

Quantitative Image Analysis for Enhanced Diagnostics in Patient Care Employing Radiomics and Deep Learning Techniques

Friederike Jungmann

Vollständiger Abdruck der von der TUM School of Medicine and Health der Technischen Universität München zur Erlangung einer

Doktorin der Medizin (Dr. med.)

genehmigten Dissertation.

Vorsitz: apl. Prof. Dr. Bernhard Haslinger

Prüfende der Dissertation:

1. apl. Prof. Dr. Rickmer Braren
2. Priv.-Doz. Dr. Alexandra Gersing
3. apl. Prof. Dr. Jan S. Kirschke

Die Dissertation wurde am 09.11.2023 bei der Technischen Universität München eingereicht und durch die TUM School of Medicine and Health am 04.12.2024 angenommen.

Acknowledgements

I would like to thank all whose efforts guided the work for this thesis and who helped me make this dissertation possible through continuous support, motivation, and discussions.

My greatest gratitude goes to my mentor, Dr. med. Georgios Kaissis, who always supported me and whose expertise was invaluable in formulating research questions and developing a suitable methodology. Thank you for introducing me to different kinds of image analysis, machine learning, and medical studies in general. You are the person who sparked my fascination for research and who has inspired me to always keep working on extending my skill set and making myself better during the last few years. Thank you for your continuous support and encouragement, for practicing the open-door policy like no other, and for your valuable feedback and advice on any matter.

I would also like to thank my supervisor, Dr. med. Rickmer Braren, who always supported me in my doings and gave me the freedom to develop my skills and ideas. Thank you for making the past years such a rewarding experience and for giving me the opportunity to work in your research group. Thank you for your support and the continuous input you provided me that always pushed me in the right direction.

Furthermore, I would like to thank my boyfriend Christian, who supported me in any possible way. Thank you for all the inspiring discussions, your curiosity, and for making me think outside the box. Thank you for always believing in me and for your valuable advice and continuous encouragement during difficult and exhausting times. Thank you for being there, thank you for making me a better person.

Last but not least, this thesis is dedicated to my parents, Ines and Andreas, who have always supported me, have always been there with a sympathetic ear, and have never stopped believing in me. Thank you!

Abstract

Medical imaging plays an essential role in patient care, aiding in achieving a timely and correct diagnosis of patients or to monitor treatment response, for example in chemotherapeutic cancer treatment. Nowadays, the analysis of medical images is mostly done qualitatively by radiologists, who focus on describing certain parts of an image, like the size or texture of pathologic findings. While qualitative descriptions have the advantage of good understandability and are commonly used in patient care, this approach has some significant limitations, including high inter- and intra-reader variability, lacking standardization, and missing findings not visible to the human eye.

Quantitative image analysis has emerged over recent years to address the limitations of qualitative image descriptions. Here, images are analyzed using machine learning algorithms or deep learning applications like neural networks.

This thesis aims to improve diagnostic radiological performance and enhance patient care through machine learning-based assistance, employing two key quantitative image analysis methods: Radiomics analysis and a deep learning approach. Additionally, this thesis provides comprehensive insights into both techniques, their advantages, and challenges. It discusses their applications in medical contexts as well as their potential to contribute to clinical practice. Both methodologies were applied to medical problems to assess their suitability for routine clinical use. The thesis summarizes the work conducted in the underlying publications, outlining the main advantages and challenges of radiomics and deep learning as quantitative imaging techniques in medical settings. It also provides an outlook on future applications of quantitative image analysis and potential improvements.

The first included publication concentrates on radiomics-based quantitative image analysis and employs a multiparametric approach to predict overall survival in pancreatic ductal adenocarcinoma (PDAC) patients. Radiomics features of the primary tumor were obtained based on automated segmentation of the pancreas and tumor, followed by manual correction if needed. By incorporating both imaging features, clinical and histopathologic data of patients, this study presented a novel methodology to assess risk biomarkers for PDAC. The obtained radiomics features emerged as independent survival predictors, emphasizing the potential role of medical imaging in the course of personalized medicine. However, limitations arise from the laborious nature of radiomics analysis and histopathologic specimen workup resulting in a modest cohort size.

The second included publication focuses on deep learning techniques, investigating their impact on radiologists' performance in classifying breast lesions. Through the training of a neural network on mammograms and a subsequent reader study, the study explores if artificial intelligence (AI) assistance can elevate radiologist performance in the classification of breast lesions. It demonstrates that AI-based assistance significantly enhances diagnostic accuracy and suggests combining human and artificial intelligence yields superior results to either group alone. Notably, enhanced confidence in the algorithm with consequent

trust and improved cooperation between radiologists and the algorithm was observed when mechanisms illuminating its classification process were available. This finding emphasizes the crucial role of explainable AI within the medical domain and the potential of combining human with artificial intelligence for enhanced patient care, while emphasizing the need for further investigation.

Both publications underlying this thesis indicate the potential of both radiomics and deep learning in quantitative image analysis. Radiomics provides numerical features extracted from images, suitable for tasks such as developing imaging biomarkers. Meanwhile, the capability of deep learning applications to process entire images enables a more comprehensive image analysis with automated pipelines.

Overall, considerable benefits of quantitative image analysis include the potential to improve diagnostic accuracy, provide objective measurements, enhance diagnostic efficiency, and reduce costs in both research and the healthcare system. Additionally, these techniques can support clinical decision-making and improve patient outcomes, which is especially beneficial in personalized medicine. With the increasing availability of high-resolution imaging, the collection of large datasets, and the development of more advanced algorithms, the field of quantitative imaging is expected to continue to grow and significantly impact both research and future patient care.

While quantitative image analysis techniques offer distinct benefits compared to qualitative image analysis, several challenges still need to be addressed to enable broad implementation of quantitative image analysis techniques in clinical routine, including standardization, generalizability, and reproducibility of findings and integration into existing clinical workflows. Furthermore, overcoming challenges related to data privacy and lacking trust of clinicians in AI models, as well as compliance with regulations, such as the Artificial Intelligence Act of the European Union will be critical for the safe and reliable implementation of quantitative image analysis in healthcare.

Despite these challenges, future perspectives of quantitative analysis of medical images, including radiomics and deep learning approaches, are promising. Ongoing efforts in standardization and regulation, technology advances, development of transparency and explainability methods as well as secure and privacy-preserving AI are expected to improve the performance and clinical applicability of quantitative image analysis methods. Additionally, generative models, such as generative adversarial networks (GANs) and large language models, have the potential to further advance medical image analysis. Integrating these techniques with other medical data, such as clinical, genetic, or histomorphologic information, could pave the way for personalized medicine, ultimately improving patient outcomes.

Zusammenfassung

Medizinische Bildgebung spielt in der Patientenversorgung eine wesentliche Rolle, da sie unter anderem zu einer rechtzeitigen und korrekten Diagnose von Patienten beiträgt oder deren Therapieansprechen überwacht, zum Beispiel bei der chemotherapeutischen Behandlung von Tumoren. Heutzutage erfolgt die Analyse medizinischer Bilder durch Radiologen meist qualitativ, indem bestimmte Teilaspekte der Bilder beschrieben werden, zum Beispiel die Größe oder Textur pathologischer Befunde. Qualitative Beschreibungen haben zwar den Vorteil, dass sie gut verständlich sind und deshalb häufig in der Patientenversorgung eingesetzt werden, doch hat dieser Ansatz einige erhebliche Einschränkungen, darunter eine hohe Variabilität zwischen Radiologen, mangelnde Standardisierung oder das Missen von Befunden, die für das menschliche Auge nicht sichtbar sind.

Das Forschungsfeld der quantitativen Bildanalyse hat sich in den letzten Jahren entwickelt, um diese Limitationen der qualitativen Bildbeschreibung zu überwinden. Dafür werden Bilder mit Hilfe von Algorithmen des maschinellen Lernens oder Deep-Learning-Anwendungen, wie neuronalen Netzen, numerisch analysiert.

Die vorliegende Arbeit befasst sich mit der Anwendung quantitativer Bildanalysetechniken im medizinischen Kontext mit dem Ziel, die radiologische Diagnostik durch maschinelles Lernen zu unterstützen und die Patientenversorgung so zu verbessern. Hierfür werden zwei wichtige quantitative Bildanalysemethoden untersucht: Radiomics-Analysen und Deep-Learning-Ansätze. Darüber hinaus werden beide Techniken hinsichtlich ihrer Stärken und Herausforderungen untersucht sowie ihre potentielle Anwendungen im medizinischen Kontext erörtert. Beide Methoden wurden auf medizinische Probleme angewandt, um ihre Eignung für den routinemäßigen klinischen Einsatz zu bewerten.

Die Arbeit fasst die in den zugrundeliegenden Publikationen geleistete Arbeit zusammen und skizziert die wichtigsten Vorteile und Herausforderungen von Radiomics und Deep Learning als quantitative Bildgebungsverfahren im medizinischen Umfeld. Außerdem wird Ausblick auf zukünftige Anwendungen der quantitativen Bildanalyse und mögliche Verbesserungen gegeben.

