

Diplomový projekt

# Automatická detekce ischemické léze z MR obrazových dat u cévní mozkové příhody

**Autor:**  
Jakub Šmíd

**Vedoucí:**  
prof. MUDr. Jakub Otáhal, Ph.D.  
Ing. David Kala

8. února 2024

## 1 Popis projektu

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 resonance. V současné klinické praxi se segmentace provádí manuálním obkreslováním a kvůli tomu je velmi časově 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 neurovědcí z EpiReC zabývající se dlouhodobě problematikou cévní mozkové příhody a souvisejících onemocnění.

Téma propojuje technické znalosti (zpracování obrazu, návrh algoritmů) s lékařským prostředím neurověd.

## 2 Evaluační metriky

Před tím, než se začneme věnovat publikacím a aktuálnímu stavu poznání, je nutné uvést metriky, na základě kterých budeme moci vyhodnotit úspěšnost segmentačních algoritmů. Díky uvedeným metrikám je možné srovnat kvalitu dvou segmentací stejného vzoru - typicky segmentaci experta (radiologa) a automatickou segmentaci.

Pro porovnání dvou segmentací jsou používány následující míry: dice similarity coefficient, average symmetric surface distance. Pro vyhodnocení **over- a under-segmentation** se používají metriky známe pod pojmy precision a recall. [15, 7]

## 2.1 Dice similarity coefficient

Dice similarity coefficient dále označovaný jako DC je mírou přesnosti segmentace.

$$DC = 2 \frac{|A \cap B|}{|A| + |B|} \quad (1)$$

Vztah pro výpočet DC je dán rovnicí 1, kde A a B jsou dvě poměřované segmentace. DC koeficient nabývá hodnot mezi 0 a 1 a popisuje překryv mezi jednotlivými segmentacemi. Porovnávání dvou různých DC je citlivé na velikost segmentovaného prostoru (léze). DC dosahuje vyšších hodnot u segmentací s větším objemem.

## 2.2 Average symetric surface distance

Pro výpočet average symetric surface distance (dále ASSD) je nutné nejprve vypočítat tzv. average surface distance (ASD).

$$ASD(A_S, B_S) = \frac{\sum_{a \in A_S} \min_{b \in B_S} d(a, b)}{|A_S|} \quad (2)$$

ASD je definován vztahem 2, kde  $d(a, b)$  je 3D matice obsahující Euklidovské vzdálenosti mezi segmentacemi A a B. Potom ASSD se vypočítá podle vztahu 3.

$$ASSD(A_S, B_S) = \frac{ASD(A, B) + ASD(B, A)}{2} \quad (3)$$

ASSD je tedy míra vyjadřující průměrnou vzdálenost všech Euklidovských vzdáleností mezi jednotlivými osegmentovanými objemy. Standardně se uvádí v milimetrech a nižší ASSD znamená lepší úspěšnost segmentace - resp. segmentace jsou si vzájemně podobnější.

## 2.3 Hausdorffská vzdálenost

Hausdorffská vzdálenost (Hausdorff's distance, dále HD) je míra vyjadřující maximální vzdálenost mezi odpovídajícími body, které tvoří hranici jednotlivých segmentací. Tato míra je tedy citlivá na **outliery**.

$$HD(A_S, B_S) = \max\left\{\max_{a \in A_S} \min_{b \in B_S} d(a, b), \max_{b \in B_S} \min_{a \in A_S} d(b, a)\right\} \quad (4)$$

HD je definována vztahem 4, a podobně jako ASSD se i HD běžně uvádí v milimetrech a nižší vzdálenost HD znamená lepší úspěšnost.

## 2.4 Precision

## 2.5 Recall

## 2.6 Leave-One-Out Cross-Validation

# 3 Doporučená literatura

## 3.1 ISLES 2015

Benchmarková studie [15] zmiňuje problém komparability jednotlivých algoritmů pro segmentaci ischemické léze z důvodu různorodosti používaných datasetů a evaluačním procesům. Tento problém se autoři rozhodli vyřešit na soutěži Ischemic Stroke Lesion Segmentation challenge (ISLES) v roce 2015. Článek představuje evaluační proces, zmiňuje veřejně dostupné datasety a hlavně nabízí srovnání výsledků z ISLES, které se zúčastnilo 16 vědeckých týmů.

Soutěž byla rozdělena na dvě kategorie. První z nich řešila Stroke Perfusion Estimation (SPES) - řešitelé se zaměřili na interpretaci obrazů akutní fáze mozkové mrtvice (0-24 hodin). Druhou kategorií byla Sub-acute Ischemic Stroke lesion Segmentation (SISS), kde bylo cílem analyzovat obrazy z pozdějších fází mrtvice (24 hodin - 2 týdny). Pro účely této práce je důležitá hlavně kategorie SISS, proto se dále budu věnovat pouze jí.

Data pro SISS byla nasbírána ze dvou zdravotnických zařízení, byla pořízena na stroji s magnetickým polem 3 T. Snímky jednotlivých pacientů ale byly pořízeny s různými parametry, což zapříčinilo, že mají různé rozlišení a různou míru artefaktů. Celkem bylo do datasetu zahrnuto 64 pacientů.

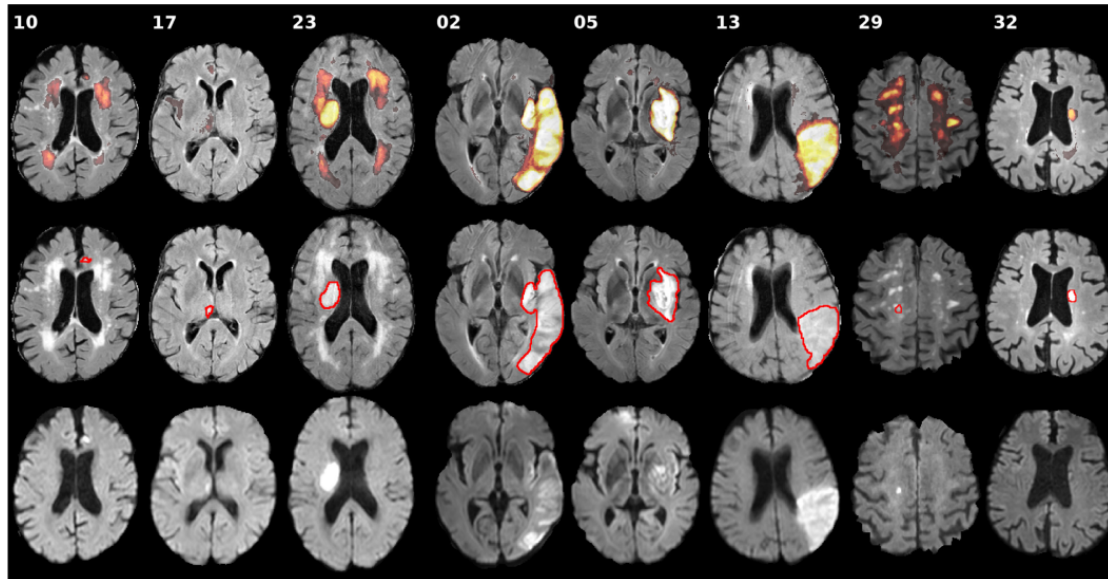
Snímky pacientů, kteří byli zařazeni do datasetu, musely obsahovat alespoň sekvence T1-vážený, T2-vážený, DWI ( $b = 1000 \text{ s/mm}^2$ ) a FLAIR. V rámci preprocessingu autoři soutěže provedli skull-stripping pomocí algoritmu BET2 [8] a provedli manuální korekturu. Pomocí spline jsou transformovány všechny snímky na izotropní velikost voxelu  $1 \text{ mm}^2$ . Byla také provedena rigidní koregistrace na snímky FLAIR pomocí Elastix [12].

