

MOŽNOSTI PŘÍSTUPU K OBRAZOVÝM DATŮM V RÁMCI PROJEKTŮ ÚBMI VE SPOLUPRÁCI S KLINICKÝMI PRACOVIŠTI

Roman Jakubíček¹, Petra Nemčková¹, Petr Ouředníček^{2,3}, Jiří Chmelík¹

¹Vysoké učení technické v Brně, Ústav biomedicínského inženýrství, Brno

²Klinika zobrazovacích metod, Fakultní nemocnice u sv. Anny v Brně, Brno

³Philips Česká republika s.r.o., Praha

ABSTRAKT

Tento článek zkoumá výzvy a možnosti spojené se zpracováním a sdílením obrazových dat v kontextu biomedicíny a počítačem podporované diagnostiky. S rostoucím výpočetním výkonem a využitím strojového učení se metody analýzy obrazů stávají stále efektivnějšími, ale potýkají se s problémy dostupnosti dat a obtížnou interpretovatelností. Autoři diskutují legislativní a etické aspekty ochrany osobních údajů a upozorňují na význam spolupráce mezi akademickými institucemi a klinickými pracovišti. Článek také prezentuje dva aktuální výzkumné projekty ÚBMI v oblasti analýzy obrazů, tj. analýzu trombu v CT mozku a kardiovaskulární zobrazování magnetickou rezonancí, ve kterých se aktuálně využívají pokročilé algoritmy strojového učení. Spolupráce mezi ÚBMI a klinickými pracovišti přináší nové možnosti pro zlepšení diagnostiky a léčby pacientů.

KLÍČOVÁ SLOVA

Zpracování obrazů, sdílení dat, výzkumné projekty, strojové učení

ÚVOD

Dnešním trendem ve výzkumu i v klinických aplikacích počítačem podporované diagnostiky prakticky ve všech odvětvích biomedicíny je využití nástrojů strojového učení [1]. Společně s neustále narůstajícím výpočetním výkonem a rostoucími nároky na množství zpracovávaných dat se tyto metody stávají efektivnější jak z hlediska rychlosti zpracování informací, tak i z hlediska úspěšnosti, robustnosti a reprodukovatelnosti. Nevýhodou těchto modelů však stále zůstává nutnost přístupu k učebním databázím s anotacemi a také obtížná interpretovatelnost, což je jedním z hlavních důvodů, proč je vždy nutná kontrola výsledků a konečné rozhodnutí leží i na dále na lékařích a jiných expertech.

Nařízení Evropského parlamentu a Rady (EU) 2016/679, schválené dne 27. dubna 2016, týkající se ochrany osobních údajů v kontextu jejich zpracování a volného pohybu těchto údajů, zrušující směrnici 95/46/ES (obecné nařízení o ochraně osobních údajů), vstoupilo v platnost 25. května 2018. V anglické verzi je známé jako "General Data Protection Regulation" [2]. Pro práci s obrazovými daty je třeba mít podepsané souhlasy pacientů s využitím jejich dat pro výzkumné účely. Navíc v případě nestandardního protokolu snímání je také nutné mít souhlas lokální etické komise k provádění experimentálních snímání. Několik let se na národní i evropské úrovni diskutuje o vzniku „evropského prostoru pro sdílení zdravotnických údajů (EHDS)“, které by výrazně podpořily konkurenceschopnost výzkumných institucí na globálním trhu medicínských inovací.

PŘEHLED MOŽNOSTÍ PŘÍSTUPU A SDÍLENÍ DAT

Vzhledem k silné regulaci přístupu k osobním a medicínským datům je zřejmé, že i přes dodržení všech legálních postupů (informovaný souhlas, etická komise, pseudo/anonymizace, šifrování) není možné jen tak přenést pacientská data z klinického pracoviště na jiné místo za účelem trénování modelů strojového učení nebo k další analýze. Pro další účely předpokládáme platné informované souhlasy a schválení etickou komisí. Následující seznam pak ukazuje několik možností přenosu a sdílení dat včetně jejich výhod a úskalí:

- Přenos dat na pevném médiu (např. flash disk) – výhodou může být umístění na jednom místě s možností šifrování, nicméně klinické pracoviště ztrácí kontrolu nad daty (kopírování, šíření, smazání). Dalším problémem je riziko ztráty a zneužití v případě prolomení nebo nepoužití šifrování. Nutná je úplná anonymizace dat.
- Sdílení přes cloudové úložiště – oproti pevnému disku je zde nevýhodou riziko hackerských útoků a problematika důvěryhodnosti provozovatele úložiště. Výhodou je možnost sledování přístupů k datům, jejich správa včetně možnosti jejich smazání. I zde je však nutná úplná anonymizace dat.
- Z hlediska ochrany dat je nejbezpečnější, aby daný externí pracovník měl pracovní smlouvu přímo na klinickém pracovišti a na datech pracoval fyzicky na místě bez toho, aby jakákoliv data opustila zabezpečený systém pracoviště. Tento postup nicméně není vždy možný a je vyžadován výkonný výpočetní server přímo na pracovišti. V tomto případě (dle charakteru pracovní smlouvy) není nutná plná anonymizace dat, případně postačí pseudoanonymizace.
- Volnějším řešením je smlouva o spolupráci mezi pracovišti, přičemž externí pracovník přistupuje do chráněné interní sítě přes zabezpečené připojení VPN (Virtual Private Network) vzdáleně. Data tak zůstávají stále uvnitř vnitřní sítě a externí pracovník k nim pouze v rámci legálních možností přistupuje. Tato varianta však vyžaduje vhodné síťové a hardwarové řešení na straně klinického pracoviště. Zde je opět nutná anonymizace dat.
- V dnešních dobách je často diskutovanou možností tzv. federovaný přístup/učení (Federated Learning) [3]. Na každém klinickém pracovišti je umístěn výpočetní server, který má přístup k datům. Neumožňuje však jejich sdílení ani prohlížení, pouze dokáže data filtrovat a třídit dle metadat a zadaných požadavků. Externí pracovník navrhne model strojového učení a jeho architekturu a parametry vzdáleně nahraje na interní výpočetní server, kde dojde k jeho natrénování. Připravený model je poté odeslán zpět externímu pracovníkovi. Hlavní myšlenkou také je, že v případě zapojení více pracovišť dojde k naučení několika nezávislých modelů (na každém pracovišti jeden), které jsou poté zkombinovány v jeden výsledný, což zvyšuje jeho robustnost. Existuje celá řada možností, jak fúzi modelů provádět. Nevýhodou jsou vysoké požadavky na hardwarové vybavení a síťovou správu (každé zapojené pracoviště musí disponovat odpovídajícím výpočetním serverem se zabezpečeným a kontrolovaným přístupem k datům).

Jak je z výše uvedeného zřejmé, neexistuje jedna obecně aplikovatelná varianta vhodného a legálního sdílení dat pro potřeby trénování modelů strojového učení. Přístup k těmto datům tak často zůstává velmi omezený.

PROBLEMATIKA SDÍLENÍ A ANONYMIZACE OBRAZOVÝCH DAT

Obrazová data mají svá specifika, která je nutné brát na zřetel. Standardizovaný formát pro obrazová medicínská data je DICOM [4], který se skládá z metadat v hlavičce souboru (rodné číslo, datum narození, akviziční parametry atd.) a samotných obrazových hodnot. Běžně pro anonymizaci postačí odstranění citlivých informací z hlavičkové části souboru, nicméně zejména pro tomografická obrazová data je nutné uvažovat také citlivé či dokonce biometrické údaje přímo v měřených hodnotách (tvář, pohlaví apod.). Například u snímků mozku je tedy pro úplnou anonymizaci nutné odstranit také obličejovou část obrazu, aby nebyla možná zpětná identifikace pomocí např. rekonstrukce tváře či zubů pacienta. Dalším úskalím zejména tomografických dat jsou jejich velké paměťové nároky, kdy mohou trojrozměrná data často obsahovat více parametrických kanálů (magnetická rezonance, více-energetická počítačová tomografie atd.) nebo mohou být snímky pořizovány v čase (snímání srdeční činnosti, perfúzní vyšetření, angiografie apod.). Všechny tyto faktory znesnadňují přenos, sdílení a uchování těchto dat.

