trénování a evaluace však vyřadili z důvodu přítomnosti artefaktů.

Preprocessing dat se skládal z odstranění lebky ze snímků (u části datasetu, kvůli konzistenci), koregistrací na společný atlas GIN-IMN a normalizací intenzit přes jednotlivé snímky. Pro každý ze tří zdvojových datasetů byly vygenerovány pravděpodobnostní mapy tkání.

Segmentace je rozdělena do dvou fází. V první z nich je použita diskrétní 2D Gaborova transformace, která se aplikuje konvolucí Gaborovských okének, vztah pro výpočet okének je navržen v článku. Autoři záměrně volí 2D transformaci, jelikož sekvence FLAIR má obvykle velký krok řezu, tedy rozlišení v této ose je poměrně špatné. Výstupem transformace je informace o energii v obraze pro daný směr a danou frekvenci. Z výstupu transformace je naučený Bayesovský klasifikátor, jehož výstupem je hrubá pravděpodobnostní mapa výskytu léze. Parametry okének jsou během učení optimalizovány maximalizací vzdáleností Gaborovských koeficientů voxelů s lézí a bez léze. Druhou fází je korekce prvotní segmentace pomocí Markovovských náhodných polí se zohledněním voxelu a jeho okolí (8 voxelové okolí v rovině 2D řezu a korespondující okolí v sousedních řezů).

Evaluace proběhla formou křížové validace s vynecháním jednoho pacienta pro každý ze tří vstupních datasetů. Výsledky navržené metody autoři srovnávají s výsledky týmu UK-Imp2 (později DeepMedic [50–51], konvoluční síť) a s CN-Neu (fuzzy c-means) z ISLES 2015 [2]. Zhodnocují, že metoda, přestože predikuje pouze na základě FLAIR, dosahuje podobných výsledků jako výsledky multimodálních metod v ISLES 2015 [2]. Výsledky v tabulce 5.5 odpovídají stejné validační proceduře. Obě porovnávané metody z ISLES 2015 [2] dosahují signifikantně horších Dice koeficientů než navrhovaná metoda. Výsledky ukazují, že metoda spíše podsegmentovává velké léze a malé léze byly často špatně segmentovány. Ukázalo se, že metoda je poměrně citlivá na artefakty vzniklé approximací po koregistraci nebo po převzorkování na jiné rozlišení.

Metoda	ASD [mm]	DC	HD [mm]	P	S
MRF	7,56 ± 9,29	0,582 ± 0,250	35,38 ± 25,97	0,658	0,509
UK-Imp2	-	0,541 ± 0,272	-	-	-
CN-Neu	-	0,524 ± 0,287	-	-	-

Tabulka 5.5. Výsledky segmentace subakutní CMP pomocí MRF (průměr ± směr. odch.) v porovnání s nejlepšími metodami ISLES 2015. ASD: Průměrná vzdálenost ploch. DC: Dice koeficient. HD: Hausdorffova vzdálenost. P: Precisnost. S: Sensitivita. [43]

Přestože jsou autoři velmi optimističtí z dosažených výsledků, sami zmiňují, že metoda zatím není vhodná pro klinické použití.

Kapitola 6

Neuronové sítě

V současné literatuře řešící segmentaci CMP převažují metody založené na neuronových sítích, proto je v této kapitole teoreticky uvedu. Metody použité v praktické části této závěrečné práce jsou také založeny na neuronových sítích.

Umělá inteligence je v současnosti populární termín, pod který se řadí i technologie neuronových sítí. Jedná se o poměrně univerzální algoritmy, které dokážou řešit celou řadu složitých úloh.

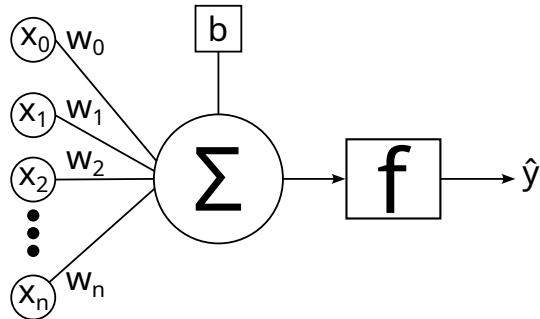
V klasických metodách programátor definuje extrakci příznaků ze vstupních dat a následně nad příznaky definuje rozhodovací pravidla pro vytváření závěrů. Pro naprogramování klasické metody typicky není potřeba mít vstupní data, jelikož rozhodovací mechanismy jsou determinované programátorem v kódu. U neuronových sítí má programátor typicky k dispozici vstupní data a jeho úkolem je navrhnout proces zpracování těchto dat neuronovou sítí. Jakmile tohle programátor vyřeší, tak se ve fázi trénování neuronové sítě načítají vstupní anotovaná data a neuronová síť se iterativně naučí příznaky a rozhodovací parametry pro vytváření predikcí. Tyto naučené informace jsou uloženy jako tzv. váhy neuronové sítě. Z toho vyplývá, že u klasických metod musí být pro programátora viditelné klíčové příznaky, které jsou důležité pro jeho rozhodnutí. U neuronových sítí tomu tak není, jelikož neuronová síť může vyhledat i lidem skryté nebo jinak složité příznaky, které jsou však pro rozhodování klíčové.

Dobrým příkladem může být klasifikace objektů v obrazu. Pokud bychom neměli k dispozici metody strojového učení a chtěli bychom např. klasifikovat jablka na obrázku, nejjednodušší možností by bylo sledovat poměr červených a zelených pixelů vůči ostatním barvám (protože barva jablek bývá červená nebo zelená). Toto řešení však není robustní, jelikož bere v potaz pouze barvu objektu na obrázku, takže např. zelená rostlina by byla také klasifikována jako jablko. Mohli bychom tedy přidat další pravidlo, které by na základě detekce hran hledalo kruhové objekty ve vstupním obrázku, což by zvýšilo robustnost. Nicméně rajče je červené a kulaté, takže to by bylo opět klasifikováno jako jablko. Takto bychom mohli postupně vymýšlet a přidávat další pravidla, která však začnou být velmi složitá. Zvláště, pokud bychom řešili klasifikaci geometricky komplexnějších objektů. Neuronové sítě se dokážou naučit nespočet takových pravidel a výsledkem často bývá kvalitnější výstup. Podmínkou však zůstává, že ve vstupních datech musí existovat příznaky, které jsou zásadní pro vytvoření rozhodovacích pravidel.

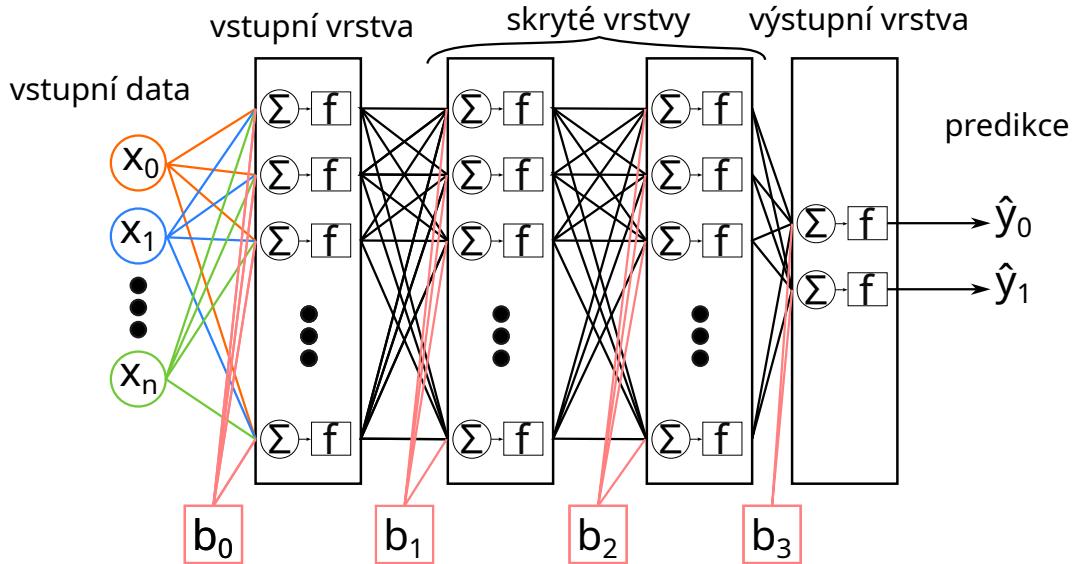
Základ neuronových sítí tvoří tzv. perceptron na obrázku 6.1. Matematicky je definován jako

$$\hat{y} = f\left(\sum_{i=0}^n (x_i w_i) + b\right) \quad (1)$$

kde n je počet vstupů, x je vstup, w je váha, b je bias, f je aktivační funkce a \hat{y} je predikce. Jednu položku vstupu si můžeme představit např. jako intenzitu konkrétního voxelu MRI snímku, správně naučené váhy přeškálují důležitost jednotlivých voxelů. Bias bývá kromě vah další učící se parametr a slouží pro změnu aktivace celého perceptronu nezávisle na vstupních datech. Výstup perceptronu (aktivace) pak může být např. pravděpodobnost, že se na snímku nachází léze.

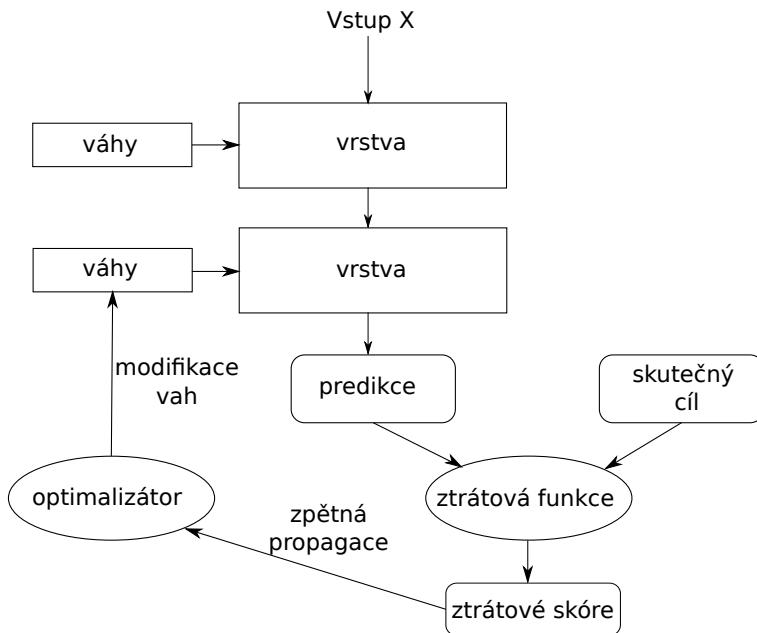
**Obrázek 6.1.** Model perceptronu.

Takto definovaný perceptron má omezené množství vah v závislosti na vstupu. Pro komplexnější problémy nastane situace, kdy váhy jednoduchého modelu neumožní dosáhnout požadované přesnosti. Kromě toho bude perceptron špatně generalizovat. Proto se definují vrstvy těchto perceptronů a tyto vrstvy se vzájemně zapojují za sebou (dopředná architektura). Model takové architektury je na obrázku 6.2. Každá vrstva obsahuje programátorem definovaný počet perceptronů - toto číslo též obsahuje velikost výstupu vrstvy. Vstupní data jsou lineárně zkombinována s váhami jednotlivých perceptronů a přeskálovány aktivačními funkcemi. Výstup takového vrstvy vstupuje do dalších vrstev označovaných jako skryté vrstvy. Poslední vrstva je výstupní a počet perceptronů definuje rozměry predikce. Metody obsahující skryté vrstvy se označují jako metody hlubokého učení.

**Obrázek 6.2.** Dopředná neuronová síť složená ze n vstupů, 2 skrytých vrstev a dvou výstupů.

Čím hlubší neuronovou síť máme a čím větší jsou jednotlivé vrstvy, tím složitější vztahy je neuronová síť schopna se naučit. Nutnou podmínkou je však použití aktivační funkce. Jedná se o nelineární derivovatelné funkce a kdybychom je nepoužívali, neuronová síť by byla schopna řešit pouze lineárně separabilní problémy a nebyla by příliš užitečná. Aktivační funkce proto do modelu vnáší nonlinearity.

Celý algoritmus neuronové sítě je uveden na obrázku 6.3. Zatím bylo popsáno, jak jsou data propagována sítí od vstupu k výstupu (predikci). Aby bylo možné síť natrénovat (nalézt optimální váhy a bias), je nutné dále definovat ztrátovou funkci. Tato funkce



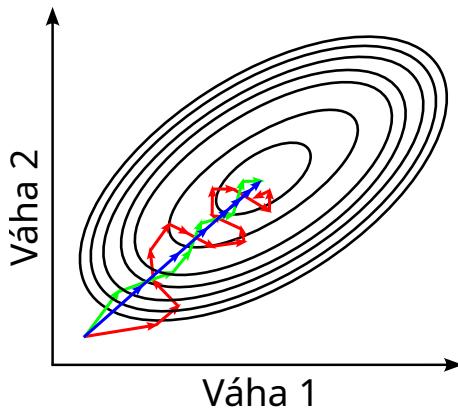
Obrázek 6.3. Kompletní schéma propagace informace dopřednou neuronovou sítí.

musí být derivovatelná a je závislá na predikci a expertní anotaci. Výstupem funkce je ztrátové skóre, což je skalární veličina udávající vzdálenost predikce od anotace.

Vzhledem k derivovatelnosti funkcí použitých v neuronové síti lze spočítat derivaci ztrátového skóre v závislosti na vahách a tím minimalizovat ztrátu. V praxi se definiují jednotlivé vrstvy a jejich derivace. Pro výpočet derivace se pak používá řetízkové pravidlo. Kvůli použitým nelinearitám a mnoha vrstvám nelze najít analytické řešení. Buď vůbec neexistuje nebo je čas pro jeho výpočet příliš velký. Z toho důvodu se používají numerické metody pro hledání optimálních vah. Jedná se o gradientní sestup, jehož principem je malá změna vah proti směru gradientu. Nejjednodušším případem je sestup v opačném směru gradientu škálovaném velikostí kroku (tzv. learning rate). Po aktualizaci vah jsou načtena nová vstupní data a proces se opakuje.

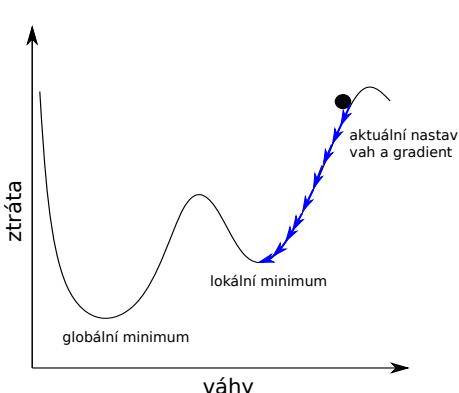
Předpokládáme, že trénovací dataset obsahuje dostatečné množství rozmanitých vzorků dat z populace, aby i během reálného nasazení neuronová síť dobře fungovala. Ideálně by se proto ztrátová funkce měla počítat z průchodu všech trénovacích dat a teprve poté by měla být provedena aktualizace vah. Datasetsy však bývají příliš velké na to, aby se všechny vešly do paměti grafické karty. Alternativou ke gradientnímu sestupu je stochastický gradientní sestup (SGD). SGD není totiké náročný na paměť ani na čas výpočtu, jelikož se vahy aktualizují po průchodu jediného vzorku místo průchodu všech vzorků trénovacího datasetu. Nevýhodou však je, že aktualizace vah nesměřuje přímo do minima ztrátové funkce. Třetí a nejčastěji používanou variantou je tzv. mini-batch kombinující dva předešlé přístupy. Během mini-batch se používá definovaná velikost tzv. batche - vzorků datasetu, které naráz projdou sítí a ze kterých se vypočítá ztrátové skóre. Aktualizace vah pak odpovídá sestupu do společného minima ztráty takového batche. Metody optimalizace jsou vizualizovány na obrázku 6.4.

Neuronová síť má mnoho lokálních minim. Váhy jsou typicky inicializovány náhodně s průměrem kolem nuly. Gradientním sestupem se tedy vahy přibližují nejbližšímu lokálnímu minimu ztrátové funkce nebo nějakému jinému, ale blízkému lokálnímu minimu. Téměř s jistotou nenajdeme globální optimum. Každé lokální minimum má však jiné hodnoty ztrátové funkce, proto každé nové učení stejné architektury dosáhne rozdíl-

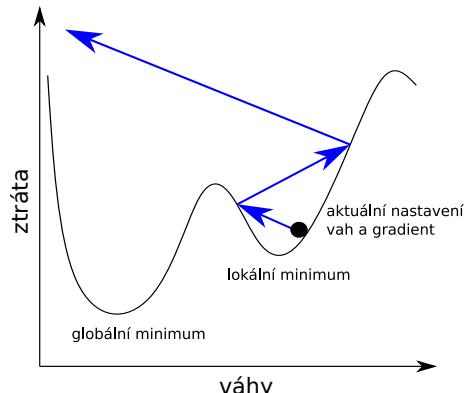


Obrázek 6.4. Vrstevnicový graf ztrátové funkce celého datasetu v závislosti na nastavení dvou vah. Modré šipky znázorňují gradientní sestup (je patrná viditelnost kroku v závislosti na gradientu). Červené šipky odpovídají SGD a zelené šipky odpovídají sestupu pomocí mini-batch.

ných výsledků, které závisí na náhodě. Pokud je zvolen příliš malý krok učení, bude konvergence do lokálního minima trvat příliš dlouho. Naopak pokud bude krok příliš velký, váhy mohou začít oscilovat kolem optima nebo se od něj začít vzdalovat. Tyto jevy jsou patrné z obrázků 6.5 a 6.6.



Obrázek 6.5. Malý krok učení a dlouhá doba konvergence.



Obrázek 6.6. Velký krok učení a divergencia.

Intuitivně můžeme zlepšit optimalizaci určením pseudo-hmotnosti vahám a zavedením hybnosti - tzv. Momentum, které se přičítá ke klasickému gradientu. Hybnost pomůže vahám překlenout mělká lokální minima nebo plateau ztrátové funkce. Názorně je to patrné na obrázku 6.5, kdy pomocí Momentum by váhy překlenuly lokální minimum, protože by využily hybnost nabytou při sestupu do onoho lokálního minima a celý algoritmus by dosáhl ještě nižší ztráty. Existuje celá řada sofistikovaných optimalizátorů, které používají gradient ztrátové funkce a Momentum a automaticky škálují krok učení a jiné parametry v závislosti na předchozím průběhu učení a aktuálním gradientu. Při použití adaptivních optimalizátorů jako jsou AdaGrad, RMSProp nebo Adam tak dojde často k rychlé konvergenci, která nepříliš závisí na úvodním nastavení parametrů optimalizátoru.

6.1 Vrstvy neuronových sítí

Pokud bychom používali schéma uvedené na obrázku 6.2, nároky na paměť při výpočtu by byly enormní. Z důvodu úspory paměti, zlepšení generalizace nebo pro zohlednění prostorového kontextu se používají různé druhy vrstev.

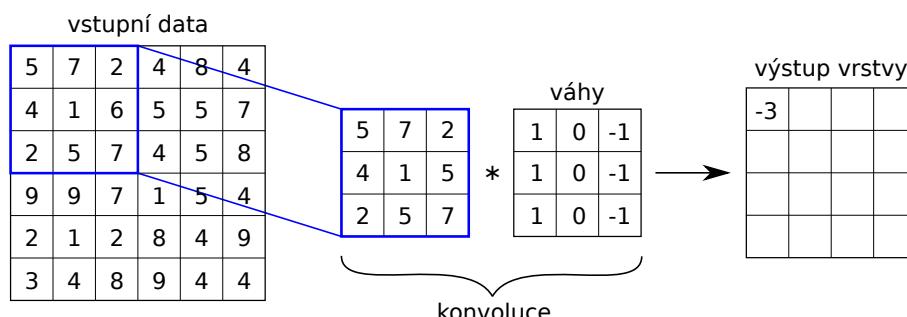
6.1.1 Lineární vrstvy

Lineární vrstvy (označováno jako FC - Fully Connected) obsahují programátorem definovaný počet perceptronů, které fungují přesně tak, jak byly doposud v této práci popisované neuronové sítě a odpovídají obrázku 6.2. Každý výstup z předchozí vrstvy vstupuje do každého perceptronu v této vrstvě. Výsledkem je struktura sítě schopná zpracovávat obecná serializovaná data, nicméně je extrémně náročná na paměť i na výpočet vzhledem k počtu vah. Běžně se FC vrstvy používají jako poslední vrstvy neuronových sítí počítačového vidění.

6.1.2 Konvoluční vrstvy

Konvoluční vrstvy tvoří hlavní prvek neuronových sítí zpracovávající obraz. Konvoluční vrstvy utilisují konvoluční kernely, které jsou trénovatelné. Tyto kernely jsou tensory o stejné dimensi jako vstupní data - tedy pro zpracování 2D obrazu jsou 2D. Vrstva je definována počtem výstupních kanálů (filtrů), velikostí kernelů, velikostí kroku kernelu a velikostí paddingu. Kanály si lze představit jako R, G a B složku u obrázků nebo jako jednotlivé sekvence MRI. Jedná se o třetí nebo čtvrtý rozměr vstupního tensoru. Jestliže používáme multimodální síť, často kanály odpovídají jednotlivým modalitám.

Každý výstupní kanál konvoluční vrstvy tedy obsahuje počet kernelů rovnající se počtu kanálů vstupujících do vrstvy a jeden bias. Každý kernel provede konvoluci s odpovídajícím vstupním kanálem a součet aktivace všech kernelů a biasu daného výstupního kanálu definuje příznak na dané pozici. Příklad 2D konvoluce s jedním filtrem, vstupním rozměrem (1, 6, 6) a rozměrem kernelu (3, 3) je na obrázku 6.7.



Obrázek 6.7. Příklad aplikace konvolučního kernelu na jeden kanál 2D obrázku. Váhy jsou nastaveny (naučeny) na detekci vertikálních hran. Výstupem tedy bude vysoká intenzita v místech vertikálních hran přecházejících do temna.

Pokud chceme zmenšit rozměry (nikoliv kanály, ty typicky zvětšujeme) obrazu dané vrstvy, definujeme krok konvoluce. Konvoluční jádra díky tomu přeskocí definovaný počet prvků. Chceme-li konvolovat i okraje obrazu, je nutné použít tzv. padding, který uměle zvětší obraz tak, aby střed konvolučního jádra odpovídal okraji obrazu.

Výhoda použití konvoluční sítě je, že kernely se naučí prostorový kontext. Konvoluční jádra jsou schopna snadno detektovat hrany a zřetězení více konvolučních vrstev extrahuje ze vstupních dat stále složitější vzory důležité pro rozhodování sítě. Zároveň množství trénovatelných parametrů není příliš vysoké, takže lze vrstvy efektivně

používat. Konvoluční vrstvy se používají jako první vrstvy u sítí zpracovávající obraz, jejich úkolem je postupně zvětšovat počet kanálů a hledat podstatné prostorové jevy v obraze. Tyto nalezené příznaky nakonec vstupují do lineárních vrstev.

6.1.3 Transponovaná konvoluční vrstva

Transponovaná konvoluční vrstva je opakem konvoluční vrstvy - tedy z příznaků s větším počtem kanálů vytvoříme menší počet kanálů, ale často společně s tím zvětšujeme prostorové rozlišení těchto příznaků. Vrstvy mají opět trénovatelné kernely, optimalizaci se kernely učí efektivně odhadovat původní obraz ze vstupu s nižším rozlišením.

6.1.4 Pooling vrstvy

Pooling vrstvy se nejčastěji používají ke snižování rozměru příznaků a k efektivnímu zvětšování receptivního pole. Vrstva neobsahuje trénovatelné parametry. Jejím principem je, že pomocí jádra vyselektuje postupně části obrazu pro jednotlivé vstupní kanály. Výstupem vrstvy je nějaká matematická funkce nad těmito příznaky. Běžně se používá MaxPooling (výběr maxima) nebo Average Pooling (průměr z prvků jádra). Např. zvolíme-li MaxPoolingové jádro o velikosti (2,2) s krokem 2, ze 4 vstupních příznaků překrytých jádrem vytvoříme 1 výstupní příznak pro každý kanál odpovídající maximální položce překrytých vstupů.

6.1.5 Ostatní vrstvy

Pro regularizaci se používá jednoduchá netrénovatelná vrstva Dropout, která náhodně přepíše aktivaci některých příznaků na nulu. Vrstva ovlivňuje síť pouze během trénovací fáze. Díky ní se přeruší závislosti perceptronů napříč různými vrstvami, což vede k přerušení memorování trénovacích dat a zlepšení generalizace.

Další používanou vrstvou je BatchNorm, který normalizuje množinu příznaků napříč všemi kanály v určité fázi průchodu neuronovou sítí. Vrstva obsahuje dva trénovatelné parametry pro každý kanál: škálovací konstantu a posuv, sloužící pro nalezení optimální transformace dat po normalizaci.

6.2 Výzvy neuronových sítí

Při aplikaci neuronových sítí na konkrétní úlohu může dojít k mnoha problémům, se kterými se musíme v reálném světě potýkat a které budou popsány v následujících sekcích.

6.2.1 Mizení gradientu

Ještě před začátkem trénování neuronové sítě je potřeba zvolit aktivační funkce na jednotlivých vrstvách. Protože učení sítí funguje díky derivacím, aktivační funkce společně s hloubkou modelu zásadně ovlivňují gradient vah. V případě příliš hlubokých modelů dochází kvůli řetízkovému pravidlu k mizení gradientu. Dále může dojít k saturaci aktivační funkce a derivace v bodě je téměř nulová nebo nulová. V případě nulové derivace hovoříme o tzv. umírání neuronů.

Příkladem aktivační funkce, kde hrozí mizení gradientu je sigmoida, která se běžně používá jako aktivační funkce poslední vrstvy. Sigmoida je definována vztahem (2).

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (2)$$

Graf sigmoidy je na obrázku 6.8 a jako aktivační funkce má dva zásadní problémy. Její derivace je všude menší než 1, konkrétně maximum derivace je 0,25, což velmi přispívá mizení gradientu. Dále pokud počítáme derivaci v bodech menších než -5 nebo větších než 5, je z grafu patrné, že vyjde téměř nulová. Pokud na začátku učení nastavíme velikost kroku příliš velkou, gradientním sestupem se dostaneme do poměrně odlehlych hodnot od počátku a váhy se poté nebudou téměř vůbec měnit.

Často se používá aktivační funkce ReLU, která je definována vztahem (3).

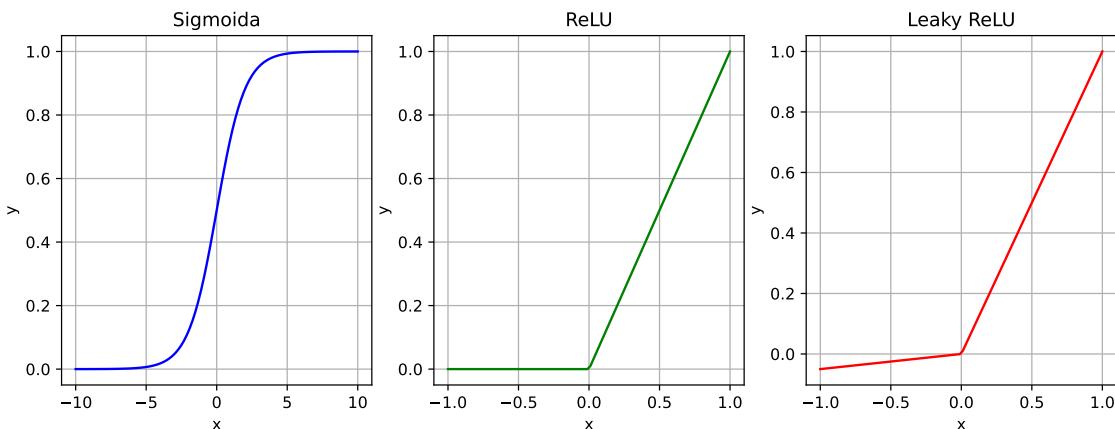
$$ReLU(x) = \max(0, x) \quad (3)$$

Graf funkce je na obrázku 6.8. Derivace aktivační funkce na intervalu (x, ∞) je 1, díky čemuž je funkce mnohem vhodnější než sigmoida. Problém je v její nekladné části, kde hodnoty funkce jsou nulové a derivace je také nulová. Pokud se váhy ocitnou v nastavení, kdy vstup do ReLU je nekladný, dojde k umírání neuronů. Výstupy těchto neuronů budou pro jakýkoliv vstup vždy nulové a jejich váhy se již nikdy neaktualizují.

Asi nejběžnější aktivační funkcí je Leaky ReLU, která opravuje původní ReLU. Je definována vztahem

$$ReLU(x) = \max(ax, x) \quad (4)$$

kde a je parametr negativního sklonu. Běžně se používá 0,01. Graf funkce je opět na obrázku 6.8. Pokud je vstup do Leaky ReLU nekladný, její výstup není nulový a je zde malý gradient, který dovolí opět aktualizovat váhy a tím pádem iterativně optimalizovat konkrétní perceptron.



Obrázek 6.8. Aktivační funkce sigmoida, ReLU a Leaky ReLU (sklon 0,05 v negativní části).

Dalším řešením mizení gradientu je použití residuální sítě. Taková síť obsahuje residuální spojení. Informace se stále propaguje postupně po vrstvách, ale vstupy do některých vrstev dopředně propagace se sečtou s výstupy některé z předešlých vrstev (toto propojení „přeskočí mezilehlé vrstvy“). Místo sumace lze v některých případech použít spojení tensorů dat přes dimenzi kanálů.

Pro boj s mizením gradientu lze také použít BatchNorm, což je vrstva, která normalizuje příznaky napříč batchem. Tím pádem nedojde k extrémně velkým nebo extrémně malým aktivacím během průchodu sítí.

■ 6.2.2 Přeúčení

Dataset, který máme k dispozici typicky rozdělíme na trénovací a validační data. Tyto množiny musí být striktně odděleny, abychom mohli nezávisle ověřit výkonnost modelu na nových datech. Dataset se rozděluje typicky náhodně v poměru 80:20.

Pokud navrhнемe velkou nebo hlubokou architekturu neuronové sítě, zpočátku trénování bude růst přesnost predikce jak na datech trénovací množiny, tak validační. Pokud sít budeme trénovat dostatečně dlouho - tj. uděláme dostatečný počet průchodů trénovacích dat a aktualizací vah, dojde k tomu, že dosáhneme bodu, kdy se stále bude zvyšovat přesnost na trénovacích datech, ale již začne klesat přesnost na validačních datech. Tomuto jevu se říká přeúčení.

Aby nedošlo k ovlivnění výkonnosti modelu, je dobré včasné zastavit učení v momentu, kdy začne přesnost na validační množině klesat (technika brzkého zastavení).

Dále je možné, že učení sítě bude korektně probíhat, ale rozdíl mezi trénovací a validační přesnosti bude příliš velký. V takovém případě opět dochází k přeúčení na trénovací data a je vhodné zmenšit hloubku modelu nebo velikost jednotlivých vrstev. Případně je možné použít augmentaci pro rozšíření trénovacího datasetu nebo použít Dropout vrstvy.

■ 6.2.3 Nevyvážený dataset

Nevyvážený dataset je obvyklý problém, se kterým je nutné se potýkat. Často se jedná o nevyváženosť tříd. Ve smyslu této práce může jít o nevyváženosť objemu zdravé tkáně vůči objemu léze u snímku vstupujícího do neuronové sítě. Pokud používáme pro segmentaci jako ztrátovou funkci křížovou entropii, lze zavést vyvažovací faktor, který škáluje vliv členů jednotlivých kategorií v křížové entropii. Zjednodušeně může vyvažovací faktor značit poměr objemu zdravé tkáně a objemu léze u vstupního snímku nebo batche.

Při segmentaci mrtvice lze považovat za nevyváženosť datasetu i oblast s převládajícím výskytem lézí nebo převažující objem nebo tvar léze.

Kromě toho může být dataset nevyvážený z hlediska intenzit. Jestliže by většina snímků měla průměrně vyšší intenzitu voxelů než menšina, došlo by k přeúčení modelu na hyperintenzitní skupinu. Z toho důvodu je vhodné použít nějakou z forem normalizace (např. normalizace snímku, normalizace batche, odečtení průměru intenzit trénovacího datasetu od vstupního snímku a vydelení směr. odchylkou intenzit trénovacího datasetu).

■ 6.2.4 Prostorové zarovnání

Konvoluční vrstvy vyžadují, aby prostorové rozměry voxelu byly konzistentní napříč všemi daty. Proto se v rámci preprocessingu data musí převzorkovat na stejné rozměry (typicky isotropní). Pokud by tomu tak nebylo, konvoluční kernely se nedokážou naučit parametry lézí a nebudou schopny kvalitně predikovat.

Pokud jsou multimodální data v nativním prostoru, pak MRI souřadnice jedné modality reálně neodpovídají souřadnicím jiné modality - obrasy jsou vůči sobě více či méně posunuté nebo otočené.

Proces při kterém dojde k zarovnání sekvencí vůči sobě se nazývá koregistrace. Tu řeší algoritmy pomocí optimalizace ztrátové funkce založené na vzájemné informaci nebo na míře podobnosti. Důležité je, jaký stupeň volnosti během koregistrace používáme. Výstupem koregistrace je buď matice afinního zobrazení nebo deformační pole definující posuvy jednotlivých voxelů původního obrazu. Tyto transformace je možné řetězit.

Základní transformací je rigidní, která zachovává Euklidovskou vzdálenost mezi každou dvojicí bodů. Jde o affinní transformaci skládající se z translace nebo rotace. Další transformací je obecná affinní transformace, která umožňuje navíc změnu měřítka a zkosení. Takové transformace lze stále realizovat maticemi affinního zobrazení tj. zachovávající přímky. Nejsložitější transformace je elastická, kterou lze realizovat deformačním polem, kde se přímky zobrazují na křivky.

■ 6.3 Augmentace dat

Často není k dispozici dostatečné množství anotovaných dat - ať už z důvodu finanční náročnosti nebo časové náročnosti. Nedostatek dat vede k přeúčení neuronové sítě na trénovacích datech nebo obecně ke špatným výsledkům na datech validačních. Z toho důvodu se někdy používá augmentace - umělé navýšení dat. Abychom realizovali augmentaci, běžně aplikujeme nějakou transformaci původních dat. To se děje buď online během učení nebo lze augmentace vygenerovat před začátkem učení v závislosti, zda chceme šetřit úložný prostor nebo výpočetní čas. Pokud augmentace generujeme předem, tak je potřeba zajistit, aby se augmentovaná data nacházela pouze v trénovací množině. Jinak by došlo k promíchání trénovací a validační množiny a výsledky dosažené na validační množině by byly optimističtější než realita.

Augmentaci MRI dat lze rozdělit na prostorovou augmentaci a augmentaci vycházející pouze z intenzit voxelů. Mezi prostorovou augmentaci můžeme započítat:

- Zrcadlové převrácení.
- Převrácení dat podél anatomické osy (snímek musí být s osami koregistrován).
- Náhodná affinní transformace (náhodná rotace, posun, změna měřítka nebo zkosení).
- Převzorkování obrazu na nižší rozlišení a zpět na původní.

Do kategorie augmentace pouze na základě intenzit spadá:

- Simulace pohybových artefaktů.
- Simulace artefaktu duchů vznikající periodickým pohybem.
- Simulace clonného pole.
- Gaussovské rozostření.
- Gaussovský šum.
- Změna kontrastu snímku.

■ 6.4 Fúze dat

Při použití multimodálních modelů je nutné zvážit, jak informace z jednotlivých modalit bude sjednocena v rozhodovacím procesu. Obecně multimodální modely dosahují o 1.2 % – 27.7 % lepší přesnosti proti unimodálním modelům [61]. Rozlišujeme 3 fúze: vstupní (brzká), in-layer a výstupní (pozdní) fúze.

