DELIVERABLE D3.1
State-of-the-Art of AI techniques used for personalized diagnosis

Project number:	ITEA 20044
Document version no.:	v 1.0
Edited by:	LUMC
Date:	29-06-2022

ITEA Roadmap challenge:

Smart Health

HISTORY

Document version #	Date	Remarks
V0.1	29-03-2022	Starting version, template
V0.2	20-06-2022	Compilation of first input by partners
V1.0	29-06-2022	Final version

Deliverable review procedure:

- **2 weeks before due date**: deliverable owner sends deliverable –approved by WP leader– to Project Manager
- **Upfront** PM assigns a co-reviewer from the PMT group to cross check the deliverable
- **1 week before due date**: co-reviewer provides input to deliverable owner
- **Due date:** deliverable owner sends the final version of the deliverable to PM and co-reviewer

TABLE OF CONTENTS

1 Glossary

AI	Artificial Intelligence
AUC	Area under the ROC curve
BPH	Benign Prostatic Hyperplasia
CBCT	Cone Beam Computed Tomography
CG	Central Gland
CIU	Contextual Importance and Utility
CNN	Convolutional Neural Network
CT	Computed Tomography
DCNN	Deep Convolutional Neural Network
DSA	Digital Subtraction Angiography
DSBN	Domain-Specific Bulk Normalization
DSC	Dice Similarity Coefficient
EDH	Extradural Hemorrhage
FCN	Fully Convolutional Network
GDPR	General Data Protection Regulation
GGG	Gleason grade groups
GNN	Graph Neural Network
ICA	Inter Cranial Arteries
ICH	Acute Intracranial Hemorrhage
LUTS	lower urinary tract symptoms
MR	Magnetic Resonance
MRI	Magnetic Resonance Imaging
MS-Net	Multisite Network
PAE	Prostate artery embolization
PDAC	Pancreatic Ductal Adenocarcinoma
PZ	Peripheral Zone
QoL	Quality of Life
SDH	Subdural Hemorrhage
TURP	Transurethral resection of the prostate
VGG	Visual Geometry Group
XAI	Explainable Artificial Intelligence

2 Introduction

2.1 Aim of the activity

This deliverable focusses on improving personalized diagnosis by Integrating all relevant data streams, in particular biomedical images because they have the capability to reflect the underlying pathophysiology, into mineable high-dimensional patient-specific datasets,. Through intelligent analysis of this data, using state-of-the-art machine learning techniques personalized (and precise) diagnosisis possible.

In this report the state of the art for the use Artificial Intelligence (AI) techniques for personalized diagnosis is reported for different use-cases of the ASSIST project. The following sections each discuss the use of AI techniques for a specific use-case.

Section 2 discusses the use of deep-learning to segment the prostate and surrounding arteries which is important for prostate artery embolization which is a minimally invasive treatment option for prostate enlargement.

Section 3 discusses the use of deep convolutional neural networks and explainable AI for the automated detection, segmentation and classification of intracranial hemorrhage.

Section 4 discusses the importance of tumor segmentation in radiology images and the current state of the art deep-learning architectures used in automatic brain tumor segmentation.

Section 5 discusses how deep learning is seen as the best approach to detect pulmonary nodules in CT images and how it might also aid in the analysis of pathology images of nodule biopsy samples.

Section 6 discusses different deep learning architectures for the segmentation of the liver and the pancreas as well as tumors inside these organs.

Finally in Section 7 we will draw some conclusions regarding the use of AI, in particular deep learning, in personalized diagnosis in the clinical use-cases of the ASSIST project.

2.2 Contributors

Several authors contributed to the production of this document. Each of those authors was responsible for one of the clinical use-cases.

Contribution	Authors
Prostate Enlargement	Fortearge

Intracranial Hemorrhage	Innova
Brain Tumors	Linkoping University
Lung diseases	Philips, Thirona
Hepato Pancreato Biliary Oncology	LUMC
Global editing	LUMC
Reviewing	Philips

3 Prostate Enlargement

3.1 Introduction

Prostate enlargement, also called benign prostatic hyperplasia (BPH), is a benign enlargement of the prostate BPH, or the proliferation of glandular and stromal tissue in the transition zone of the prostate, can lead to lower urinary tract symptoms (LUTS) and bladder outlet obstruction.

The prevalence of LUTS increases with age, and 25% of men over 70 years old have moderate to severe LUTS that significantly affect their Quality of Life (QoL). BPH affects about 105 million men globally. It usually starts after the age of 40, half of males aged 50 or over are affected the incidence of male-pattern baldness increases with age.
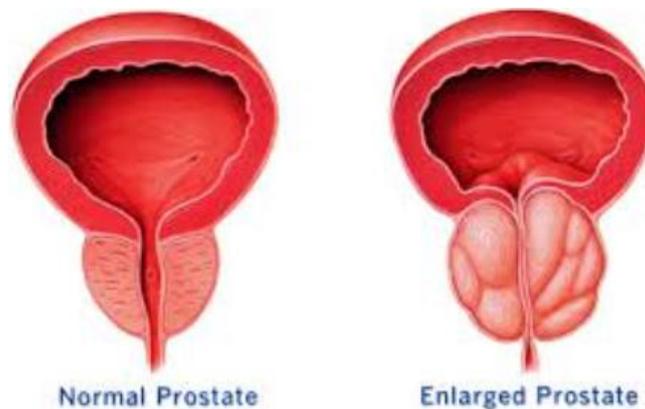
Figure 1: Normal vs Enlarged Prostate (A1, 2014)

There are many different medical and surgical options available for the treatment of BPH with LUTS. In patients with moderate to severe LUTS who are not responsive to medical management, more invasive treatments may be considered. Transurethral resection of the prostate (TURP) and open prostatectomy (OP) are the most effective treatments for large prostate glands. However, these procedures have significant morbidity rates including retrograde ejaculation, erectile dysfunction, urethral strictures, urinary retention, blood transfusion requirements and incontinence in patients with existing comorbidities, increasing age and large prostate volume are associated with higher complication rates, which limits the eligibility for surgical therapies.

Prostate artery embolization (PAE) is a minimally invasive treatment option that has a lower risk of urinary incontinence and sexual side effects. The PAE procedure involves delivering embolic materials to block the blood vessels supplying the hypertrophied transitional zone in the prostate gland. This will reduce the size of the gland and prevent it from growing further. In order for PAE treatment to be successful, a thorough analysis of the patient should be carried out before the procedure. For the procedure to be technically successful, accurate determination of the anatomy of the prostate arteries and adequate embolization of the target are required. This procedure

should avoid off-target embolization of other tissues, soft tissues of the bladder, rectum and penis, and other critical pelvic structures.

When considering PAE the technique is the challenging part requiring experience in recognizing prostatic arteries and avoiding non target embolization. As male internal iliac anatomy is prone to variations so are prostatic arteries which they can vary in origins, number. Besides, usually during treatment, cone beam CT is required especially with less experienced operators/angiographers. Many different protocols exist among centers for workup before the procedure to recognize and plan treatment including pre-operative CT imaging, MR imaging or no pre-imaging at all.

Studies in the field of prostate artery embolization and AI are limited in the literature. The important thing in PAE is the detection of the prosthesis artery, so after a study on PAE, we will focus on the studies on artery classification and artery identification.

Gurgitano et al. investigated the effect of artificial intelligence on application areas in interventional radiology. In their research, they have sh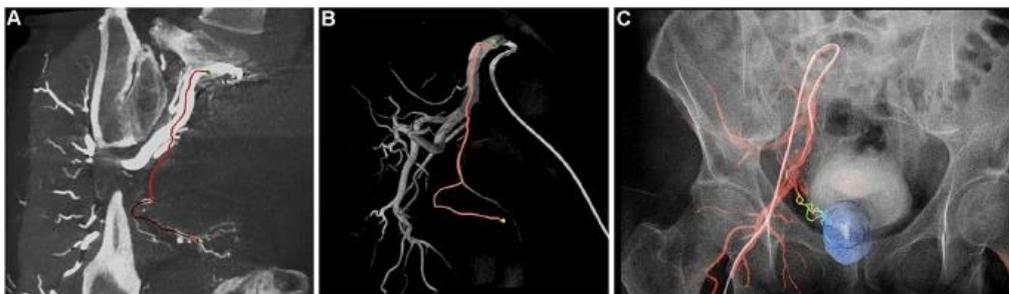own what the use of artificial intelligence in various application areas is good for. In the Prostate Artery Embolization method used in the treatment of benign Prostate Hyperplasia disease, it has been shown that artery detection can be made from 3D CBCT images with artificial intelligence application (see Figure 5). (Gurgitano, 2021)

Figure 2: "Automatic 3D detection of prostatic arteries using Cone-Beam CT during Prostatic Arterial Embolization"—a CBCT identification of prostatic arteries; b Realization of 3D roadmap; c Overlap on fluoroscopic images" (Gurgitano, 2021)

They stated that virtual 3D anatomical data can be obtained using augmented reality and AI-based CT, CBCT or MRI. It is a combination of real-world 2D visual images that create a virtual device trajectory superimposed on visual surface anatomy. Theoretically, they explained that accurate navigation can be achieved without the need for fluoroscopy. (Gurgitano, 2021)

Thanks to the integrated mapping AI software, automatic landmark recognition and motion compensation can be activated using reference marks linked by a computer algorithm. This system can be applied in lesion targeting/localization, spinal/paraspinal injections, arthrograms, tumor ablation, bone biopsies and more recently minimally invasive surgical procedures. (Gurgitano, 2021)

Chen et al. constructed a comprehensive dataset with 729 Magnetic Resonance Angiography scans and proposed a Graph Neural Network (GNN) method to label arteries by classifying the types of nodes and edges in an ascribed associative graph. Additionally, they developed a hierarchical improvement framework to further refine the GNN outputs to incorporate structural and relational information about Intra Cranial Arteries (ICA). The GNN developed in the application takes a graph with node and

edge properties as input and returns a graph with additional properties for the node and edge types as output. In addition to its superior performance compared to the methods described in the literature, the study demonstrated robustness and generalizability over a variety of challenging anatomical variations. (Chen)

Li et al proposed a two-stage vessel classification to improve the performance of existing automated methods for retinal image analysis. They adopted a UNet-based model, SeqNet, to accurately segment vessels from the background and predict vessel type. SeqNet mainly consists of two streams, bottom and top. Upstream is for segmentation. IterNet is adopted, which iteratively improves segmentation results with smaller UNets after initial segmentation. Adam was used as the optimizer. Two popular general datasets, DRIVE and LES-AV were used. It has been experimentally shown that the method achieves the most advanced performance on two general datasets, including SeqNet and postprocessing. (Li, 2020)

As a result of our research, the scarcity of studies using prostate Digital Subtraction Angiography (DSA) image has attracted attention. Due to the scarcity of data on our research topic, we will benefit from it in our own study by examining different studies that may be useful for our project. We hope that the following literature studies will guide us in our application, which will be used in the diagnosis and planned treatment of Benign Prostate Hyperplasia treatment to be used in the ASSIST project.