Řešením legálních problémů může být využití veřejně dostupných databází pro trénování modelů, jejichž množství s posledními roky stále narůstá. I zde je však třeba překonat několik překážek, jako je jejich kvalita, správnost anotací a zejména licenční podmínky. Často se můžeme setkat s veřejnými datasety, které nemohou být dle licence použity například ke komerčním nebo dokonce ani k výzkumným účelům, ale pouze v rámci některé konkrétní soutěžní výzvy konference nebo k ověření již natrénovaných modelů. Veřejně dostupné datasety jsou také často příliš specifické a nemusí odpovídat přesně našemu řešenému problému. Problematikou je i původ datasetů, kdy pochází například z jednoho pracoviště a jsou snímány jedním přístrojem; to má za následek nízkou robustnost a přenositelnost naučených modelů na jiná data.

SPOLUPRÁCE ÚBMI S KLINICKÝMI PRACOVIŠTI

K navázání společné interdisciplinární výzkumné spolupráce v oblasti počítačové analýzy lékařských obrazových dat a medicínské interpretace a hodnocení výsledků mezi FEKT VUT a FNUSA byla mezi odděleními vytvořena rámcová smlouva uzavřená v souladu s ust. § 1746 odst. 2 zákona č. 89/2012 Sb., občanského zákoníku. Tato smlouva na dobu neurčitou právně upravuje předmět spolupráce, práva a povinnosti obou stran, náplň práce, publikování, či způsoby financování.

V rámci ní jsou uzavřeny dílčí smlouvy specifikující konkrétní výzkumné projekty řešící vědecko-výzkumné problémy, které vyžadují těsnou spolupráci obou smluvních stran, a v rámci které obě strany profitují z dostupnosti potřebných medicínských dat (na straně FNUSA) a současně z dostupnosti pokročilých matematických a biomedicínských algoritmů, vytváření programů a jejich testování (na straně VUT).

Díky této spolupráci a za finanční podpory firmy Philips se zakoupil a zapůjčil výpočetní server s přístupem do nemocničního intranetu a PACSU, a který se fyzicky nachází přímo v budovách FNUSA. Přístup do PACSu a intranetu z tohoto serveru má pouze zdravotnický personál FNUSA s platným přístupem, které je schopen stáhnout a uložit potřebná anonymizovaná obrazová data přímo na tento server. Ostatní, dokonce nelékařští, uživatelé tohoto serveru s platným přístupem prostřednictvím připojení přes VPN mají k těmto staženým datům přístup. Lze je poté zpracovávat navrženými softwary nebo přímo softwary vyvíjet na tomto serveru. Ten disponuje grafickou kartou NVIDIA P4000, výkonnými procesory Intel Xeon E5-2643v4 a 32 GB RAM pamětí a licencemi Microsoft Windows Server 2019, NVIDIA a MATLAB. Výhodou tohoto způsobu sdílení dat je zejména to, že obrazová data fyzicky neopustí nemocniční virtuální prostor a v případě úzké spolupráce mezi spolupracovníky prakticky „okamžitý“ a „neomezený“ přístup k anonymizovaným obrazovým datům.

Velmi přínosné pro vědecko-výzkumnou práci na ÚBMI s obrazovými daty je také zapůjčení výpočetního serveru s licencovanou nejnovější verzí softwarových doplňků pro práci s CT a MR daty, většinou dostupné s licencí u akvizičních systémů přímo na medicínských pracovištích. Přístup k těmto moderním softwarům umožňuje srovnání dostupných komerčních algoritmů nebo získání výstupních obrazových či parametrických map z těchto analytických programů. Těchto serverů se využívá také pro výuku zobrazovacích metod, vizualizaci a zpracování obrazů.