■ 6.4.1 Vstupní fúze

Vstupní fúze dat je v drtivé většině nejpoužívanější fúzí [62]. Informace ze vstupních kanálů resp. modalit se fúzuje před vstupem do neuronové sítě. Jinými slovy neuronová síť hledá příznaky z fúze modalit. Tuto fúzi můžeme rozdělit na fúzi 1. typu, kde fúzujeme originální data a na fúzi 2. typu, kdy fúzujeme již extrahované příznaky každé z modalit. Fúze 1. typu se nejčastěji realizuje spojením dvou tensorů podél dimenze kanálů nebo poolingem. Fúze 2. typu zahrnuje značně komplikovanější algoritmy, které budou mohou být opět natrénované (např. hrubě segmentují jednotlivé modality) [61] nebo na základě expertně definovaných pravidel extrahuje užitečné příznaky.

■ 6.4.2 In-layer fúze

Do in-layer fúze vstupují modality jednotlivě do jediné neuronové sítě. Informace je postupně fúzována během dopředné propagace a učení prostorových příznaků pomocí konvoluční části sítě. Proto architektury realizující in-layer fúzi mají nejčastěji dvě vstupní cesty a mezi nimi probíhá dynamická výměna prostorových map příznaků [62]. Dle článku [61] lze opět fúzi rozdělit: 1. typ extrahuje pomocí neuronové sítě příznaky všech modalit, zatím co fúze 2. typu extrahuje příznaky pouze z některých modalit.

■ 6.4.3 Výstupní fúze

Pro realizaci výstupní fúze je třeba natrénovat více samostatných unimodální neuronových sítí na jednotlivých modalitách. Výstupem téhoto sítí je typicky pravděpodobnostní mapa. Výstupní fúze tedy řeší, jak spojit více pravděpodobnostních map, které tvoří výslednou segmentaci. Výhoda této metody je, že optimalizace sítí probíhá se zaměřením na unikátní vlastnosti daných modalit. Fúze se často realizuje průměrováním pravděpodobnostní mapy, poolingem nebo většinovým hlasováním. Případně lze použít jednoduchý klasifikátor pro naučení se fúzování posteriorní pravděpodobnosti. [62, 61]

Kapitola 7

Metodika

Vzhledem k provedené rešerši je zřejmé, že metody založené na neuronových sítích s postupem času převládají nad klasickými metodami. Výsledky nových článků používající tradiční metody, které mohou být značně složité, pouze dorovnávají hluboké učení. Nejenom, že metody postavené na hlubokém učení dosahují nejlepších výsledků, ale staly se globálně populárními, pro rozmanitost problémů, jakou dokážou řešit. Postupně s rozvojem grafických karet a obecně výkonností hardwaru je možné realizovat komplexnější neuronové sítě, které mohou řešit složitější problémy s excellentními výsledky.

Z popsaného důvodu v této práci jsou ověřeny dvě metody založené na neuronových sítích: DeepMedic [50–51], který vyhrál soutěž ISLES 2015 [2] a BRATS 2017; a nnU-Net [53], ze kterého vycházelo mnoho metod (vč. první umístěné) na soutěži ISLES 2022 [40]. Dále je v této práci implementována vlastní metoda na architektuře 3D U-Net [63]. Vstupními modalitami do těchto metod jsou vždy dvě sekvence FLAIR a DWI. Výstupem je pak jedna segmentace určující polohu léze. Při experimentování s vlivem fúzí v této práci byly vybrány nejčastější typy fúzí v literatuře. Vstupní fúze je realizována pomocí spojení tensorů podél osy kanálů. In-layer fúze je realizována pomocí spojení příznaků v enkodéru sítě. Výstupní fúzi jsem provedl průměrováním pravděpodobnostních map. Jelikož vstupní modality jsou dvě, nelze použít většinové hlasování.

Neuronové sítě se nejčastěji implementují v jazyce Python, pro který jsou napsané i dvě populární deep-learningové knihovny TensorFlow a PyTorch. Pro experimenty je nutné používat obě tyto knihovny. Z toho důvodu byl celý proces implementován pouze v jazyce Python (tj. analýzu datasetů, preprocessing, trénování metody i evaluaci). Byla použita knihovna pynrrd pro práci s NRRD soubory; knihovna ANTsPy pro načítání NIfTI souborů, transformace a koregistrace; knihovna NiBabel pro ensembling; knihovny connected-components-3d a Nilearn pro analýzu datasetů; knihovna TorchIO pro vzorkování dat po preprocessingu. Vývoj probíhal na sdíleném grafickém serveru s osmi grafickými kartami NVIDIA GeForce RTX 2080 Ti (11 GB RAM), 377 GB RAM, procesorem Intel Xeon Silver 4214R a se systémem Debian. Při práci s GPU byla používána jediná GPU.

Knihovny ANTsPy a NiBabel (asi nejpopulárnější knihovny) obsahují zhruba stejné funkce, kromě koregistrace, kterou nepodporuje NiBabel. Načtení dat a operace nad nimi jsou poměrně časově náročné, proto jsem nejprve porovnal rychlosti operací, které jsou v ANTsPy i v NiBabel. Autoři ANTsPy jej otestovali proti NiBabel. Doba běhu NiBabel při testu rychlosti načítání souborů byla 56,121 s, ale doba ANTsPy byla pouze 33,601 s. Tento test proběhl na počítači Macbook Pro. Ten stejný test na poskytnutém grafickém serveru ukázal výsledky: ANTsPy 22,462 s a NiBabel 19,778 s. Dále byl proveden vlastní test, který se skládal z načtení FLAIR, DWI a masky mozku, převzorkování DWI na FLAIR a z odstranění lebky vymaskováním. Vykonání kódu v NiBabel trvalo 3,380 s, zatím co v ANTsPy 5,654 s. U ANTsPy byla použita lineární interpolace. NiBabel používá pouze b-spline interpolaci, proto byla nastavena b-spline 1. řádu, která by měla odpovídat lineární interpolaci u ANTsPy. Přestože ANTsPy je pomalejší,

je nutné ho použít z důvodu potřeby koregistrace. Proto se nabízí varianta provádět většinu operací v NiBabel a pro koregistraci převést data do ANTsPy a následně zpět do NiBabel. Převod dat mezi knihovnami je ale časově dražší než provádět všechny operace přímo v ANTsPy. Na základě závěrů tohoto testování byla využita knihovna ANTsPy.

Kroky nutné ke splnění cílů této práce jsou: analýza dostupných datasetů a návrh jejich rozdělení; ověření dvou zmiňovaných metod; implementace a ověření 3D U-Net [63]; ověření vlivu fúze dat; zvolení nejvhodnější metody pro segmentaci dat z Motola; a nakonec srovnání dosažených výsledků jak mezi sebou, tak s literaturou. Jelikož se tato závěrečná práce zabývá více datasety a více metodami, je vhodné si stanovit co možná nejjednotnější postup preprocessingu dat a následné evaluace, aby vstupní data i výsledky jednotlivých metod byly co nejkonzistentnější.

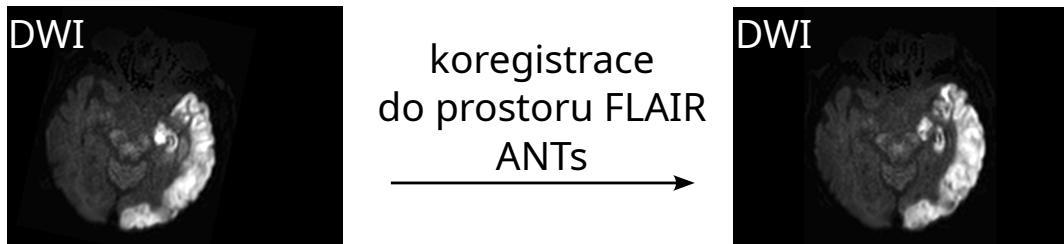
Z výše zmíněného důvodu je repozitář rozdělen do několika funkčních modulů. Nejdůležitějším modulem je složka datasets, která obsahuje soubory datasetů v původní podobě a soubor *dataset_loaders.py*, který definuje funkce pro načítání dat z jednotlivých datasetů a definuje třídu Subject, která obsahuje funkce důležité pro manipulaci s daty v této práci (maskování mozků, aplikace transformací pro koregistrování, převzorkování, normalizace, ...). Idea ověřování automatických metod v práci je, že nejprve budou vygenerovány pro každou segmentační metodu snímky pomocí třídy Subject, které budou v požadovaném formátu uloženy a připraveny pro natrénování segmentační metody.

Před zahájením preprocessingu třídou Subject je však nutné vygenerovat pomocná data, která se uloží ke každému skenu ve složce *datasets*. Zaprve je potřeba vygenerovat masky mozku. Pro segmentaci mozku existují již hotová poměrně robustní řešení. Já jsem k segmentaci použil software HD-BET [64], který segmentuje mozek ze sekvence FLAIR. Příklad vymaskování mozku je na obrázku 7.1. Složka *skull_stripping* obsahuje skripty pro přípravu dat pro HD-BET a import výstupu z BET zpět do repozitáře. Zadruhé je potřeba vygenerovat transformace koregistrování modality a expertní segmentace. Protože sekvence FLAIR mají vyšší rozlišení, byla vygenerována rigidní transformace z prostoru DWI do prostoru FLAIR. Příklad koregistrace z prostoru DWI do prostoru FLAIR je na obrázku 7.2. Dále byly uloženy soubory SyN elastické transformace z prostoru FLAIR do standardního prostoru MNI (cílová šablona byla získána z práce [65]). Příklad SyN koregistrace je na obrázku 7.3. Popsané transformace se vygenerují spuštěním skriptu *generate_transforms.py*. Vygenerované transformace jsou poté používány při volání funkcí třídy Subject.

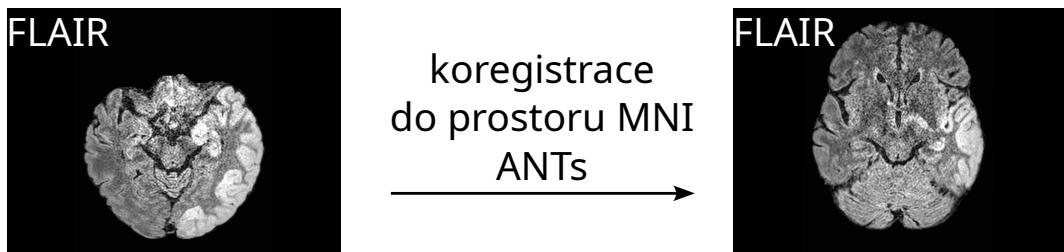


Obrázek 7.1. Odstranění lebky aplikací masky z HD-BET. Snímek datasetu Motol 20091_Anat_20210510.

V repozitáři závěrečné práce se pak nachází složky s kódem již experimentálně ověřovaných metod a složky s příponou *_workspace*. V těchto složkách se nachází skripty pro preprocessing dat. Ten se typicky skládá z načtení dat, transformací DWI sekvence do



Obrázek 7.2. Rigidní koregistrace z prostoru DWI do prostoru FLAIR. Snímek datasetu Motol 20091_Anat_20210510.



Obrázek 7.3. Koregistrace SyN z prostoru FLAIR do prostoru MNI. Snímek datasetu Motol 20091_Anat_20210510.

prostoru FLAIR (případně i expertní segmentace), odstranění lebky díky vymaskování segmentace mozku z HD-BET, převzorkování snímků na rozměry (200, 200, 200) a rozměry voxelu (1, 1, 1) mm. Případně je aplikována normalizace snímků nebo jiné funkce, které vyžaduje konkrétní testovaná metoda. Pro zajištění co nejlepší konzistence se provádí koregistrace u všech datasetů, přestože některé snímky mohly být distribuovány již koregistrované. Preprocessing končí uložením takto zpracovaných snímků do struktury, kterou vyžaduje konkrétní testovaná metoda. Dále se ve složce s příponou *_workspace* nachází konfigurační soubory příslušné metody.

Jakmile jsou připravena data, natrénují se modely příslušné metody, provede se predikce validačních dat na těchto modelech. V kořeni repozitáře se nachází skript *evaluate.py*, kterým je možné nakonec porovnat predikce vytvořené naučenými modely s původními daty. Protože preprocessingem originální data mírně ztratí kvalitu, neporovnávám predikce validačních dat s výsledky z preprocessingu. Skript *evaluate.py* proto vrátí predikované segmentace zpět do nativního prostoru a evaluace se provádí proti originálním expertním segmentacím. Došlo u nich pouze v případě potřeby ke koregistraci DWI části masek do FLAIR prostoru (jelikož i vstup do metod je koregistrován na příslušný FLAIR). Evaluační skript vytvoří csv soubor obsahující TP, FP, TN, FN, Dice koeficient a predikovaný objem. V článcích z rešerše se často považuje za selhání metody, pokud neexistuje žádný pravdivě pozitivní voxel (nulový Dice koeficient). Vzhledem k provedené rešerši evaluačních metrik jsem se však rozhodl, že počet selhání metody budu sledovat porovnáváním Dice koeficientu s hodnotou 0,1.

Jednotlivé csv soubory jsou uloženy ve složce *results*, kde se navíc nachází soubor *results.ipynb*, který jsem používal pro zpracování csv souborů a ke generování grafických výsledků.

Kapitola 8

Analýza datasetů

V této kapitole popíší tři datasety, se kterými jsem v této práci disponoval a byly použity pro trénování nebo validaci modelů. Skripty pro generování statistik se nachází ve složce *stats*.

V kapitole jsou uváděny tzv. komponenty - jako komponenta je označovaná ohraničená oblast jedné léze (jedno ložisko). Komponenta je determinována svým 26 voxelovým okolím - tedy aby byl nějaký voxel ohraničený, nesmí navazovat plochou, hranou ani vrcholem na další segmentovaný voxel.

Pro vytvoření analýz byl využit atlas MNI Structural Atlas se segmentacemi 9 regionů mozku, vytvořený Jackem Lancasterem [66–67]. Stejný atlas je používám i softwarem FSL. Atlas je koregistrován do prostoru MNI152, pro analýzy byla použita segmentace s rozlišením 1 mm a nulovým prahem (označení MNI-maxprob-thr0-1mm).

Pro grafickou vizualizaci lézí byly vytvořeny NIfTI soubory v prostoru MNI, jejichž intenzita voxelů představuje počet skenů, u nichž se vyskytuje léze v daném místě. Tyto soubory mohly být následně převedeny na obrázek pomocí funkce `plot_glass_brain` knihovny Nilearn [68].

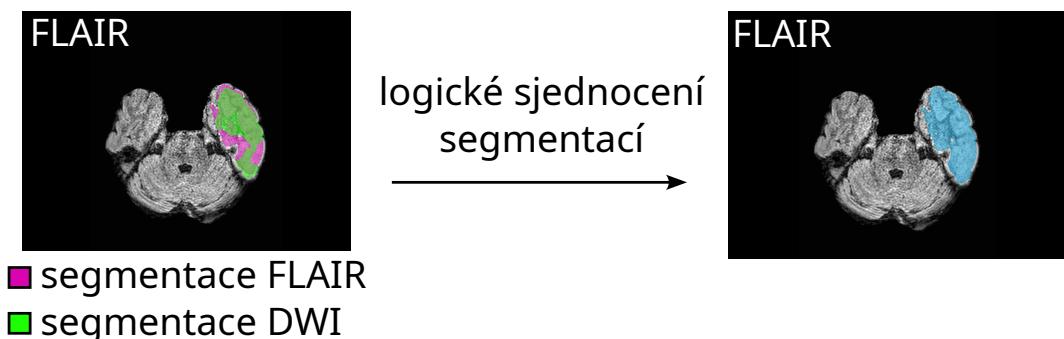
8.1 Dataset Motol

Výchozí MRI dataset s ischemickou CMP byl poskytnut 2. Lékařskou fakultou UK (dataset je dále označen jako Motol). Jedním z cílů diplomové práce bylo dosáhnout co nejlepší automatické segmentace pro tento dataset.

Dataset je složen ze snímků celkem 19 pacientů a každý z nich byl skenován ve třech časových bodech. Původní struktura složek, v nichž byly uložené NIfTI soubory byla: `{číslo pacienta}/NI/Processing/Anat_{datum skenu}/`. Data byla uložena na NAS v Motole, takže nejprve bylo nutné data stáhnout a pro účely diplomové práce byla cesta zkrácena na `{číslo pacienta}/Anat_{datum skenu}/` a data byla uložena do složky `datasets/Motol`. Případně konkrétní sken je někdy označen jako `{číslo pacienta}_Anat_{datum skenu}`.

Dataset obsahuje sekvence: DWI1, DWI2, FLAIR, SWI, kvantitativní T1, T1-w a T1 s kontrastní látkou. V práci dále používám pouze sekvence DWI2 (označováno jako DWI) a FLAIR. Všechny modality jsou koregistrovány do společného prostoru pomocí SPM. Snímky obsahují lebku. Dále jsou k dispozici expertní segmentace v NRRD souboru. NRRD soubor obsahuje jednu segmentaci léze na sekvenci FLAIR a druhou segmentaci na sekvenci DWI; pokud není specifikována modalita segmentace v kontextu tohoto datasetu, jedná se o logické sjednocení segmentací FLAIR a DWI (obrázek 8.1). Dále by v souboru měly být obsaženy segmentace tkání a maska mozku z SPM, nicméně ne u všech pacientů tomu tak je. Masku mozku z SPM navíc nelze použít pro odstranění lebky, jelikož neobsahuje mozkomíšní mok, ale pouze šedou a bílou hmotu.

Expertní segmentace jsou vytvořeny pomocí softwaru 3D Slicer a doinstalovaného modulu SegmentEditorExtraEffects [69]. Expert si nejprve pro koregistrované snímky vygeneruje segmentace tkání bílé hmoty, šedé hmoty a mozkomíšního moku pomocí



Obrázek 8.1. Na levém obrázku se nachází originální expertní segmentace v NRRD souboru, vpravo je segmentace po sjednocení. Snímek datasetu Motol 20091_Anat_20210510.

toolboxu Statistical Parametric Mapping (SPM) na modalitách, na kterých bude segmentovat lézi. Následně si zobrazí příslušné skeny v 3D Sliceru a kombinací Gaussovského rozostření, prahováním a logickými operacemi nad segmentacemi tkání a prahovanou maskou získá hrubou masku ischemické léze, kterou nakonec štětcem v 3D Sliceru zdokonalí.

Bohužel NRRD soubor 2290867_Anat_20230109 je poškozen a nelze načíst expertní segmentaci, proto byl dále vyřazen. Pro 18 pacientů tedy byly k dispozici snímky i segmentace v 1., 2. i 3. sběru dat. Pro 1 pacienta byl k dispozici pouze 2. a 3. sběr. Celkem jde tedy o 56 skenů.

Odstup druhého sběru od prvního je průměrně \pm směr. odch. $6,67 \pm 0,69$ dne, jednotlivé četnosti jsou v tabulce 8.1. Odstup třetího od druhého sběru je průměrně $6,95 \pm 0,71$ dne, četnosti jsou v tabulce 8.2. O datasetu lze tedy říci, že nejvíce odpovídá subakutní fázi ischemické mrtvice.

Charakteristika dat v nativním prostoru je ukázána v tabulce 8.3, nicméně platí pouze pro skeny - ne všechny NRRD segmentace respektují dimenze tensorů příslušných skenů. Deskriptivní statistika lézí je uvedena v tabulce 8.4, je v ní zahrnut i DC značící míru zkreslení po preprocessingu, během něhož došlo k mírné ztrátě kvality expertní segmentace.

Rozestup ve dnech	Počet pacientů
5	1
6	5
7	11
8	1

Tabulka 8.1. Počty pacientů s daným rozestupem 2. sběru dat od 1.

Rozestup ve dnech	Počet pacientů
5	1
6	1
7	16
9	1

Tabulka 8.2. Počty pacientů s daným rozestupem 3. sběru dat od 2.

Počet voxelů snímku	Rozměry voxelu [mm]	Počet snímků
(224, 512, 512)	(0,464, 0,464, 0,9)	15
(224, 512, 512)	(0,464, 0,464, 0,8)	15
(448, 448, 208)	(0,75, 0,422 0,422)	26

Tabulka 8.3. Nativní velikosti tensorů datasetu Motol. Počet voxelů označuje dimenze tensoru dat.

	Průměr	Směr. odch.	Min	Max
Objem léze	42,39	40,94	1,73	1,73
Objem léze na FLAIR	36,42	37,10	0,16	133,81
Objem léze na DWI	23,79	32,37	0,00	140,64
DC po preprocessingu	0,97	0,01	0,94	0,99

Tabulka 8.4. Deskriptivní statistika lézí datestu Motol přepočítaná na jeden sken. Uváděné objemy jsou v ml. DC byl vypočítáním porovnáním expertní segmentace před a po preprocessingu.

Komponenty	Průměr	Směr. odch.	Min	Max
Objem absolutně	9,97	26,53	0,00	153,13
Počet na sken	4,25	5,01	1	28

Tabulka 8.5. Statistika komponent datasetu Motol. První řádek hodnot obsahuje statistiku objemů v ml počítanou přes celý dataset. Druhý řádek hodnot značí četnost komponent na jeden sken.

Dataset obsahuje celkem 238 komponent lézí vypočítaných ze sjednocení FLAIR a DWI segmentace. Analýza komponent se nachází v tabulce 8.5.

Data byla koregistrována na šablonu prostoru MNI. Díky tomu bylo možné udělat analýzu laterality z tabulky 8.6. Uvedená data odpovídají sumě objemů voxelů ležících v dané hemisféře napříč datasetem. Je evidentní, že dataset je z tohoto hlediska silně nevyvážený.

Hemisféra	Objem léze [ml]
Levá	1971,31
Pravá	402,42

Tabulka 8.6. Analýza laterality datasetu Motol, kumulativní objem pro dané hemisféry.

Dále byla provedena analýza výskytu léze v jednotlivých částech mozku, výsledky jsou v tabulce 8.7. Toho bylo dosaženo sumou objemů voxelů ležících v patřičné oblasti použitého atlasu. Voxely ležící mimo atlas jsou označeny položkou Neuvedeno.

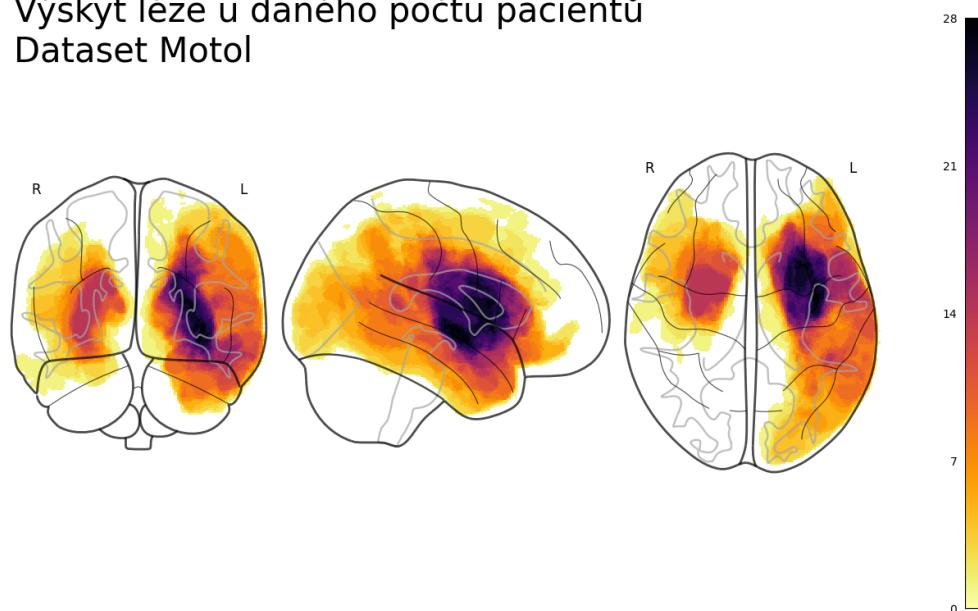
Rozložení lézí v datasetu je vizualizováno pomocí funkce Glass brain na obrázku 8.2.

Dataset Motol má sice 56 skenů, což je poměrně dost kvalitních a kvalitně anotovaných skenů, jenže na sobě nezávislých skenů je v datasetu málo (19 pacientů). Na základě znalostí z provedené rešerše jsem se rozhodl data rozšířit o další dva veřejně dostupné datasety ISLES 2015 [2] a ISLES 2022 [40]. 19 poskytnutých pacientů nemusí pokrývat všechny případy, které v reálném světě mohou nastat a které bude třeba v

Část atlasu	Objem léze [ml]
Neuvedeno	295,15
Caudate	138,74
Mozeček	0,02
Frontální lalok	349,26
Insula	261,35
Týlní lalok	156,63
Parietální lalok	282,82
Putamen	231,81
Spánkový lalok	630,00
Thalamus	27,95

Tabulka 8.7. Analýza datasetu Motol podle atlasu, kumulativní objem pro danou část mozku.

Výskyt léze u daného počtu pacientů Dataset Motol



Obrázek 8.2. Vizualizace rozložení lézí datasetu Motol po transformaci do prostoru MNI. Barva reprezentuje maximální počet skenů, jenž má ve stejném místě lézi, projektovaný do daného směru.

budoucnu segmentovat. Díky přidání nových dat by mělo dojít k lepšímu zobecňování automatického algoritmu a tedy i k lepším výsledkům v praktickém nasazení.

8.2 Dataset ISLES 2015

Ze soutěže ISLES 2015 [2] jsem využil trénovací množinu zveřejněného datasetu. Tato množina obsahuje 28 snímků, které byly pořízeny univerzitním zdravotnickým střediskem UKSH Universitätsklinikum Schleswig-Holstein v Lübecku. Všechny byly skenovaly na rezonanci Philips s indukcí 3 T, přesto jsou snímky relativně různorodé, jelikož mají různé rozlišení nebo různou kvalitu (vlivem artefaktů).

Dataset je různorodý, autoři do něj zahrnuli léze s jedním i více ložisky, léze s různými velikostmi atd.

Dataset obsahuje sekvence T1-w, T2-w, FLAIR a DWI ($b=1000$) a příslušnou segmentaci jednoho experta. Dataset obsahuje pouze pacienty v subakutní fázi CMP (po vizuální kontrole expertem, kdy léze je patrná na DWI i na FLAIR). Pacienti s WMH nebo chronickými lézemi vyřazeni nebyli, ale tyto oblasti nejsou oblastmi zájmu v segmentaci subakutní ischémické CMP.

Data byla autory datasetu anonymizována odstraněním lebky softwarem BET2 z roku 2005, tento krok byl posléze manuálně opraven. Voxely byly normalizovány na izotropní velikost 1 mm^3 . Snímky každého pacienta byly koregistrovány do prostoru FLAIR pomocí elastix toolboxu.

Autori zvolili, že jako hlavní sekvenci pro segmentaci použijí FLAIR, jelikož segmentace více expertů na této modalitě vykazují nejlepší shodu. Ostatní modality sloužily pouze jako podpora pro experts.

Charakteristika dat v nativním prostoru je ukázána v tabulce 8.8. Deskriptivní statistika lézí je uvedena v tabulce 8.9, z uvedeného DC je patrné, že aplikovaným preprocessingem v této práci nedošlo ke zhoršení kvality segmentace experts.

Počet voxelů snímku	Rozměry voxelu [mm]	Počet snímků
(230, 230, 153)	(1,0, 1,0, 1,0)	12
(230, 230, 154)	(1,0, 1,0, 1,0)	16

Tabulka 8.8. Nativní velikosti tensorů datasetu ISLES 2015. Počet voxelů označuje dimenze tensoru dat.

	Průměr	Směr. odch.	Min	Max
Objem léze	43,79	60,71	0,11	233,55
DC po preprocessingu	1,00	0,00	1,00	1,00

Tabulka 8.9. Deskriptivní statistika lézí datasetu ISLES 2015 přepočítaná na jeden sken. Uváděné objemy jsou v ml. DC byl vypočítán porovnáním expertní segmentace před a po preprocessingu.

Tento dataset obsahuje celkem 59 komponent lézí, analýza komponent se nachází v tabulce 8.10.

Komponenty	Průměr	Směr. odch.	Min	Max
Objem absolutně	20,78	46,63	0,02	233,55
Počet na sken	2,11	2,31	1	13

Tabulka 8.10. Statistika komponent datasetu ISLES 2015. První řádek hodnot obsahuje statistiku objemů v ml počítanou přes celý dataset. Druhý řádek hodnot značí četnost komponent na jeden sken.

Stejně jako u předchozího datasetu i zde byla provedena analýza laterality zobrazená v tabulce 8.11 a analýza výskytu léze podle atlasu zobrazená v tabulce 8.12. Ukázalo se, že i tento dataset je z pohledu laterality silně nevyvážený.

I v tomto případě byl vytvořen Glass brain obrázek 8.3 vizualizující rozložení lézí napříč trénovacím datasetem ISLES 2015 [2].

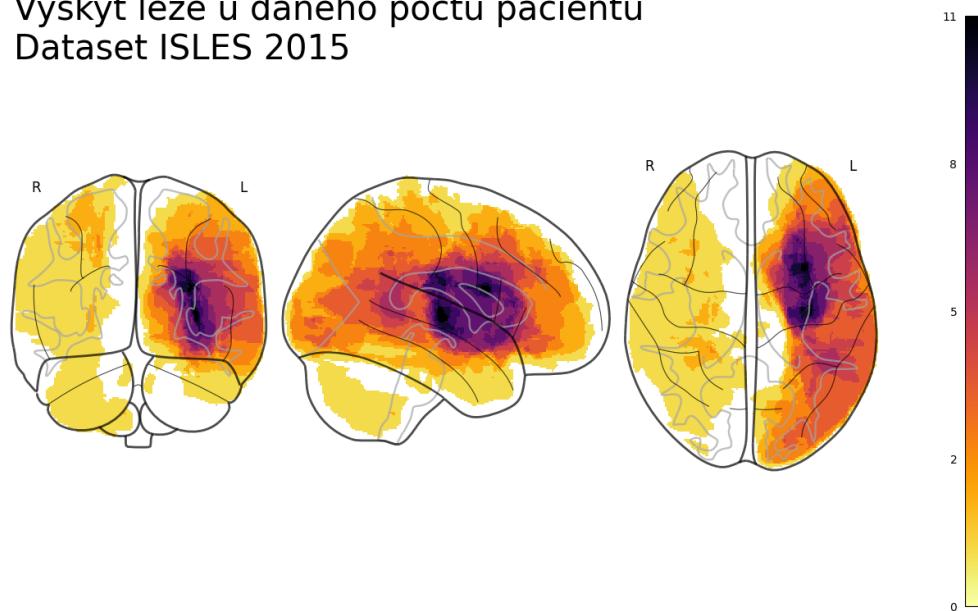
Hemisféra	Objem léze [ml]
Levá	1061,73
Pravá	164,53

Tabulka 8.11. Analýza laterality datasetu ISLES 2015, kumulativní objem pro dané hemisféry.

Část atlasu	Objem léze [ml]
Neuvedeno	127,80
Caudate	37,03
Mozeček	24,20
Frontální lalok	321,89
Insula	93,13
Týlní lalok	138,64
Parietální lalok	218,28
Putamen	53,18
Spánkový lalok	208,72
Thalamus	3,39

Tabulka 8.12. Analýza datasetu ISLES 2015 podle atlasu, kumulativní objem tkáně zasazený pro danou část mozku.

Výskyt léze u daného počtu pacientů Dataset ISLES 2015



Obrázek 8.3. Vizualizace rozložení lézí datasetu ISLES 2015 po transformaci do prostoru MNI. Barva reprezentuje maximální počet skenů, jenž má ve stejném místě lézi, projektovaný do daného směru.

8.3 Dataset ISLES 2022

Ze soutěže ISLES 2022 [40] jsem také využil trénovací dataset obsahující subakutní i akutní ischemické léze. Obsahuje celkem 250 skenů pocházejících ze 4 skenerů od 2

výrobců. Data byla pořízena ve dvou zdravotnických centrech. Oproti ISLES 2015 [2] tato data obsahují i malé mnohočetné kortikální a embolické infarkty.

198 snímků bylo pořízeno univerzitní nemocnicí Technické univerzity v Mnichově. Tyto snímky odpovídají 1-3 týdnům od počátku onemocnění a jsou pořízené po mechanické trombektomii.

52 snímků bylo pořízeno univerzitní nemocnicí Inselspital v Bernu. Snímky odpovídají 0 až 7 dnům od počátku onemocnění a jsou pořízeny před mechanickou trombektomií.

Aby data reprezentovala praxi, důvodem pro nezahrnutí skenu do datasetu byly pouze velké pohybové artefakty, které by byly nepoužitelné ani pro manuální diagnostiku. Naopak nebyly vyřazeny skeny s chybami kvality vzniklé poškozením signálu nebo např. pacienti s gliovými jizvami.

V rámci datasetu jsou k dispozici sekvence DWI ($b=1000$), ADC a FLAIR. Dále jsou k dispozici expertní segmentace léze a metadata ke skenům, jež uvádějí zdravotnické centrum; pohlaví, věk a váhu participantů. K vytvoření skenů byly použity přístroje 3 T Philips (Achieva a Ingenia), 3 T Siemens Verio a 1,5 T Siemens MAGNETOM Avanto a Aera.

Vytvoření expertních segmentací v tomto případě bylo složitější než v případě ISLES 2015 [2]. Složitost byla dána primárně velikostí datasetu. Z důvodu urychlení anotování byla autory natrénována vlastní neuronová síť 3D U-Net [63] na DWI snímcích z Technické univerzity v Mnichově. Tyto snímky byly již anotovány pro účely jiných projektů, přesto segmentace před trénováním 3D U-Net [63] nebyly kvalitní a museli je opravit neuroradiologové.

Po natrénování sítě došlo k předsegmentování zbylých dat, ty byly následně editovány a revidovány studenty medicíny, poté byly provedeny korektury předtestačním neuroradiologem a nakonec byly schváleny neuroradiologem s více než desetiletou praxí. V případě, kdy byla algoritmická segmentace příliš špatná, mohli studenti segmentovat plně manuálně. S přibývajícími daty autoři postupně dotrénovávali 3D U-Net [63], čímž získávali postupně lepší a lepší predikce, které bylo třeba méně upravovat.

Primární modalita pro segmentaci tedy byla DWI a ostatní modality poskytovaly podpůrnou informaci v případě nejistoty experts.

	Průměr	Směr. odch.	Min	Max
DC Radiolog A x ISLES22	0,90	0,09	0,69	0,98
DC Radiolog B x ISLES22	0,86	0,13	0,61	0,98
DC Radiolog A x Radiolog B	0,83	0,12	0,57	0,98
VD Radiolog A x ISLES22	2,37	2,59	0,11	7,34
VD Radiolog B x ISLES22	6,56	13,37	0,06	45,66
VD Radiolog A x Radiolog B	7,22	11,32	0,06	38,32

Tabulka 8.13. Analýza anotací ISLES 2022 proti dvěma nezávislým neuroradiologům. DC: Dice koeficient. VD: Rozdíl objemů. [40]

Autoři ISLES 2022 [40] také vybrali 10 skenů (pravděpodobně z trénovací množiny, kterou používám i z testovací), které byly předloženy dvěma zkušeným radiologům pro studium překryvu segmentací napříč různými experty (tzv. inter-rater analýza) viz obrázek 8.13. Ukázalo se, že hranice segmentace v datasetu ISLES 2022 [40] jsou spíše podsegmentované vzhledem k témtu dvěma neuroradiologům. Pravděpodobně tedy tito testovací neuroradiologové kladli větší důraz na modalitu FLAIR, ze které lze jasněji

hranice léze rozlišit. Dice koeficient i rozdíl v objemech, při srovnání těchto dvou neuroradiologů navzájem, je horší než při srovnání každého neuroradiologa zvlášt proti anotacím v datasetu ISLES 2022 [40]. Přesto je Dice koeficient ve všech případech lepší než inter-rater Dice koeficient u předešlého ISLES 2015 [2].