Die erste Veröffentlichung konzentriert sich auf die Radiomics-basierte quantitative Bildanalyse und verwendet einen multiparametrischen Ansatz zur Vorhersage des Gesamtüberlebens von Patienten mit duktalem Adenokarzinom der Bauchspeicheldrüse. Die Radiomics-Merkmale des Primärtumors wurden auf Grundlage einer automatischen Segmentierung des Pankreas und des Tumors ermittelt, die bei Bedarf manuell korrigiert wurde. Durch Einbeziehung sowohl bildgebender Merkmale als auch klinischer und histopathologischer Daten der Patienten wurde in dieser Studie eine neuartige Methode zur Bewertung von Risikobiomarkern für Patienten mit duktalem Adenokarzinom der Bauchspeicheldrüse vorgestellt. Die gewonnenen Radiomics-Merkmale erwiesen sich als unabhängige Überlebensprädiktoren, was die potenzielle Rolle medizinischer Bildgebung im Zuge personalisierter Medizin unterstreicht. Einschränkungen der Studie ergeben sich jedoch aus der arbeitsintensiven Natur der Radiomics-Analyse und der histopathologischen Aufarbeitung der Gewebeproben sowie der daraus folgenden limitierten Kohortengröße.

Die zweite Veröffentlichung konzentriert sich auf Deep-Learning-Techniken und untersucht deren Auswirkungen auf die Leistung von Radiologen bei der Mammographieanalyse. Durch Training eines neuronalen Netzwerks und anschließendes Experiment mit Radiologen untersucht die Studie, ob die Unterstützung durch künstliche Intelligenz (KI) die Leistung von Radiologen bei der Klassifizierung von Brustläsionen steigern kann. Sie zeigt, dass KI-basierte Unterstützung die diagnostische Genauigkeit deutlich verbessert und die Kombination von menschlicher und künstlicher Intelligenz zu besseren Ergebnissen führt als eine der beiden Gruppen allein. Größeres Vertrauen in den Algorithmus und eine damit verbundene bessere Zusammenarbeit zwischen Radiologen und dem Algorithmus wurde beobachtet, wenn Mechanismen zur Verfügung standen, die den Klassifizierungsprozess des neuronalen Netzes beleuchteten. Dies unterstreicht die entscheidende Rolle erklärbarer KI im medizinischen Bereich und das Potenzial menschliche und künstliche Intelligenz für eine verbesserte Patientenversorgung zu kombinieren, wobei gleichzeitig die Notwendigkeit weiteren Forschungs- und Entwicklungsbedarfes betont wird.

Die beiden dieser Arbeit zugrunde liegenden Veröffentlichungen zeigen das Potenzial von Radiomics und Deep Learning im medizinischen Kontext. Radiomics-Analysen liefern numerische Merkmale, die aus Bildern extrahiert werden und sich für Aufgaben wie die Entwicklung bildgebender Biomarker eignen. Gleichzeitig ermöglicht die Fähigkeit von Deep-Learning-Anwendungen, ganze Bilder zu verarbeiten, eine umfassendere Bildanalyse mittels automatisierten Pipelines.

Zu den größten Vorteilen der quantitativen Bildanalyse gehört das Potenzial, die diagnostische Genauigkeit von Radiologen zu verbessern, objektive Messungen zu liefern, die diagnostische Effizienz zu steigern und Kosten sowohl in der Forschung als auch im Gesundheitssystem zu senken. Darüber hinaus können diese Techniken die klinische Entscheidungsfindung unterstützen und so die Patientenversorgung verbessern, insbesondere im Hinblick auf personalisierte Medizin. Mit der zunehmenden Verfügbarkeit hochauflösender Bildgebung, der Erfassung zunehmend großer Datensätze und der Entwicklung fortschrittlicherer Algorithmen ist zu erwarten, dass der Bereich der quantitativen Bildgebung weiter wachsen und sowohl die medizinische Forschung als auch die künftige Patientenversorgung erheblich beeinflussen wird.

Quantitative Bildanalysetechniken bieten zwar deutliche Vorteile gegenüber der qualitativen Bildanalyse, doch müssen noch viele Herausforderungen bewältigt werden, um die Einführung quantitativer Bildanalysetechniken in der klinischen Routine zu ermöglichen. Dazu gehören Standardisierung von Prozeduren, Verallgemeinerbarkeit und Reproduzierbarkeit von Ergebnissen sowie die Integration solcher Techniken in bestehende klinische Arbeitsabläufe. Darüber hinaus sind die Bewältigung von Herausforderungen im Zusammenhang mit Datenschutz, mangelndem Vertrauen der Kliniker in KI-Modelle sowie die Einhaltung von Vorschriften, wie dem Gesetz über künstliche Intelligenz der Europäischen Union, von entscheidender Bedeutung für eine sichere und zuverlässige Anwendung quantitativer Bildanalyse im Gesundheitswesen.

Trotz dieser Herausforderungen sind die Zukunftsaussichten quantitativer Bildanalyse, einschließlich Radiomics und Deep-Learning-Ansätzen, vielversprechend. Aktuelle Bemühungen um Standardisierung und Regulierung, technologische Fortschritte, die Entwicklung von Transparenz- und Erklärungsmethoden sowie sichere und datenschutzkonforme KI werden voraussichtlich die Leistung und klinische Anwendbarkeit dieser Verfahren verbessern. Darüber hinaus haben generative Modelle wie GANs und große Sprachmodelle das Potenzial, die medizinische Bildanalyse weiter zu verbessern. Die Integration dieser Techniken mit anderen medizinischen Daten, wie klinischen, genetischen oder histomorphologischen Informationen, könnte den Weg für eine personalisierte Medizin ebnen und letztlich die Patientenversorgung erheblich verbessern.

Table of Contents

Acknowledgements	iii
Abstract	v
Zusammenfassung	vii
List of Abbreviations	xi
List of Publications	xv
1 Introduction	1
1.1 Quantitative Image Analysis in Medical Settings	1
1.2 Thesis Aim	2
1.3 Thesis Outline	2
2 Background	5
2.1 Computed tomography	5
2.2 Development of Quantitative Image Analysis	6
2.3 Radiomics	7
2.3.1 Workflow of Radiomic analysis	7
2.3.2 PyRadiomics	8
2.4 Neural Networks	11
2.4.1 Workflow in medical applications	12
2.4.2 Main network architectures	15
2.5 Pancreatic ductal adenocarcinoma	21
2.5.1 Epidemiology	21
2.5.2 Pathology	23
2.5.3 Diagnosis	25
2.5.4 Therapy	25
2.6 Breast Cancer	28
2.6.1 Epidemiology	28
2.6.2 Pathology	29
2.6.3 Diagnosis	32
2.6.4 Mammography Screening	33
2.6.5 Therapy	36
2.6.6 Endocrine Therapy	37
2.6.7 Targeted Therapy	38
3 Methodology	39
3.1 Methodology of the first publication P-I	39
3.1.1 Study design	39
3.1.2 Clinical and histopathological data	39

3.1.3	Imaging Data	40
3.1.4	Survival Analysis	42
3.1.5	Statistics	42
3.2	Methodology of the second publication P-II	42
3.2.1	Study Design	43
3.2.2	Training, validation, and independent test datasets	43
3.2.3	Network training	44
3.2.4	AI based assistance	44
3.2.5	Reader study	45
3.2.6	Statistics	46
4	Journal Publications	47
4.1	Journal Publication P I: Multiparametric Modelling of Survival in Pancreatic Ductal Adenocarcinoma Using Clinical, Histomorphological, Genetic and Image-Derived Parameters	47
4.1.1	Short summary	47
4.1.2	Author's Contributions	48
4.2	Journal Publication P II: Algorithmic transparency and interpretability measures improve radiologists' performance in BI-RADS 4 classification	49
4.2.1	Short summary	49
4.2.2	Author's Contributions	50
5	Discussion	51
5.1	Findings and limitations of the publications	51
5.1.1	First publication P-I	51
5.1.2	Second publication P-II	52
5.2	Advantages and current challenges of quantitative image analysis	54
5.2.1	Benefits	55
5.2.2	Current challenges and improvement efforts	60
5.3	Contemporary research topics and future trends of quantitative image analysis	71
5.3.1	Deep generative models	71
5.3.2	Large language models	77
6	Conclusion	81
6.1	Summary of underlying publications	81
6.2	Current state of quantitative image analysis	83
6.3	Future work	84
A	References	87
B	Lists	133
B.1	List of Figures	133
B.2	List of Tables	134

List of Abbreviations

Short	Long
5-FU	5-Fluorouracil
ACR	American College Of Radiology
Adam	Adaptive Moment Estimation
AI	Artificial Intelligence
BI-RADS	Breast Imaging-Reporting And Data System
BRCA1	Breast Cancer 1
BRCA2	Breast Cancer 2
CA19-9	Carbohydrate Antigen 19-9
CAD	Computer Aided Diagnosis
CBIS-DDSM	Curated Breast Imaging Subset Of The Digital Database For Screening Mammography
CC	Craniocaudal
CDKN2A (p16)	Cyclin-dependent Kinase Inhibitor 2A (P16)
CEA	Carcinoembryonic Antigen
CHEK2	Checkpoint Kinase 2
CI	Concordance Index
CNN	Convolutional Neural Network
CT	Computed Tomography
DCIS	Ductal Carcinoma In Situ
DNA	Deoxyribonucleic Acid
EGFR	Epidermal Growth Factor Receptor

continued on next page ...

... continued from last page

Short	Long
ER	Estrogen Receptor
GAN	Generative Adversarial Network
GLCM	Gray Level Co-occurrence Matrix
GLDM	Gray Level Dependence Matrix
GLRLM	Gray Level Run Length Matrix
GLSZM	Gray Level Size Zone Matrix
GnRH	Gonadotropin-releasing Hormone
GPU	Graphics Processing Unit
Her2-neu	Human Epidermal Growth Factor Receptor 2
HNF1A	Hepatocyte Nuclear Factor 1A
HU	Hounsfield Unit
IBSI	Image Biomarker Standardization Initiative
IDC	Invasive Ductal Carcinoma
IHC	Immunohistochemistry
ILC	Invasive Lobular Carcinoma
K-RAS	Kirsten Rat Sarcoma Virus
KRT81	Cytokeratin 81
MLO	Mediolateral Oblique
MRI	Magnetic Resonance Imaging
Nadam	Nesterov-accelerated Adaptive Moment Estimation
NAG	Nesterov's Accelerated Gradient
NGTDM	Neighbouring Gray Tone Difference Matrix
OS	Overall Survival
PACS	Picture Archiving Communication System
PCA	Linear Principle Component Analysis

continued on next page ...