Expertní segmentace byla provedena pouze na sekvenci FLAIR, jelikož tato sekvence vykazuje nejnižší variabilitu při segmentování mezi různými experty. Zbylá data poskytovala pouze podpůrnou informaci. Kromě pravidla, že byla segmentována pouze sub-akutní ischemická léze se experti řídili dalšími pravidly: částečně obklopené fissury nejsou segmentovány, ale malé nebo z velké části obklopené fissury jsou osegmentovány. Segmentace nejsou duté (nemají díry) a mají hladký tvar ("prostorově konzistentní").

Průměrný počet lézí na snímku je 2.46 s tím, že jsou v datasetu zahrnuty snímky s 1 až 14 lézemi. Průměrný objem léze je 17.59 ml, minimum je 1.00 ml a maximální objem léze je 346.06 ml. Snímky pacientů byly rozděleny následovně: 28 pacientů pro trénink a 36 pro testování. Pro testovací data nejsou zveřejněny expertní segmentace, které byly pro účely testování pořízeny dvěma experty. Trénovací množina dat je segmentována pouze jedním expertem a obsahuje snímky pouze z jednoho zdravotnického zařízení.

Výsledky kategorie SISS jsou zobrazeny v tabulce 1. Jelikož dataset byl anotován dvakrát - podruhé jiným expertem, obsahuje tabulka výsledků na posledním řádku po-

rovnání těchto dvou segmentací. Vizualizace výsledků je na obrázku 1. Je nutné zmínit, že žádný z navržených algoritmů nedosáhl kladného Dice koeficientu u všech 36 pacientů z testovací sady. To znamená, že u některých snímků nedošlo k překryvu segmentací mezi expertem a algoritmem.



Obrázek 1: Příklady snímků a výsledků z SISS. První řádek zobrazuje distribuci segmentace navržených metod na sekvenci FLAIR, druhý řádek zobrazuje stejný snímek segmentovaný expertem a na třetím řádku se nachází odpovídající DWI sekvence.

Autoři článku uvádí, že nejsou schopni jasně rozhodnout o rozdílech ve výkonnosti navržených metod - a to ani v typech jednotlivých přístupů (deep learning nebo klasické metody). Dvě nejlépe hodnocené metody, vykazovaly statisticky stejnou přesnost: metoda UK-Imp2 (později pojmenovaná DeepMedic) [9] používala hluboké učení - konkrétně **neuronové konvoluční sítě**, metoda CN-Neu [3] byla založená na fuzzy c-means. Třetí nejlepší metoda byla založená na náhodných lesech (**Random Forest**). Přesto nelze říci, že by nějaká z nich pracovala výrazně lépe. Autoři zmiňují, že klíčem k úspěchu je spíše odladění hyperparametrů modelu a jeho adaptace. Nejlepší tři řešení využívaly kombinaci dvou různých algoritmů, aby vykompenzovaly slabá místa každého z nich.

Metody soutěžících byly rovněž otestovány na datech z jiných zdravotnických center. Zajímavým výstupem soutěže je zjištění, že ani jeden z algoritmů nebyl schopen dobře fungovat na cizích datech. Přestože metody vykazovaly dobrou generalizaci na testovacích datech pořízených ze stejného datacentra jako data trénovací. Hypotéza, proč nedocházelo k **overfittingu** je, že kvůli vysokému inter-rater Dice koeficientu (tzn. variabilitě segmentací napříč experty) nebyly schopny se ani modely přeučit na trénovací data.

Automatická segmentace obecně fungovala dobře na velkých lézích se silným FLAIR signálem, naopak byl problém v segmentaci malých lézí se slabým FLAIR signálem. Dal-

Method	Cases	ASSD (mm)	DC [0,1]	HD (mm)
UK-Imp2	34/36	05.96 $\pm$ 09.38	0.59 $\pm$ 0.31	37.88 $\pm$ 30.06
CN-Neu	32/36	03.27 $\pm$ 03.62	0.55 $\pm$ 0.30	19.78 $\pm$ 15.65
FI-Hus	31/36	08.05 $\pm$ 09.57	0.47 $\pm$ 0.32	40.23 $\pm$ 33.17
US-Odu	33/36	06.24 $\pm$ 05.21	0.43 $\pm$ 0.27	41.76 $\pm$ 25.11
BE-Kul2	33/36	11.27 $\pm$ 10.17	0.43 $\pm$ 0.30	60.79 $\pm$ 31.14
DE-UzL	31/36	10.21 $\pm$ 09.44	0.42 $\pm$ 0.33	49.17 $\pm$ 29.6
US-Jhu	33/36	11.54 $\pm$ 11.14	0.42 $\pm$ 0.32	62.43 $\pm$ 28.64
UK-Imp1	34/36	11.71 $\pm$ 10.12	0.44 $\pm$ 0.30	70.61 $\pm$ 24.59
CA-USher	27/36	09.25 $\pm$ 09.79	0.35 $\pm$ 0.32	44.91 $\pm$ 32.53
BE-Kull	30/36	12.24 $\pm$ 13.49	0.37 $\pm$ 0.33	58.65 $\pm$ 29.99
CA-McGill	31/36	11.04 $\pm$ 13.68	0.32 $\pm$ 0.26	40.42 $\pm$ 26.98
SE-Cth	30/36	10.00 $\pm$ 06.61	0.38 $\pm$ 0.28	72.16 $\pm$ 17.32
DE-Dkfz	35/36	14.20 $\pm$ 10.41	0.33 $\pm$ 0.28	77.95 $\pm$ 22.13
TW-Ntust	15/36	07.59 $\pm$ 06.24	0.16 $\pm$ 0.26	38.54 $\pm$ 20.36
Inter-observer	34/36	02.02 $\pm$ 02.17	0.70 $\pm$ 0.20	15.46 $\pm$ 13.56

Tabulka 1: Výsledky jednotlivých týmů na ISLES 2015 pro SISS.