3.2 Prostate Segmentation

There is no evidence that prostate treatments using DSA imaging techniques can be used to treat Benign Prostate Hyperplasia (BPH). For this reason, we looked at the studies that were conducted using different imaging techniques. Below, we've listed some of the most important research into artificial intelligence that we think will be helpful for our own project.

Liu et al. aimed to perform prostate segmentation from heterogeneous multisite MRI data. They developed a multisite network (MS-Net) for prostate segmentation. They proposed a Domain-Specific Bulk Normalization (DSBN) layer in the network backbone to minimize heterogeneity in images. This allowed the network to estimate statistics and feature normalization for each region separately. An adapted 2D Residual-UNet is adopted as the segmentation network backbone, providing remarkable performance in prostate segmentation problem. Experiments by Liu et al. show the superiority of the approach. (Liu Q. D., 2020)

Liu et al presented a new shape-sensitive meta-learning scheme to improve model generalization in prostate MRI segmentation. The learning scheme is based on gradient-based meta-learning by explicitly simulating field shift with virtual meta-training and meta-testing during training. In this study, it is aimed to improve the meta-optimization by considering the shortcomings encountered when applying a segmentation model to invisible areas, especially the shape compactness and shape smoothness of segmentations under simulated area shift. In the experiment, prostate T2-weighted MRI from 6 different data sources with distribution shift was used. An adapted Mix-residual-UNet has been implemented as the segmentation backbone. Due to the large differences in slice thickness between different regions, 2D architecture was used. (Liu Q. D.)

Gillespie et al have listed the latest technological advances in prostate segmentation and provided insight into the field by discussing the limitations and strengths of MR prostate segmentation and proposed an optimized 2D U-Net for MR prostate segmentation. According to research by Gillespie et al., (Gillespie)
• Yu et al. (2017) improved this by adding residual links to the 3D segmentation network that helped improve prostate segmentation. It also used an aggregation process instead of aggregation features to the upsampling layer, transforming it into a ResNet-U-Net hybrid.
• To improve the accuracy of prostate segmentation, Zhu et al. (2019b) suggested the use of a staggered U-Net. It emerged from the first mesh, segmented the entire prostate gland, and the segmented gland was fed into the second mesh to compartmentalize the peripheral region.
• Liu et al. (2019b) proposed a network to segment prostate regions using FCN with a feature pyramid attention mechanism. They used a feature pyramid network and a simple decoder and a ResNet50 as the backbone of their network to capture features at multiple scales.

Gillespie et al. implemented a 2D U-Net architecture using Ranger optimizer Wright (2019) and Mish Activation Misra (2020) to segment the prostate from MRI data. We analyzed the capability of these minor changes in U-Net configuration and its impact on performances in four publicly available datasets, namely Promise 12, Prostate X, NIC ISBI 2013 and Decathlon Medical Dataset. Models trained on each data set and a combination of all were evaluated in a test set with Dice Similarity Coefficient (DSC). DSC scores obtained in separate test sets were calculated for each data set. It is observed that the model trained with all data outperforms all other models with higher DSC scores. This research provides a new perspective on MR prostate segmentation and, importantly, provides standardized experimental settings for researchers to evaluate their algorithms. (Gillespie)

Pellicer-Valero et al propose a fully automated system based on Deep Learning that takes prostate mpMRI from a patient with suspected prostate cancer and uses the Retina U-Net detection framework to locate, segment and predict the most probable Gleason grade groups (GGG) of prostate cancer lesions. In this study, it was developed for automatic segmentation of the central gland (CG) and peripheral zone (PZ), which are defined as the two main regions of the prostate. Uses 490 mpMRI for training/validation and 75 patients for testing from the ProstateX and IVO dataset.There is a data preprocessing stage, which consists of the first CNN where the images are entered and the second CNN which outputs the first CNN as the CG segmentation mask. After preprocessing the data, it was used to train a Retina U-Net CNN architecture that allows simultaneous detection, segmentation and classification of prostate cancer lesions. The Retina U-Net architecture combines the Retina Net detector with U-Net segmentation CNN and is specifically designed for application to medical images. (Pellicer-Valero OJ, 2022)

3.3 Artery Segmentation

As we mentioned before, there is no BPH study using DSA imaging technique. To aid our study of vessel segmentation, we reviewed a vessel segmentation study using a DSA image of brain blood vessels.

Zhang et al have presented a deep learning approach to automatically segment brain blood vessels in DSA. In this study, a U-net was used to detect brain blood vessels in DSA and tested on real DSA images.

In the studies, rotation, translation and scaling were applied to the image data for magnification.
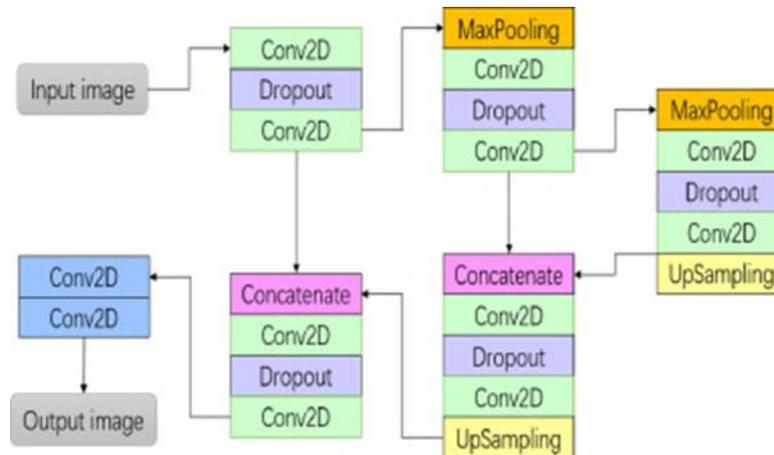
Figure 3: Structure of deep learning network (Zhang, 2020)

They used a 12-layer U-net structure to segment DSA images and trace brain blood vessels by calculating the probability that each pixel in the original image was part of a blood vessel. The structure of the deep learning network is shown in Figure 2.
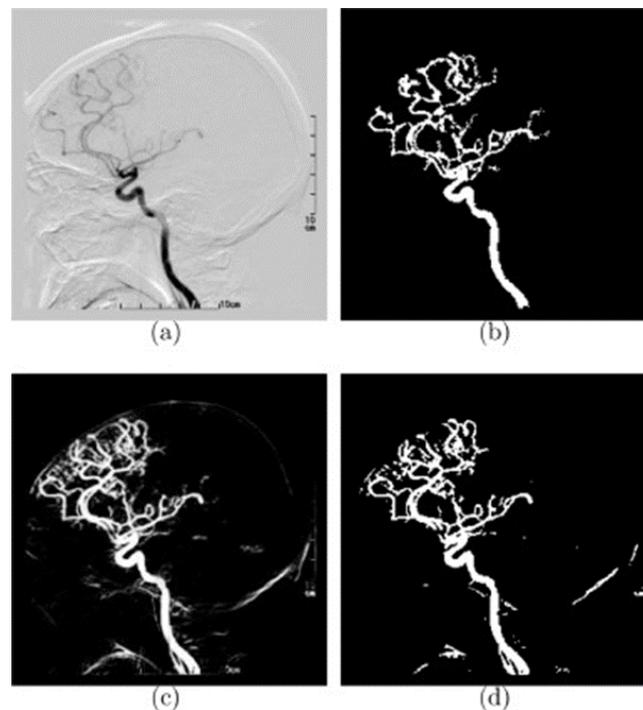
Figure 4: (a), an original DSA image. (b) manually marked ground truth. (c) Probability map given by deep learning network with pixel assigned to blood vessels class. (d) the final result of our approach after thresholding the probability map. (Zhang, 2020)

Figure 3(a) shows an original DSA image. The hand-marked ground reality is shown in Figure 3(b). The segmentation result in the form of probability map given by the trained

deep learning network is shown in Figure 3(c). From this example, it has been observed that the deep learning network can faithfully segment blood vessels of many different sizes, from the large blood vessel in the middle of the image to the small vessels near the skull. The final result is shown in Figure 3(d) with the threshold of the probability map set to 0.5. The result shown in Figure 3(d) is significantly improved compared to Figure 3(c). Figure 4 shows the overlay of the segmentation result on the original image. *(Zhang, 2020)*

Figure 5: The segmentation result is superimposed on the original DSA image. (Zhang, 2020)

3.4 Artery Identification

Identifying the artery for prostate embolization is one of the most important points in our use case. For this reason, we aim to research studies that define arteries and to benefit from the parts that may be useful in our own project.