... continued from last page

Short	Long
PDAC	Pancreatic Ductal Adenocarcinoma
PET	Positron Emission Tomography
PFS	Progression Free Survival
PR	Progesterone Receptor
QIBA	Quantitative Imaging Biomarkers Alliance
QM	Quasi-mesenchymal
RECIST	Response Evaluation Criateria In Solid Tumors
ReLU	Rectified Linear Unit
ResNet	Residual Neural Network
ROI	Region Of Interest
SGD	Stochastic Gradient Descent
SMAD4	Mothers Against Decapentaplegic Homolog 4
TCIA	The Cancer Imaging Archive
TGF- β	Transforming Growth Factor Beta
TP53	Tumor Protein P53
UICC	Union For International Cancer Control
VAE	Variational Autoencoder
VOI	Volume Of Interest
WHO	World Health Organization

List of Publications

List of included Journal Publications

The doctoral thesis is based upon the following journal publications with shared first authorship:

- P-I** Kaissis, G. A., **Jungmann, F.**, Ziegelmayr, S., Lohöfer, F. K., Harder, F. N., Schlitter, A. M., Muckenhuber, A., Steiger, K., Schirren, R., Friess, H., Schmid, R., Weichert, W., Makowski, M.R. and Braren, R. F. (2020, apr). Multiparametric Modelling of Survival in Pancreatic Ductal Adenocarcinoma Using Clinical, Histomorphological, Genetic and Image-Derived Parameters. *Journal of Clinical Medicine*, 9(5), 1250. doi: 10.3390/jcm9051250
- P-II** **Jungmann, F.**, Ziegelmayr, S., Lohöfer, F.K., Metz, S. Müller-Leisse, C., Englmaier, M., Makowski, M.R., Kaissis, G.A., Braren, R.F. (2022, oct). Algorithmic transparency and interpretability measures improve radiologists' performance in BI-RADS 4 classification. *European Radiology*. doi: 10.1007/s00330-022-09165-9

List of other Publications

- P-1** Kaissis, G. A., Ziegelmayr, S., Lohöfer, F. K., Harder, F. N., **Jungmann, F.**, Sasse, D., Muckenhuber, A., Yen, H., Steiger, K., Siveke J., Friess, H., Schmid, R., Weichert, W., Makowski, M. R. and Braren, R. F. (2020, mar). Image-Based Molecular Phenotyping of Pancreatic Ductal Adenocarcinoma. *Journal of Clinical Medicine*, 9(3), 724. doi: 10.3390/jcm9030724
- P-2** Burian, E., **Jungmann, F.**, Kaissis, G. A., Lohöfer, F. K., Spinner, C. D., Lahmer, T., Treiber, M., Dommasch, M., Schneider, G., Geisler, F., Huber, W., Protzer, U., Schmid, R.M., Schwaiger, M., Makowski, M. R. and Braren, R. F. (2020, may). Intensive Care Risk Estimation in COVID-19 Pneumonia Based on Clinical and Imaging Parameters: Experiences from the Munich Cohort. *Journal of Clinical Medicine*, 9(5), 1514. doi: 10.3390/jcm9051514
- P-3** Ziegelmayr, S., Kaissis, G., Harder, F., **Jungmann, F.**, Müller, T., Makowski, M. and Braren, R. (2020, dec). Deep Convolutional Neural Network-Assisted Feature Extraction for Diagnostic Discrimination and Feature Visualization in Pancreatic Ductal Adenocarcinoma (PDAC) versus Autoimmune Pancreatitis (AIP). *Journal of Clinical Medicine*, 9 (12), 4013. doi: 10.3390/jcm9124013

- P-4** Dou, Q., So, T. Y., Jiang, M., Liu, Q., Vardhanabhuti, V., Kaissis, G.A., Li, Z., Si, W., Lee, H. H. C., Yu, K., Feng, Z., Dong, L., Burian, E., **Jungmann, F.**, Braren, R.F., Makowski, M.R., Kainz, B., Rueckert, D., Glocker, B., Yu, S. C. H. and Heng, P. A. (2021, mar). Federated deep learning for detecting COVID-19 lung abnormalities in CT: a privacy-preserving multinational validation study. *npj Digital Medicine*, 4(1). doi: 10.1038/s41746-021-00431-6
- P-5** **Jungmann, F.**, Kaissis, G. A., Ziegelmayr, S., Harder, F., Schilling, C., Yen, H.-Y., Steiger, K., Weichert, W., Schirren, R., Demir, I.E., Friess, H., Makowski, M.R., Braren, R.F. and Lohöfer, F.K. (2021, apr). Prediction of Tumor Cellularity in Resectable PDAC from Preoperative Computed Tomography Imaging. *Cancers*, 13 (9), 2069. doi: 10.3390/cancers13092069
- P-6** Knolle, M., Kaissis, G.A., **Jungmann, F.**, Ziegelmayr, S., Sasse, D., Makowski, M.R., Rückert, D. and Braren, R.F. (2021, apr). Efficient, high-performance semantic segmentation using multi-scale feature extraction. *PLOS ONE*, 16(8), e0255397. doi: 10.1371/journal.pone.0255397
- P-7** Kaissis, G., Ziller, A., Passerat-Palmbach, J., Ryffel, T., Usynin, D., Trask, A., Lima, I., Mancuso, J., **Jungmann, F.**, Steinborn, M., Saleh, A., Makowski, M., Rueckert, D. and Braren, R. (2021, may). End-to-end privacy preserving deep learning on multi-institutional medical imaging. *Nature Machine Intelligence*. doi: 10.1038/s42256-021-00337-8
- P-8** Harder, F.N., **Jungmann, F.**, Kaissis, G.A., Lohöfer, F.K., Ziegelmayr, S., Havel, D., Quante, M., Reichert, M., Schmid, R.M., Demir, I.E., Friess, H. Wildgruber, M. Siveke, J. Muckenhuber, A., Steiger, K., Weichert, W., Rauscher, I., Eiber, M., Makowski, M.R. and Braren, R.F. (2021, aug) [18F] FDG PET/MRI enables early chemotherapy response prediction in pancreatic ductal adenocarcinoma. *EJNMMI Res*11, 70. doi: 10.1186/s13550-021-00808-4

1 Introduction

1.1 Quantitative Image Analysis in Medical Settings

Medical imaging techniques have been a substantial part of diagnosis and patient monitoring in medical practice since the 1970s. Especially for abdominal solid cancers, computed tomography (CT) or magnetic resonance imaging (MRI) are most commonly used for diagnosis and therapy monitoring. Medical images can provide information about localization, size, invasiveness, and often type of cancer in a non-invasive way (van Griethuysen et al., 2017). Thus, medical imaging is a standard-of-care tool for analyzing cancer and metastasis in baseline and follow-up phases.

Historically, radiologists have performed image analysis primarily qualitatively by describing specific details in the images, such as the size or texture of pathological findings. While these pictorial descriptions have the advantage of good understandability and are commonly used in clinical routine, this approach has some significant drawbacks, such as lack of standardized description vocabulary (van Griethuysen et al., 2017), low inter- and intra-reader reproducibility (Kumar et al., 2012) and often time-consuming procedures involved in analyzing large datasets (Aerts, 2016; Limkin et al., 2017).

Quantitative image analysis has emerged in recent years to overcome the issues mentioned above. Here, the aim is to quantitatively describe images by computing either handcrafted features, as done in the field of radiomics, or by analyzing more complex image characteristics using machine learning algorithms and neural networks. Recent breakthroughs of AI in medical fields, especially in medical image analysis, have shown the possible benefits of quantitative over qualitative image analysis by achieving performances on par or even superior to human readers (Ardila et al., 2019; Gulshan et al., 2016; Haenssle et al., 2018; Lakhani & Sundaram, 2017; Lotter et al., 2021; McKinney et al., 2020; Ribli et al., 2018; Roy et al., 2020).

However, AI relies on large amounts of standardized, high-quality data which can be processed automatically. This dependence is still a main issue when working with medical data due to the lack of large, curated, multi-center, open accessible datasets (Kaissis et al., 2020b). Another aspect impeding the integration of AI based assistance into the clinical routine is many algorithms' lack of transparency and quantification of uncertainty. The resulting inability of quality control invokes mistrust in clinicians and raises ethical problems concerning AI based decisions in medical fields. This aspect, as well as the lack of standardization and privacy issues hindering multi-center studies, are problems addressed by current research (Jungmann et al., 2022; Kaissis et al., 2020a)

Although methods of quantitative image analysis are currently impeded from sufficient integration into clinical practice, it will most likely become an essential aspect of future patient treatment in the process of personalized medicine (Afshar et al., 2019) with many states investing in the development of AI research and applications (Söder, 2019). The National Cancer Institute defines personalized medicine as follows: '*In cancer, personalized medicine uses specific information about a person's tumor to help make a diagnosis,*

plan treatment, find out how well treatment is working, or make a prognosis' (National Cancer Institute, 2020a, Dictionary of Cancer Terms). Here, non-invasive automated quantitative analysis of tumor features such as histopathological subtype (Kaissis et al., 2019; Zhu et al., 2018), genetic markers (Mei et al., 2018) or survival predictions (Huang et al., 2016; Kaissis et al., 2020a) will be crucial.

