



Multimodal Medical Image Processing Methods for Computer-Aided Diagnosis Support System of Brain Tumors

Submitted by
Eстера Kot, MSc Eng

Main Advisor
Krzysztof Siwek, PhD, DSc, University Professor

Secondary Advisor
Zuzanna Krawczyk, PhD, DSc

Warsaw, 2023

Abstract

Warsaw, 2023

Doctor of Philosophy

**Multimodal Medical Image Processing Methods for Computer-Aided Diagnosis
Support System of Brain Tumors**

by Estera Kot, MSc Eng

Faculty of Electrical Engineering

Ph.D. Thesis

**Multimodal Medical Image Processing Methods for Computer-Aided Diagnosis
Support System of Brain Tumors**

by Estera Kot, M.Sc. Eng.

The PhD thesis presents the proposed multi-modal medical image processing methods based on multi-stage deep learning pipeline that combine multiple neural networks of various architectures which can perform meaningful detection and segmentation of brain tumors and can be utilized to support radiologists during both the diagnostic stage and treatment planning process. The driving force behind the research was the realization that radiological assessments were being conducted with limited pools of data and the contention that collective knowledge could be leveraged with Artificial Intelligence and Data-Driven approaches. Taking into consideration glioblastoma multiforme, which is a devastating cancer with the potential to cause death within six months, the goal of any established treatment plan is to extend life expectancy while maintaining the patient's quality of life and functionality. Radiologists use computed tomography (CT), positron emission tomography (PET) and magnetic resonance imaging (MRI) scans to determine the tumor localization, size, and stage of advancement, as well as the amount of radiopharmaceutical which should be injected into the tumor during invasive examinations or postoperatively after resection. These substances are intended to damage tumor cells along with normal, noncancerous brain cells. Radiologists and oncologists evaluate the volume of radiopharmaceutical administered and the probability of local cure TCP (Tumor Control Probability) based on microdosimetry methods, which involve dosimetric measurements of radiation exposure. A volume of radiopharmaceutical which is too high has the potential to damage healthy brain cells, while a volume that is too low will not effectively eradicate cancer cells. At present, the measurement of glioma volume is typically performed manually by radiologists, which is a time-consuming process and prone to human error. The challenge that remains to be addressed is the accurate determination of the percentage of the entire tumor tissue volume that has been reached by the radiopharmaceutical. This value is closely related to the tumor volume and thus is critical to the success of the modern approach. From the author's position, there appeared to be a lack in the processing of all multimodal medical data from MRI, CT and PET (on which one can observe the metabolic activity of tumors) examinations for which calculations of the surface area and volume could be conducted specifically for patients with gliomas. Additionally, most of the work being conducted deals with the models themselves as opposed to the holistic implementation of the models. This gap includes the fundamental area of "explainability" of the models and applicability of the methods to healthcare providers and the medical industry at large.

To this end, computer-based methods have been developed that automatically detect tumors, define their localization, visualize their shape, and calculate their volume. The thesis introduces the standard for the development of medical data processing solutions by defining a universal six-stage MeDAPR framework. The author explored classical active contour algorithms and implemented methods to compute tumor volume based on Edge and Chan-Vese approaches. The author designed and implemented a novel method for removing bone tissues from CT-correlated PET scans. Additionally,

the author designed and implemented a Phantom-based Verification and Calibration Method (PVCN) to validate and evaluate the accuracy of algorithms and assistance in setting the correct thresholds. Furthermore, the author implemented a method combining U-Net and the Morphological Geodesic Active Contour algorithm. To fuse multimodal images such as CT, PET, and MR produced by various diagnostic devices, the author introduced the CNN-based methods. The author created tools which generate additional training data consistent with the original and to unify labels provided for the medical data. The author designed and implemented Tumor Area Cropper with U-Net architecture generator as, based on the authorial novel equation that allows calculation of the U-Net architecture depth, based on the image input size. The author created a series of research experiments with the usage of the U-Net model to evaluate the accuracy of the segmentation based on the fused CT and PET scans. Additionally, the author introduced multiple methods based on convolutional neural networks (CNNs), YOLO v4, Mask R-CNN, and U-Net, combining the modeling into pipelines with one or two stages, to perform efficient and accurate image segmentation and classification tasks. The author elaborated on hyperparameter tuning of Mask R-CNN, including verification of backbones. The originality of the proposed solution relates also to the utilization of different multimodal medical data in the experiments.

Moreover, the author states that cloud-based technologies are necessary to build a complete solution targeting healthcare institutions - the universal automated cloud architecture developed, which allows for training and for running models in batch mode, can be implemented in both medical and research units working on medical data. The author proposed methods for parallel processing. The proposal of a novel architecture for federated learning - allowing multiple parties to collaboratively train a model without sharing their data - is also made in the dissertation. Additionally, the research explores the use of explainable AI methods, such as Grand-CAD, to enhance the interpretability of the CAD system and ensure trustworthiness and transparency. The work presents innovative solutions based on knowledge of the fields of computer science, computer vision, image processing, and design of CAD systems which utilize introduced state-of-the-art techniques. The work proposes an original solution in the field of applying the results of the author's scientific research in the healthcare industry.

In this work the author proposes several ideas to continue the work of redefining the application of new technologies in healthcare, particularly in the context of medical imaging. The newly introduced methods and results obtained in the research as presented in this thesis have a high potential for implementation in industrial and commercial practice. This research may lead to the development of new products and services, further algorithms, or tools in the field of medical imaging, and development of new methods in the field of computer vision image analysis with robotics or autonomous machines where precise object detection based on grayscale images is required. The author's deep belief and intention was to develop methods, algorithms, and procedures for the fully automated detection and segmentation of specific objects – mainly brain tumors, and the prostate – while integrating the necessary work of medical health professionals along with best practices for patients.

Keywords: medical image processing, computer vision, tumor detection, tumor segmentation, CAD, deep learning

Abstract

Warsaw, 2023

Doctor of Philosophy

**Multimodal Medical Image Processing Methods for Computer-Aided Diagnosis
Support System of Brain Tumors**

by Estera Kot, MSc Eng

Wydział Elektryczny Politechniki Warszawskiej

Rozprawa doktorska

Metody przetwarzania wielomodalnych obrazów medycznych do systemu komputerowego wspomagającego diagnozowanie guzów mózgu

mgr inż. Estera Kot

Rozprawa doktorska przedstawia metody przetwarzania wielomodalnych obrazów medycznych, bazujące na kilkuetapowym procesie opartym na modelach sieci neuronowych o różnorodnych architekturach, a umożliwiające skuteczną detekcję i segmentację guzów mózgu. Metody te mogą być stosowane jako wsparcie komputerowe dla radiologów - zarówno podczas diagnostyki obrazowej, jak i planowania leczenia. Motywacją do podjęcia tego tematu w pracy doktorskiej była świadomość tego, jak czasochłonnym zadaniem jest dokonywanie ocen radiologicznych przeprowadzanych manualnie przez wysoko kwalifikowanych ekspertów na podstawie ograniczonych danych medycznych. Współcześnie dostępne algorytmy sztucznej inteligencji przy zastosowaniu podejścia opartego na przetwarzaniu i analizie danych medycznych mogą posłużyć jako źródło niezależnej opinii przy detekcji zmian, opisie, ocenie właściwości choroby i interpretacji stopnia jej zaawansowania.

Jednym z najbardziej agresywnych nowotworów mózgu jest glejak wielopostaciowy IV stopnia (nowotwór gleju gwiaździstego), który cechuje się złym rokowaniem (większość chorych nie przeżywa 6 miesięcy od momentu postawienia diagnozy). Celem leczenia pacjentów ze zdiagnozowanym glejakiem jest przedłużenie ich oczekiwanej długości życia przy jednoczesnym podtrzymaniu jakości życia oraz sprawności.

Radiolodzy bazują na wynikach badań tomografii komputerowej (CT), pozytonowej tomografii emisyjnej (PET) oraz obrazowania rezonansem magnetycznym (MRI) w celu określenia umiejscowienia, wielkości i stadium zaawansowania nowotworu, jak również docelowo objętości radiofarmaceutyku, który jest podawany do łoża pooperacyjnej powstajej po resekcji guza. Podana substancja ma na celu uszkodzenie komórek nowotworowych. Objętość podanego radiofarmaceutyku i prawdopodobieństwo eradykacji nowotworu (TCP - Tumor Control Probability) radiolodzy i onkolodzy oceniają na podstawie metod mikrodozymetrii, które obejmują pomiary dawek radiacyjnych. Jeśli objętość radiofarmaceutyku jest zbyt duża, to na skutek jego podania mogą zostać uszkodzone również zdrowe komórki mózgu. Tymczasem zbyt mała objętość radiofarmaceutyku nie będzie w stanie skutecznie zniszczyć wszystkich komórek nowotworowych. Obecnie pomiar objętości glejaka wykonywany jest ręcznie przez radiologów przy zastosowaniu oprogramowania pozwalającego na zaznaczanie jedynie prostych form geometrycznych, przez co proces ten jest czasochłonny i obciążony ryzykiem błędu. Problemem, którego rozwiązanie należy opracować, jest tu określenie procentowo całkowitej objętości tkanki guza, do którego dotarł radiofarmaceutyk.

Wspomniana wartość procentowa jest ściśle związana z objętością guza i ma kluczowe znaczenie dla powodzenia współcześnie stosowanej metody leczenia. Autorka zwraca uwagę, że brakuje obecnie metod przetwarzania badań obrazowych mózgu pochodzących z różnych modalności, tzn. MRI, CT i PET (obrazującego aktywność metaboliczną guzów i ich przerzutów), na podstawie których wykonywane jest automatyczne obliczanie powierzchni i objętości guzów. Ponadto większość opracowań traktuje o samych modelach oraz ich skuteczności, pomija zaś omówienie możliwości

wdrożenia tych modeli przez dostawców usług opieki zdrowotnej i wyjaśnienia dotyczące późniejszych działań.

W opisanym kontekście autorka opracowała zestaw metod komputerowych automatycznie wykrywających umiejscowienie guzów, wizualizujących ich kształt i obliczających objętość. W niniejszej pracy zdefiniowano uniwersalny sześciostopniowy system (MeDAPR) rozwiązań przetwarzania danych medycznych przy przeprowadzaniu detekcji i segmentacji guzów. Na podstawie klasycznych algorytmów aktywnego konturu autorka przygotowała algorytmy obliczania objętości glejaków oparte na podejściu typu Edge i Chan-Vese. Opracowano metodę usuwania tkanki kostnej z obrazów PET skorelowanych z badaniami CT. Ponadto autorka stworzyła nowatorską metodę ewaluacji parametrów algorytmów realizujących mierzenie pola powierzchni i objętości guzów mózgu oraz ocenę ich dokładności - metodę wykorzystującą wyniki badań prowadzonych w toku doświadczeń z użyciem fantomu fizycznego PET-CT (PVCm). Opracowano również metodę łączącą sieć neuronową U-Net i algorytm morfologicznego geodezyjnego aktywnego konturu służącą do precyzyjnego wyliczania objętości glejaka. By umożliwić łączenie wielomodalnych obrazów medycznych takich jak CT, PET i MRI, tj. pochodzących z różnych urządzeń diagnostycznych, wprowadzono metody fuzji wspomnianych badań bazujących na sieciach konwolucyjnych. Autorka opracowała własne narzędzia generujące dodatkowe dane treningowe oraz narzędzia umożliwiające ujednolicenie etykiet (nazw klas) dostarczanych dla danych medycznych przez radiologów. Zaprojektowano i implementowano narzędzie służące wycinaniu obszaru guzów ze skanów, połączone z generatorem architektury U-Net, który z kolei oparty jest na autorskim nowym równaniu pozwalającym obliczyć głębokość architektury modelu U-Net na podstawie wielkości wejściowego obrazu. Autorka przeprowadziła serię eksperymentów badawczych z użyciem modelu U-Net w celu dokonania oceny dokładności segmentacji opartej na danych będących fuzją badań CT i PET. Wprowadziła ponadto zestaw metod bazujących na konwolucyjnych sieciach neuronowych (CNN) takich jak YOLO v4, Mask R-CNN i U-Net, łącząc modele w jedno- lub dwustopniowe procesy, co ma służyć skuteczniejszemu i dokładniejszemu przeprowadzaniu zadania klasyfikacji i segmentacji. W niniejszej pracy omówiono dobieranie hiperparametrów dla modelu Mask R-CNN, w tym weryfikację różnych kręgosłupów sieci. Oryginalność proponowanego rozwiązania związana jest również z faktem, że opracowane metody każdorazowo wykorzystano i zweryfikowano na wielomodalnych danych medycznych.

Autorka kładzie szczególny nacisk na fakt, że technologie wykorzystujące chmurę obliczeniową mają rolę kluczową dla zbudowania kompleksowego rozwiązania, które skierowane będzie do instytucji opieki zdrowotnej. Opracowano tu uniwersalną architekturę chmurową umożliwiającą trenowanie oraz uruchamianie modeli sztucznej inteligencji w trybie wsadowym. Architektura ta może znaleźć zastosowanie zarówno w jednostkach medycznych jak i badawczych, pracujących nad danymi medycznymi. Zaproponowano także metody równoległego przetwarzania danych. Niniejsza praca przedstawia także propozycję nowatorskiej architektury dla uczenia federacyjnego — pozwalającej licznym instytucjom na wspólne trenowanie modeli sztucznej inteligencji bez konieczności udostępniania swoich danych. Dodatkowo autorka poruszyła kwestię użycia metod sztucznej inteligencji takich jak Grand-CAD w celu zwiększenia interpretowalności (tj. zrozumienia tego, w jaki sposób model uczenia maszynowego podjął

decyzję) systemu komputerowego wspomagania diagnozy CAD (computer-aided diagnosis) oraz poprawienia wiarygodności oraz przejrzystości wyników zwracanych przez opracowane modele.

Niniejsza praca doktorska prezentuje innowacyjne rozwiązania oparte na wiedzy z dziedzin informatyki, wizji komputerowej, przetwarzania obrazów oraz projektowania systemów CAD, bazujące na nowoczesnych i opracowanych metodach. Zawiera ponadto propozycje oryginalnych rozwiązań w dziedzinie wdrażania i zastosowania wyników własnych badań naukowych w sektorze służby zdrowia.

Autorka przedstawia wreszcie propozycje kontynuacji prac badawczych w obszarze zwiększania zastosowania nowych technologii komputerowych w opiece zdrowotnej, zwłaszcza w kontekście przetwarzania danych medycznych. Opracowane metody i uzyskane wyniki mają potencjał wdrożeniowy w praktyce przemysłowej i komercyjnej, mogą ułatwić implementację nowych produktów i usług, kolejnych algorytmów, czy narzędzi w dziedzinie przetwarzania obrazów medycznych oraz rozwijania nowych metod w dziedzinie analizy obrazów z zakresu wizji komputerowej, robotyki czy autonomicznych maszyn - a także innych obszarów, w których prace wymagają precyzyjnego wykrywania obiektów na podstawie obrazów w skali szarości. Intencją autorki było opracowanie metod, algorytmów i procedur automatycznego komputerowego wykrywania i segmentacji obiektów – głównie guzów mózgu i prostaty – przy jednoczesnej współpracy ze specjalistami opieki zdrowotnej, ostatecznie w celu uzyskania poprawy skuteczności w diagnozowaniu pacjentów.

Słowa kluczowe: przetwarzanie obrazów medycznych, widzenie komputerowe, wykrywanie guzów, segmentacja guzów, komputerowe wspomaganie diagnostyki, głębokie uczenie.

Contents

Abstract	i
1 Introduction	1
1.1 Research Aim and Thesis Statement	3
1.2 Overview of the Thesis Contents	5
2 Fundamentals of Medical Imaging	6
2.1 Computed Tomography: A Fundamental Component	6
2.2 The Basics of Positron Emission Tomography	7
2.3 The Foundations of Magnetic Resonance Imaging	8
2.4 Microdosimetry in the Treatment of Brain Tumors	9
2.5 Novel Treatment Strategy	12
3 Literature Review	15
3.1 Computer-Aided Detection Applications	15
3.2 Multi-modal Image Processing and Segmentation Methods	17
3.3 Critical Appraisal	22
4 Introduction of a Medical Data Processing Framework	24
4.1 The Key Components of MeDAPR Framework	24
4.2 Capabilities of Convolutional Neural Networks in Medical Image Processing	25
4.2.1 Key Concepts of CNNs	26
4.2.2 U-Net Model	28
4.2.3 Mask R-CNN Model	30
Mask R-CNN Critical Assessment	31
4.2.4 YOLOv4 Model	34
4.2.5 Key Differences and Strengths	35
5 Active Contour Approach for Image Segmentation	37
5.1 Bone Tissue Removal and Active Contour	38
5.1.1 Edge-based Method	42
5.1.2 Chan-Vese-based Method	42
5.1.3 Appraisal	43
5.2 Phantom-based Verification and Calibration Method (PVCm)	44
6 Convolutional Neural Network-based Approach for Segmentation	49
6.1 U-Net with Morphological Geodesic Active Contour	49
6.1.1 Improved Morphological Geodesic Active Contour	51
6.1.2 Results and Evaluation	53

6.2	Multimodal Data Fusions Algorithms	56
6.2.1	CT and PET Scans Fusion Algorithm	56
	VGG19 and VGG19-BN	56
6.2.2	CT, PET and MRI Scans Fusion Algorithm	58
	Fusion of PET and CT	59
	Fusion of CT-PET with MRI	60
	Final Fusion of CT, PET and MRI	62
6.3	Data Augmentation and Annotations	64
6.3.1	Mask-aware Medical Data Augmentation	64
6.3.2	Medical Annotations Transformer (MAT)	68
6.4	Tumor Area Cropper with Connected U-Net Architecture Generator . .	69
6.5	U-Net-based Segmentation on CT and PET Scans	71
6.5.1	Processing Methods and Experiments	73
6.5.2	Training on the Original Dataset	73
	Single Tumor Extracted	73
	Multiple Tumors Extracted	74
6.5.3	Training on the Augmented Dataset	75
	Augmentation with AugLy	75
	Single Tumor Extracted Augmented	77
	Multiple Tumors Extracted Augmented	77
6.5.4	Results and Discussion	77
6.6	Multi-stage Pipeline	79
6.6.1	Prostate Detection based on YOLOv4 and U-Net (YU-Med) . . .	79
6.6.2	Brain Tumor Segmentation based on Mask R-CNN and U-Net (MU-Med)	83
	Tumor Detection by Mask R-CNN	84
	Tumor Segmentation by U-Net	86
	Update of the MeDAPR Framework	88
6.6.3	Mask R-CNN based Hyperparameters Tuning Procedure	90
	Datasets and Preprocessing	91
	Model Training and Transfer Learning	91
	Hyperparameter Optimization and Results	92
7	Proposed Novel Architecture for the Healthcare Industry	97
7.1	Cloud-native Proposed Architecture	99
7.1.1	Medical Data Ingestion and Storing	101
7.1.2	Training Machine Learning Models for Healthcare Applications and Cloud Deployment	102
7.1.3	Parallel Processing	102
	Proposed Method	103
7.1.4	Visualizing Medical Data	107
7.1.5	Targeting User Needs with Persona-Based Approach	108
7.1.6	Operationalization of ML Models	109
7.1.7	An Applicable Approach to Federated Learning	111
7.1.8	Cost-Effective Cloud Development Strategy	114
7.2	Ensuring Security in Medical Imaging	115
7.3	Explainability in Enhancing Transparency and Trust	116

7.4	An Ethical Framework for Medical Image Analysis	117
8	Conclusions	119
8.1	List of Novelties	119
8.2	Summary	121
8.2.1	Broader Impact	123
8.3	Future Directions	124
9	Acknowledgements	128
	Publications	129

List of Figures

1.1	The impact of increasing applications across different industries and regulations published by Allied Market Research in 2022 [1].	3
2.1	MRI of the head with contrast illustrating a glioblastoma multiform in a 15-year-old patient. Figure provided by the collaborative radiologist, Czwarnowski, P.	10
2.2	The intracerebral administration method described above necessitates a flexible container-catheter set. The container is designed to be flexible and the catheter is surgically inserted into the tumor volume or post-operative cavity after tumor resection (Fig. 2.4). The container remains outside the skull, positioned just beneath the skin surface. Created by the author based on the initial draft provided by the collaborative radiologist, Czwarnowski, P.	11
2.3	An example of reflux resulting from a clogged catheter or too rapid administration of a radiopharmaceutical. Created by the author with the usage of RadiAnt [2] software.	11
2.4	Treatment environment for glioma tumors. Created by the author based on the initial materials provided by the collaborative radiologist.	13
2.5	PETsyngo 6.7.3, VB42A from Siemens, the resulting image obtained by an Subject Matter Expert (SME). Figure provided by the collaborative radiologist, Czwarnowski, P.	14
4.1	Implementation of the U-Net architecture. Legend starting from the left: red - transformation to the input, blue - Conv2D - convolution layer with 3x3px filter and ReLu activation function, yellow - MaxPooling2D with 2x2px - down-samples the input size, violet - dropout - decrease the input size, green - UpSampling2D - 2x2px up-convolution layer, gray - concatenation of the output of the layer, turquoise - Conv2D layer with sigmoid activation function [3].	29
4.2	Original masks for two patients.	32
4.3	Detected objects (classes) with their corresponding IoU scores returned by Mask R-CNN model presented on top of the original scans.	32
4.4	Segmented detected objects (classes) with their corresponding IoU scores returned by Mask R-CNN model.	33
4.5	Two results returned by Mask R-CNN model with top anchor scores.	33
4.6	Detections returned by Mask R-CNN model.	34

4.7	Implementation of YOLOv4 tiny custom architecture. Legend starting from the left: green - input images in NxN size, blue - convolutional layer with leaky-relu activation function, yellow - route layer's output concatenation, orange - max pooling that down-samples the input size, pink - convolutional layer with linear activation function, gray - YOLO object detection layer [4]	35
5.1	Active contour-based created algorithm created initially by the thesis author during the development of author's master thesis [5].	39
5.2	CT scans with automatically removed bone tissue.	40
5.3	The effect of automatic image binarization.	41
5.4	Removing bone structures from a CT image. Bone tissue is found on the left. In the middle, a CT image with dissected bone tissues is shown. On the right, an original CT image is presented.	41
5.5	Picture of the PET-CT phantom BIODEx MEDICAL SYSTEMS available at the Nuclear Medicine Department of the Medical University of Warsaw.	45
5.6	A bird's-eye view of the PET-CT phantom BIODEx MEDICAL SYSTEMS available at the Nuclear Medicine Department of the Medical University of Warsaw.	46
5.7	Diagram of the developed algorithm to assess it's accuracy based on the data from physical phantom.	47
5.8	Developed algorithm detects tumors on the phantom. Visualization done with the usage of RadiAnt [2].	48
5.9	Image obtained using the developed algorithm. On the left side, a two-layer PET image is visible with tumors identified using the active contour method. On the right side, the PET-CT fusion with the active contour overlaid is presented.	48
6.1	Flow diagram for developed method based on U-Net and active contour.	49
6.2	Qualitative results presented by the dissertation author during International Joint Conference on Neural Networks (IJCNN) in 2020 [6].	51
6.3	Histogram for fused images published. Results and methods have been published by the author in The Bulletin of the Polish Academy of Sciences Technical Sciences in 2021 [7]. As of June 2023, the article had received 10 citations.	58
6.4	Three different data sources. The left picture shows the CT picture, the central picture the PET and the right picture the MRI for the identical slice of a single patient. The CT and MRI data differ from each other in terms of the color of healthy and tumor-infected tissues.	59
6.6	The CT-PET fusion. On the right side of the above figure is an image that is the result of applying a fusion. The two images on the left show the components of the fusion. Looking at the fusion image, a CT image with the tumor sites indicated by PET is seen. Thus, the conclusion can be drawn that the obtained image combines the diagnostic features of two component images.	60

6.7	Transformation performed by the detection of diagnostic features. The left photo shows image before transformation, the right photo shows image after features extraction.	61
6.8	This figure shows final fusion. On the left the picture shows the CT-PET image, the central picture is of the MRI and the right picture of the fusion result.	63
6.9	The result of the work is an algorithm that allows for automatic fusion of images from CT, PET and MRI.	63
6.10	CT-PET-MRI fusion diagram.	64
6.11	Authorial universal tool named Medical Annotations Transformer (MAT) which automate using labelled datasets by different deep learning models within the cloud environment.	68
6.12	General diagram of the 7th stage pipeline. Tumor area cropping is performed before tumor segmentation, for example via U-Net models. . . .	72
6.13	Image on the left side: Fused CT with PET scans. In the middle - Ground truth. On the right - Predicted mask.	74
6.14	Image on the left side: Fused CT with PET. In the middle - Ground truth. On the right - Predicted mask.	74
6.15	Image on the left side: Fused CT with PET. In the middle - Ground truth. On the right - Predicted mask for single slice of fused CT-PET where Dice Co-Eff equals to 0.6638.	75
6.16	Augmented fused CT-PET scans with corresponding mask with a usage of AugLy published by the thesis author in [8].	76
6.17	Image on the left side: Fused CT with PET. In the middle - Ground truth. On the right - Predicted mask for a single slice of fused CT-PET where Dice coefficient equals to 0.8945.	78
6.18	Area under the ROC Curve (ROC-AUC) value over epochs for the experiments.	79
6.19	Mean Squared Error (MSE) value over epochs for the experiments. . . .	81
6.20	Distribution of Dice coefficient for two experiments: Single Tumor Extracted and Single Tumor Extracted Augmented.	82
6.21	Post-processing workflow for prostate detection image. Legend starting from the left: input image in NxN size, YOLOv4 convolutional neural network, prostate detection output with yellow bounding-box, prostate detection output with enlarged turquoise bounding-box (15px padding), cropped prostate detection output and resized into 256x256px as U-Net input, U-Net convolutional neural network to segment prostate.	82
6.22	The values of the ROC-AUC per epoch for each experiment.	83
6.23	Over-detection on slices where glioma cannot be present due to its nature. .	86
6.24	Glioma detection based on YOLOv4 trained on fused CT-PET-MR scans. . .	89
6.25	Key components of the proposed framework.	90
6.26	The specific stages reflecting the cooperation between SME (e.g., radiologists) who provide the labelled data, and the continuous retraining of the model performed by the author.	91
6.27	Models with the highest mAP value. The value of mAP began to stabilize around 0.9 after the 20th epoch.	93

6.28	Training charts for the chosen configuration of hyperparameters (colored in blue at the Table 6.16).	95
6.29	Segmentation results are presented with bounding boxes, which localize the tumor area.	96
7.1	The architecture diagram of the healthcare analytics platform is provisioned on Azure that consists of Azure Services and Open-Source technologies. The solution based on a similar architecture won Microsoft Global Hackathon 2022 Executive Challenge - Hack for Industry - 1st Place.	99
7.2	Authorial solution of using Azure Data Factory (ADF) moves the medical data from on-premises hospital environment to Azure Data Lake Storage, breaking new ground in usage for ADF.	101
7.3	Sequential medical data processing.	103
7.4	Concurrent medical data processing.	104
7.5	Proposed architecture based on the containers approach.	105
7.6	Proposed architecture based on the containers approach with data partitioning.	105
7.7	OHIF Medical Imaging Viewer extended by adding triggering the organ scanning or specific tumor detection. Created by the author.	108
7.8	Brain visualizations for a single patient build on top of VTK library.	109
7.9	The designed operationalization procedure.	110
7.10	The designed operationalization procedure with the focus on researcher persona.	111
7.11	Federated learning - machine learning on decentralized data.	112
7.12	Federated learning solution diagram. Created by the author.	113
7.13	Federated learning designed case. Created by the author.	114
7.14	Costs analysis for an AKS cluster.	115
7.15	Predictions explained with Grad-CAM (LayerGradCam). Predicted: glioma tumor (0.6499). Mask R-CNN based on architecture ResNet50 trained on only 349 images. Created by the author.	117

List of Tables

5.1	Results (in cm^3) for detected and measured brain tumors provided by SME and computed by Edge and Chan-Vese methods.	43
5.2	Table with theoretical values, calculated according to the phantom's catalogue card. The total volume of all tumors is $21.316cm^3$	45
6.1	Results for 5 methods morphological active contour.	53
6.2	Visual results of applying morphological methods.	54
6.3	Results of Tumor Volumes per specific methods, expressed in cm^3	55
6.4	Calculations of RE and AE for received results for specific methods.	55
6.5	MSE, RMSE and MAE for achieved results for specific methods.	55
6.6	PET and corresponding CT scans and the resulted fusion done by methods based on VGG19 and VGG19-BN.	57
6.7	Jaccard coefficient results for nine outputs of the fusion method.	58
6.8	Results for the U-Net training on a dataset with single tumor extracted.	74
6.9	Results for the U-Net training on a dataset with multiple tumors extracted.	75
6.10	Results of U-Net training on an augmented dataset with a single extracted tumor.	77
6.11	Results for the U-Net training on an augmented dataset with multiple tumors extracted.	77
6.12	Performance Metrics Means for The U-Net. The Best Result per Metric in Bold and Blue.	78
6.13	Dice Co-Eff Results. Per Experiment the Highest Mean Value in Bold and Blue, the Highest Max Value in Bold and Teal.	80
6.14	Source images and detected and cropped tumors for six patients.	85
6.15	Source image and segmented tumor area achieved by method based on U-Net.	87
6.16	F1-score results for Mask R-CNN.	94
7.1	Comparing the two types of executors in the proposed solution.	106
7.2	Results for Single and Partition Executor Types.	107

List of Abbreviations

AI	Artificial Intelligence
AUC	Area Under the Curve
CAD	Computer-Aided Diagnosis
CNN	Convolutional Neural Network
CPU	Central Processing Unit
CT	Computed Tomography
DICOM	Digital Imaging and Communications in Medicine
DNN	Deep Neural Network
DSC	Dice Similarity Coefficient
FPR	False Positive Rate
GPU	Graphics Processing Unit
HU	Hounsfield Units
IOU	Intersection over Union
JI	Jaccard Index
ML	Machine Learning
MAE	Mean Absolute Error
MGAC	Morphological Geodesic Active Contours
MRI	Magnetic Resonance Imaging
MSE	Mean Squared Error
PET	Positron Emission Tomography
ROC-AUC	Area under the ROC Curve
ROI	Region Of Interest
SPECT	Single-Photon Emission Computed Tomography
YOLO	You Only Look Once

This thesis is dedicated to the memory of those who have passed away after receiving a late diagnosis or a misdiagnosis which resulted in a delayed course of treatment. It is my sincere desire that the use of emerging technologies in novel and highly effective ways in medical care will lead to a brighter future saving patients from suffering the same loss.

Chapter 1

Introduction

It is well accepted within even the general population that cancer as a grouping of related diseases is a significant global health issue. The statistics across 204 nations bear this out [9, 10]. Narrowing the scope to the United States, the probability in each person's lifetime of receiving an invasive cancer diagnosis is 40.2% in men and 38.5% in women [11]. In 2022, it is estimated that 1.9 million new cancer cases will be diagnosed, and 609,360 people in the United States will die from cancer, which corresponds to almost 1,700 deaths per day [11]. In the United Kingdom, the National Institute for Health and Care Excellence (NICE) predicts that one in two people will develop cancer at some point in their lives [12].

The COVID-19 (coronavirus) pandemic had an irreparable impact during 2020, with one consequence being hindering cancer diagnosis and treatment [11]. Adding the ensuing necessary adjustment in research efforts into the COVID/Cancer equation, one cannot deny the medical research landscape with regards to cancer has been permanently altered [9]. Regardless of the complications, solutions must be proposed in order to drive diagnoses and treatments in manners which positively impact patient processes and outcomes. Groundbreaking research is entirely possible, regardless of the circumstances, as this thesis proposes.

According to reports from Cancer Research UK, brain, other central nervous system (CNS), and intracranial tumors are the ninth most common group of cancers in the UK [13]. The central nervous system regulates the functions of the human body, therefore, damage or impairment to this area has the potential to affect all bodily functions [10, 11]. Since the 1990s, there has been a 39% increase in brain tumor incidence rates. Gliomas are a type of tumor which occurs in the brain or spinal cord, deriving from glial cells that form the stroma of nerve tissue. Gliomas make up about 70% of all intracranial tumors, with glioblastoma (GBM) being the most aggressive primary brain tumor [14]. GBM is classified as a grade IV tumor by the World Health Organization with a mortality rate of 100% if left untreated [15]. GBM is a devastating cancer with the potential to cause death within six months, having a median survival time of 14.6 months for patients who follow a treatment plan [16]. The goal of any established treatment plan is to extend life expectancy while maintaining the patient's quality of life and functionality. Despite efforts to prevent cancer and advances in highly specialized medical equipment for cancer treatment, the number of cancer patients (which includes GBM) continues to increase every year [17].

One of the oldest codes of conduct for all members of the medical profession is the Declaration of Geneva, based on the Hippocratic Oath, states in part "The health of

my patient will be my first consideration” [18]. One is compelled to apply this declaration in the face of GBM, given the mortality rate. Obtaining clinically relevant and meaningful diagnostic information in the case of gliomas demands that doctors, nuclear medicine physicians, and radiologists rely heavily on medical imaging to determine diagnoses and prognoses. The broad use of various imaging modalities such as Computed Tomography (CT), Computed Tomography-Positron Emission Tomography (CT-PET), and Magnetic Resonance (MR) imaging is commonplace.

This modern healthcare approach focusing on various data has led to an increase in overall patient data, including medical images. Artificial intelligence techniques have become increasingly successful in radiology, specifically for image-based diagnosis, disease detection, and treatment planning [19]. Medical imaging techniques are divided into structural and functional methods. Structural methods such as MRI and CT produce high-resolution scans with structural and anatomical information. Functional methods such as PET and single photon emission computed tomography (SPECT) produce low-resolution scans with functional information. Medical image fusion combines data from one or more modalities into a single image while maintaining all essential data and image quality. This fused image is then used for further manual or automated image analysis. Fusion increases clinician capacity providing additional context during the critical stage of the diagnosis and treatment process. Clinicians are assisted by computerized solutions in tumor area detection and localization with glioma tissue segmentation – measurement of size and volume, and analysis utilizing a blending of CT, PET and MRI scans.

Vast improvements have been made in the diagnostic efficiency in assessing abnormal tissue by the use of computer solutions. Due to the high incidence of central nervous system cancers worldwide, there is a critical need to develop computer-aided diagnosis (CADx) and computer-aided detection (CADE) methods that can significantly impact healthcare approaches and patient care. Further to this end, the Department of Nuclear Medicine of the Warsaw Medical University, located in the Central Clinical Hospital of the University Clinical Center of the Warsaw Medical University, has been conducting clinical trials for a novel treatment strategy for patients with stage four glioblastoma multiforme (GBM). The research presented in that dissertation was developed in collaboration with and for these clinicians.

The demand for image recognition technologies has increased due to their ability to identify and analyze visual data, especially in medical imaging. The market for image recognition is experiencing significant growth, with investment in the healthcare industry being the highest among all industries at 28.4% [1]. The global image recognition market is projected to grow at a compound annual growth rate (CAGR) of 21.8% over the period 2017–2025 which is presented in Fig. 1.1. This growth is due to the increasing adoption of image recognition technology in various industries, including healthcare, automotive, retail, and security, among others.

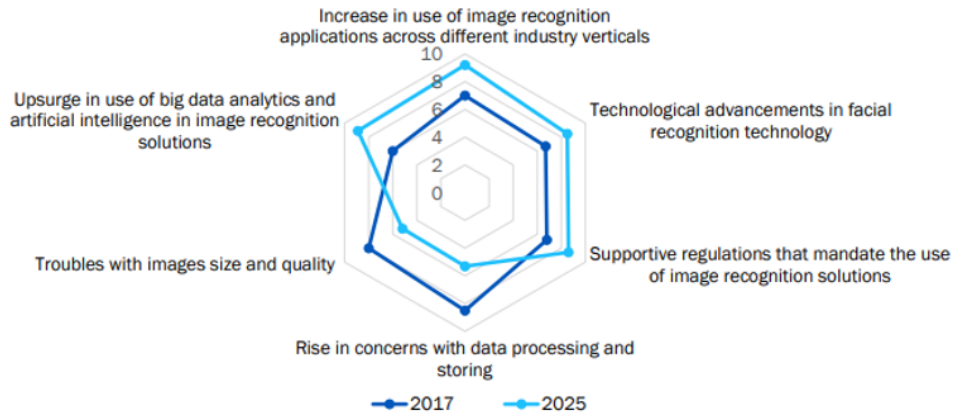


FIGURE 1.1: The impact of increasing applications across different industries and regulations published by Allied Market Research in 2022 [1].

1.1 Research Aim and Thesis Statement

The problem addressed in this thesis involves the processing of multimodal medical images, which varies from patient to patient. A comprehensive view of the patient's scans is required, including fusion of data from distinct imaging modalities, such as PET and CT, to facilitate better visualization and diagnosis. Further, the thesis aims to develop methods for measuring the surface area and volume of brain tumors. A supplemental objective is the creation of a system for conducting screening tests, to improve the radiologist's accuracy and efficiency by focusing attention on specific identifiable structures. The primacy of the medical problem addressed in this dissertation is undeniable, as the aggressive rate of progression of GBM and a 100% mortality rate amplifies the significance of any effort to prolong and improve the quality of patient's lives, during concerted efforts to cure them.

The objective of this research was to develop algorithms, methods, and procedures for the automated detection and segmentation of brain tumors based on fused multimodal medical images. The author utilized advanced techniques from AI, computer vision, software engineering, and platform-based development to devise innovative solutions for measuring brain tumor size and volume, while also providing a reproducible process that can serve as a decision support system. Several methods, algorithms, and ideas from the body of the thesis research have been published in relevant academic journals and presented at conferences to showcase the originality and innovative approach, as well as to gather feedback from experts in the field.

The conducted research has been divided into assessment and introducing new methods grounded in:

1. Classical active contour algorithms.
2. Convolutional neural networks, including critical assessment of multiple models.
3. Multimodal medical data fusion strategies.

4. Cloud-based technologies to build a complete solution targeting healthcare institutions.

The conducted examinations, created innovations and developed methods presented as results of the research done between the years 2016 and 2023, aim to prove the primary objective of the thesis: **The methods based on a proposed multi-stage deep learning pipeline that combine multiple neural networks of various architectures are capable of performing meaningful detection and segmentation of brain tumors and are able to be utilized in order to support radiologists during both the diagnostic stage and treatment planning process.**

In addition to the primary objective of the thesis, the author also set out to achieve several additional goals contributing to the computer science field applied to the healthcare industry, which are as follows:

- The introduction of MeDaPr - a Medical Data Processing framework - to structure the approach when working with medical data.
- The proposal of a novel architecture for Federated learning in collaboration with three institutions from the USA, Poland, and Germany.
- Evaluation of an approach which explains the results returned by deep neural network models used for the automated detection and segmentation of brain tumors.
- The leading of a discourse regarding the significance and potential impact of the thesis contributions for the field of computer science and its applications in healthcare and beyond.

The expertise and collaboration of medical professionals were provided by a team of radiologists and nuclear medicine physicians from the Nuclear Medicine Department of the Medical University of Warsaw, at the Central Clinical Hospital (*Banacha Szpital*), as well as a team of radiologists and oncologists from the Lower Silesian Oncology Center in Wrocław (*Dolnośląskie Centrum Onkologiczne*). The data was provided by the aforementioned institutions under conditional agreements and is not publicly available due to privacy and ethical considerations. In the context of the thesis, both qualitative and quantitative results play important roles in evaluating the effectiveness of the proposed methods and algorithms. Qualitative results focus on the visual inspection of the segmentation results, with experts in the field analyzing and evaluating the accuracy of the proposed methods. Qualitative results additionally serve to identify the strengths and weaknesses of the proposed methods, providing insights into areas that may require further improvement. Alternately, quantitative results involve measuring the performance of the proposed methods utilizing metrics such as accuracy, sensitivity, specificity, and the Dice score. These metrics provide a more objective evaluation of the proposed methods and are employed to compare the performance of different algorithms. Overall, both qualitative and quantitative results are vital in accurately evaluating the effectiveness of the proposed methods and provide valuable insights for future methodological improvements.

1.2 Overview of the Thesis Contents

The aim of this thesis was to create computer vision techniques, methodologies, and processes for identifying and segmenting brain tumors from multimodal medical images. The following content and issues are included in the following chapters:

1. **Introduction** – The first chapter describes the field of computer aided decision support systems in medicine. The main objectives of the dissertation are presented, and the thesis is described in detail.
2. **Fundamentals of Medical Imaging** – The second chapter emphasizes the significance of brain tumors, including glioma, as an extreme challenge even for modern medicine. The chapter covers the methods of diagnosis of brain tumors and the role of medical imaging currently applied in the diagnostic process.
3. **Literature Review** – The third chapter presents an overview of existing research on the computer-aided diagnosis of brain tumors. The chapter encompasses tools that allow for automated detection and segmentation of tumors and assesses which areas of research still require exploration and identification of new solutions.
4. **Introduction of a Medical Data Processing Framework** – The fourth chapter presents a structured approach towards medical data processing, categorizing all activities into defined stages.
5. **Active Contour Approach for Image Segmentation** – The fifth chapter presents the initial study and methods for developing computer vision algorithms using active contour methods. The chapter describes the methodology used and presents the results of the study.
6. **Convolutional Neural Network-based Approach for Segmentation** – The sixth chapter presents the secondary study and methods for developing computer-vision algorithms using deep neural networks. The chapter describes the methodology used and presents the results of the study.
7. **Proposed Novel Architecture for the Healthcare Industry** – The seventh chapter describes the industry deployment and architecture which was created as part of the research. The chapter covers the technical details of the deployed system and the challenges faced during deployment.
8. **Conclusions** – The eighth chapter summarizes the results and presents the conclusions from the conducted research. The chapter also provides a short summary of each novelty and highlights their potential impact on the field of computer-aided diagnosis.

Chapter 2

Fundamentals of Medical Imaging

Nuclear medicine is an essential medical imaging technique that makes use of a small quantity of radioactive material for diagnostic and therapeutic purposes. It is utilized for the diagnosis and treatment of various medical conditions, including but not limited to cancer, heart disease, and neurological disorders. Nuclear medicine imaging affords unique insights into the human body that are typically inaccessible via other imaging modalities such as X-rays, computed tomography (CT) scans, and magnetic resonance imaging (MRI). The ability to detect the presence of disease, determine the extent of the ailment and monitor the effectiveness of treatment is unparalleled. Furthermore, nuclear medicine can detect the presence of abnormal cells or tissues, such as tumors, and measure organ activity, such as the heart and lungs.

Notably, X-ray and gamma-ray photons are applied to provide unique insights into the human body; however, these types of radiation can cause significant harm to living cells and tissues. This holds true for other particle radiation such α and β . Radiation induced DNA damage can result in cell death, mutation, and carcinogenesis as the main target for radiation-induced biological effects is deoxyribonucleic acid (DNA), located in the nucleus of mammalian cells. Due to the human body's constant exposure to cosmic and terrestrial radiation, several DNA repair mechanisms have been developed to combat radiation damage. Thus, a meticulous evaluation and balancing of the potential benefits against the hazards of ionizing radiation must be conducted before any radiological examination [20].

This chapter provides an overview of the fundamental concepts of nuclear medicine as a diagnostic and therapeutic tool, as well as the types of medical imaging modalities and their acquisition. The chapter further describes novel treatment strategies and clinical trials currently being conducted at the Nuclear Medicine Department of the Medical University of Warsaw, located at the Central Clinical Hospital, focusing on patients diagnosed with glioblastoma.

2.1 Computed Tomography: A Fundamental Component

Computed tomography (CT) involves X-rays and the measuring of their attenuation in the human body. The emitted wave becomes weaker as it passes through human tissue. The design of the CT is constructed in such a manner that it is possible to observe even the smallest changes in the energy of the X-ray wave. On this basis, an absorption map of the cross-section of the human body is created, used for two-dimensional visualization. The output images are a series of equidistant sections of the human body, the most common images are 512x512 pixels, which correspond to voxels on the detector

array. A characteristic feature attributed to computed tomography is favorable contrast sensitivity, compared to magnetic resonance imaging (MRI).

In CT, the stored image in a DICOM file has up to 65,536 brightness levels. Humans are able to distinguish up to approximately 60 levels of gray, which in combination with the 65,536 levels in the stored image, makes the simple rewriting of the pixel value to the file in a lossless conversion an unattractive option. Such an image is of little use as it does not bear any distinguishable diagnostic features. In order to assess how the creators of medical software address this issue, it was necessary to become acquainted with the available software for displaying medical images. One example of such a program is Radiant [2]. Working with the application revealed an option in which it is possible to adjust the window of displayed gray levels. Such an operation, in essence, involves narrowing the brightness levels to a specific window, and then saving that window to a specified number of gray levels which is distinguishable to the human eye. Pixels with a frequency exceeding the indicated window are catalogued to the highest and lowest level in the applied scale. The aforementioned method was implemented in the Python programming language. The use of such an image formatting operation allows for the highlighting of image structures and emphasizes their characteristic features. In the case of the used files containing computed tomography data, the window was defined within the limits [1015-1105] which was recorded at 255 levels.

This selection of parameters allowed for the visualization of a satisfactory number of diagnostic features. The formula describing the applied transformation is as follows in Eq. 2.1.

$$\frac{1015}{x_{max}} \cdot 255 \leq \frac{x}{x_{max}} \cdot 255 \leq \frac{1115}{x_{max}} \cdot 255 \quad (2.1)$$

where x is the brightness of a single pixel.

CT scans present intracranial images of the soft tissues and structure of the brain surrounded by cerebrospinal fluid (CSF). The aspects of GBM make them nearly imperceptible on CT scans. Therefore, PET scans are typically utilized during brain examinations. The most accurate visualization of a tumor inside the brain utilizes the fusing of PET and CT results from the same CT-PET scanner. CT scans produce results which are reflections of the structure of the brain revealed as a level of grey, indicated by a measure of Hounsfield units, the degree of absorption of X-rays. The digital CT scan is represented in 12–16-bit greyscale (from 4096 to 65536 grey levels). The Hounsfield Scale standardizes CT scans, though it is not applicable to MRI scans.

2.2 The Basics of Positron Emission Tomography

Positron emission tomography (PET) is one of the lesser invasive methods used in brain medical imaging. This method of identifying cancer cells is based on the intravenous administration of a radioactive substance that breaks down B+ (a positron-emitting radionuclide used in the radioactive substance injected for imaging purposes). The preparation is accumulated in neoplastic cells, due to the fact that metabolism in infected places is faster than in healthy cells, and neoplastic networks require an increased demand for nutrients. Post accumulation, this substance decomposes, and positrons are released. The design of PET ensures the precise localization of positron

emission and its subsequent location in the patient's body. Images obtained after PET examination represent a series of equidistant sections of the patient's body, most often recorded at a resolution of 336 x 336 pixels. PET images are negative compared to CT images. The feature that PET brings to the final image is the unambiguous determination of the nature of the tissues, their precise location in the body, indication of their boundaries and the provision of information on the molecular characteristics of infected sites.

By analyzing the origin of PET images, one can see the similarity to CT, which may indicate the necessity to use the brightness window excision method. Considering that the use of a PET is aimed at determining the location of the tumor, a different strategy for image transformation is indicated. A larger window of brightness levels must be retained to reduce the amount of information lost. This strategy must be chosen because the retained brightness levels will have a decisive influence on the subsequent quality of the fusion created. The Radiant PET data visualization software was again used to adjust the brightness window. In places where the brain tumor cannot reside, a large amount of noise from the device was observed. The collected experiments allowed the determination of the lower range of the brightness window in such a manner as to filter out the generated noise. The upper brightness points were located on a level above which the information is not essential and will be condensed in order to clearly define the tumor's location. Overall, the strategy was adopted to visualize the tumor's contour, and the interior of this contour was mapped to the maximum brightness value to facilitate tumor localization. This approach allows for simple detection of the location of the tumor in the fusion image without losing vital information regarding the extent of the tumor. The final window was set to [500-10000] which was saved at 255 levels of the grayscale. The formula describing the applied transformation is as follows in Eq. 2.2.

$$\frac{500}{x_{max}} \cdot 255 \leq \frac{x}{x_{max}} \cdot 255 \leq \frac{10000}{x_{max}} \cdot 255 \quad (2.2)$$

where x is the brightness of a single pixel.

2.3 The Foundations of Magnetic Resonance Imaging

Magnetic resonance imaging (MRI) works by inducing a resonance phenomenon in a strong magnetic field. The test object is placed inside an electromagnet, the power of which can oscillate within 3 Tesla. Additionally, an electromagnetic radio frequency wave is generated at a point on the patient's body. The tissue in the patient's body that absorbs the radio wave beam begins to resonate, which causes a change in the magnetization parameters in this area. After the end of the radio wave transmission, the excited area returns to its initial state, emitting energy in the form of a wave with the excitation frequency. Magnetic resonance has been designed in such a manner that it is possible to observe all magnetic field anomalies and register all waves emitted from the patient's body. In this way, the relaxation time is recorded for individual body tissues. The difference in these times allows for the determination of the type and density of the studied area. During an examination, the indicated area of the human body is scanned, and a two-dimensional image is reconstructed based on the collected relaxation times.