Pu et al present a new integrative computed solution to automatically identify and differentiate pulmonary arteries and veins shown on lung computed tomography (CT) without iodinated contrast agents. They first described the central extrapulmonary arteries and veins using a convolutional neural network (CNN) model. They then used a computational differential geometry method to automatically identify tubular-like structures in the lungs with high density, which we believe are intrapulmonary vessels. A dataset of 120 chest CT scans obtained on different subjects using various protocols was used to develop, train and test the algorithms. CT scans without iodinated contrast agents were randomly selected by the Cancer Imaging Archive (TCIA) Lung Image Database Consortium (LIDC) and the Image Database Resource Initiative (IDRI) (LIDC-IDRI). The computer algorithm achieved a sensitivity of ~98% in labelling pulmonary artery and vein branches compared with the results of a human expert, demonstrating the feasibility of the computerized solution for pulmonary artery/vein labelling. (Pu, 2022)
The scheme developed consisted of four main components (Fig. 2): (1) defining extrapulmonary arteries and vessels using a U-Net architecture, (2) defining intrapulmonary vessels using a computational differential geometry solution, (3) skeletonizing intrapulmonary vessels that guide the tracing of adjacent vessel branches, and (4) tracing the skeletons of intrapulmonary vessels to differentiate between arteries and veins, starting from the extrapulmonary arteries and veins. (Pu, 2022)
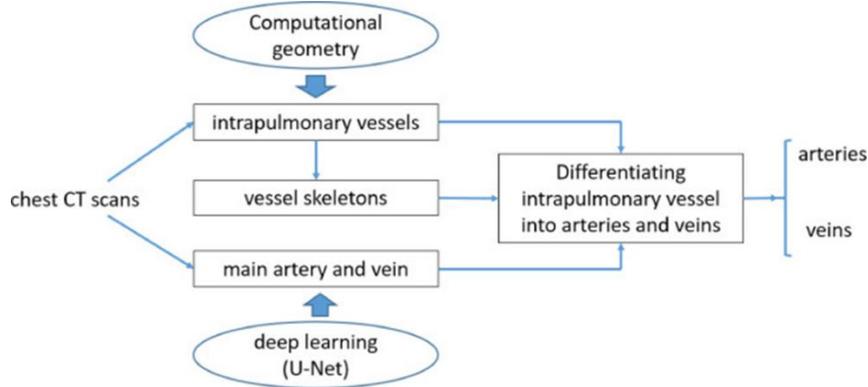
ASSIST
ITEA 20044
WP3 Deliverable D3.1
Page 14 of 35

Figure 6: Schematic flow chart to identify pulmonary arteries and veins (Pu, 2022)

Various U-Net models have been applied and tested to segment the central extrapulmonary arteries and veins, including the classic U-Net, R2Unet, Attention U-Net and U-Net ++.

Small vessels are progressively labelled as arteries or veins in the lungs, starting from the extrapulmonary veins. Next, the algorithm automatically segmented extrapulmonary arteries and veins and correlated strongly with manual segmentation by a radiologist.

3.5 References

A1, F. (2014).

Chen, L. H.-N. (2020). Automated Intracranial Artery Labeling using a Graph Neural Network and Hierarchical Refinement.

Gillespie, D. K. (2020). Deep learning in magnetic resonance prostate segmentation: A review and a new perspective.

Gurgitano, M. A. (2021). Interventional Radiology ex-machina: impact of Artificial Intelligence on practice. *La Radiologia Medica*, 998–1006.

Li, L. V. (2020). Joint Learning of Vessel Segmentation and Artery/Vein Classification with Post-processing. *In Proceedings of Machine Learning Research*.

Liu, Q. D. (2020). MS-Net: Multi-Site Network for Improving Prostate Segmentation With Heterogeneous MRI Data. *IEEE Transactions on Medical Imaging*, 2713–2724.

Liu, Q. D. (2020). Shape-aware Meta-learning for Generalizing Prostate MRI Segmentation to Unseen Domains.

Pellicer-Valero OJ, M. J.-P. (2022). Deep learning for fully automatic detection, segmentation, and Gleason grade estimation of prostate cancer in multiparametric magnetic resonance images. *Scientific Reports*.

Pu, J. L. (2022). Automated identification of pulmonary arteries and veins depicted in non-contrast chest CT scans. *Medical Image Analysis*.
Zhang, M. Z. (2020). A neural network approach to segment brain blood vessels in digital subtraction angiography. *Computer Methods and Programs in Biomedicine*.

4 Intracranial hemorrhage

4.1 Introduction

Intracranial hemorrhage refers to any bleeding within the intracranial vault, including the brain parenchyma and surrounding meningeal spaces (Caceres & Goldstein, 2012). Acute intracranial hemorrhage (ICH) is a potentially life-threatening condition that requires fast and accurate detection because of its frequently rapid progression during the first several hours.

Intracranial hemorrhage (ICH), a subtype of stroke, can be classified into five sub-types according to bleeding location: Intraventricular (IVH), Intraparenchymal (IPH), Subarachnoid (SAH), Epidural (EDH) and Subdural (SDH). The ICH that occurs within the brain tissue is called Intracerebral Hemorrhage (Figure 7). Although ICH are less frequent than ischemic stroke, it presents higher mortality rate. The degrees of severity and interventions vary with bleeding types (Ye et al., 2019).
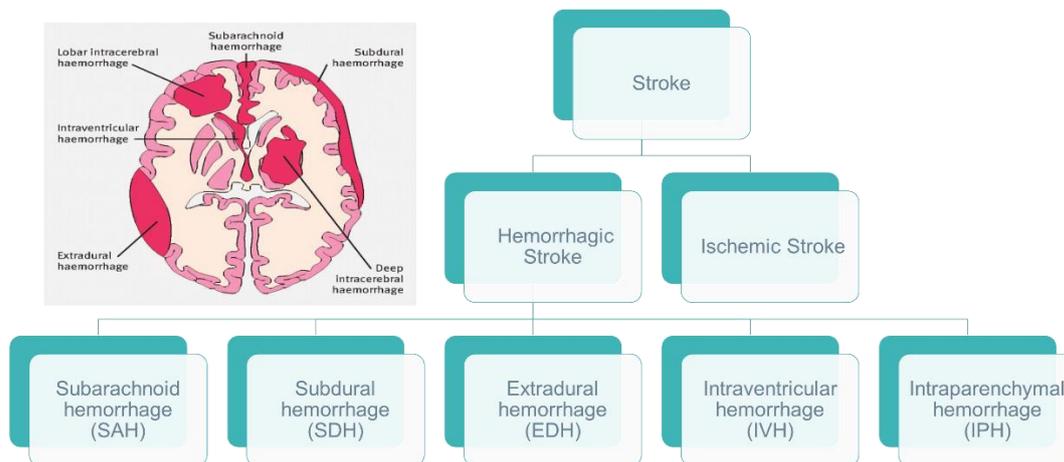
Figure 7: Sub-types of stroke and hemorrhagic stroke

Classification of ICH and distinguishing it from ischemic stroke is critical due to prompt appropriate treatment and mitigate neurological deficit, and mortality. In ischemic strokes, therapy with drugs that can break up a clot has to be given within 4.5 hours from when symptoms first started if given intravenously. Intravenous tissue-type plasminogen activator (IV-tPA) is the gold standard treatment for ischemic stroke. It improves outcomes in ischemic stroke but is associated with certain risks such as potential bleeding in the brain. Differentiating extradural hemorrhage from subdural (SDH) hemorrhage in the head is also important. While extradural hemorrhage is treated with expedient evacuation via a craniotomy, SDH has various management strategies depending on the size, location and extent of mass effect.

4.2 ICH Diagnosis

Non-contrast Computed Tomography (CT) scan is usually the first imaging method used to assess patients with suspected ICH and distinguish ICH from ischemic stroke as it can be performed fast and has high sensitivity for hemorrhage. Hemorrhage and its sub-

types can be recognized on non-contrast CT since blood has slightly higher density (Figure 8). CT scans generate a sequence of images using X-ray beams. Depending on the amount of tissue X-ray absorbency, brain tissues are captured with different intensities. CT scans are displayed using a windowing method. Different features of the brain tissues can be displayed in the grayscale image by selecting different window parameters. In the CT scan images, the ICH regions appear as hyperdense regions with a relatively undefined structure (Hssayeni et al., 2020). However, there are difficulties in using CT scan to detect hemorrhages due to their similar appearance with the parenchyma and complexity in distinguishing mass effect and edema (Mirza & Gokhale, 2017). Even highly trained experts may miss subtle life-threatening findings and many hospitals do not have trained neuro-radiologists, especially at night and on weekend.
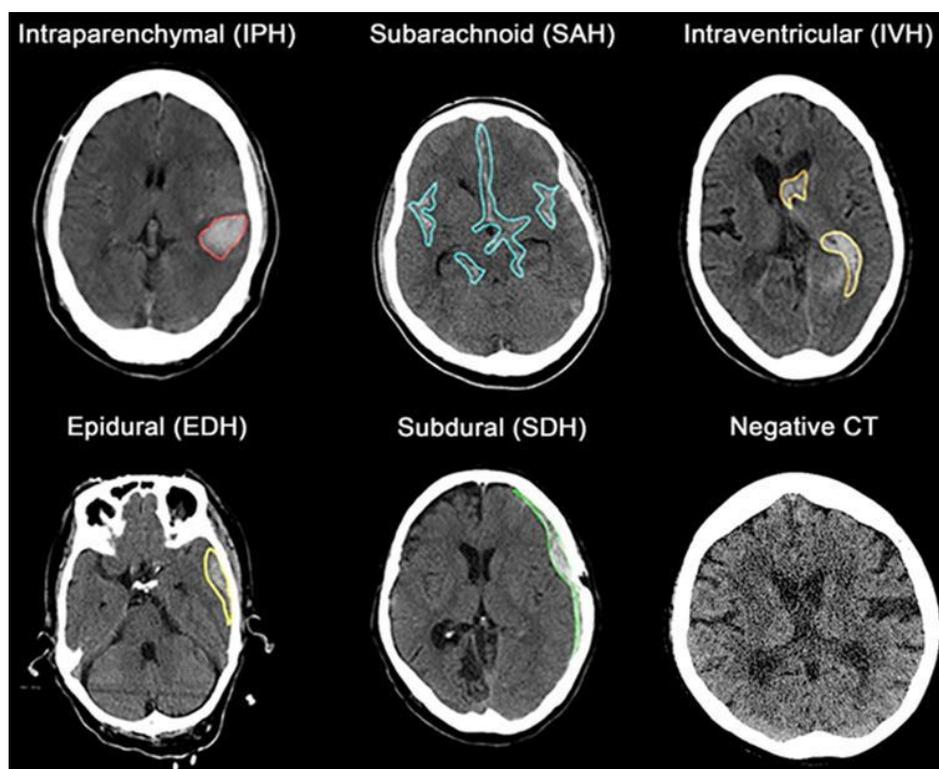
Figure 8: Non-contrast Computed Tomography scans for ICH sub-types. Note. Reprinted from "Clinical usefulness of deep learning-based automated segmentation in intracranial hemorrhage", 29(5), pp.881-895