Another possible application for deep learning-based assistance in medical fields is the analysis of medical screening images. The growing number of these examinations, such as mammography screenings or dermatoscopic images to determine skin cancer, create the demand for faster and more efficient analysis. In recent studies, neural networks have shown promising results in successfully tackling these applications by achieving performance levels superior to human readers (McKinney et al., 2020; Tschandl et al., 2019). Furthermore, collaborating with AI based assistance has increased clinicians' performance in recent studies (Jungmann et al., 2022; Tschandl et al., 2020).

1.2 Thesis Aim

The aim of this thesis is to investigate the possible value of quantitative image analysis in medical settings. For this, the performance and use in the clinical routine of two different approaches were assessed: Radiomics analysis and a deep learning approach using a neural network.

This topic was chosen for two reasons: Firstly, for patients with different tumor entities, individual information about their tumor in personalized medicine is crucial for prognosis. Secondly, the field of quantitative image analysis in medicine has developed rapidly over the past years yielding numerous publications reporting features associated with different medical topics. However, only a few publications consider different types of quantitative image analysis, assess the added value of these to the established clinical prediction models or consider human-AI-collaboration. Thus, a more critical examination of quantitative image analysis in medical fields is needed.

1.3 Thesis Outline

This cumulative thesis is based on two publications dealing with quantitative image analysis in medical research, which are listed in the List of Publications. The contents of this thesis are given in the following.

In Chapter 2, the different methods of medical image analysis and their development are presented. Radiomics as a field of current medical research is described, and an overview of artificial intelligence, especially deep learning as a part of it, is given. Furthermore, pancreatic ductal adenocarcinoma and mammography screening as clinical applications used in the publications are presented.

The methods used for both publications are reported in Chapter 3. The study layout, clinical data used, and the implementation of both quantitative image analysis techniques are described.

Chapter 4 provides the two publications on which the doctoral thesis is based and highlights the contribution of the doctoral candidate to each of them.

In Chapter 5, the methods of quantitative image analysis used in the publications are discussed with respect to their possible benefits and challenges. The results of the two publications and their limitations are reviewed.

Chapter 6 concludes the thesis with a conclusion. Here, the advantages and disadvantages of both methods are summarized, and points for improvement are explored. Furthermore, possible future applications are presented.

2 Background

2.1 Computed tomography

Computed tomography (CT) is an imaging modality based on the attenuation of x-rays and enables the non-invasive three-dimensional representation of organs without superposition of structures (Beitzel et al., 2017). The first CT scanner was developed by Godfrey Hounsfield in 1967, and the first clinical CT scanner was deployed in England in 1971 (Goldman, 2007). Since then, multiple improvements and innovations have occurred, such as multi-slice or dual-source CT scanners or, recently, photon-counting detectors (Flohr et al., 2020).

A CT scanner consists of a gantry containing, amongst others, the x-ray source and detector, a patient bed, a control console, and a computer (Beitzel et al., 2017). The arrangement of the x-ray source and detector and their motion varies depending on the generation of the CT scanner. In the currently used fourth generation, the x-ray tube revolves around the patient while the patient bed moves through the gantry, although first attempts to create CT scanners with stationary, photocathode-based x-ray sources have shown promising results (Cramer et al., 2018).

After passing through the patient, the x-rays are detected by a 360° detector system, which is usually composed of multiple rows to enable the computation of multiple image slices as one (multi-slice CT). In the case of dual-source CT scanners, two x-ray tubes are used (Matlaga, Kawamoto, & Fishman, 2008). The attenuated x-rays are indirectly converted into electric signals by attenuation in the ceramic scintillators. The recently introduced photon-counting CT scanners can provide higher image resolution with almost no electronic noise by directly quantifying photon energy using crystal semiconductors rather than ceramic scintillators (Flohr et al., 2020).

After converting the attenuated x-rays into electric signals, images are reconstructed from the data using back projection combined with filtering or iterative reconstruction. Each axial slice of the body of slice thickness z is divided into a matrix of $x \times y$ volume elements, called a voxel. Usually, a matrix of 512×512 voxels is used. The attenuation of x-rays depends on the density, the atomic number of the tissue, and the energy of the x-rays used. By detecting the attenuation of x-rays for each angle of incidence, the density value of each voxel can be computed using the linear attenuation coefficient. The density value is then translated into gray scale levels for each voxel using Hounsfield unit (HU), which represents the attenuation coefficient for each voxel relative to the attenuation coefficient of water ($HU = 0$) (Beitzel et al., 2017; Goldman, 2007, 2008; Hounsfield, 1980).

CT is an imaging modality widely used in medical applications. Especially in evaluating the spreading of cancer and its response to treatment, CT plays an important role. In this thesis, CT images were used for evaluating the predictive value of quantitative imaging parameters in patients with pancreatic cancer.

2.2 Development of Quantitative Image Analysis

At the time of this thesis, radiologic analysis of medical images is primarily qualitative by describing details in the images, for example, the shape or texture, like necrosis or heterogeneity, of pathological findings (Aerts, 2016). Although this approach has the advantage of being more human-understandable than quantitative analysis, due to its pictorial description, it has some significant drawbacks.

Firstly, descriptions of findings are non-standardized as there is no universally used vocabulary. As a result, radiologic reports of the same finding may vary drastically between departments or radiologists. Secondly, studies have shown low inter- and intra-reader agreeability and reproducibility for qualitative image descriptions (Kumar et al., 2012; Tixier et al., 2014). Thirdly, many procedures of qualitative image analysis, like segmentation tasks, are very time-consuming (Aerts, 2016; Kumar et al., 2012; Limkin et al., 2017). Thus, qualitative image analysis is not suited for automated analysis of large amounts of data.

Another drawback of qualitative image analysis is the –in parts– poor performance of established, image-based evaluation criteria, as demonstrated in the following example of cancer monitoring. There are two widely used methods to monitor the therapy response of solid tumors. When using the response evaluation criteria in solid tumors (RECIST), one measures the tumor across its longest diameter in the image and compares the result with the tumor length measured before the start of the treatment. The second option, the World Health Organization (WHO) method, is based upon the same principle but measuring the tumor's longest diameter and its longest perpendicular diameter. Both methods use predefined categories to assign the change of linear tumor measurement to the tumor's response to therapy (Jaffe, 2006).

However, the RECIST and WHO criteria do not always correlate well with the clinical outcome of the patients or with the actual tumor response to therapy, which may be due to several reasons (Tirkes et al., 2013). If the original tumor site is replaced with other tissue, such as scarring or desmoplasia, which may be difficult to distinguish from the original tumor, the size of the tumor is likely to be overestimated. Also, the inter- and intra-reader reproducibility varies considerably (Kumar et al., 2012). Furthermore, new drugs developed in cancer therapies do not always lead to an initial shrinkage of the tumor, even if the tumor responds positively to the therapy (Burton, 2007).

This method is unsuitable for medical research as the qualitative analysis of radiological images often does not produce reliable, reproducible data. Another disadvantage of qualitative analysis is the considerable time needed when dealing with large data sets. Quantitative image analysis has gained importance in medical research over the last years to meet the need for a more reproducible method of extracting information from medical images (Lambin et al., 2012). Quantitative image analysis aims to numerically capture the qualitative characteristics of images. This is done by calculating geometrical, morphological, or other mathematical features derived or calculated from the images (Handels, 2009).

The first approaches in computer aided diagnosis (CAD) were developed in the 1960s, but it was not until the 1980s that successful CAD techniques were proposed. The concept of CAD is such that computers could help radiologists to analyze medical images. One example of the successful application of CAD is detecting probably malignant lesions in mammograms (Doi, 2007). However, CAD systems are not widely used in clinical practice or medical research as they failed to increase the radiologists' performance in clinical settings (Lehman et al., 2015). This can be explained by the fact that CAD systems are prone to false-positive findings. The large number of regions marked by the software each need to be checked by

the radiologist, which is a very time-consuming task (Kohli & Jha, 2018). Accordingly, CAD software did not improve the sensitivity of trained radiologists in several studies (Cole et al., 2014; Gur et al., 2004; Lehman et al., 2015).

2.3 Radiomics

Nowadays, radiomics as a development of CAD is a substantial field of radiological research. Radiomics refers to the automated extraction and analysis of quantitative imaging features derived from standard-of-care medical images. It provides descriptive or predictive models which could yield important additional information to help make a diagnosis, monitor treatment response, or assess a patient's prognosis (Kumar et al., 2012). Here, large amounts of quantitative features are derived from the region of interest (ROI) or volume of interest (VOI). ROI or VOI describe the parts of an image marked by a person or algorithm which are of interest to the medical analysis (Handels, 2009). An example of a VOI is shown in Figure 2.1.

2.3.1 Workflow of Radiomic analysis

The workflow of radiomics can be divided into five steps (Lambin et al., 2017):

1. Data selection
2. Imaging
3. Feature extraction
4. Analysis
5. Modeling

The protocol for image acquisition and ROI are defined during data selection. Furthermore, the prediction target is identified, meaning the event or characteristic to be analyzed or predicted (Lambin et al., 2017).

In the step of imaging, the medical images are obtained, and the ROI or VOI is marked afterward, as shown in Figure 2.1. The marking of the voxels defining the ROI or VOI is called segmentation. This can be done manually or (semi-) automatically using different kinds of segmentation algorithms (Aerts, 2016). Figure 2.1 shows an example of a VOI obtained by manual segmentation, in this case of a pancreatic tumor in CT data. In cross-sectional images, each pixel represents a cuboid with the slice thickness of those images and is therefore called a voxel. The value assigned to the voxel is calculated as the arithmetic mean value of the represented cuboid.