ším problémem byla segmentace WMH (**white matter hyperintensity**), místo ischemické léze. Toto zjištění je podle autorů překvapující, protože WMH není rozpoznatelná na DWI sekvenci (na rozdíl od léze), jelikož má stejnou intenzitu jako okolní tkáň - není viditelná.

Článek hodnotí SISS jako velmi obtížnou úlohu - nejen na základě přesnosti navržených metod, ale i na velkém rozptylu Dice koeficientu napříč experty. Podle autorů by kliničtí výzkumníci neměli očekávat spolehlivé a plně automatické řešení v blízké budoucnosti, jelikož je úloha příliš složitá. Automatické segmentační algoritmy by měly sloužit spíše jako podpůrný nástroj pro radiology.

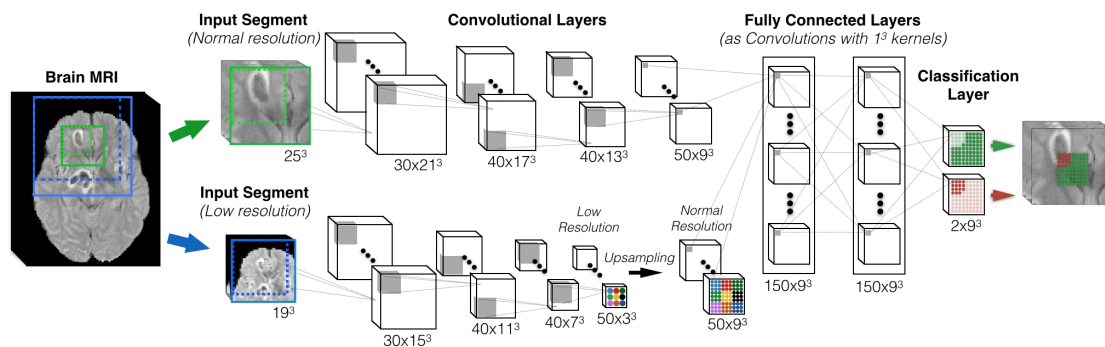
Soutěž ukázala, že aktuální stav automatické segmentace sub-akutní léze postrádá jak přesnost, tak robustnost, která je potřebná pro reálné nasazení. Dalším zjištěním bylo, že žádný z algoritmů nefungoval významně lépe proti ostatním.

Pro budoucí výzkum autoři doporučují kombinovat různé algoritmy pro vylepšení segmentace. Evaluace by nikdy neměla být prováděna pouze na privátním datasetu a měl by být kladen důraz na adaptaci algoritmů na data z jiných center.

### 3.1.1 DeepMedic [9]

Segmentační metoda je založená na 3D konvoluční neuronové síti (CNN) s 11 vrstvami. Dataset, na kterém autoři trénovali neuronovou síť byl augmentován zrcadlením snímků přes sagitální rovinu. V rámci preprocessingu autoři normalizují snímky na nulovou střední hodnotu a rozptyl 1, to se společně s augmentací děje paralelně s trénováním, tedy je to již implementované v navrhované metodě. Před samotnou segmentací je 3D snímek rozdělen do několika menších 3D patchů, díky tomu autoři zmenší počet tréno-

vaných parametrů sítě. Následně vstupuje daný 3D batch do dvoucestné konvoluční sítě - do první cesty v plném rozlišení a do druhé cesty v nízkém rozlišení. Snížení rozlišení je realizováno podsamplováním původního skenu. Díky tomu batch vstupující do druhé cesty obsahuje širší okolí než batch vstupující do cesty první. Poté jsou obě cesty spojeny a příznaky vstupují do fully connected vrstev (FC). Nákres metody se nachází na Obrázku 2.



Obrázek 2: Jedna z prvotních verzí DeepMedic, která ilustruje architekturu sítě a přístup autorů k problému. V levé části vidíme snímek s více kanály (modalitami) vstupující do sítě, tento snímek je rozdělen do několika menších patchů s různým rozlišením. Architektura je čistě konvoluční a na konci výstup jader vstupuje do Fully Connected vrstev, které jsou realizovány jako konvoluce s jednotkovým kernelem.

Autoři používají malá konvoluční jádra a tím snižují nároky na GPU paměť. Přestože architektura pracuje na 3D snímcích, lze ji natrénovat na grafické kartě, která má pouze 3 GB paměti.

Tato metoda po vyhodnocení ISLES 2015 byla a je stále aktivně vyvíjena. V roce 2017 DeepMedic byl součástí výherního řešení soutěže BRATS 2017, jejímž cílem je segmentace nádorů na mozku. [10, 11]

### 3.1.2 CN-Neu (Chaolu Feng) [3]

Tato metoda nepotřebuje žádná trénovací data, jelikož k segmentaci používá fuzzy c-means. Pro každou vstupní sekvenci proběhne segmentace na bílou hmotu, šedou hmotu mozkovou, mozkomíšní mok a lézi. Za lézi je následně určen ten voxel, který byl označen za lézi ve FLAIR sekvenci a zároveň alespoň v nějaké další sekvenci. Nakonec autoři upraví hranice léze podle vlastního algoritmu, jež je matematicky formulován v citovaném článku.

## 3.2 Segmentace chronické mrtvice z T1-váženého obrazu

Jelikož většina rehabilitačních metod po mozkové mrtvici závisí na T1-váženém obrazu, rozhodli se autoři článku [7] provést srovnání jedné semi-automatické a tří plně au-

tomatických segmentačních algoritmů lézí z T1-vážených obrazů. Motivací bylo jejich systematické otestování na velkém datasetu.

Chronickou fázi autoři definují jako časové období od několika týdnů po roky od mozkové příhody. Pro rehabilitační účely se z časových a finančních důvodů pacient typicky snímkuje pouze sekvencí T1-w ve vysokém rozlišení. Tato sekvence je zároveň citlivá na zobrazení nekrózy vzniklé od 2 týdnů po příhodě, proto je vhodná k detekci chronických lézí.

Na druhou stranu segmentace akutní mrtvice, kde jsou snímky typicky pořízeny během prvních několika dnů, snímkování zahrnuje multimodální data - typicky DWI, FLAIR, PWI (perfusion weighted imaging). Sběr multimodálních dat přispívá k lepšímu rozhodování o léčbě pacienta v akutní fázi a zároveň k hrubé predikci klinických dopadů mrtvice. Proto se algoritmy zaměřují na rychlou segmentaci a pracují s multimodálními daty. Nicméně autoři článku se dále soustředí pouze na segmentaci, která je určená pro rehabilitační výzkum.