The resolution of the received images depends on the diagnostic equipment used and the parameters used by the operator, with the most common being a 392×448 pixel table. A feature that distinguishes MRI from other devices for visualizing the human body is the high level of detail, therefore MRI often supplements the information from computed tomography scans. MRI offers a more accurate representation of the soft tissues of the human brain than CT. As much data as possible from this image should be preserved since the tumor is the most common tissue among the tissues. In this case, the decision was made to scale the entire brightness range offered by the device to 255 brightness levels. The formula describing the applied transformation is as follows in Eq. 2.3.

$$y = \frac{x}{x_{max}} \cdot 255 \quad (2.3)$$

where x is the brightness of a single pixel.

Such a number of brightness levels will allow for the maintenance of the features of the MRI image to a satisfactory degree. The narrow brightness scale will allow the differences to be noticed in the fusion images and determine the usefulness of the created fusion. Another issue worth mentioning is the fact that a large number of brightness levels are lost, which may later translate into errors in the operation of deep neural networks in which the fusion is to be used. The application will have the option to modify each of the parameters presented in this chapter, which allows the operator to designate the optimal settings.

2.4 Microdosimetry in the Treatment of Brain Tumors

Glioblastoma multiforme can be characterized by infiltrative growth, rapid tumor spread, and intense migration within the surrounding nerve tissue. The tumor infiltrates along nerve fibers, blood vessels, and soft meninges before surrounding nerve cells, impeding complete surgical resection and resulting in tumor recurrences. Nonetheless, gliomas generally lack metastatic capacity. Fig. 2.1 presents the MRI scan for a 15-year-old patient with glioblastoma multiform detected.

The treatment of glioblastoma primarily targets (a) tumor tissues located within the volume of the tumor and (b) tumor tissue remaining within the surgical cavity or around its periphery.

Therapeutic methods assume:

1. The use of isotopes that emit radiation, particularly alpha particles, is an effective strategy for selectively destroying cancerous tissues while minimizing damage to healthy tissues. Alpha particles fulfill the necessary conditions for this approach.
2. The use of a radiopharmaceutical (rad-farm) that incorporates a ligand with rapid and selective accumulation in tumor tissues, including both the tumor site and surgical cavity, and has slow metabolism, represents a promising approach.
3. The feasibility of repeat therapy.
4. The ability to confirm the accuracy of radiopharmaceutical administration, for instance by using PET scans, is crucial for optimizing treatment outcomes.

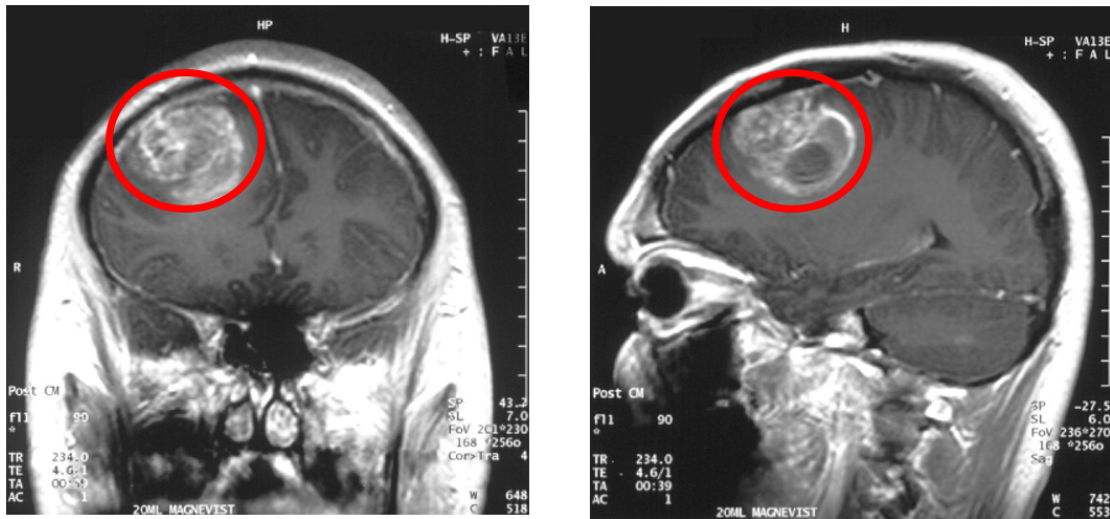


FIGURE 2.1: MRI of the head with contrast illustrating a glioblastoma multiform in a 15-year-old patient. Figure provided by the collaborative radiologist, Czwarnowski, P.

Direct methods of radiopharmaceutical administration include:

- Vein injection, which can result in high levels of internal radiation in the patient.
- Direct injection of the radiopharmaceutical into the tumor volume or the surgical site, which can significantly reduce internal radiation exposure to the patient, limited to the treated volume. However, this method necessitates surgical intervention.

The intracerebral administration method described above necessitates a flexible container-catheter set. The container is designed to be flexible, and the catheter is surgically inserted into the tumor volume or postoperative cavity post tumor resection. The container remains outside the skull, positioned just beneath the skin surface. Fig. 2.2 depicts the intracerebral administration set.

During the treatment phase of targeted alpha radionuclide therapy, a solution with a high concentration of rad-farm comes into contact with a tissue microstructure that has no rad-farm concentration, initiating a diffusion-like process that mainly utilizes microcirculation in the tumor tissues. This process transports the alpha-particle emitting isotope and results in the destruction of tumor tissues [21]. The efficacy of the therapy can be significantly influenced by the controlled compression of the radium pharm set container because the force and timing of the pressure determine the contact surface's shape between the solution containing the rad-farm and the surrounding tissues. This contact surface's shape changes the rate at which rad-farm is introduced into the microcirculation in tumor tissues, which is the same as the isotope's rate of entry and subsequent transport. It is crucial to keep in mind that the therapeutic effect depends on the isotope, while the rad-farm properties determine the isotope's transport in radioisotope therapy. Figure 2.3 presents an example of reflux resulting from a clogged catheter or too rapid administration of a radiopharmaceutical.

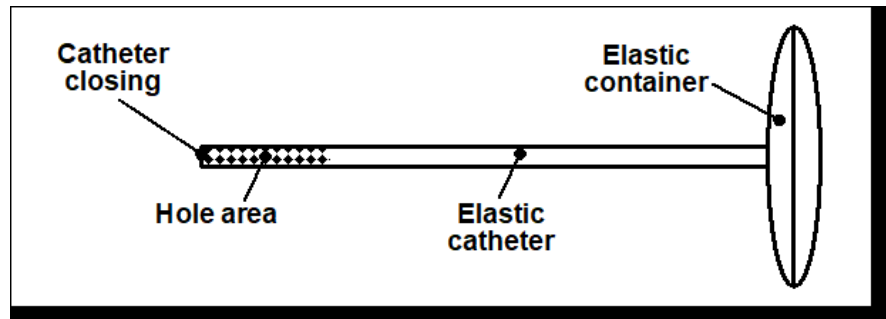


FIGURE 2.2: The intracerebral administration method described above necessitates a flexible container-catheter set. The container is designed to be flexible and the catheter is surgically inserted into the tumor volume or postoperative cavity after tumor resection (Fig. 2.4). The container remains outside the skull, positioned just beneath the skin surface. Created by the author based on the initial draft provided by the collaborative radiologist, Czwarnowski, P.



FIGURE 2.3: An example of reflux resulting from a clogged catheter or too rapid administration of a radiopharmaceutical. Created by the author with the usage of RadiAnt [2] software.

The rate of introduction of the radiopharmaceutical into the microcirculation of the tumor tissues changes as a result of the treatment and is dependent on the rate of administration and subsequent transport of the isotope. One should be constantly aware that the therapeutic effects of isotopic radiotherapy are dependent upon the isotope's properties, while the transport of the isotope is established by the attributes of the Radiopharmaceutical.

In that method, both alpha and beta radiation are used as a result of subsequent isotope transformations, with ^{213}Bi being the primary isotope of interest. This isotope is part of the uranium (^{233}U) radioactive chain and its physical properties, including a half-life $T_{1/2}$ of 46 minutes and emitted radiation, are considered relevant for therapy.

The reported activity of ^{213}Bi -DOTA-substance P is approximately 2 GBq combined with about 10 MBq of ^{68}Ga . During the initial phase, the radiopharmaceutical fills the entire volume of the container-catheter set, emitting alpha particles that are mostly absorbed by the walls of the set. The amount of radiopharmaceutical penetrating through the small holes in the walls of the catheter and reaching the tumor or postoperative bed tissues is minimal.

Through the controlled compression of the flexible container, the radiopharmaceutical fills the space between the catheter and the tissues of the tumor or postoperative bed, significantly increasing the contact area between the radiopharmaceutical and tissues. This process results in the diffusion of the isotope, emitting alpha particles, into the tumor, and ultimately leading to the destruction of the diseased tissues.

The course of the procedure of controlled compression of the container-catheter set with radiopharmaceutical can significantly impact the effectiveness of therapy, mainly due to the pressure force and its timing, which determine the shape of the contact surface of the solution containing the radiopharmaceutical with the surrounding tissues.

Isotope radiotherapy is applied in both the case of the tumor and the postoperative tumor bed, with the tissues filling these volumes constituting an isotropic environment. The metabolism in tumor tissues is relatively slow, compared to the half-life, therefore it can be assumed that the therapy's effectiveness depends on the radiation's efficacy on these tissues and not on the metabolism.

In the context of this dissertation, it is important to note that certain details, such as determining the energy deposited in Bismuth-213 tissue, calculating the absorbed dose in a given volume [22], and assessing the tumor control probability (TCP), are not the primary focuses of the research and therefore are not extensively discussed. While these background details are important to consider, the objective of this thesis is to establish a method for calculating the volume of radiopharmaceutical to be injected and to evaluate tumor progression based on this volume, providing a key metric for radiologists.

2.5 Novel Treatment Strategy

At the Nuclear Medicine Department of the Medical University of Warsaw, located at the Central Clinical Hospital, clinical trials are being conducted to investigate state-of-the-art treatment strategies for patients diagnosed with glioblastoma. The standard medical procedure following diagnosis is to remove the tumor and initiate a radio and

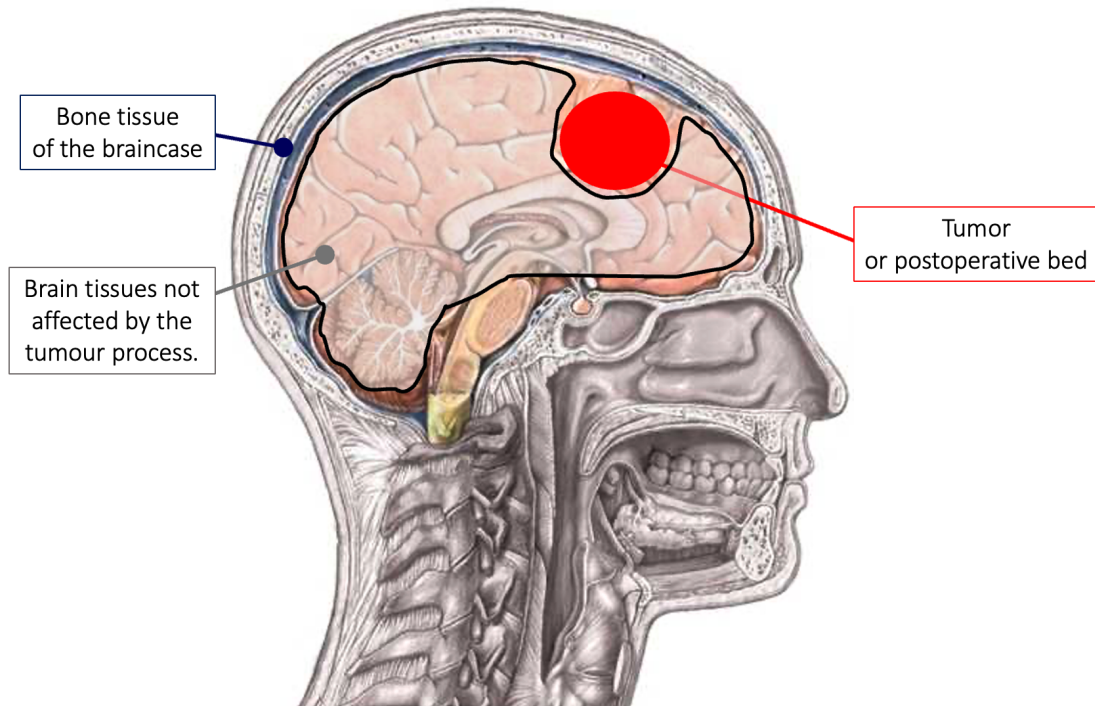


FIGURE 2.4: Treatment environment for glioma tumors. Created by the author based on the initial materials provided by the collaborative radiologist.

chemotherapy treatment plan. Radiologists use CT and PET scans to determine the tumor localization, size, and stage of advancement, as well as the amount of radiopharmaceutical injected into the tumor during invasive examinations or postoperatively after resection. These substances are intended to damage tumor cells along with normal, noncancerous brain cells. Radiologists and oncologists evaluate the volume of radiopharmaceutical administered and the probability of local cure TCP (Tumor Control Probability) based on microdosimetry methods, which involve dosimetric measurements of radiation exposure. A volume of radiopharmaceutical which is too high has the potential to damage healthy brain cells, while a volume that is too low will not effectively eradicate cancer cells [6].

At present, the measurement of glioma volume is typically performed manually by radiologists, which is a time-consuming process and prone to human error. The analysis of PET-CT scan results requires the expertise of certified nuclear medicine physicians who have years of experience in PET imaging [23]. To address this challenge, software tools have been developed to assist radiologists and nuclear medicine physicians in their measurements. Fig. 2.5 shows an example of software that is commonly used in this context.

The challenge that remains to be addressed is the accurate determination of the percentage of the entire tumor tissue volume that has been reached by the radiopharmaceutical. This value is closely related to the tumor volume and thus is critical to the success of the modern approach. To this end, computer-based methods have been developed that automatically detect tumors, define their localization, visualize their



FIGURE 2.5: PETSyngo 6.7.3, VB42A from Siemens, the resulting image obtained by an Subject Matter Expert (SME). Figure provided by the collaborative radiologist, Czwarnowski, P.

shape, and calculate their volume. These methods are designed to operate on various image modalities, such as CT, PET, and MR scans, and require the fusion of data from different sources.

All brain scans used in the development of the algorithms presented in the research were obtained from the Department of Nuclear Medicine at the Medical University of Warsaw. A Siemens Biograph 64 PET-CT scanner was utilized between 2012 and 2019 to acquire the images.

Chapter 3

Literature Review

3.1 Computer-Aided Detection Applications

Due to the significant effectiveness of computer-aided systems in detection and classification of diseases, CAD has become a central pillar of medical research in medical image analysis. Implementing computer-based techniques has a substantial impact on improving various stages of radiology workflow, such as order scheduling and triage, clinical decision support systems, identification and interpretation of findings, post-processing and dose estimation, examination quality control, and radiology reporting [24].

CAD systems by design are able to help medical professionals make more accurate diagnoses in addition to reducing the amount of time and effort required to make a discovery. Specific CAD systems have been used to detect and diagnose a wide range of diseases: cancer, cardiovascular disease, neurological disorders, and infectious diseases. The exponential growth of CAD use is observed in the detection of prostate cancer, which alone constitutes 26% of all diagnoses [25].

Addressing the greater aspect, Obermeyer and Emanuel (2016) [26] highlight the transformative potential of machine learning (ML) algorithms in medicine, specifically in improving prognosis, radiologist and pathologist displacement, and improving diagnostic accuracy. They emphasize that ML excels in handling large numbers of predictors and complex data thanks to advances in computing power and techniques such as clustering and non-linear modeling, as well as handling large data types. However, they also caution against challenges such as overfitting predictions, biases in data collection, and causal inference issues, which need to be addressed through rigorous validation, careful data curation, and responsible model building. Their perspective fully envisions the role of CADs: “Increasingly, the ability to transform data into knowledge will disrupt at least three areas of medicine: machine learning will dramatically improve prognosis, displace much of the work of radiologists and anatomical pathologists, and improve diagnostic accuracy” [26].

In 2022 Strzelecki and Badura published an article titled “Machine Learning for Biomedical Application” [27] which examined 10 papers that focus on machine learning methods for biomedical applications. The authors highlighted significant advancements in technology that have impacted the healthcare industry and medical providers. They emphasized that medical imaging, including biomedical signal acquisition, has become increasingly crucial not only in diagnostics but also in therapeutic solutions, monitoring its effects and rehabilitation progress markers. The authors noted that algorithms are trained to discover patterns and correlations in large datasets in order to

make optimal decisions and predictions based on the results of such analyses. Furthermore, Strzelecki and Badura contend that machine learning systems become more efficient over time, with accuracy increasing with access to greater data. The authors revealed their predictive models based on classification and regression models to automated multiscale feature extraction from 2D, 3D, and higher-dimensional data structures, contextual analysis, and segmentation, representing the future of the software branch of biomedical engineering.

Esteva et al. [28] demonstrated that deep neural networks, specifically CNNs, have the ability to achieve a dermatologist-level classification of skin cancer. The 129,450 clinical images from the authors included more than 2,200 different skin diseases. Their primary objective was to classify the most common and deadly skin cancer. In order to test the performance of the deep neural network, the authors examined the performance of the deep neural network against twenty-one Board-Certified Dermatologists. The results showed that the CNN performance was on par with or better than that of the dermatologists. The study confirmed the potential of Artificial Intelligence (AI) and deep learning in medical diagnosis. It opened the door for further research and development of artificial intelligence-based diagnostic tools in healthcare, with the potential to improve the accuracy and efficiency of disease identification, help clinicians make decisions, and ultimately improve patient care.

Gulshan et al. [29] developed and validated a deep learning algorithm to detect diabetic retinopathy, a leading cause of blindness, in retinal fundus photographs. The authors used the power of convolutional neural networks to create an automated system capable of analyzing retinal images and identifying diabetic retinopathy with high precision. The research team used a dataset of 12,175 retinal images graded by ophthalmologists to diagnose the presence of diabetes retinopathy. The data set was divided into a training set and a validation set. The authors used a deep CNN architecture that was fine-tuned and optimized for this specific task. The depth and arrangement of the network is critical to accurately capture complex patterns and features in retinal images indicative of diabetic retinopathy. After training, the algorithm was validated on two separate datasets: one from EyePACS (a web-based platform for retinal image grading) and another from three hospitals in India. The results demonstrated that the deep learning algorithm achieved high sensitivity and specificity in detecting diabetic retinopathy and diabetic macular edema. The algorithm showed robust performance compared to manual retinal image grading by ophthalmologists, with some metrics exceeding human experts.

In the study from 2017 by Rajpurkar et al. [30], the authors introduced CheXNet, a deep learning model that can diagnose pneumonia and 13 other thoracic diseases by analyzing chest radiographs. The researchers used the 121-layer DenseNet architecture, a convolutional neural network known for its effectiveness in image classification tasks. The authors trained CheXNet with a dataset of 102,120 frontal chest X-ray images from the National Institutes of Health (NIH) data set labeled with up to 14 chest diseases in a data set of chest X-ray images. The authors used models that use weaker supervised learning methods to address the challenges of limited annotations and data augmentation strategies to expand the model's generalization capacity. To evaluate CheXNet performance, the authors used the F1 score, which combines precision and recall, as the primary evaluation metric. They compared CheXNet's results with those

of four practicing radiologists using a 420-image test set. CheXNet demonstrated superior performance in detecting pneumonia by achieving F1 scores of 0.435, and radiologists' F1 rating ranged from 0.387 to 0.427.

Computer-aided detection systems are more closely related to how to the provision of a solution to connecting providers with computer technologies. From the perspective of the computer science discipline, CAD systems represent the encapsulation of the ultimate computational solutions that are applicable across a multitude of scenarios or use cases. In this research, the specific objects of the investigation are the algorithms, methods, and mechanisms that make up the core part of CAD, making the CAD systems composed of modular pipelines of algorithms.

3.2 Multi-modal Image Processing and Segmentation Methods

Modern Active Contour Approach

Active contour methods have been known since the early 1990s. Although improved methods are being used by researchers, many researchers still use the standard active contour method. Chen et al. [31] developed a CNN model based on conventional Dense U-Net which uses the loss function that combines geometric information: contour length and region information. The authors evaluated 2,000 cardiac magnetic resonance images, confirming that their approach produced a higher Dice Score than the traditional approach with AC as a loss function.

Marcos et al. [32] combined the Active Contour Model (ACM) with CNN in the form of feedback loop, where CNN is used to predict the parameterized energy functions used as input for ACM.

Mo et al. [33] introduced the Dense U-Net deep active contour network based on the Chan-Vese model for image segmentation for precise boundary detection. Results validated on two public datasets in total 2917 slices 256x256. The proposed methods gave significantly better results for U-Net and Dense U-Net for the Harlev datasets, but for the ISIC dataset, the methods results were comparable to Dense U-Net and FocusNetAlpha.

CNN-based Approach

Contemporary scientists have extensively covered the topic of segmentation of medical images. Dong et al. [34] presented the use of U-Net [3,35] in fully automated segmentation of brain tumors from MRI images. MRI processing was also performed in the paper written by Kaldera et al. [36] based on the faster R-CNN model to classify and segment tumors.

A typical prerequisite for therapeutic planning is the position of the voxel along with the identification and parameters of the areas with respect to diseased conditions. The deep learning method, ROI-based CNN, YOLO, was used successfully by Al-masni et al. [37] to automatically detect changes in mammography images. The obtained results concluded with a 96.33% accuracy of the model which carried out automatic detection and classification. A model capable of processing a single image in 43 ms was built by Yin et al. [38]. Based on the possibilities presented in that article and

maintaining exactness for statistical reasons it can take up to 10 seconds for a complete analysis of a study for a single patient. Detection and location of tumors in the lungs using CT images can also be done on the YOLO network as Sindhu et al. presented “Using YOLO based deep learning network for real time detection and localization of lung nodules from low dose CT scans” [39].

A description of the use of the AlexNet model and faster R-CNN using transfer learning with the detection of early-stage brain tumors was published by Ezhilarasi and Varalakshmi [40] and by Ren et al. [41]. Liu et al. [42] presented the completed new XMasNet architecture for the classification of senile neoplasms based on MRI. Additionally, using CWT, DWT and SVM there was an implementation of a brain tumor detection and classification system done by Gurbină et al. [43].

Applications of U-Net to Prostate cancer were proposed by Chen et al. [44]. The clinical implementation of MRI-only radiotherapy requires a method to derive a synthetic CT image (SCT) for dose calculation. This study examined the feasibility of building a deep convolutional neural network to produce S-IMRTs based on MRI and evaluated the dosimetric accuracy to plan prostate IMRT. Combined MR images and T2-weighted MR were acquired in each of 51 patients with prostate cancer. Fifteen pairs were randomly selected for testing and the remaining 36 pairs for training. The U-Net represents a nonlinear function with an input MR slice and an output corresponding S-CT slice. The mean absolute error (MAE) of the Hounsfield unit between the true CT and S-CT images was used to evaluate the precision of the HU estimation. IMRT plans with a dose of 79.2 Gy prescribed for PTV were applied using true CT images. The true CT images were then replaced by the S-CT images, and the dose matrices were recalculated on the same plan and compared with the one obtained from the true CT using gamma index analysis and absolute point dose discrepancy. The authors concluded that U-net can generate SCT images from conventional IMRT images in seconds with high dosimetry accuracy for a prostate cancer plan.

Huang et al. proposed an application of U-Net with MRI fusion to prostate cancer [45]. By learning a neural network, weight distributions are generated based on the relationship between image feature information and multifocal training targets. The prostate cancer MRI image pair (axial T2-weighted and ADC map) is fused using a strategy based on local similarity and Gaussian pyramid transformation. The experimental results show that the fusion method can improve the appearance of prostate cancer in terms of both visual quality and objective evaluation. The dataset included a total of 17 patients with prostate cancer confirmed by biopsy. Of these 17 patients, 12 had prostate cancer in the peripheral zone (PZ), 3 had prostate cancer in the central gland (CG), and 2 had infiltrating prostate cancer in both the PZ and CG areas. An experienced radiologist segmented the prostate and prostate cancer areas in the image. There are some limitations to this study. This study is based on a set of single-site magnetic resonance images. The method requires further validation for multi-center image trials with various anatomical regions, image resolution, or image quality with a more robust and generalized deep learning network through the adjustment of multi-center images with different anatomical domains, image resolution, or image quality. Another limitation is that the study focuses mainly on imaging fusion techniques for the ADC map and the T2W sequence. Additional mpMRI images have been developed to detect tumors, such as DCE MRI, magnetic resonance elastography, and chemical

exchange saturation transfer MRI. Future studies may investigate how the proposed fusion method can be applied effectively to mpMRI images in addition to T2W and ADC images.

As of June 2023, the IEEE Xplore database alone has published 303 articles about U-Net, 11 about Mask R-CNN, and 297 about YOLO. Despite the passage of years, one article that stands out was presented at the 2nd International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA) by Chen P. and Xu L. [46]. The U-Net model, which has low computing performance in the ARM microprocessor, can have its hardware computing significantly improved with the help of a field programmable gate arrays (FPGA), according to the cited research. The authors utilized the U-Net network in FPGA mode and employed circular flow and circular unfolding to further accelerate its hardware computing. They demonstrated the potential for offloading computation from the CPU and GPU to the FPGA, although their performance evaluation remains somewhat ambiguous. In addition, the authors improved model performance by modifying the networks, pre-processing data, and using cross-validation. This paper is worth mentioning as it points to the next direction in computational hardware optimization, especially considering the model's inference mode and the requirement to return predictions in near real-time.

Fusion

Sedghi et al. [47] proposed another application of fusion U-Net to prostate cancer. Detection of clinically significant prostate cancer (PCa) has been shown to greatly benefit from ultrasound fusion MRI biopsy, which involves overlaying pre-biopsy MRI volumes (or targets) with real-time ultrasound images. The article explores an information fusion approach between MRI and ultrasound to improve PCa targeting in biopsies. The authors built fully convolutional network models (FCN) using data from a newly proposed ultrasound modality, temporal enhanced ultrasound (TeUS), and apparent diffusion coefficient (ADC) from 107 patients with 145 biopsy cores. The model architecture is based on U-Net and U-Net with attention gates. Models are created by joint training by fusion of intermediate and late data through data fusion. The authors also built models containing data from each modality separately for use as an initial model. Using their proposed deep learning framework and inter-fusion, the integration of TeUS and ADC outperforms individual diagnostic models using the TeADC and ADC framework. The authors achieved an AUC of 0.76 for the detection of all PCa foci and 0.89 for PCa with larger foci. The results show that shared representations between multiple modalities outperform average uni-modal predictions.

Cheng et al. in 2021 made a study on the diagnosis of bone metastases in patients with prostate cancer in Taiwan [48]. The research aimed to investigate the efficient means of early diagnosis of bone metastases by bone scintigraphy images using negative mining, pretraining, convolutional neural networks, and deep learning. The author studied 205 prostate cancer patients and 371 breast cancer patients and used bone scintigraphy data from breast cancer patients to pretrain a YOLOv4 with a false positive reduction strategy. The pre-trained model is a model for detecting and identifying metastatic sites using bone scintigraphy with transferred learning applied to

prostate cancer patients. A ten-fold cross-validation was performed. The mean sensitivity and precision rates for detection and classification (based on lesions) in the chests of prostate patients were 0.72 ± 0.04 and 0.90 ± 0.04 , respectively. The mean sensitivity and specificity rates for the classification of bone metastases (based on patients) in the chests of prostate patients were 0.94 ± 0.09 and 0.92 ± 0.09 , respectively. The authors concluded that the developed system has the potential to provide pre-diagnostic reports which aids physicians in making definitive decisions.

Masoudi et al. conducted another study of bone metastases in patients with prostate cancer published by American Society of Clinical Oncology [49]. CT scans from 56 patients with histopathologically proven prostate cancer were included. An experienced radiologist annotated the extent of each bone lesion ($N = 4217$) and characterized all areas of the bone as benign or malignant. Prior to training, all scans were anonymized and normalized at the patient-level before training. The proposed method can be summarized as a two-stage framework. The first stage is the detection algorithm inspired by Yolo-v3 with a darknet-53 backbone pretraining on the Coco dataset and four final scaling blocks to compensate for a wide range of lesion diameters. The second stage is a ResNet-50-based binary classifier that is also pre-trained using the ImageNet dataset. The authors used a train / validation split equal to 90%/10% for this study. To facilitate the learning process, horizontal flipping, relative zooming, and mean weighted averaging were used for data augmentation in stage one. Instead, the classification algorithm took advantage of patches synthesized generated by the deep-convolutional generative adversary network (DC-GAN) for augmentation. The overall performance of the detection algorithm was 81% sensitivity with 86% positive predictive value. In stage 2, authors obtained an accuracy of 89% for correct classification of benign bone lesions from malignant without augmentation, which was improved to 91% when the authors incorporated the augmented data for training. The authors concluded that the proposed 2-stage algorithm sequentially detects and classifies bone lesions on CT of patients with prostate cancer with significant performance.

Applications of Mask R-CNN to prostate cancer and specifically, segmentation of intraprostatic lesions (IL) on bp-MRI using entire prostatectomy specimen-based delineations were developed by Dai et al. and published in the International Journal of Radiation Oncology, Biology, Physics [50]. Suspicious lesions were defined for 158 patients on MRI according to bp-MRI, ILs were defined for 64 patients on MRI as referenced by whole-point prostatectomy sample sections, with the remaining 40 patients being unlabeled. In order to improve segmentation accuracy, a nonlocal mask R-CNN was proposed. A model trained using MRI-based delineations was fine-tuned with prostatectomy-based delineations to investigate transfer learning. A non-local mask R-CNN with fine-tuning and self-training greatly improved all evaluation metrics with a refinement of the R-CNN mask by a prostatectomy-based delineation. Eighty point five percent (80.5%, 33/41) of lesions for the model with the highest detection rate and DSC were detected in all Gleason Grade Groups (GGG) with DSC of 0.548 ± 0.165 , 95 HD of 5.72 ± 3.17 and TPR of 0.613 ± 0.193 . Among them, 94.7% (18/19) of lesions with GGG > 2 were detected with DSC of 0.604 ± 0.135 , 95 HD of 6.26 ± 3.44 and TPR of 0.580 ± 0.190 . The conclusion of the authors was that a high detection and accuracy of segmentation of prostate cancer can be achieved in bp-MRI based on histological image annotations.

Liang et al. [51] discuss the importance of fusion, making the statement "The fused

image contains more information and precision than any source images". Multimodal medical image fusion can combine information from source images of the same organ to generate a composite image with better visualization [52]. Juanying Xie and Ying Peng introduced the 3D scSE nnU-Net model for Head and Neck (H&N) primary tumor segmentation based on positron emission tomography and computed tomography images [53]. The developed model is based on the 3D nnU-Net and published by Isensee et al. [54] with the extension of the spatial and channel squeeze and excitation (scSE) blocks by following encoder and decoder blocks. The scSE blocks adaptively recalibrate the directions of the learned feature maps to boost meaningful features and suppress weak ones. The model was trained on 160 images and validated on 41. As a result, the proposed 3D scSE nnU-Net model got better results than the original 3D nnU-Net by 1.4% of the Dice Similarity Coefficient (DSC) and achieved 0.735 of the DSC on the test dataset. Furthermore, the inference time increased from 1.83 s to 2.64 compared to the original 3D nnU-Net.

The nnU-Net architecture is based on U-Net [3]. To get the model to operate on a volumetric dataset, the authors applied the combination of three models: a U-Net 2D and 3D [4], and a cascade of two 3D U-Net models generated a segmentation at low resolution that the second model subsequently refines. On the basis of the cross-validation results, the model automatically chooses the best prediction model.

The 3D U-Net [55] is a volumetric segmentation network that learns from sparsely annotated volumetric images. The model served as the backbone network for medical image segmentation due to its performance.

Chen S., Hu G. and Sun J. in 2020 utilized the 3D U-Net with the channel attention mechanism block extension and added the channel attention mechanism to each layer on the 3D U-Net decoding path [56]. Models have been trained and evaluated by the Brats 2018 dataset. On average, the proposed model achieved 0.901 IoU when 3D U-Net got 0.736 and 2D U-Net 0.567. Woo et al. [57] compared 2D U-Net with 3D U-Net for brain structures segmentation based on MRI - classifying brain tissue as cerebrospinal fluid (CSF), gray matter or white matter. The mean DSCs results for the 2D U-Net are comparable to 3D U-Net for gray matter and white matter. However, the mean DSC for CSF was noticeably lower (89. 6% vs. 94. 4%).

Milletari et al. [58] proposed a novel approach for MRI prostate volume segmentation named V-Net. When U-Net refers to a U-shape of the network, V-Net refers to the size of the volumetric input. At each layer, V-Net contains a residual function.

Soares et al. [59] used 3D U-Net and 3D V-Net to automatically segment pulmonary nodules in CT images of the chest. Two architectures with different configurations were compared, and as of the given dataset, the 3D V-Net significantly outperformed the 3D U-Net, achieving 0.99 IoU compared to 0.74.

An outstanding open-source project requires special mention - Project MONAI [60]. It was originated by NVIDIA and King's College London to accelerate the development and knowledge sharing in utilizing AI in healthcare. As one of many results, the Monai (Medical Open Network for AI) was formulated and published as open source and aims to standardize the process of creation and model evaluation. Based on the MONAI, a novel transformer-based model architecture intended for the objective of semantic segmentation of volumetric (3D) medical images. That model reformulates the task of 3D segmentation as a sequence-to-sequence prediction problem [61] inspired by

NLP trends. UNetR is the first 3D segmentation neural network that utilizes a transformer as an encoder. That resulted in creating the architecture independent of CNNs for feature extraction. The effectiveness of UNetR was evaluated in several volumetric segmentation tasks in CT and magnetic resonance modalities using the BTCV and MSD datasets. The model obtained a new state-of-the-art performance and outperformed the top model by at least 1.0% DSC.

Zhou et al. [62] present a broad summary of deep learning-based approaches to multimodal medical image segmentation. The main challenge in this area is data. Obtaining a large amount of medical image data where medical cases and classes are balanced can be a challenge for researchers. Without diverse and numerous slices, training can produce an over-adjusted model. Publicly available multimodal medical image datasets with labelled tumors are few. The BraTS dataset, the most commonly used dataset in papers, is described in The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS) [63]. Wang et al. used the BraTs2018 dataset with four modalities (T1, T1-c, T2, and FLAIR) to develop a fusion model that can potentially be used to detect abnormal brain regions [64]. Muzammi et al. introduced Convolutional Sparse Image Decomposition (CSID) methods that fuse CT and MR images [65]. The sixth step algorithm consists of different transformation techniques such as contrast enhancement, edge detection, cartoon and texture decomposition, enhancement of CSC-based sparse coding, fusion of sparse coefficient maps and fused image reconstruction. The authors verified their approach in a data set published by Zhu et al. [66] that also leads to a publicly available data set for brain tumors published by Harvard Medical School [67]. The deep learning approach for fusion of MRI-PET scans based on extract image features using a pre-trained network (VGG19) was presented by Guo et al. at the 15th International Symposium on Biomedical Imaging [68] and evaluated on a data set containing soft tissue sarcoma. A single machine can do MRI/PET fusion by integrating PET detectors into an MRI scanner, allowing simultaneous data acquisition, resulting in the blend of functional and morphological structures with soft tissue contrast [69]. Hardware-based fusion is the future, and in that of realization, the authors of that paper want to address CT, PET, and MRI scan fusion realized by software.

3.3 Critical Appraisal

The adoption for implementing computer and deep learning-based models for the accurate detection and segmentation of tumors is correlated with three major factors:

1. By observing data one can see the size and extent of cancer has been misclassified and consistently underestimated [70].
2. The convolutional operations used in most DL model processes within a local neighborhood of the input as well as the localization and segmentation accuracy are discounted within the CNN approach. Information loss in segmentation models is often caused by pooling and resizing operations with the lost information typically containing the opaque regions between the lesion and surrounding normal tissues which leads to inaccurate definition of boundaries.
3. Registering alternative modalities to MRIs, CTs and labeling them in a multidisciplinary collaboration is extremely time-consuming, not to mention challenging.

Due to the small volume of annotated data [70] there has been little progress on DL-based segmentation using cross-modality image domains. Obtaining diverse and representative medical datasets is one of the biggest challenges due to data privacy regulations, data ownership, and data sharing issues. Additionally, obtaining different modalities (CT, PET and MRI scans) for the same patient can have blocking factors, and as data availability includes aspects of covering different demographics and types and states of tumors, it is significantly limited. No public sources were found for multi-modal scans for the same patients. The data used in this research were obtained from two hospitals in Poland after signing official agreements and joint cooperation during granted projects. Furthermore, during development, additional data were obtained from various publicly available repositories such as the BraTS (Brain Tumor Segmentation Based on Magnetic Resonance Scanning) dataset and the MNIST-Med dataset. All used data have been annotated by subject-matter experts (radiologists), making them referential as ground truth.

Summary

Medical data vary in aspect of modalities, numbers, and sizes. Different CT scanners are used to perform tests, resulting in variances in the Hounsfield scale. The lack of enough data has a profound impact.

Active contour methods are becoming obsolete mainly due to the difficulty in implementation, single-threading, and the amount of time to process individual images, as well as the possibility of final implementation. For several years, methods based on neural networks have been leading in this area. Currently, the author has dozens of different models used in 2D and 3D image processing. It can be said that 3D processing is yet an emerging trend, with 4D processing being the future. Researchers use various techniques to train and modify networks to improve the performance of algorithms. Researchers use various metrics to evaluate these algorithms and at the end of the day perform validations against certain data. At times there is publicly available data, with there being reference data; however, there are times that data that cannot be shared. The problematic aspect in the case of the latter is there is no way to compare the effectiveness of the proprietary algorithm with the published one as the data is not shared. While in the case of the former, the data most often refer to rather generic diseases and are poor in terms of modalities. Overall, few works deal with the subject of data and network throughput in the context of medical data including topics related to potential divination or security of the data itself.

Medical image fusion is a very distinct topic compared to segmentation and classification alone. The reason is the data, not only that it is unreleased, but that it is uncommon to have an MRI, CT and PET performed. Evaluation of effectiveness in terms of qualitative and quantitative methods of measuring volume is rare.

Chapter 4

Introduction of a Medical Data Processing Framework

4.1 The Key Components of MeDAPR Framework

The author of the research developed a universal six-stage framework named MeDAPR for structuring the research process in healthcare use cases. MeDAPR stands for Medical Data Processing and the framework's stages are as follows:

- Stage 1: Medical Image Assessment;
- Stage 2: Medical Data Preprocessing;
- Stage 3: Medical Data Fusion;
- Stage 4: Medical Data Augmentation;
- Stage 5: Deep Neural Networks for Classification and Segmentation;
- Stage 6: Volumetric calculations and 3D visualizations.

Although the MeDAPR framework has been extended and adjusted to the specific research experiments, the main pillars remain unalterable.

Stage 1: Medical Image Assessment – This first stage is critical in the overall research process. It provides the foundation for subsequent stages by involving the analysis of medical images to identify the distinct modalities present in the data. Appropriate methods are then developed to detect and segment the region of interest. This is the query/response stage where questions and answers are essential to inform detection, procedure, visibility, and modality.

Stage 2: Medical Data Preprocessing – The second stage involves a series of procedures to prepare raw data for further analysis as the data may come in various forms and modalities. Cleaning and transforming the data into a more usable format is the primary objective of this stage. A variety of techniques will be deployed to achieve that objective, namely, noise reduction, normalization, and feature extraction. Additionally, data preprocessing is vital to ensure the data is consistent, clean, and fit for further analysis. Critical aspects such as the removal of unnecessary data, format and size must also be addressed.

Stage 3: Medical Data Fusion – The third stage mostly involves the integration of multiple modalities of medical imaging, such as CT, MRI, and PET, to provide a more comprehensive understanding of a patient's condition. It is critical to accurately

identify and segment tumors or other abnormalities. The stage addresses aspects such as whether all necessary data have been obtained and combined to achieve the end goal of accurate tumor segmentation, and whether additional data sources should be included for a more complete picture of the patient's condition.

Stage 4: Medical Data Augmentation – The fourth stage involves the generation of additional synthetic images based on the existing ones to increase the size of the medical image dataset. To achieve generation of the images, the original images are rotated, scaled, cropped, and flipped to create clinically relevant variations. Data augmentation is critical to providing additional examples for training and validating machine learning algorithms, which also helps to mitigate overfitting and biases in the algorithms. This stage ensures the dataset is sufficiently large and diverse enough to train algorithms that can generalize well to new data.

Stage 5: Deep Neural Networks for classification and segmentation – The fifth stage involves training neural networks to recognize patterns in the data in order to make subsequent predictions. Classification of different types of tissue or lesions in an image, or segmentation of specific regions of interest is achieved by this method. The choice of deep learning architecture, the selection of hyperparameters, and the optimization of the model to achieve the best performance are addressed. This stage also involves determining the effectiveness and accuracy of classification and segmentation by evaluating the performance of the model along with comparing it to other methods.

Stage 6: Volumetric calculations and 3D visualizations – The sixth stage is primarily focused on the generation of a three-dimensional model of an object to support various applications, such as surgical planning, diagnosis, and monitoring disease progression over time. The process involves calculating the volume of the object based on its dimensions and then generating a 3D model of the object. The end user is able to utilize this 3D model to visualize and interact with the object in a more intuitive way, allowing for a deeper understanding of its characteristics and properties. This stage ensures the end user possesses the optimal manner in which to view and analyze the computed results generated during the preceding stages.

MeDAPR framework is designed to solve several problems that researchers encounter while working with medical data. Moreover, the framework assists in the automation and streamlining of the medical data processing pipeline, which allows researchers to focus on the analysis and interpretation of results rather than on the manual processing of data. The MeDAPR framework brings clarity and structure to the research process, making replication and result comparison across different studies less complicated. Moreover, the framework provides a clear, standardized procedure for working with medical data, which is especially important in the field of healthcare, where the accuracy and reliability of results has a significant impact on patient care.

4.2 Capabilities of Convolutional Neural Networks in Medical Image Processing

Convolutional Neural Networks (CNNs) are a type of deep learning neural network used for image recognition and classification tasks. They are composed of a series of convolutional layers, which are used to extract features from an input image. Each convolutional layer consists of a set of filters, which are used to detect patterns in the

input image. The output of each convolutional layer is then passed to a pooling layer, which is used to reduce the dimensionality of the feature map. Finally, the output of the pooling layer is passed to a fully connected layer, which is used to classify the input image. CNNs are able to learn complex features from the input image and are used for a variety of tasks such as object recognition, image segmentation, and image generation.

Convolutional Neural Networks (CNNs) are a powerful class of deep learning models that have been widely used in a variety of applications, such as computer vision, natural language processing, and speech recognition. CNNs are composed of multiple layers of neurons that are connected in a hierarchical structure, allowing them to learn complex patterns from large amounts of data. This chapter discusses the fundamentals of CNNs, including their architecture, training, and applications. Further to the discussion will be an exploration of selected recent advances in CNN research highlighting the potential for further development.

CNN refers to a network architecture composed of several stacked convolutional layers [71]. The convolution identifies and extracts locally connected information – among adjacent voxels or pixels. By combining the pooling layers between the convolution layers, the network's perspective can be expanded through the presence of a large area of view. It takes part in the locally connected nodes of the input layer and results in an output having a smaller spatial footprint [71]. A series of fully connected layers integrates all activations of previous layers together to form one. Upon completion of processing, the model outputs the final set of feature values related to a particular task.

Convolutional Neural Networks (CNNs) are extensively and successfully utilized in 2D and 3D medical image processing, even with the knowledge of CNN's pooling layer tendency to discard position, and the sensitivity of the models to rotation and affine transformation.

4.2.1 Key Concepts of CNNs

The convolution operation is the core building block of CNNs, enabling the extraction of local features from the input data. For a two-dimensional input matrix I and a filter (also known as kernel) K of size $m \times n$, the convolution operation can be defined as the Eq. 4.1:

$$C(i, j) = \sum_{m=-\frac{M}{2}}^{\frac{M}{2}} \sum_{n=-\frac{N}{2}}^{\frac{N}{2}} (I(i + m, j + n) * K(m, n)) \quad (4.1)$$

where $C(i, j)$ is the output of the convolution operation at the position (i, j) .

After the convolution operation, an activation function is applied elementwise to introduce non-linearity into the network. The Rectified Linear Unit (ReLU) is a common activation function and is defined as the Eq. 4.2:

$$f(x) = \max(0, x) \quad (4.2)$$

where x is the input value and $f(x)$ is the output value after applying the activation function.

Pooling is used to reduce the spatial dimensions of the feature maps while retaining important information. One common pooling operation is max-pooling, which can be defined for a two-dimensional input matrix P and a pooling window of size $m \times n$ as the Eq. 4.3:

$$Q(i, j) = \max(P(i * m : (i + 1)m, jn : (j + 1) * n)) \quad (4.3)$$

where $Q(i, j)$ is the output of the pooling operation at the position (i, j) .

Eq. 4.4 express the form of calculating the output spatial dimensions after applying padding.

$$O = (W - K + 2P) / S + 1 \quad (4.4)$$

where:

- O is the output spatial dimension (height or width) of the feature map.
- W is the input spatial dimension (height or width) of the image.
- K is the kernel size (height or width) of the convolution filter.
- P is the amount of padding applied to the input image.
- S is the stride, which represents the step size the convolution filter moves.

In Eq. 4.4 if one wants to calculate the padding (P) needed to preserve the input dimensions (W) after applying the convolution operation, one can rearrange the formula:

$$P = ((O - 1) * S + K - W) / 2 \quad (4.5)$$

This equation assumes the use of the same padding and stride for both height and width dimensions. If one applies a different padding or alternate stride values for height and width, one should calculate them separately using the appropriate dimensions. In the padded image, the pixel value at location (i, j) is $Padded_I(i, j)$. The original input image is I . The width and height of the input image are W and H respectively. The padding width is P , with the assumption of equal padding on all sides. This equation states that if the indices (i, j) fall within the original image boundaries, considering the padding width P , then the pixel value of the padded image is taken from the input image I . Otherwise, if the indices fall outside the original image boundaries and within the padding area, the pixel value is set to 0 (in the case of zero-padding). This equation assumes a zero-padding scheme. There are other padding schemes such as reflection padding, symmetric padding, and constant-value padding. The equation for $Padded_I(i, j)$ would need to be adjusted accordingly based on the chosen padding scheme. Padding is often applied to the input before the convolution operation to control the spatial dimensions of the output feature maps or to maintain the original input dimensions. Zero-padding can be represented as Eq.4.6:

$$Padded_I(i, j) = \begin{cases} I(i - P, j - P) & \text{if } P \leq i \leq W + P - 1 \text{ and } P \leq j \leq H + P - 1 \end{cases} \quad (4.6)$$

where $Padded_I$ is the padded input matrix, I is the original input matrix, W is the width and H is the height of the input matrix.

Stride controls the step size used during the convolution operation, determining how the filter moves across the input. The output dimensions after applying stride can be calculated as the Eq. 4.7:

$$\begin{aligned} \text{Output_Height} &= \left\lfloor \frac{H - F_H + 2 * \text{Padding_Height}}{\text{Stride_Height}} + 1 \right\rfloor \\ \text{Output_Width} &= \left\lfloor \frac{W - F_W + 2 * \text{Padding_Width}}{\text{Stride_Width}} + 1 \right\rfloor \end{aligned} \quad (4.7)$$

where H and W are the input dimensions, F_H and F_W are the filter dimensions, $\text{Padding}_{\text{Height}}$ and $\text{Padding}_{\text{Width}}$ are the padding dimensions, $\text{Stride}_{\text{Height}}$ and $\text{Stride}_{\text{Width}}$ are the stride dimensions, and $\text{Output}_{\text{Height}}$ and $\text{Output}_{\text{Width}}$ are the output dimensions after applying stride.

In the final layer of a CNN used for classification tasks, the softmax activation function is often applied to produce class probabilities. The softmax function can be defined as Eq. 4.8:

$$\text{softmax}(x_i) = \frac{\exp(x_i)}{\sum_j \exp(x_j)} \quad (4.8)$$

where x_i is the input value for class i , x_j is the input value for class j , and $\text{softmax}(x_i)$ is the output probability for class i .