PROBÍHAJÍCÍ VÝZKUMNÉ PROJEKTY

Díky výše uvedeným možnostem přístupu k datům naše výzkumná skupina řeší ve spolupráci s klinickými pracovišti několik výzkumných projektů, jejichž stručný popis následuje.

ANALÝZA TROMBU CT DAT MOZKU

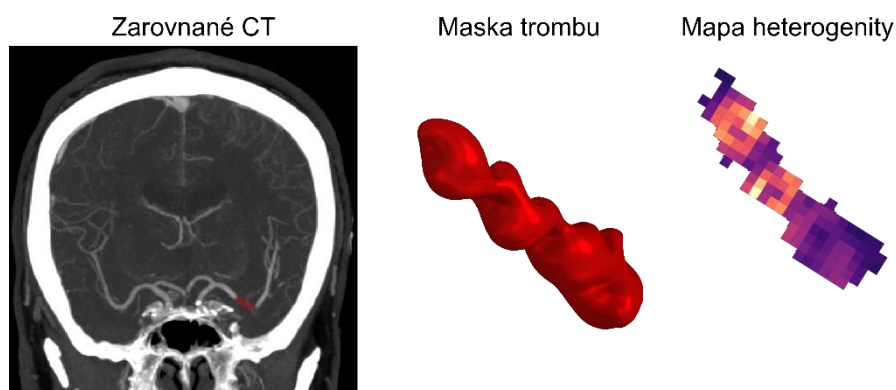
Cílem tohoto projektu ve spolupráci s týmem docenta Jiřího Vaníčka z FNUSA v Brně je prokázat, že permeabilita trombu úzce souvisí s jeho pórovitostí, což může ovlivnit jeho mechanické vlastnosti, a tak může ovlivnit úspěch mechanické trombektomie u pacientů s ischemickou cévní mozkovou příhodou. Pro hodnocení permeability trombu používáme bezkontrastní počítačovou tomografii (CT) a vícefázovou CT angiografii hlavy u pacientů s ischemickou cévní mozkovou příhodou s okluzí velkých cév, kde všechna kritéria pro zařazení povedou k léčbě mechanickou trombektomií.

V současné době je k dispozici 80 patientských dat a aktuálně probíhá manuální značení trombů lékařskými experty pro jejich následnou analýzu. Ve spolupráci s lékařskými experty z FNUSA byla navržena obecná metodika manuálního značení eliminující vliv anotátora a zajišťující maximální reprodukovatelnost. Členy technického týmu byl navržen přístup pro předzpracování multifázických dat, zahrnující vzájemnou registraci a zarovnání do radiologických rovin a vytvoření fúzovaných dat pro snadnou anotaci; ten byl publikován na mezinárodní konferenci [5]. Ukázka kroků manuální segmentace trombu společně s kroky automatického lícování obrazových dat je znázorněna na Obrázku 1.



Obrázek 1: 3D vizualizace manuálně segmentovaného trombu v CT datech mozku. Vlevo: překrytí nativního CT obrazu (červený) s obrazem s kontrastní látkou v první fázi (zelený) před registrací. Uprostřed: překrytí registrovaných snímků. Vpravo: 3D vizualizace manuálně segmentovaného trombu.

S ohledem na absolvovanou stáž členky týmu Ing. Petry Nemčkové v roce 2023 na zahraničním pracovišti Amsterdam University Medical Centers probíhá nyní úzká spolupráce s vedoucím týmu profesorem Henkem Marqueringem na výzkumu heterogenity trombu (Obrázek 3) a extrakce relevantních obrazových příznaků ve vztahu k fyzikálním vlastnostem trombu, a tedy i jeho složením. Současný stav řešení návrhu prvotní analýzy relevantnosti obrazových příznaků trombu byl prezentován příspěvkem konference [6].