Interpretation of non-contrast CT images is difficult due to the following challenges:
- Image noise, artefacts and cerebral parenchyma with similar appearance and density make segmentation of ICH challenging
- Differentiating extradural (EDH) from subdural (SDH) hemorrhage in the head can be challenging as SDHs are more common and there are a few distinguishing features which are usually reliable
- Gray scale images are limited by low signal-to-noise, poor contrast, and a high incidence of image artifacts. A unique challenge is to identify tiny subtle abnormalities in a large 3D volume with near-perfect sensitivity

4.3 Deep Neural Networks in ICH detection and classification

Recent advances in deep convolutional neural networks (DCNN) have showed that the method has a great potential in automating ICH detection and segmentation and can assist junior radiology trainees when experts are not available. DCNNs with their capability of self-learning of nonlinear image filters and the self-extraction of relevant features are superior to methods that demand complicated engineering feature including skull stripping, image registration, and feature extraction from voxel intensity and local moment information (Muschelli et al., 2017; Ye et al., 2019).

Lee et al. (2018) proposed a high-performance system for the detection and classification of ICH system from small and imbalanced data using ImageNet pretrained DCNNs of VGG167, ResNet-508, Inception-v39 and Inception-ResNet-v210. The system achieved a performance similar to that of expert radiologists (sensitivity of 98% and specificity of 95%). A method based on 3D joint convolutional and recurrent neural networks was able to accurately detect ICH and its subtypes (> 0.8 AUC across all subtypes) with fast speed (< 30s), suggesting its potential for assisting radiologists and physicians in their clinical diagnosis workflow (Ye et al., 2019). Kuo et al. (2019) demonstrated that a fully convolutional network trained with 4,396 head CT scans could detect ICH with high accuracy (> 0.99 AUC). Cho et al. (2019) reported 80.19% precision and 82.15% recall with their cascade deep learning model constructed using two convolutional neural networks (CNNs) and dual fully convolutional networks (FCNs). Nemcek, Jakubicek and Chmelik (2020) developed CNN based classifiers with a designed cascade parallel architecture that enables localization and classification of ICHs with average Jaccard coefficient of 53.7%.

4.4 Explainable AI in ICH diagnosis

In most clinical centers, initial interpretations of head CT is usually provided by junior radiologists, radiology trainees, or emergency physicians and initial interpretations will be reviewed later by senior or more-experienced radiologists. Several studies have confirmed that discrepancies exist between the initial and final interpretations and some misinterpretations might even cause clinical consequences. Diagnosis process relies on the availability of a subspecialty-trained neuroradiologist, and as a result, could be time inefficient and even inaccurate, especially in remote areas where specialized care is scarce (Patel et al., 2019; Burduja, Ionescu, & Verga, 2020; Unnithan & Mehta, 2022; Ye et al., 2019; Hssayeni et al., 2020). Visualizing the model decision and increasing interpretability is especially helpful for users with insufficient experience with ICH. However, there are a few studies in literature that aims to enhance interpretability of ICH detection models. Lee et al. (2018) used an attention map and prediction basis retrieved from training data. Alis et al. (2022) implemented a modified version of Gradient-based class activation maps, a well-established saliency map generating method.

Explainable AI (XAI) aims to to shift the traditional black-box approach to a white-box one for greater transparency, interpretability, and explainability. Medical diagnosis and treatment selection are responsible for human life and healthcare professionals need to

be confident enough to treat a patient as instructed by a black-box model. From perspective of healthcare providers "omitting explainability in clinical decision support systems poses a threat to core ethical values in medicine and may have detrimental consequences for individual and public health" (Amann et al, 2020). XAI is critical not only for clinicians but also for patients and for any stakeholder in healthcare. Regulations like the European General Data Protection Regulation (GDPR) are making it harder for the use of black-box models in healthcare as retraceability of the decisions is now a requirement. Explainability is the key to safe, ethical, fair, and trust-able use of AI and a key enabler for its deployment in the real world.

There is often a perceived trade-off between the performance of a model and its ability to produce explainable predictions (Antoniadi et al, 2021). Prediction accuracy is usually the first requirement of AI systems in medicine and currently AI models in healthcare are often developed with only predictive performance. Therefore, the majority of the medical XAI literature is devoted to explaining the previously developed model.

XAI has a great potential to increase trust and lead to the adoption of deep learning methods in medical imaging where explanation is defined as a set of domain features such as pixels of an image that contribute to the output decision of the model. Standard attribution-based methods and architecture or domain specific techniques are two types of broadly used approaches to explain the results of DNNs in medical imaging (Singh, Sengupta, & Lakshminarayanan, 2020).

The goal of an attribution method is to determine the contribution of an input feature to the target neuron which is usually the output neuron of the correct class for a classification problem. The arrangement of the attributions of all the input features in the shape of the input sample forms heatmaps known as the attribution maps. The attribution methods can be applied on a black box convolutional neural network (CNN) without any modification to the underlying architecture making them a convenient yet powerful Explainable AI (XAI) tool.

Importance scores, decision rules, decision trees, dependency plots are the most common types of explanation families that enable information content can easily understandable by end users. Importance scores (aka saliency heatmaps) are perhaps the most common type of explanation families. For instance, Lundberg and Lee proposed SHAP (SHapley Additive exPlanations), a unified framework for generating post-hoc local explanations in the form of additive feature attribution. Local Interpretable Model-Agnostic Explanation (LIME) is another well-validated, model-agnostic local XAI approach that can provide an explanation for a complex deep learning model in the neighborhood of an instance. LIME method can explain each individual prediction by investigating contribution of each pixel (Yang, Ye, & Xia, 2022). Gradient weighted class activation mapping produces activation maps using the gradients of the target concept as it flows to the final convolutional layer but can only be applied to CNNs (Singh, Sengupta, & Lakshminarayanan, 2020).

Contextual Importance and Utility (CIU), which does not build any intermediate interpretable model like LIME, make it possible explain results of any AI system with any level of abstraction using semantics that are independent of the internal mechanism of AI system and can provide more expressive and flexible explanations than LIME and Shapley values (Främling et al, 2021). The use of CIU for image recognition and importance scores are also promising.

Performance of these explainable methods vary in terms of their time needed for generating explanations. LIME and SHAP need around 11 and 10 seconds per image respectively. In comparison to SHAP and IME, the running time of CIU method is about 8.5 seconds per image (Knapič, Malhi, Saluja, & Främling, 2021).

4.5 References

Alis, D., Alis, C., Yergin, M. et al. (2022) A joint convolutional-recurrent neural network with an attention mechanism for detecting intracranial hemorrhage on noncontrast head CT. Sci Rep 12, 2084 (2022). https://doi.org/10.1038/s41598-022-05872-x

Amann, J., Blasimme, A., Vayena, E., Frey, D., & Madai, V. I. (2020). Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC Medical Informatics and Decision Making, 20(1). https://doi.org/10.1186/s12911-020-01332-6

Antoniadi, A. M., Du, Y., Guendouz, Y., Wei, L., Mazo, C., Becker, B. A., & Mooney, C. (2021). Current Challenges and Future Opportunities for XAI in Machine Learning-Based Clinical Decision Support Systems: A Systematic Review. Applied Sciences, 11(11), 5088. https://doi.org/10.3390/app11115088

Burduja, M., Ionescu, R. and Verga, N. (2020). Accurate and Efficient Intracranial Hemorrhage Detection and Subtype Classification in 3D CT Scans with Convolutional and Long Short-Term Memory Neural Networks. Sensors, 20(19), p.5611.

Caceres, J.A., & Goldstein, J.N. (2012). Intracranial Hemorrhage. Emerg Med Clin North Am. 30(3), 771–794. https://doi.org /10.1016/j.emc.2012.06.003

Cho, J., Park, K., Karki, M., Lee, E., Ko, S., Kim, J., Lee, D., Choe, J., Son, J., Kim, M., Lee, S., Lee, J., Yoon, C. and Park, S. (2019). Improving Sensitivity on Identification and Delineation of Intracranial Hemorrhage Lesion Using Cascaded Deep Learning Models. Journal of Digital Imaging, 32(3), pp.450-461.

Främling, K., Westberg, M., Jullum, M., Madhikermi, M., & Malhi, A. (2021). Comparison of Contextual Importance and Utility with LIME and Shapley Values. Explainable and Transparent AI and Multi-Agent Systems, 39–54. https://doi.org/10.1007/978-3-030-82017-6_3

Hssayeni, M., Croock, M., Salman, A., Al-khafaji, H., Yahya, Z. and Ghoraani, B. (2020). Intracranial Hemorrhage Segmentation Using a Deep Convolutional Model. Data, 5(1), p.14.

Kim, C., Hahm, M., Lee, D., Choe, J., Ahn, J., Park, S., Lee, S., Kwak, Y., Yoon, S., Kim, K., Kim, M., Chang, S., Son, J., Cho, J., Park, K. and Kim, J. (2021). Clinical usefulness of deep learning-based automated segmentation in intracranial hemorrhage. Technology and Health Care, 29(5), pp.881-895.