4.2.2 U-Net Model

The objective of the semantic segmentation is to assign an object category label for each pixel in the image. The U-Net is the extension of Fully Convolutional Network (FCN) [72] - the first end-to-end architecture proposed for semantic segmentation. Olaf Ronneberger et al. developed the deep learning architecture known as U-Net, for biomedical image segmentation, in the Computer Science Department at the University of Freiburg in 2015 [3] for medical image segmentation. This convolutional neural network (CNN) architecture is widely used for biomedical image segmentation tasks. U-Net is a fully convolutional network (FCN) that consists of an encoder and a decoder. The encoder is a CNN that extracts features from the input image. The decoder is a symmetric network which uses the features extracted by the encoder to generate a segmentation mask. The architectural design comprises a contracting pathway (which adheres to the conventional architecture of a CNN) to facilitate the acquisition of contextual information, and a symmetric expanding pathway to enable accurate localization. The proposed methodology involves the iterative implementation of two 3x3 convolutions, which are subsequently subjected to a rectified linear unit (ReLU) and a 2x2 max pooling operation with stride 2 to achieve down-scaling. At every step of down-scaling, there is a doubling of the number of feature channels. Each stage within the extensive trajectory involves an up-scaling of the feature map, succeeded by a 2x2 convolution that reduces the number of feature channels by half, a fusion with the correspondingly trimmed feature map from the contracting trajectory, and two 3x3 convolutions, each of which is succeeded by a rectified linear unit (ReLU). Cropping is requisite, owing to the depletion of border pixels during each convolution. A 1x1

convolutional operation is employed at the consummate stage to transform every feature vector consisting of 64 components into the intended quantity of categories. The resultant of the neural network is a segmentation map of the input image with high resolution.

U-Net has been used for medical image segmentation tasks such as brain tumor segmentation, liver segmentation, and lung segmentation. Additional applications in the segmentation of images include satellite and microscopy imagery. U-Net has been shown to outperform other segmentation methods, such as the Mask R-CNN. U-Net has several advantages over other segmentation methods. Firstly, it is a fully convolutional network, which means that it can be used for segmentation of images of any size. Secondly, it is a symmetric network, meaning the encoder and decoder are mirror images of one another which allows for efficient training and inference. Thirdly, U-Net is a deep learning architecture, meaning that it has the ability to learn complex features from the input image. Finally, U-Net is a fast and efficient architecture, making it suitable for real-time applications.

The network architecture is illustrated in Figure 4.1. The configuration presented is of a typical architecture of the convolutional network.

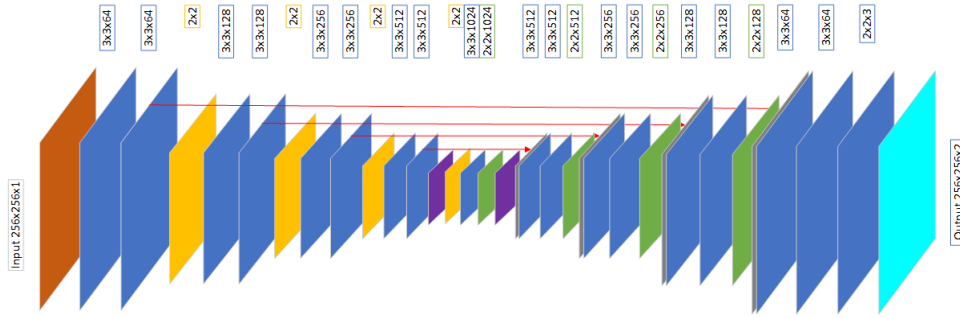


FIGURE 4.1: Implementation of the U-Net architecture. Legend starting from the left: red - transformation to the input, blue - Conv2D - convolution layer with 3x3px filter and ReLu activation function, yellow - Max-Pooling2D with 2x2px - down-samples the input size, violet - dropout - decrease the input size, green - UpSampling2D - 2x2px up-convolution layer, gray - concatenation of the output of the layer, turquoise - Conv2D layer with sigmoid activation function [3].

The U-Net is broadly applied for brain tumor segmentation on MR imaging [73,74]. Naser et al. [75] used PET-CT for automatic segmentation of head and neck cancer based on the U-Net, with resizing of the images to 144x144 pixels. The gap addressed in this dissertation regards the lack of U-Net training on upfront fused full-sized dual modalities: CT and PET.

In the U-Net architecture, max-pooling (Eq. 4.9) is used into encoder (contracting) path for two principal reasons: (a) max-pooling reduces the spatial dimensions (height and width) of the feature maps while retaining the most significant information. This reduction in dimensionality helps to manage computational complexity and memory requirements, especially when dealing with high-resolution images; (b) by down-sampling the feature maps, max-pooling helps the network to learn and capture contextual information from larger receptive fields. This is important for segmentation

tasks, as it allows the network to recognize larger structures and contextual patterns within the input image, which contributes to better segmentation performance. However, it is worth noting that max-pooling may lead to the loss of spatial information due to the down-sampling process. U-Net employs skip connections between the encoder and decoder paths, mitigation this issue. These connections transfer high resolution feature maps from the contracting path to the expanding path, enabling the network to recover the spatial information necessary for accurate segmentation at the output. Transposed convolution operation known as deconvolution or up-convolution (for up-sampling) is used into to decoder (expanding) path for up-sampling. The primary goal of the transposed convolution operation is the increase of the spatial dimensions (height and width) of the feature maps simultaneously learning to reconstruct the segmented output from the lower resolution feature maps generated by the encoder path. The transposed convolution operation works by sliding a kernel across the input feature map, performing an element-wise multiplication followed by a sum, similar to the regular convolution operation. However, unlike a regular convolution, transposed convolution involves inserting zeros (also known as stride) between the input elements before the convolution operation. This results in an expanded output feature map.

$$Q(i, j) = \max(P(i * m : (i + 1)m, jn : (j + 1) * n)) \quad (4.9)$$

In the context of U-Net, the transposed convolution operation (Eq. 4.10) allows the decoder path to up-sample the lower resolution feature maps generated by the encoder path, progressively increasing the spatial resolution. Combined with the skip connections, which transfer high-resolution feature maps from the encoder path, the transposed convolution operation enables the U-Net architecture to recover detailed spatial information necessary for accurate segmentation.

$$C_{\text{transposed}}(i, j) = \sum_{m=-\frac{M}{2}}^{\frac{M}{2}} \sum_{n=-\frac{N}{2}}^{\frac{N}{2}} (I(i - m, j - n) * K(m, n)) \quad (4.10)$$

where $C_{\text{transposed}}(i, j)$ is the output of the transposed convolution operation at position (i, j) , $I(i - m, j - n)$ is the input feature map, and $K(m, n)$ is the kernel.

4.2.3 Mask R-CNN Model

Mask R-CNN can be considered as an extension of Faster R-CNN [76]. The model is flexible and can be utilized for object classification, object detection and instance segmentation. For each object in an image, Mask R-CNN returns the class label, bounding box coordinates, and the object mask.

Except for U-Net, Mask R-CNN (a mask regional convolutional neural network) model developed by He, K. is known for the semantic and instance segmentation [77]. These two networks were compared with U-Net for segmentation of nuclei [78]. Liu et al. [79] proposed an automatic and accurate cervical nucleus segmentation method based on Mask R-CNN, which extracts the nuclei's multiscale features and the coarse segmentation bounding box of the nuclei. To reach the accurate segmentation, the authors used a local fully connected conditional random field (LFCCRF) for that task.

Based on the references, experiments were executed to verify the usability of Mask R-CNN for segmentation of doses distribution.

Mask R-CNN operates in two stages:

1. Region proposal network (RPN) –examines the initial image and generates regions of interest (RoI) or region proposals (concepts developed for Faster R-CNN [80]).
2. RoI-pooling –which down-samples the feature map using the nearest-neighbor approach [81], generates the bounding boxes, predicts the class for each object, and segments the doses distribution area inside the proposed regions.

Both stages are connected to the backbone structure, another convolutional neural network that initializes the image's feature map. The total loss is a weighted sum of the classification loss, bounding box regression loss, and mask loss, which is minimized during training.

Mask R-CNN Critical Assessment

The thesis' main objective assumes the use of different models in a pipeline, for a higher quality of the final volumetric results. The usability of U-Net is indisputable. At the same time, the fact that U-Net requires the size of the input to be a multiple of the max pool size to the power of U-Net depth, limits the potential of the network. Moreover, during the U-Net research and methods development and assessment, it resulted that its struggle with class imbalance and its struggles in differentiating distinct entities belonging to the same category, which in case of brain tumor detection, can be split into multiple smaller ones, is a serious issue. These reasons impacted the further examination of the Mask R-CNN model and attempts to identify the weak points of the network first.

To provide objective research, the author evaluated the Mask R-CNN in the case of dose distribution during radiotherapy for breast cancer [82]. Mask R-CNN's Keras implementation was adjusted to train with grayscale images by loading weights of all layers except conv1 (first convolutional layer of the network). Image channel numbers were decreased to one, and equivalent image means were an array of one element. Additionally, adjusting the implementation of loading image and masks was executed. Per each patient's results, there were equivalent eighteen classes of distributed doses. Each class was represented by a separate file with a binary mask, which in general could be a bitmap or polygon layer that identifies specific areas. The models were initialized with ResNet50 backbone [83].

Developed method's source code and trained models are available on GitHub [84]. Fig. 4.2 presents original masks for two patients. Fig.4.3 and 4.4 present objects detected by the trained Mask R-CNN model, and fig. 4.5 presents Top anchors results. Fig. 4.6 presents final detections.

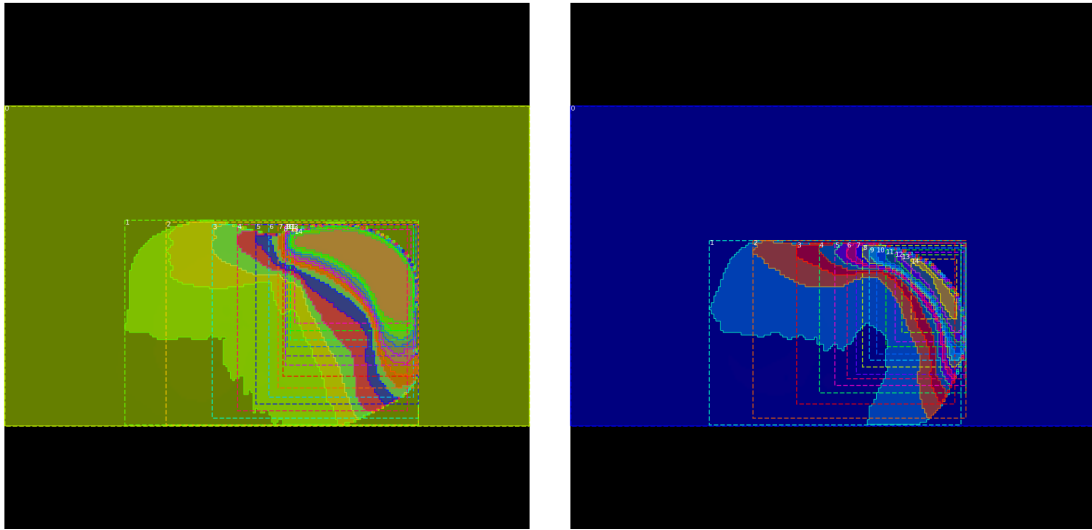


FIGURE 4.2: Original masks for two patients.

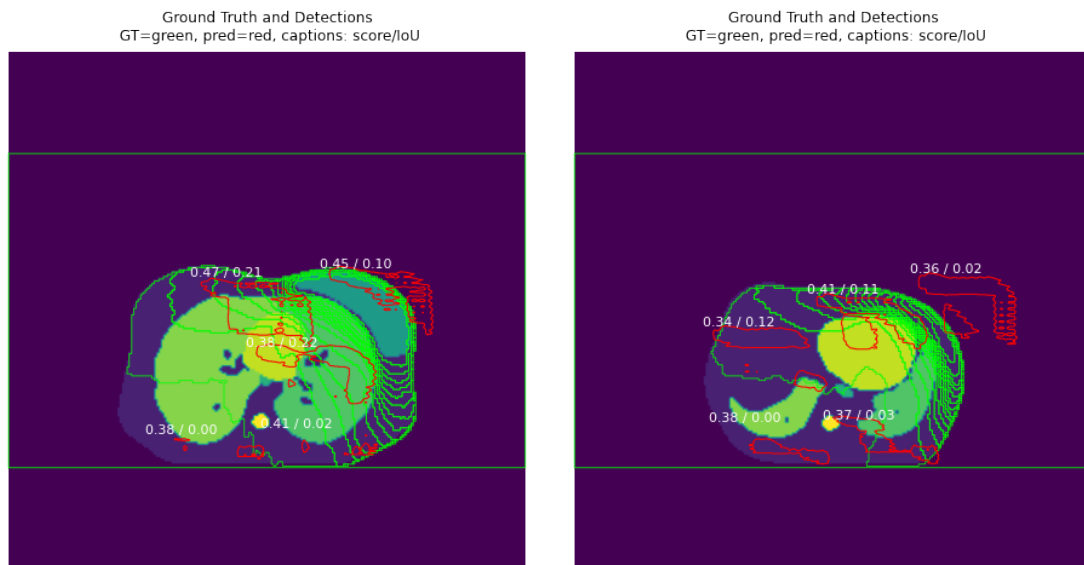


FIGURE 4.3: Detected objects (classes) with their corresponding IoU scores returned by Mask R-CNN model presented on top of the original scans.

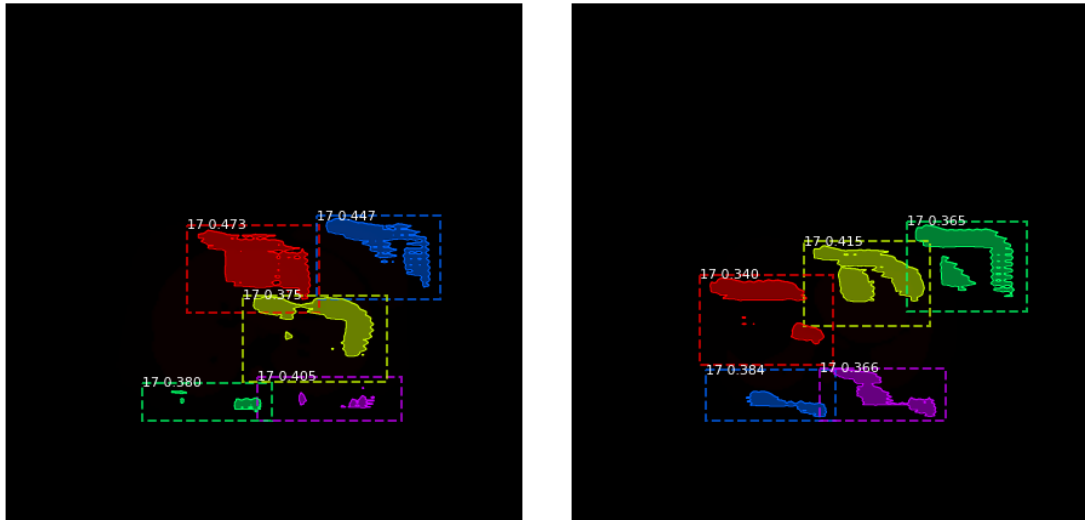


FIGURE 4.4: Segmented detected objects (classes) with their corresponding IoU scores returned by Mask R-CNN model.

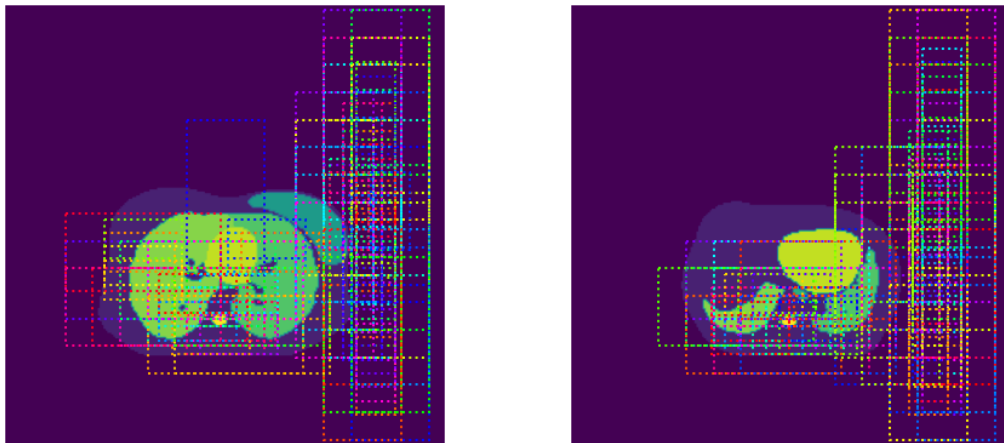


FIGURE 4.5: Two results returned by Mask R-CNN model with top anchor scores.

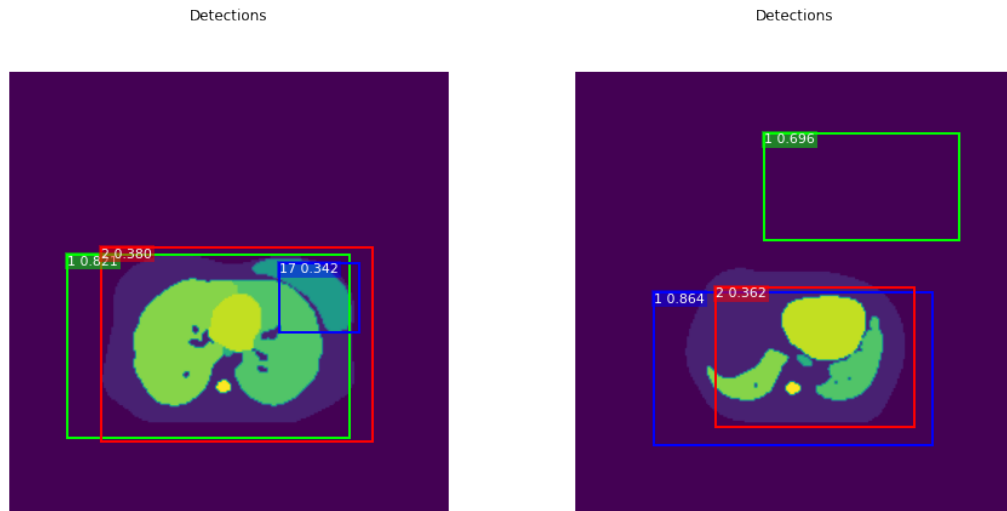


FIGURE 4.6: Detections returned by Mask R-CNN model.

The Mask R-CNN showed a high degree of capability to predict bounding boxes for objects but struggles with performing a considerable segmentation when the objects contain each other within the same or different classes. The nature of grayscale images introduces additional difficulties to train the model successfully. Moreover, shapeless doses distribution classes are adjacent to each other what together with epoch number lower than 500, could result in inaccurate prediction for this specific case. Additional Mask R-CNN implementation tuning is required to get acceptable results. Experiments with training the model in more than 1500 epochs and converting grayscale images to pseudo-RGB, where the first and third channel will contain a standard grayscale image. The second will represent an inverted grayscale image, perhaps leading to results which will include that network in subsequent research. At the same time, the accuracy of the detection and segmentation was outstanding, which is why the author proceeded with Mask R-CNN as the main model for brain tumor detection.

4.2.4 YOLOv4 Model

YOLOv4 is the latest version of the YOLO (You Only Look Once) object detection system. It is a state-of-the-art real-time object detection system that is both accurate and fast. YOLOv4 is an improved version of YOLOv3, which was released in 2018. YOLOv4 uses a variety of techniques to improve accuracy and speed. It uses a new architecture called SPP-Net, which is a combination of Spatial Pyramid Pooling and Feature Pyramid Network. This architecture allows YOLOv4 to detect objects of different sizes in an image. It also uses a new feature called Cross-Stage Partial Network (CSPNet), which helps to reduce the number of parameters and improve accuracy. YOLOv4 also uses a variety of data augmentation techniques to improve accuracy. These include mixup, mosaic, and cutmix. It also uses a new loss function called CIOU (Complete Intersection over Union) loss, which helps to improve accuracy. Moreover, YOLOv4 uses a new training strategy called Mosaic Data Augmentation, which helps to improve accuracy

and speed. This strategy uses a combination of data augmentation techniques to create a large number of training images.

Eq. 4.11 expresses the bounding box prediction.

$$\begin{aligned} b_x &= \sigma(t_x) + c_x \\ b_y &= \sigma(t_y) + c_y \\ b_w &= p_w \exp(t_w) \\ b_h &= p_h \exp(t_h) \end{aligned} \quad (4.11)$$

where b_x, b_y, b_w, b_h represent the predicted bounding box coordinates, t_x, t_y, t_w, t_h are the network output values for the box coordinates, c_x, c_y are the top-left coordinates of the grid cell p_w, p_h are the anchor box width and height, and σ is the sigmoid function.

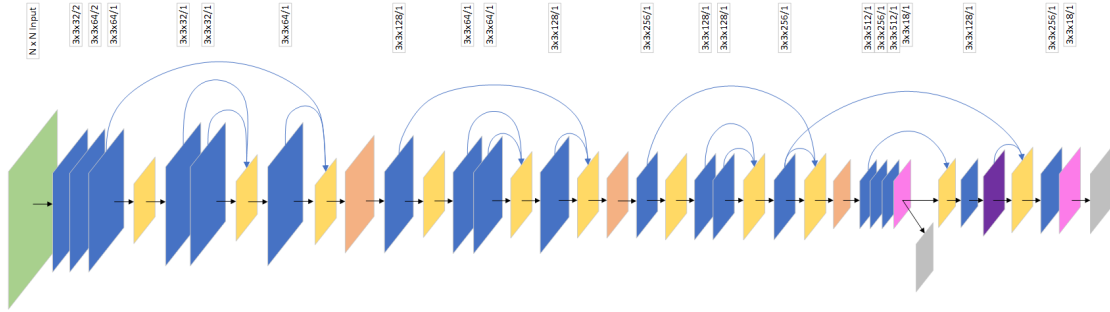


FIGURE 4.7: Implementation of YOLOv4 tiny custom architecture. Legend starting from the left: green - input images in $N \times N$ size, blue - convolutional layer with leaky-relu activation function, yellow - route layer's output concatenation, orange - max pooling that down-samples the input size, pink - convolutional layer with linear activation function, gray - YOLO object detection layer [4]

YOLOv4 (Fig. 4.7) was trained on MSCOCO-2014 which is labeled with crowds=1, therefore the crowd of people is marked as one object in the MSCOCO-dataset, as is seen in yolov4detection image. This can greatly reduce the detection accuracy of individual persons in a crowd. Therefore, YOLOv4 should be retrained with correct labels.

4.2.5 Key Differences and Strengths

U-Net, Mask R-CNN, and YOLOv4 all have available pre-trained models, making it uncomplicated to commence medical image analysis tasks. This is beneficial in cases where data is limited or when developing new models from the ground up is prohibitive. Additionally, all architectures are implemented using popular deep learning libraries such as TensorFlow and PyTorch, making integration into existing workflows and pipelines seamless.

Specific requirements for medical image processing and brain tumor detection and segmentation will determine the choice among U-Net, Mask R-CNN, and YOLOv4.

- U-Net's architecture: Featuring a contracting and expanding path with skip connections, this architecture is specifically designed for image segmentation tasks

making it a natural choice for medical image analysis, including brain tumor segmentation, where accurate segmentation is critical.

- The high accuracy of Mask R-CNN: Demonstrating a high accuracy in object detection and segmentation tasks, makes this architecture well-suited for medical imaging applications where accurate detection and localization of tumors and other anomalies is essential.
- Efficient real-time processing with YOLOv4: Known for its high accuracy and efficiency, YOLOv4 is ideal for real-time medical imaging applications such as surgical robotics or image-guided therapy. Its efficiency is achieved through a combination of advanced data augmentation techniques with a more complex backbone network.

Chapter 5

Active Contour Approach for Image Segmentation

The active contour model, initially proposed in 1988 [85], is a method for detecting object boundaries between areas of different intensities. Over the years, this method has been modified and extended to incorporate techniques such as neural networks [86], dynamic programming [87], genetic algorithms [88], wavelet analysis [89], and fuzzy logic [90]. The active contour method involves identifying pixels that could potentially belong to the object's boundary, and then grouping these points into lines that follow the shape of the object. The objective of the active contour method is to minimize the contour energy, which can be computed from various factors such as the image gradient, distance between individual points, or curvature value. One of the key features of the active contour method is its ability to trace the continuation of the edge, regardless of the physical continuity.

Active contour methods, also known as snakes, are a class of image segmentation techniques that rely on the concept of energy minimization. The energy function represents a cost function that measures how well the contour fits the boundaries of the objects of interest in the image. The contour is represented by a set of connected points that define a parametric curve, which can be deformed iteratively to minimize the energy function. The deformation of the contour is controlled by internal and external forces, which balance the attraction to the object boundaries and the smoothness of the contour, respectively. The internal forces are related to the elasticity and rigidity of the contour, while the external forces are derived from the image gradient, or other features, and act as a driving force that pushes the contour towards the object boundaries. The optimization of the energy function is typically performed using gradient descent or level-set methods, which allow the contour to evolve continuously in time until a stable state is reached.

This research is centered on the active contour techniques of Chan-Vese and Edge, which are noted for their unique and favorable attributes in handling diverse image categories. The Chan-Vese technique was selected due to its exceptional performance in region-based image segmentation, which is particularly advantageous in situations where boundaries are indistinct or blurred, as is frequently the case in medical imaging. On the other hand, the Edge technique was chosen based on its adeptness in identifying acute image transitions or discontinuities, rendering it exceptionally efficient for images that possess prominent object boundaries.

5.1 Bone Tissue Removal and Active Contour

In medical image processing, removing bone tissue from medical images is an important preprocessing step. This is owing to the fact that bones are non-cancerous tissues therefore the presence of bones can interfere with the diagnosis and detection of tumors in medical images. The main reason behind this is the high X-ray absorption of bones, which causes a high signal intensity in medical images, making the bone structure dominant in the image. As a result, the tumor tissue may be obscured by the bone tissue and can be easily missed during the diagnostic process. Therefore, removing bone tissue can improve the accuracy and reliability of tumor detection in medical images. In addition, the removal of bones can also help in reducing the computational complexity of image processing algorithms and can significantly speed up the processing time, as the bone structure is complex and requires more processing power to handle.

Algorithm

The block diagram (Fig. 5.1) presents the tasks performed in the designed algorithm, assuming that the radiologist loads a folder with pre-prepared CT and PET images. Subsequently, patient data is loaded, and the global threshold necessary for binarization is automatically determined. The bone tissue extraction described prepares the CT image for the image scaling and fusion stage. The extraction of the skull regions from the PET image is based on the processed CT image and local thresholding followed by the activation of the active contour method on the PET image. Additionally, the area of the identified object, which must be larger than 10 pixels, is verified to be classified as a tumor or tumor part. Furthermore, tumor calculations and display are performed.

Two thresholds can be distinguished: the global threshold, whose value is calculated based on the entire image, and in the case of the algorithm for calculating the volume of glioma, based on a series of 148 brain images obtained using positron emission tomography; and the local threshold, whose value is calculated based on a single image. The threshold can be chosen manually based on the histogram, or automatically.

Bone Tissue Removal based on Binarization

A substance injected into the tumor site or into the postoperative cavity may pass into subcutaneous structures near the skull, which can mistakenly indicate the presence of a tumor [91]. One of the objectives of the designed algorithm is to extract bone tissue, with particular emphasis on the skull [92]. The first step in bone tissue extraction is to set a threshold value, determined as 1190 (using a histogram), due to the fact that the skull is brighter than the brain and air. Subsequently, from the black and white image, the five largest objects are selected and removed from the two-layer image. Fig. 5.2 presents the results of the algorithm which removed bone tissue from CT scan.

The result of binarization (Fig. 5.3) plays a crucial role in determining the accuracy of the final processed image as it directly ascertains the presence or absence of a tumor in a given area. Thresholding, which is a key component of the binarization process, solely takes into account the pixel intensity values 36. Thus, the same algorithm with

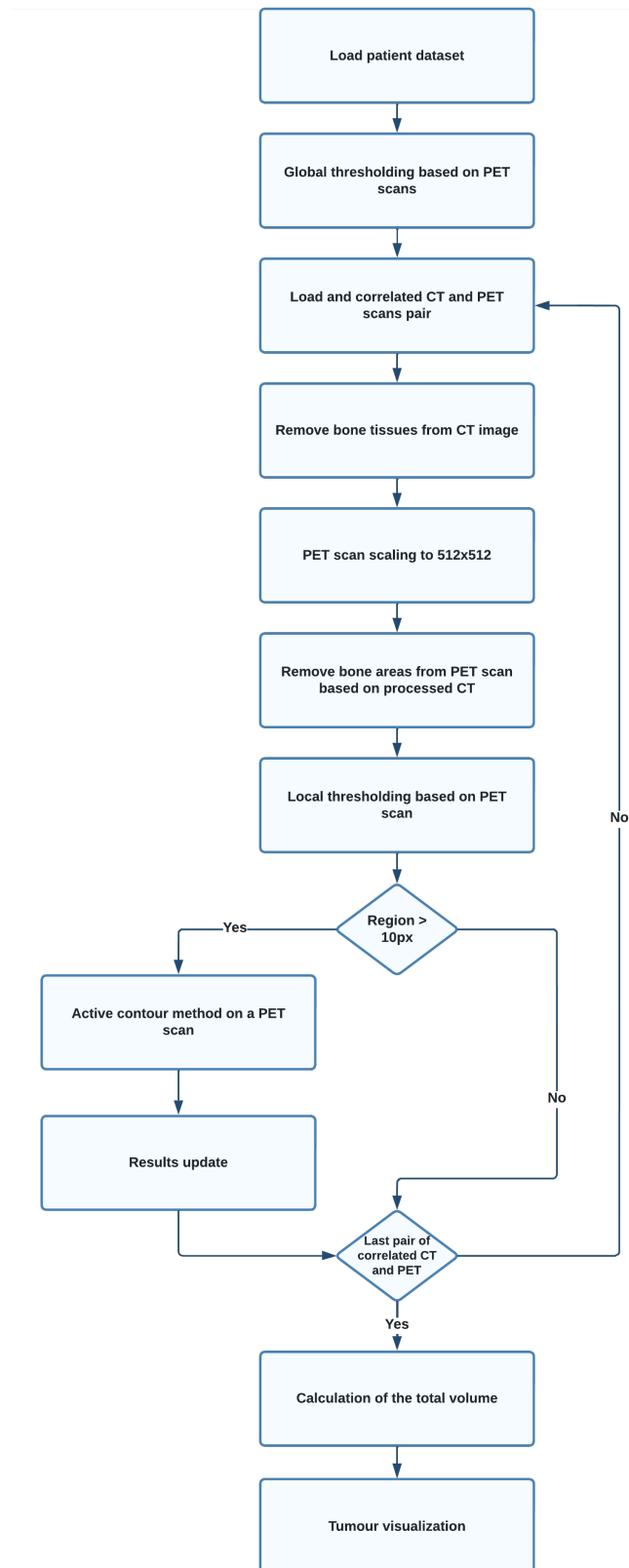


FIGURE 5.1: Active contour-based created algorithm created initially by the thesis author during the development of author's master thesis [5].

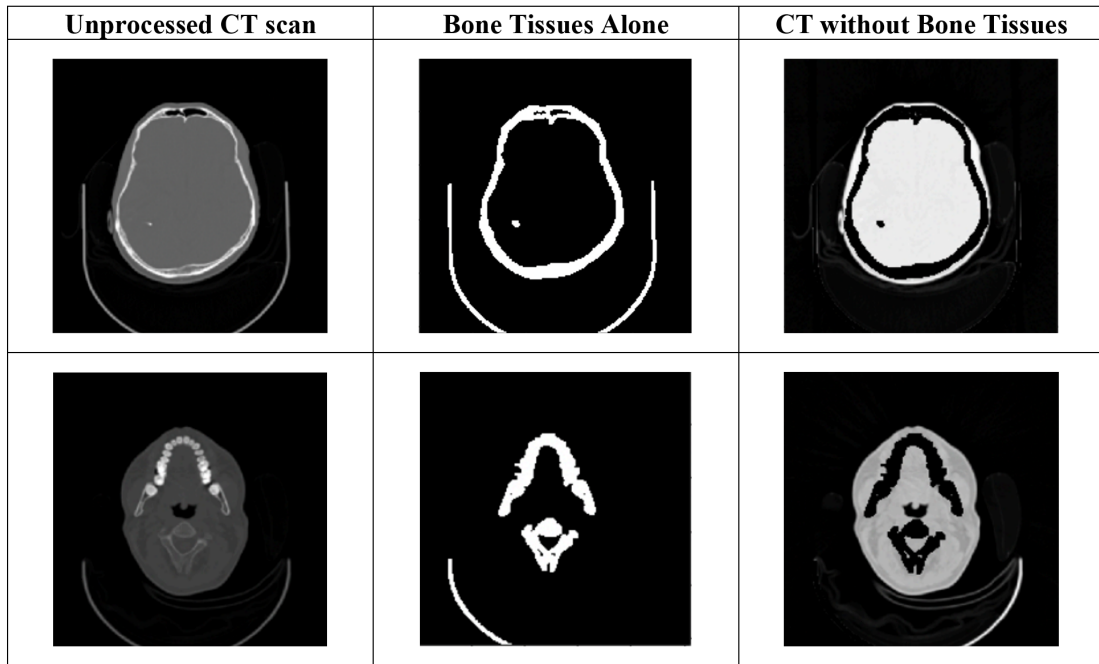


FIGURE 5.2: CT scans with automatically removed bone tissue.

different threshold values can identify differing structures. Hence, finding the optimal binarization threshold value is a crucial challenge.

In this work, two types of thresholds were utilized: global and local thresholds. The global threshold value was computed based on the entire image. For instance, in a study on calculating the volume of glioma, the global threshold value was determined from a series of 148 brain images captured using the positron emission tomography method. On the other hand, local thresholds were calculated for each image before segmentation. In this approach, the global threshold was selected from the calculated local thresholds in the initial stage of segmentation. Phantom tests were conducted, and values of 0.8 and 0.4 were obtained. It is worth noting that images acquired after surgery or with contrast agents may contain substances that could be falsely interpreted as tumors. Therefore, one of the objectives of the designed algorithm is to eliminate bone tissue, especially the skull (Fig. 5.4). To accomplish this, the threshold was shifted to 1190 based on the histogram, which was then followed by the selection and removal of the five largest objects from the resulting binary image.

In this work, the PET images have a size of 336x336 and the CT images have a size of 512x512. In order to merge these images, scaling of larger images to smaller ones is required. This scaling is possible due to the parameter defined in DICOM images known as pixel spacing. Pixel spacing represents the distance between the center of each pixel in the DICOM image, and is determined by a numeric pair representing adjacent spacing between the lines and adjacent spacing between the columns, given in millimeters. Since the pixel spacing differs for PET and CT images, PET must be scaled to the parameters of the CT image in order to obtain a fusion of images. The next step in the preparation phase to run the active contour method is to eliminate all pixels corresponding to the skull or tumor. The skull appears in black in the images

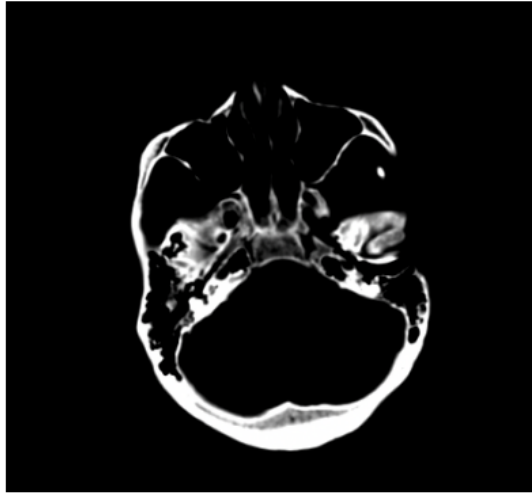


FIGURE 5.3: The effect of automatic image binarization.

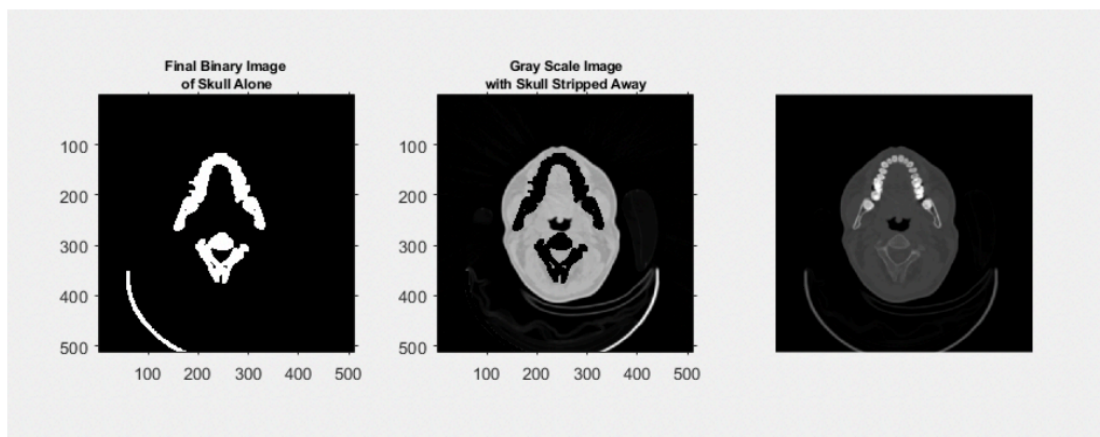


FIGURE 5.4: Removing bone structures from a CT image. Bone tissue is found on the left. In the middle, a CT image with dissected bone tissues is shown. On the right, an original CT image is presented.

due to binarization and thresholding, similar to the tumor. Thus, the pixel value in the CT image is examined and in the case of finding a value of 0 (indicating black), the corresponding pixel in the PET image is excluded. Following these transformations, the PET image is fit for the application of the active contour method.

Initially, the patient data is read from individual folders containing CT and PET examination files. The data is then processed to automatically determine a global threshold for further binarization. To prepare the CT image for scaling to PET size and fusion with PET images, bone tissue extraction is performed as a second step. This is followed by the extraction of the skull areas from the PET image based on the processed CT image, and local thresholding of the processed image. The active contour method is subsequently applied to the PET image to identify the area of the found object, which must be larger than 10 pixels and classified as a complete tumor or portion of one. Finally, the algorithm calculates the necessary parameters and visualizes the tumor.

5.1.1 Edge-based Method

The Edge-based active contour method was first introduced by Caselles et al. in 1993 [93]. The method was based on the idea that the energy functional should depend on both the image data and the length of the curve. The energy functional is minimized using a gradient descent approach to evolve the contour. The method uses the edge information of the image to detect and segment objects. The method starts with an initial contour, which is iteratively updated to the object boundary. The energy functional is defined as the sum of two terms: an internal energy term that constrains the curve to be smooth, and an external energy term that attracts the curve towards the edges of the image. The external energy term is computed as the gradient magnitude of the image.

The geodesic active contour was utilized in the algorithm through the method called: "activecontour(PET scan, mask (a binary image that specifies the initial state of the active contour), 20 as maximum number of iterations, Edge method, SmoothFactor (The default smoothness value is 1 for the "Edge" method), 1.0, ContractionBias, -0.35)". The Edge method was described by Caselles et al. [94], and an additional parameter, ContractionBias, was introduced, which indicates the contour's tendency to grow outward (negative values) or merge inward (positive values).

5.1.2 Chan-Vese-based Method

Chan-Vese is an active contour model for image segmentation proposed by Tony F. Chan and Luminita A. Vese in 2001 [95]. It is a level-set method which can segment images into multiple regions based on differences in intensity and texture. In this approach, the image is divided into two regions based on a threshold value, and an energy functional is defined which attempts to minimize the difference between the regions while maintaining the smoothness of the boundaries. The energy functional is represented as a sum of two terms, an internal energy term, which penalizes the curvature of the contour, and an external energy term, which is based on the difference between the intensity inside and outside the contour. The external energy term can be modified to include additional information such as texture or color.

The parametric active contour method was used in the algorithm through the invocation of the activecontour() method with parameters such as PET scan, mask, 100

TABLE 5.1: Results (in cm^3) for detected and measured brain tumors provided by SME and computed by Edge and Chan-Vese methods.

ID	SME	Edge	Chan-Vese
1	17.11	38.183	17.3745
2	18.63	69.6465	25.0241
3	30.22	19.3572	11.1329
4	25.5	80.0574	36.7369
5	13.32	27.2416	12.7943
6	-	12.6758	7.7618
7	15.02	17.802	11.6712

(maximum number of iterations), Chan-Vese method, and SmoothFactor (regularity of the boundaries of the segmented regions). Higher values of SmoothFactor result in smoother boundaries, but at the risk of losing finer details. Lower values allow more irregularities but enable finer details to be included. The default value for Chan-Vese is 0, which was changed to 2 for this algorithm. The parameter 100 represents the maximum number of iterations for the active contour method. If the current contour position is the same as the position of the contour in one of the last five iterations, the active contour method stops its evolution.

5.1.3 Appraisal

Tab. 5.1 presents obtained results for detected and measured brain tumors for seven patients. The SME's values represent the ground truth and the reference for other results, also acknowledging that SME could have been in error during measurements.

In the case of patient with ID 1, the SME has a value of $17.11cm^3$, whereas Chan-Vese has a closer value to the SME at $17.3745cm^3$. For ID 3, both Edge and Chan-Vese methods have much lower values than the SME. In the cases of patient 2 and patient 4, the Edge method gives values significantly higher than the SME and the Chan-Vese. For ID 5, all methods seem to perform similarly.

Assessment

Active contour methods in general do not take into account the anatomical context of the processed image. Although bone tissue removal impacted that ability, the results, especially for the Edge methods, are significantly divergent from the ground truth. Moreover, during the research, the features of active contour methods such as: heavily reliant on the placement of the initial contour, blurry boundaries, lack of support to segment multiple tumors simultaneously on the same slice and heavy parametrization for the specific tumor and medical results concluded that research must be extended on the modern convolutional neural network processing. Two modalities have been used. The qualitative evaluation with the cooperating hospital confirms that the solution only partially meets their needs, as many patients also have MRI scans, but for some there is simply a CT scan. The processing of a single patient's data took around three minutes. Furthermore, the technology used to create the algorithms does not support scaling

the solution as the technology (MATLAB) supports execution of the final program in a single node (single machine). Active contour methods seem to currently have a didactic or historical significance.

5.2 Phantom-based Verification and Calibration Method (PVCMM)

The phantom is a cylinder composed of plexiglass (a transparent synthetic material). Within the phantom are spherical holes of a known diameter, which simulate tumors within a patient's body (Fig. 5.5 and Fig. 5.6). The primary purpose of the phantom is the annual calibration of medical equipment aimed at monitoring, maintaining, and improving the quality of the results provided. The image produced in the study is recorded in the DICOM format. It includes the `DateOfLastCalibration` parameter, which provides the ability to monitor the time since the last verification of the given medical equipment.

The thesis research introduced a novelty, of using a phantom to calibrate the algorithm - validate and evaluate the accuracy of algorithms and assistance in setting the correct thresholds.

The NEMA IEC Body phantom [96], commonly used to evaluate the performance of computed tomography (CT), has limitations in detecting micro-lesions. This study proposes the NIM PET-CT phantom, designed to simultaneously test the performance of PET and CT systems, and evaluate imaging quality. The NIM CT-PET phantom consists of a PET imaging module and a CT imaging module connected through bolts. Hot spheres filled with ^{18}F -fluorodeoxyglucose and cold spheres filled with non-radioactive water were used for imaging. The imaging quality of the NIM PET/CT phantom and the NEMA IEC Body phantom was relatively consistent. The NIM PET/CT phantom detected 7 mm spheres without influencing imaging quality. The NIM PET/CT phantom outperformed the NEMA IEC Body phantom in evaluating PET image quality of micro-lesions and the performance parameters of CT. The study "Brain-mimicking phantom for biomechanical validation of motion sensitive MR imaging techniques" Ozkaya et al. [97] presented a novel brain-mimicking phantom test setup that enables the validation of motion-sensitive MR imaging techniques. Due to the frequency-dependent behavior of brain mechanics, there is a need to develop brain phantom models that can mimic the broadband mechanical response of the brain in order to validate these techniques.

Moreover, it is of significant interest to note that the spatial distribution of Bismuth-213 is essentially identical to that of Gallium-68. However, for the PET-CT phantom test, the calculations were conducted using Gallium-68 due to its lower cost. The production process of Bismuth-213 is comparatively more complex, and its short half-life restricts its applications, thereby leading to a considerably higher cost compared to Gallium-68. The Phantom-based Verification and Calibration Method (PVCMM) serves as an effective tool for validating and calibrating algorithms. This approach utilizes a phantom, which is a physical object specifically designed to simulate the properties of a real-world object. The phantom is utilized to evaluate the accuracy of the algorithm by comparing the algorithmic outputs to the actual measurements derived from the phantom. This technique is particularly useful for validating and calibrating algorithms employed in medical imaging, such as CT scans and MRI scans. PVCMM is deemed a



FIGURE 5.5: Picture of the PET-CT phantom BIODEX MEDICAL SYSTEMS available at the Nuclear Medicine Department of the Medical University of Warsaw.

TABLE 5.2: Table with theoretical values, calculated according to the phantom's catalogue card. The total volume of all tumors is 21.316cm^3 .

ID	Sphere diameter [mm]	Sphere Volume [mm3]
1	10	523.5988
2	13	1150.3465
3	17	2572.4407
4	22	5575.2798
5	28	11494.0403

reliable and accurate method for ensuring that algorithms perform as anticipated and that their outputs are precise.

The block diagram of the algorithm used to determine the parameters for the main algorithm is presented in Fig. 5.7. The main differences compared to Fig. 5.1 is that the test data on the phantom required automatic differentiation based on size: $\text{img.bytes} > 520000 \& \& \text{img.bytes} < 530000$. The size of the image obtained by the PET method is smaller than that of CT. Subsequently, the scaling factor was calculated to obtain the actual tumor area using the code: $\text{scalingFactor} = \text{dicom.PixelSpacing}(2) * \text{dicom.PixelSpacing}(1)$. Fig. 5.8 presents the results of the algorithm with the visualization created by RadiAnt software, and Fig. 5.9 presents own visualization developed in MATLAB.

The algorithm calculated the tumor volume from the obtained images to be 21.4157cm^3 , which results in a relative error value of 0.047% based on the given scans. The results for each sphere are presented in Tab. 5.2.

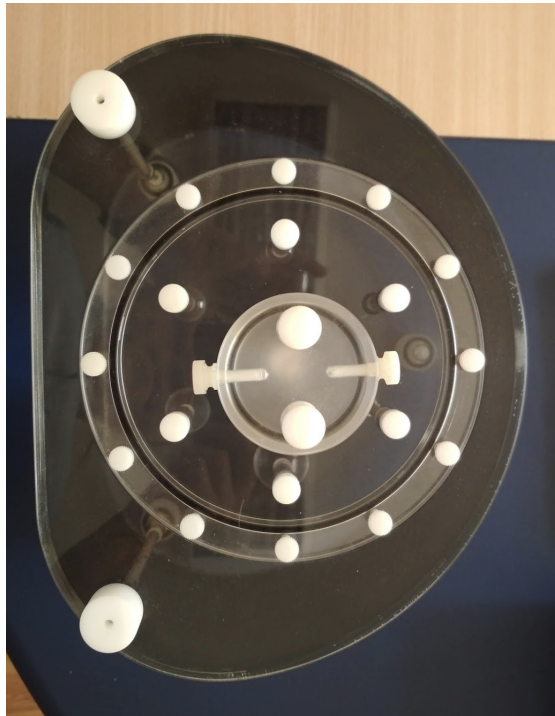


FIGURE 5.6: A bird's-eye view of the PET-CT phantom BIODIX MEDICAL SYSTEMS available at the Nuclear Medicine Department of the Medical University of Warsaw.

The research confirmed that phantoms can be used to generate synthetic data that closely mimic authentic medical images, providing a controlled environment for testing and evaluation. Additionally, phantoms can be used to assess the robustness of models to variations in image acquisition and processing parameters.