In the next step, quantitative features are extracted from the defined ROI using computational image analysis tools. These features include statistical information and shape-based data (Aerts, 2016). The features are then combined with non-radiomic data, such as clinical or demographic traits, for further analysis.

In the analysis step, the number of features is systematically reduced by identifying strongly correlating items by clustering or substituting them with the single most representative feature. Additionally, feature robustness should be assessed, for example, through multiple segmentations. This should be done to exclude inconsistent features from further analysis so that the study's findings are reproducible.

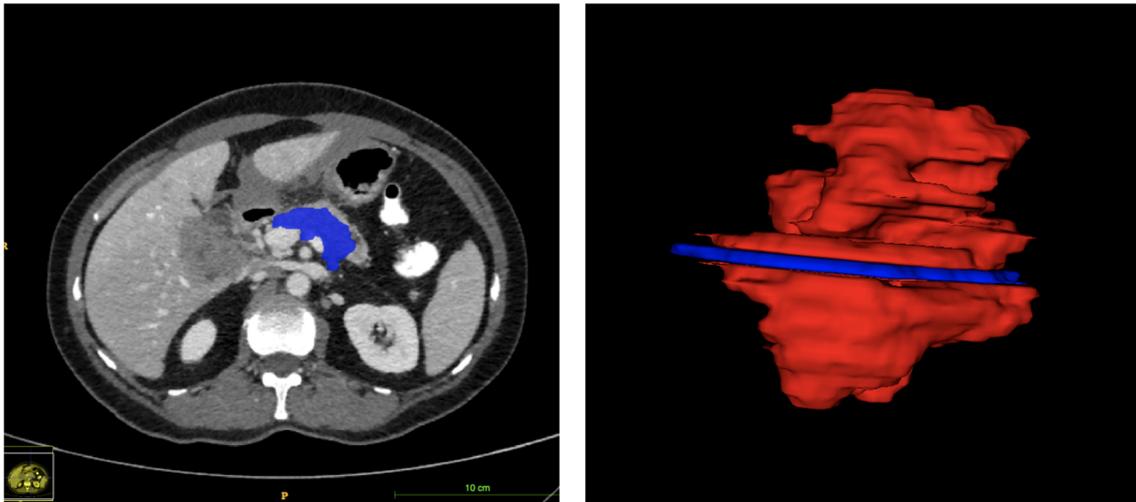


Figure 2.1 Manual segmentation of a volume of interest (VOI) on a CT, in this case a tumor of the pancreas. Left: axial view of a slice of the segmentation displayed over the corresponding CT image. Right: 3D-rendering of the whole tumor segmentation with the slice from the left colored in blue.

Finally, models to predict or describe the target of the study on different datasets are developed based on the feature analysis. This is usually done by computing a predefined selection of radiomic features, called a radiomic signature, or applying machine-learning algorithms. The obtained model should be validated, ideally on an external cohort, to assess its generalizability and overall performance (Lambin et al., 2017).

As of writing this thesis, the mathematical description for features derived in radiomic studies is inconsistent over different software used and has often not been well-documented in publications. To assess this problem, radiomics frameworks like PyRadiomics (van Griethuysen et al., 2017) or R0diomiX (Bagher-Ebadian & Chetty, 2021) have been developed, which represent standardized platforms for quantitative image analysis. However, many institutions still use their in-house developed software without public access and thus limiting the fidelity and reproducibility of their work (van Griethuysen et al., 2017).

2.3.2 PyRadiomics

This thesis's radiomic analysis of data was done using package PyRadiomics. It was developed at Harvard Medical School and is a Python-based, free and open-source framework for radiomic feature extraction with detailed documentation about the mathematical definition of these features (van Griethuysen et al., 2017). To provide a common standard for radiomic analysis, the documentation of all features extracted and filters applied are available to the public. This, as well as the number of publications that have successfully used PyRadiomics analysis in quantitative image analysis, are the main reasons why PyRadiomics was used in the first publication (Kaissis et al., 2020a), in the following referred to as P-I, on which this thesis is based. PyRadiomics can either be used directly from the command line, in interactive Python scripts, or as a front-end extension of the medical image computing application 3D Slicer. In the work for this thesis, PyRadiomics was used from the command line.

Workflow

The workflow of PyRadiomics looks as follows: Firstly, images and their corresponding segmentation masks are loaded into the platform. The segmentation masks define the ROI or VOI and thus the voxels of the image from which the features will be calculated. Creating segmentation masks in publication P-I (Kaissis et al., 2020a) was done using SimpleITK. SimpleITK provides an interface to The National Library of Medicine Insight Segmentation and Registration Toolkit, also referred to as the Insight Toolkit (ITK) for different programming languages (Johnson et al., 2019). PyRadiomics can handle different types of image formats, including JPEG or PNG. For medical images, the NIfTI format is supported. The step of image loading also includes the option to pre-process the data. For example, voxels can be scaled to be isotropic.

Several built-in filters can be applied to the images in the second step. The filters are primarily implemented using NumPy and include square, square root, logarithm, and exponential filters. More complex filters like Laplacian of Gaussian or Wavelet filters are implemented using PyWavelet.

In the next step, the features to be extracted from the images can be specified. PyRadiomics offers seven feature classes: Shape descriptors, first-order statistical data, gray level co-occurrence matrix (GLCM), gray level run length matrix (GLRLM), gray level size zone matrix (GLSZM), gray level dependence matrix (GLDM) and neighbouring gray tone difference matrix (NGTDM) features. All these features can be extracted from the original image as well as filtered images. The only exception are shape-based features, independent from the voxel values and thus calculated only once from the segmentation mask using the original images. The information about features to be extracted, filters to be applied, and the type of pre-processing are stored in an additional parameter file.

Subsequently, the features can be extracted either from 2D or 3D segmentation. PyRadiomics stores the calculated features and information about the version and settings used as an ordered dictionary. Here, every feature can be identified via its name and the image it was extracted from.

Radiomic Features

The following section briefly explains the features derived from PyRadiomics. All feature descriptions were obtained from the PyRadiomics Documentation (pyradiomics community, 2016).

The shape and size of the ROI or VOI are described by calculating shape features. This is done either in 2D or 3D, depending on the dimensionality of the segmentation mask. As shape features are independent of the voxel values inside the ROI, they are calculated using only the unfiltered original image and segmentation mask. In the case of a 3D segmentation, the shape of the VOI is approximated by triangle mesh, built by applying a marching cubes algorithm (Lorensen & Cline, 1987). In the case of a 2D segmentation, a circumference mesh based upon an adapted marching cubes algorithm is used for this task. The mesh consists of triangles (3D) or lines (2D) created by connecting previously calculated points halfway on an edge between a voxel inside and its neighboring voxel outside the segmentation mask. Further shape features are calculated using the mesh volume of the ROI or VOI. These features cover surface area, surface volume ratio, sphericity, compactness, and different diameters of the segmentation and their relationship. In the case of a 2D analysis, the perimeter of the segmentation mask replaces the volume in the 3D calculation, and no compactness is calculated.

Features calculated in the class 'first order features' characterize the distribution of the different values assigned to the voxels of the ROI or VOI in the image. This class comprises features describing statistical aspects of the voxel value distribution such as the minimum, maximum, mean, and median value, the standard distribution, interquartile range, absolute and standard deviation, as well as the root mean squared value. Furthermore, energy and entropy values are calculated. Energy features refer to the squared magnitude of voxel values in the analyzed image region and are volume confounded, while entropy is used to evaluate the disorder of voxel values analyzed.

Gray level co-occurrence matrix (GLCM) features are calculated based on the gray level co-occurrence matrix of the ROI or VOI, which illustrates the homogeneity or heterogeneity of the image texture. For this matrix, the gray levels of the image analyzed are pooled into equally spaced bins, each representing a range of possible voxel values. The number of bins N_g is specified manually in the settings file by the parameter `binWidth`. The GLCM is of the size $N_g \times N_g$. It gives information about how often two voxels with the distance δ along an angle θ have a specific combination of bin values. For example, the $(i, j)^{th}$ element of the matrix codes how many combinations of voxels with bin value i and j exist in a given distance δ from each other along the specific angle θ . Thus, the GLCM is symmetrical. The distance δ from the voxel analyzed is calculated using the infinity norm, which is defined as the maximum of its components and calculated using the following formula: $\|v\|_\infty = \max\{|v_i| \mid i = 1, \dots, n\}$ (Karpfing, 2014, p.422). Thus, the distance between two voxels is the greatest of their differences along any coordinate dimension. Afterward, features describing the skewness and symmetry of the GLCM, as well as the correlation, the relationship between different matrix entries, and patterns, are determined. Each feature is calculated separately on the gray-level co-occurrence matrices for the different angles, and the mean of those calculations is returned as the feature value.

The gray level size zone matrix (GLSZM) characterizes the number of gray level zones in the image analyzed. A gray level zone consists of connected voxels which share the same bin value and thus form a homogeneous zone in the binned image. Consequently, the GLSZM features are dependent on the `binWidth` parameter, like the GLCM features. In 2D analysis, each voxel can be directly connected to a maximum of 8 other voxels, including diagonals. In 3D, each voxel can have up to 26 neighbors. As the GLSZM is not calculated along a specific axis, it is rotation-independent. The features calculated from the GLSZM mostly correspond to the GLCM features and describe the gray-level zones' distribution, variability, and heterogeneity.