Obrázek 2: Vizualizace heterogenity segmentovaného trombu. Vlevo: zobrazení segmentovaného trombu na sfúzovaných multifázických datech. Uprostřed: 3D maska extrahovaného trombu. Vpravo: zobrazení míry heterogenity okolo daného voxelu. Fialová značí nízkou míru heterogenity, žlutá vysokou.

V následujícím období bude probíhat vývoj a testování segmentačního algoritmu založeném na segmentačních architekturách konvolučních neuronových sítí jakými jsou U-net [7] nebo nnUNet [8]. Od projektu se ve finální fázi očekává softwarový nástroj umožňující predikovat složení či jen fyzikální vlastnosti trombu, a tím lékařům usnadnit rozhodovací proces ve volbě zavedeného typu léčby.

ANALÝZA KARDIOLOGICKÝCH OBRAZOVÝCH MR DAT

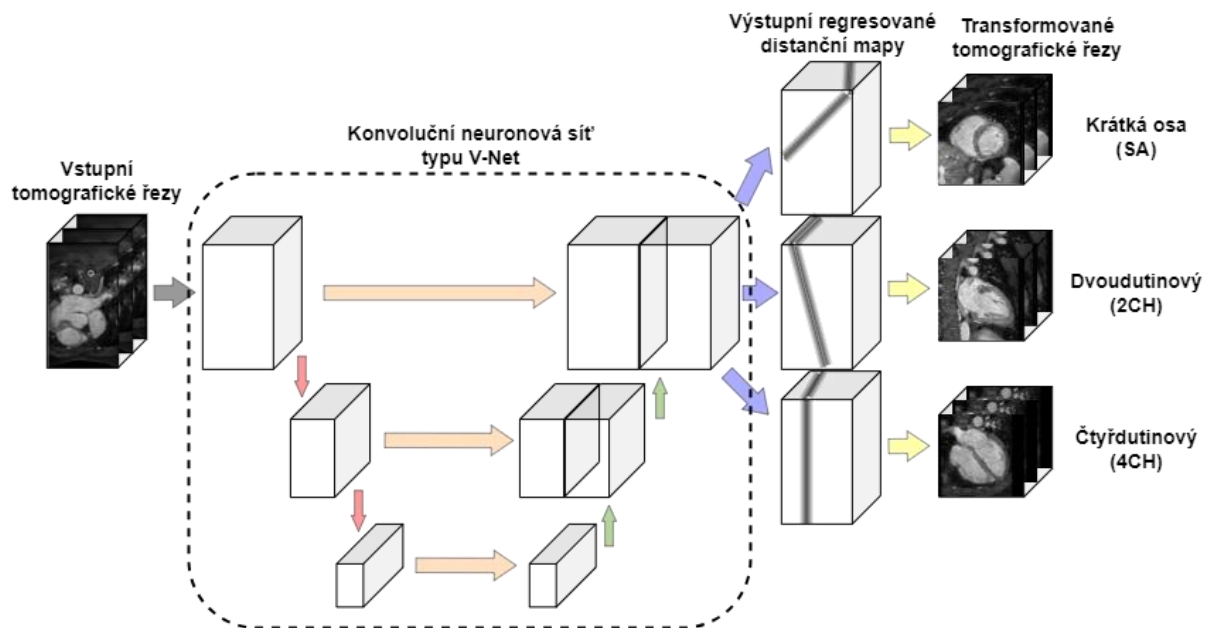
Hlavními cíli tohoto projektu je použití algoritmů hlubokého učení ke zjednodušení a zkrácení doby zobrazování a k eliminaci variability plánovacího postupu v kardiovaskulárním zobrazování magnetickou rezonancí (MR). V této fázi projektu se zabýváme automatizovaným určením základních radiologických pohledů během plánování.

Očekává se, že budou shromážděna data od kohorty 100 pacientů naskenovaných standardním protokolem včetně hlavních srdečních rovin určených radiologem; patří mezi ně například krátká osa (SA), 4dutinový (4CH), 2dutinový (2CH), a 3D přehledový snímek. Tato data budou použita pro trénování a testování algoritmů hlubokého učení.