Snímky byly anonymizovány odstraněním lebky, což byl jediný krok, kterým byla data upravena. Data se tudíž nachází v nativním prostoru. Obsahují mnoho kombinací velikostí tensorů dat i mnoho kombinací velikostí voxelů.

Deskriptivní statistika se nachází v tabulce 8.14. V datasetu se nachází jak extrémně velké léze, tak tři pacienti v expertní segmentaci nemají označený žádný pozitivní voxel. U těchto pacientů bylo podezření na CMP, ale po MRI se to neprokázalo. Pacienti přesto byli pro různorodost zařazeni do datasetu. Poměrně nízké je minimum Dice koeficientu po preprocessingu, to však přičítám velmi malým lézím v datasetu, kdy malé zmenšení kvality má velký důsledek na DC.

	Průměr	Směr. odch.	Min	Max
Objem léze	23,42	47,00	0,00	482,15
DC po preprocessingu	0,95	0,04	0,75	1,00

Tabulka 8.14. Deskriptivní statistika lézí datestu ISLES 2022 přepočítaná na jeden sken. Uváděné objemy jsou v ml. DC byl vypočítán porovnáním expertní segmentace před a po preprocessingu.

Tento dataset obsahuje celkem 2506 komponent lézí, analýza komponent se nachází v tabulce 8.15. Zvláště z průměrného počtu ložisek na jednoho pacienta je patrné, že tento dataset obsahuje značné množství mnohočetných lézí.

Komponenty	Průměr	Směr. odch.	Min	Max
Objem absolutně	2,30	15,88	0,00	477,08
Počet na sken	10,01	16,22	0	184

Tabulka 8.15. Statistika komponent datasetu ISLES 2022. První řádek hodnot obsahuje statistiku objemů v ml počítanou přes celý dataset. Druhý řádek hodnot značí četnost komponent na jeden sken.

Hemisféra	Objem léze [ml]
Levá	2692,81
Pravá	3163,07

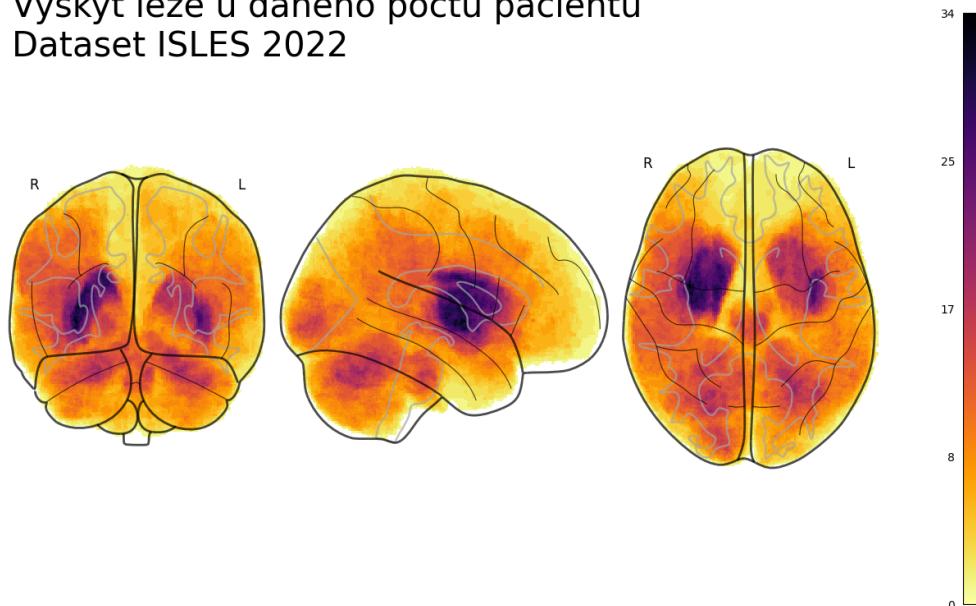
Tabulka 8.16. Analýza laterality datasetu ISLES 2022, kumulativní objem pro dané hemisféry.

Analýza laterality se nachází v tabulce 8.16 a analýza výskytu léze podle atlasu zobrazená v tabulce 8.17. Na rozdíl od předchozích je tento dataset relativně vyvážený jak z pohledu laterality, tak z pohledu lokace léze. To potvrzuje i vizualizace Glass brain na obrázku 8.4.

Část atlasu	Objem léze [ml]
Neuvedeno	488,04
Caudate	196,73
Mozeček	996,67
Frontální lalok	1165,75
Insula	321,49
Týlní lalok	589,48
Parietální lalok	997,30
Putamen	268,40
Spánkový lalok	759,55
Thalamus	72,47

Tabulka 8.17. Analýza datasetu ISLES 2022 podle atlasu, kumulativní objem tkáně zasažený pro danou část mozku.

Výskyt léze u daného počtu pacientů Dataset ISLES 2022



Obrázek 8.4. Vizualizace rozložení lézí datasetu ISLES 2022 po transformaci do prostoru MNI. Barva reprezentuje maximální počet skenů, jenž má ve stejném místě lézi, projektovaný do daného směru.

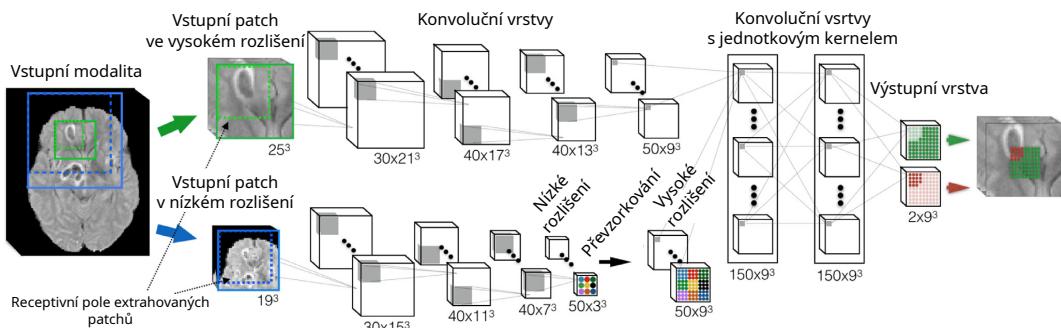
Kapitola 9

Experimentální ověření metody DeepMedic

První experimentálně ověřovanou metodou je DeepMedic [50–51], který byl uveden na soutěži ISLES 2015 [2], a tu vyhrál, stejně tak jako soutěž BRATS 2017. Od roku 2015 je postupně vyvíjen. Zdrojový kód metody je k dispozici na serveru GitHub. Metoda je implementována ve frameworku TensorFlow.

9.1 Popis metody

V této sekci krátce popíšu metodu v původní konfiguraci a její vývoj v čase.



Obrázek 9.1. Částečný nákres metody DeepMedic vzorkující v prvním enkodéru 25 voxelové patche v 3D prostoru, které následně procházejí konvolučními vrstvami s 30, 40, 40 a 50 filtry. Na výstupu jsou dvě konvoluční vrstvy s jednotkovým 3D kernelem a výstupní vrstva. Obdobně to platí pro druhý enkodér, který je před spojením příznaků s prvním enkodérem převzorkován na kompatibilní rozlišení. [51]

9.1.1 Konfigurace ISLES 2015

Neuronová síť se skládá z 11 3D konvolučních vrstev. Z důvodu příliš velké velikosti vstupních tensorů jsou extrahovány subtensity (tzv. patche). Vzhledem k tomu, že větší prostorový kontext vede k lepším výsledkům, DeepMedic [50–51] používá dva enkodéry. Do prvního z nich vstupují patche vzorkované z původního rozlišení. Druhý enkodér zpracovává patche prostorově centrované na stejné místo, jako první enkodér, patche jsou však vzorkované z nižšího rozlišení (podvzorkováno v poměru 3:1). Díky tomu patch o stejném počtu voxelů obsahuje větší prostorový kontext, stejně tak to platí pro konvoluční jádro. Tento způsob je výhodný, jelikož bez zvětšení patche nebo bez přidání vrstev se zvětšuje receptivní pole.

Konvoluční síť používá jednotkový krok konvolučního jádra. Počet výstupních kanálů prvních šesti vrstev je 40 (15 pro druhý enkodér) o velikosti 3D jádra 5, následující dvě vrstvy mají počet kanálů 80 (30 pro druhý enkodér) o velikosti jádra 3. Po této vrstvě probíhá převzorkování tensoru ve druhém enkodéru na velikost kompatibilní s příznaky v prvním enkodéru. Následují konvoluční vrstvy, které mají jednotkové kernely (učí se selektovat podstatné kanály předchozích vrstev) se 150 výstupními příznaky a obsahují

50 % Dropout, poslední vrstva obsahuje aktivační funkci softmax a výstupem jsou dvě pravděpodobnostní mapy daného patche odpovídající abnormální tkáni nebo zdravé tkáni. Vstupní data jsou metodou automaticky rozdělena do patchů a nakonec jsou segmentace opět spojeny do původního prostoru vstupního tensoru. ReLU je použita jako aktivační funkce vrstev. Jako optimalizátor je použit RMSProp. Pro zrychlení konvergence je použit BatchNorm. Nákres architektury je na obrázku 9.1.

Každý snímek musí být před vstupem do sítě normalizován na nulový průměr a jednotkovou směr. odchylku. Autoři použili augmentaci převrácením dat vzhledem k sagitální rovině. Načítání, augmentace a podvzorkování dat probíhá online během učení, tj. časově nezatěžuje. Pravděpodobnost, že centrální voxel patche patří do třídy lézí nebo zdravé tkáně je stejná tj. 50 %, čímž autoři řeší problém nevyváženého datasetu.

Při použití 4 vstupních modalit (T1-w, T2-w, FLAIR a DWI) a GPU Nvidia GTX Titan Black trvalo učení 1 den a predikce 3 minuty.

Výsledná metoda, která byla vyhodnocena autory na trénovacích ISLES 2015 [2] datech, se skládala z 5 modelů používajících ensembling průměrováním pravděpodobnostních map. Výsledky byly ověřeny 5 násobnou křížovou validací a dosáhly DC 0,64. Pravděpodobnostní mapy obsahovaly falešně pozitivně segmentované malé oblasti nebo naopak díry uvnitř pravdivě pozitivních segmentací. Proto autoři implementovali podmíněné náhodné pole (CRF; obecná verze Markovovského náhodného pole) a vylepšili DC na 0,66. Takto připravená metoda dosáhla v soutěži ISLES 2015 [2] DC 0,59.

Problémy metody DeepMedic [50–51] se objevují při segmentaci malých lézí, proto autoři doporučují rozdělit data do několika kategorií (např. podle velikosti lézí) a dle těchto kategorií jednotlivě natrénovat modely.

9.1.2 Pozdější vývoj

Od roku 2015 autoři metodu stále vyvíjí. Díky tomu se mění např. její efektivita, ale hlavně její doporučená konfigurace. Nyní DeepMedic [50–51] používá PReLU (Leaky ReLU s trénovatelným parametrem sklonu) jako aktivační funkci, jiné nastavení počtu kanálů a dalších parametrů.

Autoři v roce 2016 [51] provedli další experimenty s touto metodou. K natrénování DeepMedic [50–51] použili 45 snímků a k validaci 15 snímků pacientů po traumatickém poranění mozku. Sledoval se mj. vliv hyperparametrů (hloubka sítě, vliv dvou enkodérů, ...) na Dice koeficient.

Zajímavým a logickým testem bylo ověření metody vzorkování patchů. Dosavadní vzorkování pomocí rovné pravděpodobnosti pro obě třídy dosáhlo v důsledku vysoké sensitivity, ale nízké specificity, relativně nízkého Dice koeficientu. Na druhou stranu uniformní vzorkování z prostoru mozku (nezávislé na třídě centrálního voxelu) vykazuje lepší DC, extrémně vysokou specificitu a nízkou sensitivitu. Postupným zvětšováním velikosti patche však autoři našli konfiguraci, vyvažující sensitivitu a specificitu původního vzorkování a tím nejlepšího DC.

Vizuální kontrolou aktivací vrstev autoři zjistili, že první enkodér se naučil rozpoznávat druhy tkání a polohy komor, aktivace vrstev v druhém enkodéru vnáší do sítě prostorový bias přesně tak, jak autoři předpokládali.

9.2 Ověřované konfigurace DeepMedic

Pro dosáhnutí co nejkvalitnějšího výstupu ve smyslu přesnosti validace jsem se rozhodl pro trénování na datech z ISLES 2015 [2] a 2022 [40] a k validaci na datasetu Motol. Předpokládám, že reálná data budou obdobná datasetu Motol. Použití odlišných dat

pro trénink a validaci pouze na datasetu Motol by měla dát střízlivou představu o výkonnosti modelu použitého v reálném prostředí. Data proto byla rozdělena na trénovací a validační v poměru 278:56.

Preprocessing dat byl proveden v souladu s popsanou metodikou pomocí třídy `Subject`. Data tedy byla koregistrována pomocí uložených transformací, byl vymaskován mozek, snímky byly následně převzorkovány na rozměry tensoru (200, 200, 200) a na isotropní velikost voxelu (1, 1, 1) mm. Následně byla data normalizována na nulový průměr a jednotkovou směr. odchylku v rámci masky mozku (nikoliv v rámci celého tensoru). Metodě DeepMedic [50–51] byly poskytnuty masky mozku, díky čemuž mohlo proběhnout efektivní vzorkování patchů resp. bylo minimalizováno vzorkování volného prostoru neobsahující tkáň.

Konfigurační soubory metody DeepMedic [50–51] předpokládají definování počtu extrahovaných patchů na jeden sken, počet načtených skenů (subepocha) a počet subepoch v rámci jedné epochy. Kromě toho se definuje batch složený z několika patchů jednoho pacienta. Pro všechny testované konfigurace jsem zvolil 1000 extrahovaných patchů a 50 skenů na subepochu. Jedna epocha se skládala z 20 subepoch. Celé trénování trvalo 15 epoch. Jeden batch obsahoval 10 patchů během trénování sítě.

Úplná validace je časově extrémně náročná, proto je vhodné provádět průběžně částečnou validaci. Ta poskytne orientační pohled na přesnost sítě. Proto po dokončení každé subepochy bylo extrahováno 5000 patchů o velikosti (17, 17, 17), které byly validovány. Výsledek validace byl následně zalogován. Kompletní predikce a validace proběhla vždy po 5 epochách. Celkem proto byly provedeny 2 kompletní validace v průběhu učení a jedna na konci učení. Vzhledem k výpočetní náročnosti tedy nelze efektivně použít brzké zastavení, jelikož validace neprobíhá na celém datasetu. Validační Dice koeficient má značný rozptyl závislý na extrahované části dat.

■ 9.2.1 Testované modely

Konfigurace všech testovaných modelů se nachází v repozitáři ve složce `deepmedic_workspace`.

Aktuální autory doporučený model je označený jako *vanilla*. Obsahuje tři enkodéry. Všechny z nich mají vrstvy s filtry [30, 30, 40, 40, 40, 40, 50, 50] a kernely o velikosti 3. Druhý a třetí enkodér používá podvzorkování faktorem 3 resp. 5. Po sjednocení příznaků z enkodérů navazuje konvoluční vrstva s 250 filtry a kernelem velikosti 3. Následuje vrstva s 250 filtry a výstupní vrstva, obě s jednotkovým kernelem. Síť obsahuje ve všech enkodérech i mezi vrstvami s 250 filtry residuální spojení. Velikost vstupního patche je (37, 37, 37) voxelů. V rámci extrahovaného patche je použita augmentace modifikací histogramu intenzit definovaná Gaussovým rozdělením.

Dále byl otestován model *no_augmentation* vycházející z předchozího modelu *vanilla*. V tomto případě byla odstraněna veškerá augmentace. Obdobně další model *image_augmentation* vychází z modelu *vanilla* a přidává k němu augmentaci, kdy na načtený sken je s pravděpodobností 50 % náhodně rotován o max. 45 ° v každé ose a částečně je změněno měřítko.

Dále byl ověřen model *ISLES_2015*, který konfigurací sítě odpovídá co nejvíce původním modelům natrénovaným autory v soutěži ISLES 2015 [2]. Tzn. obsahuje pouze dva enkodéry, hlavní má počty filtrů [40, 40, 40, 40, 40, 40, 80, 80], druhý má počty filtrů [15, 15, 15, 15, 15, 15, 30, 30] a je podvzorkován faktorem 3. Poslední vrstvy nemají 250 filtrů, ale pouze 150. Je extrahována velikost patche (25, 25, 25). Augmentace změnou intenzit patche je mírnější než u *vanilla* modelu.

Model *150FC* značí, že oproti modelu *vanilla* byl zmenšen počet filtrov v posledních vrstvách z 250 na 150. Model *150FC_1L* používá pouze jednu vrstvu se 150 filtrov, celý model je tedy mělčí o jednu vrstvu. Analogicky je vytvořena konfigurace modelu *75FC*.

Další ověřovaný model *7layers* opět vychází z modelu *vanilla* a má odstraněnou jednu vrstvu se 40 filtrov ve všech třech enkodérech. Stejně tak model *6layers* má odstraněné proti modelu *vanilla* celkem 2 vrstvy, které obsahovaly 40 filtrov. Pro doplnění informace vlivu Dice koeficientu na konfiguraci sítě jsou testovány dva modely *7layer_150FC* a *6layers_150FC*, které vychází z modelů *7layers* a *6layers*, ale místo 250 filtrov ve vrstvách s jednotkovým kernelem obsahují pouze 150 filtrov.

9.3 Vstupní fúze

První experimenty proběhly s defaultní vstupní fúzí, při níž byly modality FLAIR a DWI spojeny přes dimenzi kanálů. Vstup do sítě tedy byl o rozložení (2, 200, 200, 200). Takto bylo ověřeno celkem 11 různých modelů s cílem dosáhnout co nejlepšího DC.

Natrénování modelu v konfiguraci *vanilla* trvalo celkem necelých 7 hodin. Úplná validace na datasetu Motol trvala necelých 8 minut. V případě konfigurace *image_augmentation* bylo trénování výrazně pomalejší a trvalo 19 hodin. Rychlosť validace nebyla ovlivněna. Časy trénování ani validace ostatních modelů se zásadně nelišily od modelu *vanilla*. Výrazné zpomalení trénování v důsledku augmentace přičítám poměrně výkonné grafické kartě, která provede tensorové operace velmi rychle a následně se čeká na procesorové dopočítání augmentace. Dle autorů totiž DeepMedic [50–51] dokáže načítat a augmentovat data online. Augmentace by tedy neměla zpomalovat trénování, přesto se tak děje.

Na konci trénování, tj. po 15 epochách všechny modely dosahují Dice koeficientu na trénovací množině od 0,88 (konfigurace *ISLES_2015*) po 0,93 (konfigurace *no_augmentation*). Výsledky evaluace v tabulce 9.1 odpovídají segmentaci původních dat datasetu Motol (segmentace jsou převzorkovány do nativního prostoru datasetu Motol).

Model	DC	Prec	Sens	Spec	F
75FC	0,427±0,276	0,422	0,626	0,962	10
150FC	0,411±0,279	0,414	0,586	0,960	10
6layers_150FC	0,408±0,275	0,361	0,670	0,954	10
ISLES2015_setup	0,408±0,279	0,354	0,681	0,953	10
7layers_150FC	0,404±0,273	0,360	0,655	0,956	10
no_augm.	0,399±0,277	0,352	0,659	0,953	10
150FC_1L	0,390±0,277	0,362	0,637	0,946	11
7layers	0,386±0,270	0,332	0,668	0,949	10
6layers	0,378±0,273	0,325	0,675	0,943	10
vanilla	0,366±0,261	0,307	0,682	0,939	11
image_augm.	0,356±0,265	0,290	0,699	0,935	11

Tabulka 9.1. Výsledky 11 ověřovaných modelů metody DeepMedic ve formátu průměr ± směr. odchylka. Modely byly natrénovány na sloučení datasetů ISLES 2015 a ISLES 2022. Validace proběhla na datasetu Motol proti expertním segmentacím v jejich nativním prostoru. DC: Dice koeficient. Prec: Precisnost. Sens: Sensitivita. Spec: Specificita. F: Počet neúspěšných segmentací (DC < 0,1).

Nejlepších výsledků dosáhl model v konfiguraci *75FC*, kde Dice koeficient dosáhl $0,427 \pm 0,276$, precisnosti $0,422 \pm 0,321$ a sensitivity $0,626 \pm 0,239$. Naopak nejhorských vý-

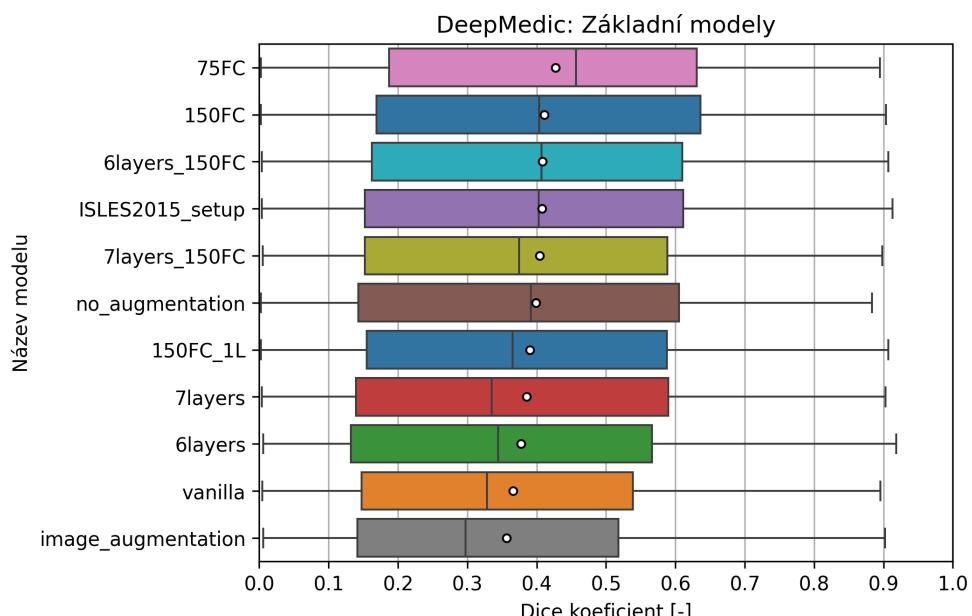
sledků dosáhl model *image_augmentation* a to pouze DC $0,356 \pm 0,265$, což je poměrně podstatný rozdíl. Křivka DC na trénovací množině v závislosti na subepochách ale ukázala, že i přes augmentaci měl model dostatečný počet iterací a spolehlivě dokončoval. Nízké hodnoty DC tedy nejsou způsobeny nedostatečným počtem epoch při učení modelu.

Je zajímavé, že sensitivita modelů se pohybuje od 0,59 po 0,70, ale precisnost se pohybuje pouze od 0,29 po 0,42. U nejlepšího modelu *75FC* došlo k vybalancování precistnosti a sensitivity (konkrétně ke zvýšení přesnosti a snížení sensitivity), čímž se zvýšil DC. Naopak hůře vybalancovaný je nejhorský model *image_augmentation*, který má sice nejlepší sensitivitu, ale má velmi špatnou precistnost.

Většina modelů nedokázala správně osegmentovat 10 z 56 skenů, z toho tři metody selhaly u 11 skenů.

Z výsledků je patrné, že největší vliv na zvýšení Dice koeficientu má snižování počtu filtrů vrstev s jednotkovými kernely. Odstraněním konvolučních vrstev na začátku sítě se nedosáhlo až takového zlepšení. Konfigurace sítě doporučená autory dosáhla druhého nejhoršího výsledku z pohledu Dice koeficientu.

Graf zobrazující rozložení Dice koeficientů napříč datasetem Motol je na obrázku 9.2.

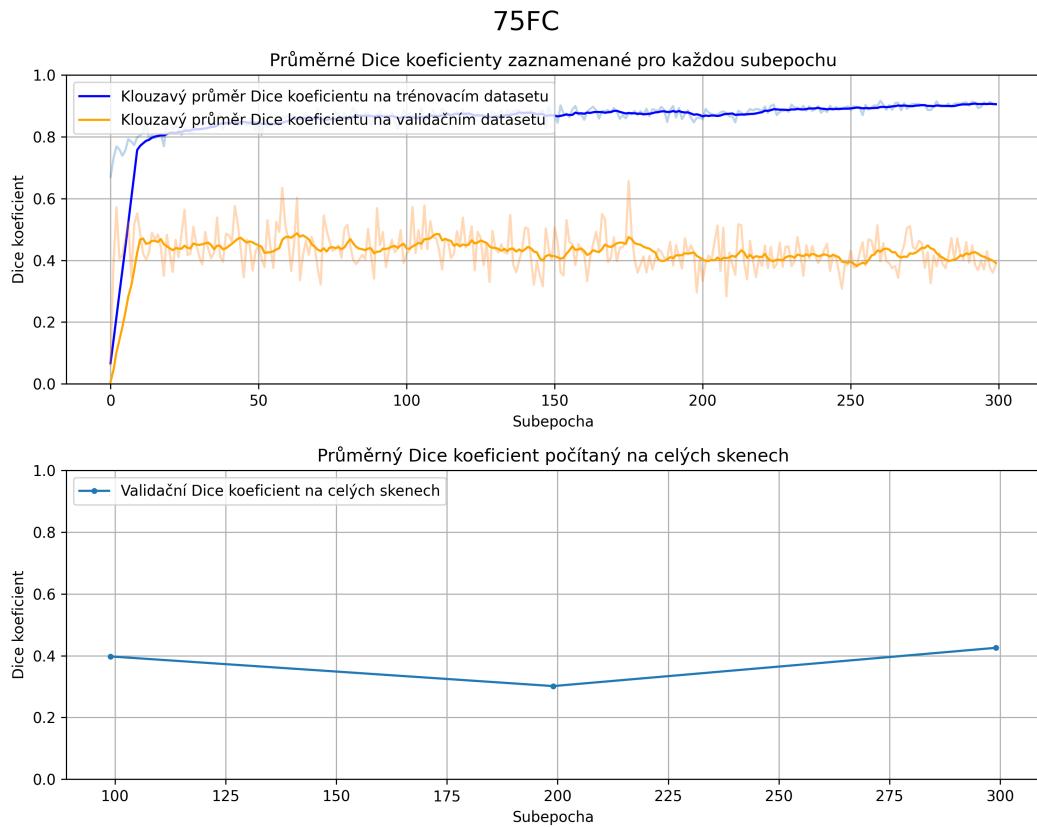


Obrázek 9.2. Výsledky 11 ověřovaných modelů metody DeepMedic se vstupní fúzí. Modely byly natrénovány na sloučení datasetů ISLES 2015 a ISLES 2022. Validace proběhla na datasetu Motol proti expertním segmentacím v jejich nativním prostoru. Box označuje interkvartilové rozpětí (IQR), vousy 1,5 IQR, čára v boxu označuje medián a kolečko průměr.

A nakonec graf závislosti Dice koeficientu v průběhu trénování pro nejlepší konfiguraci *75FC* je zobrazen na obrázku 9.3.

9.4 Výstupní fúze

Vzhledem k tomu, že model *75FC* dosáhl nejlepších výsledků se vstupní fúzí z pohledu Dice koeficientu, rozhodl jsem se tento model ověřit i při výstupní fúzi, k čemuž není nutná modifikace kódu již implementované metody DeepMedic [50–51].



Obrázek 9.3. Výsledky modelu 75FC v závislosti na subepoše. Horní graf odpovídá trénovacímu a částečnému validačnímu DC. Spodní graf poté odpovídá úplné validaci, která byla provedena dvakrát během trénování.

Proto byly natrénovány další dva modely označené jako *75FC_FLAIR* a *75FC_DWI*. Tyto nové modely se od modelu *75FC* liší pouze počtem vstupních modalit, kdy do každého z nich vstupoval tensor pouze s modalitou odpovídají jejich názvu.

Po natrénování modelů byla provedena predikce léze na datasetu Motol na daných instancích modelů tj. pro každou sekvenci zvlášt. Pravděpodobnostní mapy vystupující z poslední vrstvy s aktivační funkcí softmax byly místo prahování uloženy.

Tyto dvě pravděpodobnostní mapy byly fúzovány průměrováním pravděpodobností v jednotlivých voxelech a následně prahovány fixní hodnotou 0,5. Takto vznikly predikce označené *ensemble_DWI_FLAIR*.

Kromě toho byly uloženy i pravděpodobnostní mapy původního modelu *75FC*. Fúzováním predikcí modelů *75FC*, *75FC_FLAIR* a *75FC_DWI* vznikla další nová sada predikcí označená jako *ensemble_75FC_DWI_FLAIR*.

Byla provedena i evaluace nefúzovaných unimodálních modelů *75FC_FLAIR* a *75FC_DWI* na datasetu Motol. Pro konzistenci (a v souladu s uvedenou metodikou výše) byly predikce modelů srovnány opět proti logickému sjednocení expertních segmentací FLAIR a DWI.

Tabulka 9.2 zobrazuje statistiku dosažených výsledků. Je zde uveden pro srovnání i původní model se vstupní fúzí *75FC*.

Nejlepší predikce metody DeepMedic [50–51] vznikly výstupní fúzí dvou modelů natrénovaných separátně na sekvenci FLAIR a DWI. Takovýto model označovaný v práci jako *ensemble_DWI_FLAIR* dosáhl DC $0,510 \pm 0,259$. Významně také klesl počet skenů,

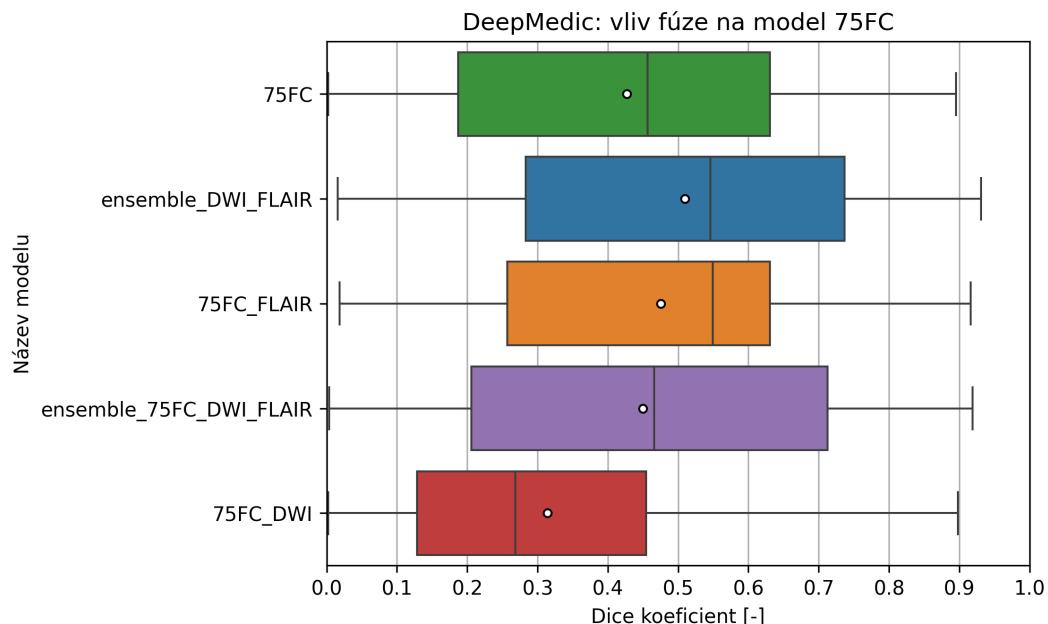
Model	DC	Prec	Sens	Spec	F
75FC	0,427±0,276	0,422	0,626	0,962	10
ens._DWI_FLAIR	0,510±0,259	0,602	0,545	0,991	4
75FC_FLAIR	0,476±0,229	0,631	0,457	0,994	3
ens._75FC_DWI_FLAIR	0,450±0,287	0,475	0,607	0,969	9
75FC_DWI	0,314±0,246	0,271	0,619	0,924	12

Tabulka 9.2. Výsledky výstupní fúze metody DeepMedic ve formátu průměr \pm směr. odchylka. Fúzované predikce (předpona ens) vznikly kombinací pravděpodobnostních map modelů 75FC, 75FC_FLAIR a 75FC_DWI. Validace proběhla na datasetu Motol proti expertním segmentacím v jejich nativním prostoru. DC: Dice koeficient. Prec: Precisnost. Sens: Sensitivita. Spec: Specificita. F: Počet neúspěšných segmentacích (DC < 0,1).

u kterých DeepMedic [50–51] selhal. Pro kvalitativní zhodnocení jsou zobrazeny predikce modelu ensemble_DWI_FLAIR v příloze B.

Predikce ensemble_DWI_FLAIR společně s predikcemi modelu 75FC_FLAIR dosáhly jako jediné modely vyšší precisnosti než sensitivity. Na druhou stranu model 75FC_FLAIR dosahuje nejhorší sensitivity ze všech testovaných modelů metody DeepMedic [50–51]. Proti tomu model 75FC_DWI dosahuje sice nejhoršího DC i nejhorší precisnosti, naopak má poměrně vysokou sensitivitu. Fúze pravděpodobnostních map vygenerovaných modely 75FC_FLAIR a 75FC_DWI tak dokáže efektivně zúročit přednosti obou z těchto modelů.

Graf s rozložením Dice koeficientů evaluace po výstupní fúzi je na obrázku 9.4.



Obrázek 9.4. Výsledky výstupní fúze v porovnání s nejlepším modelem se vstupní fúzí 75FC. Validace proběhla na datasetu Motol proti expertním segmentacím v jejich nativním prostoru. Box označuje interkvartilové rozpětí (IQR), vousy 1,5 IQR, čára v boxu označuje medián a kolečko průměr.

Kapitola 10

Experimentální ověření metody nnU-Net

Druhou experimentálně ověřovanou metodou je nnU-Net („no new Net“) [53, 70]. Je to obecná metoda umožňující segmentovat rozličné medicínské problémy. Jedná se o naprostě univerzální nástroj pohybující se v mnoha oblastech segmentace na přední hranici současného stavu poznání. Týmy používající nnU-Net [53] skončily na soutěži ISLES 2022 [40] mezi prvními a metoda nnU-Net [53] společně s architekturami typu U-Net tvořily drtivou většinu použitých metod. Zdrojový kód nnU-Net [53] je opět k dispozici na serveru GitHub a metoda je implementována ve frameworku PyTorch, který je v současnosti hojně používaným nástrojem pro vývoj neuronových sítí.

10.1 Popis metody

Standardní postup při implementaci nové metody založené na dobře známé architektuře U-Net [71] (mj. popsané v další kapitole) zahrnuje ladění preprocessingu, konkrétní nastavení nebo úpravy architektury, ladění hyperparametrů jako je krok učení, velikost mini-batch, velikost extrahovaných patchů, nastavení augmentace, atd. Tyto manuálně nastavované parametry generují poměrně vysokodimenzionální prostor a nalezení optimálních parametrů pro danou úlohu není vůbec jednoduché. Úspěšnost takto vytvořeného modelu je vysoce závislá na individuální expertní znalosti a zkušenosti programátora hlubokého učení. Aby bylo dosaženo ještě kvalitnějších predikcí po manuální estimaci hyperparametrů, často se používá metoda pokus-omyl, kdy se aktualizuje nějaký z hyperparametrů neuronové sítě a následně se určuje vliv této změny na přesnost predikce. Tímto způsobem nalezené parametry jsou obvykle suboptimální a v případě nového datasetu je nutné hyperparametry hledat znovu.