The gray level run length matrix (GLRLM) is determined by measuring the number of successive voxels along a specific angle θ belonging to the same voxel value bin. The $(i, j)^{th}$ element of the GLRLM represents the number of gray level runs containing j voxels with the gray level i along the angle θ . Like the GLCM, the GLRLM is rotation dependent. The features based upon this matrix are calculated for each angle θ individually, and the mean value is returned. GLRLM features measure the similarity, distribution, and variance of the gray level runs in the image and thus provide information about the area's texture.

Features calculated on the neighbouring gray tone difference matrix (NGTDM) also give information about the image's texture. In contrast to the GLRLM, each voxel gray level is compared to its neighbors' arithmetic mean gray level in the distance δ . The sum of the absolute differences is then stored in the NGTDM. Thus, NGTDM describes local changes in texture.

In the gray level dependence matrix (GLDM), the image texture is described by quantifying the gray-level dependencies in the image. The matrix covers the number of consecutive voxels in a distance δ dependent on the voxel analyzed. Dependency is defined as $|i - j| \leq \alpha$. This means the absolute difference between gray level i and j of two voxels is smaller than a defined number α . The $(i, j)^{\text{th}}$ element of the GLDM shows how many voxels are contained in the image, which have the gray value i and j dependent neighboring voxels. Features calculated on the GLDM characterize the gray-level dependencies' distribution, similarity, and variance.

2.4 Neural Networks

To overcome the dependencies of feature extraction mentioned in Section 2.3 and provide clinicians with an end-to-end approach, quantitative image analysis using deep learning methods is another objective in medical imaging research. Deep learning is a subgroup of machine learning which in turn is a subclass of AI. Definitions of the terms are given in the following, and their relation is shown in Figure 2.2, as adapted from (Chartrand et al., 2017; Goodfellow, Bengio, & Courville, 2016; Li et al., 2021a).

- Artificial intelligence (AI): A field of computer vision to perform tasks requiring human intelligence for resolving (Goodfellow, Bengio, & Courville, 2016). Examples include natural language processing and knowledge bases (Goodfellow, Bengio, & Courville, 2016; Li et al., 2021a).
- Machine learning: A subclass of AI where algorithms learn patterns from given data, usually relying on hand-crafted features and human input. Thus, it depends on the quality and type of representation of the data provided. Examples include logistic regression, random forest, and naive Bayes algorithms (Goodfellow, Bengio, & Courville, 2016).
- Deep learning: A subfield of machine learning where relevant features are extracted directly from the raw input data without needing hand-crafted features. Here, complex features are derived from simpler ones by the algorithms (Chartrand et al., 2017; Soffer et al., 2019).
- Neural network: Computational model inspired by the function of natural brains, which consists of multiple interconnected processing units, called nodes (Zakaria, Al-Shebany, & Sarhan, 2014). If the model consists of multiple connected layers of nodes reflecting a hierarchy of representations learned from the data, it represents a deep learning application (Goodfellow, Bengio, & Courville, 2016).

Deep learning encompasses models which can learn relevant features for the set task by reflecting the hierarchy of structures in the data. This is done directly from the raw data without requiring previous human feature engineering (Chartrand et al., 2017). A subgroup of deep learning models called deep neural networks consists of multiple layers of representations. In each layer, non-linear functions transform the representation into a higher, more abstract level. By composing multiple transformations, complicated functions can be learned (LeCun, Bengio, & Hinton, 2015). For example, objects can be identified based on edges, corners, contours, and pixel values (Goodfellow, Bengio, & Courville, 2016).

Today, neural networks are widely used in everyday applications, for example, in analyzing handwriting (LeCun et al., 1989; Simard, Steinkraus, Platt, et al., 2003), facial recognition (Lawrence et al., 1997), self-driving cars (Hadsell et al., 2009), virtual personal assistants (Kepuska & Bohouta, 2018), and chatbots

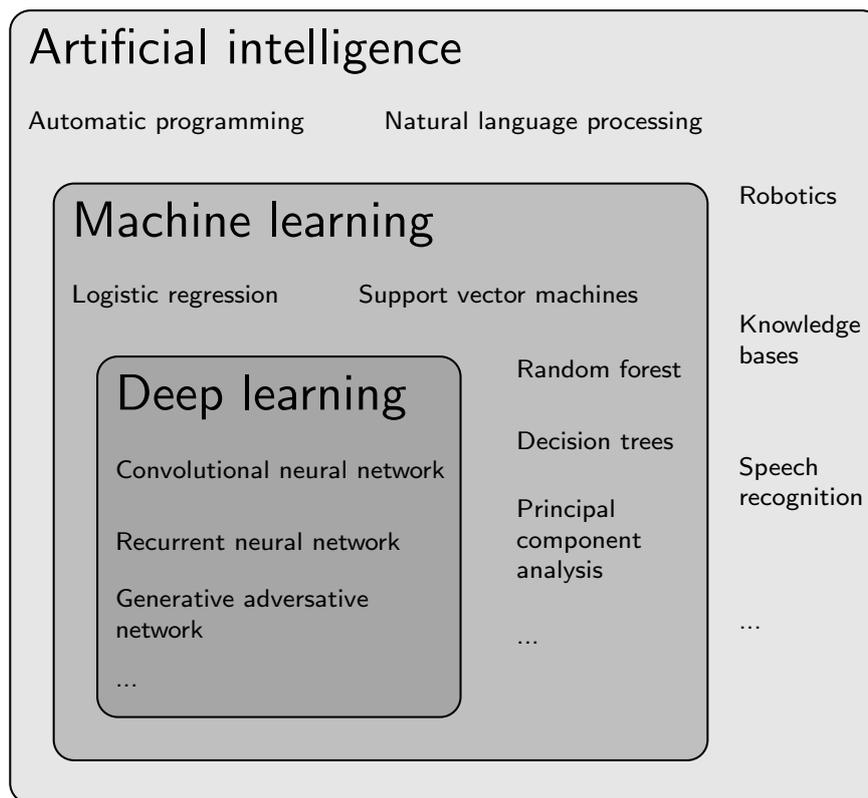


Figure 2.2 Relationship between artificial intelligence, machine learning and deep learning with examples for each field.

(OpenAI, 2023). They even outperform humans in complicated tasks like the game of Go (Silver et al., 2017).

Recent methods of accelerating the training process, like the usage of graphics processing units (GPUs) or application-specific integrated circuits (ASICs), the availability of software frameworks such as PyTorch, Keras, and TensorFlow for easy implementation and training of network architectures as well as the growing amount of labeled medical data have lead to the introduction of neural networks in medical research (Chartrand et al., 2017; Hwang, 2018).

2.4.1 Workflow in medical applications

Early neural networks used for image analysis in medical research have achieved results comparable to or even better than clinicians in multiple medical image analysis tasks (Esteva et al., 2017; McKinney et al., 2020). This demonstrates their potential application in the clinical routine. For image analysis, a special kind of neural network called convolutional neural network (CNN) is mostly used (Li et al., 2021a) due to its beneficial properties, which are explained in Section 2.4.2. As all experiments in this thesis were conducted on medical image analysis tasks, this section will explain the workflow for developing neural networks with respect to medical image analysis only. Although neural networks can be applied to solve tasks which are also suitable for radiomics analysis, the workflow of developing a neural network differs significantly from the workflow of radiomics analyses. To recall, the workflow of radiomics consists of five distinct steps: Data selection, imaging, feature extraction, analysis, and modeling (Lambin et al., 2017).

Data collection and pre-processing

Similar to the workflow of radiomics analyses, the first step of studies including neural networks comprises the definition of protocols for image acquisition and collection of data for the given task. Depending on the task, manual segmentation of the ROI or VOI or a limage labeling may be needed (Soffer et al., 2019). However, no hand-crafted feature selection is needed as neural networks are trained to extract relevant features directly from the raw input data. This is advantageous concerning human effort and tasks such as lesion detection and image synthesis. However, using whole images instead of defined regions of interest comes at the cost of more noisy data, which requires more comprehensive training data sets (Li et al., 2021a). To obtain more extensive data sets and thus improve the model's performance and robustness, augmentation is often applied to the training data. This means creating additional data by generating slightly different versions of the existing training data, for example, by rotating or shifting the image (Shorten & Khoshgoftaar, 2019).

Training

The radiomics steps of feature extraction, analysis, and modeling are merged into a step of model training, where the neural network is trained on the input data. Training a deep learning algorithm is a repetitive process, intending to capture the training data's underlying structure and relevant features to ultimately predict the correct labels on former unseen data. The training process is further explained in Section 2.4.2. There are different types of training categories, which require different amounts of labeled data: Supervised, semi-supervised and unsupervised learning (Li et al., 2021a).

In supervised learning, the algorithm is trained on a training set containing input-label pairs (Alloghani et al., 2020; Rashidi et al., 2019). Here, the aim is to learn a specific task and capture the patterns and relevant features in the training data by comparing the generated output \hat{y} to the actual label y and to updating the network's parameters based upon the value of the loss function, as explained in Section 2.4.2 (Goodfellow, Bengio, & Courville, 2016). The generalizability of the trained algorithm is dependent on the quantity and quality of training data, as it learns the relevant features with respect to loss minimization of the training label. Training data not representative of the test data have a low diversity or a set size too small to capture the underlying relationships will lead to poor generalizability of the neural network (Jabbar & Khan, 2015; Sheller et al., 2020). Supervised learning is the most common form of machine learning used in the medical context due to the possibility of using accurate labels, easy validation methods of the algorithm, and common medical tasks like image classification (Penny & Frost, 1996; Rashidi et al., 2019).