Paralelně se získáváním vlastních klinických dat je využívána veřejně dostupná databáze MR kardiologických snímků [9] ke stanovení standardního předzpracování obrazu – formátování, standardizace a způsob anotace dat. Je také vyvinut vlastní programový nástroj, který umožňuje anotaci pozice všech základních os expertem dle platného postupu při radiologickém plánování.

Implementace konvoluční neuronové sítě (CNN) jako regresního enkodéru-dekodéru pro odhad všech základních os je hlavní myšlenkou. Jako architekturu sítě jsme použili regresní 3D V-Net [10], jejímž vstupem jsou 2D MR tomografické řezy 3D objemem a výstupem jsou odhadnuté distanční mapy určující pozici každé hledané roviny ve vstupních datech. Momentálně probíhá odhad SA, 4CH a 2CH os současně. Základní schéma zpracování obrazů je ukázáno na Obrázku 3. Výsledky byly přijaty k prezentaci na mezinárodní konferenci [11].

Následně je v plánu tento postup rozšířit o další běžně používané pohledy aby pracoval jak s isotropickými voxely, tak i s řídkým vzorkováním ve směru snímání (velmi nízký počet tomografických řezů). Testování algoritmů bude probíhat jak v rámci spolupráce s FNUSA na klinických snímcích, tak i na dalším externím pracovišti ve VFN Praha, kde jsou aktuálně řešeny rychlé snímací protokoly umožňující získ isotropických 3D snímků ve vysokém rozlišení.



Obrázek 3: Schéma odhadu základních radiologických pohledů MR srdce. Vstupem jsou původní axiální tomografické řezy, které jsou zpracovány modelem strojového učení (architektura typu V-Net). Výstupem z modelu jsou odhadnuté distanční mapy každé hledané roviny, na základě nichž jsou z původních dat interpolovány tomografické řezy v požadovaných pohledech.

ZÁVĚR

Rozvoj metod strojového učení je patrný ve všech odvětvích lidského snažení a není výjimkou ani oblast biomedicínské inženýrství a zpracování obrazů. Stále je však třeba překonat celou řadu technických překážek (interpretovatelnost a důvěryhodnost modelů, dostupnost řádně anotovaných dat, výpočetní a paměťová náročnost) a zejména legislativních (oprávněný přístup k patientským datům a ochrana osobních dat) a etických překážek, které dle názoru autorů stále nejsou dostatečně řešeny.

Během posledních let se naši výzkumné skupině na ÚBMI podařilo vyřešit většinu technicko-legislativních problémů za pomoci smlouvy o spolupráci mezi klinickým pracovištěm a univerzitou, kdy neoprávněné osoby nemají přístup k neanonymizovaným datům a současně žádá (ani anonymizovaná) data neopouštějí vnitřní síť klinického pracoviště. Díky této spolupráci se nám otevřely možnosti společného vědeckého výzkumu a řešení absolventských prací, a také možnosti klinického testování vyvinutých algoritmů.

PODĚKOVÁNÍ

Autoři článku by rádi poděkovali profesorovi Jiřímu Janovi, docentu Jiřímu Vaníčkovi a docentu Romanu Panovskému za iniciaci a formulaci dílčích výzkumných projektů, a také smlouvy o spolupráci. Dále Ing. Tomáši Holečkovi a všem zaměstnancům FNUSA za aktivní a produktivní vědecko-výzkumnou spolupráci, dále. V neposlední řadě děkujeme společnosti Philips za dlouholetou finanční a technickou podporu.