Další možností je použití nějaké z metod automatického strojového učení (AutoML), která automaticky prohledává prostor hyperparametrů a optimalizuje je na základě empirických poznatků. Vzhledem k počtu možných kombinací hyperparametrů a vzhledem k tomu, že optimalizace probíhá na základě empirického poznání, jsou takovéto metody časově velmi neefektivní.

Článek o nnU-Net [70] proto v roce 2018 přichází s přelomovou metodou, která hyperparametry nastavuje v závislosti na tzv. otisku datasetu. Autoři nnU-Net [70] shledávají, že úspěšnost metody záleží jen minimálně na modifikacích architektury neuronové sítě, ale je extrémně důležité nastavení parametrů preprocessingu a parametrů učení. Tyto parametry jsou metodou nnU-Net [70] odhadovány právě z datasetu, na který má být metoda nnU-Net [70] natrénována. Tím pádem lze metodu použít univerzálně na jakýkoliv dataset a adaptace hyperparametrů sítě téměř nevytíší dobu učení.

Autoři metody identifikovali v rámci soutěží pořádaných pod společností MICCAI, že mnoho soutěžících využívá pro segmentace ledvin a nádorů ledvin populární architekturu U-Net [71] nebo její 3D variantu [63]. Přestože soutěžící používají nemodifikované architektury nebo velice podobné architektury vůči původnímu U-Net, jejich dosahované úspěšnosti se velice liší a tím se liší i jejich umístění v soutěži. To vedlo autory nnU-Net [53] k dedukci, že není důležité soustředit se na změny v architektuře jako je

implementace residuálních spojení, zapojení více konvolučních vrstev v rámci stejného stupně enkodéru, atd. Za naprostoto klíčové autoři shledávají správnou konfiguraci parametrů sítě jako je velikost vstupu, velikost batche, normalizace dat, volba ztrátové funkce, atd. Metoda nnU-Net [53] se proto snaží vytvořit nový framework, jenž nepotřebuje k získání kvalitních výstupů zkušeného programátora v oblasti segmentace medicínských dat, který dokáže odhadnout volitelné parametry.

nnU-Net [53] identifikuje a rozděluje volitelné parametry na tři základní skupiny. První z nich jsou fixní parametry, tzv. volba těchto parametrů by neměla být ovlivněna vstupním datasetem. Tyto parametry jsou tedy napříč různými datasety neměnné, jejich výchozí nastavení by mělo zaručovat robustní nastavení pro segmentaci medicínských dat. Další skupinou jsou parametry závislé na vstupních datech. Tyto parametry jsou nastaveny pomocí pravidel (heuristik), které vychází z otisku vstupního datasetu. Autoři se snaží najít co nejvíce takovýchto pravidel, jelikož nastavení parametrů pomocí nich nezpůsobí znatelné prodloužení doby vývoje modelu. Vstupní dataset je na začátku automaticky analyzován, je vytvořen otisk tohoto datasetu a na základě něj jsou aplikována pravidla pro nastavení všech důležitých parametrů sítě, které závisí na vstupních datech. Třetí skupinou parametrů jsou tzv. empirické parametry, jež jsou nastaveny až po natrénování neuronové sítě. Tyto parametry se získávají až díky evaluaci vytvořených predikcí. Autoři se tyto empirické parametry snaží minimalizovat na nulu, jelikož jsou časově nákladné.

Co se týče fixních parametrů, autoři preferují instanční normalizaci batche před BatchNorm. Na rozdíl od BatchNorm probíhá tato normalizace pouze přes konkrétní kanál a konkrétní vzorek v batch, nikoliv přes celý batch. Aktivační funkce ReLU v U-Net jsou nahrazeny Leaky ReLU. Pro lepší propagaci chyby jsou přidány podpůrné ztrátové funkce do každé úrovně dekodéru (kromě dvou nejnižších). Konvoluce v první úrovni obsahují 32 kanálů a počet je vždy zdvojnásoben po MaxPoolingu s tím, že maximální počet kanálů je zastropován na 320 pro 3D U-Net. Další fixní parametr je nastavení učení ve smyslu počtu epoch (1000), kdy každá z nich je definována jako 250 průchodů mini-batch. Ztrátová funkce je součet vážené křížové entropie s Dice koeficientem. nnU-Net [53] používá iterativní snižování velikosti kroku učení v závislosti na epoše a Momentum. Nevyváženosť datasetu je řešena tím, že 2/3 patchů se vzorkují z náhodné nenulové oblasti a 1/3 patchů obsahuje nějakou ze segmentovaných tříd. V případě, že batch má velikost 2, poté je zaručeno, že jeden patch obsahuje expertem segmentované voxely. Mezi fixní parametry také patří nastavení augmentace, která je složená z rotace, změny měřítka, Gaussovským zašuměním nebo rozostřením, posuvem kontrastu nebo jasu, snížením rozlišení, zrcadlením, atd. Poslední fixní parametry se týkají predikcí. Patche predikcí se překrývají jednou polovinou a je aplikována tzv. test-time augmentace zrcadlením obrazu podél os. Tato augmentace zlepšuje globální přesnost predikce. Během agregace pravděpodobnostní (až 64 různých predikcí pro jeden vstupní patch díky test-time augmentaci) je Gaussovou funkcí přerozdělena váha pravděpodobností tak, že centrální voxel patche má maximální váhu.

Během vytvoření otisku datasetu jsou data oříznuta tak, aby minimálně obsahovala prostor s nulovou hodnotou. Následně jsou z datasetu extrahovány údaje jako velikost tensorů dat před a po oříznutí, rozměry voxelů, druhy modalit, počet tříd, počet snímků, průměr a směr. odch. intenzit, percentily intenzit, atd. Tato sesbíraná data jsou uložena jako otisk datasetu a určují nastavení tzv. rule-based parametrů. Prvním parametrem pro data (kromě CT) je z-normalizace každého snímku; v případě, že došlo k ořezu velkého množství voxelů, děje se normalizace pouze na nenulových voxelech. V případě, že data nemají stejnou velikost voxelů, jsou data převzorkována na mediánovou hod-

notu velikosti voxelu pro každou osu. Principem nnU-Net [53] je maximalizovat velikost patche za dodržení podmínky alespoň 2 patchů v jednom mini-batch. Vstupní tensor do neuronové sítě má mediánové rozměry tensorů v datasetu a v případě, že velikost není kompatibilní s U-Net, dojde k paddingu. Batch dat prochází sítí U-Net do doby, než se zmenší rozměr některé z os na méně než 4 voxely. Pokud má nějaká z os vyšší rozlišení, dojde k MaxPoolingu pouze přes tuto osu, aby došlo k vyrovnání rozlišení s osami v menším rozlišení.

nnU-Net [53] ve výchozím nastavení natrénuje dvě až tři konfigurace, které odpovídají 2D, 3D U-Net a v případě malého objemu patche vůči objemu snímku se natrénuje i kaskáda 3D U-Net (tedy nejprve 3D U-Net v nízkém rozlišení a posléze v původním rozlišení). V rámci empirických parametrů nnU-Net [53] vyhodnotí přesnost jednotlivých konfigurací nebo jejich kombinace pomocí výstupní fúze průměrováním a automaticky vybere nejlepší konfiguraci nebo kombinaci (ensemble). Jako druhý empirický krok nnU-Net [53] vybere největší komponentu lézí a ostatní minoritní komponenty odstraní. Pokud tento krok vede ke zlepšení predikce, automaticky jej bude aplikovat i v rámci reálného nasazení.

Autoři bohužel v článcích neuvádí implementační detaile, nicméně podrobnosti k pravidlům automatické konfigurace učení pomocí metody nnU-Net [53] lze nalézt v článku [53] na obrázku 2 nebo v kapitole Methods nebo v přílohách k tomuto článku.

Metoda nnU-Net [53] byla vyvinuta na celkem 10 datasetech zahrnujících segmentaci orgánů (segmentace z MRI: jater, slinivky, sleziny a segmentace z CT: levé síně, hipokampu, prostaty) a také segmentaci abnormalit (segmentace z MRI: nádorů jater, plicních uzlů, nádorů slinivky, rakoviny tlustého střeva a segmentace z CT: nádorů mozku). Autoři v článku z roku 2021 [53] demonstrovali výsledky na dalších 13 datasetech. Demonstrace proběhla na celkem 53 segmentačních úlohách zahrnujících i segmentace buněčných struktur ze snímků z elektronového nebo fluorescenčního mikroskopu. U 33 úloh metoda nnU-Net [53] předčila dosavadní nejlepší výsledky a ve zbylých případech byly výsledky nnU-Net [53] velice blízko nejlepším výsledkům ostatních metod. To značně naprostou unikání univerzálnost této automatické metody, která dokáže efektivně extrahat podstatné informace ze vstupních datasetů a díky nim efektivně škálovat architektury typu U-Net. nnU-Net [53] přináší robustní řešení, které potvrzuje, že ve většině případů je zásadní nastavení zpracování dat nebo nastavení globálních parametrů celé metody než úpravy relativně staré architektury neuronové sítě U-Net.

10.2 Vstupní fúze

Konfigurace nnU-Net [53] jsou uložené ve složce *nnunet_workspace*. Preprocessing dat je proveden pomocí třídy Subject a skládá se z obdobného postupu jako u metody DeepMedic [50–51], avšak je zde vynechána normalizace intenzit (jelikož je implementována interně v nnU-Net [53]). Preprocessing se tedy skládal z koregistrování snímků a expertních segmentací, extrakcí lebky z MRI, převzorkování na rozměry tensoru (200, 200, 200) voxelů a na isotropní velikost voxelu (1, 1, 1) mm. K trénování byly využity snímky z datasetů soutěží ISLES 2015 [2] a ISLES 2022 [40].

Dokumentace nnU-Net [53] je částečně v rozporu s publikovanými články autorů, jelikož se v ní uvádí, že metoda vyžaduje jednotný rozměr vstupních tensorů, jejich koregistrování a pokud možno jednotné rozměry jednotlivých voxelů. Co se týče formátu vstupních dat do metody, tak je možné snadno zaměnit jednotlivé modality (které jsou

identifikovány číslem a názvem skenu v rámci názvu souboru). Proto je i v dokumentaci kladen apel na to, aby nedošlo k jejich záměně, zvláště při reálném nasazení metody.

Ve složce *nnunet_workspace* se také nachází složka s výstupem nnU-Net [53] - s predikcemi, s otisky datasetů, nastavením architektury, s natréovanými modely a s predikcemi.

Nutno podotknout, že trénování bylo oproti navržené metodě podle článku upraveno. Trénování probíhalo pouze na 500 epochách místo 1000 epochách. Dále nebylo vůbec využito empirických znalostí, jelikož nnU-Net [53] byl použit pouze v konfiguraci 3D U-Net (3d_fullres), jelikož 2D U-Net nedokáže zpracovat prostorovou informaci a tudíž obecně dosahuje horších výsledků. Pro kaskádní zpracování zase nejsou snímky v dostatečně velkém rozlišení.

K trénování veškerých instancí nnU-Net [53] byly použity nové šablony „nnU-Net ResEnc“ [72–73], které implementují residuální bloky při konstruování U-Netu na základě pravidel. Díky těmto blokům se autorům zlepšila přesnost v řádu desetin procent Dice koeficientu. Konkrétně byl využit nejmenší plán „M“, který je vhodný pro grafické karty s pamětí 9–11 GB. Autorům nnU-Net [53] trvalo trénování 1000 epoch takové konfigurace metody na grafické kartě NVIDIA A100 zhruba 12 hodin.

nnU-Net [53] používá pro načítání dat knihovnu SimpleITK. Vstupní patch byl o velikosti (128, 128, 128) voxelů při isotropní velikosti voxelu (1, 1, 1) mm. Residuální 3D U-Net obsahuje celkem 6 úrovní v nichž jsou konvoluční vrstvy o následujících počtech kanálů [32, 64, 128, 256, 320, 320]. Velikost konvolučního kernelu je vždy (3, 3, 3). V dekodéru byla vždy pouze jedna konvoluční vrstva.

S uvedeným nastavením architektury byly natrénovány dvě instance do nichž tedy vstupovaly modality FLAIR a DWI sloučené podél osy kanálů, jakožto vstupní fúze. První instance s označením *Strokes* a druhá instance *StrokesMNI*. Přípona MNI v tomto případě znamená, že data byla koregistrována do prostoru MNI v rámci preprocessingu třídu Subject. Koregistrace byla provedena na FLAIR šablonu [65] o rozměru (193, 229, 193) a následně byla standardně pomocí třídy Subject převzorkována na velikost tensoru (200, 200, 200) a jednotkovou velikost voxelu. Výsledky evaluace na datasetu Motol jsou v tabulce 10.1.

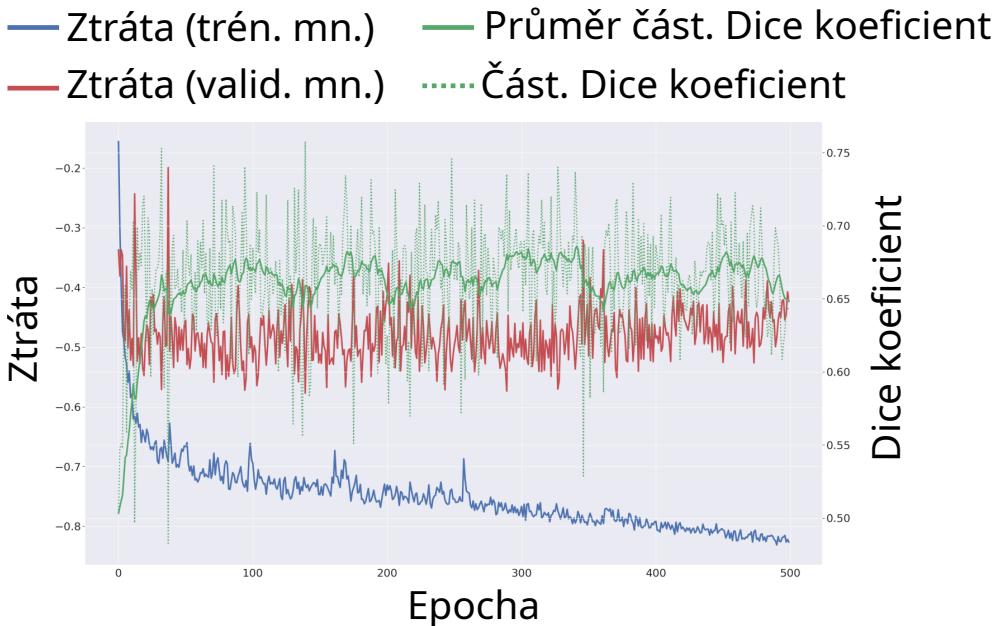
Model	DC	Prec	Sens	Spec	F
StrokesMNI	0,559±0,265	0,650	0,609	0,991	5
Strokes	0,554±0,262	0,675	0,563	0,993	4

Tabulka 10.1. Výsledky metody nnU-Net pro vstupní fúzi ve formátu průměr ± směr. odchylka. Modely byly natrénovány na sloučení datasetů ISLES 2015 a ISLES 2022. Validace proběhla na datasetu Motol proti expertním segmentacím v jejich nativním prostoru. DC: Dice koeficient. Prec: Precisnost. Sens: Sensitivita. Spec: Specificita. F: Počet neúspěšných segmentacích (DC < 0,1).

Natrénování jedné instance modelu trvalo necelých 15 hodin na poskytnutém výpočetním serveru. Vytvoření všech predikcí validačního datasetu Motol trvalo přes 4 minuty. Nutno podotknout že predikce odpovídá času bez preprocessingu a data jsou tedy v prostoru trénovací množiny.

Z pohledu Dice koeficientu je mírně lepší model *StrokesMNI*, který dosahuje DC $0,559 \pm 0,265$. Nicméně například při pohledu na počet neúspěšných segmentací, tak zde je o jeden sken horší. Obecně jsou výsledky velmi podobné. Dle tabulky lze soudit, že opět lepší Dice koeficient získal model s menším odstupem precisnosti a sensitivities.

Průběh trénování modelu *StrokesMNI* je zachycen na obrázku 10.1.



Obrázek 10.1. Výsledky modelu StrokesMNI v závislosti na epoše. Výpočet ztrátové funkce probíhá pouze na malé části datasetu, proto má uváděná ztráta i Dice koeficient extrémní variabilitu.

10.3 Výstupní fúze

S naprostou totožným přístupem byly natrénované modely, na jejichž vstupu byla pouze jedna ze vstupních modalit. Tedy byl natrénován model *FLAIR* a *DWI* s úplně shodnou architekturou založenou na 3D U-Net vygenerovanou na základě otisku vstupního datasetu, který tentokrát obsahoval buď pouze FLAIR nebo DWI sekvence. Analogicky byly natrénovány modely *FLAIRMNI* a *DWIMNI*, kde byly snímky nejprve koregistrovány do prostoru MNI.

Výstupní fúze je realizována stejnou funkcí a stejným skriptem jako v předchozí kapitole, tj. pomocí průměrování pravděpodobnostní a prahováním fixní hodnotou 0,5. Takto fúzované pravděpodobnostní mapy jsou uložené ve složkách *ensemble_FLAIR_DWI* a *ensemble_FLAIR_DWI_MNI* pro fúzované mapy z modelů *FLAIR* a *DWI*, resp. *FLAIRMNI* a *DWIMNI*. Kromě toho byly také fúzovány tři modely pomocí průměrování výstupů modelů *Strokes*, *FLAIR* a *DWI*. Tyto segmentace jsou uloženy ve složce *ensemble_Strokes_FLAIR_DWI*. Analogicky byly fúzovány tři modely natrénované na snímcích v MNI do složky *ensemble_Strokes_FLAIR_DWI_MNI*.

Výsledky dílčích unimodálních instancí a výsledky po fúzování jsou zobrazeny v tabulce 10.2. Z tabulky je patrné že nejlepších výsledků lze dosáhnout výstupní fúzí modelů *FLAIR* a *DWI*. Výstup této fúze při evaluaci dosáhl Dice koeficientu až $0,646 \pm 0,238$ - to platí pro instanci jež zpracovávala snímky koregistrované do prostoru MNI *ensemble_FLAIR_DWI_MNI*. Kvalitativní rozbor nejlepšího modelu *ensemble_FLAIR_DWI_MNI* je zobrazen v příloze C.

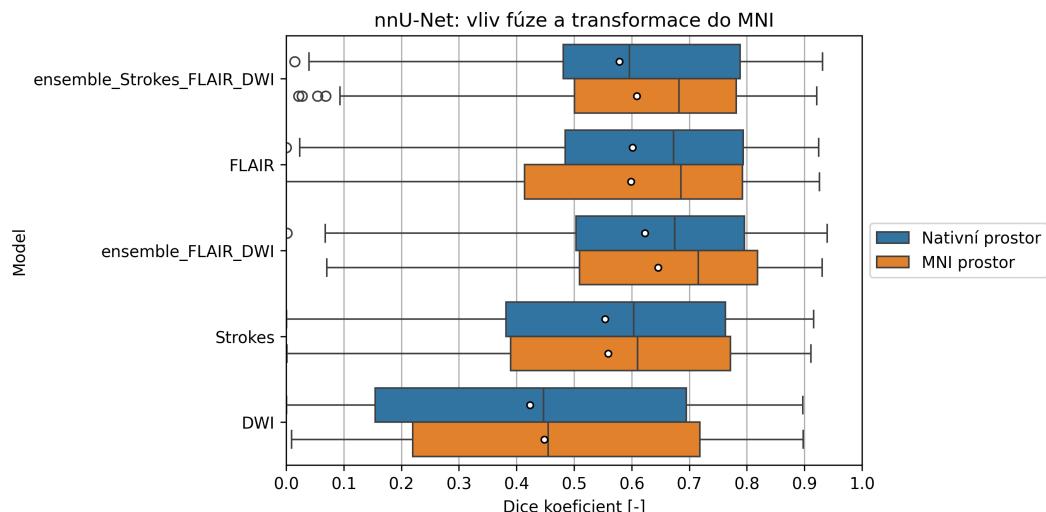
I v tomto případě platí, že unimodální modely natrénované na sekvenci DWI dosahují velmi nízké precisnosti a naopak modely natrénované pouze na FLAIR dosahují vysoké precisnosti, ale poměrně nízké sensitivity. Dále je poměrně zajímavé, že model *ensemble_Strokes_FLAIR_DWI* dosahuje poměrně nízkého Dice koeficientu oproti jeho obdobě v prostoru MNI *ensemble_Strokes_FLAIR_DWI_MNI*. Přičemž všechny

Model	DC	Prec	Sens	Spec	F
StrokesMNI	0,559±0,265	0,650	0,609	0,991	5
Strokes	0,554±0,262	0,675	0,563	0,993	4
ens._FLAIR_DWI_MNI	0,646±0,238	0,783	0,631	0,995	1
ens._FLAIR_DWI	0,623±0,241	0,816	0,577	0,996	3
ens._Str._FLAIR_DWI_MNI	0,609±0,253	0,676	0,668	0,991	5
FLAIR	0,601±0,255	0,752	0,574	0,995	3
FLAIRMNI	0,599±0,259	0,714	0,598	0,993	3
ens._Str._FLAIR_DWI	0,579±0,262	0,699	0,585	0,993	5
DWIMNI	0,448±0,288	0,407	0,746	0,941	10
DWI	0,423±0,293	0,392	0,627	0,946	13

Tabulka 10.2. Výsledky výstupní fúze metody nnU-Net ve formátu průměr ± směr. odchylka. Fúzované predikce (předpona ens.) vznikly kombinací pravděpodobnostních map modelů *Strokes*, *FLAIR* a *DWI* a analogicky pro instance natrénované v prostoru MNI. Validace proběhla na datasetu Motol proti expertním segmentacím v jejich nativním prostoru. DC: Dice koeficient. Prec: Precisnost. Sens: Sensitivita. Spec: Specificita. F: Počet neúspěšných segmentacích (DC < 0,1).

ostatní dvojice Dice koeficientů jsou si velmi blízko (*Strokes* a *Strokes_MNI*, *FLAIR* a *FLAIRMNI*, ...), přesto výstupní fúze všech tří modelů mimo prostor MNI nevykazuje tak kvalitní výsledky jako fúze v MNI prostoru.

Grafické porovnání Dice koeficientů během evaluace je na obrázku 10.2.



Obrázek 10.2. Výsledky metody nnU-Net, modely s předponou ensemble označují výsledky výstupní fúze. Validace proběhla na datasetu Motol proti expertním segmentacím v jejich nativním prostoru. Box označuje interkvartilové rozpětí (IQR), vousy 1,5 IQR, čára v boxu označuje medián a kolečko průměr.

10.4 Křížová validace

V případě metody nnU-Net [53] bylo dodatečně provedeno trénování metody pouze na datasetu Motol. Preprocessing byl proveden naprosto shodným postupem jako v

předchozích případech. Tentokrát však byla realizována křížová validace z důvodu nedostatku trénovacích dat. Modality DWI a FLAIR byly fúzovány vstupně.

Bыло vytvořeno pět různých posloupností snímků datasetu Motol, jež v každé z nich zhruba pětina snímků odpovídala validačnímu datasetu dané posloupnosti. Konkrétní snímek se mohl vyskytnout ve validační množině pouze jedenkrát. Data byla náhodně rozdělena, ale vždy tak, aby se všechny tři sběry daného pacienta vyskytovaly všechny buď v trénovací nebo ve validační množině dané posloupnosti. Konkrétní rozdělení je uloženo v souboru *results/metadata_Motol.ods*.

Bylo natrénováno pět různých instancí metody nnU-Net [53] v klasickém prostoru a pět různých instancí natrénovaných v prostoru MNI. Průměrné výsledky daných pěti modelů jsou v tabulce 10.3. Oproti modelům nnU-Net [53] natrénovaným na datech z ISLES zde dosahuje vyššího DC model, jež není natrénován na datech v MNI.

Model	DC	Prec	Sens	Spec	F
MotolOnly	0,745±0,182	0,804	0,763	0,996	0
MotolOnlyMNI	0,724±0,212	0,763	0,783	0,995	0

Tabulka 10.3. Průměrné výsledky metody nnU-Net ve formátu průměr \pm směr. odchylka. Modely byly natrénovány na části datasetu Motol a modely byly křížově validovány v nativním prostoru datasetu Motol. DC: Dice koeficient. Prec: Precisnost. Sens: Sensitivita. Spec: Specificita. F: Počet neúspěšných segmentacích ($DC < 0,1$).

Kvalitativní rozbor predikcí všech pěti instancí natrénovaných v rámci *MotolOnly* je zobrazen v příloze D.

Vzhledem k excelentním výsledkům byl model vytvořený křížovou validací otestován ještě na datech ISLES 2015 [2] a ISLES 2022 [40]. Opět proběhla predikce všech pěti modelů *MotolOnly* a proběhla výstupní fúze pěti pravděpodobnostních map průměrováním. Bylo dosaženo průměrného Dice koeficientu $0,394 \pm 0,332$, precisnosti 0,387, sensitivity 0,514 a specificity 0,990. Predikce selhala ($DC < 0,1$) u 84 snímků z ISLES 2022 [40] a u 7 snímků z ISLES 2015 [2]. Dle očekávání se ukázalo, že *MotolOnly* pracuje na ISLES velmi špatně, jelikož vychází z velmi konzistentních dat Motol (na rozdíl od variability snímků v ISLES).

Kapitola 11

Experimenty s architekturou 3D U-Net

Vzhledem k filosofii nnU-Net [53] a vzhledem k provedené rešerši jsem se rozhodl pro vytvoření a ověření vlastní implementace 3D U-Net [63] vycházející z článku [63]. Vlastní implementace dovolí snadnou modifikaci neuronové sítě a tím pádem umožní snadné ověřování metod fúzí nebo jiných parametrů trénování.

11.1 Popis metody

Architektura 3D U-Net [63] pochází z roku 2016 a je na ní založena drtivá část metod dosahující kvalitních segmentací medicínských dat.

Architektura pracuje na 4 stupních - tedy enkodér a dekadér zpracovává informace ve 4 rozlišeních. Každý stupeň obsahuje posloupnost dvou konvolucí s jádrem (3, 3, 3). Za každou konvolucí následuje BatchNorm a aktivační funkce ReLu. Nakonec je v části enkoderu stupeň zakončen MaxPoolingovou vrstvou s krokem (2, 2, 2). Na druhou stranu v dekadéru se místo MaxPoolingu uplatňují transponované konvoluce s jádrem (2, 2, 2) a krokem také (2, 2, 2). Jinak blok dekadéru zůstává stejný jako u enkoderu. Klíčovou roli v architektuře také hrají residiální spojení realizovaná spojením tensorů. Tato spojení jsou v rámci jednoho stupně a převádí informaci z enkodéru do dekadéru. Kdyby tato spojení v síti nebyla, dekadér by přišel o prostorovou informaci, která se s použitím MaxPoolingu vytrácí. Před vytvořením výstupu se nachází ještě jedna konvoluční vrstva (navíc oproti klasickým blokům) s jednotkovým kernelem a počtem požadovaných výstupních kanálů. Jako výstupní aktivace je použitý softmax a ztrátová funkce je vážená křížová entropie. Trénování autory proběhlo s jednotkovou velikostí batch. Nákres architektury je na obrázku 11.1.

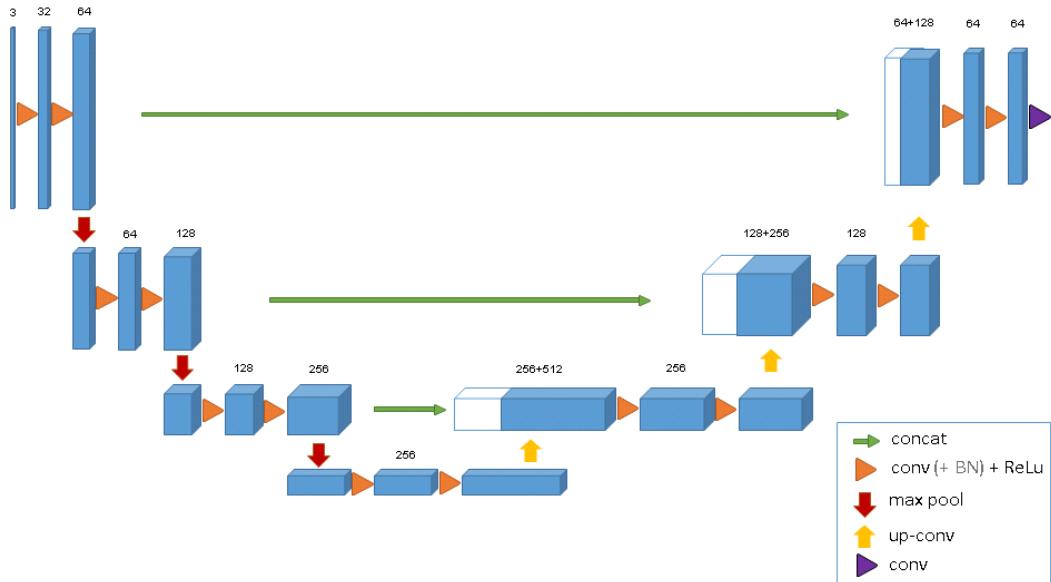
Autor metodu navrhl pro vstupní tensor o velikosti (132, 132, 116) se třemi vstupními kanály a třemi výstupními kanály. Vzhledem k tomu, že původní implementace nepoužívá padding ve vrstvách, predikce odpovídá velikosti pouze (44, 44, 28) a během residiálního spojení se zároveň provádí příslušné oříznutí spojovaného tensoru. Sít byla vyvinuta během segmentování embryí drápatky vodní z dat z mikroskopu. Velikost voxelu proto byla pouze (1,76; 1,76; 2,04) μm .

Takto navržený model autoři trénovali na grafické kartě NVIDIA TitanX tři dny.

11.2 Implementace a ověření 3D U-Net

Soubor s definicí modelů a moduly pro preprocessing, trénování a predikce jsou ve složce *3dunet*. Vzhledem k tomu, že jde o vlastní implementaci, pro tuto metodu není v kořeni repozitáře složka s příponou *_workspace* a veškerá pracovní data (preprocessing, váhy a predikce) jsou uložena přímo ve složce se zdrojovým kódem implementace.

Před začátkem experimentování byly uloženy do příslušné složky soubory po preprocessingu. Ten byl totožný s preprocessingem dat pro metodu nnU-Net [53], lišila se pouze struktura uložených dat po zpracování. Identifikace modalit pomocí čísla byla nahrazena jménem modality.



Obrázek 11.1. Původní architektura 3D U-Net se třemi vstupními a třemi výstupními kanály. [63]

Pro implementaci 3D U-Net [63] jsem použil knihovnu PyTorch a pro načítání dat po preprocessingu a augmentaci jsem použil knihovnu TorchIO [74].

Nejprve byla implementována architektura 3D U-Net [63] odpovídající architektuře z článku [63], byl však využit padding, aby nedošlo k významné ztrátě segmentovaného prostoru. Tím pádem není nutné používat ořezání tenzoru v residuálním spojení. Pro tuto implementaci byla nalezena maximální velikost tensoru a velikost batch, který je možno společně s modelem nahrát na GPU výpočetního serveru. Bylo dosaženo závěru, že maximální velikost patche obsahující dva kanály je (96, 96, 96) a pro danou velikost patche je maximální velikost batch 2. Pro trénování všech modelů byl použit optimalizátor Adamax a trénování probíhalo vždy 20 epoch. Jedna epocha je definována načtením všech vstupních skenů s tím, že z každého vstupního skenu se po načtení extrahuje 10 patchů, ze kterých probíhá trénování. Vzorkování dat probíhalo na pozadí pomocí subprocesů a díky tomu nezvětšovalo dobu trénování. Jako ztrátová funkce byla použita binární křížová entropie s integrovanou sigmoidou (BCEWithLogitsLoss), která vykazuje dobrou numerickou stabilitu. Při každém načtení vzorku je ztrátová funkce vážena poměrem pozitivních segmentací obsažených v celém batch.

Před pořízením vzorků byl na každý snímek aplikován padding o poloviční velikosti, než je velikost patch z toho důvodu, aby bylo možné centrovat vzorkovaný patch i na okraje skenu. Po paddingu proběhla normalizace skenu na nulový průměr a jednotkovou směr. odchylku.

Predikce jsou počítány aplikací stejných transformací (tj. padding a normalizace). Na výstupu sítě při predikcích je zapojena sigmoida jejíž výstup může být uložen (tj. pravděpodobnostní mapa modelu) nebo jsou hodnoty prahovány hodnotou 0,5 pro získání segmentace.

Nakonec pomocí vstupní fúze byly empiricky hledány vhodné parametry vzorkování dat, vhodná ztrátová funkce, vhodná velikost modelu, atd. Po vyhodnocení vlivu těchto parametrů byla ověřena i výstupní fúze a po modifikaci modelu byla ověřena in-layer fúze dat. K trénování byly použity skeny z datasetů ISLES 2015 [2] a 2022 [40], evaluace probíhala v nativním prostoru datasetu Motol.

■ 11.2.1 Vliv vzorkování

Ze všeho nejdříve byly ověřeny metody vzorkování se vstupní fúzí. TorchIO poskytuje framework, který dokáže z načteného souboru vzorkovat požadované patche. Přístup k výběru vzorku se však zásadně liší. Patch je možné vzorkovat z prostoru s uniformní pravděpodobností; je možné je vzorkovat z celého objemu skenu, poté, co je objem rozdělen na požadované velikosti; je možné vzorkovat pomocí pravděpodobnostní mapy; nebo je možné vzorkovat data tak, aby bylo zaručeno, že patch obsahuje vždy alespoň jeden pozitivně segmentovaný voxel.

Segmentace z celého objemu se uplatňuje primárně při validaci, resp. při vytváření predikcí, kdy není determinován prázdný prostor kolem pacienta a je třeba, aby nebyl vynechán žádný voxel. Vzorkování s uniformní pravděpodobností zase extrémně zkomplikuje trénování a neuronová síť nedosáhne požadované sensitivity, jelikož bude přeúčená na tom, že se v trénovacích datech téměř pozitivní voxely nevyskytují (jelikož poměr pozitivních voxelů vůči zdravé tkáni je extrémně malý).

Z výše uvedeného tedy vyplývá, že kvalitní výsledky lze dosáhnout buď vzorkováním tak, že centrální voxel vždy obsahuje expertní segmentaci a tím je zaručeno, že v každém patch je alespoň jeden expertem segmentovaný voxel. Tuto metodu vzorkování používá model s názvem *LabelSampler*. Dále byl ověřen model *WeightedSampler1*, jež vzorkoval data tak, aby pravděpodobnost třídy centrálního voxelu patche byla stejná pro zdravou tkáň a pro tkáň postiženou CMP pro jeden sken (pravděpodobnost výběru centrálního voxelu mimo prostor mozku byla nulová). Obdobně byl natrénován model *WeightedSampler2*, u nějž vzorkování probíhalo tím způsobem, že prostor mozku byl vzorkován uniformní pravděpodobností, zatímco prostor mimo mozek nebyl vzorkován vůbec. V případě modelu *LabelSampler* a *WeightedSampler1* nebyly využity tři skeny z datasetu ISLES 2022 [40], které neobsahovaly expertní segmentace.

Vstupní fúze tedy byla realizována sloučením tensorů modalit podél osy kanálů a tím pádem vstup do sítě byl o rozměrech (2, 2, 96, 96, 96). Predikce byly vytvořeny pomocí osminového překrytí patche a překrývající se predikce byly oříznuty. Později se ukázalo, že toto nastavení není nejhodnější, jelikož na hranách vzorkování se na pravděpodobnostní mapě zobrazují artefakty.