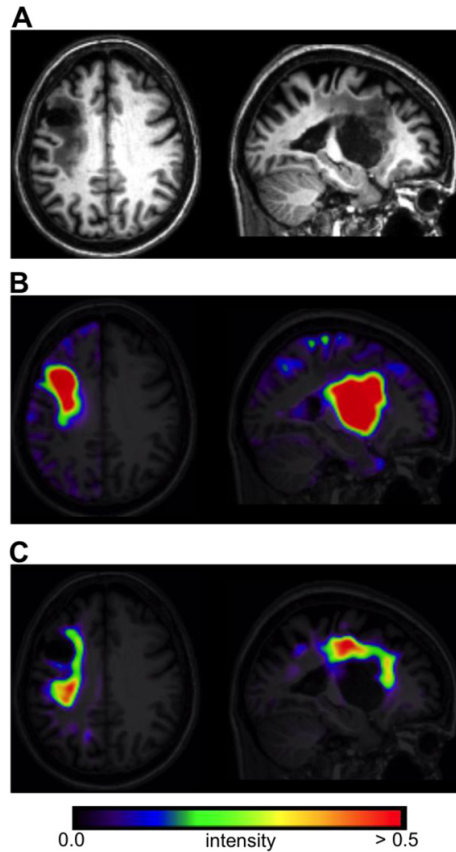
Semiautomatická metoda **Clusterize** byla původně vyvinutá pro sledování ztráty myelinu při leukodystrofii. Ukázalo se však, že metoda založená na analýze T2-w, funguje i pro segmentaci akutní a chronické mrtvice z T1-w. [2] Clusterize nejprve vyhledává lokální maxima intenzity voxelů na každém snímku. Následně přiřadí zbylé voxely podle jejich intenzity do klastrů identifikovaných podle vyhledaných maxim. Poté musí člověk manuálně vybrat v každém řezu požadovaný klaster a případně udělat korekci masky.

První plně automatickou metodou zkoumanou v článku [7] je **Automated Lesion Identification (ALI)** [18], která je založená na detekci outlierů. Outliery jsou získány fuzzy clusteringem (c-means) pravděpodobnostních map bílé a šedé hmoty mozkové pacienta a kontrolních map. Kontrolní mapy jsou vygenerované ze zdravé tkáně - proto je třeba vlastnit dataset obsahující zdravé MRI snímky. Segmentace tkání je realizována iterativně softwarem **SPM**, autoři uvádí novou třídu tkáně nazvanou "extra", která je průměrem již osegmentované pravděpodobnostní mapy bílé hmoty a mozkomíšního moku. Tato nová třída tkáně vstupuje do další iterace SPM segmentace a napomáhá ke správné segmentaci bílé a šedé hmoty. Po poslední iteraci se mapy bílé a šedé hmoty mozkové gaussovsky vyhladí a na takovýchto datech se hledají výše popsáním způsobem outliery.

Druhá automatická metoda [4] implementuje **Gaussovský naivní Bayes** klasifikátor. Pomocí softwaru SPM jsou získány pravděpodobnostní mapy tkání (bílá hmota, šedá hmota a mozkomíšního moku). Tyto mapy jsou společně s originálním T1-w obrazem transformovány do prostoru atlasu Montreal Neurological Institute a vyhlazeny gaussovským kernelem. Poté jsou z pravděpodobnostních map vygenerovány dvě mapy příznaků zobrazené na obrázku 3. Pro vygenerování map je potřebná znalost, v jaké hemisféře se léze nachází. První příznaková mapa  $F_1$  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/šedou hmotu mozkovou. Druhá příznaková mapa  $F_2$  obsahuje informaci o abnormální tkáni a vztah vychází z faktu, že tato tkáň má podobné hodnoty T1-w jako šedá hmota. Mapy příznaků jsou následně použity pro natrénování naivního Bayes klasifikátoru. Bylo po-



užito 30 snímků pacientů a trénování probíhalo metodou, kdy jeden z pacientů sloužil jako validační data a zbytek jako trénovací, v další učící iteraci byl vybrán jiný pacient jako validační a zbylé snímky pacientů sloužily opět jako trénovací data. Tato metoda se v odborně označuje jako leave-one-case-out cross validation. Pro získání segmentace byl výstup thresholdován.



Obrázek 3: A zobrazuje T1-w obraz jednoho z pacientů. B znázorňuje příznaky  $F_1$ , tedy chybějící tkáň pro stejného pacienta. C znázorňuje příznaky  $F_2$ , tedy abnormální tkáň pro stejného pacienta.

Poslední zkoumanou metodou byla **LINDA** [17], kterou tvoří série **náhodných lesů**. Metoda se skládá ze tří náhodných lesů, kdy každý z nich je navržen na jiné rozlišení dat. Vstupní MRI obraz je nejprve registrovaný na univerzální šablonu mozku, následně jsou z něj vypočítány různé **příznaky** (např. odchylka od atlasu, levo-pravá asymetrie, gradient). Získané příznaky jsou poté převedeny do nižšího rozlišení. Každý řádek matice dat, která je použita pro trénink a predikci v náhodném lese, se skládá z příznaků pro daný **voxel** a jeho okolí. Rozhodování v lese tedy závisí nejen na příznacích jednoho voxelu, ale i na jeho okolí. Výstupem rozhodování v náhodném lese je **mapa posteriorní pravděpodobnosti**, která je předána do další iterace, kde se používá již větší rozlišení příznaků, které opět vstupují do náhodného lesa. Výstupní pravděpodobnostní mapa ze



třetího náhodného lesa se převede na segmentaci zdravá/postižená tkáň. Pro použití této metody je nutné, aby se léze nacházela v levé hemisféře, proto autoři srovnávací studie zrcadlově obracely obrazy, na kterých se léze nacházela v pravé hemisféře.

Tato srovnávací studie používala již předučené modely, které dodali autoři srovnávaných metod. Validace proběhla na [datasetu ATLAS](#), z něhož bylo použito 181 T1-w snímků pacientů. Každý ze snímků byl anotován jednou osobou ze skupiny jedenácti lidí, kteří byli předem podrobeni tréninku.

Z porovnání metod mezi sebou byly odstraněny ty snímky, kde některá z metod nedokázala osegmentovat žádný voxel, přestože na snímku byla léze. Také byly odstraněny snímky, u kterých všechny tři metody osegmentovaly voxely tak, že nenastal průnik se segmentací experta (misclassified). Selhání jednotlivých metod je možné vidět v tabulce 2.

	<b>prázdná segmentace</b>	<b>prázdný průnik se segmentací experta</b>
ALI	24	28
naive Bayes	0	39
LINDA	23	45
	8 překryvů pro LINDA a ALI	10 překryvů mezi všemi metodami

Tabulka 2: Počty pacientů (z celkových 181), u kterých segmentace kompletně selhala.