Knapič, S., Malhi, A., Saluja, R., & Främling, K. (2021). Explainable Artificial Intelligence for Human Decision Support System in the Medical Domain. Machine Learning and Knowledge Extraction, 3(3), 740–770. https://doi.org/10.3390/make3030037

Kuo, W., Häne, C., Mukherjee, P., Malik, J. and Yuh, E. (2019). Expert-level detection of acute intracranial hemorrhage on head computed tomography using deep learning. Proceedings of the National Academy of Sciences, 116(45), pp.22737-22745.

Lee, H., Yune, S., Mansouri, M., Kim, M., Tajmir, S., Guerrier, C., Ebert, S., Pomerantz, S., Romero, J., Kamalian, S., Gonzalez, R., Lev, M. and Do, S. (2018). An explainable deep-learning algorithm for the detection of acute intracranial haemorrhage from small datasets. Nature Biomedical Engineering, 3(3), pp.173-182.

Mirza, S., & Gokhale, S. (2017). Neuroimaging in Intracerebral Hemorrhage. In A. Agrawal (Ed.). Hemorrhagic Stroke - An Update. https://doi.org/10.5772/67303. Available from: https://www.intechopen.com/books/hemorrhagic-stroke-an-update/neuroimaging-in-intracerebral-hemorrhage

Muschelli, J., Sweeney, E., Ullman, N., Vespa, P., Hanley, D. and Crainiceanu, C. (2017). PItcHPERFeCT: Primary Intracranial Hemorrhage Probability Estimation using Random Forests on CT. NeuroImage: Clinical, 14, pp.379-390.

Nemcek, J., Jakubicek, R. and Chmelik, J. (2020). Localization and Classification of Intracranial Hemorrhages in CT Data. 8th European Medical and Biological Engineering Conference, pp.767-773.

Patel, A., Schreuder, F., Klijn, C., Prokop, M., Ginneken, B., Marquering, H., Roos, Y., Baharoglu, M., Meijer, F. and Manniesing, R. (2019). Intracerebral Haemorrhage Segmentation in Non-Contrast CT. Scientific Reports, 9(1).

Singh, A., Sengupta, S., & Lakshminarayanan, V. (2020). Explainable Deep Learning Models in Medical Image Analysis. Journal of Imaging, 6(6), 52. https://doi.org/10.3390/app11115088

Unnithan AKA, Mehta P. Hemorrhagic Stroke. [Updated 2022 Feb 5]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK559173/

Yang, G., Ye, Q., & Xia, J. (2022). Unbox the black-box for the medical explainable AI via multi-modal and multi-centre data fusion: A mini-review, two showcases and beyond. Information Fusion, 77, 29–52. https://doi.org/10.1016/j.inffus.2021.07.016

Ye, H., Gao, F., Yin, Y., Guo, D., Zhao, P., Lu, Y., Wang, X., Bai, J., Cao, K., Song, Q., Zhang, H., Chen, W., Guo, X. and Xia, J. (2019). Precise diagnosis of intracranial hemorrhage and subtypes using a three-dimensional joint convolutional and recurrent neural network. European Radiology, 29(11), pp.6191-6201.

5 Brain tumors

5.1 Introduction

Brain tumors compose about 2% of the cancer incidences, affect some 300,000 subjects globally each year (Leece et al., 2017), with a low survival rate and a high morbidity for the patients. Though not being the most prevalent cancer type, brain tumors are prone to complicated and challenging treatment procedures that are often a combination of surgery, radiotherapy and chemotherapy, where treatment planning and follow up of the treatment is highly dependent on radiology images. The best treatment for a specific patient depends on if there is one tumor or many small metastases, and the size and location of each tumor or metastasis. Furthermore, the size of the tumor is required to calculate how much radiation to apply to kill the cancer cells. MRI is normally used to obtain this information, and to plan the treatment, as MRI provides very good contrast between soft tissue types (and different MR sequences provide slightly different information / contrast). It is also necessary to segment important risk organs (e.g. the optic nerve) which should not be damaged by the radiation, see Figure 9. The treatment plan, i.e. how much radiation to apply to different parts of the brain, can be generated manually, through mathematical optimization or through machine learning.
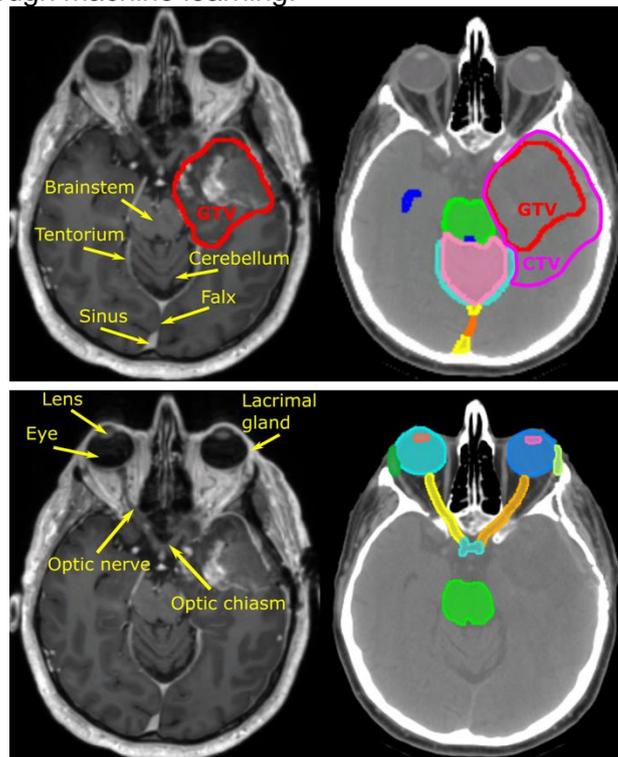

Figure 9 Illustration of brain tumor (red, to be killed by radiation) and risk organs (yellow, which should receive as little radiation as possible). GTV = gross tumor volume, CTV = clinical target volume (CTV). Deep learning can reduce the treatment planning time time substantially, by performing automatic segmentation of tumor(s) and risk organs (instead of doing manual time-consuming segmentations). Image from an open dataset in the cancer imaging archive (see references).

5.2 Automatic segmentation

To segment tumor(s) and risk organ(s) is currently often performed manually or semi-automatically by a neuro radiologist, medical physicist or radiation oncologist. Manual segmentations can be very time consuming, e.g. 10 – 60 minutes per patient, especially for many metastases and risk organs. Automatic brain tumor segmentation using deep learning is a very active area of research, as a trained network can perform the segmentation in 10 – 30 seconds. The annual BraTS (brain tumor segmentation) challenge (Menze et al., 2014, Baid et al., 2021) provides a large training dataset (2020: 369 subjects) as well as separate validation and test datasets (2020: about 100 subjects each), which has been very important for development in this field. See Figure 10 for an example of the MR images available in BraTS, for each subject there are also tumor annotations. The segmentation can be performed using a single MR modality (e.g. a T1-weighted image, T1W) or by simultaneously showing several types of MR images to a multi-channel CNN (e.g. T1W, T1W with gadolinium contrast, T2W, FLAIR).
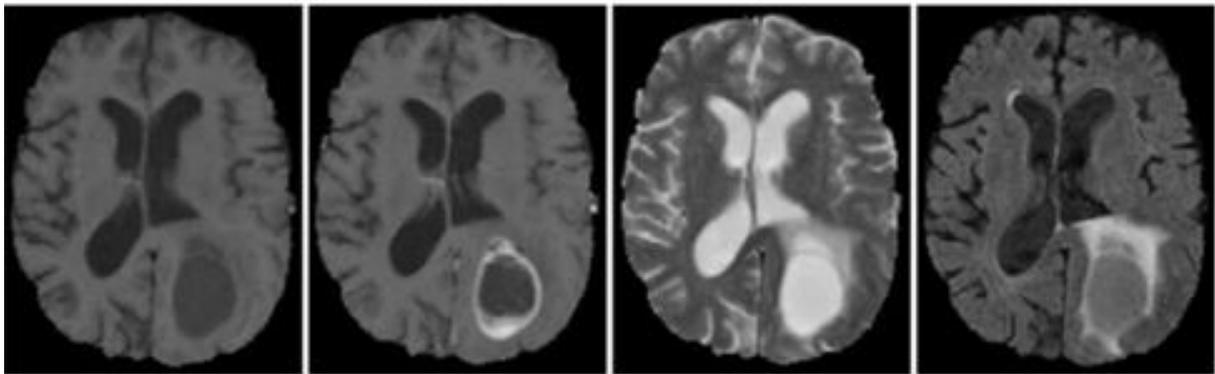

Figure 10 .MR images of glioblastoma multiforme taken from the BraTS dataset (Menze et al., 2014). Images taken with (from left): T1W, T1W Gd contrast, T2W and T2 FLAIR.