LITERATURA

- [1] SCHEETZ, Jane; ROTHSCCHILD, Philip; MCGUINNESS, Myra; HADOUX, Xavier; SOYER, H. Peter et al., 2021. A survey of clinicians on the use of artificial intelligence in ophthalmology, dermatology, radiology and radiation oncology. Online. *Scientific Reports*. Roč. 11, č. 1. ISSN 2045-2322. Dostupné z: <https://doi.org/10.1038/s41598-021-84698-5>. [cit. 2023-08-31].
- [2] TĚŠITELOVÁ, Vladimíra; POLICAR, Radek; BLAHA, Milan; KLIMEŠ, Daniel a DUŠEK, Ladislav, 2018. Jak implementovat Nařízení evropského parlamentu a rady (EU) 2016/679. Online. In: <https://www.mzcr.cz/>. Ústav zdravotnických informací a statistiky ČR. [cit. 2023-09-06].
- [3] DARZIDEHKALANI, Erfan; GHASEMI-RAD, Mohammad a VAN OOIJEN, P.M.A., 2022. Federated Learning in Medical Imaging: Part I. Online. *Journal of the American College of Radiology*. Roč. 19, č. 8, s. 969-974. ISSN 15461440. Dostupné z: <https://doi.org/10.1016/j.jacr.2022.03.015>. [cit. 2023-08-31].
- [4] PARISOT, Charles, 1995. The DICOM standard. Online. *The International Journal of Cardiac Imaging*. Roč. 11, č. 3, s. 171-177. ISSN 0167-9899. Dostupné z: <https://doi.org/10.1007/BF01143137>. [cit. 2023-08-31].
- [5] NEMCEKOVA, Petra; HOLECEK, Tomas; CHMELIK, Jiri; OUREDNICEK, Petr; VALIS, Katerina; and JAKUBICEK Roman, 2023. Establishing the Optimal Standard for Preprocessing Head CT Data in Diagnostic Analysis. In: MEDICON'23 & CMBEBIH'23. Accepted paper.
- [6] NEMCEKOVA, Petra; SKRVAN, Adam; MARQUERING, Henk; CHMELIK, Jiri a JAKUBICEK, Roman, 2023. Analýza vzťahov medzi radiomickými príznakmi heterogenity trombu v akútnych ischemických mozgových príhodách. In: Trendy v biomedicínskom inženýrství 2023. Accepted paper.

- [7] RONNEBERGER, Olaf; FISCHER, Philipp a BROX, Thomas, 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. Online. In: NAVAB, Nassir; HORNEGGER, Joachim; WELLS, William M. a FRANGI, Alejandro F. (ed.). *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Lecture Notes in Computer Science. Cham: Springer International Publishing, s. 234-241. ISBN 978-3-319-24573-7. Dostupné z: https://doi.org/10.1007/978-3-319-24574-4_28. [cit. 2023-09-05].
- [8] ISENSEE, Fabian; JAEGER, Paul F.; KOHL, Simon A. A.; PETERSEN, Jens a MAIER-HEIN, Klaus H., 2021. NnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Online. *Nature Methods*. Roč. 18, č. 2, s. 203-211. ISSN 1548-7091. Dostupné z: <https://doi.org/10.1038/s41592-020-01008-z>. [cit. 2023-09-05].
- [9] TOBON-GOMEZ, Catalina; GEERS, Arjan J.; PETERS, Jochen; WEESE, Jurgen; PINTO, Karen et al., 2015. Benchmark for Algorithms Segmenting the Left Atrium From 3D CT and MRI Datasets. Online. *IEEE Transactions on Medical Imaging*. Roč. 34, č. 7, s. 1460-1473. ISSN 0278-0062. Dostupné z: <https://doi.org/10.1109/TMI.2015.2398818>. [cit. 2023-09-05].
- [10] MILLETARI, Fausto; NAVAB, Nassir a AHMADI, Seyed-Ahmad, 2016. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. Online. In: *2016 Fourth International Conference on 3D Vision (3DV)*. IEEE, s. 565-571. ISBN 978-1-5090-5407-7. Dostupné z: <https://doi.org/10.1109/3DV.2016.79>. [cit. 2023-09-05].
- [11] JURCA, Jan; HARABIS, Vratislav; JAKUBICEK, Roman; HOLECEK, Tomas; NEMCEKOVA, Petra; OUREDNICEK, Petr a CHMELIK, Jiri, 2023. Deep-learning based automatic determination of cardiac planes in survey MRI data. In: *2023 MEDICON'23 & CMBEBIH'23*. Accepted paper.