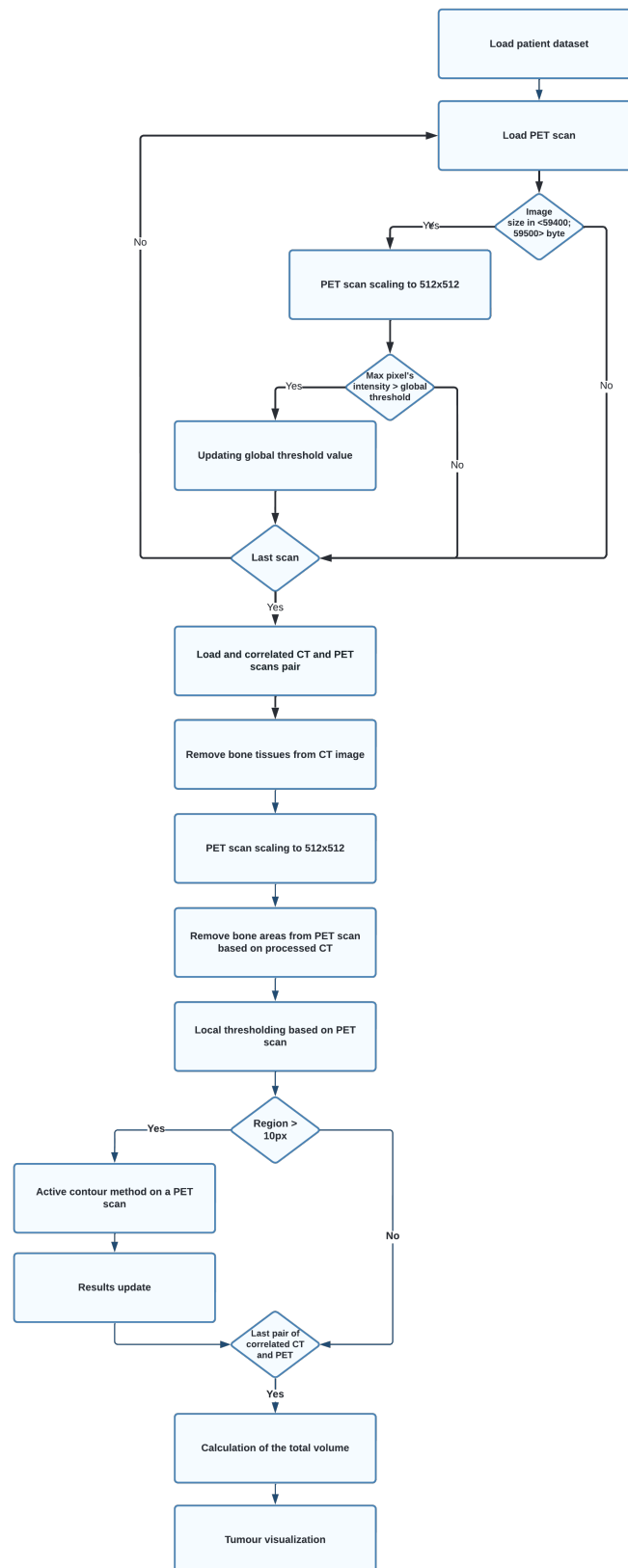


FIGURE 5.7: Diagram of the developed algorithm to assess its accuracy based on the data from physical phantom.

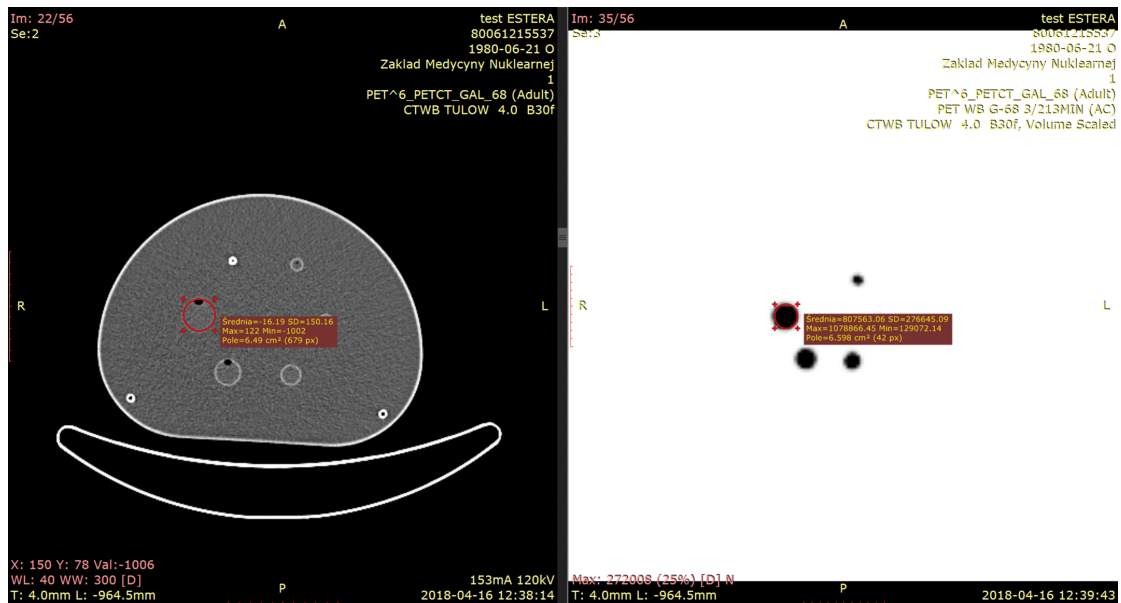


FIGURE 5.8: Developed algorithm detects tumors on the phantom. Visualization done with the usage of RadiAnt [2].

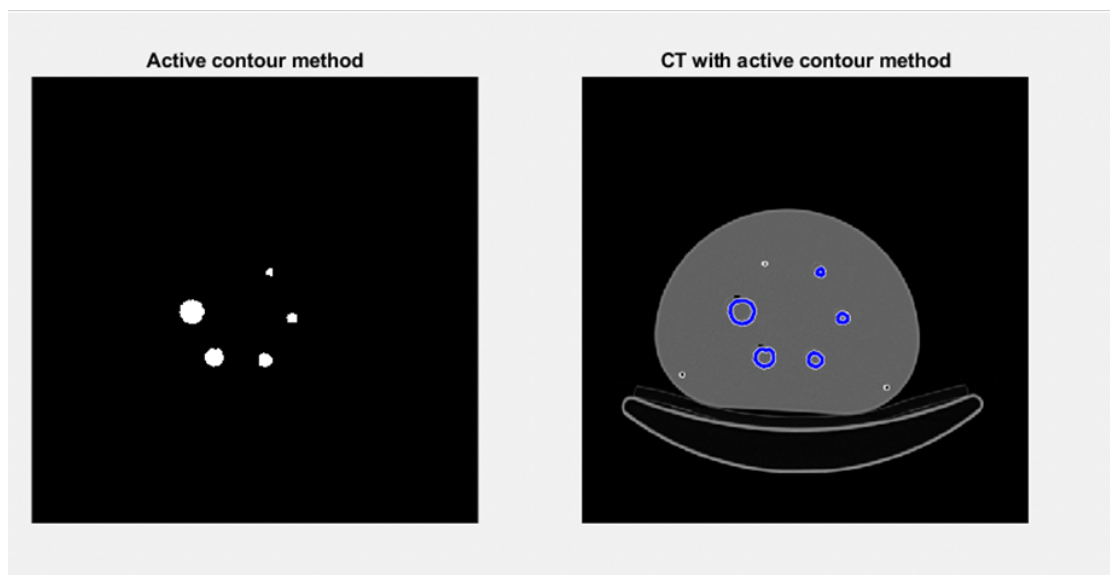


FIGURE 5.9: Image obtained using the developed algorithm. On the left side, a two-layer PET image is visible with tumors identified using the active contour method. On the right side, the PET-CT fusion with the active contour overlaid is presented.

Chapter 6

Convolutional Neural Network-based Approach for Segmentation

6.1 U-Net with Morphological Geodesic Active Contour

The proposed algorithm follows a block diagram, illustrated in Figure 6.1, that outlines the sequence of tasks performed.

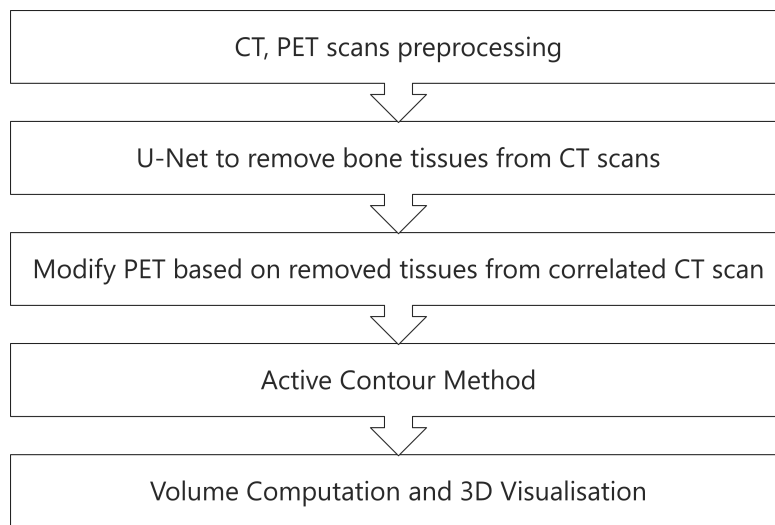


FIGURE 6.1: Flow diagram for developed method based on U-Net and active contour.

Morphological Geodesic Active Contour (MGAC) is a type of image segmentation technique that uses a combination of morphological operations (to preprocess the image to remove noise and enhance the edges of the object) and geodesic active contours (to detect and delineate objects in the image).

Brain scans were performed on 22 male and female patients. For each patient, 148 PET and 148 CT images were saved during a single examination. An additional examination of an undetermined number of patients was performed. Data augmentation was applied due to the low number of patient scans. The ImageDataGenerator from the Keras library was used. A generator was implemented for both CT images and the

corresponding masks for each slide. Elastic deformation was achieved by practicing rotation, shear intensity, interval, zoom, and horizontal flip inputs. Therefore, 13,480 files were created. Bone tissue removal was performed using the U-Net model. The training portion was completed using expert-labeled scans. Of the total dataset, 20% of the images were used for that phase. The training objective is to decrease the differences between the gray levels of pixels (Hounsfield scale) and to increase the ROI of the areas labeled by the expert. The activation function was ReLU, except for the last layer, where sigmoid activations were applied. The Adam optimizer and categorical cross-entropy as the loss function were used. The algorithm with U-Net was developed using Python 3.7, Keras 2.3.1 and Tensorflow 2.0.0a0. The algorithm was trained and run on a single virtual machine with 64 virtual CPU(s), 259GiB of RAM. The best model achieved 97.39% accuracy and 7.76% loss – giving a comparable result to others that implement U-Net [98]. After the removal process of bone tissue, the Python active geodesic morphsnakes method was used to segment the tumor area (Edge Active Contour Method with 20 maximum iterations. For the Chan-Vese method, 100 iterations). The behavior of the method utilized is similar to that of Active Contour such as Geodesic Active Contour [94]. Quantitative results are presented in 6.3 and 6.4. Qualitative results are presented in Fig. 6.2. The complete description of the methods and experiments was published in 2020 [99]. For comparison, the brain size of the women participating in the studies was 1406.62 ± 101.41 cm³ at the age of 43.88 ± 14.74 , and men $1406.57 \pm 101, 69$ cm³ at 42.96 ± 12.31 [100].

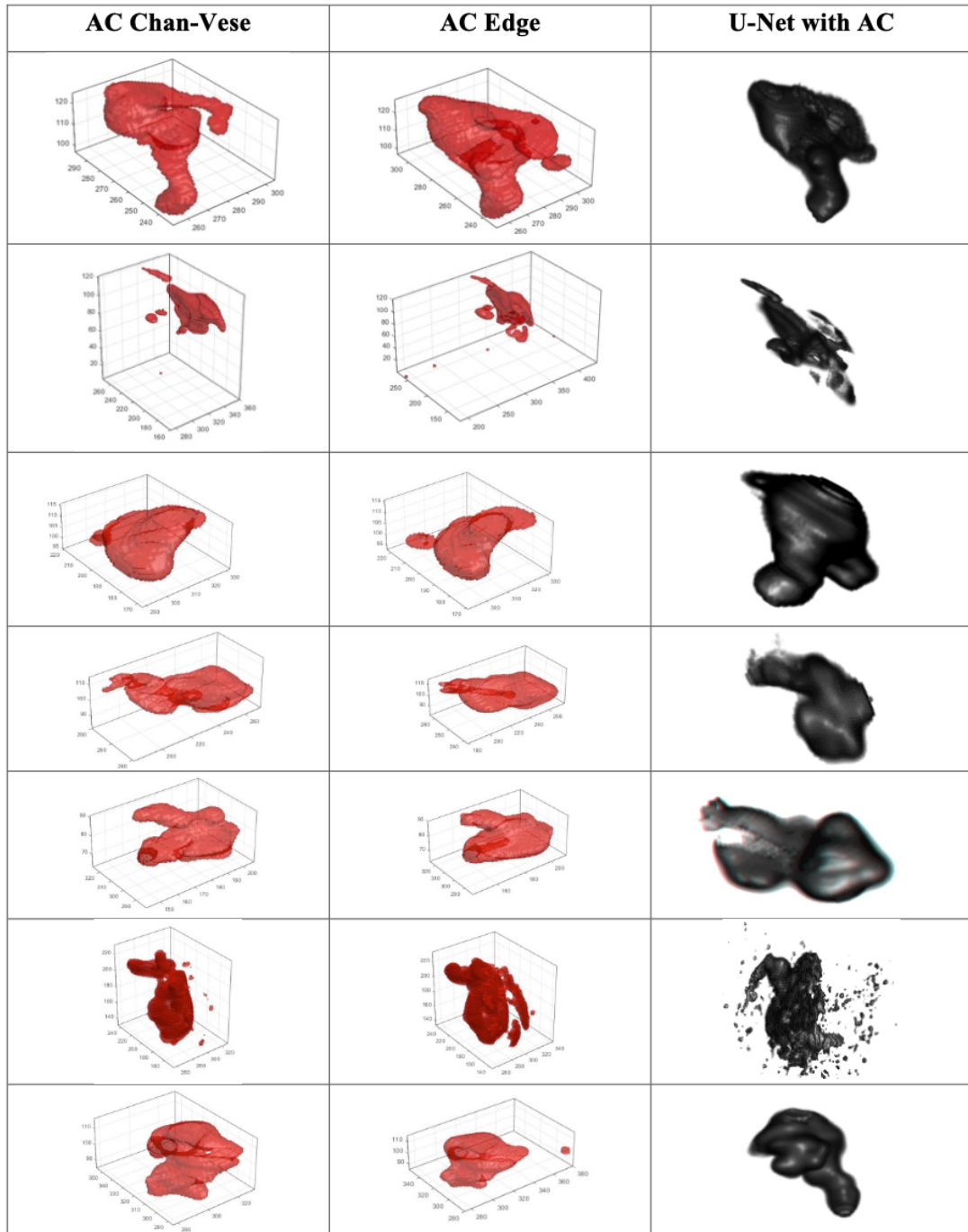


FIGURE 6.2: Qualitative results presented by the dissertation author during International Joint Conference on Neural Networks (IJCNN) in 2020 [6].

6.1.1 Improved Morphological Geodesic Active Contour

The following methods were evaluated during the research:

1. Method 1 - Morphological Active Contour without Edges was applied, where the initial level-set was created using the checkerboard level set function with the image shape and 6 as parameters, and the morphological Chan-Vese function with smoothing=3 and iter callback = callback was used to calculate the final level set.
2. Method 2 - Morphological Geodesic Active Contour evolution was performed by first using inverse gaussian-gradient function on the image, then creating an initial level-set with np.zeros function and setting a region of interest, and finally using morphological geodesic active contour function with smoothing=1, balloon=-1, threshold=0.69, and iter callback callback.
3. Methods 3, 4 and 5 are improved versions of Morphological Geodesic Active Contour that include additional image processing techniques with parameters selected empirically. Method 3 involves image erosion using disk(3), threshold = 0.8, and threshold otsu function. Method 4 involves closing the image using disk(3), threshold=0.8, and threshold otsu function. Method 5 involves image dilation using disk(2), threshold = 0.8 and the threshold function of the otsu function.

The Jaccard-Needham dissimilarity between 1-D boolean arrays u and v , is defined as Eq. 6.1.

$$JN = \frac{c_{TF} + c_{FT}}{c_{TT} + c_{FT} + c_{TF}} \quad (6.1)$$

where c_{ij} is the number of occurrences of $u[k] = i$ and $v[k] = j$ for $k < n$.

The Jaccard-Needham coefficient was used to evaluate the similarity of the pixel values between the original image and the processed one. It works by comparing the number of pixels that are common to both images with the total number of pixels in both images. The resulting score is a value between 0 and 1, with 1 indicating a perfect match. The author is aware that the Jaccard-Needham coefficient has some limitations. It does not take into account the intensity or color values of the pixels, but only their presence or absence. It also does not account for the spatial relationships between the pixels. Therefore, it may not be the most suitable method to compare images with complex textures or patterns. But it is not the case here, as two binary images were compared. Data for 22 patients from PET and CT scans have been used. The first stage of that pipeline is the same as detailed in the previous chapter. Once U-NET produces results and PET is modified based on the removed tissues from the correlated CT scan, then an Improved Morphological Geodesic Active Contour is run.

During the research on Improved Morphological Geodesic Active Contour, five methods alone have been assessed. Tab. 6.1 presents quantitative results. From each patient, 10 random slices with visible tumor were selected which provided 220 slices. Each method has been executed on the selected data to produce the segmentation of the tumor. Then, for each result the author calculated Jaccard-Needham and Dice Coefficient. The table represents the average results, when comparing the result from the methods and original image – essentially, the author's novel method was utilized which is intended to improve the default Morphological Geodesic Active Contour by adding morphological operations inside.

TABLE 6.1: Results for 5 methods morphological active contour.

Method	Jaccard-Needham AVG	Dice Coefficient
Method 1	0.75	0.86
Method 2	0.83	0.91
Method 3	0.86	0.92
Method 4	0.71	0.83
Method 5	0.85	0.91

Notably, Dice coefficient (DSC) can be expressed as: $DSC = \frac{2 * JN}{1 + JN}$ where JN is the Jaccard-Needham coefficient.

Jaccard-Needham coefficient is a measure of how similar two sets are meaning the proportion of shared elements between the two sets. It is calculated by taking the size of the intersection of the sets and dividing it by the size of the union of the sets, The Dice Coefficient instead of dividing by the size of the union of the sets, divides by the average size of the two sets. While these two metrics often produce similar results, they can differ in specific situations. The Jaccard index is more sensitive to the size of the union of the sets, while the Dice coefficient is more sensitive to the size of the intersection. Therefore, reviewing Tab. 6.1 - the Dice Coefficient is higher because it normalizes by the average set size, whereas the Jaccard Index is lower because it normalizes by the larger union size.

Qualitative results are presented in 6.2. Method 3 (Improved Morphological Geodesic Active Contour with erosion) has the highest Dice coefficient, indicating that it may be the most effective method for detecting tumors on PET-CT scans of gliomas.

Assessment

For gliomas, they are typically heterogeneous in nature and can have irregular and poorly defined borders. They may also have varying degrees of contrast enhancement on CT scans, ranging from no enhancement to strong enhancement. Additionally, gliomas can have areas of necrosis, hemorrhage, and cyst formation, which can further complicate their appearance on imaging. Therefore, it may be beneficial to apply morphological methods such as erosion and dilation to help better define the borders of the tumor and remove any surrounding artifacts or noise. Additionally, the use of evolution-based techniques may also be useful in accurately identifying the boundaries of the tumor. Overall, a combination of multiple morphological methods may be necessary to improve the detection and characterization of gliomas on CT scans.

6.1.2 Results and Evaluation

Then, based on the most accurate methods, the same steps have been repeated from the previous chapter. Results are presented in 6.3. To properly assess the results, the calculation of AE - Absolute Error, RE - Relative Error, and based on that MSE, RMSE and MAE is needed (Tab. 6.5).

TABLE 6.2: Visual results of applying morphological methods.

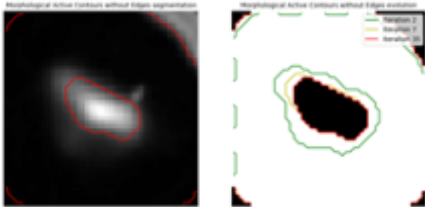
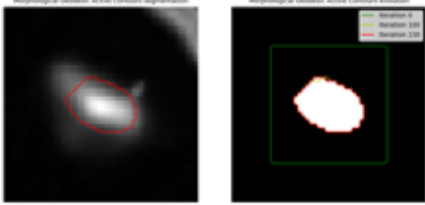
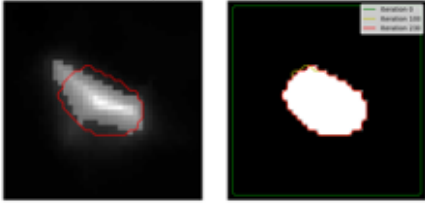
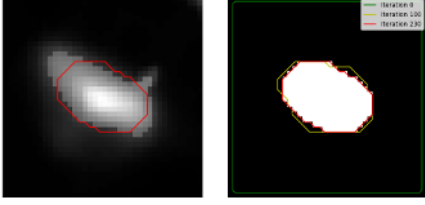
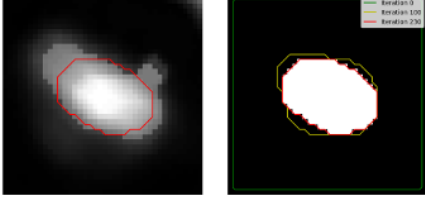
Method	Visualization	
Method 1	<div> <div>Morphological Active Contours without Edge Segmentation</div>  </div>	
Method 2	<div> <div>Morphological Geodesic Active Contours Segmentation</div>  </div>	
Method 3	<div> <div>Morphological Geodesic Active Contours Segmentation</div>  </div>	
Method 4	<div> <div>Morphological Geodesic Active Contours Segmentation</div>  </div>	
Method 5	<div> <div>Morphological Geodesic Active Contours Segmentation</div>  </div>	

TABLE 6.3: Results of Tumor Volumes per specific methods, expressed in cm^3 .

ID	SME	Edge	Chan-Vese	U-Net with MGAC	U-Net with IGMAC
1	17.11	38.183	17.3745	13.76	15.34
2	18.63	69.6465	25.0241	97.97	78,32
3	30.22	19.3572	11.1329	29.26	29.84
4	25.5	80.0574	36.7369	18.95	21.23
5	13.32	27.2416	12.7943	11.92	12.64
6	-	12.6758	7.7618	7.82	8.42
7	15.02	17.802	11.6712	15.97	16.03

TABLE 6.4: Calculations of RE and AE for received results for specific methods.

ID	Edge		Chan-Vese		U-Net with MGAC		U-Net with IMGAC	
	RE	AE	RE	AE	RE	AE	RE	AE
1	0.5519	123.16%	0.0152	1.55%	0.2435	19.58%	0.1154	10.34%
2	0.7325	273.84%	0.2555	34.32%	0.8098	425.87%	0.7621	320.40%
3	0.5612	35.95%	1.7145	63.16%	0.0328	3.18%	0.0127	1.26%
4	0.6815	213.95%	0.3059	44.07%	0.3456	25.69%	0.2011	16.75%
5	0.511	104.52%	0.0411	3.95%	0.1174	10.51%	0.0538	5.11%
6	0	0	0	0	0	0	0	0
7	0.1563	18.52%	0.2869	22.30%	0.0595	6.32%	0.063	6.72%

Assessment

IGMAC methods outperformed the typical method. At the same time, it still does not produce better results than Chan-Vese. The author did not run the visualization because the qualitative results are simply to view and compare, without making further decisions. Both methods proved to be efficient and accurate. However, knowing that a further research objective is to process more multimodal images, manual parametrization of active contour methods is unacceptable. Morphological operation showed potential in postprocessing of the images.

TABLE 6.5: MSE, RMSE and MAE for achieved results for specific methods.

Measure	Edge	Chan-Vese	U-Net with MGAC	U-Net with IGMAC
MSE	151.0194	12.9293	151.2558	85.3783
RMSE	30.10177	8.807713	30.12532	22.63338
MAE	25.70222	6.809517	15.425	11.3

6.2 Multimodal Data Fusions Algorithms

6.2.1 CT and PET Scans Fusion Algorithm

CT are rarely able to detect tumor structures because the soft tissues as well as the structure of the brain are presented. A PET scan must be executed during a medical examination via the same CT-PET scanner in order to base a diagnostic decision on a CT. Two series of slices per each of the types - CT and PET - must be executed. These slices are not merged, only correlated. Fusion of the CT and PET scans must be executed in order for the radiologist or researcher to conduct further analysis. The highest value pixels reveal themselves as white, with the lowest (meaning 0) as black. The degree of absorption of X-rays, determined by Hounsfield units (HU), is indicated as a level of gray, which reflects the structure of the brain as reflected in the CT scan. In order to determine whether the developed model distinguishes between abnormal and normal brain matter regions as well as the correct classification of benign or malignant tumor along with healthy tissue is a matter of proper training sets and validating data. The Nuclear Medicine Department acquired and delivered a dataset of 22 patients used for the purposes of this research. A Siemens Biograph 64 PET-CT scanner was used during the years 2016-2018 to take the images, 148 slices for both PET and CT scans. The weight of each slice was approximately 520KB, which netted around 110MB for a single CT examination with its correlated PET scan.

Traditional machine learning algorithms may be incapable due to the increase in the volume of data, especially while working with MR scans that can weigh 15 GB for a single examination. The most promising developments on the horizon are modern deep learning models. The concept of the neuron as the most basic computational unit grew from the study of the human brain which takes numerous signals as inputs, uses weights to combine them in linear fashion and then transfers those combined signals using nonlinear means to generate output signals [101]. Deep learning models require massive amounts of training datasets with annotations made by medical experts in order to produce results that are accurate, which makes it harder to obtain a medical dataset.

Two approaches can be taken to mitigate this difficulty. One can apply transfer learning which allows transfer features from non-medical data. Additionally, data augmentation allows for increasing the number of training samples by way of cropping, transposing, elastic transform and random notations, etc.

VGG19 and VGG19-BN

The method described by Hui et al. [102] which was initially developed to fuse infrared and visible images was applied to fuse the CT and PET scans. Confirmation that the method could be successfully applied to medical image fusion was achieved through experimentation. The first stage in the method is to take the source images, namely the CT and PET scans, in order to decompose them into detail content and base parts. The base parts are then fused using a weighted-averaging fusion strategy. A pre-trained deep learning network (VGG19 or VGG19BN) is applied to compute multi-layer features in order to extract the detail content. The paper by Simonyan, K. and Zisserman,


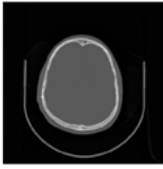
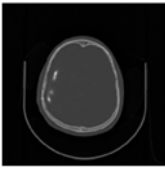
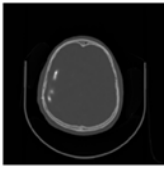

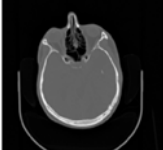
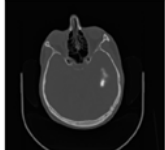
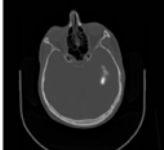
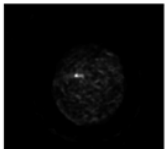
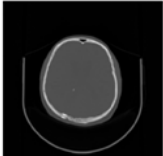
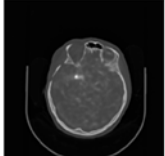
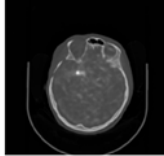
PET scans	CT scans	Fusion with VGG19	Fusion with VGG19-BN
			
			
			

TABLE 6.6: PET and corresponding CT scans and the resulted fusion done by methods based on VGG19 and VGG19-BN.

A. [103] proposed VGGNet - a Deep CNN of which, VGG19 is a variant. The research group Visual Geometry Group or VGG, was the inspiration for the model's name.

VGG19 is composed of 19 layers. As an extended VGG19 model, VGG19-BN contains batch normalization which is a layer between layers which occurs between the convolutional and activation unit layer and the inner product and activation unit layer [104]. The implementation is publicly available on GitHub [105]. ImageNet is used to train both models to extract deep features. Various types of objects which can be used for training purposes are contained in the ImageNet dataset. Utilizing the transfer learning from a pre-trained ImageNet network will significantly improve results on every dataset which attempts to compensate for a lack of adequate training data [106]. Implementing the original methods in MATLABR2016a on 3.2GHz Intel(R) Core(TM) CPU with 12 GB RAM significantly shortcuts the usability.

The authorial impact is related to the implementation. The author of the thesis implemented the fusion methods in Python utilizing the PyTorch library and making it publicly available [107] with the following changes: PET scans are rescaled to the size of the CT scans, with the usage of the inner cubic interpolation method. Table 6.6 presents a selection of obtained graphical results.

For VGG19, processing each slice takes approximately 6 to 7s of computation, with approximately 8s for VGG19-BN on 2 nodes of Virtual Machine with 224 GB of memory and 4 GPUs (Tesla K80). Tab. 6.7 shows the results comparing nine fusion outcomes.

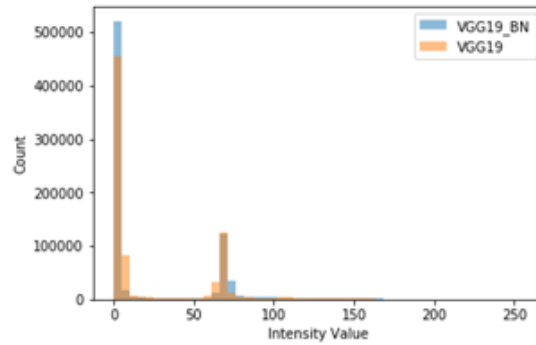


FIGURE 6.3: Histogram for fused images published. Results and methods have been published by the author in The Bulletin of the Polish Academy of Sciences Technical Sciences in 2021 [7]. As of June 2023, the article had received 10 citations.

TABLE 6.7: Jaccard coefficient results for nine outputs of the fusion method.

Method	Slice no.								
	1	2	3	4	5	6	7	8	9
Jaccard coefficient	0.9823	0.9712	0.9837	0.9864	0.9851	0.9841	0.9931	0.9864	0.9479

The impact of applying batch normalization [108] in a fusion mode are confirmed by the results as long as the identical deep learning model architecture was chosen. In order to confirm if there is any significant noise affecting the image, utilization of the histogram is valuable. In the case of the original examination results being saved in DICOM format, the histograms should reflect intensities in HU. Any transformation of an image, and especially conversion to PNG/JPG formats can change pixel intensities values. The maximum pixel intensity value obtained from the PNG files was 255 when on the original CT was 2692, and on the original PET DICOM file, 23706.

The choice of alternative deep learning models to fuse images can lead to different output images and modified pixel intensities values being produced. Pixel intensities, independent of image format, are utilized to detect tumors and especially in segmenting tumor areas. Pixel intensity is crucial for volume computation to determine a further treatment plan which is why the choice of the correct method for fusion will influence results.

6.2.2 CT, PET and MRI Scans Fusion Algorithm

The objective is to fuse images from various diagnostic devices. An original 22-patient dataset was acquired and shared by the cooperating Nuclear Medicine Department. Images were taken using a Siemens Biograph 64 PET-CT True Point Scanner during the years 2016–2018, which included 148 PET and CT scan slices. The MRI dataset consisted of seven patients who also had the CT and PET examinations.

For this research there are three data sources: CT, PET and MRI. All images were in DICOM format. For the MRI there are four modalities: T1-weighted (T1), contrast

enhanced T1-weighted (T1c), T2-weighted (T2), and T2 Fluid Attenuation Inversion Recovery (FLAIR). Only the T1 and T2 modalities were processed by the author.

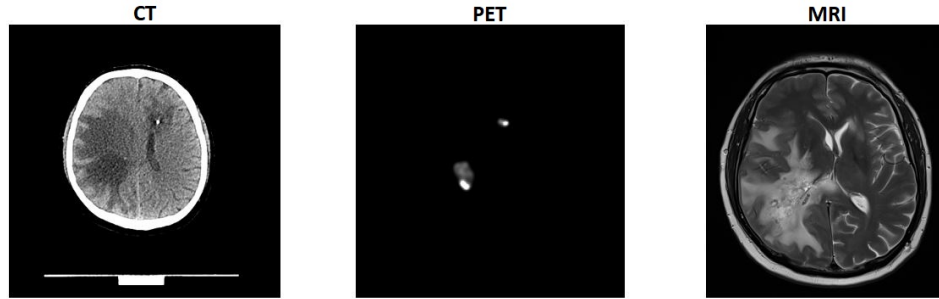


FIGURE 6.4: Three different data sources. The left picture shows the CT picture, the central picture the PET and the right picture the MRI for the identical slice of a single patient. The CT and MRI data differ from each other in terms of the color of healthy and tumor-infected tissues.

Fig. 6.4 illustrates the source elements of the fusion. The CT image with the T1 method and the MRI image with the T2 method can be identified. These photos carry various diagnostic features, therefore combining them by means of fusion may result in an increase in the value of information carried by the image.

Fusion of PET and CT

In the first stage of work, the fusion attributes are data from one measuring device. Taking PET and CT images for the same sections of the patient's head has become a standard in the performed examinations, which clearly allows for pairing of corresponding images. As the recording was done with the use of different technology, it should be noted that the images are created by registering the variances in the distribution of two different radioactive decays. This subsequently results in different resolution along with diversity in the characteristics of the input images. In order for the fusion to be performed, the input images need to be adjusted in terms of the represented resolution, which required appropriate scaling of the PET images with the use of Inner Cubic interpolation. A proven method was applied and modified in order to combine CT images and PET images [7]. The method combines infrared scans and images in the visible space. As the author of the article points out, the applied fusion process allows for the correct combination of images from two diagnostic tests of the brain. In the fusion process, source images are introduced and separated into high and low image frequencies, which, in effect, allows for the separation of basic and detailed elements, as presented in Fig. 6.5. The details are merged through the use of the pre-trained VGG19 deep learning network, the origin of which is owed to the VGGNet [109]. The resultant image of the fusion effect is an image with a resolution equal to computed tomography examination with clearly marked tumor sites indicated by PET examination. Fig. 6.6 illustrates the CT-PET fusion result.

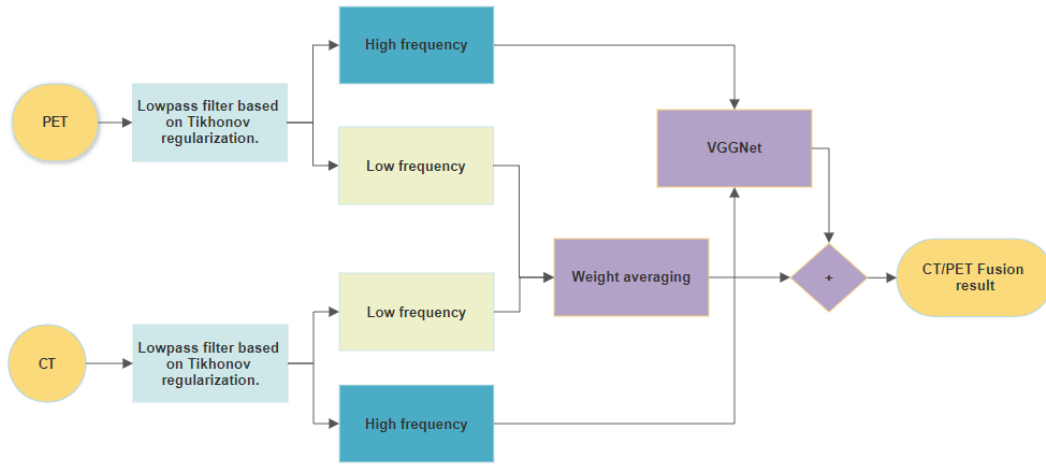


FIGURE 6.5: CT-PET fusion diagram. Applied lowpass filter based on Tikhonov regularization published in “The filtering effect of the Tikhonov regularization: Application to eddy-current problem” by Ribeiro et al. [110].

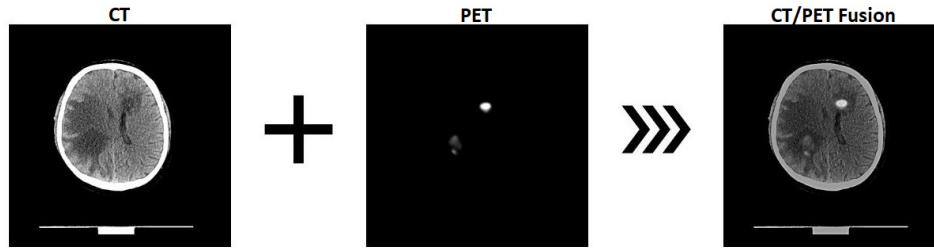


FIGURE 6.6: The CT-PET fusion. On the right side of the above figure is an image that is the result of applying a fusion. The two images on the left show the components of the fusion. Looking at the fusion image, a CT image with the tumor sites indicated by PET is seen. Thus, the conclusion can be drawn that the obtained image combines the diagnostic features of two component images.

Fusion of CT-PET with MRI

The fusion process of CT-PET with MRI was performed in the following sequence:

1. Reading the patient's position from the DICOM file
2. Preparation of data for further transformations
3. Binarization in a specific frequency window
4. Filtration
5. Labeling
6. Detection of diagnostic features
7. Determination of the base point on the series of data

8. Sorting and matching images
9. Scaling and fusion

Implementation details have been published on GitHub [111].

Features Extraction

The labeling task is to detect the places where the details are present (logical value 1) and to define the diagnostic features in the area under consideration. The following features were taken into account in the conducted research: center of the area, the size of the area, filling an area, circuit, and circularity. The circularity has been calculated according to Eq. 6.2.

$$K = 4 \cdot \pi \cdot \left(\frac{P}{O^2} \right) \quad (6.2)$$

Where K is circularity, P is area of the test, O is circuit.

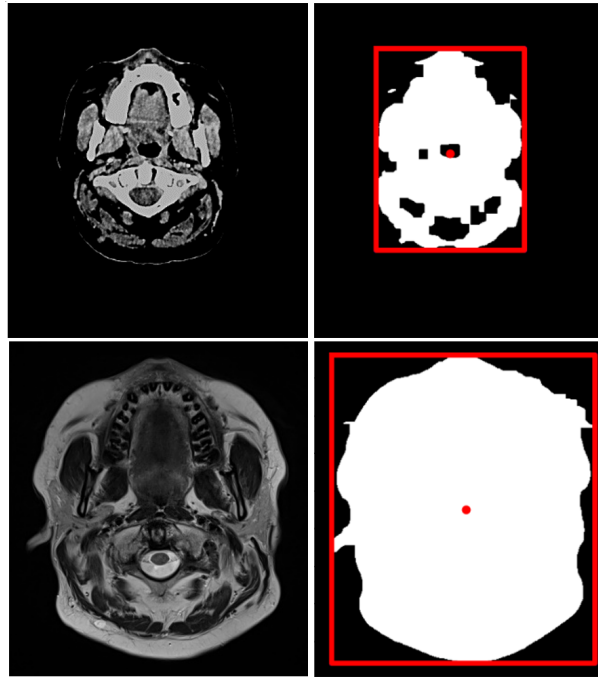


FIGURE 6.7: Transformation performed by the detection of diagnostic features. The left photo shows image before transformation, the right photo shows image after features extraction.

Fig. 6.7 presents the effect obtained for MRI images by performing binarization in the light band within the range of 7–25; then a median filter with a square 5-dimensional mask was applied. In order to remove all openings inside the patient's head, the operation of closing the morphological mask 5x5 was performed. Such an operation was performed three times. Filtered images were labeled and diagnostic features determined. The designated parameters are the center of the head, width and height, area

and roundness. Determining these parameters will allow for a further stage of image processing. Note, however, that different filters should be applied to images obtained from fusion as the nature of the image is different. The same binarization band is used on the fusion image, with a double median filtering in a 7 square frame, and triple morphological occlusion with a 5x5 frame were used to obtain MRI identical effects. The manner of the processing method determined was tested on several datasets. The effects obtained thanks to the developed method allows researchers to consider this set of filters as optimal and adapted to automatic operation. The applied filters do not affect the further quality of the fusion because the filtered images serve only to determine the characteristic dimensions of the head.

Image Matching

The head image diagnostic parameters determined in the previous stage have been saved in the table along with the name and paths leading to the original images. To solve the key problem of matching images from two diagnostic devices, information about the patient's location during the examination was added to the existing function table. Thanks to this information, it is possible to unambiguously align the images from the two diagnostic devices. This is done by finding a base point on a series of fusion and MRI data. For the head, the base point was defined as its apex, determined by finding the cross-section with the smallest area while maintaining a high circularity parameter. The determined cross-section has a parameter of the patient's position in the diagnostic apparatus, with respect to which the shift of the remaining images in the series should be determined. Baseline points should be established for both fused images and MRI images. After such operations, there are two parameter tables with specified distances from the base point for all sections. To find the correct pairs of images, the lists should be searched for images with a similar section point.

Not every series of data can be combined with each other due to the different location of the patient during the tests. The scans may be performed in cross-sections in locations which exceed the adopted tolerance limits. During CT / PET and MRI scans, a series of data is captured with a certain shift of the patient between successive segments. It may occur that one type of scan is able to capture an image which is not possible in another form of scan.

Final Fusion of CT, PET and MRI

The final step before the fusion takes place is selection of pre-processed images to represent an identical scene to their counterpart from another device. To do this, it is necessary to utilize predetermined diagnostic features, which are: the center of the head, height and width. The input images should be processed so that they have the same resolution, position, and dimensions of the head. For this purpose, it is necessary to apply scaling and image cutting operations using designated parameters. MRI images are characterized by a high level of detail, which is higher than in computed tomography, therefore it was decided to match the CT-PET fusion images with the MRI images. The input data prepared in this way was subjected to fusion, the method of which was identical to that of the CT-PET fusion and used the pre-trained VGG19 network.

Fig. 6.8 shows the results of the developed algorithm. The first two images in the figure show the matching prior to the fitting operation. The third image is the result

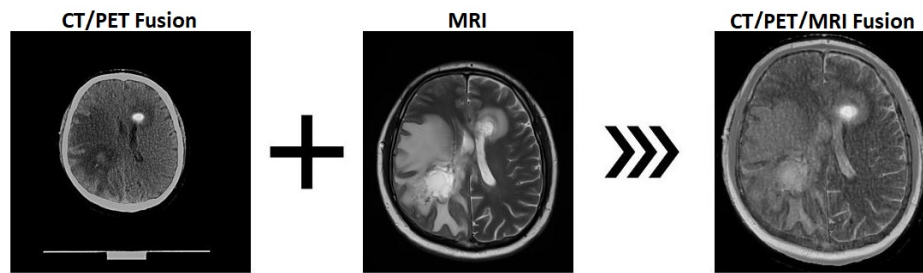


FIGURE 6.8: This figure shows final fusion. On the left the picture shows the CT-PET image, the central picture is of the MRI and the right picture of the fusion result.

image, which contains the sum of the characteristics of the component images, which are computed tomography, positron emission tomography, and magnetic resonance images.

The model to be ultimately used for fusion has been described in detail by Zhang et al. [112]. The structure of the neural network used in the model is based on the convolutional architecture and was named by the creators IFCCN with the utilization of methods IFCNN-MAX (selecting the higher value). After merging the images, the resulting image is reconstructed by two convolution layers.

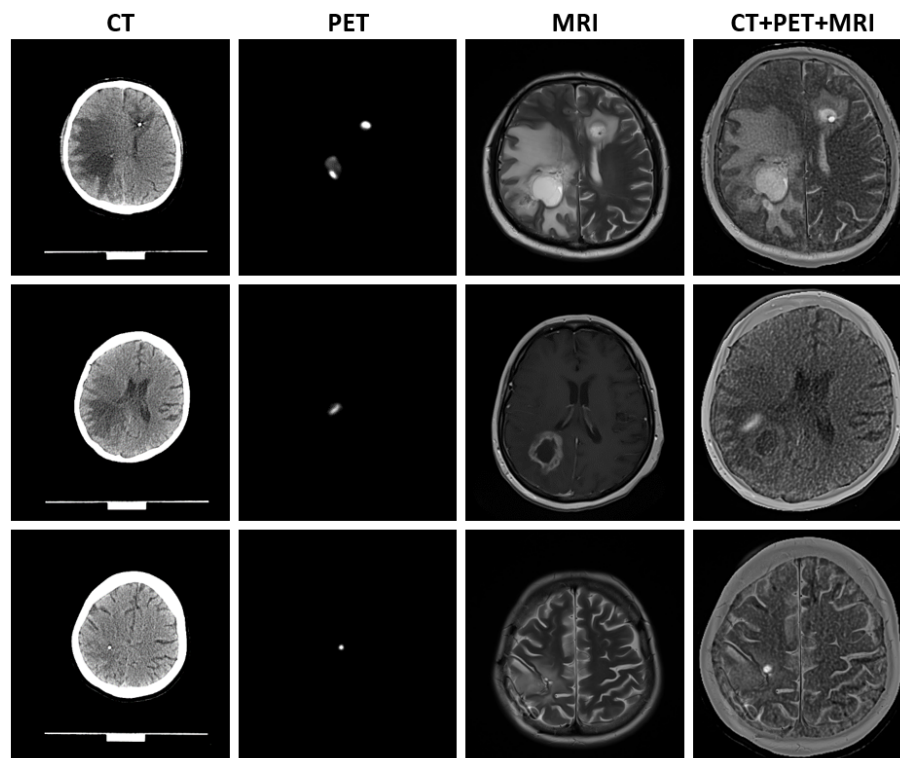


FIGURE 6.9: The result of the work is an algorithm that allows for automatic fusion of images from CT, PET and MRI.

The result of the work is an algorithm that allows for automatic fusion of images

from CT, PET and MRI. Visual results illustrate Fig. 6.9, and Fig. 6.10 presents the process diagram. In the designed program, using, among others, the libraries: PyDicom, OpenCV, Numpy, Skimage, Torch, Torchvision, Matplotlib, Pymongo and test data from the Central Clinical Hospital of the Medical University of Warsaw, the author managed to achieve the output image being the sum of the features from the component images. The data created can be used in an unchanging form as training and testing data for deep neural networks.

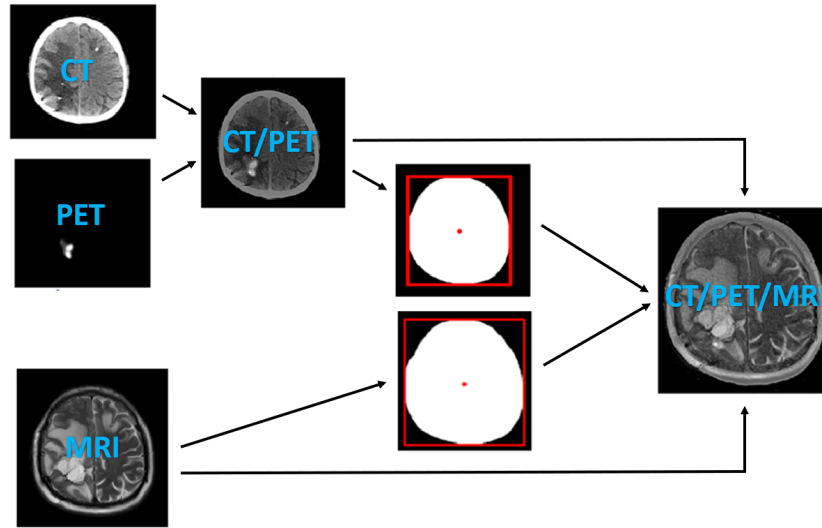


FIGURE 6.10: CT-PET-MRI fusion diagram.

6.3 Data Augmentation and Annotations

Data augmentation is a widely used technique for increasing the size of the dataset available for training ML models. The approach involves manipulating existing data points by means of various transformations, such as rotation, scaling, translation, shearing, flipping, and cropping. Through this process, new data points are generated from the existing ones, thus expanding the dataset. The primary objective of data augmentation is to mitigate the risk of overfitting while improving the generalization of the model. In particular, image augmentation has become increasingly popular in recent years, with a variety of tools available for implementing transformations such as zoom, shear, rotation, and flipping. Notable examples of image augmentation tools include imgaug, keras, smart augment, and autoaugment.

6.3.1 Mask-aware Medical Data Augmentation

Mask-aware Medical Data Augmentation (MMDA) is an authorial tool utilized to generate additional training data for medical image analysis tasks. It is designed to generate realistic data that is consistent with the original data. The tool uses a combination of image processing techniques, such as masking, cropping, and resizing, to generate new data that is similar to the original data. This allows for more accurate training

of deep learning models, which can then be used to improve the accuracy of medical image analysis tasks.

In medical imaging, masks are binary images that highlight specific regions or structures of interest within an original image. When applying traditional data augmentation techniques to medical images, the augmentation process typically alters the original image without considering the associated masks. This can result in a misalignment between the modified image and its corresponding mask, leading to inconsistencies and inaccuracies in subsequent analysis or model training. MMDA takes into account the spatial relationship between the original image and its associated mask. During augmentation, the tool ensures that any transformations or modifications applied to the original image are also applied to the mask in a consistent manner. For example, if the original image was rotated, the tool applied the same rotation to the mask to ensure that the annotated regions were positioned with the converted image. Likewise, other augmentation techniques such as translation, scaling, flipping, elastic deformation, or intensity changes, while retaining a spatial relationship between the image and its mask, are used. This means that any modifications introduced to the image are propagated to the mask, preserving the integrity of the annotated regions.