Virtually all participants in the BraTS challenge use some 2D or 3D variant of the popular U-Net architecture (Ronneberger et al., 2015). Isensee et al. (2018) demonstrated that a well-trained U-net with minor modifications (e.g., region based training and a combination of loss functions) together with additional training data produces very competitive results indicating that a well-constructed and performed training process is at least as important as focusing on novel architectural modifications when it comes to segmentation. Myronenko (2018) employed a 3D encoder-decoder architecture based on multiple ResNet-like blocks. As a novelty, the network is split into two decoding branches at the encoder endpoint output, where one of the branches is a regular decoder that produces the three tumor segmentation maps, and the other a variational decoder that reconstructs the input volume. This variational decoder branch serves as regularization for the shared encoder and is only active during training (see Figure 11).
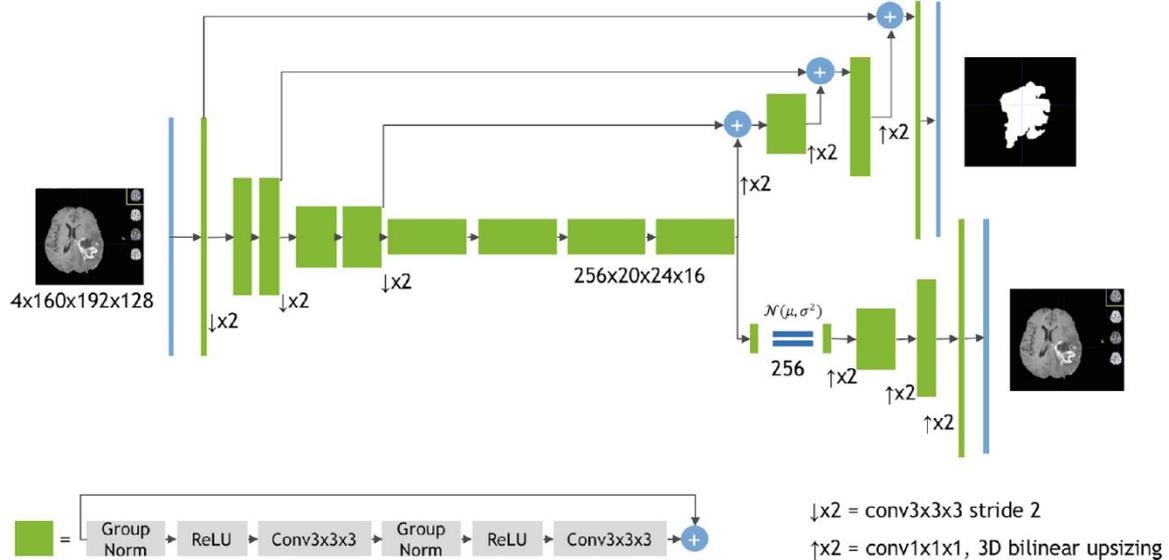
Figure 11 The encoder-decoder architecture employed in Myronenko (2018). The top decoder branch produces the tumor segmentation maps, while the bottom one reconstructs the input volume (mainly acting as regularization, to force the encoder to be good at several tasks).

More recently, segmentation architectures that include some kind of adversarial loss function (from generative adversarial networks, GANs) have become more popular (e.g. Cirillo et al., 2020), to punish segmentation maps that do not look realistic.

The main clinical challenge is to integrate different segmentation networks into the clinical workflow, as many clinics still use manual or semi-automatic segmentation. Another challenge is that a network trained on images from one MR scanner will not perform as well on images from another scanner, commonly called domain shift.

5.3 References

Baid, U., Ghodasara, S., Mohan, S., Bilello, M., Calabrese, E., Colak, E., ... & Bakas, S. (2021). The RSNA-ASNR-MICCAI BraTS 2021 benchmark on brain tumor segmentation and radiogenomic classification. arXiv:2107.02314.

Cirillo, M. D., Abramian, D., & Eklund, A. (2020, October). Vox2Vox: 3D-GAN for brain tumour segmentation. In International MICCAI Brainlesion Workshop (pp. 274-284). Springer, Cham.

Isensee, F., Kickingereder, P., Wick, W., Bendszus, M., & Maier-Hein, K. H. (2018, September). No new-net. In International MICCAI Brainlesion Workshop (pp. 234-244). Springer, Cham.

Leece, R., Xu, J., Ostrom, Q. T., Chen, Y., Kruchko, C., & Barnholtz-Sloan, J. S. (2017). Global incidence of malignant brain and other central nervous system tumors by histology, 2003–2007. Neuro-oncology, 19(11), 1553-1564.

Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., ... & Van Leemput, K. (2014). The multimodal brain tumor image segmentation benchmark (BRATS). IEEE transactions on medical imaging, 34(10), 1993-2024.

Myronenko, A. (2018, September). 3D MRI brain tumor segmentation using autoencoder regularization. In International MICCAI Brainlesion Workshop (pp. 311-320). Springer, Cham.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.

Brain tumor dataset for radiation therapy,
https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=95224486

6 Lung diseases

6.1 Introduction

Lung diseases cover lung cancer, and airway and pleural diseases.

Lung cancer is the leading cause of cancer death. Several countries have introduced lung cancer screening programs in order to detect lung cancer earlier, improving the probability for curative treatment for patients.
The lung cancer screening involves a low-dose CT scan of the chest ('lung screening CT'), which is to be checked for presence of suspicious lung nodules.
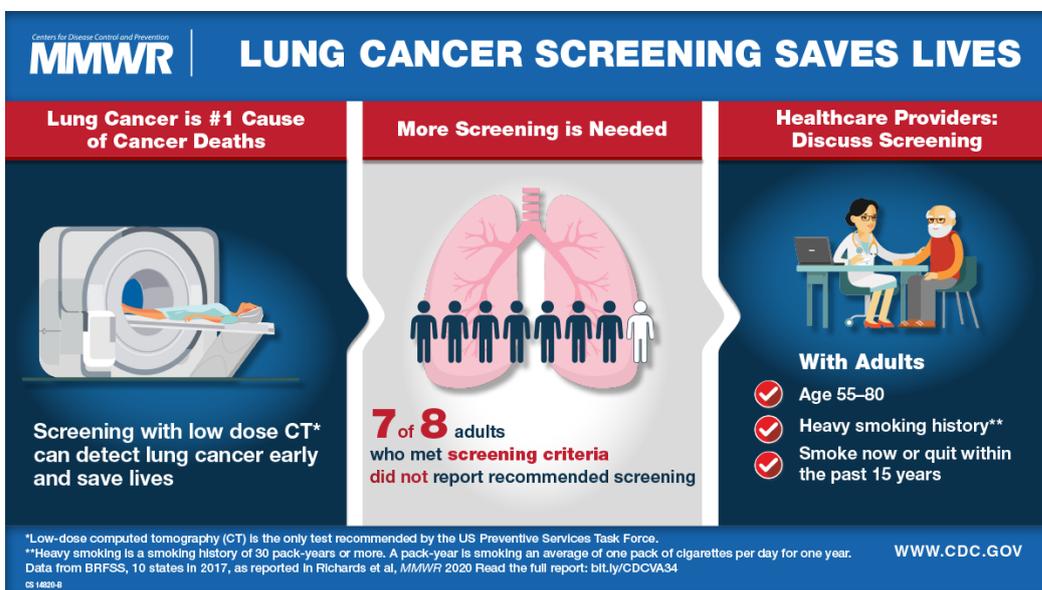
The increased volume of lung screening CTs represents an additional workload for the radiologist. Automation and AI are considered to help off-load the radiologist and reduce 'missed nodules' in the lung screening CTs.

Deep learning AI is seen as the best approach to detect nodules(Setio et al. 2017). Products with DL based lung nodule detection already are available commercially, for example (Veye Lung Nodules, 2021).

Once a suspicious nodule has been detected, typically a biopsy is scheduled to obtain tissue from the nodule for pathology. AI is also being considered to assist in analysis of the pathology images, with early research results being promising but adoption expected to take some time (Sakamoto et al., 2020).

In the next sections, we will give an overview of deep-learning AI approaches for nodule detection in lung screening CTs.

6.2 Pulmonary Nodule Detection in CT

From 2017 onwards, deep learning approaches using convolutional networks are the highest scoring networks in 'pulmonary nodule detection' competitions. In Setio et al, the DL winner achieved sensitivity 95%, less than 1% false positives. Best solutions detecting nodules which were missed by expert readers annotating the original data set (Setio et al. 2017)

Riquelme analyses various DL approaches for nodule detection in CT. Among the various approaches, 3D convolutional neural networks architectures demonstrated their usefulness, as most of the best-performing methods used them (Riquelme 2020). Specifically, densely connected networks with wide residual networks along with U-Net architecture obtained interesting results. Although 2D approaches are computationally less expensive, three-dimensional kernels detect more details about the nodules which inherently are a three-dimensional structure.

Some approaches divided the work into two stages: nodule candidate detection, and false-positive reduction, whereas others tackle the problem in a single network. Also for individual stages, 3D CNN approaches seem superior (although use of different datasets makes comparison difficult).

6.3 Digital pathology image analysis for lung adenocarcinoma

Microscopic examination of tissue slides is an essential step in cancer diagnosis. Hematoxylin and eosin (H&E) stained whole slide imaging (WSI) of tissue slides has become a routine clinical procedure, in which high resolution pathology images are captured and analysed. The limited capacity of pathology image analysis is a bottleneck in digital pathology.

Deep learning has started showing great potential in pathology image analysis task such as tumor region identification, prognosis prediction, tumor microenvironment characterization, and metastasis detection (see Table 1 below). The application of deep learning is still in the research phase.

Table 1 A summary of deep learning models for lung cancer pathology image analysis from Wang et al. 2019

Table 1. Summary of deep learning models for lung cancer pathology image analysis.

Topic	Lung Cancer Subtype	Task	Model	Prognostic Value Reported?	Year	Ref.
Lung cancer detection	ADC	Maglinant vs. non-malignant classification	CNN	Yes	2018	[69]
	NSCLC and SCLC		CNN	No	2018	[70]
	ADC and SCC		CNN	No	2019	[71]
	ADC		CNN	No	2019	[72]
	SCC		CNN	No	2019	[72]
	Not specified		CNN	No	2019	[73]
Lung cancer classification	ADC and SCC	ADC vs. SCC vs. non-malignant classification	CNN	No	2018	[74]
	ADC and SCC	Mutation status prediction	CNN	No	2018	[74]
	ADC	Histological subtype classification	CNN	No	2019	[75]
	NSCLC	PD-L1 status prediction	FCN	No	2019	[76]
	ADC and SCC	ADC vs. SCC classification	CNN	No	2019	[71]
	ADC and SCC	ADC vs. SCC classification	CNN	No	2019	[72]
	ADC and SCC	Transcriptome subtype classification	CNN	No	2019	[72]
	ADC and SCC	ADC vs. SCC vs. non-malignant classification	CNN	No	2019	[77]
	ADC	Hisotological subtype classification	CNN	No	2019	[78]

Table 1. *Cont.*

Topic	Lung Cancer Subtype	Task	Model	Prognostic Value Reported?	Year	Ref.
Micro-environment analysis	ADC and SCC	TIL positive vs. negative classification	CNN	Yes	2018	[39]
	ADC and SCC	Necrosis positive vs. negative classification	CNN	Yes	2018	[39]
	ADC	Tumor vs. stromal cell vs. lymphocyte classification	CNN	Yes	2018	[79]
	ADC	Microvessel segmentation	FCN	Yes	2018	[80]
	ADC	Computation staining of 6 different nuclei types	Mask-RCNN	Yes	2019	[81]
	ADC and SCC	TIL positive vs. negative classification	CNN	No	2019	[82]
Other	Early-stage NSCLC	Nucleus boundary segmentation	CNN	Yes	2017	[83]
	Not specified	Nucleus segmentation	Unet + CRF	No	2019	[84]

ADC: Adenocarcinoma; CNN: Convolutional neural network; CRF: Conditional random field; FCN: Fully convolutional neural network; Mask-RCNN: Mask-regional convolutional neural network; NSCLC: Non-small cell lung cancer; SCC: Squamous cell carcinoma; SCLC: Small cell lung cancer; TIL: Tumor-infiltrated lymphocytes.