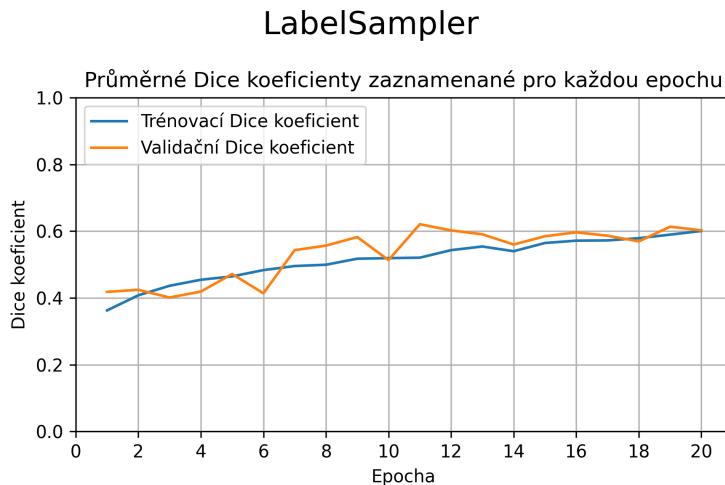
Trénování modelů trvalo zhruba 12 hodin a výsledky evaluace na datasetu Motol jsou v tabulce 11.1. Z tabulky vyplývá, že z hlediska DC funguje nejlépe konfigurace, u níž je zaručeno, že v každém patch spadá centrální voxel do expertní segmentace. Další dvě konfigurace podle analýzy segmentují více tkáně, než by bylo nezbytné. Proto pro další vývoj bylo zvoleno vzorkování použité v modelu *LabelSampler*.

Model	DC	Prec	Sens	Spec	F
LabelSampler	0,453±0,281	0,447	0,645	0,965	10
WeightedSampler2	0,442±0,269	0,373	0,753	0,955	7
WeightedSampler1	0,419±0,265	0,332	0,806	0,942	9

Tabulka 11.1. Výsledky vstupní fúze metody 3D U-Net se zaměřením na vliv vzorkování. Uvedeno ve formátu průměr ± směr. odchylnka. Modely byly natrénovány na sloučení datasetů ISLES 2015 a ISLES 2022. Validace proběhla na datasetu Motol proti expertním segmentacím v jejich nativním prostoru. DC: Dice koeficient. Prec: Precisnost. Sens: Sensitivity. Spec: Specificita. F: Počet neúspěšných segmentacích ($DC < 0,1$).

Z průběhu trénování navíc vyplývá, že průměrný Dice koeficient v rámci poslední epochy vypočítaný na trénovacím datasetu je velmi nízký pro *WeightedSampler1* (DC 0,43) i *WeightedSampler2* (DC 0,30). Oproti modelu *LabelSampler*, kde DC dosáhl 0,60.

Průběh trénování je modelu *LabelSampler* je na obrázku 11.2. Validační Dice koeficient je vypočítán jako průměr DC napříč všemi skeny datasetu Motol, z nichž byl extrahován náhodně pouze jediný patch a DC je vypočítán v prostoru trénovacích dat (po podvzorkování).



Obrázek 11.2. Výsledky modelu *LabelSampler* v závislosti na epoše. Jedná o průměrný trénovací Dice koeficient a částečný validační Dice koeficient.

■ 11.2.2 Vliv ztrátové funkce

Po zafixování metody vzorkování byl dále ověřen vliv použití ztrátové funkce. Literatura se spíše přiklání k používání kombinace Dice koeficientu a křížové entropie než k používání pouze křížové entropie. Rozhodně ale nelze vyvodit jednoznačný závěr, která z variant funguje lépe. To bylo motivací k empirickému ověření v rámci mé implementace.

Kromě modelu s výchozí ztrátovou funkcí binární křížové entropie (*LabelSampler*) byly natrénovaný modely uplatňující:

- Součet Dice koeficientu a binární křížové entropie (*LabelSampler_Dice+BCE*).
- Součet logaritmu Dice koeficientu a binární křížové entropie (*LabelSampler_logDice+BCE*).

Výsledky evaluace natrénovaných modelů jsou uvedeny v tabulce 11.2. Z výsledků vyplývá, že nejlepších výsledků lze v mé implementaci dosáhnout při použití pouze binární křížové entropie jako ztrátové funkce.

■ 11.2.3 Vliv velikosti modelu

Dalším krokem, který mohl přinést lepší výsledky bylo ladění velikosti modelu. Proto oproti doposud nejlepší instanci modelu *LabelSampler* byla ověřena jak rozšířená architektura, tak i naopak zmenšená architektura. Tyto konfigurace 3D U-Net [63] se lišily v počtu filtrů v každém stupni U-Net po provedených konvolucích. Proto konfigurace s názvem *LargeUNet3D* obsahovala následující počty filtrů: [64+32, 128+64, 256+128] a 512+256 pro hrdlo. Menší konfigurace proti původnímu 3D U-Net [63] *TinyUNet3D* obsahovala počty filtrů [64-32, 128-64, 256-128] v jednotlivých stupních a 512-256 filtrů v hrdle.

Model	DC	Prec	Sens	Spec	F
LabelSampler	0,453±0,281	0,447	0,645	0,965	10
LabelSampler_Dice+BCE	0,426±0,256	0,356	0,737	0,958	9
LabelSampler_logDice+BCE	0,392±0,254	0,326	0,732	0,948	10

Tabulka 11.2. Výsledky vstupní fúze metody 3D U-Net se zaměřením na vliv ztrátové funkce. Uvedeno ve formátu průměr ± směr. odchylka. Modely byly natrénovány na sloučení datasetů ISLES 2015 a ISLES 2022. Validace proběhla na datasetu Motol proti expertním segmentacím v jejich nativním prostoru. DC: Dice koeficient. Prec: Precisnost. Sens: Sensitivita. Spec: Specificita. F: Počet neúspěšných segmentacích (DC < 0,1).

Doba trénování pro větší síť trvala dvojnásobek oproti výchozí *LabelSampler* a musel být použitý jednotkový batch. Menší síť se trénovala dvakrát rychleji než *LabelSampler*.

Výsledky jsou uvedeny v tabulce 11.3. Konvergence trénovacího Dice koeficientu proběhla pro všechny tři případy velmi podobně, tudíž všechny tři sítě by měly být zhruba stejně natrénovány. Přestože byl použit menší model, nedošlo u něj k lepší generalizaci a výsledky evaluace nejsou lepší než u konfigurace *LabelSampler*. Stejně tak to platí pro U-Net s více filtry než standardně.

Model	DC	Prec	Sens	Spec	F
LabelSampler	0,453±0,281	0,447	0,645	0,965	10
TinyUNet3D	0,407±0,270	0,385	0,617	0,955	10
LargeUNet3D	0,388±0,240	0,375	0,590	0,969	8

Tabulka 11.3. Výsledky vstupní fúze metody 3D U-Net se zaměřením na různé počty filtrů. Uvedeno ve formátu průměr ± směr. odchylka. Modely byly natrénovány na sloučení datasetů ISLES 2015 a ISLES 2022. Validace proběhla na datasetu Motol proti expertním segmentacím v jejich nativním prostoru. DC: Dice koeficient. Prec: Precisnost. Sens: Sensitivita. Spec: Specificita. F: Počet neúspěšných segmentacích (DC < 0,1).

■ 11.2.4 Vliv augmentace

Nakonec byl posuzován vliv augmentace, díky které by mohlo dojít ke zlepšení zobecňování modelu *LabelSampler*. Při načtení skenu se po paddingu a normalizaci aplikovaly funkce augmentace v pořadí náhodná afinní funkce, Gaussovské rozostření a Gaussovský šum. Byly natrénovány a ověřeny dvě konfigurace s různou intenzitou těchto funkcí. První z nich *LabelSampler_augmentation1* umožňovala větší úpravy původního snímku (např. rotaci až o 90 st.). Druhá konfigurace *LabelSampler_augmentation2* aplikovala velmi malý šum, malé rozostření, takže nedošlo k velkému zkreslení. Zároveň bylo umožněno pouze malé škálování afinní funkcí a rotace o max. 45 st. Augmentace byly provedeny na pozadí v průběhu trénování, proto neprodloužily čas učení sítě.

Výsledky trénování jsou zobrazeny v tabulce 11.4. Je patrné, že poškození informace augmentací v konfiguraci *LabelSampler_augmentation1* je extrémní. Přestože člověk stále dokáže nalézt expertem segmentovaná ložiska, neuronová síť se nedokáže efektivně naučit pravidla pro jejich lokalizaci. Druhá konfigurace *LabelSampler_augmentation2* s mírnější augmentací sice neselhává, ale její Dice koeficient nepřekoná původní nastavení *LabelSampler*. Nicméně je zajímavé, že konfigurace *LabelSampler_augmentation2* dosahuje velmi malého počtu pacientů, u kterých instance modelu selhala (sloupec F v tabulce 11.4). Přesto nejlepších výsledků dosahuje původní model *LabelSampler* bez augmentace.

Model	DC	Prec	Sens	Spec	F
LabelSampler	0,453±0,281	0,447	0,645	0,965	10
LabelSampler_augm.2	0,400±0,228	0,420	0,546	0,975	5
LabelSampler_augm.1	0,271±0,228	0,194	0,807	0,887	14

Tabulka 11.4. Výsledky vstupní fúze metody 3D U-Net se zaměřením na augmentaci. Uvedeno ve formátu průměr ± směr. odchylka. Modely byly natrénovány na sloučení datasetů ISLES 2015 a ISLES 2022. Validace proběhla na datasetu Motol proti expertním segmentacím v jejich nativním prostoru. DC: Dice koeficient. Prec: Precisnost. Sens: Sensitivita. Spec: Specificita. F: Počet neúspěšných segmentacích ($DC < 0,1$).

■ 11.2.5 Vliv výstupní fúze

Přes modifikace architektury 3D U-Net [63], modifikace ztrátové funkce, augmentace a vzorkování byla nalezena nejlepší konfigurace vstupní fúze - tedy model s označením *LabelSampler*. Do modelu vstupovaly modality sloučené přes dimenze kanálů a instance modelu tak dosáhla během evaluace DC $0,453\pm0,281$.

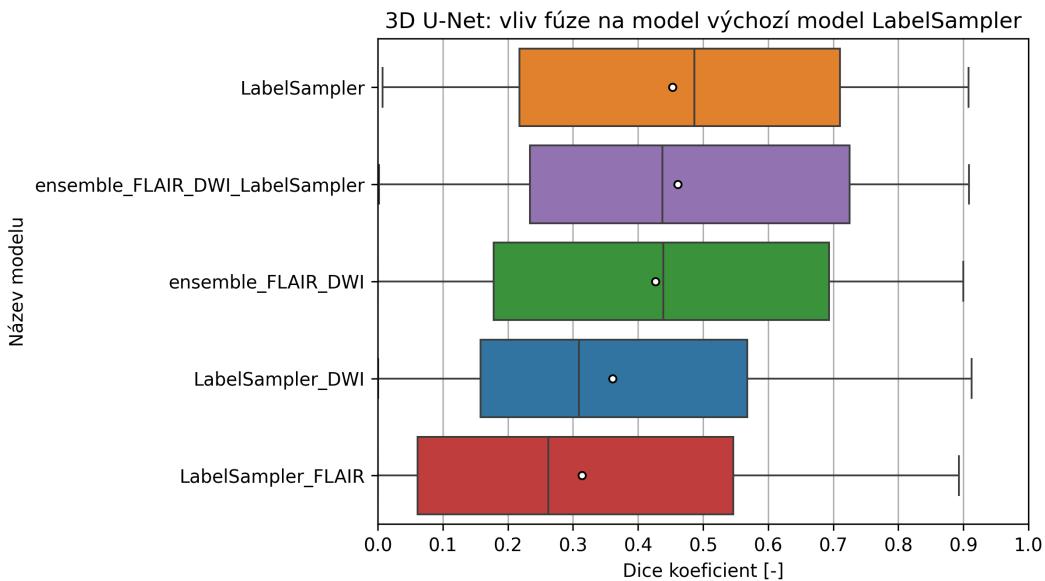
Konfigurace *LabelSampler* proto sloužila jako výchozí nastavení pro sledování vlivu výstupní fúze. Byly natrénovány dva odvozené modely s názvy *LabelSampler_DWI* a *LabelSampler_FLAIR*, kde vstupem do neuronové sítě byl pouze jediný kanál obsahující modalitu s příslušným názvem, jinak se nastavení trénování vůbec nelišilo. Výstup z poslední aktivační funkce sigmoidy byl během predikcí uložen a tyto pravděpodobnostní mapy byly fúzovány pomocí průměrování pravděpodobnosti, stejně jako u předchozích metod. Tímto způsobem vznikla predikce označená jako *ensemble_FLAIR_DWI*, která vznikla fúzí pravděpodobnostních map z *LabelSampler_DWI* a *LabelSampler_FLAIR*. A také predikce s označením *ensemble_FLAIR_DWI_LabelSampler*, která je tvořena fúzí pravděpodobnostních map *LabelSampler_DWI*, *LabelSampler_FLAIR* a *LabelSampler*.

Kvantitativní výsledky evaluace jsou v tabulce 11.5. Z pohledu DC je tedy nejlepší *ensemble_FLAIR_DWI_LabelSampler*, jež kombinuje pravděpodobnostní mapy všech tří modelů (dva z nich zpracovávají unimodální informaci a třetí uplatňuje samostatně vstupní fúzi). Tímto způsobem bylo dosaženo až Dice koeficientu $0,461\pm0,288$.

Model	DC	Prec	Sens	Spec	F
LabelSampler	0,453±0,281	0,447	0,645	0,965	10
ens._FLAIR_DWI_LabelSam.	0,461±0,288	0,511	0,560	0,977	11
ens._FLAIR_DWI	0,427±0,285	0,751	0,352	0,998	9
LabelSampler_DWI	0,361±0,260	0,318	0,591	0,955	12
LabelSampler_FLAIR	0,314±0,269	0,679	0,241	0,998	18

Tabulka 11.5. Výsledky výstupní fúze metody 3D U-Net ve formátu průměr ± směr. odchylka. Fúzované predikce (předpona ens) vznikly kombinací pravděpodobnostních map modelů *LabelSampler_DWI*, *LabelSampler_FLAIR* nebo *LabelSampler*. Validace proběhla na datasetu Motol proti expertním segmentacím v jejich nativním prostoru. DC: Dice koeficient. Prec: Precisnost. Sens: Sensitivita. Spec: Specificita. F: Počet neúspěšných segmentacích ($DC < 0,1$).

Rozložení Dice koeficientu napříč evaluačním datasetem Motol je patrné z obrázku 11.3. Kvalitativní rozbor predikcí nejlepšího modelu *ensemble_FLAIR_DWI_LabelSampler* v rámci architektury 3D U-Net [63] je zobrazen v příloze E.



Obrázek 11.3. Výsledky výstupní fúze v porovnání s nejlepším modelem se vstupní fúzí *LabelSampler*. Validace proběhla na datasetu Motol proti expertním segmentacím v jejich nativním prostoru. Box označuje interkvartilové rozpětí (IQR), vousy 1,5 IQR, čára v boxu označuje medián a kolečko průměr.

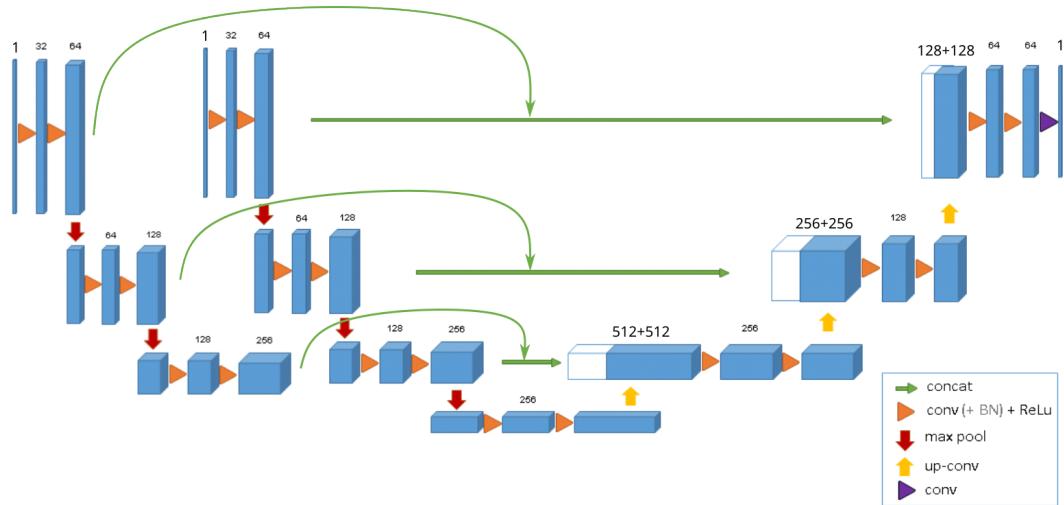
■ 11.2.6 Vliv in-layer fúze

Jako poslední test byl ověřen vliv in-layer fúze. K tomu je nutná modifikace modelu 3D U-Net [63]. Rozhodl jsem se jít cestou dvojitého enkodéru podobně, jako je uvedeno v článku [75], který 2D segmentací vytváří mapy z dálkového průzkumu Země. Cílem je analýza odlesňování za pomocí modalit, které jsou dva snímky posunuté v čase. Dle autorů při průchodu těchto snímků nezávisle dvěma enkodéry došlo ke zlepšení přesnosti sítě o 0,0023 % oproti standardní architektuře U-Net.

Nový model s označením *TwoHeadedUNet3D* tedy definuje dva enkodéry, jež na třech stupních obsahují klasické 3D U-Net [63] enkodérové bloky, jejichž výstup je před předáním do dekodéru sloučen. V hrdle sítě už ke sloučení nedochází, proto lze považovat za jednu z modalit za hlavní, jelikož architektura není symetrická. Bloky dekodérů jsou nezměněné. Obrázek modifikované architektury je na obrázku 11.4.

Díky asymetrii sítě byly natrénovány dvě verze této architektury. První z modelů je označen jako *TwoHeadedUNet3D_DWI*. U něj je hlavní modalita DWI, která vstupuje do čtyř stupňového enkodéru s hrdlem (na obrázku 11.4 enkodér vpravo) a modalita FLAIR vstupuje do třístupňového enkodéru. Analogicky byl natrénován model *TwoHeadedUNet3D_FLAIR*, u nějž je hlavní modalita FLAIR, která vstupuje do čtyřstupňového enkodéru s hrdlem a DWI vstupuje do třístupňového enkodéru.

Výsledky natrénovaných modelů jsou v tabulce 11.6. Ani jedna z instancí modelu během evaluace na datasetu Motol nedosáhla lepších výsledků než model uplatňující čistě vstupní nebo výstupní fúzi. Model *TwoHeadedUNet3D_FLAIR* dosáhl ale srovnatelných výsledků a zároveň má nejmenší počet pacientů, u kterých model selhal (sloupec F) v rámci všech instancí metody 3D U-Net [63]. Instance modelu *TwoHeadedUNet3D_DWI*, jež hlavní modalitou bylo DWI, dosáhla velmi špatných výsledků.



Obrázek 11.4. Modifikace 3D U-Net nazvaná TwoHeadedUNet3D uplatňující in-layer fúzi informace ve třech stupních enkodéru.

Model	DC	Prec	Sens	Spec	F
TwoHeadedUNet3D_FLAIR	0,446±0,251	0,662	0,396	0,997	4
TwoHeadedUNet3D_DWI	0,254±0,190	0,835	0,174	1.000	15

Tabulka 11.6. Výsledky in-layer fúze metody 3D U-Net ve formátu průměr ± směr. odchylka. Validace proběhla na datasetu Motol proti expertním segmentacím v jejich nativním prostoru. DC: Dice koeficient. Prec: Precisnost. Sens: Sensitivita. Spec: Specificita. F: Počet neúspěšných segmentacích (DC < 0,1).

Kapitola 12

Diskuse

V této kapitole budou shrnutý dosažené výsledky, bude uvedeno srovnání vlivu koregistence do prostoru MNI, srovnání vlivu fúzí, velikosti léze a TSS na úspěšnost segmentace, srovnání s provedenou rešerší a nakonec budou prezentovány návrhy na zlepšení dosahovaných výsledků z pohledu Dice koeficientu.

12.1 Dosažené výsledky

Experimentálně byly ověřeny dvě již existující metody. Během trénování na datech ze soutěží ISLES 2015 [2] a 2022 [40] a validaci na datasetu Motol dosáhla první testovaná metoda DeepMedic [50–51] Dice koeficientu až $0,510 \pm 0,259$ (označeno jako *ensemble_DWI_FLAIR*). Druhá experimentálně ověřovaná metoda nnU-Net [53] dosáhla při trénování na datech ISLES a validaci na datasetu Motol Dice koeficientu až $0,646 \pm 0,238$ (*ensemble_FLAIR_DWI_MNI*). Poslední vlastní metoda, jež byla v rámci práce implementována vycházela z architektury 3D U-Net [63] a dosáhla Dice koeficientu až $0,461 \pm 0,288$ (*ensemble_FLAIR_DWI_LabelSampler*).

Jelikož metoda nnU-Net [53] dosahovala obecně nejlepších výsledků během trénování na datasetech ISLES a validaci na datasetu Motol, bylo pomocí této metody natréno-váno dalších 5 modelů. Modely byly natrénovány pouze na datasetu Motol a validovány na něm pomocí křížové validace. Průměr DC těchto pěti modelů je $0,745 \pm 0,182$ (*MotolOnly*).

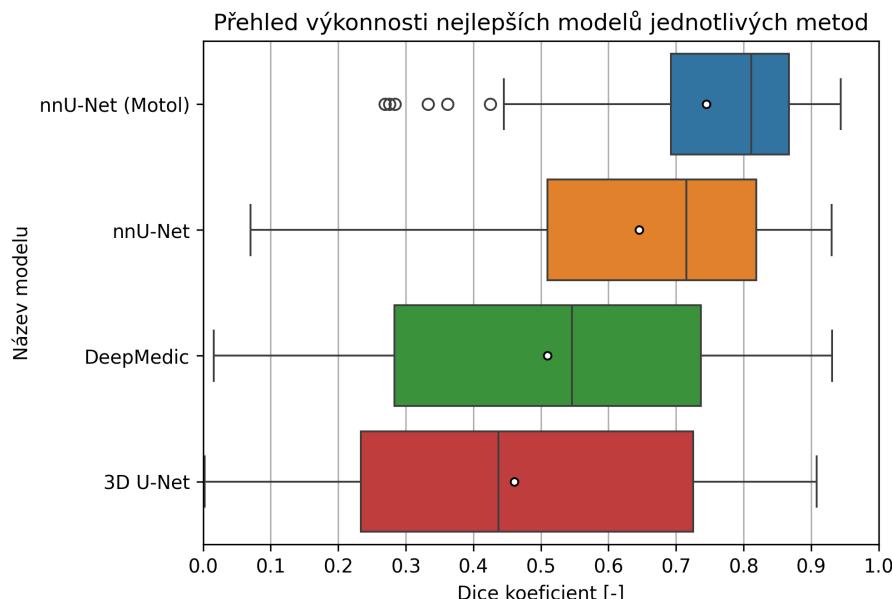
Bohužel pro dataset Motol není k dispozici ani částečná inter-rater analýza segmentací, takže nelze stanovit, jak úspěšný je člověk během segmentace datasetu Motol. Nicméně inter-rater Dice koeficienty pro datasety ISLES 2015 [2] a 2022 [40] byly 0,70, resp. 0,83. Je tedy možné, že v průměru modely nnU-Net [53] narazily na strop, který je definován lidskou přesností, pokud se i inter-rater DC datasetu Motol pohybuje v blízkých hodnotách inter-rater DC ze soutěží ISLES [2, 40].

Je zajímavé, že unimodální modely se vstupní sekvencí FLAIR dosáhly v případě metody DeepMedic [50–51] a nnU-Net [53] lepších Dice koeficientů nežli modely implementující vstupní fúzi FLAIR a DWI. Dále je důležité zmínit pozorování, že unimodální modely natrénované na FLAIR dosahují velmi špatné sensitivity, ale vysoké precisnosti. Na druhou stranu unimodální modely natrénované na sekvenci DWI mají velmi nízkou precisnost a spíše průměrnou sensitivitu dané metody.

Při pohledu na souhrnné kvantitativní výsledky v příloze A je evidentní, že nízké Dice koeficienty jsou dosahovány všemi metodami u stejných skenů či pacientů. Jinými slovy všechny metody selhávají velmi často u těch stejných skenů. Pokud bychom uvažovali, že jsou ověřované metody nezávislé, pak lze tvrdit, že důvod, proč metody selhávají, leží v samotných skenech. Ty se nějakým způsobem odlišují od většinových vzorků. Jako konkrétní příklad lze uvést sken 2179097_Anat_20211220, kde selhávají všechny modely kromě modelu natrénovaného na datech Motol. Proto se lze domnívat, že pro segmentaci tohoto skenu je třeba znalost, která chybí v datasetech ze soutěží ISLES, ale nachází se v ostatních vzorcích datasetu Motol. Dalším příkladem může být sken

2234859_Anat_20220620, kde žádný z natrénovaných modelů nedokázal dosáhnout uspokojivých výsledků.

Statistické rozdělení Dice koeficientů nejlepších vybraných modelů jednotlivých metod je na obrázku 12.1. Co se týče kvalitativního posouzení, při porovnání predikcí s expertní segmentací všechny metody více či méně pro segmentace CMP fungují. Vizuálně lze posoudit kvalitu predikce v přílohách B, C, D a E k jednotlivým metodám. Vlastní implementace architektury 3D U-Net [63] se kvalitou přibližuje k výherní metodě soutěže ISLES 2015 [2] - DeepMedic [50–51]. Excelentních výsledků pak dosahují oba nejlepší modely metody nnU-Net [53]. Dosažené výsledky v této práci v podstatě zdůrazňují hypotézy autorů nnU-Net [53] o tom, jak důležitá je samotná konfigurace trénování a v případě manuálního nastavování parametrů, jak důležitá je zkušenosť samotného programátora.



Obrázek 12.1. Rozložení DC nejlepších modelů testovaných metod při validaci na datasetu Motol. nnU-Net (Motol): Metoda nnU-Net model *MotolOnly*. nnU-Net: Metoda nnU-Net model *ensemble_FLAIR_DWI_MNI*. DeepMedic: Metoda DeepMedic model *ensemble_DWI_FLAIR*. 3D U-Net: Metoda 3D U-Net model *ensemble_FLAIR_DWI_LabelSampler*. Box označuje interkvartilové rozpětí (IQR), vousy 1,5 IQR, čára v boxu označuje medián a kolečko průměr.

Pro reálné nasazení byly vedoucímu práce doručeny ony dva nejlepší natrénované modely metody nnU-Net [53] (*ensemble_FLAIR_DWI_MNI* a *MotolOnly*). První z nich - model uplatňující výstupní fúzi natrénovaný na datech ISLES a validovaný na datasetu Motol by měl velmi dobře generalizovat (díky extrémní rozličnosti použitých datasetů), avšak za cenu nižšího Dice koeficientu. Jinými slovy prezentovaný Dice koeficient by měl být konzistentní i pro nová data, která se mohou poměrně výrazně lišit od jinak vizuálně blízkých vzorků v datasetu Motol. Druhý z nich - model složený z pěti instancí, které byly natrénovány pouze na datasetu Motol, dosahuje sice velmi vysokého Dice koeficientu, avšak takto dobré výsledky lze dosáhnout pouze při vytváření predikcí z dat podobných, jako jsou obsaženy v trénovacím datasetu Motol. Toto úskalí bylo ukázáno i v této práci během ověření výsledků tohoto modelu na datasetu ISLES. Proto se rozhodně nelze spolehnout na tak vysoké DC, jaké jsou v této práci u tohoto modelu prezentovány.

12.2 Vliv koregistrace do prostoru MNI

Z důvodu velmi kvalitních výsledků metody nnU-Net [53], jsem se rozhodl ověřit vliv koregistrace dat do prostoru MNI. Ukázalo se, že v případě, kdy modely jsou natréno-vány na ISLES koregistrovaných datech do MNI a validovány na datech Motol koregis-trovaných do MNI, téměř vždy dojde ke zvýšení Dice koeficientu. V případě, kdy byl natrénován unimodální model na DWI došlo ke zlepšení DC o 0,025, v případě FLAIR došlo k poklesu DC o 0,002. Když byly modely natrénovány vstupní fúzí v prostoru MNI, došlo ke zlepšení o 0,005. U výstupních fúzí došlo ke zlepšení DC o 0,023 a 0,03, což jsou poměrně významná zlepšení.

V případě modelů validovaných pomocí pětinásobné křížové validace v prostoru MNI na datasetu Motol již ke zlepšení Dice koeficientu nedošlo. Predikce v prostoru MNI jsou v průměru o 0,021 Dice koeficientu nižší.

Nelze proto jednoznačně říci, zda trénování a predikce modelů v prostoru MNI fungují u metody nnU-Net [53] obecně lépe nebo hůře. Jelikož u metod hlubokého učení se podílí na dosažených výsledcích částečně náhoda, jsou i výsledky jednotlivých instancí zatíženy náhodou. Pokud by přesto transformace do MNI měla vliv, pak je nutné ji ověřovat u každého datasetu zvláště. Navíc by došlo ke zlepšení pouze o nízká procenta Dice koeficientu, ale za cenu poměrně vysoké výpočetní náročnosti při generování transformací.

12.3 Vliv fúzí

U všech tří ověřovaných metod vycházelo testování ze vstupní fúze, u níž byla data sloučena do jednoho tensoru podél osy kanálů, jelikož z rešerše nebyla známa žádná pravidla, která by šlo před vstupem tensoru do sítě uplatnit. U již implementovaných metod DeepMedic [50–51] a nnU-Net [53] byla ověřena pouze výstupní fúze formou natrénování dvou identických neuronových unimodálních sítí a spojením pravděpodobnostních map pomocí průměrování. V případě metody DeepMedic [50–51] takováto fúze zlepšila DC o 0,083. V případě stejného pokusu pro metodu nnU-Net [53] jde o zlepšení DC o 0,087. Dle Studentova párového t-testu jde v obou případech o signifikantní zlepšení oproti vstupní fúzi ($\alpha=0,01$).

V případě vlastní implementace architektury 3D U-Net [63] byla také ověřena vý-stupní fúze. Nejlepších výsledků bylo dosaženo při průměrování pravděpodobnostních map ze tří predikcí - ze dvou unimodálních modelů a jednoho modelu se vstupní fúzí. Jde tedy o kombinaci vstupní i výstupní fúze. Díky tomu došlo ke zlepšení DC o 0,008. Výsledek však není statisticky významný ($\alpha=0,01$).

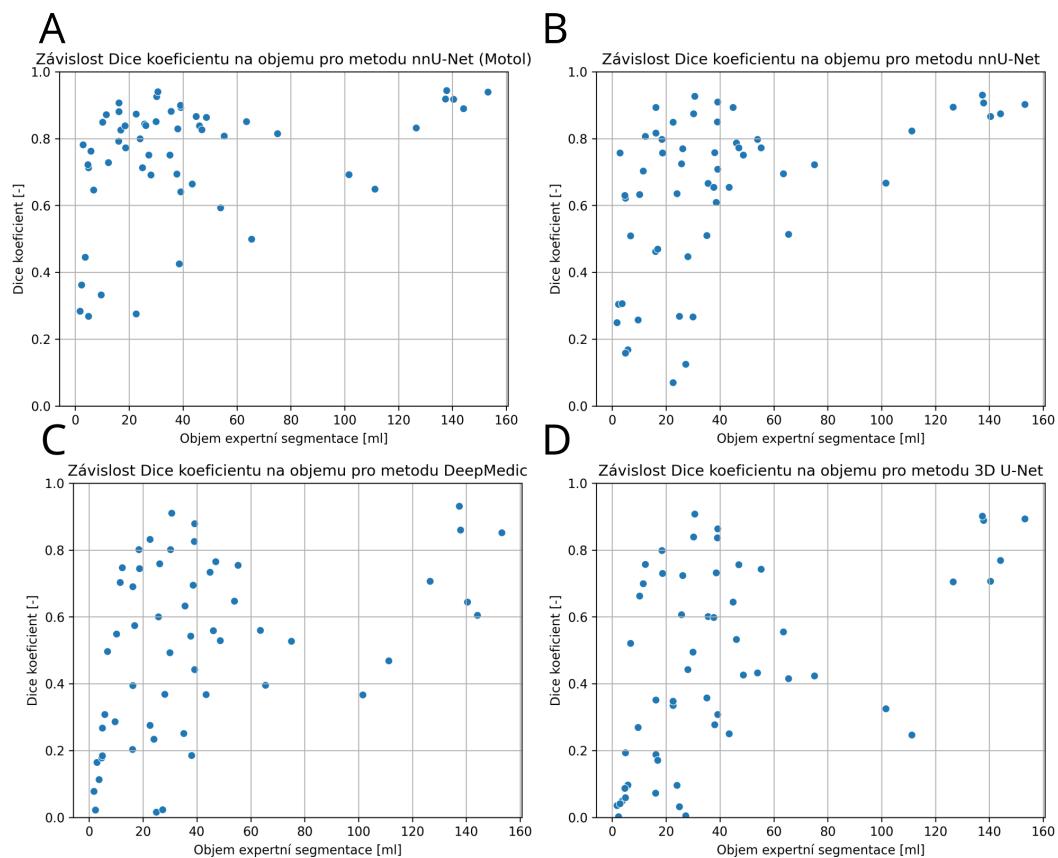
Nakonec byl implementován 3D U-Net [63] se dvěma enkodéry. V tomto případě se jednalo o in-layer fúzi, která slučovala příznakové mapy při prostupování informace enkodéry. Příznaky z obou enkodérů byly dále předány formou residiálního spojení do dekodéru. Při validaci tohoto modelu nebylo dosaženo vyššího DC, než v případě výchozího modelu implementující vstupní fúzi.

12.4 Vliv velikosti léze

Vliv velikosti lézí na Dice koeficientu je demonstrován na obrázku 12.2. Z grafů není patrná jasná závislost. Pro metodu nnU-Net [53] a její dva modely (obr. A a B) platí, že velké léze jsou segmentovány velmi přesně a nadprůměrně, co se týče DC. Při porovnání obrázků A a B lze spatřit pokles Dice koeficientů u vybraných lezí s objemem zhruba

mezi 0 a 40 ml v obou obrázcích. Pro metodu DeepMedic [50–51] (obr. C) a pro vlastní implementaci architektury 3D U-Net [63] (obr. D) je možné říci, že metody selhávají u extrémně malých lézí, naopak pro velké léze segmentují spíše průměrně. Je třeba podotknout, že DC je částečně citlivý na objem léze. Např. dojde-li k mírnému posuvu malé predikce vůči malé expertní segmentaci, má to na DC značný vliv.