The Mask-aware Medical Data Augmentation blends:

1. Keras Image Generator – The Keras ImageDataGenerator class provides real-time data augmentation and requires lower memory usage. It is a superb way to create new variations of images on the fly without having to add them to the original corpus of images. The primary method used for data augmentation is ImageDataGenerator class. The class generates augmented image data from an existing set of images by applying various image transformation operations such as Random Crop, Flip, Translation, Rotation, Zoom, Height, Width, Contrast, Brightness. The ImageDataGenerator class in Keras allows for the real-time data augmentation of images during model training. This can greatly increase the size of the dataset and help prevent overfitting.
2. AugLy – AugLy is a data augmentation library that has recently gained attention in the field of computer vision, including medical imaging. It contains both function-based and class-based transforms, with the option to provide metadata about the applied transform. It supports adding Gaussian noise, CT windowing, and histology-specific augmentations.
3. imgaug – imgaug is a library for image augmentation in ML experiments, supporting a vast range of augmentation techniques and allowing easy combination, execution on multiple cores, and augmentation of images, key-points/landmarks, bounding boxes, heatmaps, and segmentation maps. It allows for a wide range of transformations to be applied to images, such as cropping, rotation, flipping, and color adjustments. These transformations can be applied in a random or deterministic manner to generate a large number of new images from a small dataset, thereby improving the robustness and generalization of a machine learning model.

MMDA command line tool was used to augment data in the research detailed in further chapters. Assessment of the tool per se can be challenging, as the hypothesis

that should be set is rather - how the data augmented by the tool impacts the final results. On the other hand, there are certain methods which must be mentioned that can be used to evaluate the tool.

Image Adjustment Methods - Assessment

An overabundance of highly similar images in the training dataset impacts the model by biasing it toward the specific class. This results in the model becoming over-fitted to the specific pattern presented in the scan. The critical phrase here is “highly similar images”. The author of this research desires to define these commonly used words and propose methods that may assist with quantifying that opinion. Post image generation, scans must be assessed, in regard to similarity to determine the “highly similar” ratio to the original. The following should be assessed:

1. Anatomical integrity: shape of the skull, tumor and other structures, as well as location of specific organs should be consistent.
2. Scan consistency: Hounsfield scale with the same representation of HU in expressing tissues and bone structures. In typical images, the colors should be consistent, meaning the contrast, brightness, sharpness should be consistent.

Consistent, but not the same. This establishes the necessity to assess the produced images in both qualitative - anatomical integrity and quantitative - scan consistency means. The author suggests using the Mean Squared Error (MSE) method (calculate the sum of the squared differences between the two images). The lower the MSE value, the better the adjustment. Another method is the Structural Similarity Index (SSIM). This method measures the similarity between two images by calculating the luminance, contrast, and structure of the images. The higher the SSIM value, the better the adjustment. Finally, the author suggests using the Peak Signal-to-Noise Ratio (PSNR) method. This method measures the difference between the original image and the adjusted image by calculating the ratio of the peak signal power to the noise power. The higher the PSNR value, the better the adjustment.

Moreover, the author of this thesis supervised student research focused on developing a data enrichment tool for medical imaging, particularly intent on the generation of synthetic PET-CT fusion images depicting tumors in head scans. This research culminated in a master’s thesis, presented by the student titled: “Medical images generation based on generative adversarial networks” and defended by said student, namely, Michał Przemysław Maciołek, on 23 June 2022. The system was built on the Cycle Generative Adversarial Network (CycleGAN) architecture, which leverages two generators and two discriminators to transform input images into realistic, synthetic outputs. Given the input of a CT scan, a fusion of the original image and a synthetic tumor is generated by the tool; conversely, a PET scan input yields an output that retains the original tumor while generating a synthetic background. The CycleGAN system utilizes a cycle consistency loss function for gauging transformation accuracy. The numeric method used for evaluating the quality of generated images (the Fréchet inception distance) was unable to capture subtle improvements in the synthetic pictures. As this metric is based on comparing the statistics between the distributions of

generated and real images, presumably, an enlargement of the test set would cause the method to yield more accurate results. Unfortunately, performance monitoring of generative adversarial networks is not as straightforward as is the monitoring of other machine learning systems such as image classifiers. The first reason is due to the fact that GAN's loss functions solely reflect the current performance of the generator against its adversary: the discriminator. The overall quality of generated images is not assessed by the loss functions. Secondly, the judgement of generated pictures with regard to realism is extremely difficult to express mathematically as opposed to measuring accuracy on a labeled dataset. In this thesis, there are two methods used to assess the quality of generator produced output pictures: human monitoring and Fréchet inception distance². The first approach is simple; however, it is vulnerable to bias and is time-consuming. The second method, on the other hand, removes human monitoring from the loop and compares the statistics of real and generated images distributions numerically. The initial step of this solution is information extraction from each image - both real and synthetic image distributions, utilizing the inception v3 model, with the top layers removed, thus the model exclusively outputs the extracted features, not the prediction. Following thusly, for each distribution, a vector of feature-wise means for all the activation is calculated. The inception v3 model extracts 2048 features from each image, therefore this vector has a shape equal to (2048). Next, for real and unreal images distributions, a covariance matrix of features is calculated, giving the results of this step two (2048, 2048) matrices. Finally, those metrics are plugged into Eq. 6.3 in order to calculate the final Fréchet inception distance value based on the [113,114].

$$d^2 = \|\mu_1 - \mu_2\|^2 + \text{Tr} \left(C_1 + C_2 - 2 \cdot \sqrt{C_1 \cdot C_2} \right) \quad (6.3)$$

where:

- d^2 is the value of Fréchet inception distance (the squared symbol emphasizes that the distance has square units)
- μ_i , the vector of feature-wise means of i-th distribution
- C_i , the feature covariance matrix of i-th distribution
- Tr , the trace linear algebra operation, which is effectively the sum of elements along the main diagonal of the matrix

The objective is to minimize the Frechet inception distance between distributions of real images and the generator produced pictures as the greater the distance, the more the two distributions will differ.

Experimental results suggest promising outcomes when CT scans serve as the source domain and PET-CT fusion images as the target domain. Evaluation by a radiologist of the synthetic images produced by the most promising version of the CycleGAN implemented in the thesis would utilize human discrimination to perform a binary classification of real and synthetic images. The low accuracy of this prediction would serve as the necessary evidence of the high performance of the generator produced PET-CT fusion images.



FIGURE 6.11: Authorial universal tool named Medical Annotations Transformer (MAT) which automate using labelled datasets by different deep learning models within the cloud environment.

6.3.2 Medical Annotations Transformer (MAT)

Biomedical imaging has its own very specific needs towards annotation which may be based upon or extend general purpose annotation tooling. There are startups addressing this problem, the most notable examples being md.ai (funded by Google and Stanford AIMI), TrainingData.io (created by NVidia), American College of Radiology's AILab. A solid annotation tool tied closely to the underlying data files and downstream machine learning experimentation environment will make the platform sticky and a leap towards democratizing AI.

Various deep learning models require diverse types of input data and label description. For example, in the case of U-Net, binary masks are needed. In the case of Mask R-CNN, these can be binary masks linked to a JSON formatted file containing descriptions of the tumor area and the bounding box, however, could also be the COCO dataset itself. Additionally, in the case of multiple objects in a single image, the masks used for training can include one object per slice and multiple objects in a single image. There are several tools on the market that allow radiologists and SMEs to label data. In addition, a variety of tools will allow label export in different formats. The challenge is to standardize.

To make the collaboration with Subject-Matter Experts more effective, automating the new neural network models training and retraining, a universal proprietary method - Medical Annotations Transformer (MAT) has been developed. Fig. 6.11 presents the workflow of the universal tool.

MAT was implemented as a command line tool able to be run on a virtual machine, local PC or in Compute Instance in Azure. It requires access to a registered dataset in Azure ML with the labels stored in Microsoft's JSON file format Common Objects in Context dataset (COCO). The COCO format is a specific JSON structure dictating how labels and metadata are saved for an image dataset.

The MAT main features:

1. Integration with Azure ML (Cloud Compute Engine) – access to datasets registered in Azure ML Data. This requires access to Azure ML Workspace, which means providing an Azure subscription ID, a resource group where the Azure ML workspace is deployed and the workspace name. Then the MAT resolves the authentication and authorization to these services.
2. Inspection and adaptation of a COCO dataset – particularly verifying:
 - Structure of the file, which should contain a dictionary with: Images – the section that contains per each image unique id, width, height, file name and

URL of the location of the file with the capture date there the file is located with also the date of captured. Annotations – contains a list of annotated objects, specified by the segmentations section. It should contain a list of polygon vertices around the object, for each object. Categories – a dictionary with a list of categories/classes.

- Area - for each segmented object there should be an area in pixels measured. The tool checks if areas values exist, and if values are correct will calculate it. Area value is required for generating the correct masks.
 - Crowding – parameters which indicate whether the bounding box is around the crowd of other objects. If the crowd is equal to 1 then objects are marked as one group/cluster of objects.
 - Annotation unique ID (adding if not present).
 - Images dimensions – height and width if are the same as parametrized.
 - Normalize – iterate over images, segmented areas and bbox, adjusting coordinates based on width and height values if they are divergent. Normalize value types to ints and floats.
 - The COCO bounding box format is [top left x position, top left y position, width, height] if are correct.
 - The category id corresponds to a single category specified in the categories section.
3. Generating masks – Convert COCO annotation and original image to binary files. Supporting both single and multiple objects in the final mask file.
 4. Generating Pascal VOC – Convert COCO annotation (single .json file) to Pascal VOC annotations (multiple .xml files)
 5. Downloading the dataset – both to a local or cloud environment; leveraging authentication and authorization with Azure Active Directory

The MAT result is the dictionary with three subfolders containing annotation (XML VOC), downloaded images and the generated masks. Moreover, the created datasets are registered in Azure ML Data Assets for further experimentation. That baseline allows the running of different types of models on given images in parallel.

6.4 Tumor Area Cropper with Connected U-Net Architecture Generator

U-Net architecture is designed as an auto-encoder with an expansive path and a symmetric contracting path. Its design determines that the input size should be divisible by $2n$, where n is the number of max-pooling layers, to allow for the proper down-sampling and up-sampling. This determination significantly impacts U-Net ability to serve as the core network for multi deep learning model pipeline because the input dimensions of the scan can differ. The solution for this problem is U-Net's architecture dynamically generated based on the input scan size. Moreover, the output of U-Net model must also be dimensions compatible with the following processing model.

Les et al. [115] suggested Eq. 6.4 that calculates the size of the individual slice with tumor.

$$\rho = \max(\alpha, \beta) + 30 \text{ px} \quad (6.4)$$

A small surrounding area of 30 pixels was added to the tumor width of α and height of β . Moreover, ρ is the final image calculation of width and height. The downsizes of that equations are: the extra padding of 30 pixels is big and impact the cost further computing, but the main issue is that the equation do not guarantee if the ρ can be processed further by another U-Net model (the input size should be divisible by 2^n).

The author of the thesis set the main objective to define the resulting square image's side length to be the maximum of the height and width of the detected tumor, plus a 10-pixel padding, and also be divisible by 2^n for some n .

Let α be the width of the tumor, β be the height of the tumor, and δ additional padding.

$$\rho = \max((\alpha, \beta) + \delta) \quad (6.5)$$

The objective is to find ν , the smallest power of 2 greater than or equal to ρ . This can be expressed as:

$$\nu = 2^{\lceil \log_2 \rho \rceil} \quad (6.6)$$

where $\lceil \log_2 \rho \rceil$ denotes the ceiling function, which rounds ρ up to the nearest integer. The constant 10px has been chosen to add extra padding, but also to decrease the further processing capabilities. Based on that novel equation, the architecture of U-Net can be now generated.

The thesis author introduces a novel equation that allows for the calculation of the U-Net architecture depth, based on the image input size:

$$\nu = 2^{\lceil \log_2 (\max(\alpha, \beta) + \delta) \rceil} \quad (6.7)$$

Architecture Generator

If the image size is not a power of 2, there are two strategies to deal with it: resized or cropped. If applying cropping, then then it follows the 6.7 rule to calculate the area to crop. Scaling technique can be split into two methods: interpolation to increase the size and downscaling for decreasing the size. Downscaling can lead to loss of information. When scaling up, the interpolation is used to estimate the value of new pixels, that do not bring new information, simply increasing the size of the image (more heavily for computing), and the approximation algorithms may introduce the additional level of error. Therefore, the conclusion is that it is better to not scale the image. It is better to generate dynamic architecture of the U-Net model based on the input image. Especially for big resolutions images like histopathology, it has huge potential to cut the smaller areas and process them efficiently, in parallel. The critical point is to connect the the depth of a U-Net architecture with the input size of the images. The author of the thesis propose a dynamic U-Net architecture generation based on:

1. Calculation of the Depth of the U-Net. For a 256x256x1 image, the depth U-Net is: 256 -> 128 -> 64 -> 32 -> 16 -> 8 -> 4 -> 2 -> 1. In practice the U-net can be stopped on resolution 16x16, because going deeper might lead to overwriting, and it also greatly increases the computational cost.
2. Correlation of the depth and max pooling layers (each layer of depth corresponds to a stage in the encoding (contracting) path where one applies a max pooling operation. Each max pooling operation typically halves the spatial dimensions of the feature maps - assuming a pool size of 2x2 and stride of 2, which are common settings).
3. The number of filters in a U-Net typically starts small and doubles after each max pooling layer in the encoding path, and halves after each sampling layer in the decoding path.
4. Activation function: the rectified linear unit (ReLU) and a sigmoid activation in the final layer (the sigmoid function outputs a probability between 0 and 1), but for multiclass segmentation tasks, a softmax activation function might be used in the final layer instead.
5. Include batch normalization (to accelerate training and improve generalization) and dropout layers (prevent overwriting by randomly setting a proportion of input units to 0 at each update during training time).
6. Binary cross entropy loss function for a binary segmentation, categorical cross entropy for multiclass segmentation. But also, the dice coefficient or Jacquard/intersection over union (IoU) are also commonly used as loss functions.
7. The Adam optimizer as a commonly used choice.

The author's example of results from the generators has been published as open source on Github [116] for the images size as 64 px. The research results have also been applied in the final experiments when Mask-RCNN hyperparameters tuning was executed.

Fig. 6.12 presents the conceptual algorithm where the stage Tumor area cropping was added. This stage occurs prior to tumor segmentation. Tumor Area cropping and Tumor segmentation can be looped, depending on how many models in the final pipeline are used.

6.5 U-Net-based Segmentation on CT and PET Scans

An initial framework for brain tumor detection and semantic segmentation introduced in a previous work [7] has been improved. Instead of feeding U-Net 64x64x1 cropped images of a given tumor, proposed training techniques that include a network model that can learn from a dataset with original format images (512x512x1) to efficiently segment glioma.

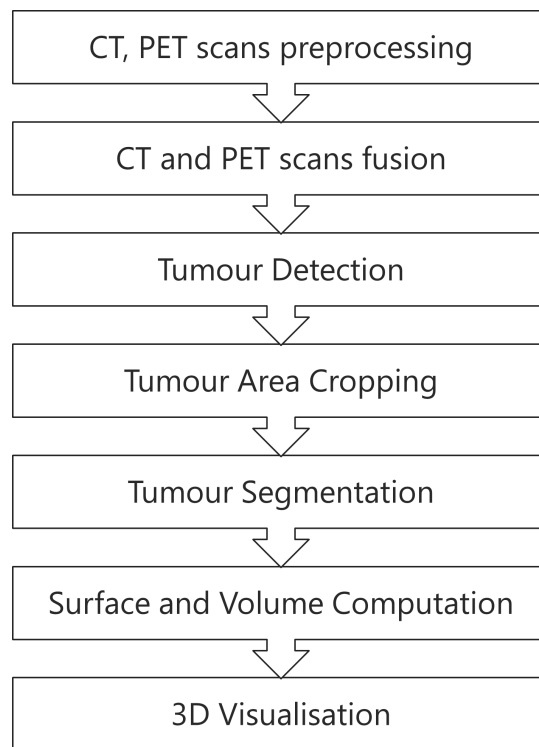


FIGURE 6.12: General diagram of the 7th stage pipeline. Tumor area cropping is performed before tumor segmentation, for example via U-Net models.

6.5.1 Processing Methods and Experiments

The fusion was accomplished by the dissertation author as a result of paper “Deep learning-based framework for tumour detection and semantic segmentation” [7] using VGG19 (chapter 6.2.1). In order to establish a conducive setting for parallel cloud-based experiments the decoupling of storage and computation was implemented to establish a conducive setting for parallel cloud-based experiments. The aforementioned dataset was migrated to Azure Data Lake Storage. A subject matter expert utilized Azure Machine Learning Data Labelling in order to annotate the dataset in question, whereby each tumor was assigned a polygon delineating its shape. Subsequently, the image labels were exported to the COCO format. As the U-Net architecture necessitates a singular binary mask for each slice, a script was implemented to extract masks from polygons.

The Lambda layer is used for the purpose of normalizing inputs. The optimization algorithm utilized for training the algorithm was Adam [117], with a learning rate of 0.001 and an epsilon value of $1e-07$. Binary cross-entropy was selected as the loss function to determine the objective that the model should aim to minimize while undergoing training. The preceding model consists of a total of 1,940,817 parameters. Throughout the training process, the primary metric under observation was accuracy, with the implementation of an early stopping strategy. Specifically, if the accuracy fails to improve during the final two epochs, the training regimen is halted to mitigate the risk of overfitting. The model architecture that has been utilized in Tensorflow2-Keras has been disseminated on GitHub [8].

6.5.2 Training on the Original Dataset

The training procedure involved a batch size of 32, 50 epochs, and 5-Fold cross-validation with a shuffle, where 80% of the data was allocated for training and 20% for validation. The K-fold method partitions a given dataset into k non-overlapping and equally sized groups or folds, each containing a subset of the samples. Following each fold, the models were evaluated using a variety of performance metrics, including the selected loss function, Accuracy, MSE, ROC-AUC, and the mean of IOU. The implementation of Mean IOU in Tensorflow 2.4 was found to be ambiguous as it did not exclude the background, which is a requirement of the Jaccard index. As a result, the Sørensen–Dice coefficient (Dice Co-Eff) was employed as an alternative measure, and the models were assessed on a per-slice basis using the test dataset. A total of 472 slices were utilized for the purpose of training, while 117 slices were allocated for testing.

Single Tumor Extracted

The first objective was to train the U-Net on the original dataset of a single tumor extracted. The machine used for training was - CPU - Compute optimized - 72 cores, 144 GB RAM. Fig. 6.13 presents the visual results obtained from the trained model run. Tab. 6.8 presents performance metrics.

For cases where the tumor was not a solid mass, the author observed that the model learned to detect correct tumor shape; however, the mask used for training presents

TABLE 6.8: Results for the U-Net training on a dataset with single tumor extracted.

Fold number	Metrics				
	<i>Loss</i>	<i>Accuracy</i>	<i>MSE</i>	<i>AUC</i>	<i>Mean IoU</i>
1	0.0314	99.5383%	0.0067	0.9918	0.4978
2	0.0420	99.5613%	0.0106	0.9911	0.4978
3	0.0219	99.5424%	0.0046	0.9929	0.4977
4	0.0336	99.5242%	0.0076	0.9932	0.4975
5	0.0088	99.4821%	0.0027	0.9915	0.4974

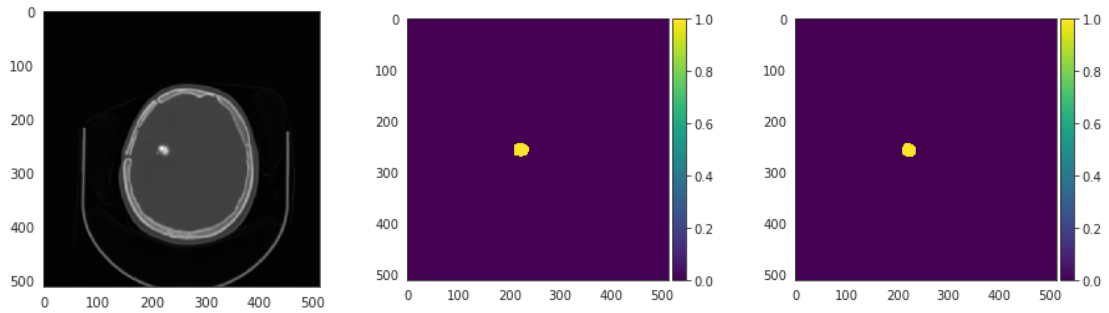


FIGURE 6.13: Image on the left side: Fused CT with PET scans. In the middle - Ground truth. On the right - Predicted mask.

only a single, massive tumor. Fragmented and scattered small tumors are not represented on the mask as is shown in Fig. 6.14, which may lower the performance metrics results, causing the metrics to perform a comparison with a ground truth mask used for training.

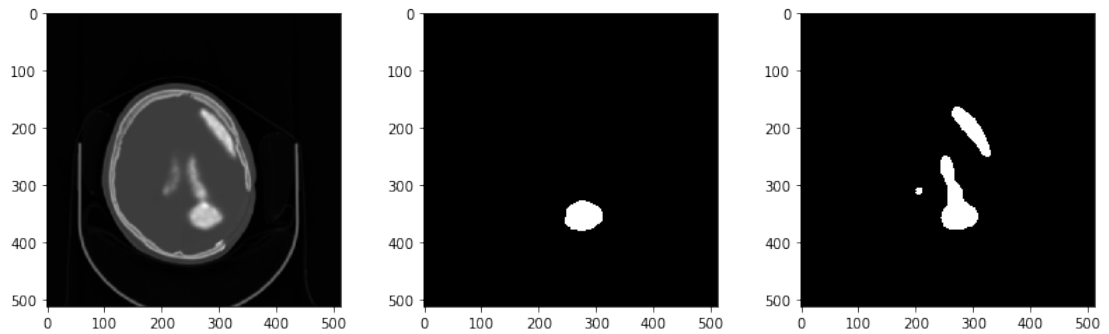


FIGURE 6.14: Image on the left side: Fused CT with PET. In the middle - Ground truth. On the right - Predicted mask.

Multiple Tumors Extracted

Automated mask extraction algorithm was improved to include all SME-labeled and marked tumors, including scattered small tumors. The second objective was repeated

TABLE 6.9: Results for the U-Net training on a dataset with multiple tumors extracted.

Fold number	Metrics				
	<i>Loss</i>	<i>Accuracy</i>	<i>MSE</i>	<i>AUC</i>	<i>Mean IoU</i>
1	0.0190	99.40%	0.0042	0.9173	0.4968
2	0.0134	99.3017%	0.0036	0.9973	0.4963
3	0.0156	99.4529%	0.0040	0.9960	0.4969
4	0.0198	99.42039%	0.0045	0.9940	0.4973
5	0.0190	99.4335%	0.0043	0.9971	0.49671

training for the same parameters of the U-Net model using an updated dataset. The machine used for that experiment was CPU - Memory optimized - 20 cores, 140 GB RAM. Results presented in Fig. 6.15 show obtained visual results of the trained model run. Tab. 6.9 presents performance metrics.

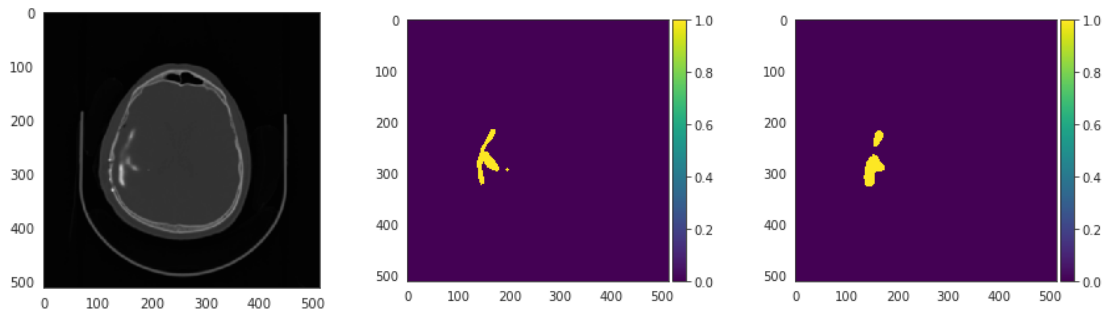


FIGURE 6.15: Image on the left side: Fused CT with PET. In the middle - Ground truth. On the right - Predicted mask for single slice of fused CT-PET where Dice Co-Eff equals to 0.6638.

6.5.3 Training on the Augmented Dataset

Augmentation is a universal method to increase the size of a dataset. The purpose was to initially train models without data augmentation to obtain the best possible metrics without artificially generated images.

Augmentation with AugLy

AugLy [118] is a novel open-source data augmentation library that was developed by researchers and engineers from Facebook AI Lab and was recently published in mid-2021. In comparison to the most commonly used Keras Image Generator, AugLy can: blur, randomly change the brightness, contrast, and saturation of an image, apply a perspective transform so results appear similar to pictures taken from another device, saturate, and sharpen. As tumor detection is firmly based on pixel and voxel intensities, as a reflection of Hounsfield units, the dataset was augmented based on transfer

perspective, blurring, horizontal and vertical flipping, and rotation. The method designed to augment the data set has been published and shared [8] and has been described in the thesis chapter 6.3.1. The data set was augmented twice, with different parameters. Each augmentation doubled the size of the data set. As a final result, the data set contains 2356 fused CT-PET scans and 2356 corresponding masks. Illustrative results of augmentation are presented in Fig. 6.16. The K-fold number was set as 5 and the training was run on 1885 pairs of images and corresponding masks were used for training with 471 for testing. Loading 2356 pairs from a decouple storage into a computer memory took 8m 58s. Two separate augmentations have been performed: one per single tumor extracted, the second for multiple tumors extracted.

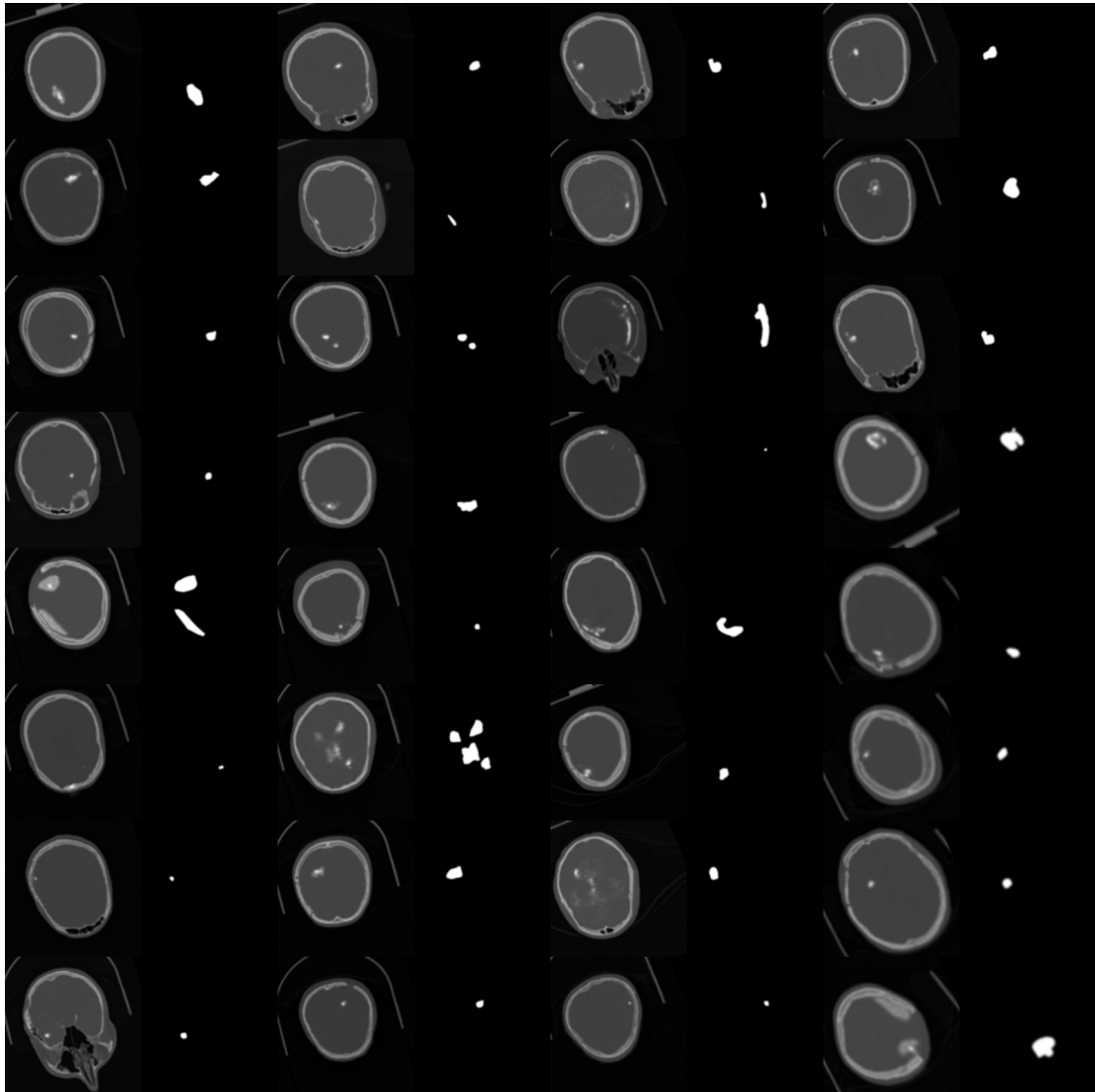


FIGURE 6.16: Augmented fused CT-PET scans with corresponding mask with a usage of AugLy published by the thesis author in [8].

Single Tumor Extracted Augmented

The third objective was to train the U-Net on an augmented dataset of a single tumor extracted. The compute setup was transferred to a machine of 6 cores, 112 GB RAM - with a NVIDIA Tesla V100 GPU card. Due to a lack of sufficient memory, the batch size was changed to 8. Results are presented in tab. 6.10.

TABLE 6.10: Results of U-Net training on an augmented dataset with a single extracted tumor.

Fold number	Metrics				
	<i>Loss</i>	<i>Accuracy</i>	<i>MSE</i>	<i>AUC</i>	<i>Mean IoU</i>
1	0.0076	99.6942%	0.0022	0.9956	0.4949
2	0.0094	99.6407%	0.0026	0.9911	0.4945
3	0.0088	99.6720%	0.0024	0.9910	0.4949
4	0.0087	99.6478%	0.0025	0.9951	0.4950
5	0.0095	99.6332%	0.0026	0.9914	0.4949

Multiple Tumors Extracted Augmented

The fourth objective was to train the U-Net on an augmented dataset of multiple tumors extracted. The computing machine used for this experiment was the same as that for the third experiment. Due to insufficient memory, the batch size was changed to 2. Fig. 6.17 shows the mask predicted for multiple tumors.

TABLE 6.11: Results for the U-Net training on an augmented dataset with multiple tumors extracted.

Fold number	Metrics				
	<i>Loss</i>	<i>Accuracy</i>	<i>MSE</i>	<i>AUC</i>	<i>Mean IoU</i>
1	0.0085	99.6850 %	0.0023	0.9933	0.4926
2	0.0082	99.6976%	0.0022	0.9936	0.4927
3	0.0081	99.6912%	0.0022	0.9953	0.4924
4	0.0085	99.6822%	0.0023	0.9943	0.4922
5	0.0086	99.668%	0.0024	0.9953	0.4922

6.5.4 Results and Discussion

Training for single tumor extracted and multiple tumors extracted was run on CPU machines. As the dataset contained 472 training samples, GPU was unnecessary. Comparable machines were used with one memory optimized and one compute optimized. There were few significant performance differences - for running experiments, the significant differences were between CPU and GPU and capacity of the machines, not optimization by name. Training on the original dataset on the CPU took approximately two hours. Training on the augmented dataset on CPU, without changing batch size,

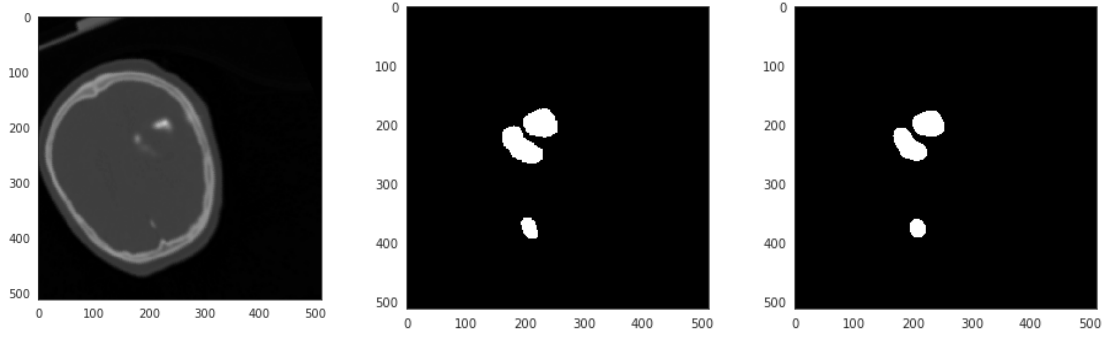


FIGURE 6.17: Image on the left side: Fused CT with PET. In the middle - Ground truth. On the right - Predicted mask for a single slice of fused CT-PET where Dice coefficient equals to 0.8945.

took nearly 56 hours. By comparison, on GPU the average training per single fold took 00:06:50.

TABLE 6.12: Performance Metrics Means for The U-Net. The Best Result per Metric in Bold and Blue.

Metrics (mean \pm std)	Experiment			
	ST^a	$ST Aug^b$	MT^c	$MT Aug^d$
<i>Loss</i>	0.0275 ± 0.0127	0.0174 ± 0.0027	0.0088 ± 0.0008	0.0084 ± 0.0002
<i>Accuracy</i>	$99.53\% \pm 0.03\%$	$99.40\% \pm 0.07\%$	$99.66\% \pm 0.03\%$	$99.68\% \pm 0.01\%$
<i>MSE</i>	0.0064 ± 0.003	0.0041 ± 0.0004	0.0025 ± 0.0002	0.0023 ± 0.0001
<i>ROC-AUC</i>	0.9921 ± 0.0009	0.9961 ± 0.0015	0.9928 ± 0.0023	0.9944 ± 0.0009
<i>Mean IOU</i>	$49.76\% \pm 0.02\%$	$49.68\% \pm 0.04\%$	$49.48\% \pm 0.02\%$	$49.24\% \pm 0.02\%$

a - ST - Single Tumor Extracted

b - ST Aug - Single Tumor Extracted Augmented

c - MT - Multiple Tumors Extracted

d - MT Aug - Multiple Tumors Extracted Augmented

Tab. 6.12 shows that training with a smaller batch size generated higher accuracy and lower loss and MSE (Fig. 6.19), which confirms the thesis from the paper “On Large-Batch Training for Deep Learning: Generalization Gap and Sharp Minima” [119], that with a larger batch, there is a significant degradation in the quality of the model, reflected as lower accuracy.

For the augmented datasets, ROC-AUC and MSE increased, MSE and loss are lower compared to original dataset results. The differences are minuscule, especially when considering the time needed to train a model on 3x larger dataset. Fig. 6.18 shows the longest training required 33 epochs to be taught. For the U-Net 10 to 15 epochs may be enough to obtain an efficient algorithm.

Results were evaluated from the fifth-fold models with a test dataset. The U-Net algorithms return the probability of the pixel being classified as a tumor – setting up a threshold is required. For all test images, with nine chosen threshold levels, Dice coefficient was calculated. Tab. 6.13 presents results.

The higher the threshold, the higher the maximal Dice coefficient. For the final system, a higher value of the mean Co-Eff is more important than the high results obtained on a single slice. Augmenting the dataset significantly increased the mean value of the Dice Co-Eff. Consequently, for a multiple tumor extracted, both the mean and max Dice CoEff values, for all thresholds, are significantly higher.

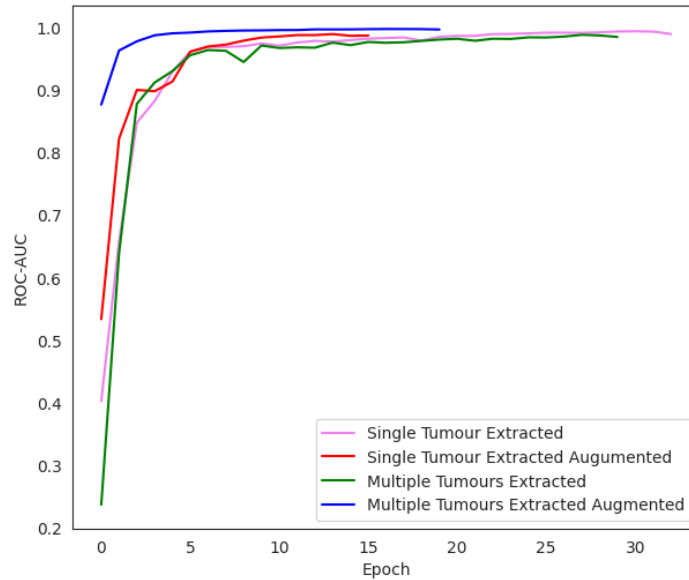


FIGURE 6.18: Area under the ROC Curve (ROC-AUC) value over epochs for the experiments.

Fig. 6.20 shows that for the dataset augmented by a novel library – AugLy – the higher metric (DSC) is – overlapping between Ground Truth and Predicted Mask is higher. With an augmented dataset, the results are better; however, simultaneously, the number of false negative cases, when the network indicated pixels as a background and by the radiologist as a tumor, is higher, and before deploying the model into an operational state, that value should be minimized.

The U-Net models were trained *ab initio*. To simplify the deep learning-based framework for tumor detection and semantic segmentation, the U-Net neural network architecture could be adjusted to be fed with two inputs, separate CT and PET scans, instead of fusion that requires additional preprocessing.

6.6 Multi-stage Pipeline

6.6.1 Prostate Detection based on YOLOv4 and U-Net (YU-Med)

Connecting two models into a single image processing pipeline, YOLOv4 and U-Net, can be an effective strategy for medical image segmentation.

The evaluation of the approach was completed and is awaiting publication as *Semantic Segmentation of the Prostate Based on Onefold and Joint Multimodal Medical Images Using YOLOv4 and U-Net* paper. As the methods evaluation is important for this thesis to provide a complete picture, the author has chosen to summarize the main results

TABLE 6.13: Dice Co-Eff Results. Per Experiment the Highest Mean Value in Bold and Blue, the Highest Max Value in Bold and Teal.

Threshold	Dice Co-Eff	Experiment			
		ST^a	$ST Aug^b$	MT^c	$MT Aug^d$
0.36	max	0.9630	0.9651	0.9051	0.9706
	mean	0.6931	0.7617	0.6079	0.8750
0.38	max	0.9637	0.9600	0.9095	0.9700
	mean	0.6939	0.7620	0.6191	0.8750
0.4	max	0.9582	0.9551	0.9210	0.9696
	mean	0.6939	0.7609	0.6304	0.8746
0.41	max	0.9629	0.9545	0.9239	0.9689
	mean	0.6923	0.7609	0.6356	0.8743
0.42	max	0.9656	0.9541	0.9281	0.9691
	mean	0.6909	0.7608	0.6407	0.8740
0.43	max	0.9624	0.9537	0.9325	0.9690
	mean	0.6890	0.7603	0.6457	0.8736
0.44	max	0.9657	0.9521	0.935	0.9709
	mean	0.6857	0.7597	0.6509	0.8732
0.46	max	0.9571	0.9539	0.9395	0.9719
	mean	0.6748	0.7584	0.6605	0.8721
0.48	max	0.9388	0.9550	0.9459	0.9713
	mean	0.6507	0.7566	0.6699	0.8707

a - Single Tumor Extracted*b* - Single Tumor Extracted Augmented*c* - Multiple Tumors Extracted*d* - Multiple Tumors Extracted Augmented

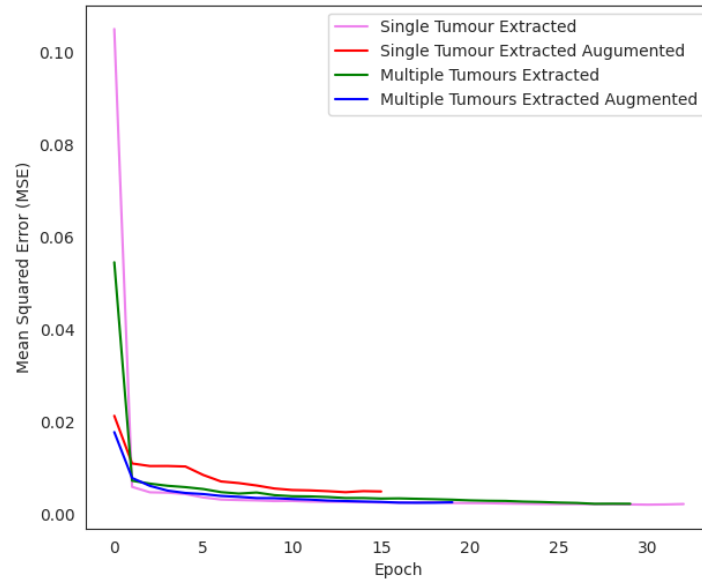


FIGURE 6.19: Mean Squared Error (MSE) value over epochs for the experiments.

here. The developed methods utilize YOLOv4 combined with U-Net for automatic prostate segmentation as presented in Fig. 6.21. The models were trained ab initio on onefold (MRI, the dataset contains 666 training and 73 validation images) and joint multimodal images (MRI mixed and CT- 739 MR and 751 CT mixed images, where the validation dataset totals 149 images). The experiments were based on data from 120 patients who had undergone MRI and CT examinations collected by the Lower Silesian Oncology Center in Wrocław. All the CTs were the size of 512x512 pixels, while MR images varied from 312x224 to 512x608. MRI were both T1 and T2-weighted. The results of the validation dataset showed that the average value of mAP for training on only MR scans was 92.3% compared to 86.9% mAP for the second model (MRI with CT dataset). For the MRI model, it was 3400 iterations, where mAP reached 95.20% and the loss function was 0.110. In the second case containing CT and MRI, the optimal learning time was established in 3300 iterations, mAP was 89.90%, and the loss function was 0.145. The AUC (Fig. 6.22) for the MRI with CT model is 0.944, and the only MRI is 0.907. After analyzing the results, it can be observed that in most cases better results were achieved by a model trained with a mixed MRI with CT dataset. These results confirm previous indications. This model is also more stable on mAP metrics while simultaneously ensuring correct results for IoU threshold values in the range 0.5-0.69. The best image detection was reached for MRI with CT model with a confidence threshold of 0.1 and avg IoU 68.30%, suggesting that the IoU threshold at level is equal to 0.6.

U-Net has been used for semantic segmentation. As the Adam optimizer algorithm was used with a learning rate of $1e-4$, the loss function was defined as binary cross-entropy and monitored as the primary metrics. The batch size was set as 2. All four training sessions were performed with 5-fold cross-validation. The experiments were executed on a machine with six cores, 56GB RAM and NVIDIA Tesla K80 GPU. In the

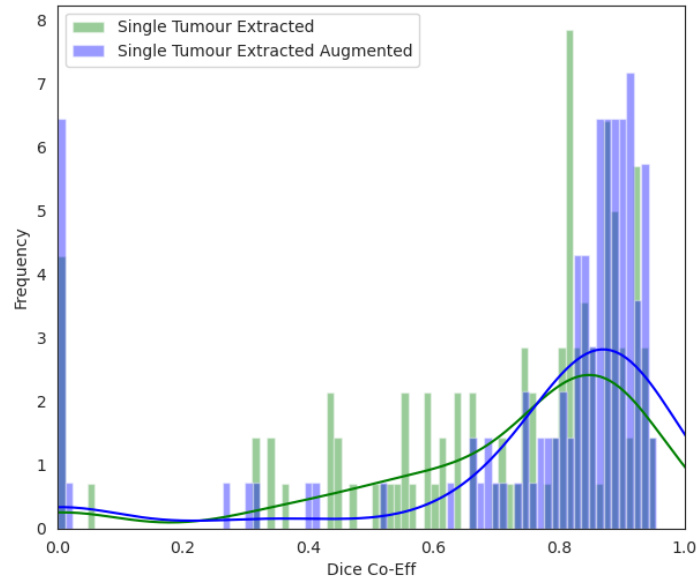


FIGURE 6.20: Distribution of Dice coefficient for two experiments: Single Tumor Extracted and Single Tumor Extracted Augmented.

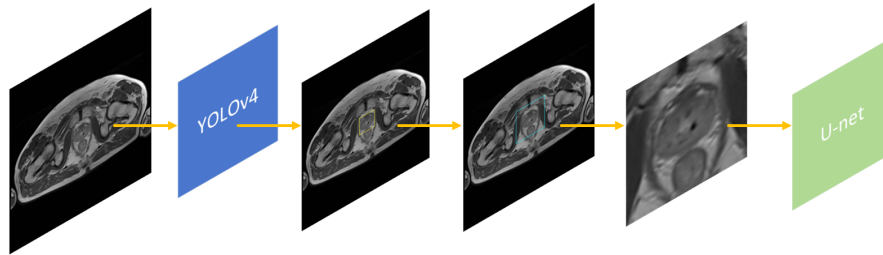


FIGURE 6.21: Post-processing workflow for prostate detection image. Legend starting from the left: input image in $N \times N$ size, YOLOv4 convolutional neural network, prostate detection output with yellow bounding-box, prostate detection output with enlarged turquoise bounding-box (15px padding), cropped prostate detection output and resized into 256×256 px as U-Net input, U-Net convolutional neural network to segment prostate.

second stage, to determine how data augmentation by transformations and enlarging dataset by adding more images of the different type (CT scans) influences training results, the following were executed:

1. MRI segmentation: the dataset of 739 scans.
2. MRI segmentation with augmented data: the dataset of 739 MRI scans with augmentation applied using Keras's ImageDataGenerator class which doubled the size of the dataset resulting in 1478 MRI.
3. MRI mixed with CT segmentation: the dataset of 739 MRI and 751 CT scans.

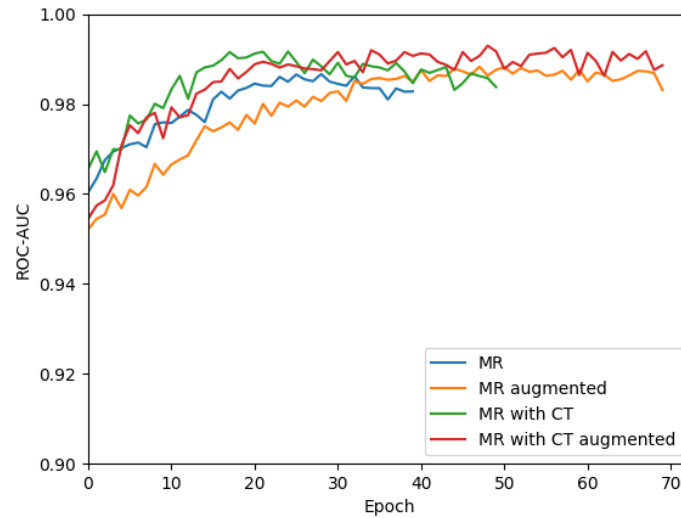


FIGURE 6.22: The values of the ROC-AUC per epoch for each experiment.

4. MRI mixed with CT segmentation with augmented data: the dataset of 739 MRI and 751 CT was augmented using Keras's ImageDataGenerator class doubling the size of the dataset resulting in 2980 images.

The training run achieved a mean speed of around 900 ms/step. The dataset for evaluation consisted of 108 images of only the MR type. Models from the best epochs were chosen for evaluation: 30th epoch for MR, 52nd epoch for augmented MR, 23rd epoch for MR with CT and 58th epoch for augmented MR with CT. Augmentation of the MR dataset led to noticeable improvement for loss, MSE and ROC-AUC. The most accurate model resulted in training in joint (mixed) modalities - MR with CT Augmented. The best model achieved the value of 0.9685 of Sørensen–Dice coefficient for the threshold value of 0.6.

That result may significantly impact the dialogue on the adoption of deep-learning based methods and applications in healthcare and remove one of the barriers defined as operational costs. Deploying and maintaining deep learning models requires a specific infrastructure, which impacts costs. The number of models increases costs linearly. Having a model that performs efficient analytics on two modalities without losing accuracy removes the financial constraint. Furthermore, the model learns on CT and can effectively identify the prostate on MRI, which is a less invasive medical examination. Successful models trained in multimodal medical images reduce the need for CT scans and the risk of radiation-induced prostate cancer.

6.6.2 Brain Tumor Segmentation based on Mask R-CNN and U-Net (MU-Med)

Connecting Mask R-CNN and U-Net for medical image segmentation has several advantages and disadvantages. One major advantage is that Mask R-CNN is capable

of generating object detection masks with high precision, while U-Net is effective in segmenting the interior of the detected objects. By combining these two models, it is possible to achieve accurate segmentation of both the object boundaries and the interior regions. Furthermore, the combination of these models allows for more efficient use of computational resources, as U-Net can be applied to the regions of interest detected by Mask R-CNN rather than the entire image.

However, one disadvantage of combining these two models, Mask R-CNN with U-Net, is that the computational complexity increases significantly, which can result in longer training and inference times. Moreover, it can be challenging to determine the optimal configuration for integrating the two models, including the selection of hyperparameters and loss functions. Additionally, while Mask R-CNN and U-Net have been widely used in medical image segmentation, their performance may vary depending on the specific imaging modality and pathology, and they may require substantial tuning to achieve optimal results.