In contrast to supervised learning, in unsupervised learning, the algorithm is trained on a data set containing no labels. Here, the aim is to identify patterns and structures contained in the data without an explicit ground truth (Abukmeil et al., 2021). Thus, unsupervised learning is generally used in exploratory data analysis. Typical applications are clustering algorithms, where data are grouped into classes based upon feature similarities, dimensionality reduction, where the number of input features is reduced to enable a more simplified data representation, or pre-processing data later used for supervised learning (Abukmeil et al., 2021; Hofmann, 2001; Jafari et al., 2020; Rashidi et al., 2019). Compared to supervised learning, the performance evaluation of trained algorithms is more complex, as no loss calculation based upon an explicit training label is possible (Palacio-Niño & Berzal, 2019; Sutskever et al., 2015).

As large amounts of label data are often time-consuming and expensive to obtain and unsupervised learning is challenging to evaluate and unsuited for specific tasks like classification or prediction, semi-supervised learning has been introduced (Hady & Schwenker, 2013; Zhou & Belkin, 2014). Here, the advantages of both training categories are combined, and the algorithm learns from both labeled and unlabeled data (Saravanan & Sujatha, 2018). A standard method used in unsupervised learning is called co-training. Here, a model is trained to predict the desired output based on a relatively small labeled data set. This model is then used to make predictions on the unlabeled training data, which act as an extended training set to improve the model's performance (van Engelen & Hoos, 2019; Zhou & Belkin, 2014). The underlying idea of semi-supervised learning is that the labeled data assist the model in recognizing patterns in the unlabeled data. Semi-supervised learning has shown to be helpful in many applications, especially when large labeled datasets are difficult to obtain. While the performance of a model trained semi-supervised still depends on the quality and quantity of labeled data, like in supervised learning, the amount of labeled data is often smaller and usually comprises around 1% to 10% of the whole training set (Ouali, Hudelot, & Tami, 2020). Supervised learning was used in the second publication P-II (Jungmann et al., 2022), on which this thesis is based. Here, expert labels were available for all images contained in the training, validation, and test set. This training method was chosen due to the availability of input-output pairs and the easy validation of the algorithm using performance metrics comparable to assessing the performance of human readers in the study.

Validation and Testing

After each training step, also called an epoch, the algorithm is validated on an additional dataset to assess its performance on former unseen data. The validation set is usually created by dividing all available training data into a training and a validation set, often using a random four-to-one ratio (Rashidi et al., 2019). The model's performance on the validation set is used to monitor for over- or underfitting, which may occur during training (Chartrand et al., 2017).

Overfitting constitutes a significant challenge in training neural networks (Srivastava et al., 2014). It occurs when a model memorizes all training data instead of capturing the underlying patterns of information and thus performs poorly on the test set while achieving perfect results on the training data (Jabbar & Khan, 2015; Ying, 2019). Many aspects can lead to overfitting, including noisy, unrepresentative, too few training data, or a too complex model for the task (Hawkins, 2003; Ying, 2019). Multiple approaches have been introduced to tackle the problem of overfitting, including expansion of training data, early stopping, dropout, network reduction, or regularization, the detailed explanation of which lies beyond the scope of this thesis (Jabbar & Khan, 2015; Shorten & Khoshgoftaar, 2019; Srivastava et al., 2014). In contrast to overfitting, a model that underfits cannot learn from the training data and performs poorly on the training or validation set (Zhang, Zhang, & Jiang, 2019).

It must be noted that due to relatively small training data sets in medical applications, models are commonly pretrained on other data, like open-source image collections like ImageNet (Russakovsky et al., 2015), rather than being built from scratch (Rashidi et al., 2019). After training the neural network is completed, the model's performance should be further evaluated on an independent test set, which contains data not used during training or validation, to assess its generalizability (Chartrand et al., 2017; Rashidi et al., 2019). Different performance measures are suited for different algorithms. One of the most commonly used metrics

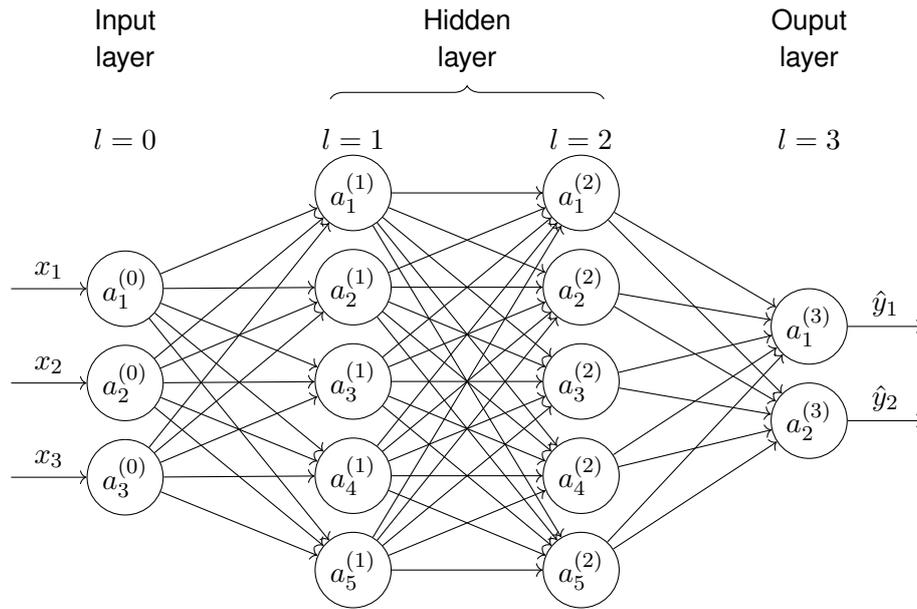


Figure 2.3 Schematic architecture of a multilayer perceptron with input x , two hidden layers and output \hat{y} . For lucidity, the input layer consists of three nodes, each hidden layer of five and the output layer of two nodes. However, most neural networks used nowadays are more complex and can have multiple hundreds of layers and nodes per layer.

is accuracy. However, there are other measures such as the area under the receiver operating curve (Penny & Frost, 1996), F-score (Goutte & Gaussier, 2005), or Matthews correlation coefficient (Chicco & Jurman, 2020).

2.4.2 Main network architectures

As explained in Section 2.4, neural networks are computational models inspired by the behavior and structure of brains and consist of multiple layers of interconnected nodes. There are numerous neural network architectures, which differ in the way information flows through different layers and in the type of input data and performance tasks, they are most suited for. The two main architecture classes based upon different input types are sequential neural networks, which are designed to handle sequential input data like video or time series data, and standard networks, which are designed to handle individual, independent data, like tabular data or images (Lipton, Berkowitz, & Elkan, 2015). This category encompasses a wide range of network architectures, two of which will be explained in the following. Firstly, multilayer perceptrons will be explained to demonstrate the general function of neural networks. Secondly, CNNs, as a predominant network architecture used in image classification tasks, will be introduced (Sultana, Sufian, & Dutta, 2018). Lastly, the residual neural network (ResNet), a subtype of CNNs, which was used in the second paper P-II (Jungmann et al., 2022), on which this thesis is based, will be described.

Multilayer Perceptron

A multilayer perceptron, also known as a fully connected neural network, is a type of feedforward neural network consisting of an input layer, one or more hidden layers, and an output layer (Goodfellow, Bengio, & Courville, 2016). Its underlying architecture is depicted in Figure 2.3.

Each layer l consists of multiple processing units $a_i^{(l)}$, called nodes, with $i = [0, 1, \dots, n_l]$ where n_l is the number of nodes in the layer l . Nodes in the input layer $l = 0$ represent the single data points of an input sample x , for example, tabular features or the pixel array of an image. Meanwhile, the output layer generates the predicted output \hat{y} (Chartrand et al., 2017; LeCun, Bengio, & Hinton, 2015). In a fully connected network, each neuron $a_i^{(l)}$ of a layer l is interconnected with each neuron of the following layer $l + 1$ (Gardner & Dorling, 1998). Each node $a_i^{(l)}$ computes a weighted sum of its input features $a^{(l-1)}$, which is, in turn, passed through a non-linear activation function ϕ to calculate the output of the node, which is then passed on to nodes of the following layer $a_i^{(l+1)}$ (Chartrand et al., 2017). Thus, each node $a_i^{(l)}$ can be represented using the following equation: $a_i^{(l)} = \phi(Z_i^{(l)}(a^{(l-1)})) = \phi(w_i^{(l)} * a^{(l-1)} + b_i^{(l)})$, where $a^{(l-1)}$ are all outputs of the previous layer $l - 1$. Each node $a_i^{(l)}$ has its own weight $w_i^{(l)}$ and bias $b_i^{(l)}$ needed for computing its output $a_i^{(l)}$ and the weights $w^{(l)}$ and biases $b^{(l)}$ of all nodes in a layer l represent the learnable parameters θ_l of this specific layer. Non-linear activation functions enable the network to perform complex tasks which could not be solved using linear functions (Sharma, Sharma, & Athaiya, 2020). Some commonly used activation functions include:

- Rectified linear unit (ReLU) $\phi(a) = \max(0, a)$. The output of this function is the input value a , for $a > 0$, or 0 otherwise. Thus, ReLU thresholds all negative values to zero. (Glorot, Bordes, & Bengio, 2011; Goodfellow, Bengio, & Courville, 2016).
- Sigmoid function $\phi(a) = \frac{1}{1+e^{-a}}$. Here, the output ranges between 0 and 1, which can be used to produce the parameters of a Bernoulli distribution (Goodfellow, Bengio, & Courville, 2016).
- Hyperbolic tangent function $\phi(a) = \tanh(a)$, which performs a linear transformation of the input a to the $(-1, 1)$ interval (Glorot, Bordes, & Bengio, 2011).
- Softmax function $\phi(a) = \text{softmax}(a) = \frac{e^{a_i}}{\sum_j e^{a_j}}$. This vector-to-vector transformation computes the probability distribution over a finite set of outputs. It is usually used in the output layer of neural networks trained for a single-label multi-class classification task (Chen et al., 2020c; Goodfellow, Bengio, & Courville, 2016).