<b>Method</b>	<b>Cases</b>	<b>DC</b>	<b>ASSD (mm)</b>	<b>HD (mm)</b>
Clusterize	152	$0.23 \pm 0.19$	$13.59 \pm 5.85$	$75.00 \pm 22.88$
ALI	132	$0.36 \pm 0.25$	$14.38 \pm 13.53$	$61.55 \pm 28.84$
naive Bayes	132	$0.39 \pm 0.23$	$10.49 \pm 6.25$	$58.00 \pm 19.73$
LINDA	132	$0.45 \pm 0.31$	$12.68 \pm 16.49$	$42.07 \pm 24.38$
inter-rater	5	$0.75 \pm 0.18$		

Tabulka 3: Porovnání Dice coefficientu (DC), average symmetric surface distance (ASSD) a Hausdorffské vzdálenosti (HD). Z evaluačního datasetu byly odstraněny některé snímky, u kterých segmentační metody nefungovaly. Poslední řádek zobrazuje DC při porovnání segmentací mezi dvěma experty u vybraných pěti pacientů.

Výsledky studie ukázaly, že všechny čtyři metody vykazují vysokou chybovost. Konkrétní hodnoty evaluace jsou k dispozici v tabulce 3. Úspěšnost segmentace léze je velmi závislá na velikosti a lokalizaci léze. Autoři článku analyzovaly případy pacientů, u kterých jednotlivé metody pracovaly velmi špatně (DC při evaluaci pod prvním kvantilem) nebo naopak velmi dobře (DC nad třetím kvantilem). Byly vytvořeny kategorie polohy léze a velikosti. Velikosti (malá, střední, velká) byly získány rozdělením celého datasetu na třetiny. Výsledky analýzy jsou uvedeny v tabulce 4, ze které vyplývá, že problém nastal nejčastěji při segmentaci malé léze v subkortikální oblasti, v mozkovém kmeni

nebo v mozečku. Naopak dobrých výsledků bylo dosaženo v případě, kdy byla léze velká nebo se vyskytovala v kortikální oblasti mozku.

Method	$DC_{Q1}$	Lokalizace	Velikost
ALI	0.00	Kortikální: 1/53 Subkortikální: 33/107 Kmen: 9/9 Mozeček: 9/12	Malá: 44/60 Střední: 8/61 Velká: 0/60
naive Bayes	0.03	Kortikální: 2/53 Subkortikální: 23/107 Kmen: 9/9 Mozeček: 11/12	Malá: 40/60 Střední: 4/61 Velká: 1/60
LINDA	0.00	Kortikální: 2/53 Subkortikální: 46/107 Kmen: 9/9 Mozeček: 11/12	Malá: 52/60 Střední: 16/61 Velká: 0/60

Method	$DC_{Q3}$	Lokalizace	Velikost
ALI	0.50	Kortikální: 34/53 Subkortikální: 11/107 Kmen: 0/9 Mozeček: 1/12	Malá: 0/60 Střední: 8/61 Velká: 38/60
naive Bayes	0.53	Kortikální: 39/53 Subkortikální: 11/107 Kmen: 0/9 Mozeček: 0/12	Malá: 0/60 Střední: 5/61 Velká: 41/60
LINDA	0.67	Kortikální: 32/53 Subkortikální: 14/107 Kmen: 0/9 Mozeček: 0/12	Malá: 0/60 Střední: 6/61 Velká: 40/60

Tabulka 4: Sloupce  $DC_{Q1}$  a  $DC_{Q3}$  označují první a třetí kvartil DC koeficientu dané metody. Sloupce lokalizace a velikost zobrazují počty lézí spadající do daných kategorií, u nichž byl DC nižší, resp. vyšší než  $DC_{Q1}$ , resp.  $DC_{Q3}$

Autoři doporučují, aby v navazujících výzkumech byla použita prior knowledge o velikosti a poloze léze. Zároveň navrhuje používání větších a variabilních datasetů. V závěru článku důrazně upozorňují, že je třeba vizuální inspekce a manuální kontrola kvality automatické segmentace pro každý snímek.

### 3.3 Segmentace pomocí Markov random fields

Studie [19] se zaměřuje na segmentaci sub-akutní ischemické léze (2-7 dní od nástupu mrtvice) pouze z FLAIR snímků metodou Markovovských náhodných polí.

Motivací autorů pro použití uvedené metody bylo to, že jiné publikované algoritmy trpí, resp. se musí vypořádat s různými problémy normalizace dat. A to zvláště v případě, kdy trénovací dataset tvoří data sesbíraná z různých zdravotnických center - typicky jsou pak snímky z jednotlivých center pořízeny s různým nastavením parametrů snímkování. Kromě toho se autoři chtěli vyhnout multimodálním datům (stejný pacient je snímkován v různých sekvencích). Důvodem bylo to, že pro použití multimodálních dat je potřeba registrovat data do společné vztažné soustavy, to vede k interpolačním artefaktům. Zároveň registrace snímků vyžaduje manuální nastavení parametrů tak, aby byl např. potlačen problém **distorse**, která je často přítomná při pořízení DWI snímků získaných metodou echo-planárního snímkování.

Výhodou použití pouze sekvence FLAIR je dle autorů to, že FLAIR je základním a nejjednodušším ukazatelem při sledování a segmentaci ischemické léze. Zároveň je možné tuto metodu použít pro historické datasey, které se používají jako kontrolní skupina v klinických studiích.

Největší chybovost segmentace mozkových lézí z FLAIR způsobují tzv. White Matter Hyperintensities, které již byly zmíněny v článku [15]. Tyto oblasti jsou algoritmy zaměňovány s lézí a falešně-pozitivně osegmentovány.

Autoři pracují s daty ze 8 zdravotnických center, které jsou získány ze tří datasetů, s nimiž autoři pracují odděleně, resp. provádí dílčí evaluaci na každém z těchto datasetů. Celkem mají k dispozici 159 FLAIR snímků pacientů. Jednotlivé snímky jsou pořízeny na zařízení s magnetickým polem 1.5 T i 3 T. Snímky mají různé tloušťky řezu i různé **in-plane** rozlišení.

V rámci preprocessingu byla ze snímků všech pacientů odstraněna lebka, nicméně autoři zmiňují, že tak dělají kvůli tomu, aby zajistili konzistenci dat, jelikož část snímků měli autoři k dispozici již s odstraněnou lebkou. Následně byly snímky registrovány na atlas [14] představující mozek lidí ve věku mezi 63 a 75 lety. Expertní segmentace autoři interpolovali do nového prostoru, aby odpovídali registrovaným snímkům. Intenzity voxelů byly normalizovány pro každého pacienta.

Navržená metoda je rozdělena do dvou fází. Spočívá v určení pravděpodobností, zda daný voxel obsahuje lézi nebo zda je tkáň zdravá. Následně mapa pravděpodobností vstupuje do druhého stupně algoritmu, který je založen na **Markovovských náhodných polích**. Výstupem je segmentace každého voxelu do jedné ze čtyř tříd - mozkomíšní mok, šedá hmota mozková, bílá hmota mozková nebo léze. Významem druhého stupně této segmentační metody je převáhování původní pravděpodobnostní mapy, která je výstupem prvního stupně tak, aby byl vzat v úvahu i lokální kontext a textura voxelů. To by dle autorů mělo vylepšit segmentaci hranice lézí.