The process described in 6.2.1 continues here. Fusion with the VGG19-BN model will produce smaller contrast results. Further, experiments utilized VGG19 results.

Tumor Detection by Mask R-CNN

The Mask R-CNN theory was described in chapter 4.2.3. Pretrained models contain the biases and weights which represent the features on which the dataset was trained. Learned features are able to be transferred to different data. The pretrained model was used on the COCO dataset [120]. The initial training was implemented (except for the initial layer which was frozen). The selected parameters were five epoch and a learning rate of 0,001. After the first run, a second training of all layers with a duration of ten epoch and a learning rate of 0,0001 was completed. The primary stage was recreation of the model in inference mode and load trained weights. Random testing followed, resulting in the removal of 10 images not used during training. The model with the best achievement has been published by the author on GitHub [121]. One the previously discussed infrastructures, tumor detection takes approximately 10s. Cropping is achieved in less than 1s. Table 6.15 presents the results.

Mask R-CNN can be successfully applied to detected tumors. This model gives accurate results and detects tumor in all slides where the tumor was clearly present. Although it should be noted that the sensitivity of the trained model is excessively high, notably for the slices where the analyzed tumor, glioma, cannot exist. Tab. 6.14 presents the results of the working model which are unacceptable. The scan results in the second (2nd) row present the algorithm behavior when two boxes are drawn - they should not be nested.

In order to solve this conundrum, one should only analyze the slices on which the tumor is visible. The tumor is not visible in slices from 0-60 and 110-148, based on statistical experiments. Exclusive processing of the scans where the tumor can be located not only reduces over-detection (Fig. 6.23) while additionally drastically improving performance by reducing the computational time by nearly 2/3. For the next step of the proposed method - segmentation, the final algorithm cropped a 64x64 sized image as the input.

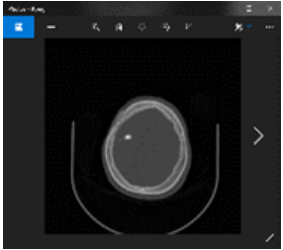
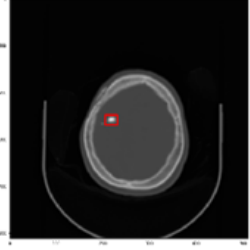
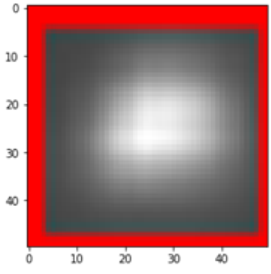
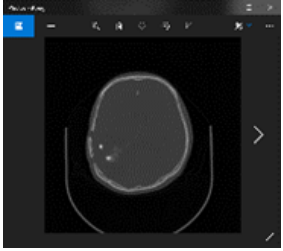
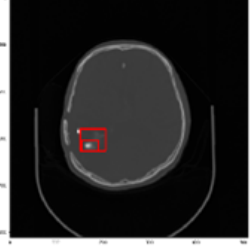
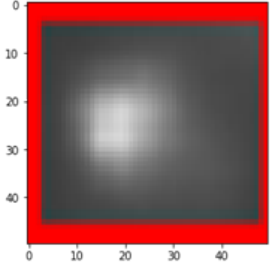
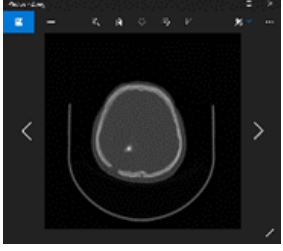
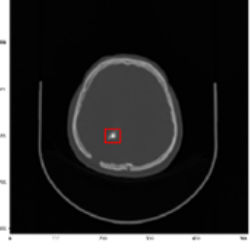
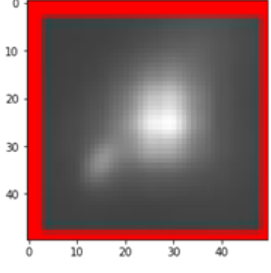
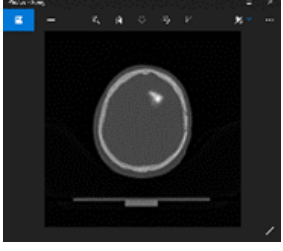
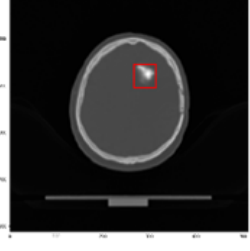
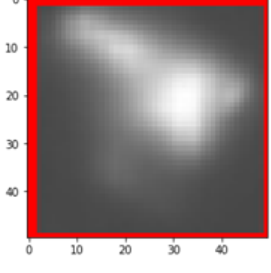

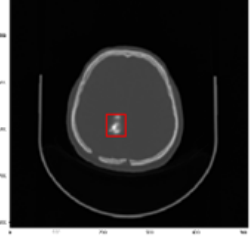
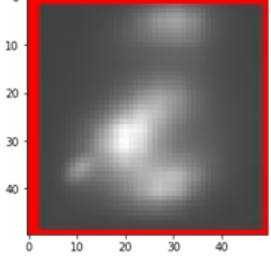
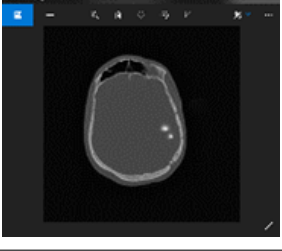
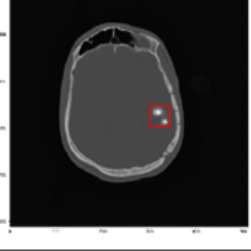
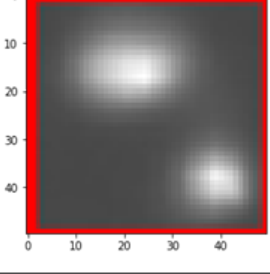
Source image	Detection	Cropped
		
		
		
		
		
		

TABLE 6.14: Source images and detected and cropped tumors for six patients.

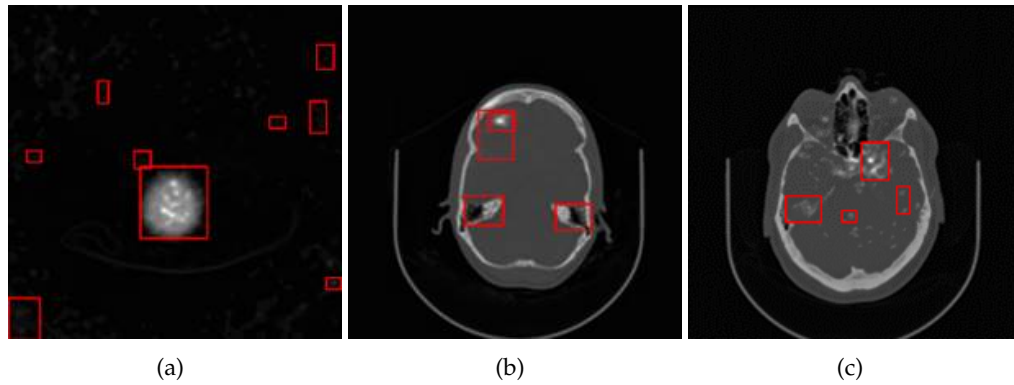


FIGURE 6.23: Over-detection on slices where glioma cannot be present due to its nature.

Tumor Segmentation by U-Net

The U-Net model which was implemented in Keras has been adjusted in order to run on smaller images than in the original design - 64x64 resolution images based on the novel equation which was introduced by the author in chapter 6.4. The implementation of the adjusted model architecture is available on GitHub [116]. The model was trained with a prepared dataset using manually masked tumor areas. The number for the steps per epoch was calculated as the number of trained images divided by a batch size set to 16. To be able to feed the deep learning neural network, data augmentation was used to increase variances in the dataset. Rotations, shear intensity, zoom, channel shifts and horizontal flops, among other techniques were used in augmenting data samples. Mask is represented as a 64x64 output from the network. Sigmoid activation function, mask pixels are in $[0,1]$ range. Loss function for the training is a simple binary cross entropy. The model achieved an accuracy of nearly 97% on the validation dataset upon training for 10 epochs.

Tumor area segmentation for a single patient (about 50 slices) takes around 15s on the same infrastructure as fusion was completed. Especially during training, the importance of infrastructure choice increased. The main inconveniences of CNNs are both the enormous amount of computational power and the amount of time necessary to train the networks. The must-have requirement for training is GPU. For inference mode it was concluded that CPUs work well and validating FPGA is necessary in order to potentially accelerate calculations.

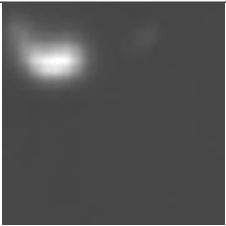
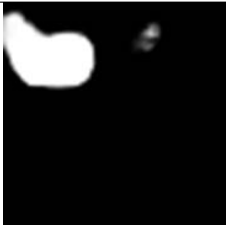
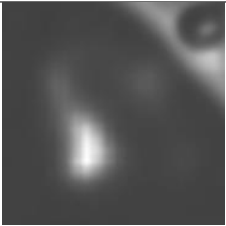

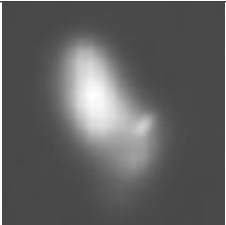
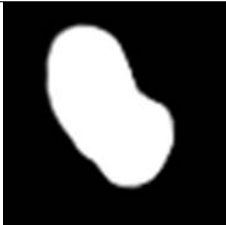
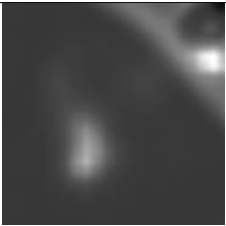
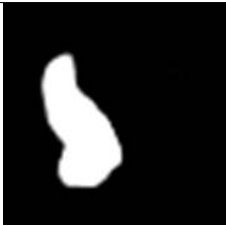
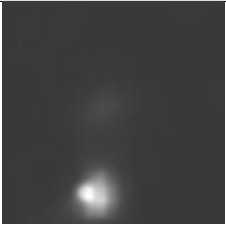
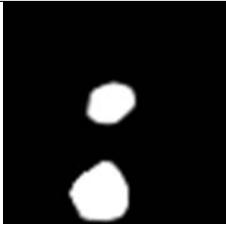
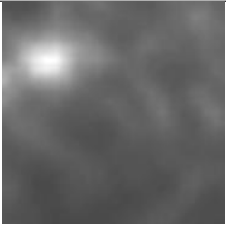
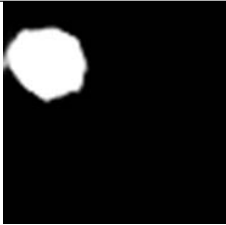
Source image	Segmented area
	
	
	
	
	
	

TABLE 6.15: Source image and segmented tumor area achieved by method based on U-Net.

Obtained results are satisfactory. Loading pre-trained weights significantly increases accuracy and reduces the time for model training. The author presents a framework for

tumor detection and segmentation on medical images. When dealing with solid structure tumors, the frames for operations involving object detection and segmentation tasks using grayscale images can be applied to other use cases such as detection of tumors from MRI images. The size of any given dataset which will feed and validate the CNNs presents the primary challenge of applying deep learning methods in medicine. The work combines CT and PET scans, fusing the images. Obtained results are similar to MRI images which gives hope that methods and models trained on MRI images, due to the transfer learning technique, can be successfully run on the fusion of CT and PET. Experimental confirmation is necessary to determine the mutuality of models trained on fused images ability to work on MRI. Compared to the identical task using RGB images, object detection on grayscale images is of greater difficulty. Mask R-CNN model is recommended while applying the presented framework to object detection in general. The segmented step can be left out when volume and area computation is unnecessary. One should consider using instance segmentation for the separate analysis of small tumors for tumor segmentation, e.g. if the primary research focus was on the main body of a tumor. Glioma can consist of a main tumor with lesser tumors in the vicinity. Using only U-Net, the objective of tumor detection and segmentation can be accomplished in a single step. Setting up training presents a challenge as a heterogeneous dataset of different shapes of glioma should be prepared. While completing that task, standards which should be applied to include local thresholding and processing of slices only when the global threshold for tumor is met. The subsequent development for these methods is following the created framework and evaluation of the performance of the proposed image segmentation algorithm using Dice coefficient. Evaluation of the proposed algorithm on MR scans with brain tumors and prostate should proceed. There are clear benefits to work on 3D tumor visualization inside a skull having color-coded tumors. Consideration has been given to releasing the algorithm as a webapp for initial use by radiologists in cooperation with the research team. Over the previous years, due to an increase of detected cancers and the SARS-COV-2 global pandemic, there is a heightened human awareness of the reliance on the healthcare sector with an overall desire for higher, faster and more accurate standards of care. However, the interpretations of medical data can only be made by medical experts. Their number is not limitless; therefore, they are in high demand in every hospital or radiological center. It can be confidently attested that all automatic systems which utilize modern deep learning techniques will be the requisite acquisition. Radiologists would be able to use the results of the model for tumor detection during screening tests thereby decreasing the time necessary while increasing the accuracy.

Update of the MeDAPR Framework

Based upon the results from the experiments, the author was led to propose a comprehensive, universal, and complex framework for two parts of the dose control process based upon medical images - tumor detection and tumor area segmentation. This framework addresses the main issues when working with medical data and applying modern deep learning techniques for automated tumor detection and segmentation. The CADx and CADe systems currently in the healthcare industry can be shaped by applying the framework to facilitate diagnosis and assist medical professionals. CADx/-CADe objectives are covered with the usage of classification, detection, and prediction

while image processing tasks are covered with image segmentation, registration, and generation [101].

The framework formed the implementation of the methods to detect glioma tumor from CT and PET scans developed and presented by the thesis author in “Brain Tumour Detection and Segmentation Using Deep Learning Methods” [122].

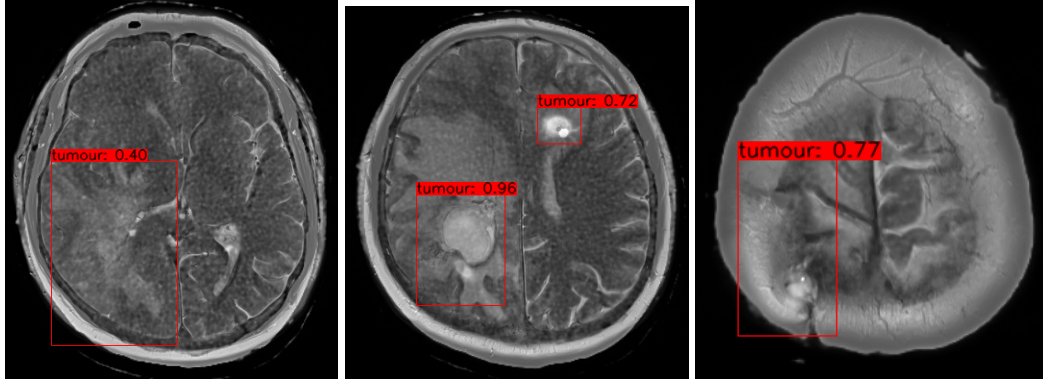


FIGURE 6.24: Glioma detection based on YOLOv4 trained on fused CT-PET-MR scans.

The methodology initiates with the pre-processing of image operations. The Python environment was utilized for the implementation. It should be noted that certain Python libraries do not provide support for image transformation in DICOM format, which is considered the preferred and most effective option for medical data. Therefore, there is a requirement for the conversion of DICOM to the PNG/JPG format. The DICOM standard encompasses fundamental data that is utilized in subsequent analytical procedures. The computation of volume necessitates the inclusion of a pixel spacing attribute. Preserving the primary files or transferring DICOM parameters to a database can aid in the advancement of subsequent research. In order to perform fusion, it is necessary to standardize the dimensions of the images by rescaling them to a uniform size. The fusion of images can be carried out through various methods, which are discussed in the following discourse. The YOLO network was examined for the purpose of tumor detection, specifically in the related work section, which contains multiple citations of this deep learning model. The outcomes obtained for the YOLOv4 architecture, which was trained on the integrated CT, PET, and MR images using the techniques expounded in chapter 6.2.2, are illustrated in Fig. 6.24. Mask R-CNN is a viable alternative to YOLO in the realm of object detection. This model has the added advantage of being capable of object instance segmentation, which can prove advantageous in subsequent research endeavors. The U-Net model was chosen for the purpose of tumor segmentation due to its specialized application in biomedical imaging and demonstrated efficacy in prior literature. The computation of volume relies on the square of the tumor present on each slice and the pixel spacing, which refers to the physical distance between the center of each pixel in the patient. The radiologist considers visualization of the tumor as a crucial component.

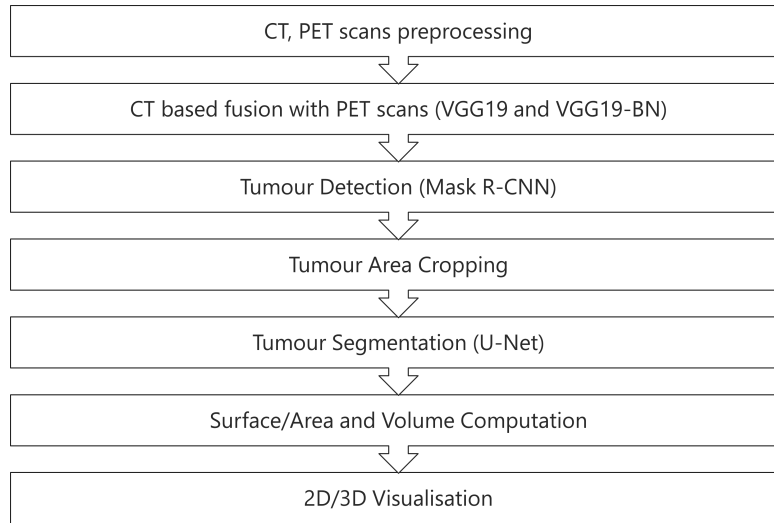


FIGURE 6.25: Key components of the proposed framework.

6.6.3 Mask R-CNN based Hyperparameters Tuning Procedure

U-Net is primarily known from the usage in medical image processing to segment tumors, organs, cells, and outputs a binary segmentation mask as it provides classes for image pixels. What needs to be pointed out, is that U-Net does not perform object localization as it focuses on pixel level segmentation. In case of multiple tumors or metastasis present on the scan, U-Net is unable to distinguish and segment individual objects of the same (glioma) class when in close proximity to each other. The U-Net model achieves high accuracy in the case of instance segmentation; however, in the case of semantic segmentation and especially panoptic segmentation, when the objective is to count the tumors and potential metastases, count their square and volume based on the detected masks, the Mask R-CNN outperform it as it includes a region proposal network (RPF) that generate bounding boxes proposals for potential objects, within the same classes, what is fundamental for any CADx that will include also tracking tumor progression.

Mask R-CNN hyperparameters tuning involves adjusting the values of the hyperparameters to optimize the performance of the model, such as: learning rate, batch size, number of epochs, and optimizer. The backbone of the network cannot be changed or tuned during the training, but the hyperparameter tuning should be considered on a broader level: considering the backbone, but also the hardware available, as well as hardware type. When tuning hyperparameters, the author considered the trade-off between accuracy and speed. Increasing the learning rate led to faster training, but the model was quickly overfitted. Increasing the batch size led to faster training but resulted in decreased accuracy.

The chapter discusses a novel, fully automated, multi-stage deep learning-based approach to brain tumor detection and segmentation that operates on fused multi-modal images. The IFCNN based fusion algorithms extended by the original approach described in chapter 6.2.2 toward detailed information extraction from fused CT and PET scans and MRI to integrate relevant aspects into the final CT-PET-MRI fusion slice.

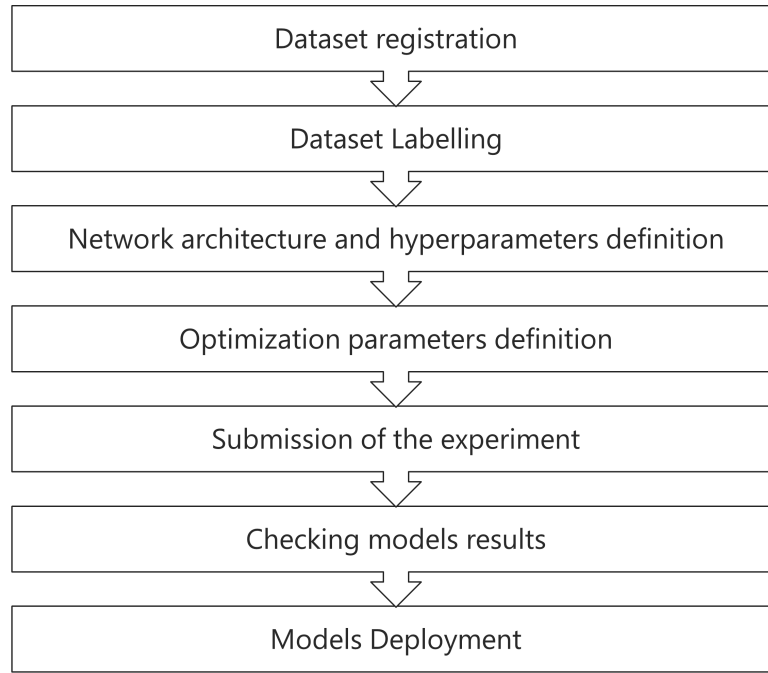


FIGURE 6.26: The specific stages reflecting the cooperation between SME (e.g., radiologists) who provide the labelled data, and the continuous retraining of the model performed by the author.

Regarding brain tumor segmentation tasks, an innovative solution toward data blending has been verified utilizing the Mask R-CNN deep learning model and three different backbones (ResNet18, ResNet50, ResNet152). One hundred and nine (109) different training configurations were created and executed to identify the most optimal hyperparameters in building the model with the highest accuracy for four datasets: CT, CT-PET, MRI, and CT-PET-MRI. The results obtained with an F1 score greater than 0.9 demonstrate the utility of created solutions and vest utilization to improve clinical diagnosis in computer-aided diagnosis (CADx) applications. Fig. 6.26 presents the specific stages of the pipeline which objective is to detect brain tumor and segment the glial tissues.

Datasets and Preprocessing

The author used four datasets: two unprocessed CT and MRI scans datasets and two results of the fusion of CT-PET and CT-PET-MRI. CT and PET series consists of 148 slices, and for MRI, the number of slices varies from 30 to 100. Data labelling of each dataset was done separately. For further processing, labelled glial tissues have been exported both as COCO and separate masks.

Model Training and Transfer Learning

For training purposes, datasets have been randomly divided into 80% for training and 20% for validation. The parameter batch size was set to 2 for all the experiments. Masters and Luschi confirmed that using small batch size achieves the best training stability

and generalization performance [123]. The best results have been obtained in all tested cases with batch size defined as 32 or less. The most common batch size parameter is set to 2 or 4. The evaluation was done after each epoch on a validation set. The Transfer Learning approach has been utilized for each selected neural network model. Models and corresponding backbones have been previously trained on the COCO dataset. Data augmentation techniques have been applied to prevent overfitting and increase the model's generalization skill, i.e. crop around bounding boxes, resizing, expanding and zooming, horizontal flip, and normalization.

Hyperparameter Optimization and Results

Hyperparameter optimization has been adapted to get the most efficient training configuration. Sweeping over the hyperparameters, i.e. different backbones, and training using various epochs, optimizer, and learning rate, has been done. The primary model is Mask R-CNN – an extension of Faster R-CNN. There are two outputs for each candidate object in Faster R-CNN – bounding box parameters and class labels. In addition to these, Mask R-CNN extends that and adds a third output – an FCN that produces a binary segmentation mask for each RoI [124]. The hyperparameters used were:

1. Models with backbones: Mask R-CNN ResNet18+FPN, Mask R-CNN ResNet50+FPN, Mask R-CNN ResNet152+FPN.
2. Optimizers: Stochastic Gradient Descent (SGD), Adam, AdamW (Adam optimizer with improved implementation of Weight decay).
3. Epochs: 15, 30, 50.

Model training was begun with different learning rates, but the loss was exploding for SGD optimizer for learning rate smaller than $1e^{-4}$. Then the attempt was made to have the identical learning rate for each training configuration equal to $1e^{-4}$. The first results showed that for SGD, that value was still not large enough (loss was slowly decreasing, and results were inaccurate). For Adam and AdamW, the value was too high (the loss too quickly stabilized). Final experiments were rerun with a new learning rate: $1e^{-5}$ for Adam and AdamW, and $1e^{-3}$ for SGD.

It was observed that training for less than 15 epochs was inefficient, and above 50 was terminated as the network did not learn more, which allowed setting epoch numbers as 15, 30 and 50 for the final experiment. For every dataset, there were twenty-seven different training combinations determined, which resulted in 108 separate training experiments. In addition, experiments have been parallelized using four different compute clusters with a separate NVIDIA Tesla K80 graphic card and the Intel Xeon E5-2690 v3 (Haswell) processor. For every training Precision, Recall and mean Average Precision (mAP) were logged.

The obtained numeric results are presented in Tab. 6.16. Fig. 6.27 presents the chart with the Mean Average Precision curves for models that obtained the highest mAP value in the last epoch per type of dataset.

Results in Table 6.16 show that for specific experiment configurations, the results for fusion are equal or almost equally good as for different data types. Fig. 6.28 presents training charts for each configuration colored in blue in the results table.

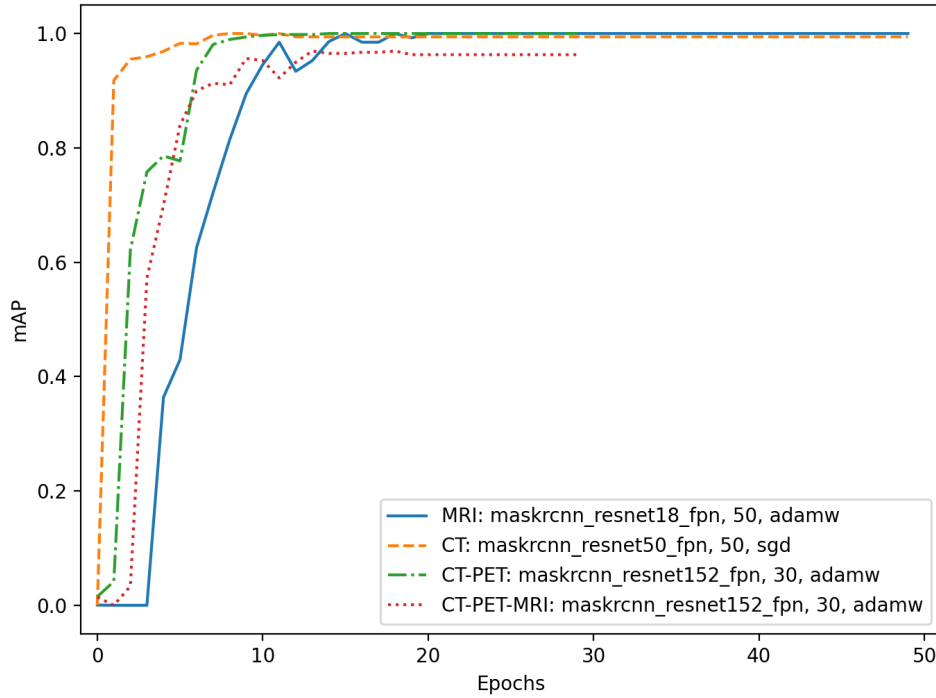


FIGURE 6.27: Models with the highest mAP value. The value of mAP began to stabilize around 0.9 after the 20th epoch.

While the number of epochs may initially appear small, the number is proportional to the transfer learning method and thus can be significantly smaller than during a full training. This approach was exemplified by the winners of RSNA-MICCAI Brain Tumor Radiogenomic Classification organized by the Radiological Society of North America aimed to predict the status of a genetic biomarker important for brain cancer treatment and used only 15 epochs [125].

Fig. 6.29 presents results obtained post implementation and used by a radiologist which confirmed the applicability and usability of the methodology via observable and identifiable structures in reviewed T2 related images.

For the configuration that achieved the highest F1-score for the final fusion of CT-PET-MRI dataset, additional evaluation metrics have been calculated as: AP@0.5 (AP at IoU=0.50) = 0.9347, and AP@0.75 (AP at IoU=0.75) = 0.8488. The results proved that the configuration based on Mask R-CNN with the backbone ResNet152+FPN trained during 50 epochs with the optimizer SGD gave the highest F1-score of 0.8649. Therefore, an additional training configuration has been created with two changes: training during 300 epochs and learning rate set as 0.00529. As a result, obtained Precision was 0.9375, Recall was 0.88235 and the F1-score was equal to 0.9091. mAP for that experiment was 0.93772. It took about 48 milliseconds (NVIDIA Tesla K80 graphic card and the Intel Xeon E5-2690 v3 (Haswell) processor) to segment the single slice, which gives the FPS value around 21 for the image size 512x512x1. The processing time for a default

TABLE 6.16: F1-score results for Mask R-CNN.

Model and backbone	Epochs	Optimizer	CT	Fusion of CT-PET	MRI	Fusion of CT-PET-MRI
			<i>F1-score</i>	<i>F1-score</i>	<i>F1-score</i>	<i>F1-score</i>
Mask R-CNN with ResNet18+FPN	15	SGD	0.641	0.7273	0.7333	0.5455
		Adam	0.4848	0.6579	0.3729	0.4545
		AdamW	0.6275	0.7632	0.3437	0.3562
	30	SGD	0.8333	0.807	0.7742	0.7027
		Adam	0.7917	0.7742	0.5455	0.7059
		AdamW	0.8333	0.7385	0.65	0.6522
	50	SGD	0.8261	0.7692	0.9032	0.6829
		Adam	0.913	0.7308	0.75	0.7442
		AdamW	0.9412	0.7164	0.6875	0.7879
Mask R-CNN with ResNet50+FPN	15	SGD	0.7018	0.7458	0.88	0.8276
		Adam	0.5843	0.5526	0.8462	0.4348
		AdamW	0.6667	0.6452	0.6486	0.4194
	30	SGD	0.8727	0.8108	0.7879	0.8108
		Adam	0.72	0.7059	0.8889	0.718
		AdamW	0.7692	0.8077	0.8	0.75
	50	SGD	0.9231	0.8276	0.8148	0.8235
		Adam	0.9231	0.7941	0.8966	0.5769
		AdamW	0.7619	0.7187	0.7742	0.7368
Mask R-CNN with ResNet152+FPN	15	SGD	0.8444	0.7273	0.4815	0.8485
		Adam	0.6462	0.7429	0.7368	0.8235
		AdamW	0.4931	0.6571	0.7857	0.4776
	30	SGD	0.9091	0.8148	0.96	0.7805
		Adam	0.7742	0.7671	0.875	0.8333
		AdamW	0.7869	0.9796	0.9032	0.8333
	50	SGD	0.871	0.7385	0.7742	0.8649
		Adam	0.9091	0.8182	1.0	0.7619
		AdamW	0.8235	0.8364	0.7368	0.7778

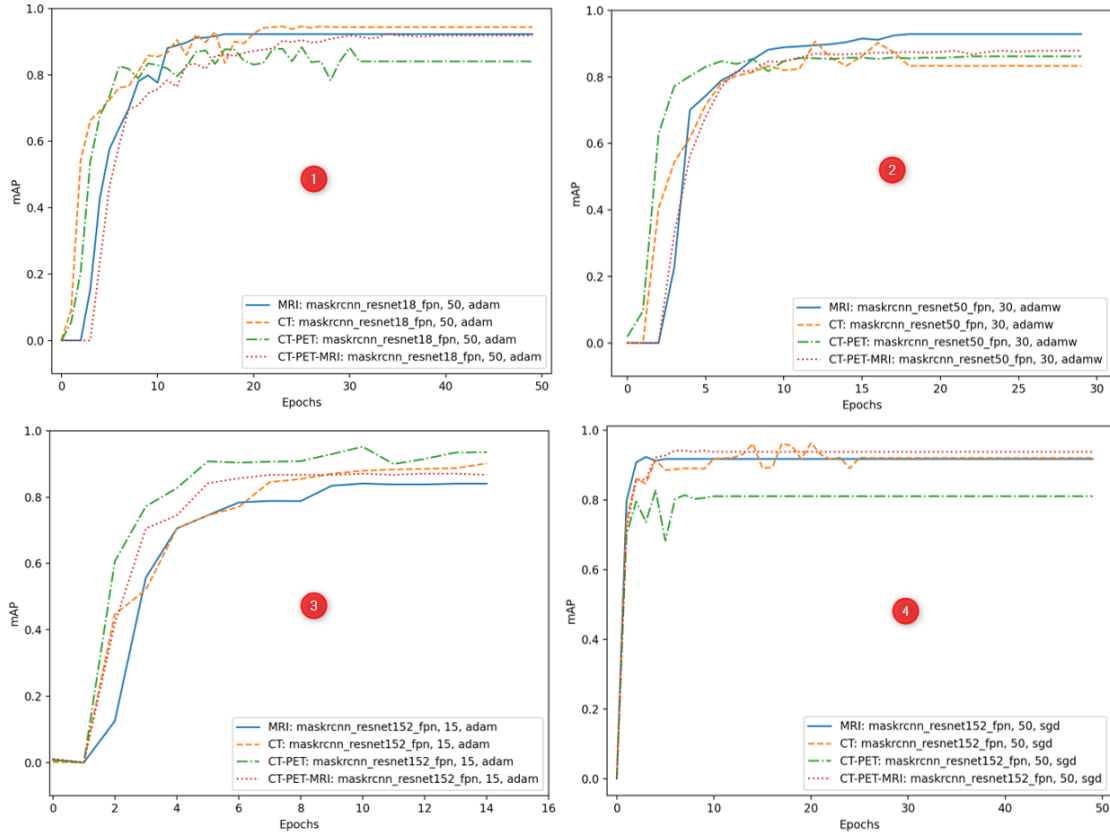


FIGURE 6.28: Training charts for the chosen configuration of hyperparameters (colored in blue at the Table 6.16).

single series for one patient (148 slices) was about 7.104s on a given compute engine. It has the potential to be used in real-time processing as well as for batch scoring as the models can be deployed as the REST API models.

The author performed CT-PET fusion for all 51 photos in the case of the test on the phantom (described in chapter 5.2). Segmentation was subsequently performed using a previously trained model - Mask R-CNN with the backbone ResNet152+FPN trained during 300 epochs. At the CT-PET fusion it was calculated that the volume of all five beads simulating brain tumors was $21.7796cm^3$ and the relative error was approximately 0.0218 or 2.18% which is an excellent result, considering the results came from the fused CT with PET scans.

Subsequently, in order to compare results and calculate the brain volume to the form of the best model, the author revised all studies and images used. A direct comparison between classical methods and CNN-based methods, including the best and most efficient model based on Mask RCNN, cannot be performed. This is due to the fact that the developed methods based on CNN work on much more advanced and detailed images such as CT-PET, MRI, and CT-PET-MRI fusion. In the case of verification of active contour methods on CT-PET, MRI, and CT-PET-MRI fusion studies, the classical methods have been proven to be incapable of segmenting tumors effectively (the phantom test and the classical algorithms were performed on PET-based images

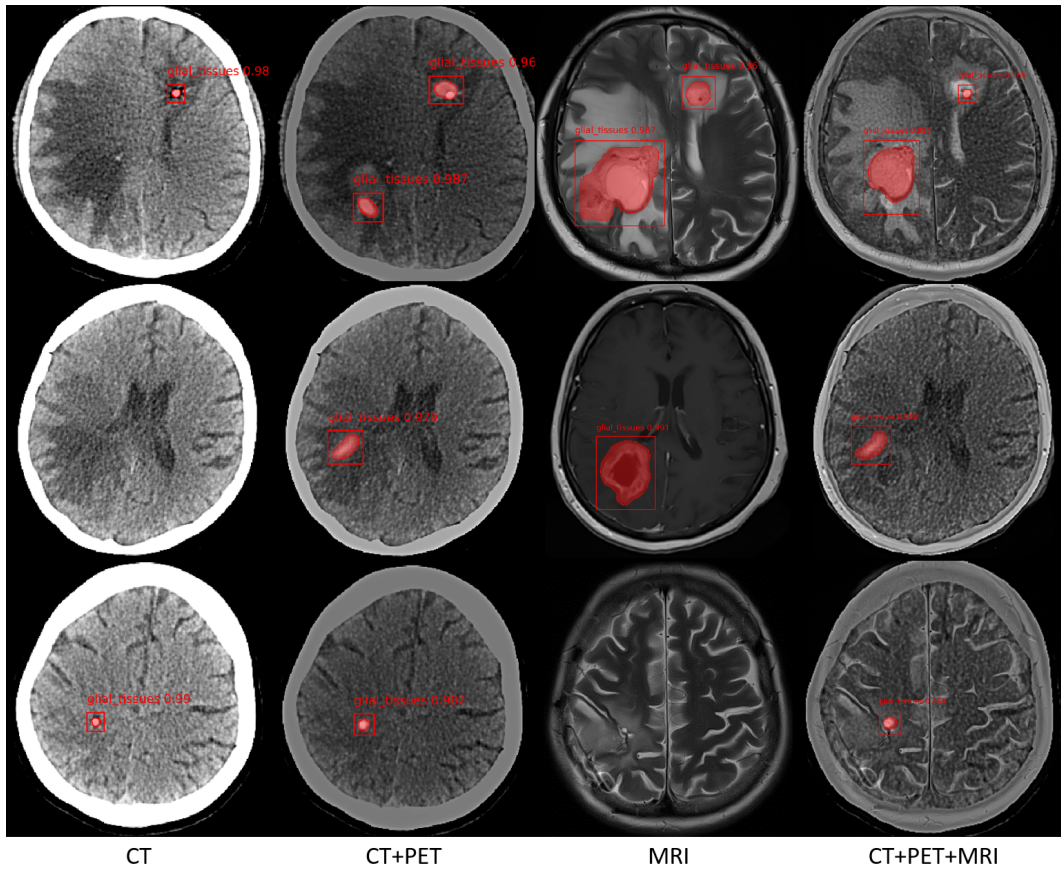


FIGURE 6.29: Segmentation results are presented with bounding boxes, which localize the tumor area.

fused with information from CT, not CT-based as in the case of the CNNs). For these reasons, active contour methods have been found to be applicable, while even though CNN-based methods are generic they are also promising due to their effectiveness.

The author postulates that the effectiveness of the segmentation itself, which in most methods oscillated at above ninety percent, will also significantly reflect the accuracy of calculating the volume. The tumor's volume can be computed by summing the segmented areas (square) and multiplying that by the volume of a single voxel, which is stored in the metadata of the DICOM files. The tumor volume can be easily calculated by possessing the segmented tumor.

Seven (7) patients for whom a CT-PET examination was carried out and volume counting methods were launched utilizing the developed methods which were based on the active contour methods based on a PET-based fused image with CT. Comparison is only possible between a CT examination and fusion CT-PET. The results for individual patient IDs were: 17.75cm^3 , 19.01cm^3 , 31.67cm^3 , 26.82cm^3 , 13.35cm^3 , 11.58cm^3 , 16.88cm^3 , which produces MAPE of 4.73% with the exclusion of patient No. 6, for whom there is no reference value.

Chapter 7

Proposed Novel Architecture for the Healthcare Industry

The medical imaging market is experiencing exponential growth from \$38.4 billion (USD) in 2020 to \$58.7 billion by 2025. Medical centers invest their research in artificial intelligence technology, capitalizing on the extensive imaging data they create through first-party research. The main examples are as follows:

- Johns Hopkins invested in a platform that allows 2,500 clinical researchers within the organization to access clinical data and use them with machine learning algorithms.
- Mass General Brigham invested in the Center for Clinical Data Science, partnering with prominent industry vendors, to develop machine learning algorithms and new tools, mainly for radiology.
- Affidea Group seeks to become a “testing ground” for AI algorithms.
- Stanford Medicine established the Center for Artificial Intelligence in Medicine & Imaging.

Although the foremost progress has been made in the development of algorithms for radiological modalities (CT, PET, XR, MRI and SPECT), medical imaging does not stop there. The field of digital pathology delivers an enormous amount of imaging data that are currently barely covered by current machine learning solutions. Thus, a Whole-Slice Imaging Scanner can process 500–1000 slides/day, potentially resulting in 1–2 terabytes of images per day, with image resolution pushing into gigapixel size. In most cases, digital pathology data is stored in the same image archiving systems as radiology data. Based on experience, investigation, and solution reviews, there are several major problems:

- Extracting and consolidating data. Data is notorious for being disorganized and spread across many disparate systems. Anyone willing to put together relevant datasets struggles with extracting and misidentifying images and correlating PACS radiological imaging data with EHR, LIS, radiation oncology, and other system data. This problem is very closely related to the IT setup of a particular hospital and has to be solved mainly within organizations. However, it is paramount to have a single standard-based way to move data into a single, scalable repository.

- Curating datasets for machine learning workflows. Once the data is extracted, there should be an straightforward way to review an image collection and filter it by condition, acquisition date, modality, and scanner parameters. There must be a way to present a single view in the longitudinal patient record where imaging data are accessible and discoverable alongside non-imaging data. Ultimately, clinical researchers who know the problem should have a way to discover the data, minimizing the need for knowledge of specific storage solutions and the underlying data infrastructure. They need a Search Engine-like interface that allows them to drive data discovery with simple structured natural language queries.
- Labeling datasets for supervised machine learning workflows. Once a dataset is put together, there should also be an straightforward way to label the dataset for subsequent training. Several kinds of labeling are typically done a) whole-image labeling (not unlike radiological reporting); b) region-level labeling for object detection problems; c) whole voxel-level segmentation. There should also be an uncomplicated way to extract and prepare a labeled dataset for subsequent training of ML algorithms. For example, in radiology, the underlying data format (DICOM) is highly redundant and is designed to minimize the chances of data corruption while providing the ability to store data from a diverse set of imaging modalities in the same form. This contradicts the way machine learning algorithms consume data, which is typically presented in the form of very dense, normalized high-dimensional vectors. In addition, curating suitable datasets for machine learning and labeling them requires a joint effort by clinical professionals who possess the necessary image interpretation skills. Furthermore, coordinating this work poses problems in organizing the work of experts who often operate from remote locations on desynchronized schedules. Thus, high-quality labeling tools that are well integrated into the dataset curation pipeline and result in versioned immutable data objects that can track their provenance to the original acquisition are paramount to the success of any annotation operation in the space.
- Executing and tracking machine learning training runs. Once training is done, researchers want to iterate over multiple versions of the model and compare the performance of models as a function of hyperparameter adjustments or dataset modifications. Medical imaging AI has its specifics in that individual data points (i.e. image series) have a high level of meaning, and it is truly valuable to have full data provenance information, i.e. being able to understand where the data that went into the model are coming from, who annotated it, and under what premise.
- Operationalization. Once the model is ready, it must be deployed in a cloud environment for inferencing. Again, when it comes to medical imaging AI, deployment for inferencing means that the model has to be exposed to clinicians in a way that does not interfere with their workflow while at the same time making it clear what output the models provide. In medical imaging, this can be done in multiple ways, to name a few: secondary capture objects, structured reporting, integration at the UI level, and integration into worklists. Additionally, it is essential to keep track of the regulatory status of a deployed model. For example,

suppose that a given regulatory compliance is applied (external or internal), in such a case, there must be mechanisms to satisfy the requirements and provide data about the model's operational characteristics in a unified way.

Currently, there are no tools on the market that provide a unified solution to all these problems. Getting all the above aspects right or even acquiring correctly integrating tools that address the individual areas of this problem space requires nontrivial software engineering skills that are not typically included in the biomedical researcher education path. As a result, the different teams approach this problem set from different angles, depending on their original core strength, but build very similar machine learning pipelines. These engineering teams, such as medical centers or startups, build redundant software systems outside of their core focus and ultimately rely on scalable cloud-based storage and computation resources.

7.1 Cloud-native Proposed Architecture

The author, based on research and industry experience, designed an innovative and complex healthcare analytics platform that connects multiple Open-Source and Microsoft technologies into a single platform that serves as a decision support system for radiologists, general practitioners, oncologists, surgeons, and other medical professionals. Fig. 7.1 presents the diagram for the architecture. The platform positions Microsoft Azure Cloud as the primary choice for processing biomedical imaging.

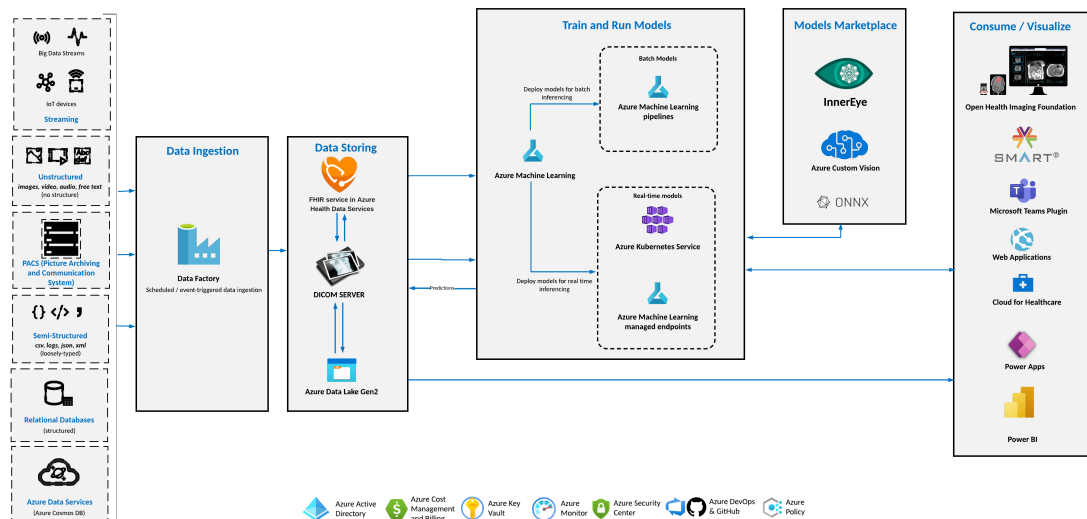


FIGURE 7.1: The architecture diagram of the healthcare analytics platform is provisioned on Azure that consists of Azure Services and Open-Source technologies. The solution based on a similar architecture won Microsoft Global Hackathon 2022 Executive Challenge - Hack for Industry - 1st Place.

The solution bridges the relationship between radiologists and AI developers, as it includes radiologists in AI design. That element of cooperation was also tested during the development of the MAT tool and is described in chapter 6.3.2.

The platform's functional features are as follows:

1. Seamless and secure data ingestion from on-premises data sources to Azure Data Lake via Azure Pipelines available in Azure Data Factory.
2. Persistent storage layer consisting of Azure Data Lake, Azure Fast Healthcare Interoperability Resources (FHIR), and DICOM server.
3. Flexible AI module that allows running different Deep Learning Models that might detect tumors or classify images.
4. A responsive and intuitive web application designed for radiologists to help them make decisions. Features of the WebApp: viewer of original and processed by AI dataset zooming comparing results patient history 2D and 3D view quick 3D rendering from a different view - layers of the body.
5. Ability to bring a personal model – the platform/marketplace is flexible and might be extended by any Machine and Deep Learning model.
6. Ability to extend, modify, and customize - the platform is built to allow developers, third-party vendors, and researchers to add their model datasets and integrate their products into the platform.
7. Ability to anonymize images on ingestion or at egress (for sharing outside), including deep anonymization of head scans via face blurring, and even going further, relying on the concepts of differential privacy.
8. Tracking provenance per patient and institution (facilitating considerations such as GDPR compliance). This can be done through one-way hashes with keys stored at the image owner's locations.
9. Security to the highest standards of Azure cloud, including compliance with ISO, HIPAA, GDPR.
10. Review and labeling tools through web-based zero footprint applications, such as the one developed for the general image domain within the Custom Vision service.
11. Version control of ground truth.
12. Organization of images into collections exposed as datasets to the machine learning models.
13. Export and ingestion in industry-wide formats such as DICOM, NIFTI, HDF5.
14. Role-Based Resources and Data Access Control Per Institution.

7.1.1 Medical Data Ingestion and Storing

The reference scenario is that the hospital or medical institution has data stored in PACS, the centralized computing device that stores medical images for various modalities. To move the data to the DICOM server, a connection with PACS, e.g., Inner Eye Gateway, must be leveraged.

The primary assumption for implementing that scenario and the DICOM Server is that there is a PACS and that the data are structured. However, based on the experience gathered, the data are only on the local PCs of physicians and thousands of CDs. Therefore, there is a need to create a repetitive procedure for moving data from the hospital to Azure Cloud and then structuring the data at the final destination. For data ingestion, the author proposes using Azure Data Factory (ADF), a service designed to integrate data with the on-premises environment, a common scenario for hospitals. Fig. 7.2 presents the created solution. Specifically, a scheduled pipeline is generated to move all data in the defined path. Moreover, integration runtime must be installed on one VM inside the hospital's infrastructure, unblocking communication with the cloud. The data can then be moved to the destination, which is Azure Data Lake Storage (ADLS).