Results from simple tasks, such as tumor detection and histology subtype classification are generally satisfactory, with an AUC around 0.9, whereas the results of more challenging tasks, including mutation and transcription status prediction, are less satisfactory with AUC ranging from 0.6 to 0.8.

6.4 References

Setio et al. (2017), Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The LUNA16 challenge, Medical Image Analysis, vol 42, pp1-13, 2017, doi.org/10.1016/j.media.2017.06.015

Riquelme Akhlufi (2020), Deep Learning for Lung Cancer Nodules Detection and Classification in CT Scans, AI 2020, 1, 28-67. doi.org/10.3390/ai1010003

Sakamoto et al. (2020), A narrative review of digital pathology and artificial intelligence: focusing on lung cancer. Transl Lung Cancer Res 2020. doi:10.21037/tlcr-20-591

Wang et al. (2019), Artificial Intelligence in Lung Cancer Pathology Image Analysis, Cancers 2019, 11, 1673 doi:10.3390/cancers11111673

Veye Lung Nodules (2021), Veye Lung Nodules product sheet, Aidence, the Netherlands, https://www.aidence.com/wp-content/uploads/2021/11/Aidence-Product-Sheet-Veye-Lung-Nodules.pdf

7 Hepato Pancreato Biliary Oncology

7.1 Introduction

Hepato Pancreato Biliary Oncology deals with malignant or cancerous tumors originating in the liver, pancreas, bile-ducts and gallbladder are some of the leading causes of cancer related deaths world-wide.

Liver cancer is the third leading cause of cancer death world-wide and pancreatic cancer is the fourth leading cause of cancer death in men and women and is projected to be the second leading cause within a decade. Early detection and complete removal of the tumor while saving as much as possible healthy tissue is important for the survival outcome and improved quality of life of the patient.

For the early diagnosis of HPB related cancers, CT and MR scans are the primary source of information and therefore automatic segmentation of tumors in CT or MR scans is of vital importance. As with most (medical) segmentation challenges nowadays deep-learning AI techniques offer the best results.

In the next sections, we will give an overview of what are currently the best deep-learning AI approaches for the segmentation of HPB related organs and the tumors therein.

7.2 Liver & Tumor Segmentation

In 2021, a study of Fernandez et al by Maastricht University and ASSIST partner Quantib, funded by the ITEA3 IMPACT project (project, nr. 17021) assessed five deep-learning architectures for liver and liver tumor segmentation. These models were 2D-UNet, 3D-UNet, Hybrid-UNet, residual encoder 3D-UNet and 3D-UNet with a different normalization of the convolutions.

The models were evaluated on the Liver Tumor Segmentation (LiTS) challenge which was organized in conjunction with the IEEE International Symposium on Biomedical Imaging 3 (ISBI) 2017 and MICCAI 2017 conferences but is still on going. The LiTS 2017 dataset contains 201 CT scans (131 for training, 70 for testing) from 7 different hospitals and research institutions (Bilic et al, 2019).

In the study the CT scans were first pre-processed by resampling them to a 1 mm^3 isotropic pixel spacing and making sure all scans have the liver in the same position inside a fixed (in X and Y directions) size bounding box.

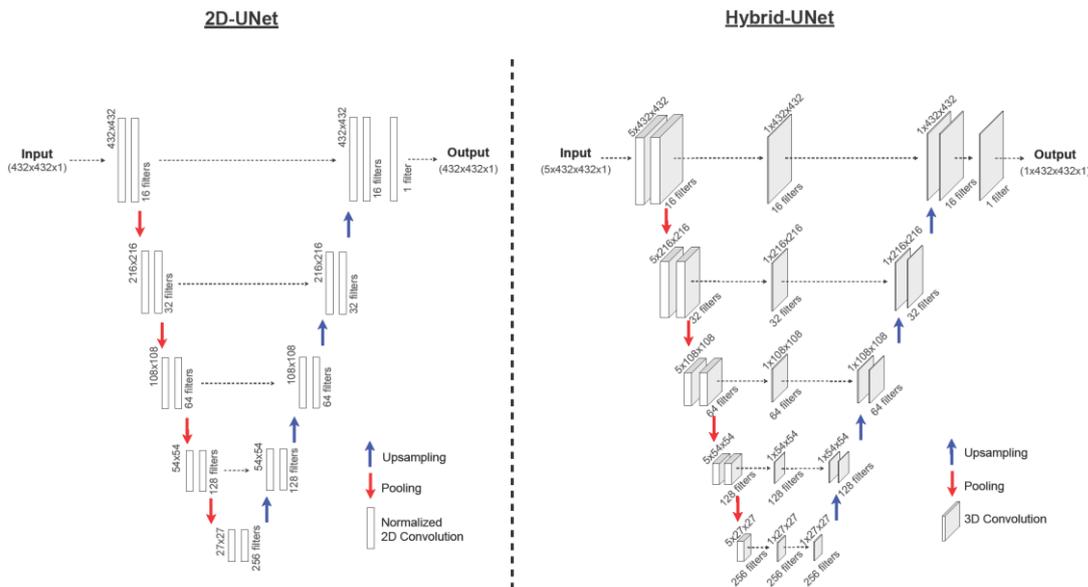The models were then trained using 5-fold cross-validation.

Figure 12 2D-UNet and Hybrid-UNet architectures. The annotations indicate the output of the data. Source: Fernandez et al. 2021

The 2D-UNet uses two convolutional layers are followed by one MaxPooling layer in each of the five levels, where the output of every convolution is normalized with batch normalization and ReLU activation function. Zero-padding is used in the convolutions and the kernel size and pooling size are 3 x 3 and 2 x 2, respectively. For the 3D-Unet the 2-dimensional operations are replaced with 3-dimensional ones.

They also tested a hybrid architecture as displayed above. This model, is UNet variation handles 3D data in the encoder and 2D data in the decoder. Thus, the model uses features from multiple slices to predict a single slice (the center slice). It uses a pooling size of 1 × 2 × 2 in every level and additional convolutions in the skip connections to reduce the z dimension from the encoder to the decoder.

Based on the 3D-UNet they also examined two additional variations. One model with residual connections from the beginning to the end of each level in the encoder and the other model where the convolutions where normalized differently.

As can be seen below, the 3D-UNet offered the best performance with the highest average dice score and the smallest deviation.
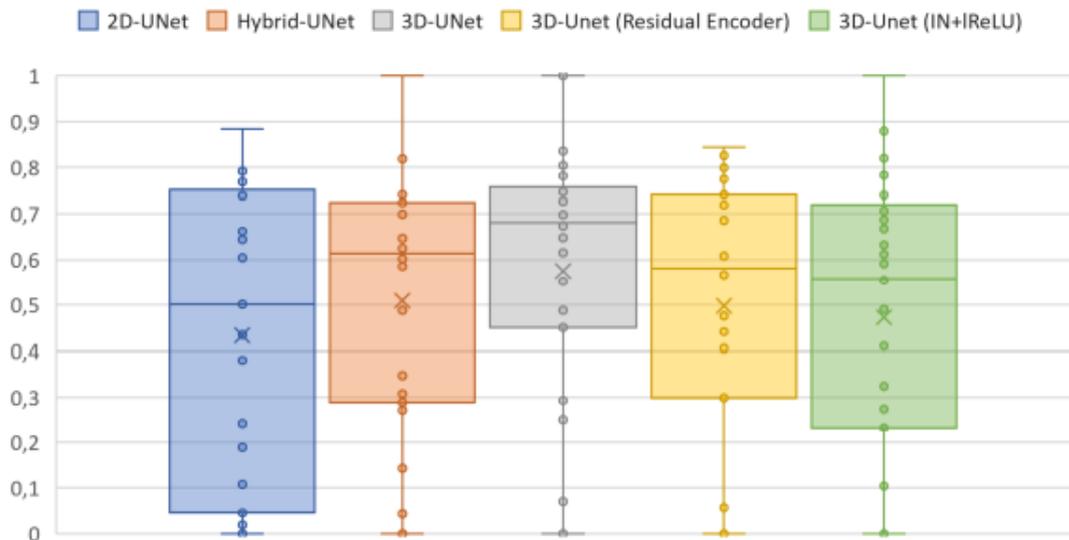
Figure 13 Dice score of the different scans from the main architectures. Source: Ferrnandez et al. 2021

The authors also further investigated adding different types of enhancements to the 3D-UNet architecture, namely attention methods, test-time inference and model ensemble, and TP/FP classification.