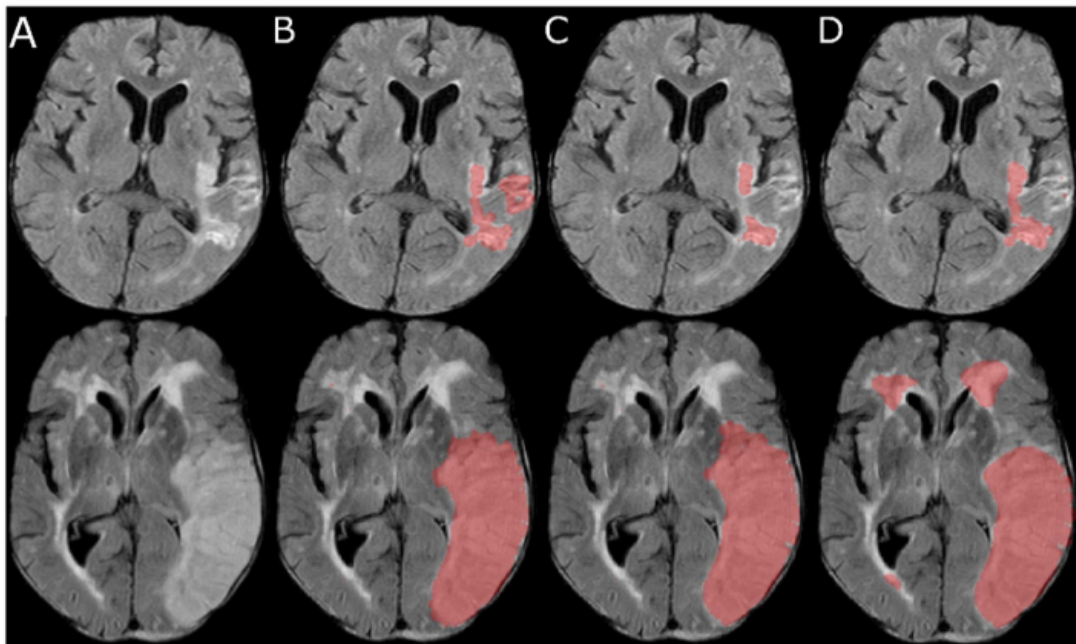
V první fázi segmentace autoři používají diskrétní 2D Gaborovu transformaci [21]. Tato transformace se aplikuje konvolucí Gaborovských okének, které jsou funkcí orientace a fáze. Výstupem této konvoluce je potom informace o energii v obraze pro daný směr a danou frekvenci. Z výstupu transformace je naučen Bayesovský klasifikátor, jehož

výstupem je pravděpodobnostní mapa.

Druhá fáze klasifikace je založena na Markovovských náhodných polích (MRF). Autoři vycházeli z tzv. Pottsova modelu, který je rozšířením MRF pro klasifikaci do více než dvou tříd [13]. Pottsův model byl přizpůsoben právě tak, aby bral v potaz okolní texturu klasifikovaného voxelu. Při trénování modelu autoři uvažují 8 sousedů in-plane a korespondující voxely o vrstvu snímku výš a vrstvu níž.

Během evaluace autoři své řešení porovnávají i se dvěma nejlepšími metodami z ISLES 2015 - DeepMedic založené na neuronových sítích a CN-Neu založené na klastrování, které jsem již v této práci představil v kapitole 3.1.1 a 3.1.2. Vyhodnocení proběhlo metodou leave-one-patient-out.

Kvalitativní výsledky studie ukázaly, že metoda je schopná úspěšně segmentovat léze různé velikosti a v různých polohách, včetně komplexních struktur léze. Proti metodě fuzzy c-means clusteringu uvedená metoda dosahuje lepších výsledků při výskytu periventrikulární hyperintenzity ve FLAIR obraze, který není součástí segmentované mrtvice. Porovnání jednotlivých výsledků včetně falešné segmentace c-means metody je ukázáno na obrázku 4.



Obrázek 4: Kvantitativní porovnání výsledků s metodami z ISLES 2015. A zobrazuje původní FLAIR obrazy. B vyznačuje segmentaci navrženou metodou Markovovských náhodných polí. C ukazuje segmentaci neuronovou sítí představenou jako DeepMedic. D ukazuje segmentaci pomocí klasteringu c-means.

Kvantitativní výsledky ukázaly, že metoda Dice koeficientem vykazuje pro všechny datasety použité ve studii podobné výsledky, avšak u jednoho z datasetů významně podhodnocuje objem segmentované léze. Dataset s podhodnocenými objemy obsahuje

Method	DC	ASD (mm)	HD (mm)
Markovovská náhodná pole	$0.582 \pm 0.250$	$7.56 \pm 9.29$	$35.38 \pm 25.97$
DeepMedic	$0.541 \pm 0.272$	-	-
C-means	$0.524 \pm 0.287$	-	-

Tabulka 5: Vážený průměr Dice koeficientu, average surface distance a Hausdorffská vzdálenost pro navrženou metodu používající Markovovská náhodná pole, metodu hlubokých neuronových sítí DeepMedic a metodu fuzzy klasteringu c-means z ISLES 2015. Vážený průměr byl získán průměrováním dílčích výsledků testování uvedených metod na různých datasetech.

102 snímků a expertní segmentace uvádí objemy lézí u pacienta v průměru  $42.95 \pm 57.6$  ml. Zatím co automaticky osegmentovaný objem činí průměrně  $26.82 \pm 41.37$  ml na pacienta. Právě objem léze je dle autorů jedním z nejdůležitějších údajů, který se snaží díky segmentaci získat, jelikož se používá při klinických testech nové léčby. Dále se ukázalo, že podhodnocení objemu je tím výraznější, čím větší je léze. Srovnání výkonu jednotlivých metod při testování na stejných datasetech je uvedeno v tabulce 5.

Měření výkonu na malých lézích (pod 1 ml) autoři provedli zkoumáním okolí dané komponenty (uzavřené oblasti léze). Pokud v 26-voxelovém okolí expertně osegmentované léze ležely alespoň tři voxely osegmentované automatickou metodou, byla segmentace považována za pravdivě pozitivní segmentaci. Na druhou stranu, pokud byla segmentována komponenta léze automatickou metodou a ležely v ní méně než 3 voxely expertní segmentace, byla segmentace považována za falešně pozitivní. Touto metrikou bylo vyhodnoceno, že 258 z 483 lézí menších než 1 ml bylo klasifikováno korektně a 81 lézí falešně pozitivně.