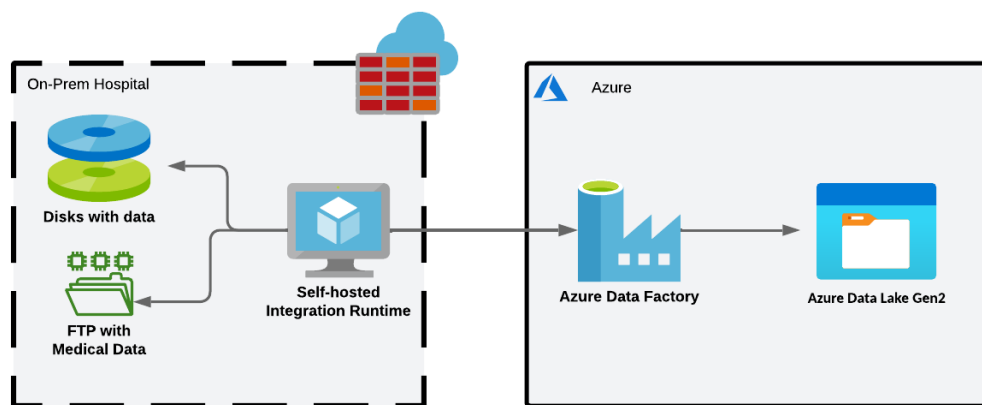


FIGURE 7.2: Authorial solution of using Azure Data Factory (ADF) moves the medical data from on-premises hospital environment to Azure Data Lake Storage, breaking new ground in usage for ADF.

Once the data are moved to ADLS and then structured by another pipeline or script, the data can be transformed into DICOM Server and use Azure Healthcare APIs (FHIR) to store protected patient information.

7.1.2 Training Machine Learning Models for Healthcare Applications and Cloud Deployment

The Model Integration Layer wraps different types of response from different models in the same API. That layer can talk to any model: ONNX, TensorFlow, and PyTorch. The integration layer has been written in Python in the form of batch-scheduled or event-triggered scripts, and it prepares data to be sent to a model and parses the response (predictions), both to be saved and to be visualized in a visualization tool. The leading service for the training and deployment stage is Azure Machine Learning (Azure ML). The main features and pros of using Azure ML for medical image processing:

- Repeatable experiment pipeline, with traceability down to individual images that contributed to the underlying datasets.
- Inferencing platform that captures metrics about operational aspects of medical imaging model, in a format that can contribute to collecting clinical performance evidence, feeding into regulatory validation/certification pipelines.
- Inference platform that is aware of data validation rules and can run protocol adherence rules that check if the data coming in for inference conforms with the capabilities stated by the ML algorithm.
- Privacy-preserving distributed learning that facilitates collaborative training projects across multiple locations.

Fig. 7.1 also illustrates that the model integration layer consists of:

- InnerEye OSS – Open-source tool developed by Microsoft Research Cambridge. InnerEye trains models on medical images and provides models for 3D segmentation and classification using 3D U-Net. The primary requirement is that the data must be in NIFTI format.
- Custom Vision – a customizable service with pre-trained computer vision models for low-code users, allowing one to create models for 2D object detection and classification.
- ONNX – any model in an ONNX open format.
- Any Models Marketplace – such as Nuance Precision Imaging Network, Blackford, IBM or Siemens.

For the deployment stage, the author recommends two main infrastructure services: Azure Kubernetes Services or Azure Container Instances. Both services assume deploying final model as the Docker container, and the difference is that AKS allows one to define and control the Virtual Machine on which the containers are running, where ACI simplifies the process, and there is only control over memory and CPU of a single containers.

7.1.3 Parallel Processing

The dissertation thesis expressed in 1.1 requires the ability to run multiple neural networks of various architectures at the same time.

Proposed Method

The inspiration for introducing a new method proving the dissertation thesis and making it possible, comes from one of the top Apache Foundation projects - Apache Spark [126]. Apache Spark is a distributed computing framework that can process large datasets in parallel across multiple worker nodes. The architecture of Spark is designed to be scalable, allowing for the processing of large datasets in parallel. The Execution Core Program architecture builds on this foundation, allowing for the processing of individual slices of data in parallel across multiple worker pods. This approach can help to improve the efficiency and speed of data processing, particularly when dealing with large datasets. Apache Spark allows to process data in two modes: batch and streaming. For the purpose of this dissertation, only batch mode has been considered.

The sequential data processing algorithm presented on 7.3 processes data in a linear order, one after the other, in a predetermined sequence. In the context of medical imaging, it involves processing all the patient data (n -slices) in a consecutive order, where each slice is processed one at a time. The sequential algorithm is particularly useful when the computation of each slice is independent and parallel computation is not possible or efficient. However, this method can be time-consuming and computationally expensive, particularly when dealing with large amounts of data. Therefore, parallel processing techniques are often preferred in order to optimize processing time and efficiency.

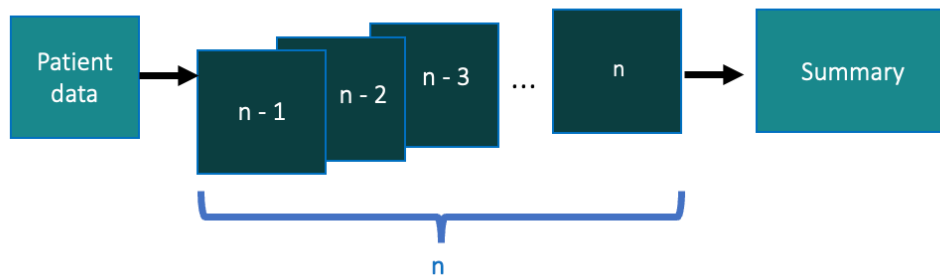


FIGURE 7.3: Sequential medical data processing.

In a concurrent data processing algorithm presented in Fig. 7.4, all patients' data are processed simultaneously by multiple worker-nodes. The data is partitioned into smaller units, and each unit is assigned to a worker-node. Each worker-node processes its assigned data independently and in parallel with other worker-nodes. This approach allows for efficient processing of large amounts of medical data, reducing the time required for analysis. Parallel processing is often implemented using distributed computing frameworks such as Apache Spark, where tasks are executed in parallel across multiple nodes in a cluster.

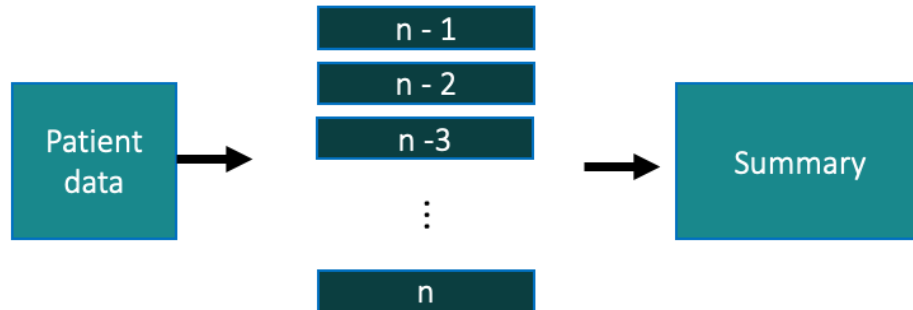


FIGURE 7.4: Concurrent medical data processing.

The proposed solution which combines both Apache Spark and Azure Kubernetes Service (AKS) and is built to process medical data in parallel across multiple worker pods running different models, include:

- The Execution Core Program is responsible for managing the processing of data. The Execution Core Program consists of a driver program and multiple worker pods, which are responsible for processing single slices of data.
- The Driver Program in Spark is responsible for coordinating the execution of tasks across all the worker pods. In the case of the Execution Core Program, the driver program would manage the processing of single slices of data, distributing them to the worker pods.
- The Worker Pods in Spark are responsible for executing the tasks assigned by the driver program. In the case of the Execution Core Program, each worker pod would be capable of processing a single slice of data (Fig. 7.5), with the driver program distributing the slices to each worker pod.

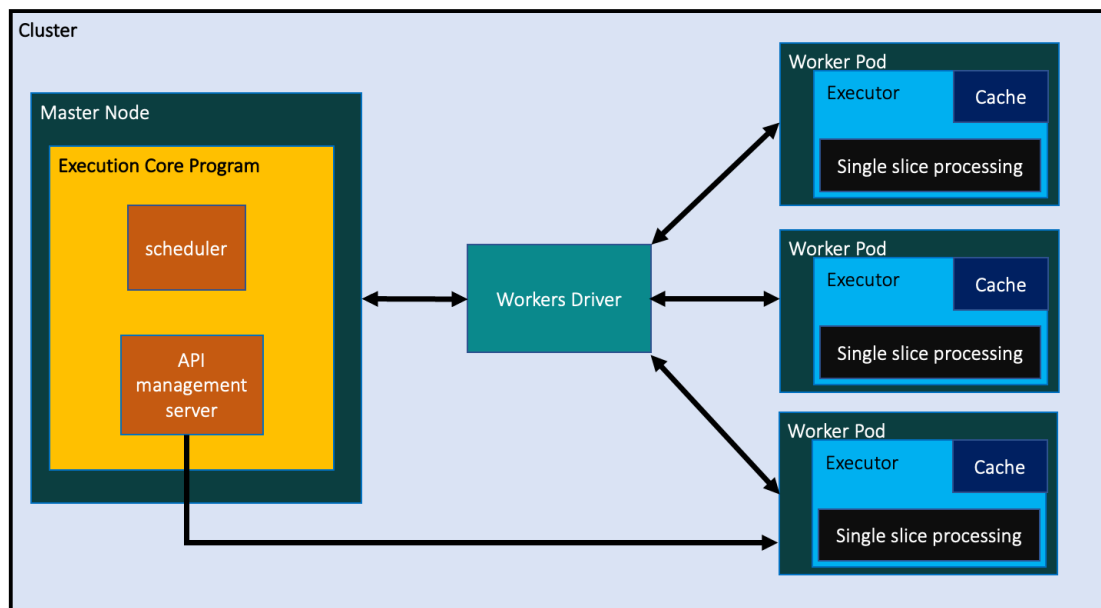


FIGURE 7.5: Proposed architecture based on the containers approach.

There are two approaches - single slice processing, the default option, and partition processing (chunk of data) as presented in Fig. 7.6.

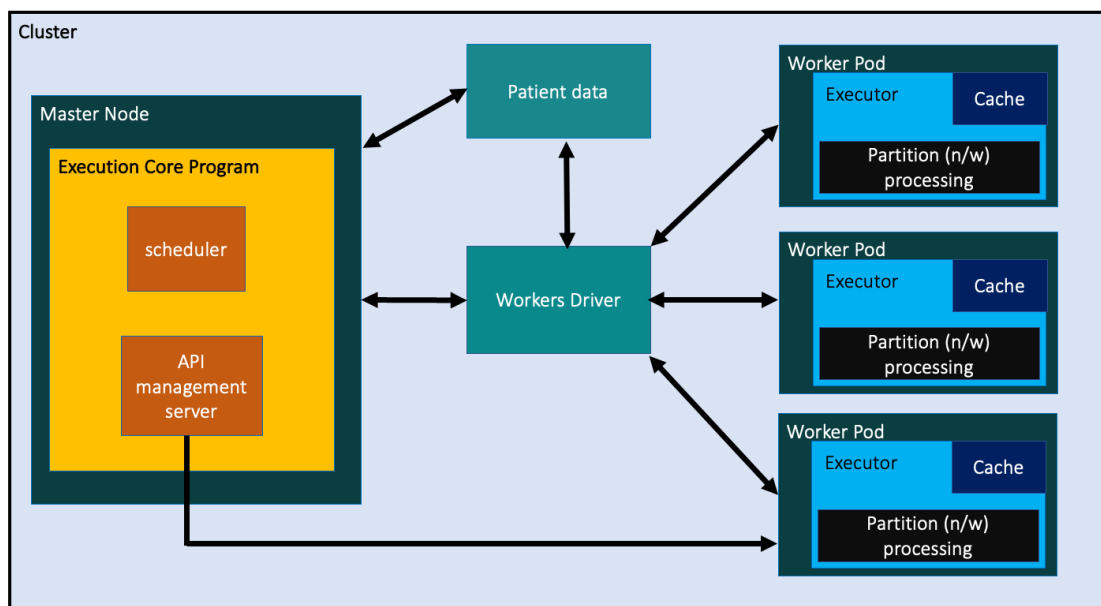


FIGURE 7.6: Proposed architecture based on the containers approach with data partitioning.

Table 7.1 provides the comparison of two types of processing. Single slice executors have low concurrency and resource utilization efficiency but are more fault-tolerant and flexible in managing resources. On the other hand, multiple slices (partitions) executors have high concurrency and resource utilization efficiency, but may have higher

TABLE 7.1: Comparing the two types of executors in the proposed solution.

Feature	Single Slice Executor	Partition (Multiple Slices) Executors
Concurrency	Low	High
Resource utilization efficiency	Low	High
Overhead in launching and managing tasks	High	Low
Fault tolerance	High	High
Scalability	Low	High
Flexibility in managing resources	High	Low

overhead in launching and managing tasks and may be less flexible in managing resources. Single task executors may provide faster processing times for individual tasks due to the reduced overhead and fewer resources needed. Multiple task executors may provide better overall throughput and efficiency due to their ability to handle multiple tasks simultaneously. Various factors such as size and complexity of tasks, the available resources, and the workload distribution may impact the performance of each approach.

Solution Environment Specification

1. AKS Cluster which is based on Standard F72s 2nd generation virtual machines powered by the second generation of Intel Xeon Scalable processors, code-named Skylake. It contains 72 virtual CPUs (vCPUs), 144GB of RAM, and 576GB of local SSD storage. Max cached and local disk IOPS (cache size in GiB) is 144000 (1520); Max. data disks (1023 GB each) is 32. Cost/Node \$3.49/hr as of Dec 2022.
2. Geographical location – both storage and compute are located in data centers in West Europe.
3. CPU only – the decision to use a CPU for CNN model inference was based on:
 - (a) limited availability of GPU resources (expensive both to buy or rent)
 - (b) inference is based on small data size in reverse to training
 - (c) powerfully CPU-based cluster able to manage CNN's model complexity (RAM capabilities)
 - (d) for the inference - ease of use and not requiring any specific configuration like Cuda or other specific drivers. Thinking about end users which are hospitals - it would be simpler and more convenient to use a CPU for inference.
4. Pre-trained model of Mask R-CNN with ResNet50+FP has been used.

Tab. 7.2 presents the results of six experiments on different executor types, worker pod numbers, and data partitions for processing medical image slices. The experiments conducted were based on single and partitioned data processing approaches, each with varying worker pod numbers and partition sizes. The set of patient data was 148 slices, fusion of CT and PET (512x512). The measured time is reflecting the end-to-end experiment execution, when triggering the job on the designed infrastructure.

TABLE 7.2: Results for Single and Partition Executor Types.

Experiment	Executor Type	Total Worker Pods	Data Partitions	Attempt 1 Results (s)	Attempt 2 Results (s)	Attempt 3 Results (s)	Avg Completion Time (s)
1	Single	4	N/A	193,53	181,72	178,54	184,59
2	Single	37	N/A	27,16	25,4	32,77	28,44
3	Single	74	N/A	12,47	13,53	12,55	12,85
4	Partition	4	37	147,4153	146,2625	122,9465	138,87
5	Partition	37	4	16,6736	17,1274	15,1963	16,33
6	Partition	74	2	7,012	5,7465	6,805	6,52

The completion time in seconds was recorded for each experiment run. The results showed that partitioned data processing, which allows for processing of data in chunks, is more efficient than single data processing. This can be attributed to the fact that for single data processing, there is an overhead related to network, storage, and scheduling. On the other hand, for partitioned data processing, these overheads occur only after the worker pods complete processing a single partition of data. This leads to a reduction in completion time, as seen in experiments 4, 5, and 6, which implemented partitioned data processing. Furthermore, as the number of worker pods increased in experiments 2 and 3 for single data processing, the completion time reduced. However, for partitioned data processing, an increase in worker pod numbers did not necessarily lead to a significant reduction in completion time. Mask R-CNN (ResNet-101), when run on a high-end GPU like the Nvidia Tesla V100, are around 5 FPS (Frames per Second) for high resolution images (e.g., 1024x1024 pixels) what gives 0.2s per each image [127]. Based on the measured results, processing the single slice varied consistently from 3s to 5s. In conclusion, partitioned data processing can lead to better processing time for medical image slices as it minimizes the overheads associated with single data processing.

7.1.4 Visualizing Medical Data

OHIF stands for Open Health Imaging Foundation. The tool can retrieve and load images from most sources and formats; render sets in 2D, 3D, and reconstruct; allows for the annotation and serialization of observations. The author's contribution is also the integration with a third-party viewer - OHIF Medical Imaging Viewer. Fig. 7.7 presents the result of integration and extension of the OHIF capabilities. Fig. 7.8 presents the visualization alone.

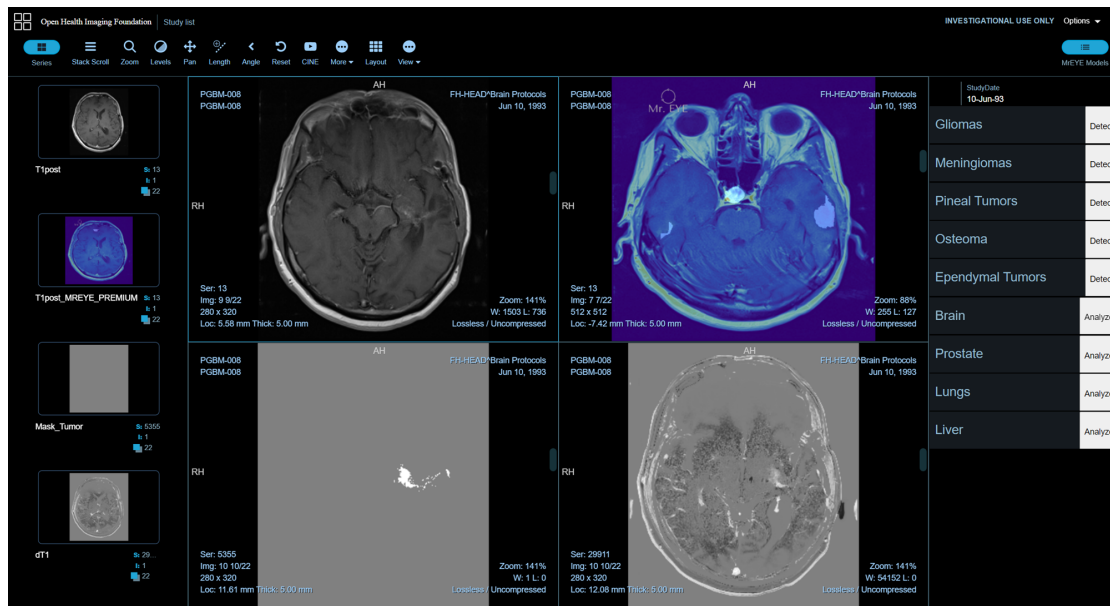


FIGURE 7.7: OHIF Medical Imaging Viewer extended by adding triggering the organ scanning or specific tumor detection. Created by the author.

7.1.5 Targeting User Needs with Persona-Based Approach

The designed platform is intended to be used by three groups of users:

1. AI developers and data scientists, researchers, designers, and creators of ML/DL algorithms; researchers and computer scientists working for software vendors (i.e., GE, Siemens) or clinical centers (CCDS, Stanford AIMI). They would have the capability to build the software infrastructure but would be willing to adopt existing solutions since they want to elevate above the infrastructure problems and are incentivized to focus on algorithm development.
2. Clinicians: Radiologists, general practitioners, oncologists, and surgeons – these users might not adopt the platform from the beginning; however, they are comfortable dealing with medical images and using software tools. They can identify clinical use cases for AI and build datasets. Given the right tools, they would be willing to use the platform and experiment with existing machine learning models by training them on their own datasets. There is growing appreciation among clinicians of the importance of medical imaging data that are generated within their hospitals.
3. IT admins at Hospitals and Clinical Research – technicians who administer hospital infrastructure. They would have the capabilities to use simplified tools to administer model enablement and infrastructure.

The predicted consequence is that a platform that facilitates medical imaging ML/DL will support the movement of massive amounts of storage and compute workloads from homegrown on-prem or cloud-based solutions. A cloud-native platform will

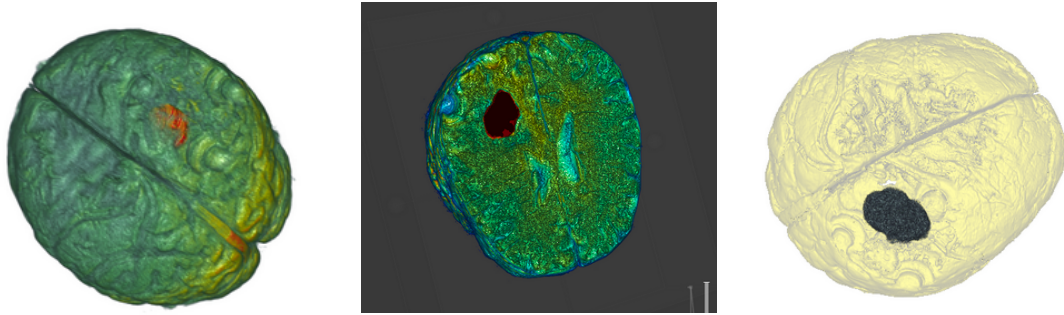


FIGURE 7.8: Brain visualizations for a single patient build on top of VTK library.

create an ecosystem of technology enthusiasts that will improve access to technology throughout the world. It will also create an opportunity to share algorithms among researchers and medical institutions and the uncertainty of model performance. Moreover, the medical imaging platform that encompasses radiological and non-radiological imaging (digital pathology, OCT, photo) can eventually expand beyond medical imaging to facilitate integrated data analytics workflows, encompassing medical text, speech, genomics, and meta-genomics.

7.1.6 Operationalization of ML Models

One of the biggest challenges facing the operationalization of ML models in medical imaging is having to integrate very closely with clinical workflows. For some scenarios, this can be done by integrating existing formats and creating various reports with data generated by ML algorithms (e.g., Secondary Capture or Structured Report formats defined by DICOM). However, some of the more advanced scenarios that integrate across clinical specialties or workflow-based decision making require new user-facing tools or updates to existing ones.

Fig. 7.9 proposes and illustrates the operationalization procedure designed and tested. It starts with data ingestion, which extracts data from various data sources and ingests the destination. That is implemented by repetitive and robust data pipelines which can also perform extraction, transformation, and loading (ETL) operations. Then, there is a data storage layer which commonly is the data lake storage, database, or other file systems (e.g., HDFS, DBFS). Then, there is a separate stage for data curation, which is based on the researched use cases focused on data quality, validation, and data preparation to be able to be annotated by SME using a tool integrated with a platform. Annotation requires defining the objective of the experiment and that smoothly leads to defining the research experiment. Once it is complete, then the third stage is a trialing model stage – a code that performs all the traditional steps in ML, such as data pre-processing, feature engineering, and feature scaling before training or retraining any model. Following this, the ML model is trained while performing hyperparameter tuning to fit the model to the dataset (training set). After training, there is a review of the model results stage. Meaning the evaluation of the performance of the trained model on a separated set of data points named test data (which were split and versioned in

the data ingestion step). The output of this step is a report on the performance of the trained model. The model deployment stage enables the ML models the author developed in the previous module (build). In this module, the model performance and behavior in a production or production-like (test) environment is tested to ensure the robustness and scalability of the ML model for production use. The models monitoring stage captures critical information to monitor data integrity, model drift, and application performance. Application performance can be monitored using telemetry data. It depicts the device performance of a production system over a period of time.

The designed operationalization procedure is practical, simplifies collaboration of research and AI developers with SME.

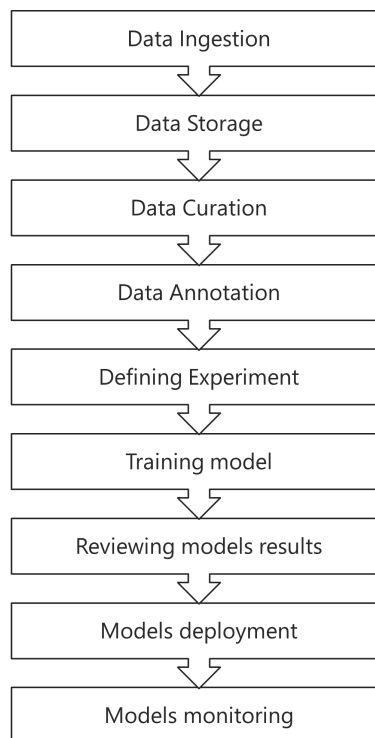


FIGURE 7.9: The designed operationalization procedure.

Based on experience, the author sees the need to specify the last four elements of Fig. 7.9, as it is more general, and the cooperation confirms that there are more stages related to only the model that must be mentioned. The improved workflow focused on the AI designer, AI developer, research personnel, and steps are illustrated in 7.10. First, a model is built or instantiated by defining its architecture. The models are then training as in the previous figure, but, once it is done, the model is packaged, serialized pickle file (Python's pickle module) or containerized (using Docker) to be exported to the other environment; or serialized to an ONNX.

The model is then evaluated on a validation dataset not used in the training. Once done, the model might be registered in the Azure Container Registry (or another container registry, e.g., Docker Hub). The model can then be deployed. Once deployed, it can be monitored for both request hitting, CPU memory, but also logs, and processing timing.

The model that was serialized or containerized in the previous step is registered and stored in the model registry. A registered model is a logical collection or package of one or more files that assemble, represent, and execute the personalized ML model. With the final points remaining the same.

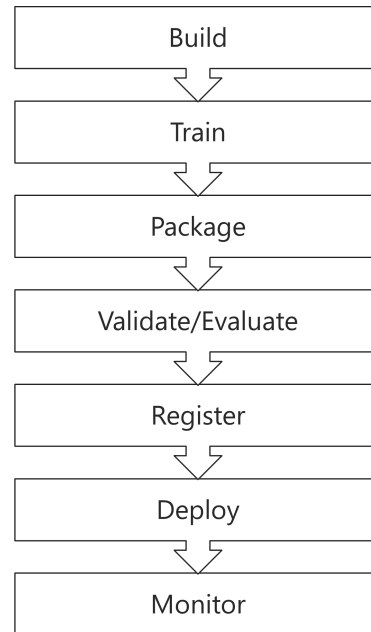


FIGURE 7.10: The designed operationalization procedure with the focus on researcher persona.

7.1.7 An Applicable Approach to Federated Learning

Federated learning is a type of machine learning that allows multiple parties to collaboratively train a model without sharing their data. Improving patient outcomes in healthcare can be achieved by utilizing federated learning to develop models, while still protecting patient privacy. For example, federated learning can be used to develop models that can predict the risk of a patient developing a certain condition, or to identify patterns in patient data that can be used to improve diagnosis and treatment. Additionally, federated learning can be effective in developing models which can identify potential drug interactions or to identify potential adverse events associated with certain medications. By using federated learning, healthcare providers would be able to develop models that improve patient care without compromising patient privacy.

The data are born at the edge, on scanners or devices/machines connected next to scanners, e.g., PACs system. The author developed a federated learning procedure visible in Fig. 7.11. Federated learning enables each model to be trained without centralizing data, retaining default privacy. Decentralized data remains with hospital departments.

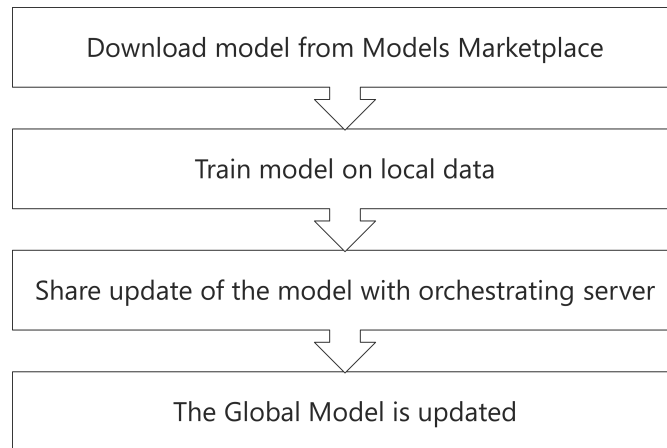


FIGURE 7.11: Federated learning - machine learning on decentralized data.

In practice, IT administrators at hospitals can download models from a Models Marketplace available for them. Then the Clinicians, Researchers or AI developers working within that environment can train the generic models on their local data. Once done, the model can share updates with the global orchestrating server, which will update the global version of the model. That is the manner in which the knowledge about a single each case detected in one institution may be immediately propagated across all medical centers.

Federated learning allows multiple hospitals to collaborate on a single machine learning model without having to share their data. In a hospital case, federated learning could be used to develop a machine learning model that can predict patient outcomes based on a variety of factors, such as medical history, current medications, and lifestyle habits. Each hospital would contribute its own data to the model, but the data would remain on the hospital's own servers. The model would then be trained on the combined data, and the resulting model would be shared among all the hospitals. This would allow each hospital to benefit from the insights of the other hospitals, while still maintaining the privacy of their own data. Fig. 7.12 presents the designed solution diagram.

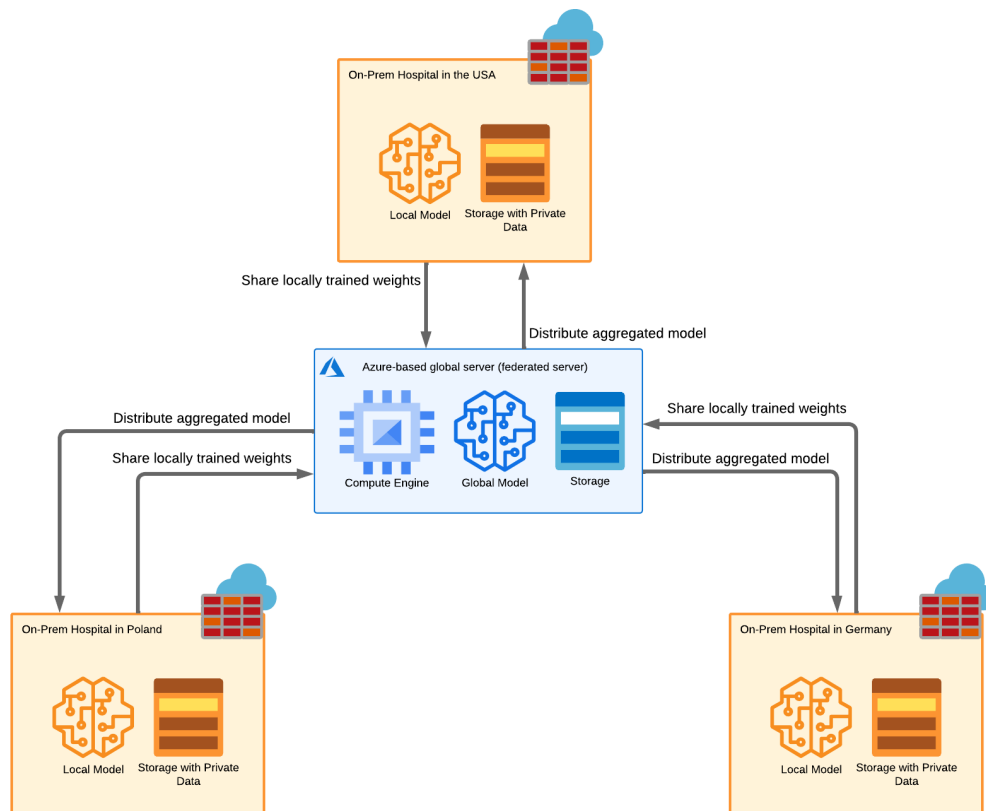


FIGURE 7.12: Federated learning solution diagram. Created by the author.

Model broadcast – clients download the current model and training instructions from the orchestrating server which leads to **Client training** – each client trains the model on its local data and then summarizes the changes to the model as a small focused update which leads to **Aggregation** – only the update to the model is sent to the orchestrating server, and it is immediately aggregated with other updates which leads to a **Model update** – the server applies the aggregate to the global model to produce the next iteration of the model

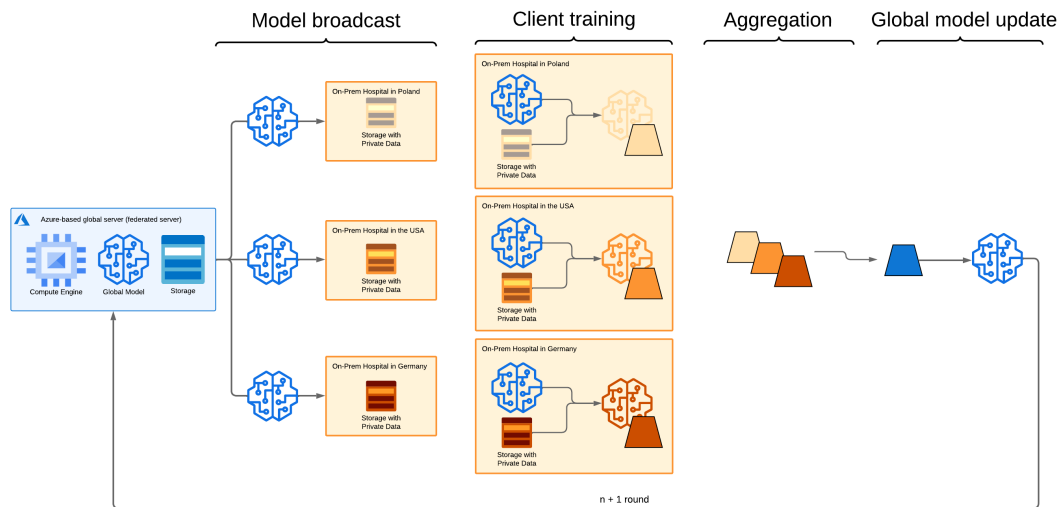


FIGURE 7.13: Federated learning designed case. Created by the author.

In terms of data sharing, Fig. 7.13 presents aggregation server topology.

7.1.8 Cost-Effective Cloud Development Strategy

The main advantage of cloud computing utilization is the financial model - no upfront costs needed and almost effortless configuration of graphic card drivers and runtime. The pricing model – “Pay as you go” – means paying for the usage. Costs are associated with data storage and processing. Each model training or running in batch or inference mode is correlated with the costs of the compute engine. Fig. 7.14 presents the costs in the middle of the month for the Azure Kubernetes Services cluster deployed in the datacenter East US 2. The Kubernetes cluster in version 1.22.6 was configured on a machine with 4 CPU cores, 28 GB memory, and 200 SSD storage.

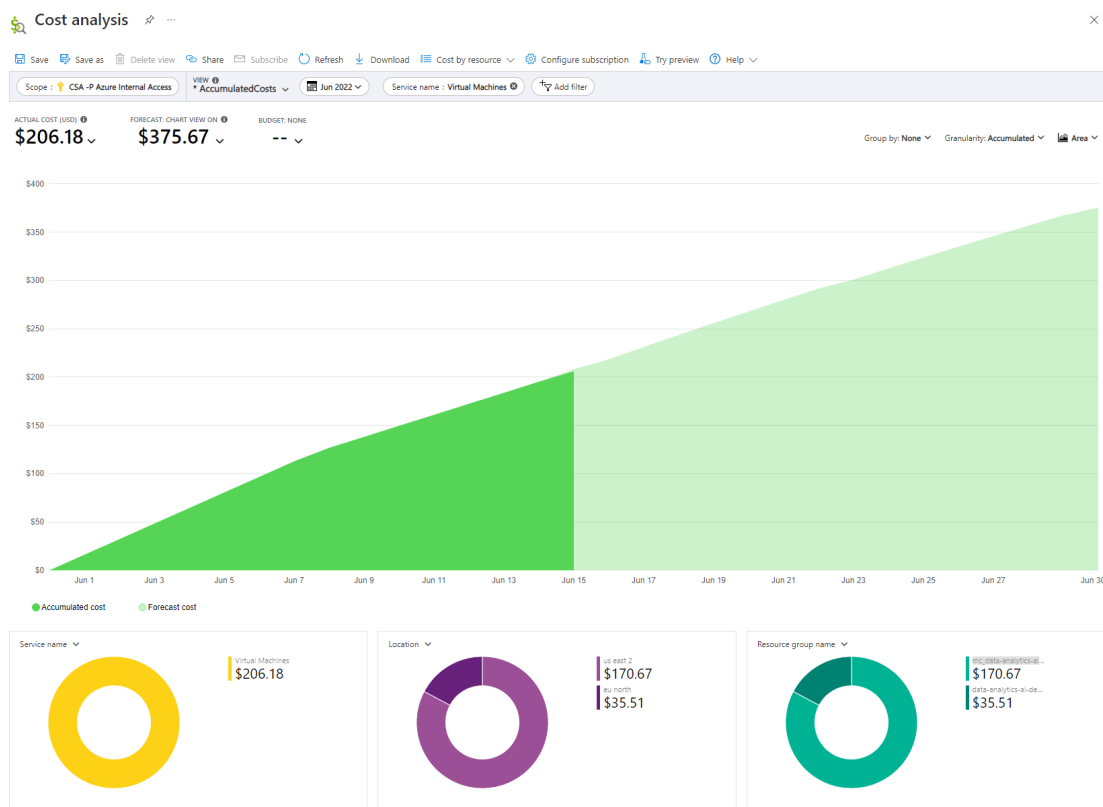


FIGURE 7.14: Costs analysis for an AKS cluster.

Dozens of models can be deployed and hosted in such a single cluster. The big advantage of the AKS cluster is automatically scaling the number of containers with the demands, which is a critical point when using that platform in the hospital's environment.

7.2 Ensuring Security in Medical Imaging

Data security issues, including data storage and transmission, are a major concern against the use of cloud platforms for medical data processing, although the division of responsibilities and guidelines against security systems have been defined [128]. Systems based on medical information are subject to a number of regulations and international standards of practice, such as ISO / ISO 27001, ISO / ISO 27002, ISO / ISO 22301, ISO 13606–1, ISO 13606–4 and ISO / ISO 27018. Using a public cloud to process medical data shifts the responsibility for security and compliance with applicable laws to the cloud provider. In the proposed architecture, the application used to transmit medical data communicates with the cloud provider's network through an Application Programming Interface (API). To ensure the security of this application, it is recommended to use HTTPS and double encryption, both on the client and server side. It is also necessary to connect the health unit network from which data is sent to the Microsoft Azure network using site-to-site VPN tunneling.

In the United States, HIPAA (the Health Insurance Portability and Accountability Act) was implemented to protect citizens from having their healthcare/medical data shared or sold for the benefit of corporations such as healthcare providers or insurance companies. Another law enacted safeguards the privacy of your personally identifiable information (PII). In Europe, GDPR (General Data Protection Regulation) sets rules for individual personal data and the movement of these data.

Before data is released, the PET-CT testing entity is required to delete or overwrite personal data from the files. Recovery of deleted metadata is not possible. For this reason, the author provided an analysis of the available variables and their values that can be readable from DICOM files. Some variables, such as pixel spacing, are necessary to calculate the tumor area. Removing such variables may prevent the algorithm from working correctly. Newhauser et al. [129] presented ways to anonymize images stored in DICOM format. Metadata can be encrypted using masking techniques [130, 131]. Metadata decryption is only possible if the encryption key is available.

7.3 Explainability in Enhancing Transparency and Trust

Hindering the widespread adoption and deployment of Deep Learning Models across industries is the perception of DLM being opaque black boxes. Particularly problematic is the lack of interpretability when the model is intended to support the final decision-making process. To counteract this, since 2016 there has been an effort in developing tools that can provide explanations for the predictions generated by DLM. These explainable AI (XAI) techniques aim to make the decision-making process of these models more transparent and comprehensible, thus enabling end-users to gain greater trust and confidence in their outputs. In the context of medical image processing, XAI techniques have the potential to improve the accuracy, reliability, and safety of DLM, thereby enhancing their usefulness in supporting clinical decision-making. Clinicians and radiologists can gain better insights into how the model arrived at its conclusion by being provided explanations for the predictions generated. Assisting them in validating the accuracy of the prediction and identifying any potential errors or biases. Additionally, XAI can help to identify regions of interest within the image that contributed most significantly to the prediction, which can also aid in clinical diagnosis and treatment planning. Zhou et al. [132] described the basic technique called CAM (Class Activation Mapping), which provides heat maps which visualize which parts of the input image were the most significant for the deep learning model during the classification. Technically the method uses a global average pooling layer and visualizes the weighted combination of the resulting feature maps at the pre-softmax layer. As the method requires retraining a linear classifier for each class, Selvaraju et al. [133] improved the method by adding the fusion mechanism of the class-discriminative property of CAM with guided backpropagation and deconvolution named Grad-CAM. Grad-CAM highlights fine-grained details of the predicted class on the images and visualizes input regions with high resolution details. Gradient-weighted Class Activation Mapping (Grad-CAM), uses the gradients of any target concept, flowing into the final convolutional layer to produce a coarse localization map highlighting important regions in the image for predicting the output. Fig. 7.15 presents exemplary results.

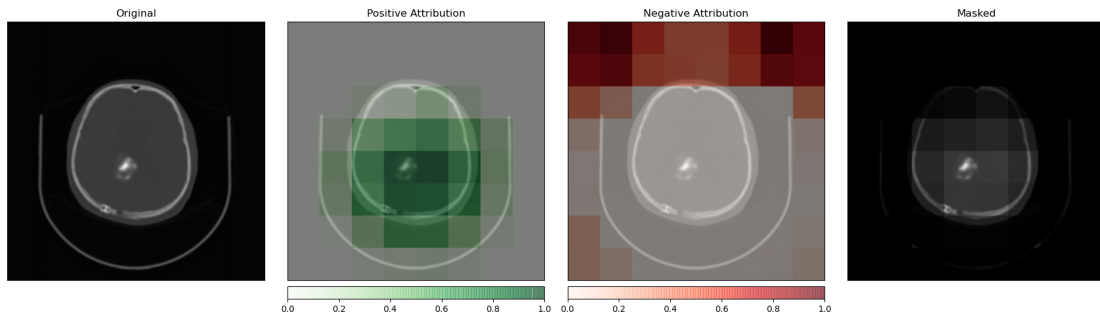


FIGURE 7.15: Predictions explained with Grad-CAM (LayerGradCam). Predicted: glioma tumor (0.6499). Mask R-CNN based on architecture ResNet50 trained on only 349 images. Created by the author.

7.4 An Ethical Framework for Medical Image Analysis

The thesis presents novel methods, algorithms, procedures, and studies that confirm effective detection and segmentation of brain tumors and can be used to support radiologists during the diagnostic stage and the treatment planning process. Thus, cloud-native architecture was proposed with industry market review and proposal of application deployment; discussing the ethics is necessary. With the deployment and use of these applications, the questions of ultimate responsibility, ethical behavior, and ethical use of the system might apply. These questions might be categorized as philosophical; however, based on the author's gathered experience during all stages of research, specific areas were identified on the engineering and science sides. The list below sums up the mandatory areas to consider when creating any ML or DL-based system dedicated to the medicine industry:

1. Bias – each application must be trained and evaluated on diverse datasets, including diversification of patients' gender, age, and race.
2. Fairness – the application must be evaluated toward intentional or unintentional unfair behavior, particularly when executing the classification or scoring tasks.
3. Transparency – the application's results must be auditable and explicable.
4. Responsibility – the application must ensure data security, and its behavior must be lawful and factual.
5. Interpretability – the application's results must be repeatable.
6. Simplicity – the application deployment and usage must be understandable for both Hospitals' IT specialists and end-users (Radiologists, Oncologists, Medical Professionals).
7. Control and maintenance – the operation of the application must be able to be controlled and straightforward to stop.

It can be observed that both science and engineering are striving to formulate ethical regulations that ensure ethically sufficient and responsible AI. Indeed, the market for

medical devices needs stricter and more precise regulations. Gundersen T. and Bærøe K. [134] examined the ways of collaboration between medical service providers, AI developers, and other stakeholders to apply AI and ML to shared decision-making in healthcare. The author strenuously underscores the insufficiency of publications and regulations on AI and accentuates the outright need to create principles that will shape both the development and the use of AI. The author claims that with AI, the clinical outcome can improve, but at the same time, patients can be deprived of their opinions, values, judgments, and treatment options if AI becomes fixed in a narrow case without human reason, consideration and capacity of care.

The thesis author stands up for defining proper evaluation of models, applications, and any new medical technology when it is deployed, as it is required for successful technology adaptation. It should be noted that the thesis and the research were intended to support medical professionals and providers, provide enhanced tools and make medical decision making more efficient, accurate, and evidence based. Collaborating with different healthcare workers and including them in the design of methods, algorithms, and procedures was essential to create valuable results that assist them but should never replace them.

The platform has been presented to medical doctor Maciej Geremek, Ph.D. from the Institute of Mother and Child, Warsaw Poland, who specialize in genetics and clinical vision. Geremek concluded the solution as: "This is the future, this kind of system. This is the way. That should be part of the entire hospital system".

Chapter 8

Conclusions

8.1 List of Novelties

The conducted research and experiments, introduced innovations and developed methods presented as results of the research aim to prove the primary objective of the thesis: **The methods based on a multi-stage deep learning pipeline that combine multiple neural networks of various architectures are capable of performing meaningful detection and segmentation of brain tumors and are able to be utilized in order to support radiologists during both the diagnostic stage and treatment planning process.** The authorial and proprietary contribution to the application of deep learning methods in the context of medical images processing is both systematic and substantial. The research and resulted novelties presented in this thesis are summarized as follows:

- Set the standard for the medical data processing solutions development by introducing the universal six-stage MeDAPR framework as described in chapter 4.
- Critical assessment of state-of-the-art CNNs models including evaluation of Mask R-CNN architecture for patient-specific dose distributions for Volumetric Modulated Arc Therapy (VMAT), as presented in chapter 4.2.3.
- Designed and implemented a method for removing bone tissues from CT-correlated PET scans as described in chapter 5.1.
- Designed and implemented Phantom-based Verification and Calibration Method (PVCMM) to validate and evaluate the accuracy of algorithms and assistance in setting the correct thresholds as described in chapter 5.2. Implemented classical active contour-based algorithm to compute tumor volume based on Edge and Chan-Vese algorithms. Test on the phantom resulted in achieving the relative error value only as 0.46%.
- Designed and implemented a method combining U-Net and Morphological Geodesic Active Contour algorithm, as described in chapter 6.1 with achieving the best U-Net model accuracy as 97.39% and 7.76% loss. The author conducted research to improve the method as described in 6.1.1 resulting in the conclusion that erosion and dilation to help better define the borders of the tumor and remove any surrounding artifacts or noise. The author evaluated results and compared them with previous methods, as presented in chapter 6.1.2.

- Implemented the algorithm to fuse CT with correlated PET scans, based on the Convolutional Neural Network model named VGG19 and VGG19 BN, as elaborated in chapter 6.2.1.
- To fuse multi-modal images such as CT, PET and MR produced by various diagnostic devices, the author introduced the CNN based method, as described in chapter 6.2.2.
- To generate additional training data, consistent with the original one, the author designed and implemented Mask-Aware Class-based Medical Data Augmentator, as presented in chapter 6.3.1, and Medical Annotations Transformer (MAT) as presented in chapter 6.3.2, to unify labels provided in COCO, VOC format with Azure ML Data Assets.
- Designed and implemented Tumor Area Cropper with U-Net architecture generator as introduced in chapter 6.4, based on the authorial novel equation (6.7) that allows to calculate the U-Net architecture depth, based on the image input size.
- Created a series of research experiments with the usage of U-Net model, to evaluate the accuracy of the segmentation based on the fused CT and PET scans, as described in chapter 6.5. The differences in results between different experiments when training models on single or multiple tumors, as well as with data augmentation and without, are minuscule especially when considering the time needed to train a model on 3x larger dataset. With an augmented dataset, the results achieved the highest accuracy; however, simultaneously, the number of false negative cases was higher, such as when the network indicated pixels as a background and by the radiologist as a tumor, is higher, and that value should be minimized and before deploying the model into an operational state, that value should be minimized.
- Evaluated pipeline with two stages combining YOLO v4 and U-Net to detect prostate and achieved the best model achieved the value of 0.9685 of Sørensen–Dice coefficient, as presented in 6.6.1.
- Designed and implemented pipeline with two stages combining Mask R-CNN and U-Net for brain tumor segmentation achieving an accuracy of nearly 97% on the validation dataset, as described in chapter 6.6.2.
- Elaborated on hyperparameters tuning of Mask R-CNN in chapter 6.6.3. Executed 108 separate training experiments to set the most optimal parameters for the models, including verification of backbones Mask R-CNN ResNet18+FPN, Mask R-CNN ResNet50+FPN, Mask R-CNN ResNet152+FPN, optimisers: SGD, Adam, AdamW, and different number of epochs on previously pre-trained models. The results proved that the configuration based on Mask R-CNN with the backbone ResNet152+FPN trained during 50 epochs with the optimizer SGD gives the highest F1-score of 0.8649. Therefore, an additional training configuration has been created with two changes: training during 300 epochs and learning rate set as 0.00529. As a result, obtained Precision was 0.9375, Recall was 0.88235 and the F1-score equal to 0.9091. mAP for that experiment was 0.93772.