Attention methods try to mitigate the weakness of CNN architectures to capture global dependencies due to the locality of convolutional operations the authors further investigated adding two different types of self-attention modules, attention gates and additive self-attention, to the 3D-UNet architecture (see below).
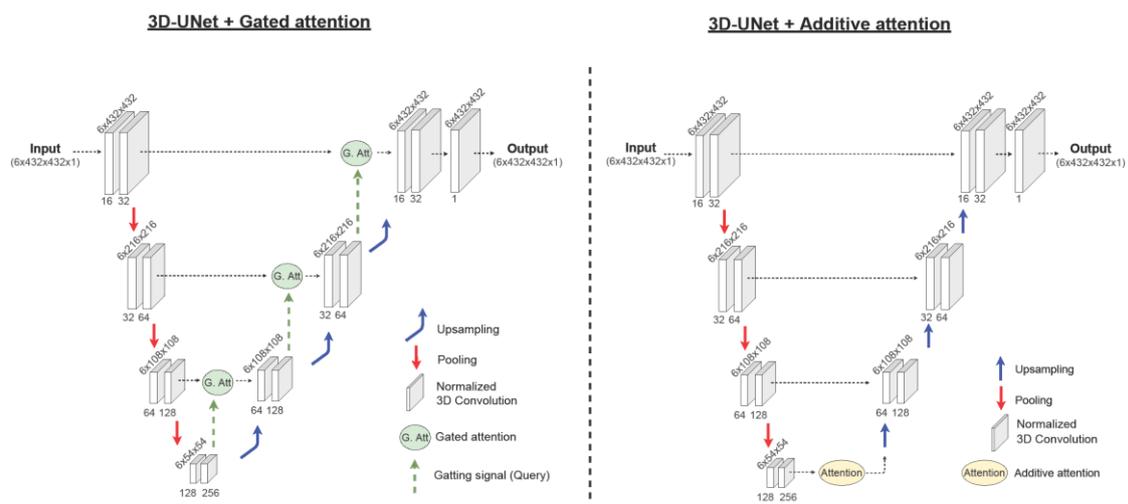

Figure 14 3D-UNet architecture enhanced with gated attention (left) or additive self-attention (right). Source: Fernandez et al. 2021

Test time inference is the procedure of applying augmentation during inference time where the input data is transformed before feeding the network. The inverse transformation is applied to the prediction in order to have the same properties as the input data. This process can be repeated using different transformations and the results averaged along with the prediction of the original image. Alternatively, model ensembling uses multiple models trained with different random seeds resulting in different versions of the same model. The outputs of the models are then averaged to produce a single result.

TP/FP classification uses a second model to classify whether the segmentation of the first model, e.g. a tumor, is segmented correctly (true positive or TP) or incorrectly (false positive or FP). It works by cropping a bounding box around a segmented "tumor" feeding it into a classifier network.
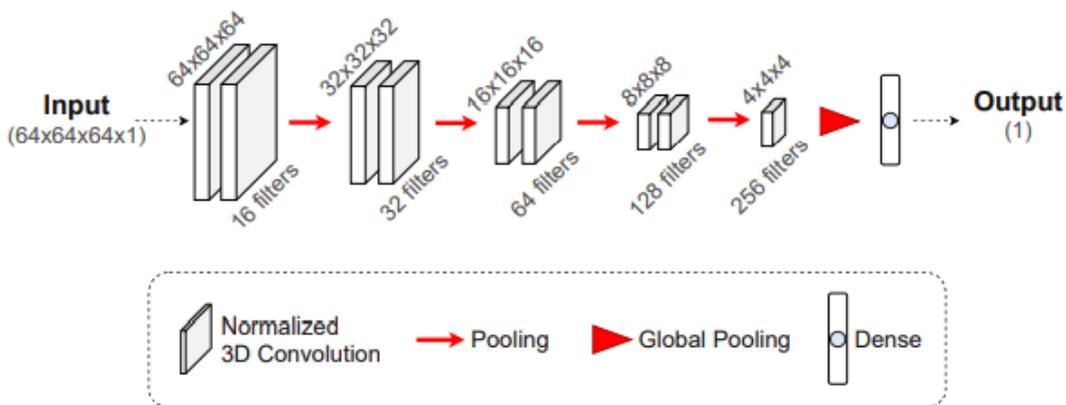

Figure 15 TP/FP classifier architecture. source: Fernandez et al. 2021

Table 2 Performance of 3D-UNets with different enhancements. Source: Fernandez et al. 2021

Performance of 3D-UNet with different additions			
Model	**Dice Score ↑**	**MSD ↓**	**HD ↓**
3D-UNet	0.575	17.98	47.59
3D-UNet + Additive attention	0.498	17.81	55.13
3D-UNet + Attention Gates	0.461	24.71	55.31
3D-UNet + TP/FP classifier	0.584	16.68	41.03
3D-UNet + Model Ensemble	0.611	17.30	38.99
3D-UNet + Test-time inference	0.609	18.26	45.73
3D-UNet + Model Ensemble + Test-time inference	0.598	17.55	42.70
3D-UNet + Model Ensemble + TP/FP classifier	0.597	17.80	35.71

As can be seen in the table above the authors concluded that a 3D-UNet with model ensemble achieved the best outcome and that averaging multiple models can reduce the errors produced by those models.

7.3 Pancreas & Tumor Segmentation

Diagnosis of pancreatic ductal adenocarcinoma (PDAC) diagnosis remains challenging because it can be difficult to differentiate them from benign lesions based on imaging features alone. Computed Tomography (CT) is currently the tool of choice for pre-operative diagnosis and follow-up.

In order to compare approaches it is important to have performance results from publically available datasets. For pancreas and pancreatic tumor segmentation, there are currently two publically available datasets available. The medical segmentation decathlon dataset containing 281 patients (Antonelli et al. 2021)   and The Cancer Imaging Archive (TCIA) Pancreas-CT dataset (Roth et al. 2016) containing 82 patients. Below, we will briefly discuss two approaches, which were tested, on these datasets.

In Liu et al (Liu et al. 2020) a modified Visual Geometry Group (VGG) network (Simonyan and Zisserman 2015) trained on data from 220 patients from the National Taiwan University Hospital image archive to classify 2D patches of the pancreas into cancerous or non-cancerous.
The model consisted of three convolutional blocks where each block consisted of two convolutional layers followed by rectified linear unit as the activation function and finally a max-polling layer. In the last convolutional block a flatten node was added at the end as well. Finally, at the end of the CNN model three fully connected (dense) layers were added. For the loss function, weighted binary cross-entropy was chosen to account for the imbalance in the number of malignant and benign patches. Patients were classified as having cancer based on the proportion of cancerous patches.

On the combined publically available datasets this deep-learning approach achieved a sensitivity of 0.790, specificity of 0.976, accuracy of 0.832, balanced accuracy of 0.883 and area under the ROC curve of 0.920, all with a 95% confidence interval.

In Alves et al (2022) A fully automatic deep-learning based framework for pancreatic ductal adenocarcinoma (PDAC) detection is described which produces tumor likelihood heat maps as well as provides segmentation of several surrounding anatomical structures such as the pancreatic duct, common bile duct, veins and arteries.

The framework was trained on 119 pathology-proven PDAC datasets and 123 non-PDAC datasets from the Radboud University Medical Center, Nijmegen.

The framework consists of self-configuring 3D nnU-Net's (Isensee et al. 2021) used to segment the pancreas and other anatomical structures. Three nnU-Net's were trained for PDAC detection and localization. The first (nnUnet_T) only segmented the tumor(s), the second (nnUnet-TP) segmented both tumor and pancreas and the third (nnUnet_MS) segmented the tumor, pancreas and surrounding anatomical structures.

The first step of the framework is to down sample the CT images and to create a low-resolution pancreas segmentation network to obtain a course segmentation, which is used to automatically extract the region of interest (ROI). Next, each of the PDAC detection nnU-Net's outputs a voxel-level tumor likelihood map and in the case of nnUnet_TP and nnUnet_MS also result in more detailed pancreas segmentation which can be used to reduce false positives by masking the tumor likelihood map.

The framework was tested on the combined publically available datasets as well. It achieved an area under the ROC curves of 0.872, 0.914 and 0.909 for the 3 networks (nnUnet_T, nnUnet_TP and nnUnet_MS) respectively. For tumors less than 2cm in size this was 0.831, 0867 and 0.876.

7.4 References

Antonelli et al. (2021) The Medical Segmentation Deathlon. https://arxiv.org/abs/2106.05735

Alves et al (2022), Fully Automatic Deep Learning Framework for Pancreatic Ductal Adenocarcinoma Detection on Computed Tomography. *Cancers* **2022**, *14*(2), 376; https://doi.org/10.3390/cancers14020376

Bilic, P. et al. (2019). The Liver Tumor Segmentation Benchmark (LiTS). *arXiv preprint arXiv:1901.04056*.

Fernandez, J.G. et al. (2021), Exploring automatic liver tumor segmentation using deep learning. International Joint Conference on Neural Networks (IJCNN). https://doi.org/10.1109/IJCNN52387.2021.9533649

Liu et al. (2020), Deep learning to distinguish pancreatic cancer tissue from non-cancerous pancreatic tissue: a retrospective study with cross-racial external validation. Lancet Digital Health 2020; 2: e303–13

Isensee et al. (2021), nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods volume 18, pages 203–211 (2021)

Roth et al. (2016). Data From Pancreas-CT. The Cancer Imaging Archive. https://doi.org/10.7937/K9/TCIA.2016.tNB1kqBU

Simonyan K and Zisserman A. (2015), Very Deep Convolutional Networks for Large-Scale Image Recognition. International Conference on Learning Representations 2015.

8 Conclusions

The previous sections have given an overview of how AI can be used in the clinical use cases of the ASSIST-project. From these descriptions it is clear that although there has been quite some research reported in literature, there are still challenges remaining for the ASSIST-project, which will be further investigated in work packages 2 and 3.

Although the clinical use cases differ a lot in terms of disease areas and organs involved, the common denominator in the use of AI in the clinical use-cases seems to be the use of deep learning neural networks both in 2D and 3D for medical image segmentation.

At the time of this writing there don't seem to be a lot of commercial software solutions for the clinical use-cases available but for pulmonary nodule detection in CT there is at least one commercial product namely Veye Lung Nodules by Aidence.