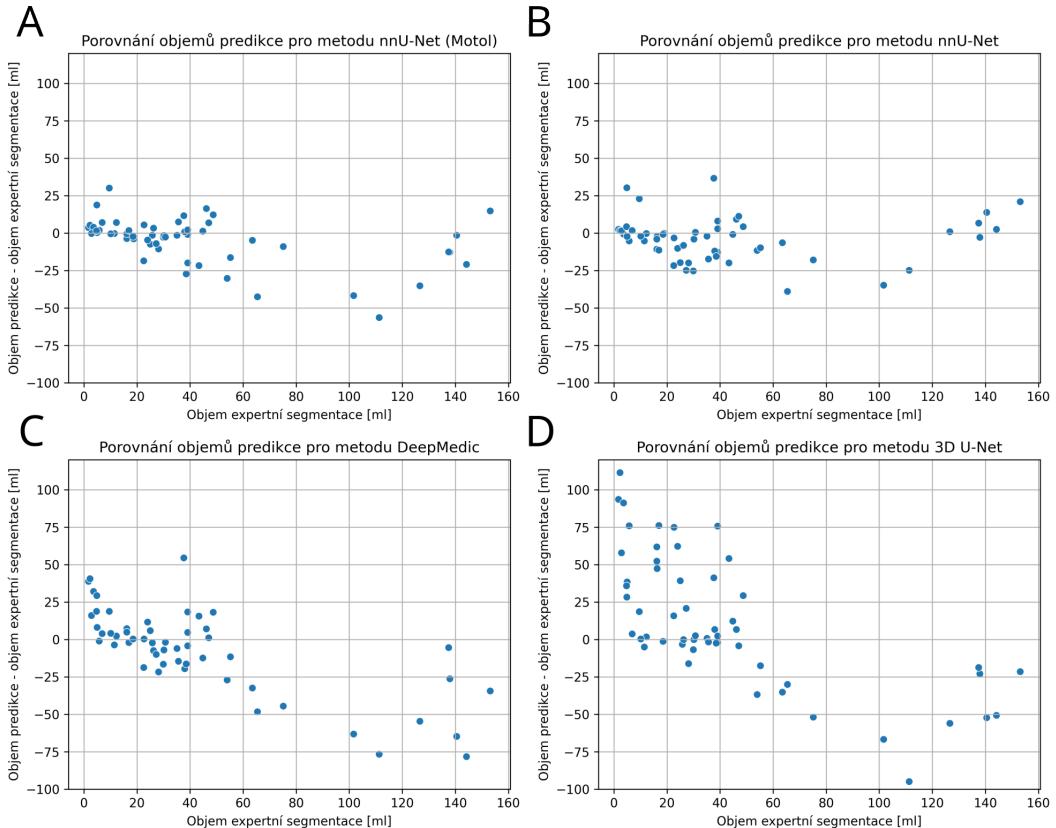
Samotná přesnost odhadu velikosti léze je patrná v obrázku 12.3. Jedná se o graf rozdílu predikovaného a expertního objemu segmentace v závislosti na expertním objemu segmentace. Z grafů vyplývá, že pro větší léze mají modely tendenci spíše jejich objemy podhodnocovat. Extrémně odlehlé hodnoty jsou na obrázku D, kde implementace 3D U-Net [63] až mnohonásobně nadhodnocuje objemy nejménších lézí v datasetu. U ostatních modelů se takovéto jevy nevyskytují. V oblasti větších expertně segmentovaných lézí dochází u všech modelů k větším odchylkám od 0, což je logické, jelikož malé nesrovnalosti expertní segmentace a predikce na okraji léze generují objemný prostor.



Obrázek 12.2. Dice koeficient v závislosti na objemu expertní segmentace. A: Metoda nnU-Net model *MotolOnly*. B: Metoda nnU-Net model *ensemble_FLAIR_DWI_MNI*. C: Metoda DeepMedic model *ensemble_DWI_FLAIR*. D: Metoda 3D U-Net model *ensemble_FLAIR_DWI_LabelSampler*.

12.5 Vliv TSS

Vliv času od počátku mrtvice je zobrazen na obrázku 12.4. U metody DeepMedic [50–51] a u architektury 3D U-Net [63] došlo k razantnímu zhoršení kvality segmentace s každým dalším týdnem od počátku mrtvice. Model metody nnU-Net [53] natrénovaný na datech

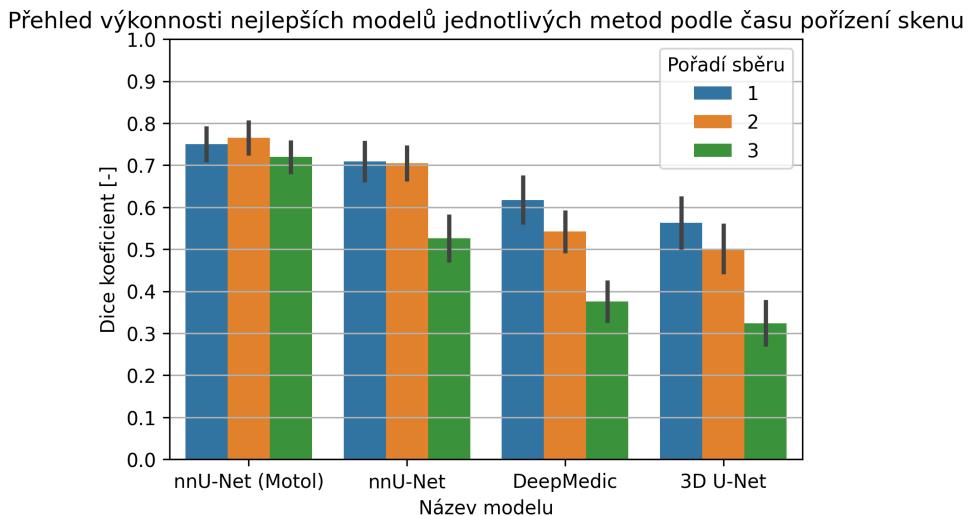


Obrázek 12.3. Rozdíl objemů predikce a expertní segmentace v závislosti na objemu expertní segmentace. A: Metoda nnU-Net model *MotolOnly*. B: Metoda nnU-Net model *ensemble_FLAIR_DWI_MNI*. C: Metoda DeepMedic model *ensemble_DWI_FLAIR*. D: Metoda 3D U-Net model *ensemble_FLAIR_DWI_LabelSampler*.

ISLES 2015 [2] a 2022 [40] se dokázal nějakým způsobem naučit generalizovat i na datech pořízených ve 2. sběru a tudiž zde není významný pokles v Dice koeficientu mezi 1. a 2. sběrem. Což je pozoruhodné, jelikož DeepMedic [50–51] i 3D U-Net [63] s tím mají problém, ale dle dosažených výsledků u metody nnU-Net [53] problém není v trénovacích datech. I přes to u modelu nnU-Net [53] natrénovaném na datech ISLES dochází k výraznému propadu kvality segmentace mezi 2. a 3. sběrem. Dosažené výsledky jsou poměrně očekávatelné, jelikož 3. sběr podle definice hraničí s chronickou lézí a soutěže ISLES se specializují na segmentaci subakutních lézí, případně pozdně akutních lézí. Je tedy očekávatelné, že v těchto datasetech chybí informace pro segmentaci 3. sběru dat.

Model metody nnU-Net [53] natrénovaný na datasetu Motol pětinásobnou křížovou validací vykazuje dokonce zlepšení mezi 1. a 2. sběrem. Pokles Dice koeficientu u 3. sběru není oproti předchozím sběrům dat nikterak velký.

Porovnáním 1. i 2. sběru obou modelů metody nnU-Net [53] zjistíme, že zde nedochází k výraznému propadu Dice koeficientu mezi modely. Teprve u třetího sběru dojde k výraznému snížení Dice koeficientu modelu, jež není natrénován na datasetu Motol. Neboli predikce brzkých subakutních lézí budou od obou modelů srovnatelné, ale predikce pozdních subakutních až chronických lézí modelu natrénovaného na ISLES budou nekvalitní.



Obrázek 12.4. Graf zobrazující průměrný dosažený Dice koeficient nejlepších modelů při validaci na datasetu Motol v závislosti na čase sběru dat. nnU-Net (Motol): Metoda nnU-Net model *MotolOnly*. nnU-Net: Metoda nnU-Net model *ensemble_FLAIR_DWI_MNI*. DeepMedic: Metoda DeepMedic model *ensemble_DWI_FLAIR*. 3D U-Net: Metoda 3D U-Net model *ensemble_FLAIR_DWI_LabelSampler*.

12.6 Srovnání s rešerší

Na začátek je nutné uvést, že přímé srovnání výsledků není možné, jelikož žádná z naštřenovaných instancí jakéhokoli modelu nebyla testována nějakým standardizovaným postupem proti neveřejným testovacím datům. Ačkoliv by to bylo možné, jelikož soutěže zaručující kvalitní evaluaci modelů jsou otevřené i po jejich vyhodnocení. Drtivá většina této práce uvádí výsledky modelů natrénovaných na sloučení datasetů ISLES 2015 [2] a ISLES 2022 [40] a validaci na datasetu Motol, který je privátní, tudíž neexistují metody, s nimiž by se daly dosažené výsledky přímo srovnat. Proto je zde poskytnuto spíše objektivní srovnání nežli absolutní.

Výsledky v této práci při použití metody DeepMedic [50–51] ukazují poměrně velký propad Dice koeficientu (z 0,59 DC na 0,510 DC) oproti výsledkům z testovací fáze soutěže ISLES 2015 [2]. Při pohledu na trénovací skeny datasetu ISLES 2015 [2] je patrné, že se značně liší od zbylých zde použitých datasetů. Zároveň si jsou skeny v ISLES 2015 [2] vizuálně blízké. Sami autoři soutěže označili tuto skutečnost jako nevýhodu a pro další soutěže doporučili použít rozmanitější data. Je tedy velmi pravděpodobné, že pokles Dice koeficientu oproti výsledkům ze soutěže je způsobený extrémní rozmanitostí skenů použitých v této práci. Tudíž na ně metoda z principu nebyla optimalizována. Kromě toho autoři DeepMedic [50–51] v soutěži (na rozdíl od této práce) použili sloučení pravděpodobnostních map z pěti natrénovaných modelů na čtyřech modalitách a CRF. Nutno podotknout, že v této práci bylo nejlepších výsledků metody DeepMedic [50–51] dosaženo při výstupním fúzování pouze dvou unimodálních instancí.

Výsledky dosažené metodou DeepMedic [50–51] a vlastní implementací architektury 3D U-Net [63] nebyly příliš vzdáleny od sebe, tudíž z pohledu Dice koeficientu by obě metody skončily na horních příčkách v soutěži ISLES 2015 [2]. Stejně tak dosahují lepších výsledků než prezentované metody segmentující chronickou CMP z přehledové studie z roku 2019 [42]. Při srovnání se soutěží ISLES 2022 [40, 52] je patrné, že by výsledky metod DeepMedic [50–51] i 3D U-Net [63] v této práci skončily z pohledu DC v poslední trojici.

Metoda nnU-Net [53] testovaná v této práci dosáhla vynikajících výsledků oproti předchozím dvěma metodám. V případě, kdy byla metoda nnU-Net [53] natrénována na sloučení datasetů ISLES a validována na datasetu Motol, její výsledky předčí všechny soutěžící ze soutěže ISLES 2015 [2] - tedy z doby, kdy nnU-Net [53] ještě neexistoval. Nicméně mezi soutěžemi ISLES 2015 [2] a ISLES 2022 [40] je tak extrémní kvalitativní skok, že i výsledky metody nnU-Net [53] v této práci by na soutěži ISLES 2022 [40] skončily mezi posledními. Velkou roli nejspíše hraje inkonzistence mezi trénovacími datasety ISLES a validačním datasetem Motol. Pokud organizátoři ISLES 2022 [40] měli neveřejný testovací dataset složený ze skenů podobných skenům ve zveřejněné části, je jasné, že soutěžní metody dosáhly lepších výsledků. Dále je třeba předpokládat, že soutěžící, kteří použili nnU-Net [53] na soutěži ISLES 2022 [40] jej více či méně upravili, aby dosáhli ještě lepších výsledků.

Při pohledu na výsledky, kdy nnU-Net [53] je natrénován a validován pětinásobnou křížovou validací pouze na datasetu Motol, dosažený DC se začíná blížit stavu poznání dle tabulky výsledků v soutěži ISLES 2022 [40]. Tam byl výherní DC 0,82 a v této práci bylo dosaženo DC 0,745. To do značné míry podtrhuje uvedený názor, že nízký výkon nnU-Net [53] při trénování na ISLES a validaci na datasetu Motol je způsobený rozmanitostí dat a různými datasety.

Mnoho autorů článků z rešerše při vývoji nových řešení optimalizuje jediný model, nicméně nejlepších výsledků dosahuje kombinací několika modelů. To platí také pro nnU-Net [53] i DeepMedic [50–51]. Autoři DeepMedic [50–51] pokaždé mírně upravili konfiguraci modelu a teprve složením pravděpodobnostních map při takto podobných modelů generovali segmentace. Díky tomu u nich došlo ke zlepšení výkonu. Obdobně to platí pro nnU-Net [53], kde se však parametry modelů nemění, pouze jsou jinak na počátku inicializovány váhy a jinak načítána data. Proto každý z natrénovaných modelů dosahuje jiného výkonu a teprve za pomoci výstupní fúze dojde ke sloučení předností jednotlivých modelů a ke zlepšení segmentace. Stejně tak se tomu děje v této práci, kdy u všech tří testovaných metod došlo ke zlepšení Dice koeficientu při použití výstupní fúze.

Přestože mnoho soutěžících ze soutěže ISLES 2022 [40] vycházelo z původního článku 3D U-Net [63], výsledky dosažené v této práci pomocí vlastní implementace jsou velmi vzdáleny jak od výsledků soutěžících v ISLES 2022 [40], tak od možného optimálního. Toto tvrzení je podloženo mj. tím, že na stejných datech s podobnou architekturou metoda nnU-Net [53] dosáhla v této práci značně lepších výsledků.

Z rešerše a ze získaných výsledků vyplývá, že pro dosažení maximálního Dice koeficientu odpovídajícímu stavu poznání, je důležité využívat 3D deep learningové metody, nejlépe založené na architektuře 3D U-Net [63] nebo na některých jeho derivátech. Důležitější nežli architektura je ale samotné nastavení celého procesu.

12.7 Návrhy na zlepšení

Co se týče zlepšení dosahovaných výsledků vlastní implementace 3D U-Net [63], bylo by vhodné za pomoci zkušeného experta identifikovat parametry, jež mají zásadní vliv na to, proč implementace nedosahuje srovnatelných výsledků jako nnU-Net [53]. V této práci byla měněna např. velikost modelu ve smyslu počtu konvolučních jader, nikoliv ve smyslu hloubky modelu. Metoda nnU-Net [53] např. aplikuje striktní pravidla, dle kterých síť škáluje, a proto je síť v případě nnU-Net [53] hlubší. V této práci se nepodařilo najít parametry, které by měly v případě 3D U-Net [63] zásadní vliv na dosahovaný DC.

Za podmínky, že bychom měli vždy k dispozici předchozí skeny daného pacienta, bylo by vhodné natrénovat tři modely neuronové sítě. První z modelů by byl natrénován stejně jako modely v této práci, ale predikoval by pouze segmentace na FLAIR a DWI skenech z prvního sběru. Druhý z modelů by používal pro rozhodování dvě modality z druhého sběru a navíc výstupní pravděpodobnostní mapu z prvního modelu - měl by tedy znalost o předchozí segmentaci. Třetí model by to dělal obdobným způsobem, kdy vstupem do něj by byly dvě modality ze třetího sběru a pravděpodobnostní mapa segmentace z druhého sběru. Jinými slovy každý další model by využíval informaci zpětně v čase, kde se léze vyskytovala. Propagací informace o segmentaci z předchozích sběrů by se dala zvětšit přesnost na datasetu Motol. Podobným způsobem postupují i lékařští experti a segmentaci aktuálního skenu podporují vizuální inspekci staršího skenu stejného pacienta, pokud je k dispozici.

Pro dosažení lepšího DC by stálo za zvážení předtrénovat segmentační modely na dalších veřejných MRI datasetech např. chronické CMP nebo na nádorech mozku. Takových dat je poměrně mnoho veřejně k dispozici, ale budou obsahují jiné modality než používané v této práci (např. T1-w) nebo podstata úlohy je jiná než v této práci. Lze však následně použít transfer learning nebo fine-tuning pro specializaci na např. malý dataset. Ověření vlivu fine-tuningu nebo transfer learningu je nad rámec zadání této diplomové práce, proto nebylo blíže zkoumáno.

V případě, kdy v této práci bylo třeba prahovat pravděpodobnostní mapy, byly prahovány striktně hodnotou 0,5. Tato hodnota je suboptimální, proto by pro další zlepšení bylo vhodné najít optimální hodnotu.

K mírnému zlepšení DC i časové náročnosti by mohlo vést použití tzv. soups [76] místo klasické fúze popsané v této práci. Soups v současnosti ukazují slibné výsledky. Principem je natrénování více modelů a zprůměrování vah napříč vrstvami těchto modelů. Neprobíhá zde tedy klasické slučování informace používané v této práci a pro účely predikce se používá jediný model.

Dále se ukázalo, že koregistrace do prostoru MNI v případě metody nnU-Net [53] nepřináší výrazné zlepšení přesnosti segmentace. Proto z hlediska rychlosti učení, ale hlavně z hlediska rychlosti predikce v ostrém provozu, by bylo výhodnější tuto koregistraci nedělat. V případě vlastního řešení je vhodnější se soustředit na optimalizaci nastavení řešení jako je např. augmentace, z toho důvodu, aby koregistrace do společného prostoru hrála minimální vliv, stejně jako u metody nnU-Net [53].

Článek metody nnU-Net [53] bohužel explicitně neuvádí, proč autoři došli ke zvolení daných pravidel, přesto metoda určuje současný stav poznání. Proto by bylo zajímavé podrobněji studovat samotnou metodu nnU-Net [53] a volbu daných pravidel a jejich vliv na výkonnost metody. Kromě toho u některých datasetů dle článku metody je vhodné provést doladění parametrů - např. lepší výsledky při segmentaci nádorů mozku jsou dosaženy při použití tří konvolucí na jeden stupeň U-Net. Chybí literatura, která by ověřovala, proč tomu tak je a jak předem odhadovat tyto jevy a případně je implementovat do nnU-Net [53] jako další pravidla závislá na otisku datasetu.

A poslední návrhy na zlepšení se týkají čistě implementační roviny. Po programování s knihovnou ANTsPy bylo objeveno několik chyb, které nebyly evidovány a týkaly se především chyb v zaokrouhlování nebo výstupy operací knihovny ANTsPy dávaly různé hodnoty pro různé vstupní datové typy hodnotově stejných tensorů. Hledání a obcházení těchto chyb trvalo značnou dobu. Neodladěnost knihovny a především zmíněné skryté chyby při čítání především tomu, že se jedná o wrapper pro program ANTs napsaný v jazyce C++. Mimo to dokumentace v některých případech není konzistentní s aktuálním kódem. Proto by bylo vhodné mít zkušenosť i s alternativní a hojně používanou

knihovnou pro vytváření koregistrací SimpleITK. SimpleITK je velmi často používána metodami z rešerše a mj. tuto knihovnu používá i metoda nnU-Net [53]. Proto by bylo zajímavé porovnat rychlosť, kvalitu a komfort používania API týchto dvou knihoven.

Pro vývoj vlastních řešení segmentace lze použiť vysokoúrovňový deep learningový framework MONAI [77]. Snad kromě koregistrování má zabudované všechny důležité funkce pro manipulaci se skenami i pro trénování modelů. Zahŕňuje funkcie knihovny TorchIO, obsahuje preddefinované běžné architektury neuronových sítí a jiné nástroje. Existuje snaha začlenit nnU-Net [53] do MONAI, kde nyní existuje DynUNet. Jedná se o reimplementaci 3D_fullres konfigurace nnU-Net [53]. Kód metody DynUNet je také poměrně odlehčený, takže v případě potřeby v něm lze snadno dělat zásahy. Pro komerční použití by bylo vhodné zvážit vytváření reálných aplikací v MONAI, který integruje vše podstatné vč. funkcí pro testování a nasazení.

Kapitola 13

Závěr

V této diplomové práci byla provedena rešerše literatury vycházející převážně ze soutěží ISLES 2015 a ISLES 2022, které zkoumají stejný problém jako tato práce. Byla provedena analýza několika metod zúčastněných na těchto soutěžích. Byly analyzovány i metody z přehledového článku řešící segmentaci chronické mrtvice.

Pro nalezení kvalitního modelu sloužícího pro segmentování pacientů na 2. Lékařské fakultě UK byly ověřeny celkem tři 3D metody implementující neuronové sítě. Modely byly natrénovány na veřejně dostupných trénovacích datasetech ISLES 2015 a ISLES 2022 a validovány na datasetu poskytnutém 2. LF. Ukázalo se, že metoda DeepMedic výrazně zaostává za architekturami typu U-Net. Proto byla dále ověřena metoda nnU-Net a vlastní implementace architektury 3D U-Net. Metoda nnU-Net fungovala ze všech testovaných nejlépe a pomocí ní bylo dosaženo Dice koeficientu až $0,646 \pm 0,238$ se sensitivitou 0,631 a specificitou 0,995. Dále byla metoda nnU-Net ověřena čistě na datasetu Motol pomocí pětinásobné křížové validace. Tím bylo dosaženo Dice koeficientu až $0,745 \pm 0,182$ se sensitivitou 0,763 a specificitou 0,996. Protože tyto dva modely fungovaly nejlépe ze všech testovaných, byly doručeny vedoucímu práce pro praktické použití. Rozdíl Dice koeficientů těchto dvou modelů poukazuje na to, že nnU-Net (i zbylé metody) mají problémy s predikcemi na datech z jiného zdravotnického centra.

Vlastní implementace 3D U-Net dosáhla špatných výsledků ve srovnání s nnU-Net. To je s největší pravděpodobnostní zapříčiněno neoptimální konfigurací trénování a vytváření predikcí, jelikož dle rešerše lze s architekturou založenou na 3D U-Net dosáhnout mnohem kvalitnějších predikcí. Přestože 3D U-Net pochází z roku 2016, lze po správném nastavení dosáhnout výsledků definující aktuální stav poznání, což ukazuje metoda nnU-Net. Neoptimální nastavení procesů má mnohem fatálnější důsledky než výběr samotné architektury neuronové sítě (viz vlastní implementace 3D U-Net).

V práci byly ověřeny vlivy fúze a ukázalo se, že výstupní fúze dosahují u všech tří testovaných metod vyšších Dice koeficientů než vstupní fúze. To je však za cenu mírně delšího času predikce a větších nároků na úložiště natrénovaných modelů.

Ukázalo se, že všechny tři metody pro velké léze vůbec neselhávají a dosahují vysokých Dice koeficientů. Zároveň jsou ale tyto velké léze objemově podhodnocovány.

Valný vliv na přesnost segmentace má čas, který uběhl od počátku onemocnění po naskenování. Čím delší čas uběhl od počátku mrtvice, tím hůře predikují modely metod, které jsou natrénované na datasetech ISLES 2015 a ISLES 2022. U modelu metody nnU-Net, který byl natrénován pouze na datasetu Motol, je tento jev minimalizován.

Dosažené Dice koeficienty modelů určených pro praktické použití jsou převážně vysoké a segmentace lézí jsou poměrně kvalitní. Přesto u některých pacientů metoda nnU-Net naprosto selhává, stejně jako metody ostatní. Proto je vždy důležitá expertní vizuální kontrola celého objemu skenu.

Díky rychlému vývoji hardwaru a rychlému výzkumu metod umělé inteligence dojdou soudobé algoritmy dosahovat obdivuhodné kvality. I přes značné výzvy, jaké klade úloha automatické segmentace subakutních lézí cévní mozkové příhody, je možné automatickou metodou alespoň částečně pomoci expertům.

Literatura

- [1] Valery L Feigin and Michael Brainin and Bo Norrving and Sheila Martins and Ralph L Sacco and Werner Hacke and Marc Fisher and Jeyaraj Pandian and Patrice Lindsay. World Stroke Organization (WSO): Global Stroke Fact Sheet 2022. *International Journal of Stroke*. 2022, 17 (1), 18-29. DOI 10.1177/17474930211065917. PMID: 34986727.
- [2] Oskar Maier, Bjoern H. Menze, Janina von der Gabletz, Levin Hani, Mattias P. Heinrich, Matthias Liebrand, Stefan Winzeck, Abdul Basit, Paul Bentley, Liang Chen, Daan Christiaens, Francis Dutil, Karl Egger, Chaolu Feng, Ben Glockner, Michael Gotz, Tom Haeck, Hanna-Leena Halme, Mohammad Havaei, Khan M. Iftekharuddin, Pierre-Marc Jodoin, Konstantinos Kamnitsas, Elias Kellner, Antti Korvenoja, Hugo Larochelle, Christian Ledig, Jia-Hong Lee, Frederik Maes, Qaiser Mahmood, Klaus H. Maier-Hein, Richard McKinley, John Muschelli, Chris Pal, Linmin Pei, Janaki Raman Rangarajan, Syed M. S. Reza, David Robben, Daniel Rueckert, Eero Salli, Paul Suetens, Ching-Wei Wang, Matthias Wilms, Jan S. Kirschke, Ulrike M. Kramer, Thomas F. Munte, Peter Schramm, Roland Wiest, Heinz Handels a Mauricio Reyes. ISLES 2015 - A public evaluation benchmark for ischemic stroke lesion segmentation from multispectral MRI. *Med. Image Anal.* 2017, 35 250–269. DOI 10.1016/j.media.2016.07.009.
- [3] Centers for DiseaseControl a Prevention. *FastStats - Leading Causes of Death*. 2024.
<https://www.cdc.gov/nchs/fastats/leading-causes-of-death.htm>. [Online; accessed 7. Mar. 2024.]
- [4] Český statistický úřad (ČSÚ). *Zemřelí podle zkráceného seznamu příčin smrti v ČR, krajích a okresech - 2014–2023*. 2024.
<https://csu.gov.cz/produkty/zemreli-podle-zkraceneho-seznamu-pric-in-smrti-v-cr-krajich-a-okresech-wj9znry8zt>. [Online; accessed 19. Sep. 2024.]
- [5] Michelle P. Lin a David S. Liebeskind. Imaging of Ischemic Stroke. *Continuum : Lifelong Learning in Neurology*. 2016, 22 (5), 1399. DOI 10.1212/CON.0000000000000376.
- [6] Eric Hartono Tedyanto, Kumara Tini a Nyoman Angga Krishna Pramana. Magnetic Resonance Imaging in Acute Ischemic Stroke. *Cureus*. 2022, 14 (7). DOI 10.7759/cureus.27224.
- [7] *MRI coils | MRI head coil | MRI foot coil | RF coil | Types of coil*. 2018.
<https://www.dofrp.com/cases-study/mri-coils>. [Online; accessed 19. Sep. 2024.]
- [8] John P. Ridgway. Cardiovascular magnetic resonance physics for clinicians: Part I. *Journal of cardiovascular magnetic resonance : official journal of the Society for Cardiovascular Magnetic Resonance*. 2010, 12 (1), 71. DOI 10.1186/1532-429X-12-71.

- [9] Vijay P. B. Grover, Joshua M. Tognarelli, Mary M. E. Crossey, I. Jane Cox, Simon D. Taylor-Robinson a Mark J. W. McPhail. Magnetic Resonance Imaging: Principles and Techniques: Lessons for Clinicians. *Journal of Clinical and Experimental Hepatology*. 2015, 5 (3), 246. DOI 10.1016/j.jceh.2015.08.001.
- [10] *Magnets, Spins and Resonances: An introduction to the basics of Magnetic Resonance*. 2023.
<https://www.magnetomworld.siemens-healthineers.com/publications/mr-basics>. [Online; accessed 22. Sep. 2024].
- [11] Cynthia B. Paschal a H. Douglas Morris. K-Space in the Clinic. *Journal of magnetic resonance imaging : JMRI*. 2004, 19 (2), 145–59. DOI 10.1002/jmri.10451.
- [12] Laura M. Allen, Anton N. Hasso, Jason Handwerker a Hamed Farid. Sequence-specific MR Imaging Findings That Are Useful in Dating Ischemic Stroke. *Radiographics : a review publication of the Radiological Society of North America, Inc.* 2012, 32 (5), 1285–97. DOI 10.1148/radiographics.325115760.
- [13] K.J. van Everdingen, J. vander Grond, L.J. Kappelle, L.M.P. Ramos a W.P.T.M. Mali. Diffusion-Weighted Magnetic Resonance Imaging in Acute Stroke. *Stroke*. 1998, 29 (9), 1783-1790. DOI 10.1161/01.STR.29.9.1783.
- [14] Yue Zhang, Shijie Liu, Chunlai Li a Jianyu Wang. Application of Deep Learning Method on Ischemic Stroke Lesion Segmentation. *Journal of Shanghai Jiaotong University (Science)*. 2022, 27 99-111. DOI 10.1007/S12204-021-2273-9/METRICS. Deep learning algorithms can notonly analyze and obtain statistical rules automaticallyby training data but also use the learned statistical rulesto predict and analyze unknown data..
- [15] Gotz Thomalla, Philipp Rossbach, Michael Rosenkranz, Susanne Siemonsen, Anna Krutzelmann, Jens Fiehler a Christian Gerloff. Negative fluid-attenuated inversion recovery imaging identifies acute ischemic stroke at 3 hours or less. *Ann. Neurol.* 2009, 65 (6), 724–732.
- [16] Kennedy R. Lees, Erich Bluhmki, Rudiger von Kummer, Thomas G. Brott, Danilo Toni, James C. Grotta, Gregory W. Albers, Markku Kaste, John R. Marler, Scott A. Hamilton, Barbara C. Tilley, Stephen M. Davis, Geoffrey A. Donnan a Werner Hacke. Time to treatment with intravenous alteplase and outcome in stroke: an updated pooled analysis of ECASS, ATLANTIS, NINDS, and EPITHET trials. *Lancet*. 2010, 375 (9727), 1695–1703. DOI 10.1016/S0140-6736(10)60491-6.
- [17] Ela Marie Z. Akay, Jana Rieger, Ricardo Schottler, Jonas Behland, Raphael Schymczyk, Ahmed A. Khalil, Ivana Galinovic, Jan Sobesky, Jochen B. Fiebach, Vince I. Madai, Adam Hilbert a Dietmar Frey. A deep learning analysis of stroke onset time prediction and comparison to DWI-FLAIR mismatch. *NeuroImage: Clinical*. 2023, 40 103544. DOI 10.1016/j.nicl.2023.103544.
- [18] Cecilie Mørck Offersen, Jens Sørensen, Kaining Sheng, Jonathan Frederik Carlsen, Annika Reynberg Langkilde, Akshay Pai, Thomas Clement Truelsen a Michael Bachmann Nielsen. Artificial Intelligence for Automated DWI/FLAIR Mismatch Assessment on Magnetic Resonance Imaging in Stroke: A Systematic Review. *Diagnostics (Basel)*. 2023, 13 (12), 2111.. DOI 10.3390/diagnostics13122111.
- [19] Andriy Fedorov, Reinhard Beichel, Jayashree Kalpathy-Cramer, Julien Finet, Jean-Christophe Fillion-Robin, Sonia Pujol, Christian Bauer, Dominique Jennings, Fiona Fennessy, Milan Sonka, John Buatti, Stephen Aylward, James V. Miller, Steve Pieper a Ron Kikinis. 3D Slicer as an Image Computing Platform

for the Quantitative Imaging Network. *Magn. Reson. Imaging.* 2012, 30 (9), 1323. DOI 10.1016/j.mri.2012.05.001.

- [20] Xiangrui Li, Paul S. Morgan, John Ashburner, Jolinda Smith a Christopher Rorden. The first step for neuroimaging data analysis: DICOM to NIfTI conversion. *J. Neurosci. Methods.* 2016, 264 47–56. DOI 10.1016/j.jneumeth.2016.03.001.
- [21] Maciej A. Mazurowski, Mateusz Buda, Ashirbani Saha a Mustafa R. Bashir. *Deep learning in radiology: An overview of the concepts and a survey of the state of the art with focus on MRI.* 2019. tasky ve strojovém učení,principy ann, cnn, tradiční vs deep, .
- [22] June-Goo Lee, Sanghoon Jun, Young-Won Cho, Hyunna Lee, Guk Bae Kim, Joon Beom Seo a Namkug Kim. Deep Learning in Medical Imaging: General Overview. *Korean J. Radiol.* 2017, 18 (4), 570–584. DOI 10.3348/kjr.2017.18.4.570.
- [23] German Gonzalez, George R. Washko a Raul San Jose Estepar. Deep learning for biomarker regression: application to osteoporosis and emphysema on chest CT scans. *Proceedings of SPIE—the International Society for Optical Engineering.* 2018, 10574 DOI 10.1117/12.2293455.
- [24] Hansang Lee, Helen Hong, Jinsil Seong, Jin Sung Kim a Junmo Kim. Survival prediction of liver cancer patients from CT images using deep learning and radiomic feature-based regression. *Medical Imaging 2020: Computer-Aided Diagnosis.* 2020, 11314 849–854. DOI 10.1117/12.2551349.
- [25] Pietro Caliandro, Jacopo Lenkowicz, Giuseppe Reale, Simone Scaringi, Aurelia Zauli, Christian Uccheddu, Simone Fabiole-Nicoletto, Stefano Patarnello, Andrea Damiani, Luca Tagliaferri, Iacopo Valente, Marco Moci, Mauro Monforte, Vincenzo Valentini a Paolo Calabresi. Artificial intelligence to predict individualized outcome of acute ischemic stroke patients: The SIBILLA project. *European Stroke Journal.* 2024, 23969873241253366. DOI 10.1177/23969873241253366.
- [26] Hossein Abbasi, Maysam Orouskhani, Samaneh Asgari a Sara Shomal Zadeh. Automatic brain ischemic stroke segmentation with deep learning: A review. *Neuroscience Informatics.* 2023, 3 100145. DOI 10.1016/j.neuri.2023.100145.
- [27] J. A. Fiez, H. Damasio a T. J. Grabowski. Lesion segmentation and manual warping to a reference brain: intra- and interobserver reliability. *Hum. Brain Mapp.* 2000, 9 (4), 192–211. DOI 10.1002/(sici)1097-0193(200004)9:4<192::aid-hbm2>3.0.co;2-y.
- [28] Kaori L. Ito, Hosung Kim a Sook-Lei Liew. A comparison of automated lesion segmentation approaches for chronic stroke T1-weighted MRI data. *bioRxiv.* 2018, 441451. DOI 10.1101/441451.
- [29] Dominik Müller, Iñaki Soto-Rey a Frank Kramer. *Towards a Guideline for Evaluation Metrics in Medical Image Segmentation.* 2022.
<https://arxiv.org/abs/2202.05273>.
- [30] Abdel Aziz Taha a Allan Hanbury. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC Med. Imaging.* 2015, 15:29. DOI 10.1186/s12880-015-0068-x.
- [31] William M. Rand. Objective Criteria for the Evaluation of Clustering Methods. *J. Am. Stat. Assoc.* 1971.
- [32] Lee R. Dice. Measures of the Amount of Ecologic Association Between Species. *Ecology.* 1945, 26 (3), 297–302. DOI 10.2307/1932409.