- Provided the extensive research about the healthcare industry and proposed a novel architecture for a decision support system for radiologists, general practitioners, oncologists and surgeons as described in 7.1 which won the Microsoft Global Hackathon 2022 Executive Challenge - Hack for Healthcare Industry - 1st Place. The universal automated cloud architecture developed, which allows for training and also for running models in batch mode, can be implemented in both medical and research units working on medical data. Additionally, this solution can be implemented in any industry using CT.
- Proposed methods for parallel processing in chapter 7.1.3. Author also elaborated on An Applicable Approach to Federated Learning based on the collaboration with three institutions from the United States, Poland, and Germany, as described in 7.1.7.

The author also provided the fundamental assessment of state-of-the-art methods, introduced algorithms, and showed awareness of the subject area by expressing critical appraisal of used methods. Moreover, the dissertation leads discourse regarding the potential impact of the thesis contributions for the field of computer science and its applications in healthcare and beyond.

8.2 Summary

The primary objective of the work contained in this dissertation was to develop methods based on a proposed multi-stage deep learning pipeline that combine multiple neural networks of various architectures capable of performing meaningful detection and segmentation of brain tumors and are able to be utilized in order to support radiologists during both the diagnostic stage and treatment planning process. The methods developed are aimed at improving the accuracy and efficiency of medical image analysis.

In surveying the author's research steps, the focus shifts to the initial methods based on traditional active contour techniques, such as Edge and Chan-Vese, which were developed based on the first data received from the Nuclear Medicine Department of the Medical University of Warsaw, at the Central Clinical Hospital (Banach Hospital). This data consisted of CT images and correlated PET images. Based on selected and developed pre-processing methods, the tumor volume was determined based on PET and CT images, where the PET image was key, as from a domain perspective, it already represented only the tumor, or tumors and their metastases, reflecting their metabolic activity. The main task was to develop effective algorithms that would perform the segmentation of the examination results. The CT image, which depicts detailed body structures (anatomical structures), was crucial to refute the places where the tumor could not be, i.e., in the bones. Measuring the volume of tumors from only the PET scans may result in an error. Therefore, whenever a PET scan is performed, either CT or MRI is performed. The developed methods for this original case turned out to be exceptionally effective and were employed in cooperation with the mentioned medical institution.

Despite solving the problem posed by the hospital, namely calculating the volume, and developing a diagnostic aid tool, the author of the thesis set a more ambitious goal

for the work. Namely - is the calculation of the surface area/volume of gliomas, brain tumors, or other tumors, based only on a CT examination possible. The central consideration was the significant impact any developed method would have on the treatment process for people with brain tumors for whom a PET scan was performed. On a global scale, CT, MRI, and ultrasound scans remain the most common. For this reason, the author was dedicated to cultivating various methods of fusing CT-PET examinations to develop segmentation methods that effectively work on images with a significant level of detail and visible structures. On each occasion, upon achieving satisfactory results of the created method, an evaluation was performed, mainly based on the segmentation result. The author contends that the evaluation of methods based on segmentation alone is sufficient. To support this contention, firstly, the methods segment the tumor on each slice, in a series. Once the methods have the segmented tumor on all slides coming from the same patient, the tumor's volume can be calculated by summing all the segmented areas (squares) and multiplying that by the volume of a single voxel, which is stored in the metadata of the DICOM files. Therefore, merely possessing the segmented tumor, the tumor volume is easily calculated.

Subsequently, delving into the domain of the problem, another goal was set before the author of the thesis. In this case, the segmentation of the glioma and calculation of the volume from the results of CT-PET examinations were insufficient because of a new method being verified on patients - MRI-PET. Medical imaging utilizing CT is harmful especially when performed frequently because it is based on ionizing radiation, which is the reality of people undergoing assessments of tumor progression. Additionally, MRI is a method that provides more advanced and more detailed results, especially for soft tissues. The author, in cooperation with the hospital, set another goal, which was to develop methods for fusing these three modalities. It was an additional extension, as the fusion of only MRI-PET was initially required, but that method was simple. The fusion of MRI and CT was complicated due to the details. One cannot allow for the loss of information even with different resolutions, contrasts, and brightness normalization. Then there is the issue related to being properly matched (patient movement). In the ensuing research and experimentation, it became possible to develop these methods. Which led the author to develop further methods that, based on any type of examination, CT, CT-PET, MRI, CT-PET-MRI, segmentation would be able to be effectively performed thus providing the correct volume.

Importantly, the SME (radiologist) provided the volume based on best knowledge and was measured on a specific day. However, the masks, or labels, were not created using the same tool, at the same time. This indicates that the volume results in cm^3 vs. the masks on which the models were taught may differ. It is not possible to evaluate the percentage of this error at this time but given a few percent would be the correct approach.

In the case of verifying the methods based on examinations from the Lower Silesian Oncology Center in Wrocław, the cooperation concerned the detection or segmentation of the prostate organ, so that in the subsequent step, within the prostate itself, the segmentation of the primary prostate tumor could be performed. Here, the effectiveness of the developed methods is the completely segmented area of the prostate, as no volume information was provided by the hospital.

The final methods which fulfil the thesis, methods based on a multi-stage deep

learning pipeline that combine multiple neural networks of various architectures capable of performing meaningful detection and segmentation of brain tumors and can be utilized to support radiologists during both the diagnostic stage and treatment planning process, were developed using the Mask R-CNN model based on various studies. The effectiveness of the U-Net was astonishing, but in the broader context of implementing the developed methods, namely in the context of ceasing the use of a CT examination for tracking tumor progression, locating tumors and their metastases turned out to be crucial. For these reasons, the author spent a significant amount of time on experiments aimed at developing effective methods, understood as methods that return the highest segmentation accuracy on different types of examinations.

AI technology is radically transforming medical imaging. Publications in the academic and private sectors continue to reveal scientific advancement in this field. Prominent professionals are advocating for the development and adoption of machine learning algorithms for medical imaging. As a generation, we are on the cusp of a new era, ultimately, the billions of medical images acquired annually will be processed by algorithms that will have the ability to extract much more measurable clinical value than is possible by human analysis alone. These images will be reused repeatedly, together with data from other diagnostic modalities, to improve ability to understand the processes within living organisms. The author's developed methods have the potential to significantly improve the diagnosis of diseases and ultimately contribute to improving patient outcomes. The modern research trend of attempting to use medical data in a healthcare environment application and transformation requires appropriate standards and frameworks in order to achieve applicable outcomes.

8.2.1 Broader Impact

The newly introduced methods and results obtained in the research as presented in this thesis have a high potential for implementation in industrial and commercial practice. This research may lead to the development of new products and services, further algorithms, or tools in the field of medical imaging, and development of new methods in the field of computer vision image analysis with robotics or autonomous machines where precise object detection based on grayscale images is required. Although this thesis' research focus was on medical data processing, at the same time, the core domain could be considered Computer Science. As a quote often misattributed to Albert Einstein states "Computers are incredibly fast, accurate, and stupid. Human beings are incredibly slow, inaccurate, and brilliant. Together they are powerful beyond imagination" [135]. The author's deep belief and intention was to develop methods, algorithms, and procedures for the fully automated detection and segmentation of specific objects – mainly brain tumors, and the prostate – while integrating the necessary work of medical health professionals along with best practices for patients.

Fusion Methods Impact

During MRI and CT examination, patients are asked to remain completely immobile. The examination process requires immobility from a few minutes up to several minutes. This "perfectly still" position is nearly impossible for patients with Movement Disorders or Epilepsy, along with patients experiencing anxiety or mental health

episodes, etc. Additionally, MRI and CT scans are performed in veterinary medicine. Sedation is the only option when attempting medical imaging examinations in these cases. The chief reason for the need to stay still is due to image quality, which is degraded by even a slight movement (blurring or distortion). The consequence of movement during the examination is the repetition of the exam, which is problematic for the patient and the provider. In the case of CT, it means double the radiation exposure with additional cost and time increases. The algorithm developed for fusion can be adjusted and applied to the post-processing of the images which were affected by patient movement during the initial exams, thus eliminating the need for repeat scans and increased patient distress.

Author's Disclosure and Transparency

The conducted research was carried out within the framework of several projects, including:

- Research funding through the Centre for Priority Research Area Artificial Intelligence and Robotics of Warsaw University of Technology within the Excellence Initiative: Research University (Inicjatywa Doskonałości Uczelnia Badawcza, IDUB) programme.
- “Design and development of algorithms for automatic identification, volume measurement and visualization of brain tumors and prostate cancer using deep learning methods” project supported by Warsaw University of Technology by the grant 504/04560/1042/43.040001 from 8 of July 2020 led by Krzysztof Siwek.
- “Development of methods for searching for the optimal dose distribution in radiotherapy planning using deep learning techniques” project supported by the Warsaw University of Technology under grant 504/04560/1042/43.040002 from 8th of July 2020 led by Zuzanna Krawczyk-Borysiak.

In the spirit of full disclosure, when the research contained in this dissertation commenced, the author was in the latter stages of achieving a master's degree while simultaneously employed by Intel. As is the case with all IT career paths, this employment was replaced with Sony, CitiBank, Proctor and Gamble and most recently Microsoft, in that order. These fluctuations in employment were to expose the author to skill development in order to grow in knowledge and pursuit of excellence. The author wishes to unequivocally state that there was no bias influencing the research or recommendations toward or for a single corporate entity, regardless of the relationship between the author and any/all corporations.

8.3 Future Directions

In this section, the author proposes several ideas to continue the work of redefining the application of new technologies in healthcare, particularly in the context of medical imaging. Continuing this research is crucial to ensure the accuracy and reliability of the results. This involves reviewing the data collected, analyzing the results, and drawing

conclusions from the findings. It is also important to identify any potential limitations of the research and areas that require further exploration. The dissemination of research results to relevant stakeholders and the wider public is also crucial.

In future research, the author plans to extend their collaboration with the medical community dealing with prostate and breast tissue tumors. The author intends to expand research in these areas, exploring new approaches to improve the accuracy and efficiency of tumor detection and segmentation. One possible way to bridge the gap between technological advancements and the medical community is to establish a collaborative platform for knowledge sharing and dissemination. However, a significant challenge with this approach is managing large amounts of medical data and ensuring the privacy and security of patient information. The author plans to explore novel solutions for addressing these challenges and to develop new methods for the efficient and secure processing of medical data.

The detailed plan includes the following points:

- The development of an **anonymization algorithm** that addresses the lack of data bottleneck in DICOM sensitive slices, such as those related to the face, breast, and reproductive organs. This algorithm aims to protect patient privacy in compliance to HIPAA regulations by removing or obscuring any identifying information from the DICOM slices, such as patient name, date of birth, and other personal information. The algorithm ensures that the data is used only for its intended medical purpose, while also reducing the risk of data breaches by making the data unidentifiable.
- The extension of existing methods to support **3D based deep neural networks (DNNs) image processing** to capture spatial and temporal relationships. 3D DNNs are particularly important in medical imaging as they can capture the spatial relationships between different parts of the image, allowing for a more detailed and accurate diagnosis of diseases. Additionally, 3D DNNs can track changes in the image over time, which is important for monitoring disease progression and treatment response. The context of the image is also important for understanding the overall structure of the image, while capturing the fine details can help in accurately diagnosing and treating diseases.
- **The application and adjustment of currently developed methods on different types of tumors, such as breast, adrenal glands, lung, colon, liver, and rectal.** This approach can lead to improved accuracy in the diagnosis and treatment of various types of tumors, as well as provide valuable insights into the underlying biology of the disease. Furthermore, this can facilitate the identification of new biomarkers that can be used to enhance predictions of disease progression in order to develop more effective treatments.
- Extending current methods to **work on data from ultrasound (USG) scans**, specifically converting USG video to slices. This expansion can facilitate the identification of new patterns and trends in the data that may not have been previously identified. By detecting abnormalities in organs, tissues, and blood vessels, this approach can help diagnose and treat a wide range of medical conditions. Additionally, ultrasound scans can be used to monitor the progress of a pregnancy,

detect fetal abnormalities, and assess the health of the placenta. The technique can also be used to diagnose and treat certain types of cancer. By analyzing the data obtained through USG scans, doctors can make more informed decisions about a patient's care.

- Developing **methods for data compression** to optimize the storage and transmission of medical data. This can have significant benefits for organizations, including reduced storage costs and improved network performance. By compressing medical data, organizations can reduce the amount of storage space required, which can lead to cost savings. Additionally, compressed data can be transmitted more quickly over a network, leading to improved network performance. This is particularly important in healthcare, where large amounts of data are generated and transmitted on a regular basis.
- Transferring the current solution to work with Industrial CT Scans. Industrial CT scans, also known as computed tomography or CT scans, are an imaging technology that produces three-dimensional images of objects. This process uses X-rays to take multiple two-dimensional images of an object from different angles and then combines them to create a three-dimensional representation of the object. The **use of Industrial CT scans** is widespread in manufacturing, where it is used to inspect internal structures of objects and detect flaws or defects that may not be visible to the naked eye. The technology can also be used to measure the dimensions of objects and analyze their material composition.
- Developing a procedure for **4D medical image analysis procedure**. 4D medical image analysis is a process of analyzing 3D medical images over time, using specialized software to analyze the changes in the 3D images. The proposed procedure aims to detect and diagnose diseases, track the progression of a disease, and monitor the effectiveness of treatments. This procedure can be used to analyze various types of medical images, including CT scans, MRI scans, PET scans, and ultrasound images. Additionally, it can be used to analyze images from other medical imaging modalities, such as X-ray and nuclear medicine. The analysis of 4D medical images can provide valuable information about the structure and function of the body, allowing doctors to make more informed decisions about diagnosis and treatment.
- Developing a **medical ROI-based image search system**, which allows efficient and accurate retrieval of images with similar regions from a database given a query of multimodal medical images from the brain, prostate, or breast with a tumor (ROI) region. The system will use advanced deep learning algorithms to analyze the image characteristics and extract features for similarity analysis. This image search system will demonstrate the ability to classify pathological characteristics for patients and allow for a more effective diagnosis and treatment. The system will also adapt semantic information from radiology reports for a more efficient and accurate search for image characteristics. Additionally, the image search algorithm will be extended to 3D and different imaging modalities, allowing for more comprehensive and accurate search results.

- **Developing innovative cross-validation techniques** designed to consider the spatial presence of tumors within given medical images. Conventional cross-validation methods indiscriminately shuffle data and divide them into k folds. This randomization neglects spatial characteristics, such as the distribution of tumors between image slices. Data used in that research on average contained tumors between the 70th and 120th slices out of total 148 scan series. Using slices without tumor for the model evaluation may result in a skewed assessment.
- Working with a patent attorney to **navigate legal requirements and procedures for patents.**

Chapter 9

Acknowledgements

I would like to express my gratitude to all the people who have supported me all who have shown their support during this lengthy research journey. Warm thanks to my family and dear friends, my mentors and colleagues, and those who encouraged my progress.

There are those who I wish to share my appreciation by name:

1. Krzysztof Siwek, my primary advisor, for more than 10 years of cooperation and for enabling me to develop my interests and scientific competence, and for supporting me at every stage of its development and crucial career decisions.
2. Zuzanna Krawczyk-Borysiak, my co-advisor, for her mentoring, support, and numerous suggestions for the research conducted.
3. Colleagues from the Faculty of Electrical Engineering, including Tomasz Leś, Radosław Roszczyk, Żaneta Świdorska-Chad, and Robert Szmurło, for their constructive comments and discussion of problems, and for not allowing me to give up when working fully remotely.
4. The employees of the Nuclear Medicine Department of the Medical University of Warsaw, especially Piotr Czwarnowski and Professor Leszek Królicki, for successful and productive cooperation, for teaching me all the flavors of radiology, and for providing me with labeled data.
5. The employees of the Lower Silesian Oncology Center in Wrocław (Dolnośląskie Centrum Onkologiczne), especially Adam Maciejczyk and his team, for successful and productive cooperation and for providing me with data.
6. Kirsten Stellick and Jessica Clark for their profound guidance in English language proficiency and relentless support in reviewing my thesis from the language perspective to prepare it for international scrutiny and potential postdoctoral pursuits in the United States.
7. My direct supervisors from Microsoft, for allowing me to take time off to fully focus and complete the thesis.

Publications

1. **U-Net and Active Contour Methods for Brain Tumour Segmentation and Visualization**, Paper First Author, *International Joint Conference on Neural Networks (IJCNN)*, 2020.
2. **Volumetric Modulated Arc Therapy Dose Distribution Prediction for Breast Cancer Patients: CNN Approach**, Paper Contributor, *International Joint Conference on Neural Networks (IJCNN)*, 2021.
3. **Deep learning based framework for tumour detection and semantic segmentation**, Paper First Author, *Bulletin of the Polish Academy of Sciences: Technical Sciences*, 2021.
4. **Segmenting prostate on onefold and joint multimodal medical imaging**, Paper First Author, *23rd International Conference on Computational Problems of Electrical Engineering (CPEE)*, 2022.
5. **The Application of a Deep Active Contour Network to the Task of Brain Tumour Segmentation**, Paper Contributor, *Progress in Applied Electrical Engineering PAEE*, 2022.
6. **U-Net training models for efficient brain tumour segmentation on multi-modality CT and PET images**, Paper First Author, *IEEE, 22nd International Conference on Computational Problems of Electrical Engineering (CPEE)*, 2021.
7. **Reducing the Uncertainty of Agile Software Development Using a Random Forest Classification Algorithm**, Paper Contributor, Springer, *International Conference on Lean and Agile Software Development*.
8. **Architektura platformy korzystającej z metod uczenia maszyny nowego do przetwarzania obrazów biomedycznych w chmurze [Architecture of a platform utilizing machine learning methods for processing biomedical images in the cloud]**, Paper First Author, 2020, *Poznan University of Technology Academic Journals. Electrical Engineering; Computer Applications in Electrical Engineering*.
9. **Segmentation of brain biomedical images to compute the volume of gliomas**, Paper First Author, 2019, *ITM Web of Conferences*.
10. **Verification of the results of medical imaging of brain via the test using phantom PET-CT**, Paper First Author, *Post-conference paper, 42th Conference on Fundamentals of Electrotechnics and Circuit Theory*.
11. **Algorytm segmentacji obrazów CT i PET do identyfikacji i wyznaczania objętości glejaka mózgu [Algorithm for segmentation of CT and PET images for identification and determination of the volume of brain glioma]**, Paper First Author, *Poznan University of Technology Academic Journals. Electrical Engineering; Computer Applications in Electrical Engineering*, 2019.

12. **Improved Morphological Geodesic Active Contours Algorithm for Automated, Voxel-wise Segmentation of Biomedical Images**, Conference Paper *ESCO 2020–7th European Seminar on Computing*
13. **Weryfikacja wyników badań obrazowania medycznego mózgu za pomocą testu na fantomie PET-CT [Verification of brain medical imaging results using a PET-CT phantom test]**, Conference Paper presented during *42th Conference on Fundamentals of Electrotechnics and Circuit Theory*.
14. **Semantic Segmentation of the Prostate Based on Onefold and Joint Multimodal Medical Images Using YOLOv4 and U-Net** - Paper First Author, forthcoming.
15. **Deep multimodal medical imaging fusion for an efficient glioblastoma multi-forme segmentation** - Paper First Author, forthcoming.

Bibliography

- [1] "Global opportunity analysis and industry forecast, 2018 - 2025," <https://www.alliedmarketresearch.com/image-recognition-market>, accessed: 2022-11-19.
- [2] Medixant, "Radiant dicom viewer," 2022. [Online]. Available: <https://www.radiantviewer.com>
- [3] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Lecture Notes in Computer Science*, vol. 9351.
- [4] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "Scaled-YOLOv4: Scaling cross stage partial network," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2021, pp. 13 029–13 038.
- [5] E. Kot, "Segmentacja obrazów biomedycznych mózgu do wyznaczania objętości glejaka [Segmentation of biomedical brain images for determining the volume of glioma]," Master's thesis, Politechnika Warszawska, Poland, 2016.
- [6] E. Kot, Z. Krawczyk, K. Siwek, and P. Czwarnowski, "U-net and active contour methods for brain tumour segmentation and visualization," in *2020 International Joint Conference on Neural Networks (IJCNN)*, Glasgow, United Kingdom, pp. 1–7.
- [7] E. Kot, Z. Krawczyk, K. Siwek, L. Królicki, and P. Czwarnowski, "Deep learning-based framework for tumour detection and semantic segmentation," *Bulletin of the Polish Academy of Sciences: Technical Sciences*, vol. 69, no. 3, p. e136750, 2021.
- [8] E. Kot, "Computer vision for biomedical images processing," Aug. 2020. [Online]. Available: <https://doi.org/10.5281/zenodo.5091735>
- [9] D. Mukherji, R. H. Murillo, M. Van Hemelrijck, V. Vanderpuye, O. Shamieh, J. Torode, C. S. Pramesh, A. Yusuf, C. M. Booth, A. Aggarwal, and R. Sullivan, "Global cancer research in the post-pandemic world," *The Lancet Oncology*, vol. 22, no. 12, pp. 1652–1654, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1470204521006021>
- [10] G. B. of Disease 2019 Cancer Collaboration, "Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life Years for 29 Cancer Groups From 2010 to 2019: A Systematic Analysis for the Global Burden of Disease Study 2019," *JAMA Oncology*, vol. 8, no. 3, pp. 420–444, 03 2022. [Online]. Available: <https://doi.org/10.1001/jamaoncol.2021.6987>

- [11] R. L. Siegel, K. D. Miller, H. E. Fuchs, and A. Jemal, "Cancer statistics, 2022," *CA: A Cancer Journal for Clinicians*, vol. 72, no. 1, pp. 7–33, 2022. [Online]. Available: <https://acsjournals.onlinelibrary.wiley.com/doi/abs/10.3322/caac.21708>
- [12] R. Rake, "National institute of health and care excellence (nice)," <https://www.nice.org.uk/Media/Default/About/what-we-do/Into-practice/measuring-uptake/End-of-life-care-impact-report/nice-impact-end-of-life-care.pdf>, accessed: 2022-06-19.
- [13] "Brain, other cns and intracranial tumours incidence statistics," <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/brain-other-cns-and-intracranial-tumours/incidenceref->, accessed: 2022-06-19.
- [14] A. Tamimi and M. Juweid, "Epidemiology and outcome of glioblastoma," in *Glioblastoma [Internet]*. Brisbane (AU: Codon Publications.
- [15] V. C. P. Jigisha P Thakkar, Pier Paolo Peruzzi, "Neurosurgical Conditions and Treatments - Glioblastoma Multiforme," <https://www.aans.org/en/Patients/Neurosurgical-Conditions-and-Treatments/Glioblastoma-Multiforme>, 2022.
- [16] R. Stupp, W. Mason, M. Bent, M. Weller, B. Fisher, and M. Taphoorn, "Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma," *N Engl J Med*, vol. 352, p. 987–96.
- [17] N. Grech, T. Dalli, S. Mizzi, L. Meilak, N. Calleja, and A. Zrinzo, "Rising incidence of glioblastoma multiforme in a well-defined population," vol. 12, 05 2020.
- [18] R. W. Parsa-Parsi, "The Revised Declaration of Geneva: A Modern-Day Physician's Pledge," *JAMA*, vol. 318, no. 20, pp. 1971–1972, 11 2017. [Online]. Available: <https://doi.org/10.1001/jama.2017.16230>
- [19] S. Wang and R. M. Summers, "Machine learning and radiology," *Medical image analysis*, vol. 16, no. 5, pp. 933–951, 2012.
- [20] T. Salditt, T. Aspelmeier, and S. Aeffner, *Biomedical Imaging: Principles of Radiography, Tomography and Medical Physics*. De Gruyter, 2017. [Online]. Available: <https://doi.org/10.1515/9783110426694>
- [21] D. Cordier, F. Forrer, F. Bruchertseifer, A. Morgenstern, C. Apostolidis, S. Good, J. Mueller-Brand, H. Maecke, J. C. Reubi, and A. Merlo, "Targeted alpha-radionuclide therapy of functionally critically located gliomas with 213bi-dota-[thi8,met(o2)11]- substance p: A pilot trial," *European journal of nuclear medicine and molecular imaging*, vol. 37, pp. 1335–44, 02 2010.
- [22] P. Czwarnowski, "Wpływ rekonstrukcji na wartość suv w badaniu pet-ct [the impact of reconstruction on the suv value in pet-ct examination]," *Inżynier i Fizyk Medyczny*, vol. Vol. 8, nr 1, pp. 38–39, 2019.
- [23] J. Kunikowska, R. Czepczyński, D. Pawlak, H. Koziara, K. Pełka, and L. Krolicki, "Imaging of glial tumor recurrence using [68ga]ga-psma-11 pet/ct," 08 2021.

- [24] G. Choy, O. Khalilzadeh, M. Michalski, S. Do, A. E. Samir, O. S. Pinykh, J. R. Geis, P. V. Pandharipande, J. A. Brink, and K. J. Dreyer, "Current applications and future impact of machine learning in radiology," *Radiology*, vol. 288, no. 2, pp. 318–328, 2018.
- [25] R. L. Siegel, K. D. Miller, H. E. Fuchs, and A. Jemal, "Cancer statistics, 2021." *CA: a cancer journal for clinicians*, vol. 71, no. 1, pp. 7–33, 2021.
- [26] Z. Obermeyer and E. J. Emanuel, "Predicting the future—big data, machine learning, and clinical medicine," *New England Journal of Medicine*, vol. 375, no. 13, pp. 1216–1219, 2016.
- [27] M. Strzelecki and P. Badura, "Machine learning for biomedical application," *Applied Sciences*, vol. 12, no. 4, 2022. [Online]. Available: <https://www.mdpi.com/2076-3417/12/4/2022>
- [28] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.
- [29] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros *et al.*, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *JAMA*, vol. 316, no. 22, pp. 2402–2410, 2016.
- [30] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya *et al.*, "Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning," *arXiv preprint arXiv:1711.05225*, 2017. [Online]. Available: <https://arxiv.org/abs/1711.05225>
- [31] X. Chen, B. M. Williams, S. R. Vallabhaneni, G. Czanner, R. Williams, and Y. Zheng, "Learning active contour models for medical image segmentation," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 11 624–11 632.
- [32] D. Marcos, D. Tuia, B. Kellenberger, L. Zhang, M. Bai, R. Liao, and R. Urtasun, "Learning deep structured active contours end-to-end," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.
- [33] M. Zhang, B. Dong, and Q. Li, "Deep active contour network for medical image segmentation," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, A. L. Martel, P. Abolmaesumi, D. Stoyanov, D. Mateus, M. A. Zuluaga, S. K. Zhou, D. Racoceanu, and L. Joskowicz, Eds. Cham: Springer International Publishing, 2020, pp. 321–331.
- [34] H. Dong, G. Yang, F. Liu, Y. Mo, and Y. Guo, "Automatic brain tumor detection and segmentation using u-net based fully convolutional networks," in *Medical image understanding and analysis*, V. H. M and G.-C. V, Eds. Cham: Springer, pp. 506–517,.

- [35] K. Hu, C. Liu, X. Yu, J. Zhang, Y. He, and H. Zhu, "A 2.5d cancer segmentation for mri images based on u-net," in *5th International Conference on Information Science and Control Engineering (ICISCE)*, Zhengzhou, pp. 6–10,.
- [36] H. Kaldera, S. Gunasekara, and M. Dissanayake, "Brain tumor classification and segmentation using faster r-cnn," in *Advances in Science and Engineering Technology International Conferences (ASET)*, Dubai, United Arab Emirates, pp. 1–6,.
- [37] M. Al-masni, M. Al-antari, J. Park, G. Gi, T. Kim, P. Rivera, E. Valarezo, S.-M. Han, and T.-S. Kim, "Detection and classification of the breast abnormalities in digital mammograms via regional convolutional neural network," in *39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Seogwipo, pp. 1230–1233,.
- [38] P. Yin, R. Yuan, Y. Cheng, and Q. Wu, "Deep guidance network for biomedical image segmentation," in *IEEE Access*, vol. 8.
- [39] R. Sindhu, G. Jose, S. Shibon, and V. Varun, "Using yolo based deep learning network for real time detection and localization of lung nodules from low dose ct scans," in *Proc. SPIE 10575, Medical Imaging 2018: Computer-Aided Diagnosis*, vol. 105751I.
- [40] R. Ezhilarasi and P. Varalakshmi, "Tumor detection in the brain using faster r-cnn," in *2nd International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)*, 2018 2nd International Conference on, Palladam, India, pp. 388–392,.
- [41] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, p. 91–99.
- [42] S. Liu, H. Zheng, Y. Feng, and W. Li, "Prostate cancer diagnosis using deep learning with 3d multiparametric mri," in *Proceedings of Medical Imaging 2017: Computer-Aided Diagnosis*, vol. 10134. Bellingham: International Society for Optics and Photonics (SPIE), p. 1013428.
- [43] M. Gurbină, M. Lascu, and D. Lascu, "Tumor detection and classification of mri brain image using different wavelet transforms and support vector machines," in *2019 42nd International Conference on Telecommunications and Signal Processing (TSP)*, Budapest, Hungary, pp. 505–508,.
- [44] S. Chen, A. Qin, D. Zhou, and D. Yan, "U-net-generated synthetic ct images for magnetic resonance imaging-only prostate intensity-modulated radiation therapy treatment planning," *Medical physics*, vol. 45, no. 12, pp. 5659–5665, 2018.
- [45] X. Huang, B. Zhang, X. Zhang, M. Tang, Q. Miao, T. Li, and G. Jia, "Application of u-net based multiparameter magnetic resonance image fusion in the diagnosis of prostate cancer," *IEEE Access*, vol. 9, pp. 33 756–33 768, 2021.
- [46] P. Chen and L. Xu, "Implementation of fpga-based hardware acceleration system for u-net networks," in *2023 IEEE 2nd International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA)*, 2023, pp. 1164–1166.

- [47] A. Sedghi, A. Mehrtash, A. Jamzad, A. Amalou, W. M. Wells, T. Kapur, J. T. Kwak, B. Turkbey, P. Choyke, P. Pinto *et al.*, "Improving detection of prostate cancer foci via information fusion of mri and temporal enhanced ultrasound," *International journal of computer assisted radiology and surgery*, vol. 15, no. 7, pp. 1215–1223, 2020.
- [48] D.-C. Cheng, T.-C. Hsieh, K.-Y. Yen, and C.-H. Kao, "Lesion-based bone metastasis detection in chest bone scintigraphy images of prostate cancer patients using pre-train, negative mining, and deep learning," *Diagnostics*, vol. 11, no. 3, p. 518, 2021.
- [49] S. Masoudi, S. Mehralivand, S. Harmon, S. Walker, P. A. Pinto, B. J. Wood, D. E. Citrin, F. Karzai, J. L. Gulley, R. A. Madan, W. L. Dahut, P. L. Choyke, and B. Turkbey, "Artificial intelligence assisted bone lesion detection and classification in computed tomography scans of prostate cancer patients." pp. e17567–e17567, 2020.
- [50] Z. Dai, I. Jambor, P. Taimen, M. Pantelic, M. Elshaikh, A. Dabaja, C. Rogers, O. Etala, P. Boström, H. Aronen *et al.*, "Accurate prostate cancer detection and segmentation using non-local mask r-cnn with histopathological ground truth," *International Journal of Radiation Oncology* Biology* Physics*, vol. 111, no. 3, p. S45, 2021.
- [51] X. Liang, P. Hu, L. Zhang, J. Sun, and G. Yin, "Mcfnet: Multi-layer concatenation fusion network for medical images fusion," *IEEE Sensors Journal*, vol. 19, no. 16, pp. 7107–7119, 2019.
- [52] H. Yin, Y. Li, Y. Chai, Z. Liu, and Z. Zhu, "A novel sparse-representation-based multi-focus image fusion approach," *Neurocomputing*, vol. 216, pp. 216–229, 2016.
- [53] J. Xie and Y. Peng, "The head and neck tumor segmentation using nnu-net with spatial and channel 'squeeze amp; excitation' blocks," in *Head and Neck Tumor Segmentation: First Challenge, HECKTOR 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Proceedings*. Berlin, Heidelberg: Springer-Verlag, 2020, p. 28–36. [Online]. Available: <https://doi.org/10.1007/978-3-030-67194-5-3>
- [54] F. Isensee, P. Jaeger, and S. Kohl, "nnu-net: a self-configuring method for deep learning-based biomedical image segmentation," *Nat Methods*, vol. 18, p. 203–211. [Online]. Available: <https://doi.org/10.1038/s41592-020-01008-z>
- [55] "3d u-net: Learning dense volumetric segmentation from sparse annotation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016. MICCAI 2016. Lecture Notes in Computer Science*, O. S., J. L., S. M., U. G., and W. W, Eds. Cham: Springer, vol. 9901. [Online]. Available: <https://doi.org/10.1007/978-3-319-46723-8-49>
- [56] S. Chen, G. Hu, and J. Sun, "Medical image segmentation based on 3d u-net," in *19th International Symposium on Distributed Computing and Applications for Business Engineering and Science*, pp. 130–133,.

- [57] B. Woo and M. Lee, "Comparison of tissue segmentation performance between 2d u-net and 3d u-net on brain mr images," in *2021 International Conference on Electronics, Information, and Communication (ICEIC)*, pp. 1–4,.
- [58] F. Milletari, N. Navab, and S. Ahmadi, "V-net: fully convolutional neural networks for volumetric medical image segmentation," in *2016 Fourth International Conference on 3D Vision (3DV)*. IEEE, p. 565–571.
- [59] A. Soares, T. Lima, R. A. L. Rabêlo, J. Rodrigues, and F. Araujo, "Automatic segmentation of lung nodules in ct images using deep learning," in *2020 IEEE International Conference on E-health Networking, Application and Services (HEALTHCOM)*, pp. 1–6,.
- [60] T. M. Consortium, "Project monai," Dec. 2020. [Online]. Available: <https://doi.org/10.5281/zenodo.4323059>
- [61] A. Hatamizadeh, D. Yang, H. Roth, and D. Xu, "Unetr: Transformers for 3d medical image segmentation," *arXiv preprint arXiv:2103.10504*, 2021.
- [62] T. Zhou, S. Ruan, and S. Canu, "A review: Deep learning for medical image segmentation using multi-modality fusion," *Array*, vol. 3-4, p. 100004, 08 2019.
- [63] B. H. Menze, A. Jakab, S. Bauer, and K.-C. et al., "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)," *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, pp. 1993–2024, 2015.
- [64] B. Wang, J. Yang, H. Peng, J. Ai, L. An, B. Yang, Z. You, and L. Ma, "Brain Tumor Segmentation via Multi-Modalities Interactive Feature Learning," *Frontiers in Medicine*, vol. 8, 2021. [Online]. Available: <https://www.frontiersin.org/article/10.3389/fmed.2021.653925>
- [65] S. R. Muzammil, S. Maqsood, S. Haider, and R. Damaševičius, "CSID: A Novel Multimodal Image Fusion Algorithm for Enhanced Clinical Diagnosis," *Diagnostics*, vol. 10, no. 11, 2020. [Online]. Available: <https://www.mdpi.com/2075-4418/10/11/904>
- [66] Z. Zhu, Y. Chai, H. Yin, Y. Li, and Z. Liu, "A novel dictionary learning approach for multi-modality medical image fusion," *Neurocomputing*, vol. 214, pp. 471–482, 2016.
- [67] J. A. B. Keith A. Johnson, "The whole brain Atlas," <http://www.med.harvard.edu/aanlib/>, 2022.
- [68] Z. Guo, X. Li, H. Huang, N. Guo, and Q. Li, "Medical image segmentation based on multi-modal convolutional neural network: Study on image fusion schemes," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, 2018, pp. 903–907.
- [69] B. Pichler, M. Judenhofer, and C. Pfannenberger, "Multimodal imaging approaches: Pet/ct and pet/mri," *Handbook of experimental pharmacology*, vol. 185, pp. 109–32, 02 2008.

- [70] V. A. O. C. D. F. M. J. M. A. R. H. T. S. Le Nobin J, Rosenkrantz AB, "Image guided focal therapy for magnetic resonance imaging visible prostate cancer: defining a 3-dimensional treatment margin based on magnetic resonance imaging histology co-registration analysis," *The Journal of urology*, vol. 194, no. 2, pp. 364–370, 2015.
- [71] A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, p. 1097–105.
- [72] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation"," in *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pp. 3431–3440,.
- [73] J. Dutta, D. Chakraborty, and D. Mondal, "Multimodal segmentation of brain tumours in volumetric mri scans of the brain using time-distributed u-net," in *Computational Intelligence in Pattern Recognition*, A. K. Das, J. Nayak, B. Naik, S. K. Pati, and D. Pelusi, Eds. Singapore: Springer Singapore, 2020, pp. 715–725.
- [74] T. Yang and J. Song, "An automatic brain tumor image segmentation method based on the u-net," in *2018 IEEE 4th International Conference on Computer and Communications (ICCC)*, 2018.
- [75] M. A. Naser, L. V. van Dijk, R. He, K. A. Wahid, and C. D. Fuller, "Tumor segmentation in patients with head and neck cancers using deep learning based-on multi-modality pet/ct images," in *Head and Neck Tumor Segmentation*, V. Andrearczyk, V. Oreiller, and A. Depeursinge, Eds. Springer International Publishing, 2021.
- [76] T. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. Zitnick, "Microsoft coco: common objects incontext," in *Computer Vision – ECCV2014*, pp. 740–55,.
- [77] K. He, "Mask r-cnn," in *Proceedings of the IEEE International Conference on Computer Vision*, Honolulu, HI, USA, Jul22–25.
- [78] A. Vuola, S. Akram, and J. Kannala, "Mask-rcnn and u-net ensembled for nuclei segmentation," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019, Venice, Italy*, pp. 208–212,.
- [79] Y. Liu, P. Zhang, Q. Song, A. Li, P. Zhang, and Z. Gui, "Automatic segmentation of cervical nuclei based on deep learning and a conditional random field," in *IEEE Access*, vol. 6, pp. 53 709–53 721,.
- [80] S. Ren, "Faster r-cnn: towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal*, no. Intell.39(6), p. 1137–1149.
- [81] T. Gulliver, S. Zahir, and Alzahir, "Image splicing detection using mask-rcnn image splicing detection using mask-rcnn," *Signal Image and Video Processing*, vol. 14, p. 10 1007 11760–020–01636–0.

- [82] Z. Krawczyk, R. Szmurło, P. Zawadzki, E. Kot, J. Starzynski, and A. Zawadzka, "Volumetric modulated arc therapy dose distribution prediction for breast cancer patients: Cnn approach," 07 2021, pp. 1–10.
- [83] K. e, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and pattern recognition*, vol. Jun 26–Jul 1, Las Vegas, NV, USA.
- [84] E. Kot, "Source code." [Online]. Available: <https://github.com/ekote/doses-mask-rcnn>
- [85] A. W. Kass, Michael and D. Terzopoulos, "Snakes: Active contour models," *International Journal of Computer Vision*, vol. 1, no. 3, pp. 321–331, 1988.
- [86] Y. Venkatesh and N. Rishikesh, "Modelling active contours using neural networks isomorphic to boundaries," vol. 3, pp. 1669–1672 vol.3, 1997.
- [87] C. Tianli, H. Chen, and D. N. Metaxas, "Shape-regularized active contour based on dynamic programming for anatomical image segmentation," in *Medical Imaging 2005: Image Processing*, vol. 5747. International Society for Optics and Photonics, 2005, pp. 1–10.
- [88] J.-J. Rousselle, N. Vincent, and N. Verbeke, "Genetic algorithm to set active contour," in *Computer Analysis of Images and Patterns*, N. Petkov and M. A. Westenberg, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, pp. 345–352.
- [89] J. Liu, W. Hwang, M. Chen, J. Tsai, and C. Lin, "Wavelet-based active contour model for object tracking," Oct. 2001, pp. 206–209, iEEE International Conference on Image Processing (ICIP) ; Conference date: 07-10-2001 Through 10-10-2001.
- [90] S. Krinidis and V. Chatzis, "Fuzzy energy-based active contours," *IEEE Transactions on Image Processing*, vol. 18, no. 12, pp. 2747–2755, 2009.
- [91] S. M. Smith, "Fast robust automated brain extraction," *Human Brain Mapping*, vol. 17, no. 3, pp. 143–155, 2002. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/hbm.10062>
- [92] J. Ashburner and K. Friston, "Image segmentation," in *Human Brain Function*, 2nd ed., R. Frackowiak, K. Friston, C. Frith, R. Dolan, K. Friston, C. Price, S. Zeki, J. Ashburner, and W. Penny, Eds. Academic Press, 2003.
- [93] V. Caselles, F. Catté, T. Coll, and F. Dibos, "A geometric model for active contours in image processing," *Numerische Mathematik*, vol. 66, pp. 1–31, 1993.
- [94] V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic active contours," *International Journal of Computer Vision*, vol. 22, no. 1, pp. 61–79, 1997.
- [95] T. Chan and L. Vese, "Active contours without edges," *IEEE Transactions on Image Processing*, vol. 10, no. 2, pp. 266–277, 2001.
- [96] S. Lu, P. Zhang, C. Li, J. Sun, W. Liu, and P. Zhang, "A nim pet/ct phantom for evaluating the pet image quality of micro-lesions and the performance parameters of ct," *BMC Med Imaging*, vol. 21, no. 1, p. 165, Nov 2021.

- [97] E. Ozkaya, R. E. Triolo, F. Rezayaraghi, J. Abderezaei, W. Meinhold, K. Hong, A. Alipour, P. Kennedy, L. Fleysher, J. Ueda, P. Balchandani, M. Eriten, C. L. Johnson, Y. Yang, and M. Kurt, "Brain-mimicking phantom for biomechanical validation of motion sensitive mr imaging techniques," *Journal of the mechanical behavior of biomedical materials*, vol. 122, p. 104680, Oct 2021.
- [98] H. Shaziya, K. Shyamala, and R. Zaheer, "Automatic lung segmentation on thoracic ct scans using u-net convolutional network," in *2018 International Conference on Communication and Signal Processing (ICCSP)*, 2018, pp. 0643–0647.
- [99] E. Kot, Z. Krawczyk, K. Siwek, and P. S. Czwarnowski, "U-net and active contour methods for brain tumour segmentation and visualization," in *2020 International Joint Conference on Neural Networks (IJCNN)*, 2020, pp. 1–7.
- [100] E. Luders, A. W. Toga, and P. M. Thompson, "Why size matters: Differences in brain volume account for apparent sex differences in callosal anatomy: The sexual dimorphism of the corpus callosum," *NeuroImage*, vol. 84, pp. 820–824, 2014. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1053811913009762>
- [101] J. Kim, J. Hong, and H. Park, "Prospects of deep learning for medical imaging," *Precis Future Med*, vol. 2, no. 2, p. 37–52, published online: June 20, 2018, DOI:. [Online]. Available: <https://doi.org/10.23838/pfm.2018.00030>.
- [102] L. Hui, X. Wu, and J. Kittler, "Infrared and visible image fusion using a deep learning framework," in *24th International Conference on Pattern Recognition (ICPR)*, Beijing, pp. 2705–2710,.
- [103] K. Simonyan and A. Zisserman, "Very deep convolutional networks for largescale image recognition," arXiv preprint arXiv:1409.1556,.
- [104] M. Simon, E. Rodner, and J. Denzler, "Imagenet pre-trained models with batch normalization," arXiv preprint arXiv:1612.01452,.
- [105] "Vgg19-bn model implementation." [Online]. Available: <https://github.com/pytorch/vision/blob/main/torchvision/models/vgg.py>
- [106] D. Jha, M. Riegler, D. Johansen, P. Halvorsen, and H. Johansen, "Doubleu-net: A deep convolutional neural network for medical image segmentation," in *2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)*, Rochester, MN, USA, pp. 558–564,.
- [107] "Fusion code." [Online]. Available: <https://github.com/ekote/computer-vision-for-biomedical-images-processing/>
- [108] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," arXiv preprint arXiv:1502.03167.
- [109] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," 2015.

- [110] A. L. Ribeiro, D. J. Pasadas, H. G. Ramos, and T. J. Rocha, "The filtering effect of the tikhonov regularization: Application to eddy-current problems," in *2015 IEEE Metrology for Aerospace (MetroAeroSpace)*, 2015, pp. 188–192.
- [111] E. Kot and M. Derszniak, "Multimodal Fusion - Feature Extractions and Hyper-Params Config Repo," <https://github.com/ekote/IJCNN22>, 2022.
- [112] Y. Zhang, Y. Liu, P. Sun, H. Yan, X. Zhao, and L. Zhang, "IFCNN: A general image fusion framework based on convolutional neural network," *Information Fusion*, vol. 54, pp. 99–118, 2020.
- [113] J. A. O'Reilly and F. Asadi, "Pre-trained vs. random weights for calculating fréchet inception distance in medical imaging," in *2021 13th Biomedical Engineering International Conference (BMEiCON)*, 2021, pp. 1–4.
- [114] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "Gans trained by a two time-scale update rule converge to a local nash equilibrium," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, ser. NIPS'17. Red Hook, NY, USA: Curran Associates Inc., 2017, p. 6629–6640.
- [115] T. Les, T. Markiewicz, S. Osowski, and M. Jesiotr, "Automatic reconstruction of overlapped cells in breast cancer fish images," in *Expert Systems with Applications*, vol. 137.
- [116] E. Kot, "The u-net architecture adjusted to 64x64 input image size," available to download from:. [Online]. Available: <http://bit.ly/unet64x64>
- [117] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [118] J. Bitton and Z. Papakipos, "Augly: A data augmentations library for audio, image, text, and video." <https://github.com/facebookresearch/AugLy>, 2021.
- [119] N. S. Keskar, D. Mudigere, J. Nocedal, M. Smelyanskiy, and P. T. P. Tang, "On large-batch training for deep learning: Generalization gap and sharp minima," *CoRR*, vol. abs/1609.04836, 2016. [Online]. Available: <http://arxiv.org/abs/1609.04836>
- [120] W. Abdulla, "Mask r-cnn for object detection and instance segmentation on keras and tensorflow," https://github.com/matterport/Mask_RCNN, 2017.
- [121] M. R.-C. model, available to download from:. [Online]. Available: <https://github.com/ekote/computer-vision-for-biomedical-images-processing/releases/tag/1.0>,
- [122] E. Kot, Z. Krawczyk, and K. Siwek, "Brain tumour detection and segmentation using deep learning methods," in *Computational Problems of Electrical Engineering*.
- [123] D. Masters and C. Luschi, "Revisiting Small Batch Training for Deep Neural Networks," 2018.

- [124] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," 2018.
- [125] F. Baba, "RSNA-MICCAI Brain Tumor Radiogenomic Classification," <https://github.com/FirasBaba/rsna-resnet10>, 2022.
- [126] M. Zaharia, R. S. Xin, P. Wendell, T. Das, M. Armbrust, A. Dave, X. Meng, J. Rosen, S. Venkataraman, M. J. Franklin, A. Ghodsi, J. Gonzalez, S. Shenker, and I. Stoica, "Apache spark: A unified engine for big data processing," *Commun. ACM*, vol. 59, no. 11, p. 56–65, oct 2016. [Online]. Available: <https://doi.org/10.1145/2934664>
- [127] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2980–2988.
- [128] K. M. Hossein, M. E. Esmaili, T. Dargahi, and A. khonsari, "Blockchain-based privacy-preserving healthcare architecture," pp. 1–4, 2019.
- [129] N. W., J. T., S. S., N. W., C. M., C. R., H. A., and Z. R, "Anonymization of dicom electronic medical records for radiation therapy," *Computers in biology and medicine*, vol. 53, pp. 134–140,.
- [130] P. Jhang, "Data masking in cloud computing," *RESEARCH REVIEW International Journal of Multidisciplinary*, vol. 3085, p. 118–121, 2019.
- [131] R. G.K., M. T.N., and H. R, "Design of data masking architecture and analysis of data masking techniques for testing," *International Journal of Engineering Science*.
- [132] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, p. 2921–2929.
- [133] R. Selvaraju, A. Das, R. Vedantam, M. Cogswell, D. Parikh, and D. Batra, "Grad-cam: Why did you say that? visual explanations from deep networks via gradient-based localization," arXiv preprint arXiv:1610.02391, 2016.
- [134] T. Gundersen and K. Bærøe, "The future ethics of artificial intelligence in medicine: Making sense of collaborative models," *Sci Eng Ethics*, vol. 28, 17. [Online]. Available: <https://doi.org/10.1007/s11948-022-00369-2>
- [135] M. Mewhinney, "Nasa science and engineering achievements to be featured," accessed: 2023-06-01. [Online]. Available: <https://www.nasa.gov/centers/ames/news/releases/2006/0679AR.html>