Při srovnání výsledků s DeepMedic bylo zjištěno 239 pravdivě pozitivně segmentovaných malých lézí, což je podobný výsledek jako u metody Markovovských náhodných polí. Vyskytuje se u ní ale značně vyšší počet falešně segmentovaných lézí, nichž je 131. Naopak metoda fuzzy c-means z ISLES 2015 našla 209 pravdivě pozitivně segmentovaných lézí, což je méně, než uvedená metoda. Fuzzy c-means měla 84 falešně pozitivních segmentací, což je zhruba stejný počet jako u uvedené metody.

Autoři zmiňují, že výsledky nelze přímo srovnávat s jinými metodami, jelikož k tréninku a evaluaci ostatních metody byla použita jiná data. Přesto poznamenávají, že došli k podobným výsledkům jako jiné publikované metody. Navíc zmiňují, že dosáhly srovnatelných výsledků jako nejlepší metody na soutěži ISLES 2015, kde však měli soutěžící k dispozici multimodální data, zatím co uvedená metoda pracuje pouze s FLAIR snímky. Nejlepší dílčí výsledky byly dosaženy touto metodou na datasetu, jehož snímky byly pořízeny pouze jedním skenerem s magnetickým polem 3 T a v diskuzi se autoři zmiňují, že právě tato konzistence a rozlišení můžou zlepšit úspěšnost při testování automatické metody.

Dále autoři zmiňují, že nejhoršího Dice koeficientu dosáhly při testování u datasetu, jenž byl převzorkovaný a díky tomu vznikly významné interpolační artefakty.

Jako největší nevýhodu autoři vidí právě ono podhodnocování objemu, díky němuž

není metoda zatím optimální. V případě dalšího vývoje této metody autoři doporučují se zaměřit na tento problém a zmiňují, že textura je důležitá pro jádro léze, zatím co kontrast voxelů je důležitý pro určení hranice léze.

### 3.4 title

## 4 Dataset

### 4.1 ISLES 2022

Kromě soutěže ISLES 2015, která byla první svého druhu a již jsem ji popsal výše, proběhlo několik dalších ročníků ISLES.

V roce 2016 byla soutěž zaměřená na predikci segmentace po 3 až 9 měsících od mrtvice a regresi predikující klinický dopad mrtvice. Tyto predikce byly vytvořeny ze snímků akutní mrtvice a klinických dat: doba od vzniku mrtvice, doba do zahájení léčby a TICI skóre, hodnotící úspěšnost léčby. V roce 2017 byla soutěž opět zaměřena na vytváření predikce segmentace z akutní mrtvice. V roce 2018 bylo cílem vytvoření segmentace akutní mrtvice, ale vstupní modalitou byla tentokrát CT perfuze. Motivací je dostupnost a rychlost vytvoření skenu z CT, díky těmto výhodám se CT používá pro triáž pacientů s mrtvicí.

Poslední a stále běžící soutěží je ISLES 2022 [16]. Ta je podobně jako ISLES 2015 rozdělena na dvě kategorie. První kategorií je multimodální fMRI segmentace akutní a subakutní mrtvice. Druhou kategorií je segmentace akutní, subakutní a chronické mrtvice ze snímků T1-vážený. Jelikož se tato práce zabývá primárně multimodální segmentací, v dalších částech této práce budu pod označením ISLES 2022 uvažovat pouze první kategorii této soutěže.

ISLES 2022 se liší od předchozích soutěží v tom, že je zaměřená i na segmentace **mnohočetných embolií a kortikální infarkty (mechanická rekanalizace)**, snímky jsou pořízeny jak před, tak po intervenční léčbě a dataset obsahuje mnohem více dat, jak dataset ISLES 2015. Výše popsané rozdíly zapříčiňují poměrně vysokou komplexitu lézí.

V trénovacím a veřejně dostupném datasetu ISLES 2022 se vyskytuje 250 pacientů. V testovacím neveřejném datasetu jsou skeny 150 pacientů. Pro každého pacienta je k dispozici DWI (až  $b=1000 \text{ mm}^2$ ) sekvence, ADC sekvence a FLAIR sekvence. Skeny byly pořízeny na dvou zařízeních se silou magnetického pole 3 T a z jednoho zařízení se silou 1.5 T. V datasetu je poznamenáno, z jakého zdravotnického zařízení byl daný sken pořízen, společně s parametry ve formátu DICOM hlavičky, kde je možné nalézt podrobnosti k nastavení, podle kterého byl sken pořízen.

Suma objemu lézí u jednoho pacienta je průměrně 26.38 ml, největší objem ischemické tkáně je 482.15 ml. Průměr objemu jedné (souvislé) léze je 2.90 ml, přičemž nejmenší léze je velká 0.01 ml a největší 477.23 ml. Průměrný počet lézí u jednoho pacienta je 9.11, pacient s největším počtem má 126 lézí. Pro diverzitu jsou v trénovacím datasetu zahrnuti tři pacienti, u kterých se mrtvice nepotvrdila.

Expertní segmentace v tomto případě byla provedena hybridně, kvůli velkému počtu pacientů zahrnutých do datasetu. Nejprve tedy autoři natrénovali 3D UNet [1] na DWI

sekvenci snímků z jednoho ze zdravotnických center, pro které již existovaly anotace. Následně autoři vytvořili prvotní segmentace pro ostatní data. Tyto prvotní segmentace byly dále upraveny mediky a následně dokorigovány neuroradiology a schváleny neuroradiology s více jak desetiletou praxí. Se získáním kvalitních segmentací byl dále dotrénován UNet pro zkvalitnění prvotních segmentací, které ještě nebyly manuálně upraveny.

Automatická segmentace pomocí UNetu byla provedena pouze na základě DWI sekvence, ale manuální expertní segmentace byla provedena s využitím znalostí ze všech dostupných modalit a proto i algoritmus, resp. soutěžící by měli využívat všech modalit, aby bylo možné dosáhnout stejného výsledku jako je anotace. Autoři explicitně zmiňují, že DWI sekvence byla použita pro segmentaci jako primární.

V rámci postprocessingu, před publikováním datasetu, bylo provedeno několik kroků. Nejprve byla provedena pomocí Elastix [12] rigidní registrace FLAIR na DWI. Díky tomu mohla být proveden skull-stripping pomocí algoritmu HD-BET [5]. Nakonec byly snímky transformovány zpět do jejich nativního prostoru.

Pro ověření a porozumění dopadů anotací mnoha experty autoři soutěže provedli inter-rater analýzu. Z původního datasetu vybrali 10 reprezentativních skenů, které předložili dvěma expertům - neuroradiologům, k segmentaci. Ukázalo se, že Dice koeficient expertů proti segmentaci ISLES 2022 je průměrně 0.90, což odpovídá rozdílu v 2.37 ml objemu léze. Druhý expert proti ISLES 2022 dosáhl Dice 0.86, což odpovídá rozdílu 6.56 ml. Dice koeficient vůči těmto dvěma expertům navzájem je průměrně 0.83, což odpovídá objemu 7.22 ml.