- [33] Paul Jaccard. THE DISTRIBUTION OF THE FLORA IN THE ALPINE ZONE. *New Phytol.* 1912, 11 (2), 37–50. DOI 10.1111/j.1469-8137.1912.tb05611.x.
- [34] D. P. Huttenlocher, G. A. Klanderman a W. A. Rucklidge. Comparing Images Using the Hausdorff Distance. *IEEE Trans. Pattern Anal. Mach. Intell.* 1993, 15 (9), 850–863. DOI 10.1109/34.232073.
- [35] Orhun Utku Aydin, Abdel Aziz Taha, Adam Hilbert, Ahmed A. Khalil, Ivana Galinovic, Jochen B. Fiebach, Dietmar Frey a Vince Istvan Mada. On the usage of average Hausdorff distance for segmentation performance assessment: hidden error when used for ranking. *Eur. Radiol. Exp.* 2021, 5 (1), 1–7. DOI 10.1186/s41747-020-00200-2.
- [36] Tobias Heimann, Bram van Ginneken, Martin A. Styner, Yulia Arzhaeva, Volker Aurich, Christian Bauer, Andreas Beck, Christoph Becker, Reinhard Beichel, Gyorgy Bekes, Fernando Bello, Gerd Binnig, Horst Bischof, Alexander Bornik, Peter M. M. Cashman, Ying Chi, Andres Cordova, Benoit M. Dawant, Marta Fidrich, Jacob D. Furst, Daisuke Furukawa, Lars Grenacher, Joachim Hornegger, Dagmar Kainmuller, Richard I. Kitney, Hidefumi Kobatake, Hans Lamecker, Thomas Lange, Jeongjin Lee, Brian Lennon, Rui Li, Senhu Li, Hans-Peter Meinzer, Gabor Nemeth, Daniela S. Raicu, Anne-Mareike Rau, Eva M. van Rikxoort, Mikaël Rousson, Laszlo Rusko, Kinda A. Saddi, Gunter Schmidt, Dieter Seghers, Akinobu Shimizu, Pieter Slagmolen, Erich Sorantin, Grzegorz Soza, Ruchaneewan Susomboon, Jonathan M. Waite, Andreas Wimmer a Ivo Wolf. Comparison and evaluation of methods for liver segmentation from CT datasets. *IEEE Trans. Med. Imaging.* 2009, 28 (8), 1251–1265. DOI 10.1109/TMI.2009.2013851.
- [37] Jacob Cohen. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement.* 1960, 20 (1), 37–46. DOI 10.1177/001316446002000104.
- [38] J. A. Hanley a B. J. McNeil. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology.* 1982, 143 (1), 29–36. DOI 10.1148/radiology.143.1.7063747.
- [39] David M. W. Powers. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *arXiv.* 2020. DOI 10.48550/arXiv.2010.16061.
- [40] Moritz R. Hernandez Petzsche, Ezequiel de la Rosa, Uta Hanning, Roland Wiest, Waldo Valenzuela, Mauricio Reyes, Maria Meyer, Sook-Lei Liew, Florian Kofler, Ivan Ezhov, David Robben, Alexandre Hutton, Tassilo Friedrich, Teresa Zarth, Johannes Burkle, The Anh Baran, Bjorn Menze, Gabriel Broocks, Lukas Meyer, Claus Zimmer, Tobias Boeckh-Behrens, Maria Berndt, Benno Ikenberg, Benedikt Wiestler a Jan S. Kirschke. ISLES 2022: A multi-center magnetic resonance imaging stroke lesion segmentation dataset. *Sci. Data.* 2022, 9 (762), 1–9. DOI 10.1038/s41597-022-01875-5.
- [41] Stefan Winzeck, Arsany Hakim, Richard McKinley, Jose A. A. D. S. R. Pinto, Victor Alves, Carlos Silva, Maxim Pisov, Egor Krivov, Mikhail Belyaev, Miguel Monteiro, Arlindo Oliveira, Youngwon Choi, Myunghee Cho Paik, Yongchan Kwon, Hanbyul Lee, Beom Joon Kim, Joong-Ho Won, Mobarakol Islam, Hongliang Ren, David Robben, Paul Suetens, Enhao Gong, Yilin Niu, Junshen Xu, John M. Pauly, Christian Lucas, Matthias P. Heinrich, Luis C. Rivera, Laura S. Castillo, Laura A. Daza, Andrew L. Beers, Pablo Arbelaezs, Oskar Maier, Ken Chang, James M. Brown, Jayashree Kalpathy-Cramer, Greg Zaharchuk, Roland Wiest a Mauricio Reyes. ISLES 2016 and 2017-Benchmarking Ischemic Stroke Lesion

Outcome Prediction Based on Multispectral MRI. *Front. Neurol.* 2018, 9 393237. DOI 10.3389/fneur.2018.00679.

- [42] Kaori L. Ito, Hosung Kim a Sook-Lei Liew. A comparison of automated lesion segmentation approaches for chronic stroke T1-weighted MRI data. *Hum. Brain Mapp.* 2019, 40 (16), 4669–4685. DOI 10.1002/hbm.24729.
- [43] Nagesh K. Subbanna, Deepthi Rajashekhar, Bastian Cheng, Gotz Thomalla, Jens Fiehler, Tal Arbel a Nils D. Forkert. Stroke Lesion Segmentation in FLAIR MRI Datasets Using Customized Markov Random Fields. *Front. Neurol.* 2019, 10:541. DOI 10.3389/fneur.2019.00541.
- [44] Bjoern H. Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, Levente Lanczi, Elizabeth Gerstner, Marc-Andre Weber, Tal Arbel, Brian B. Avants, Nicholas Ayache, Patricia Buendia, D. Louis Collins, Nicolas Cordier, Jason J. Corso, Antonio Criminisi, Tilak Das, Herve Delingette, Çagatay Demiralp, Christopher R. Durst, Michel Dojat, Senan Doyle, Joana Festa, Florence Forbes, Ezequiel Geremia, Ben Glocker, Polina Golland, Xiaotao Guo, Andac Hamamci, Khan M. Iftekharuddin, Raj Jena, Nigel M. John, Ender Konukoglu, Danial Lashkari, Jose Antonio Mariz, Raphael Meier, Sergio Pereira, Doina Precup, Stephen J. Price, Tammy Riklin Raviv, Syed M. S. Reza, Michael Ryan, Duygu Sarikaya, Lawrence Schwartz, Hoo-Chang Shin, Jamie Shotton, Carlos A. Silva, Nuno Sousa, Nagesh K. Subbanna, Gabor Szekely, Thomas J. Taylor, Owen M. Thomas, Nicholas J. Tustison, Gozde Unal, Flor Vasseur, Max Wintermark, Dong Hye Ye, Liang Zhao, Binsheng Zhao, Darko Zikic, Marcel Prastawa, Mauricio Reyes a Koen Van Leemput. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Trans. Med. Imaging.* 2015, 34 (10), 1993–2024. DOI 10.1109/TMI.2014.2377694.
- [45] Martin Styner, Joohwi Lee, Brian Chin, Matthew S. Chin a Simon Warfield. 3D Segmentation in the Clinic: A Grand Challenge II: MS lesion segmentation. *MIDAS Journal.* 2007. DOI 10.54294/lmkqvm.
- [46] Geert Litjens, Oscar Debats, Jelle Barentsz, Nico Karssemeijer a Henkjan Huismans. Computer-Aided Detection of Prostate Cancer in MRI. *IEEE Trans. Med. Imaging.* 2014, 33 (5), 1083–1092. DOI 10.1109/TMI.2014.2303821.
- [47] David W. Shattuck, Gautam Prasad, Mubeena Mirza, Katherine L. Narr a Arthur W. Toga. Online Resource for Validation of Brain Segmentation Methods. *Neuroimage.* 2008, 45 (2), 431. DOI 10.1016/j.neuroimage.2008.10.066.
- [48] K. Murphy. *Development and evaluation of automated image analysis techniques in thoracic CT.* Utrecht, The Netherlands: Utrecht University, 2011. ISBN 978-90-393-5542-8.
<https://dspace.library.uu.nl/handle/1874/204800>.
- [49] Rina D. Rudyanto, Sjoerd Kerkstra, Eva M. van Rikxoort, Catalin Fetita, Pierre-Yves Brillet, Christophe Lefevre, Wenzhe Xue, Xiangjun Zhu, Jianming Liang, Ilkay Oksuz, Devrim Unay, Kamuran Kadipaşaoglu, Raul San Jose Estepar, James C. Ross, George R. Washko, Juan-Carlos Prieto, Marcela Hernandez Hoyos, Maciej Orkisz, Hans Meine, Markus Hullebrand, Christina Stocker, Fernando Lopez Mir, Valery Narango, Eliseo Villanueva, Marius Staring, Changyan Xiao, Berend C. Stoel, Anna Fabijanska, Erik Smistad, Anne C. Elster, Frank Lindseth, Amir Hossein Foruzan, Ryan Kiros, Kartikey Popuri, Dana Cobzas, Daniel Jimenez-Carretero, Andres Santos, Maria J. Ledesma-Carbayo, Michael Helmberger, Mar-

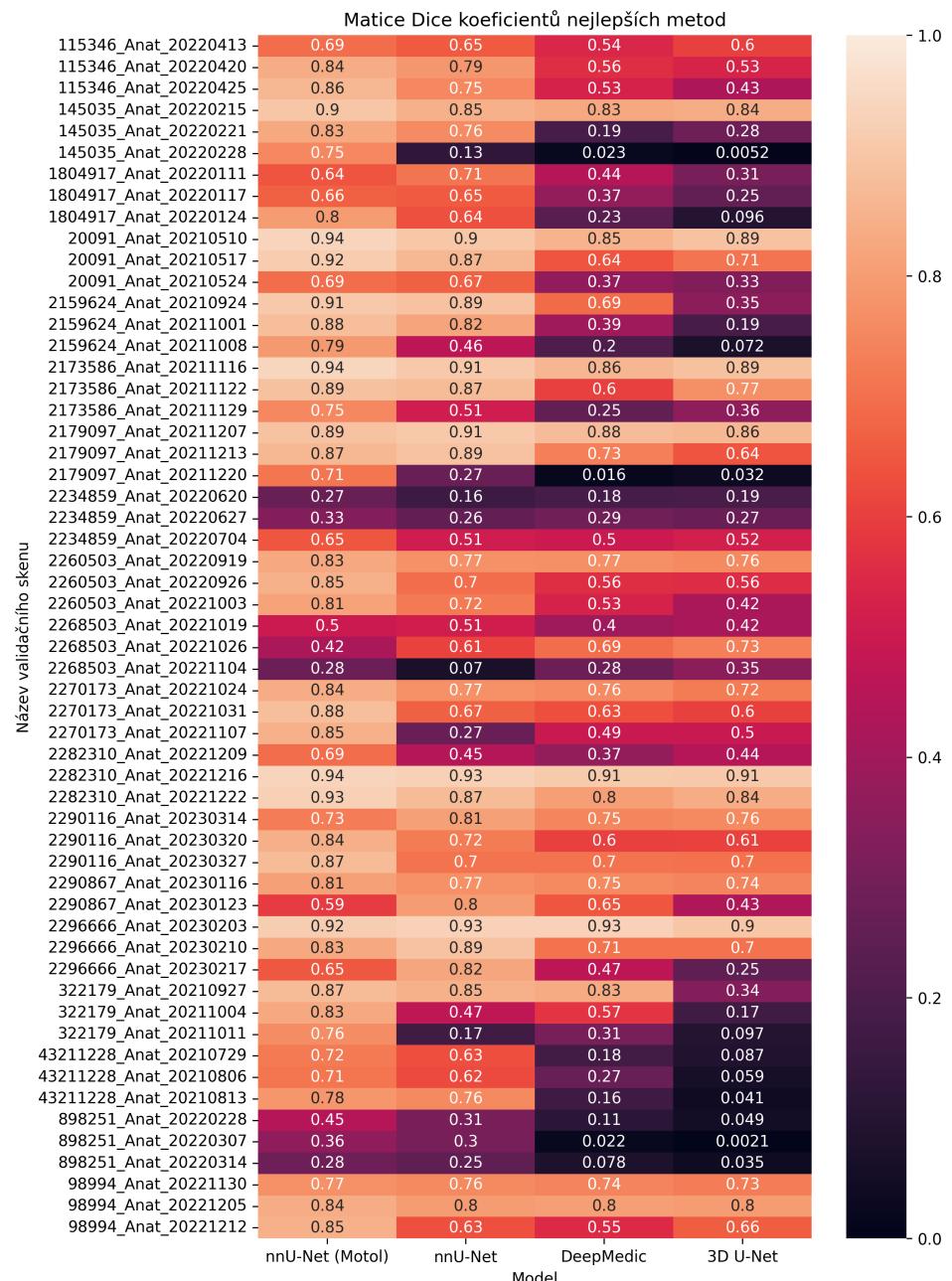
- tin Urschler, Michael Pienn, Dennis G. H. Bosboom, Arantza Campo, Mathias Prokop, Pim A. de Jong, Carlos Ortiz-de-Solorzano, Arrate Munoz-Barrutia a Bram van Ginneken. Comparing algorithms for automated vessel segmentation in computed tomography scans of the lung: the VESSEL12 study. *Med. Image Anal.* 2014, 18 (7), 1217–1232. DOI 10.1016/j.media.2014.07.003.
- [50] K. Kamnitsas, Liang Chen, C. Ledig, D. Rueckert a Ben Glocker. *Multi-scale 3D convolutional neural networks for lesion segmentation in brain MRI*. 2015. <https://www.semanticscholar.org/paper/Multi-scale-3D-convolutional-neural-networks-for-in-Kamnitsas-Chen/8a25125427b1b438311281b60b452d04252fa764>. [Online; accessed 24. Nov. 2024].
- [51] Konstantinos Kamnitsas, Christian Ledig, Virginia F. J. Newcombe, Joanna P. Simpson, Andrew D. Kane, David K. Menon, Daniel Rueckert a Ben Glocker. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med. Image Anal.* 2017, 36 61–78. DOI 10.1016/j.media.2016.10.004.
- [52] Ezequiel de la Rosa, Mauricio Reyes, Sook-Lei Liew, Alexandre Hutton, Roland Wiest, Johannes Kaesmacher, Uta Hanning, Arsany Hakim, Richard Zubal, Waldo Valenzuela, David Robben, Diana M. Sima, Vincenzo Anania, Arne Brys, James A. Meakin, Anne Mickan, Gabriel Broocks, Christian Heitkamp, Shengbo Gao, Kongming Liang, Ziji Zhang, Md Mahfuzur Rahman Siddiquee, Andriy Myronenko, Pooya Ashtari, Sabine Van Huffel, Hyun-su Jeong, Chi-ho Yoon, Chulhong Kim, Jiayu Huo, Sebastien Ourselin, Rachel Sparks, Albert Clerigues, Arnau Oliver, Xavier Llado, Liam Chalcroft, Ioannis Pappas, Jeroen Bertels, Ewout Heylen, Juliette Moreau, Nima Hatami, Carole Frindel, Abdul Qayyum, Moona Mazher, Domenec Puig, Shao-Chieh Lin, Chun-Jung Juan, Tianxi Hu, Lyndon Boone, Maged Goubran, Yi-Jui Liu, Susanne Wegener, Florian Kofler, Ivan Ezhov, Suprosanna Shit, Moritz R. Hernandez Petzsche, Bjoern Menze, Jan S. Kirschke a Benedikt Wiestler. A Robust Ensemble Algorithm for Ischemic Stroke Lesion Segmentation: Generalizability and Clinical Utility Beyond the ISLES Challenge. *arXiv*. 2024. DOI 10.48550/arXiv.2403.19425.
- [53] Fabian Isensee, Paul F. Jaeger, Simon A. A. Kohl, Jens Petersen a Klaus H. Maier-Hein. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods*. 2021, 18 203–211. DOI 10.1038/s41592-020-01008-z.
- [54] ISLES22_SEALS. 2024. https://github.com/Tabrisrei/ISLES22_SEALS. [Online; accessed 12. Nov. 2024].
- [55] Md Mahfuzur Rahman Siddique, Dong Yang, Yufan He, Daguang Xu a Andriy Myronenko. Automated ischemic stroke lesion segmentation from 3D MRI. *arXiv*. 2022. DOI 10.48550/arXiv.2209.09546.
- [56] Sook-Lei Liew, Julia M. Anglin, Nick W. Banks, Matt Sondag, Kaori L. Ito, Hosung Kim, Jennifer Chan, Joyce Ito, Connie Jung, Nima Khoshab, Stephanie Lefebvre, William Nakamura, David Saldana, Allie Schmiesing, Cathy Tran, Danny Vo, Tyler Ard, Panthea Heydari, Bokkyu Kim, Lisa Aziz-Zadeh, Steven C. Cramer, Jingchun Liu, Surjo Soekadar, Jan-Egil Nordvik, Lars T. Westlye, Junping Wang, Carolee Winstein, Chunshui Yu, Lei Ai, Bonhwang Koo, R. Cameron Craddock, Michael Milham, Matthew Lakich, Amy Pienta a Alison Stroud. A large, open source dataset of stroke anatomical brain images and manual lesion segmentations. *Sci. Data*. 2018, 5 (180011), 1–11. DOI 10.1038/sdata.2018.11.

- [57] Bianca de Haan, Philipp Clas, Hendrik Juenger, Marko Wilke a Hans-Otto Karnath. Fast semi-automated lesion demarcation in stroke. *NeuroImage: Clinical*. 2015, 9 69–74. DOI 10.1016/j.nicl.2015.06.013.
- [58] Mohamed L. Seghier, Anil Ramlackhansingh, Jenny Crinion, Alexander P. Leff a Cathy J. Price. Lesion identification using unified segmentation-normalisation models and fuzzy clustering. *Neuroimage*. 2008, 41 (4), 1253–1266. DOI 10.1016/j.neuroimage.2008.03.028.
- [59] Joseph C. Griffis, Jane B. Allendorfer a Jerzy P. Szaflarski. Voxel-based Gaussian naïve Bayes classification of ischemic stroke lesions in individual T1-weighted MRI scans. *J. Neurosci. Methods*. 2015, 257 97. DOI 10.1016/j.jneumeth.2015.09.019.
- [60] Dorian Pustina, H. Branch Coslett, Peter E. Turkeltaub, Nicholas Tustison, Myrna F. Schwartz a Brian Avants. Automated segmentation of chronic stroke lesions using LINDA: Lesion identification with neighborhood data analysis. *Hum. Brain Mapp.* 2016, 37 (4), 1405–1421. DOI 10.1002/hbm.23110.
- [61] Shih-Cheng Huang, Anuj Pareek, Saeed Seyyedi, Imon Banerjee a Matthew P. Lungren. Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. *npj Digital Med.* 2020, 3:136. DOI 10.1038/s41746-020-00341-z.
- [62] Tongxue Zhou, Su Ruan a Stephane Canu. A review: Deep learning for medical image segmentation using multi-modality fusion. *Array*. 2019, 3-4 100004. DOI 10.1016/j.array.2019.100004.
- [63] Ozgun Çiçek, Ahmed Abdulkadir, Soeren S. Lienkamp, Thomas Brox a Olaf Ronneberger. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. *arXiv*. 2016. DOI 10.48550/arXiv.1606.06650.
- [64] Fabian Isensee, Marianne Schell, Irada Pflueger, Gianluca Brugnara, David Bonekamp, Ulf Neuberger, Antje Wick, Heinz-Peter Schlemmer, Sabine Heiland, Wolfgang Wick, Martin Bendszus, Klaus H. Maier-Hein a Philipp Kickingereder. Automated brain extraction of multisequence MRI using artificial neural networks. *Hum. Brain Mapp.* 2019, 40 (17), 4952–4964. DOI 10.1002/hbm.24750.
- [65] Deepthi Rajashekhar, Matthias Wilms, M. Ethan MacDonald, Jan Ehrhardt, Pauline Mouches, Richard Frayne, Michael D. Hill a Nils D. Forkert. High-resolution T2-FLAIR and non-contrast CT brain atlas of the elderly. *Sci. Data*. 2020, 7 (56), 1–7. DOI 10.1038/s41597-020-0379-9.
- [66] D. Louis Collins, C. J. Holmes, T. M. Peters a A. C. Evans. Automatic 3-D model-based neuroanatomical segmentation. *Hum. Brain Mapp.* 1995, 3 (3), 190–208. DOI 10.1002/hbm.460030304.
- [67] J. Mazziotta, A. Toga, A. Evans, P. Fox, J. Lancaster, K. Zilles, R. Woods, T. Paus, G. Simpson, B. Pike, C. Holmes, L. Collins, P. Thompson, D. MacDonald, M. Iacoboni, T. Schormann, K. Amunts, N. Palomero-Gallagher, S. Geyer, L. Parsons, K. Narr, N. Kabani, G. Le Goualher, D. Boomsma, T. Cannon, R. Kawashima a B. Mazoyer. A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). *Philos. Trans. R. Soc. London, Ser. B*. 2001, 356 (1412), 1293. DOI 10.1098/rstb.2001.0915.
- [68] Nilearn contributors. *nilearn*.
<https://github.com/nilearn/nilearn>.

- [69] Andras Lasso. *SlicerSegmentEditorExtraEffects*. 2024.
<https://github.com/lassoan/SlicerSegmentEditorExtraEffects>. [Online; accessed 2. Oct. 2024].
- [70] Fabian Isensee, Jens Petersen, Andre Klein, David Zimmerer, Paul F. Jaeger, Simon Kohl, Jakob Wasserthal, Gregor Koehler, Tobias Norajitra, Sebastian Wirkert a Klaus H. Maier-Hein. nnU-Net: Self-adapting Framework for U-Net-Based Medical Image Segmentation. *arXiv*. 2018. DOI 10.48550/arXiv.1809.10486.
- [71] Olaf Ronneberger, Philipp Fischer a Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv*. 2015. DOI 10.48550/arXiv.1505.04597.
- [72] Fabian Isensee, Tassilo Wald, Constantin Ulrich, Michael Baumgartner, Saikat Roy, Klaus Maier-Hein a Paul F. Jaeger. nnU-Net Revisited: A Call for Rigorous Validation in 3D Medical Image Segmentation. *arXiv*. 2024. DOI 10.48550/arXiv.2404.09556.
- [73] Fabian Isensee a Klaus H. Maier-Hein. An attempt at beating the 3D U-Net. *arXiv*. 2019. DOI 10.48550/arXiv.1908.02182.
- [74] Fernando Perez-Garcia, Rachel Sparks a Sebastien Ourselin. TorchIO: A Python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning. *Comput. Methods Programs Biomed.* 2021, 208 106236. DOI 10.1016/j.cmpb.2021.106236.
- [75] Raian V. Maretto, Leila M. G. Fonseca, Nathan Jacobs, Thales S. Korting a Leonardo L. Parente. Spatio-Temporal Deep Learning Approach to Map Deforestation in Amazon Rainforest. *IEEE Geosci. Remote Sens. Lett.* 2020, PP (99), 1–5. DOI 10.1109/LGRS.2020.2986407.
- [76] Mitchell Wortsman, Gabriel Ilharco, Samir Yitzhak Gadre, Rebecca Roelofs, Raphael Gontijo-Lopes, Ari S. Morcos, Hongseok Namkoong, Ali Farhadi, Yair Carmom, Simon Kornblith a Ludwig Schmidt. Model soups: averaging weights of multiple fine-tuned models improves accuracy without increasing inference time. *arXiv*. 2022. DOI 10.48550/arXiv.2203.05482.
- [77] The MONAI Consortium. Project MONAI. *Zenodo*. 2020. DOI 10.5281/zenodo.4323059.

Příloha A

Matice predikcí testovaných metod

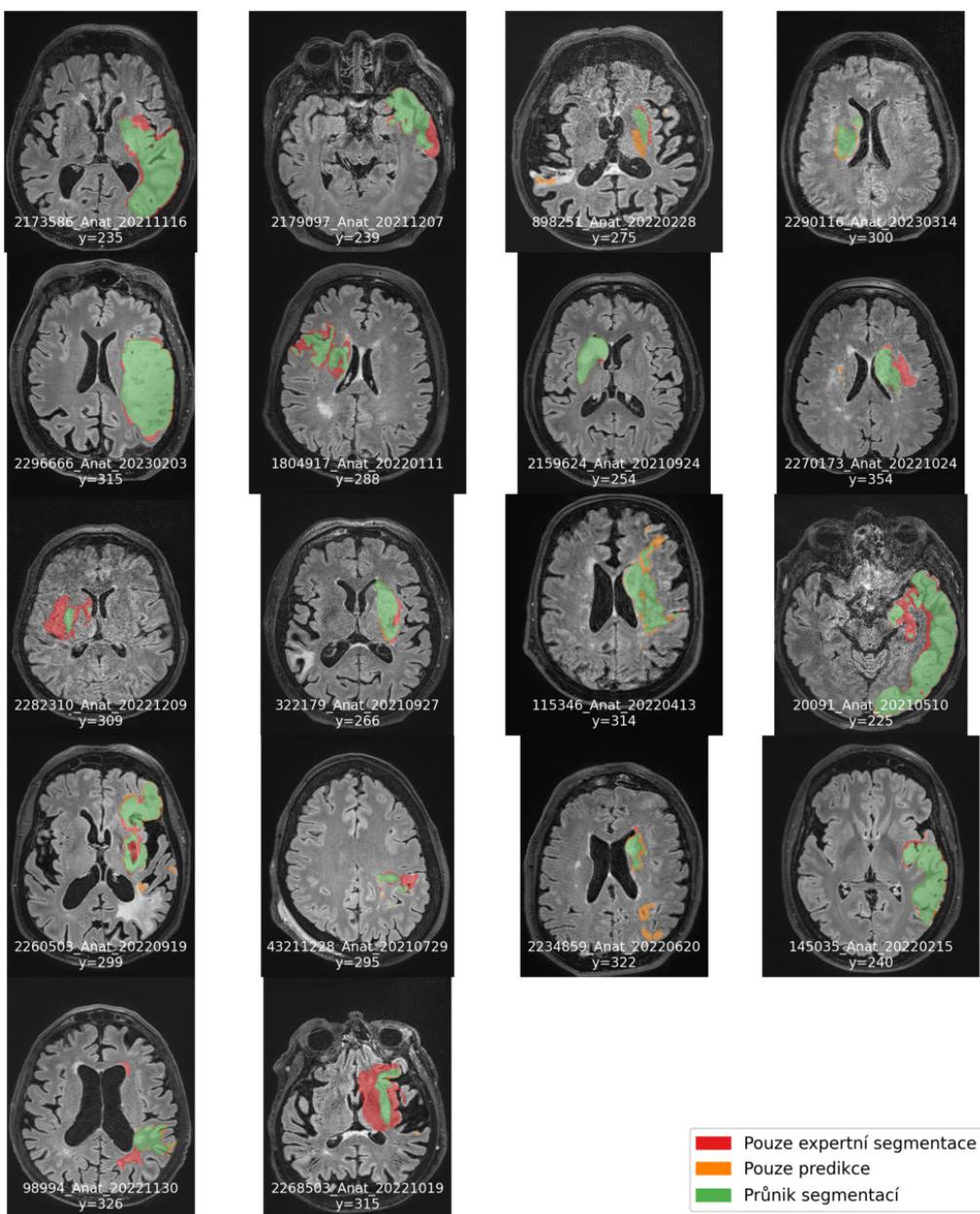


Obrázek A.1. Matice predikcí nejlepších modelů každé z testovaných metod. nnU-Net (Motol): Metoda nnU-Net model *MotolOnly*. nnU-Net: Metoda nnU-Net model *ensemble_FLAIR_DWI_MNI*. DeepMedic: Metoda DeepMedic model *ensemble_DWI_FLAIR*. 3D U-Net: Metoda 3D U-Net model *ensemble_FLAIR_DWI_LabelSampler*.

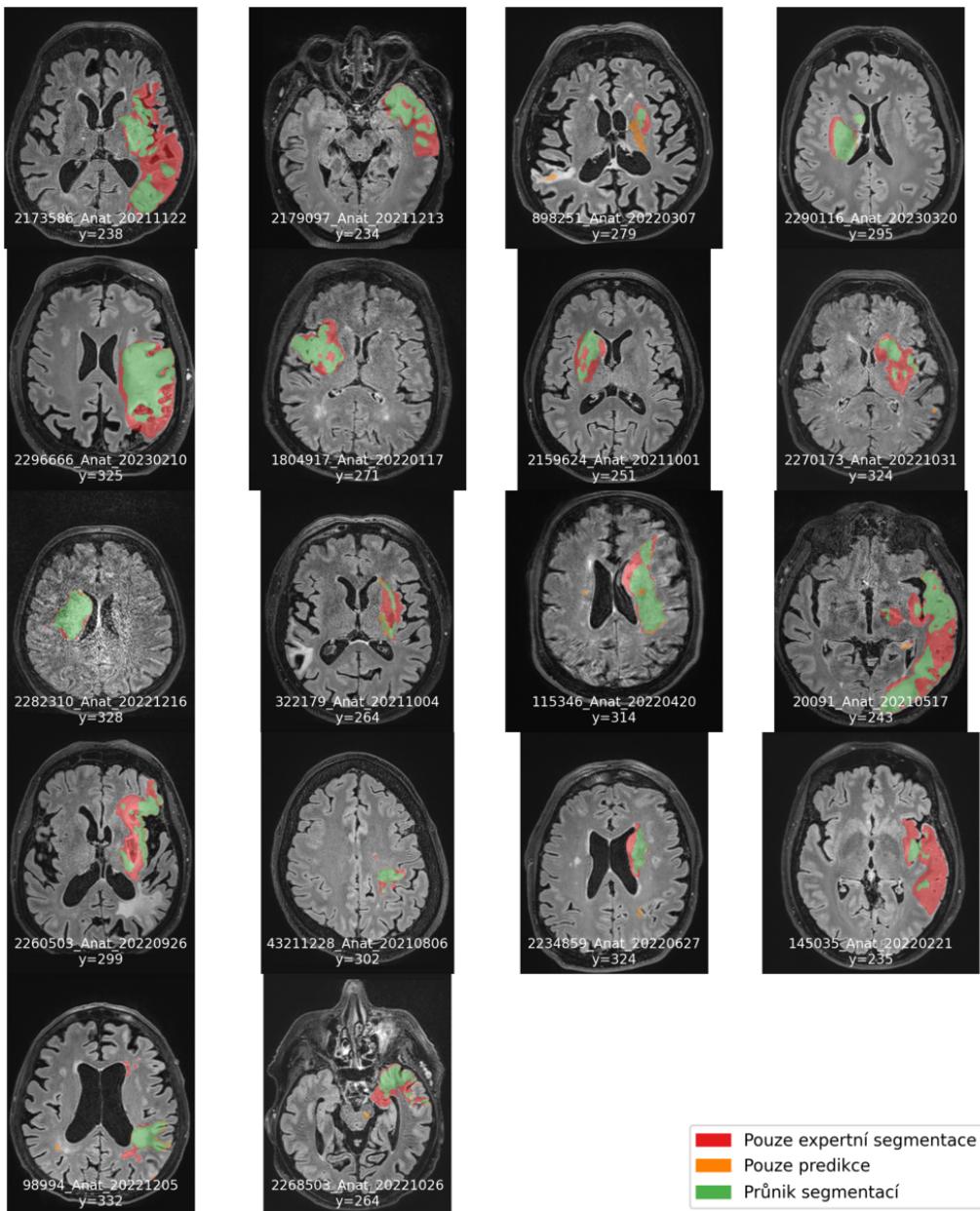
Příloha B

Predikce metody DeepMedic

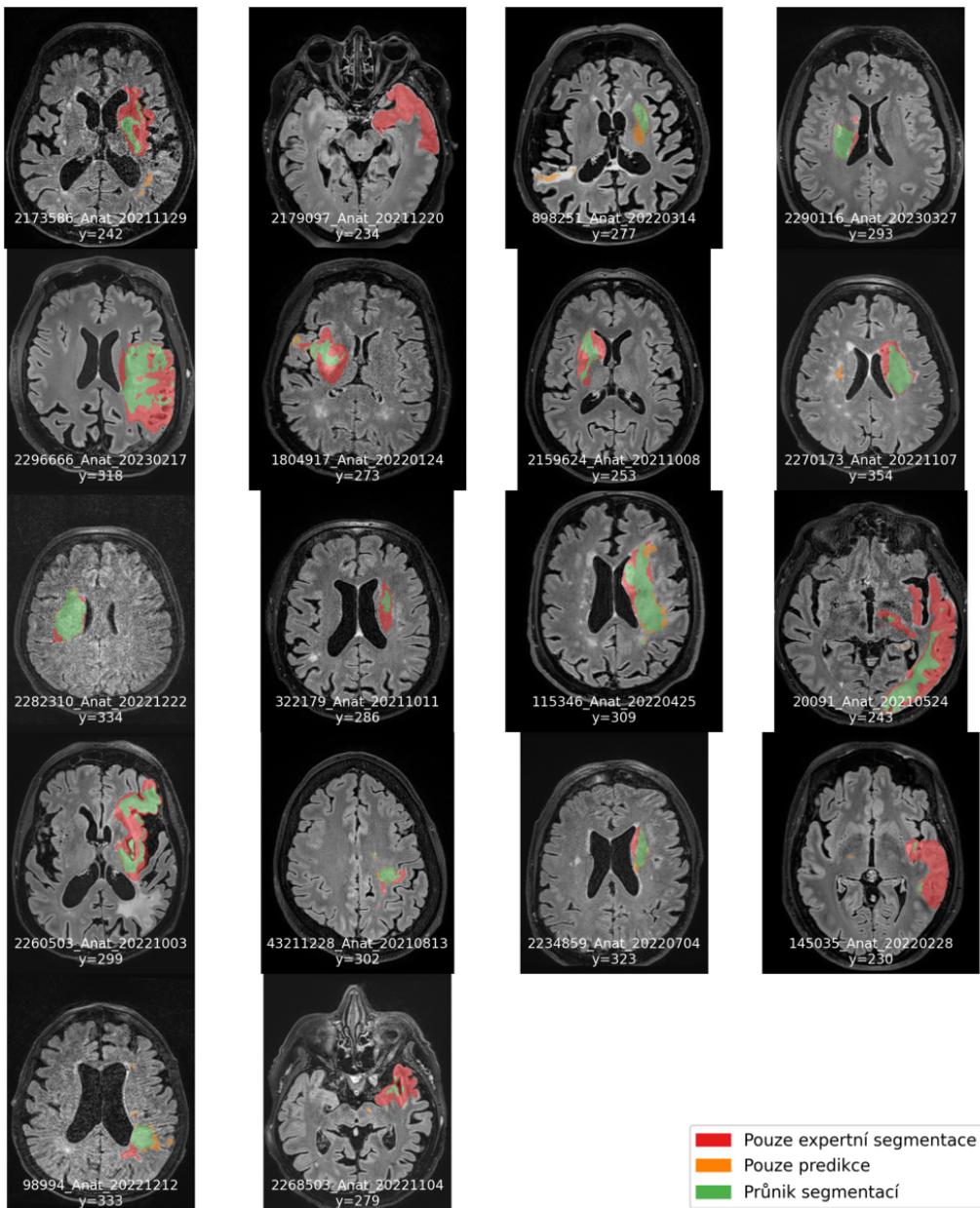
Na následujících obrázcích se nachází predikce *ensemble_DWI_FLAIR*. Jedná se o predikce, které dosáhly nejlepšího Dice koeficientu ze všech testovaných modelů metody DeepMedic.



Obrázek B.2. Predikce instance modelu *ensemble_DWI_FLAIR* metody DeepMedic. Predikce byly vytvořeny na pacientech datasetu Motol v 1. sběru dat. Zobrazení transversálního řezu v místě s maximální plochou expertní segmentace.



Obrázek B.3. Predikce instance modelu *ensemble_DWI_FLAIR* metody DeepMedic. Predikce byly vytvořeny na pacientech datasetu Motol v 2. sběru dat. Zobrazení transversálního řezu v místě s maximální plochou expertní segmentace.