Výsledky soutěže jsou sice veřejné, nicméně metody, které autoři používají, nejsou obecně zveřejněny. Nejlepší tým v soutěži dosahuje Dice koeficientu 0.96. V tabulce výsledků existuje jediná metoda, která má zveřejněnou implementaci. Tuto metodu vytvořil Shengbo Gao s týmem Seals [20] a dosahuje Dice koeficientu 0.92. Autoři uvádí jen návod, jak použít jejich metodu, ale bohužel nemají k dispozici článek ani abstrakt popisující jejich metodu. Pouze uvádí, že vychází z metody nnUNet [6].

## Odkazy

- [1] Özgün Çiçek et al. “3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation”. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*. Ed. Sebastien Ourselin et al. Cham: Springer International Publishing, 2016, s. 424–432. ISBN: 978-3-319-46723-8.
- [2] Bianca de Haan et al. “Fast semi-automated lesion demarcation in stroke”. In: *NeuroImage: Clinical* 9 (2015), s. 69–74. ISSN: 2213-1582. DOI: <https://doi.org/10.1016/j.nicl.2015.06.013>. URL: <https://www.sciencedirect.com/science/article/pii/S2213158215001199>.
- [3] Chaolu Feng, Dazhe Zhao a Min Huang. “Segmentation of Ischemic Stroke Lesions in Multi-spectral MR Images Using Weighting Suppressed FCM and Three Phase Level Set”. In: sv. 9556. Břez. 2016, s. 233–245. ISBN: 0302-9743. DOI: 10.1007/978-3-319-30858-6\_20.



- [4] 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”. In: *Journal of Neuroscience Methods* 257 (2016), s. 97–108. URL: <https://api.semanticscholar.org/CorpusID:22978547>.
- [5] Fabian Isensee et al. “Automated brain extraction of multisequence MRI using artificial neural networks”. In: *Human Brain Mapping* 40.17 (2019), s. 4952–4964. DOI: <https://doi.org/10.1002/hbm.24750>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/hbm.24750>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/hbm.24750>.
- [6] Fabian Isensee et al. “nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation”. In: *Nature Methods* 18 (2020), s. 203–211. URL: <https://api.semanticscholar.org/CorpusID:227947847>.
- [7] Kaori L. Ito, Hosung Kim a Sook-Lei Liew. “A comparison of automated lesion segmentation approaches for chronic stroke T1-weighted MRI data”. In: *bioRxiv* (řij. 2018), s. 441451. DOI: 10.1101/441451. URL: <https://www.biorxiv.org/content/10.1101/441451v1><https://www.biorxiv.org/content/10.1101/441451v1.abstract>.
- [8] Mark Jenkinson, Mickael Pechaud, Stephen Smith et al. “BET2: MR-based estimation of brain, skull and scalp surfaces”. In: *Eleventh annual meeting of the organization for human brain mapping*. Sv. 17. 3. Toronto. 2005, s. 167.
- [9] Konstantinos Kamnitsas et al. “Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation”. In: *Medical Image Analysis* 36 (2017), s. 61–78. ISSN: 1361-8415. DOI: <https://doi.org/10.1016/j.media.2016.10.004>. URL: <https://www.sciencedirect.com/science/article/pii/S1361841516301839>.
- [10] Konstantinos Kamnitsas et al. “Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation”. In: *Medical Image Analysis* 36 (ún. 2017), 61–78. ISSN: 1361-8415. DOI: 10.1016/j.media.2016.10.004. URL: <http://dx.doi.org/10.1016/j.media.2016.10.004>.
- [11] Konstantinos Kamnitsas et al. *Ensembles of Multiple Models and Architectures for Robust Brain Tumour Segmentation*. 2017. arXiv: 1711.01468 [cs.CV].
- [12] Stefan Klein et al. “elastix: A Toolbox for Intensity-Based Medical Image Registration”. In: *IEEE Transactions on Medical Imaging* 29.1 (2010), s. 196–205. DOI: 10.1109/TMI.2009.2035616.
- [13] D. Koller a N. Friedman. *Probabilistic Graphical Models: Principles and Techniques*. Adaptive computation and machine learning. MIT Press, 2009. ISBN: 9780262013192. URL: <https://books.google.co.in/books?id=7dzpHCHzNQ4C>.
- [14] Hervé Lemaître et al. “Age- and sex-related effects on the neuroanatomy of healthy elderly”. In: *NeuroImage* 26.3 (2005), s. 900–911. ISSN: 1053-8119. DOI: <https://doi.org/10.1016/j.neuroimage.2005.02.042>. URL: <https://www.sciencedirect.com/science/article/pii/S105381190500145X>.

- [15] Oskar Maier et al. “ISLES 2015 - A public evaluation benchmark for ischemic stroke lesion segmentation from multispectral MRI”. In: *Medical Image Analysis* 35 (2016), s. 250–269. DOI: 10.1016/j.media.2016.07.009. URL: <http://dx.doi.org/10.1016/j.media.2016.07.0091361-8415/>.
- [16] Moritz R. Hernandez Petzsche et al. “ISLES 2022: A multi-center magnetic resonance imaging stroke lesion segmentation dataset”. In: *Scientific Data* 2022 9:1 9 (1 pros. 2022), s. 1–9. ISSN: 2052-4463. DOI: 10.1038/s41597-022-01875-5. URL: <https://www.nature.com/articles/s41597-022-01875-5>.
- [17] Dorian Pustina et al. “Automated segmentation of chronic stroke lesions using LINDA: Lesion identification with neighborhood data analysis”. In: *Human Brain Mapping* 37.4 (2016), s. 1405–1421. DOI: <https://doi.org/10.1002/hbm.23110>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/hbm.23110>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/hbm.23110>.
- [18] Mohamed L Seghier et al. “Lesion identification using unified segmentation-normalisation models and fuzzy clustering”. en. In: *Neuroimage* 41.4 (čvc. 2008), s. 1253–1266.
- [19] Nagesh K. Subbanna et al. “Stroke lesion segmentation in FLAIR MRI datasets using customized Markov random fields”. In: *Frontiers in Neurology* 10 (MAY květ. 2019), s. 446260. ISSN: 16642295. DOI: 10.3389/FNEUR.2019.00541/BIBTEX.
- [20] *Tabrisrei/ISLES22\_SEALS: ISLES22 Acute Ischemic Stroke segmentation competition, team SEALS final submission*. URL: [https://github.com/Tabrisrei/ISLES22\\_SEALS](https://github.com/Tabrisrei/ISLES22_SEALS).
- [21] Meir Zibulski a Yehoshua Y. Zeevi. “Analysis of Multiwindow Gabor-Type Schemes by Frame Methods”. In: *Applied and Computational Harmonic Analysis* 4.2 (1997), s. 188–221. ISSN: 1063-5203. DOI: <https://doi.org/10.1006/acha.1997.0209>. URL: <https://www.sciencedirect.com/science/article/pii/S1063520397902098>.