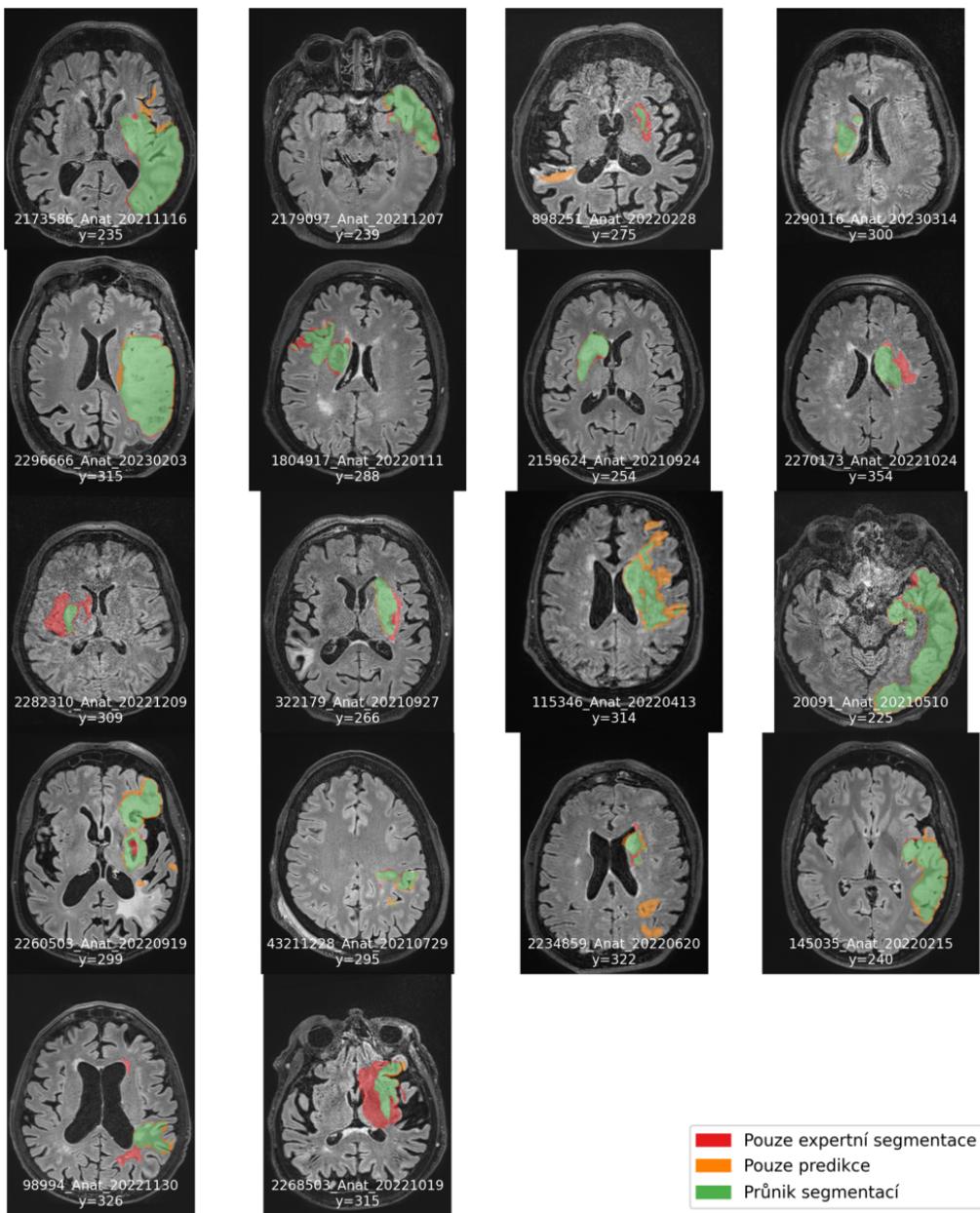


Obrázek B.4. Predikce instance modelu *ensemble_DWI_FLAIR* metody DeepMedic. Predikce byly vytvořeny na pacientech datasetu Motol v 3. sběru dat. Zobrazení transversálního řezu v místě s maximální plochou expertní segmentace.

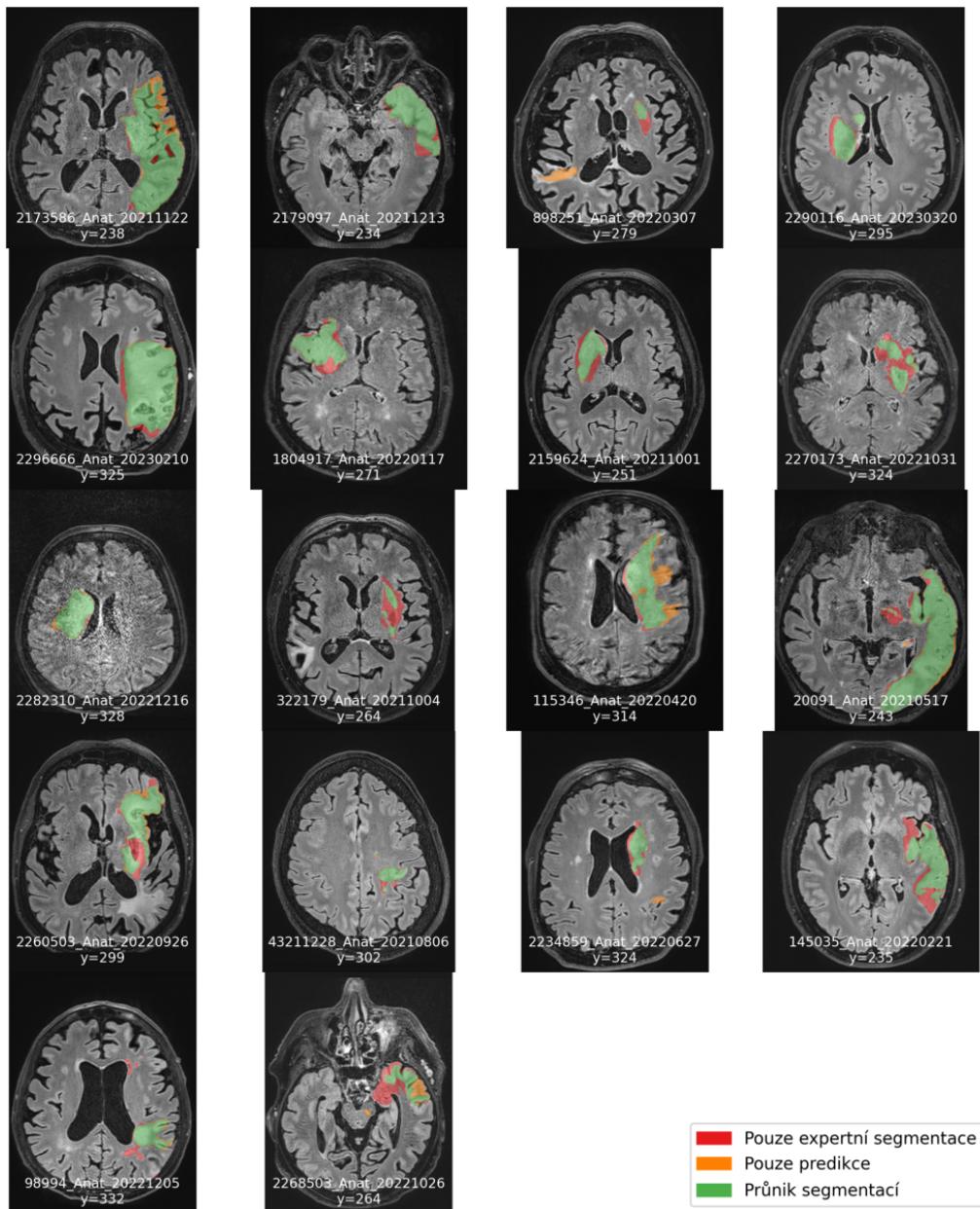
Příloha C

Predikce metody nnU-Net

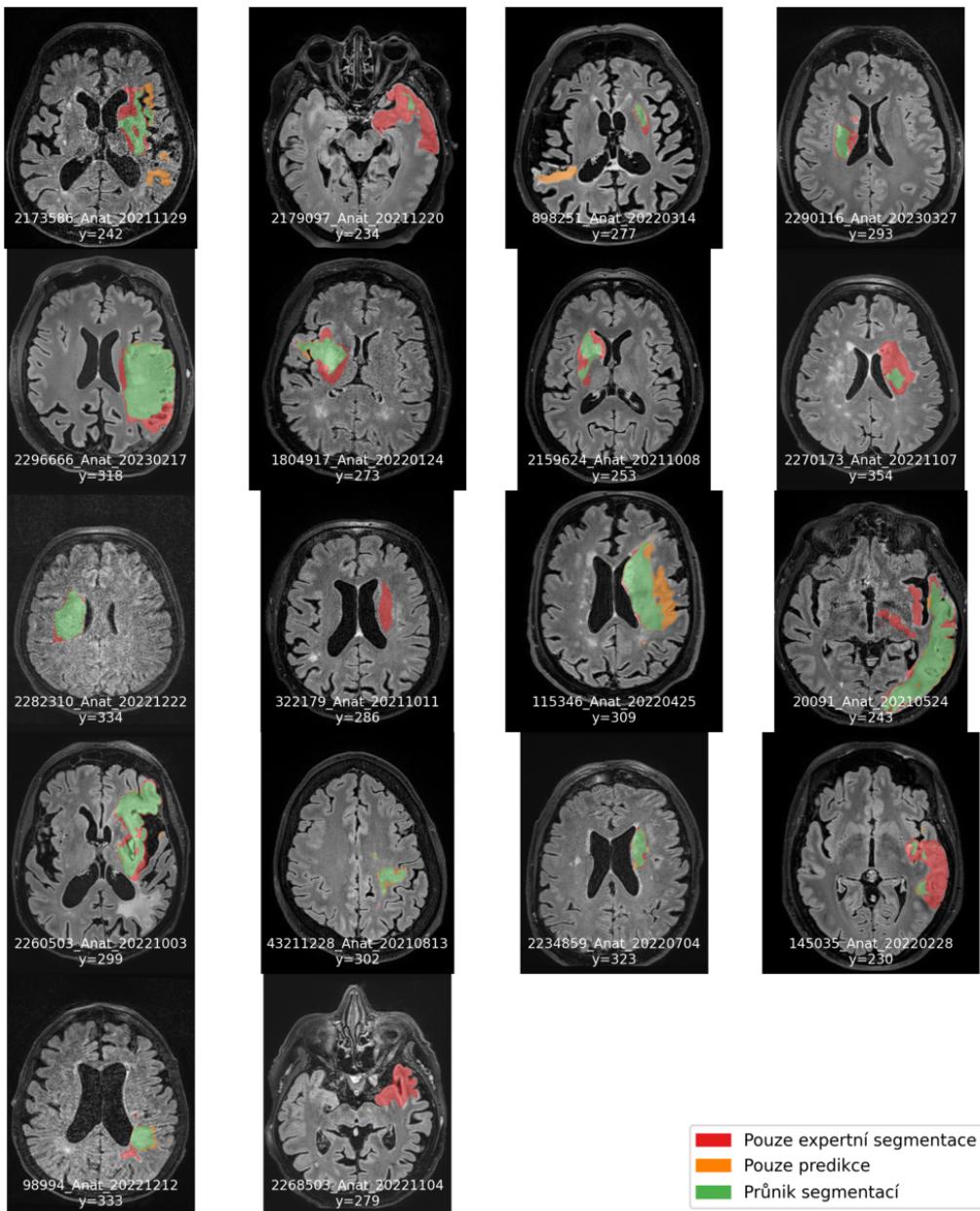
Na následujících obrázcích se nachází predikce *ensemble_FLAIR_DWI_MNI*. Jedná se o predikce, které dosáhly nejlepšího Dice koeficientu ze všech testovaných modelů metody nnU-Net natrénovaných na ISLES.



Obrázek C.5. Predikce instance modelu *ensemble_FLAIR_DWI_MNI* metody nnU-Net. Predikce byly vytvořeny na pacientech datasetu Motol v 1. sběru dat. Zobrazení transverzálního řezu v místě s maximální plochou expertní segmentace.



Obrázek C.6. Predikce instance modelu *ensemble_FLAIR_DWI_MNI* metody nnU-Net. Predikce byly vytvořeny na pacientech datasetu Motol v 2. sběru dat. Zobrazení transverzálního řezu v místě s maximální plochou expertní segmentace.

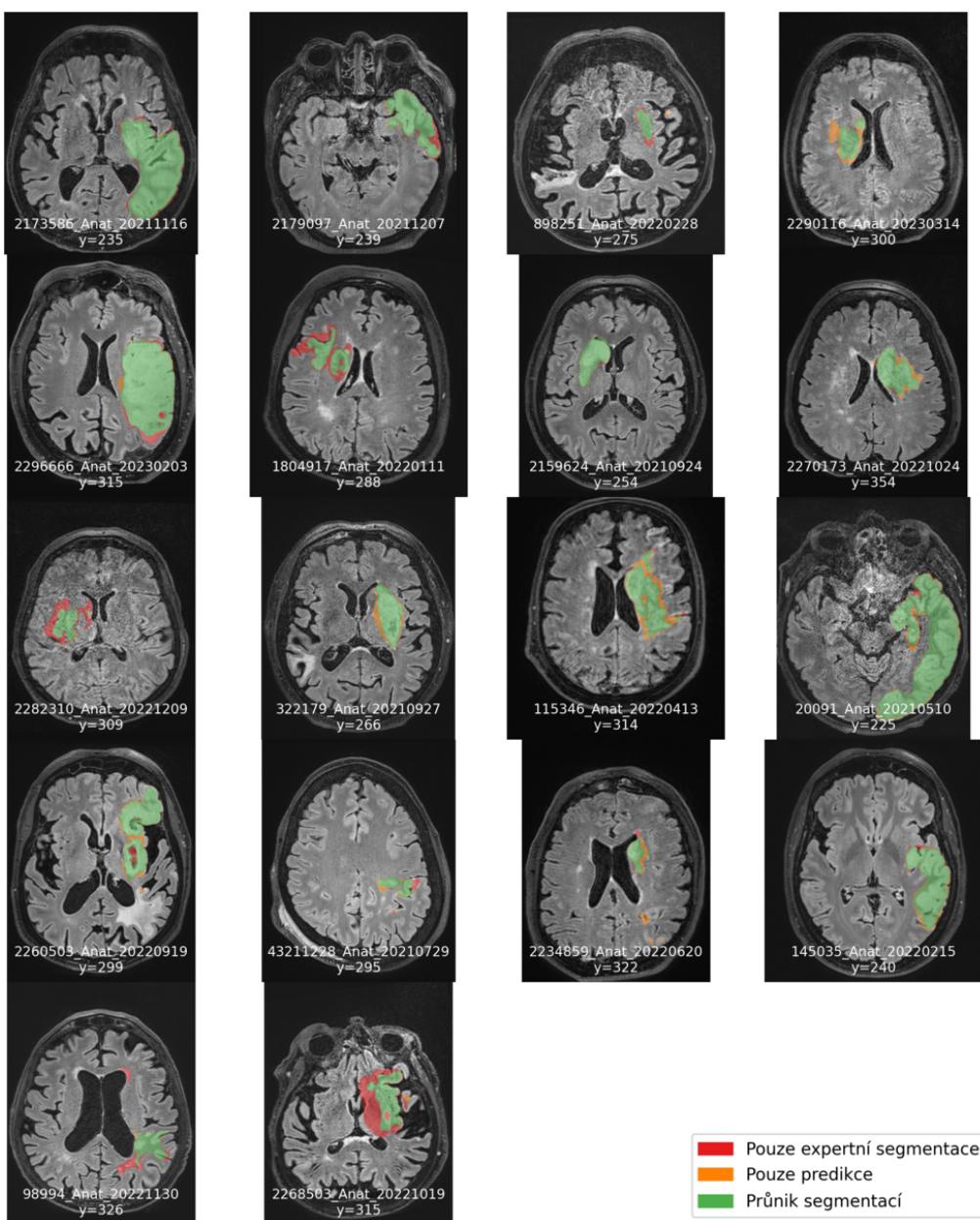


Obrázek C.7. Predikce instance modelu *ensemble_FLAIR_DWI_MNI* metody nnU-Net. Predikce byly vytvořeny na pacientech datasetu Motol v 3. sběru dat. Zobrazení transverzálního řezu v místě s maximální plochou expertní segmentace.

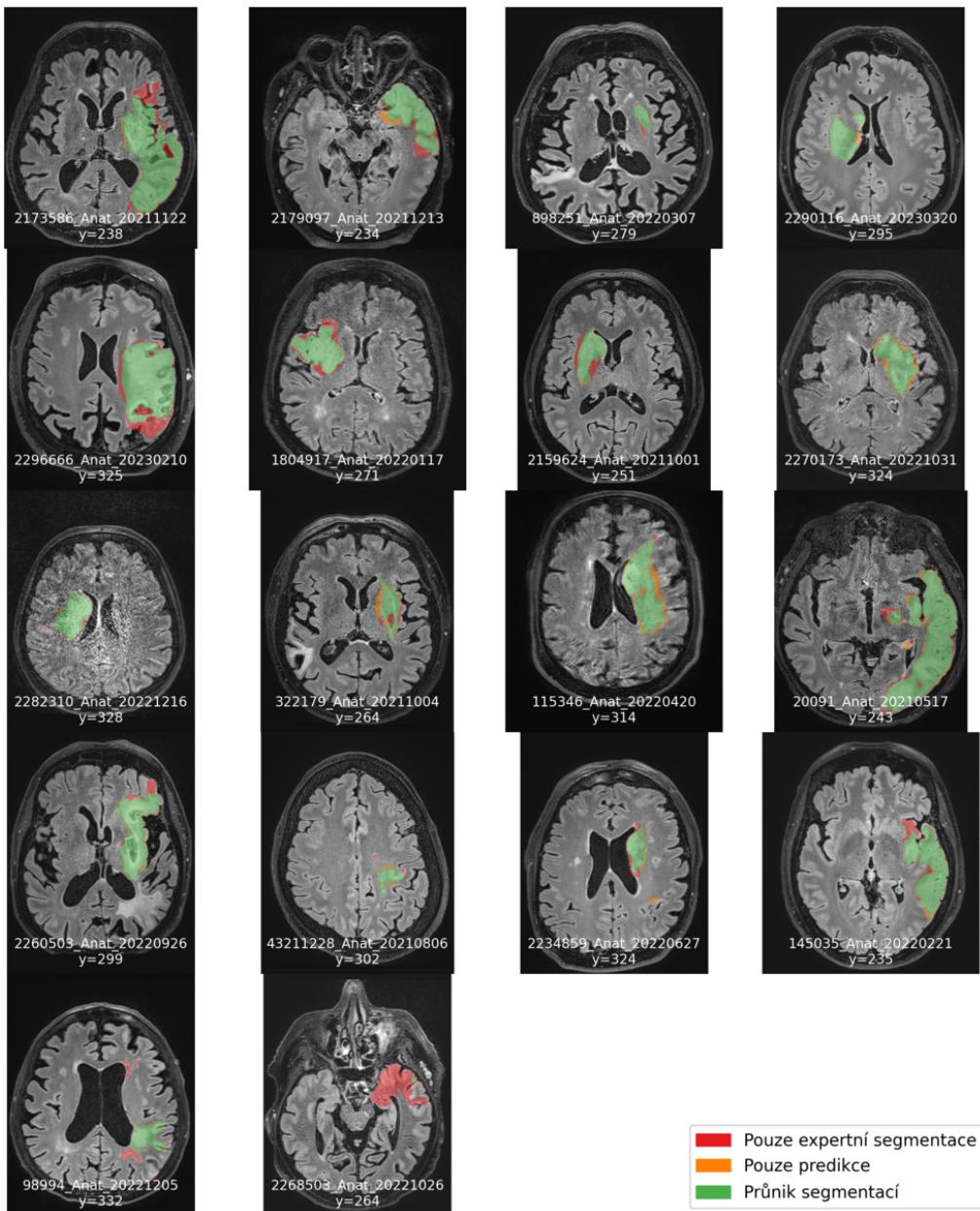
Příloha D

Predikce metody nnU-Net (Motol)

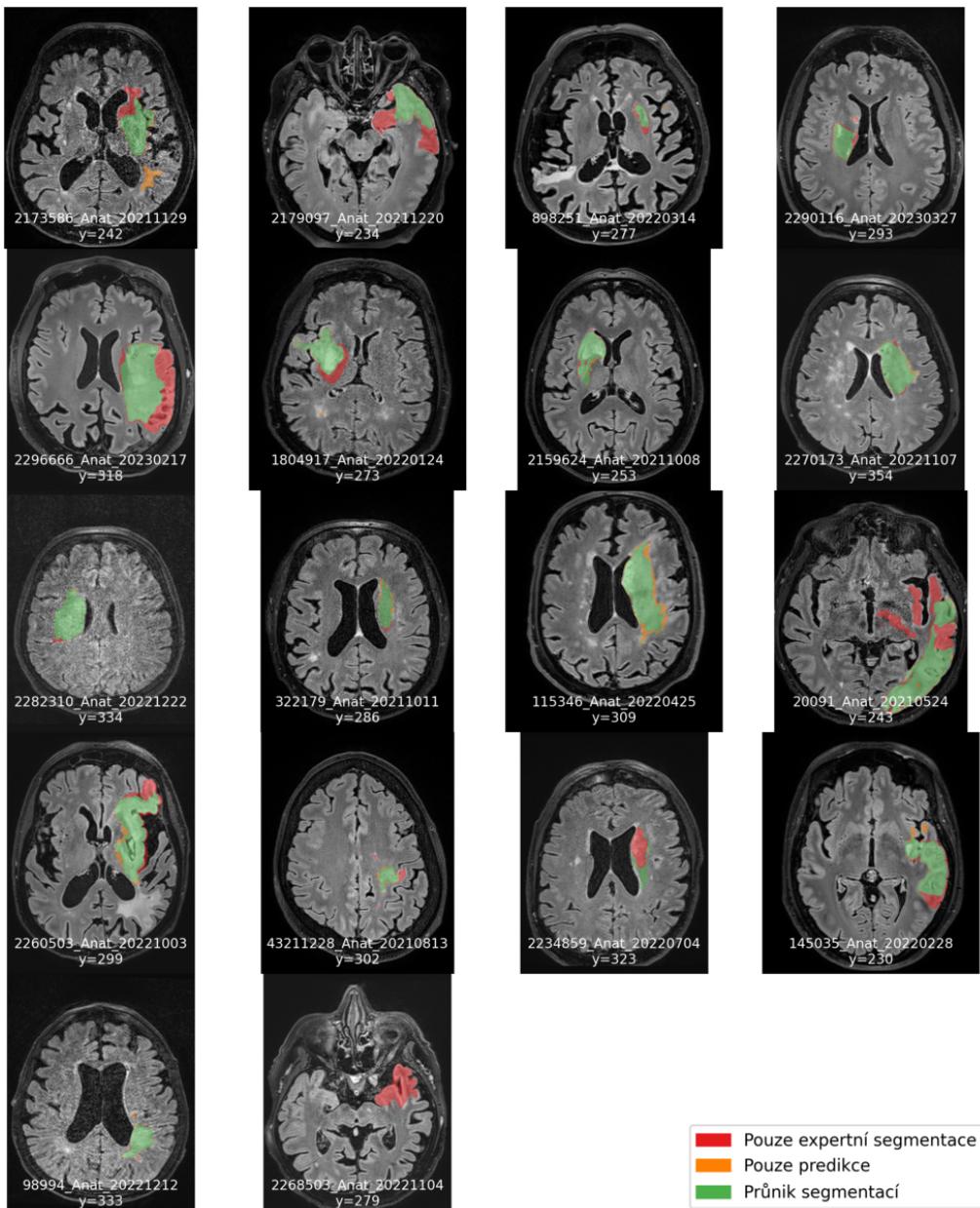
Na následujících obrázcích se nachází predikce natrénovaného modelu *MotolOnly*. Predikce jsou generovány modelem natrénovaným pouze na zhruba 4/5 datasetu Motol.



Obrázek D.8. Predikce instance modelu *MotolOnly* metody nnU-Net. Predikce byly vytvořeny na pacientech datasetu Motol v 1. sběru dat. Zobrazení transversálního řezu v místě s maximální plochou expertní segmentace.



Obrázek D.9. Predikce instance modelu *MotolOnly* metody nnU-Net. Predikce byly vytvořeny na pacientech datasetu Motol v 2. sběru dat. Zobrazení transversálního řezu v místě s maximální plochou expertní segmentace.

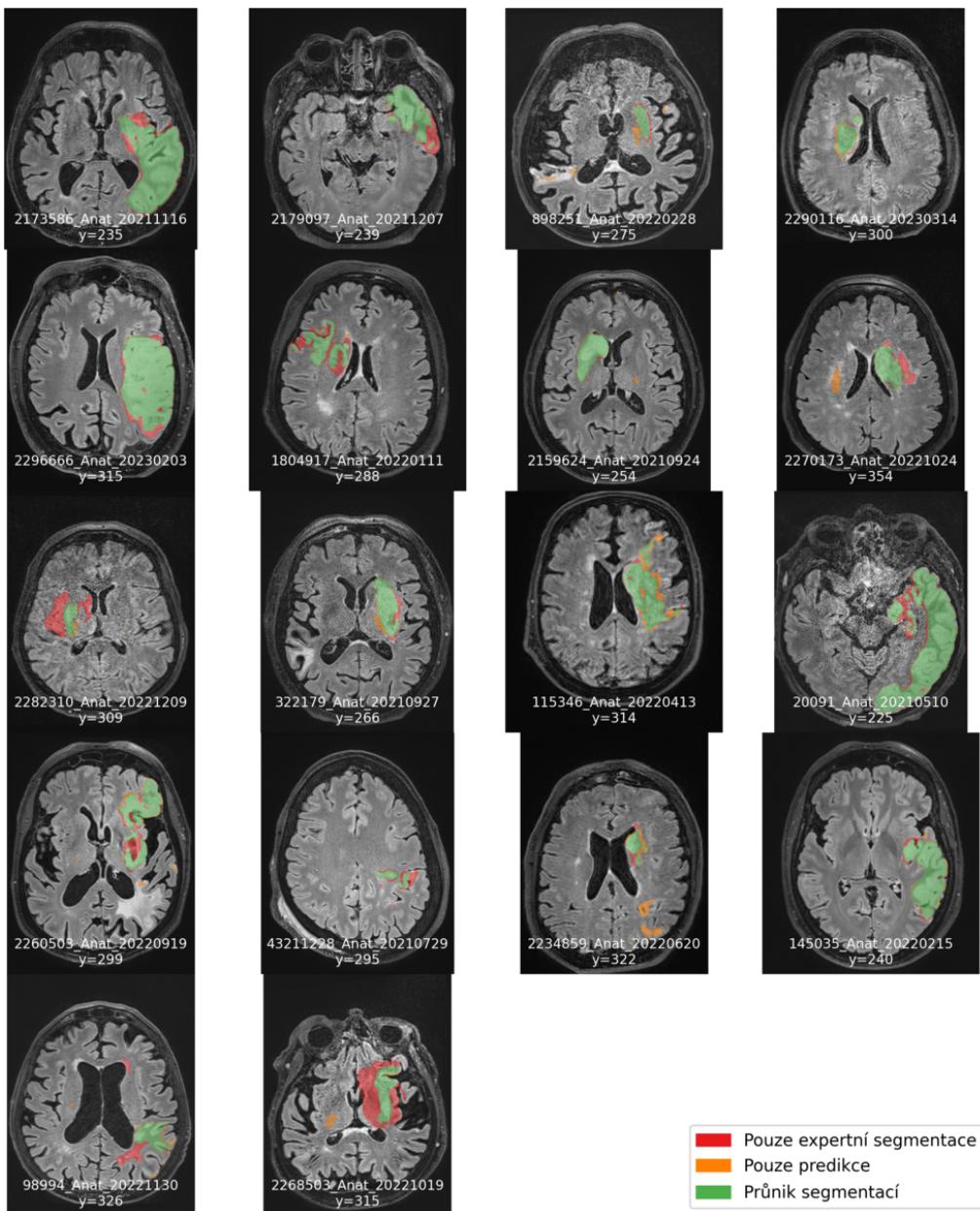


Obrázek D.10. Predikce instance modelu *MotolOnly* metody nnU-Net. Predikce byly vytvořeny na pacientech datasetu Motol v 3. sběru dat. Zobrazení transversálního řezu v místě s maximální plochou expertní segmentace.

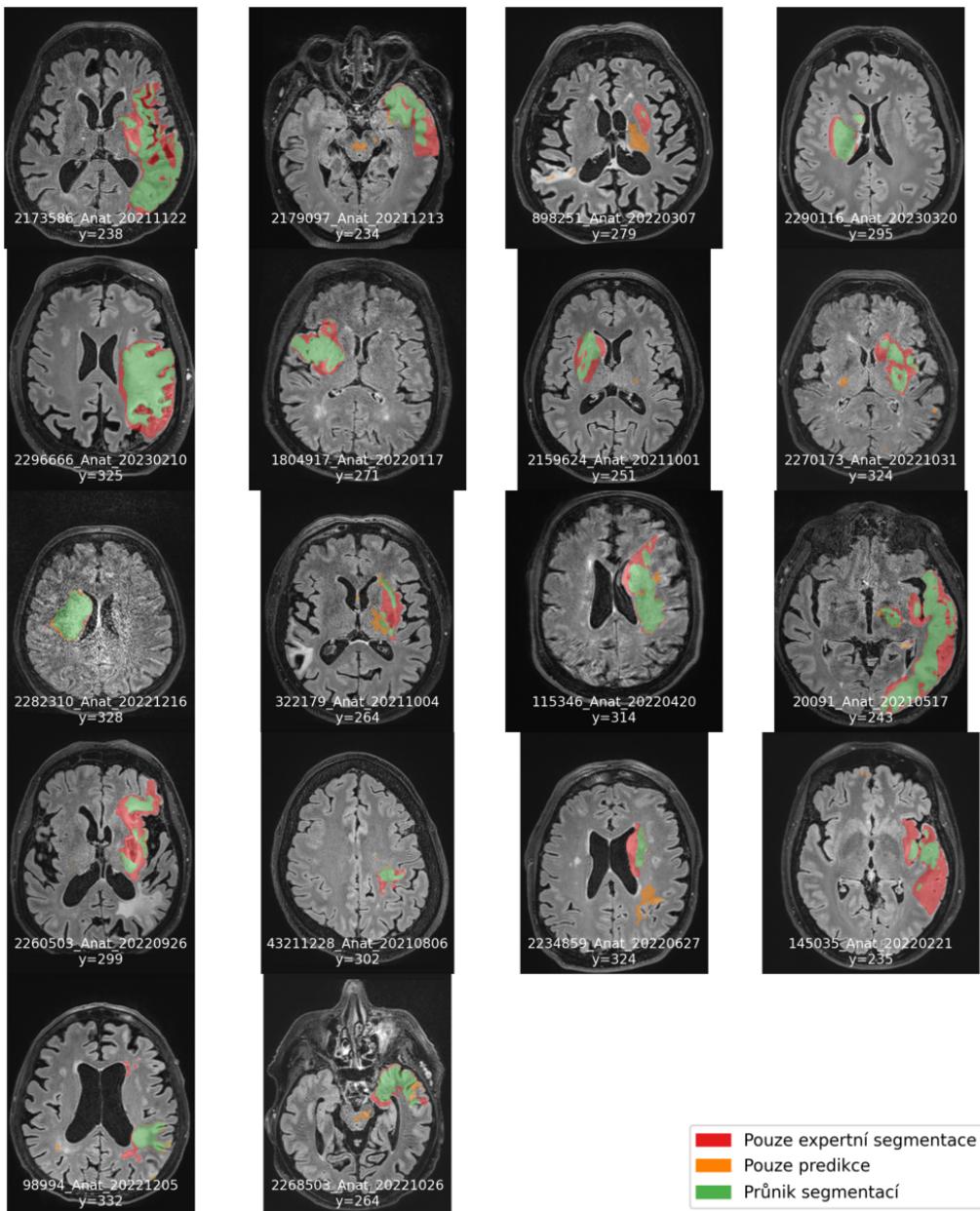
Příloha E

Predikce architektury 3D U-Net

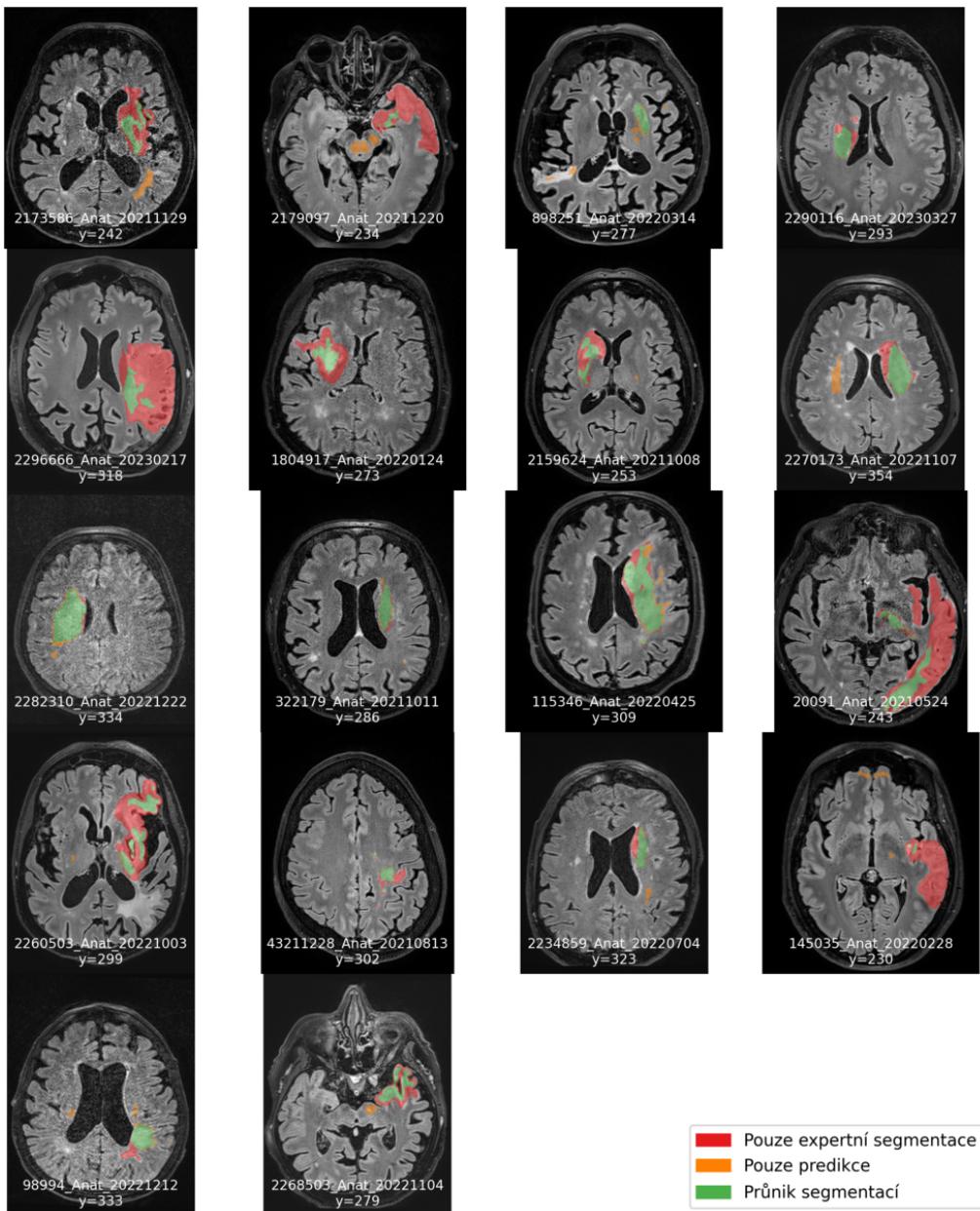
Na následujících obrázcích se nachází predikce *ensemble_FLAIR_DWI_LabelSampler*. Jedná se o predikce, které dosáhly nejlepšího Dice koeficientu ze všech testovaných modelů architektury 3D U-Net.



Obrázek E.11. Predikce instance modelu *ensemble_FLAIR_DWI_LabelSampler* architektury 3D U-Net. Predikce byly vytvořeny na pacientech datasetu Motol v 1. sběru dat. Zobrazení transversálního řezu v místě s maximální plochou expertní segmentace.



Obrázek E.12. Predikce instance modelu *ensemble_FLAIR_DWI_LabelSampler* architektury 3D U-Net. Predikce byly vytvořeny na pacientech datasetu Motol v 2. sběru dat. Zobrazení transversálního řezu v místě s maximální plochou expertní segmentace.



Obrázek E.13. Predikce instance modelu *ensemble_FLAIR_DWI_LabelSampler* architektury 3D U-Net. Predikce byly vytvořeny na pacientech datasetu Motol v 3. sběru dat. Zobrazení transversálního řezu v místě s maximální plochou expertní segmentace.