Although the computation performed in each node is quite simple, very complex functions can be learned by the superposition of multiple nodes due to their properties explained above (Chartrand et al., 2017).

Training a deep learning algorithm is a circular process, repeating the following steps: Forward pass, validation, and updating of the network's parameters. This procedure is called backpropagation (Chartrand et al., 2017). In the forward pass, an input image is passed into the first layer of the neural network to ultimately predict the desired outcome label \hat{y} , for example, a class prediction (Chartrand et al., 2017). During this process, multiple intermediate variables are calculated in the hidden layers of the neural network, and information flows from the first to the last layer to compute the output metric. Afterward, the calculated output value \hat{y} is compared against the ground truth y , meaning the actual label of the input image, by computing the loss function for the given example (Mutasa, Sun, & Ha, 2021). The nearer \hat{y} and y are, the smaller the value of the loss function is for the given example, which corresponds to a minimization problem that has to be tackled during training (Goodfellow, Bengio, & Courville, 2016).

Numerous loss functions differ in calculating the neural network error and the type of problem for which they are most suited. Common loss functions encompass:

- Mean squared error $L(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2$. It is commonly used for regression tasks and assesses the algorithm's performance by calculating the average squared difference between prediction \hat{y} and ground truth y (Goodfellow, Bengio, & Courville, 2016).
- The binary cross-entropy loss function computes the difference between the probability distribution of the predicted label \hat{y} and actual label y belonging to the positive class using the following equation: $L(y, \hat{y}) = -y \log(p(\hat{y})) - (1 - y) \log(1 - p(\hat{y}))$, where $p(\hat{y})$ is the probability of \hat{y} belonging to the positive class if $y \in \{0, 1\}$ (Goodfellow, Bengio, & Courville, 2016). It is typically used for binary classification problems and was used in the second publication P-II (Jungmann et al., 2022), on which this thesis is based.
- Categorical cross-entropy loss is similar to binary cross-entropy, but for multiple classes $y \in \{0, 1, \dots, n\}$. It is a standard loss function used in multi-class analysis and measures the difference between the actual probability distribution of the data and the predicted probability distribution of the model (Chen et al., 2020c).

To optimize the predictive performance, an algorithm minimizing the loss function is used to iteratively adjust the parameters of the neural network during training. For this, the gradient of the loss function with respect to the model's weights and biases is calculated, which are then updated accordingly (Goodfellow, Bengio, & Courville, 2016). This process is referred to as backpropagation (LeCun, Bengio, & Hinton, 2015). As with loss functions, various optimization algorithms have been developed, each with different advantages and disadvantages. Some popular optimization algorithms are (Ruder, 2016):

- stochastic gradient descent (SGD)
- Nesterov's accelerated gradient (NAG)
- adaptive moment estimation (Adam)

When using SGD, the gradient of the loss function with respect to the parameters of the network is computed for each training example (x, y) , and the parameters are updated following the slope of the loss function to reach the global minimum (Ruder, 2016). NAG additionally incorporates the previous and current gradients, called momentum, to accelerate convergence to the global minimum, as SGD tends to oscillate around local minima (Kingma & Ba, 2014). Briefly, Adam further refines the process of minimizing the loss function by storing a decaying average of past squared gradients and using adaptive learning rates (Kingma & Ba, 2014). In the second paper P-II (Jungmann et al., 2022), on which this thesis is based, Nesterov-accelerated adaptive moment estimation (Nadam) was used (Dozat, 2016). It is an optimization algorithm based upon NAG and Adam and is explained in Section 3.2.

Convolutional Neural Networks

A convolutional neural network (CNN) is a neural network that applies convolution instead of general matrix multiplication in one or more layers (Goodfellow, Bengio, & Courville, 2016). This kind of network was developed to work with input data of a grid-like structure, including time-series as 1D grid data or images consisting of a 2D pixel grid (Goodfellow, Bengio, & Courville, 2016).

A CNN usually consists of an input layer, multiple convolutional blocks, and one or more fully connected layers at the bottom, mapping the learned features to the given set of output labels (Soffer et al., 2019). The schematic architecture of a CNN is shown in Figure 2.4.

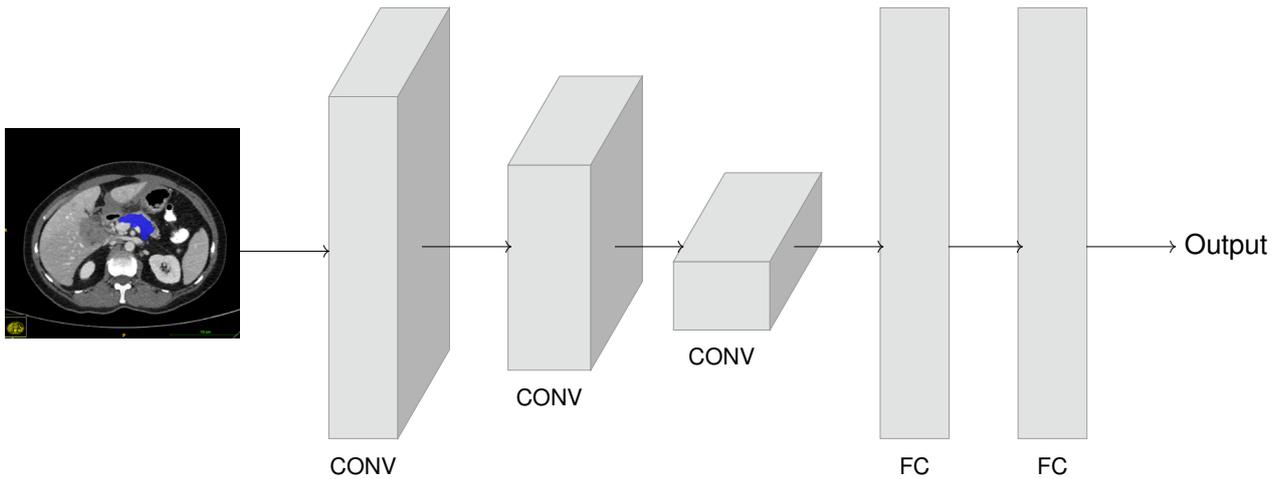


Figure 2.4 Schematic architecture of a convolutional neural network with multiple convolutional blocks (CONV) followed by fully connected layers (FC). For simplification purposes, no dimensions or missing layers are depicted. The CNN takes an image as input to predict the desired output.

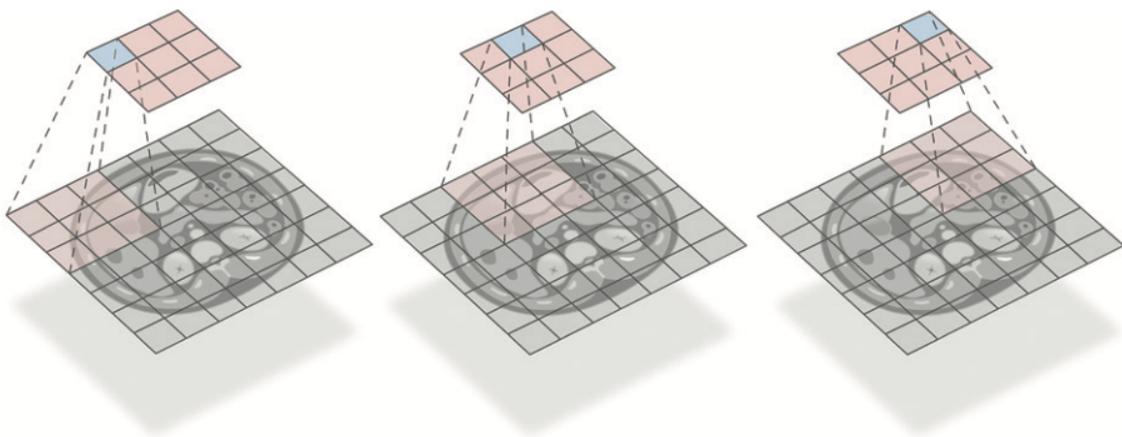


Figure 2.5 Illustration of a convolution from an input image to the output layer, from (Soffer et al., 2019).

A convolutional block usually consists of three operations: (1) The convolutional operation, commonly followed by a normalization step like Batch Normalization (Ioffe & Szegedy, 2015), (2) a non-linear activation function, and (3) a pooling layer (Goodfellow, Bengio, & Courville, 2016).

In the convolution operation, a kernel consisting of a tensor of learnable parameters is multiplied with the input matrix to produce a feature map. Convolution for discrete, two-dimensional inputs is often defined as follows:

$$s[i, j] = (I * K)[i, j] = \sum_m \sum_n I[m, n]K[i - m, j - n]$$

Here, a 2D kernel K with corresponding indices $[i, j]$ is multiplied with the 2D input I of the shape of $m \times n$ to produce the output matrix s (Goodfellow, Bengio, & Courville, 2016). For discrete convolution, the kernel K usually can be represented as a sparse (Toeplitz) matrix with several entries set to be equal to other entries, mostly zero, as K is usually smaller than the input I . Due to this, the infinite summation of convolution can be implemented as a summation of a finite number of matrix elements (Goodfellow, Bengio, & Courville, 2016).

Figure 2.5 shows a graphical explanation of this process. Here, the kernel K slides over the input image I , and element-wise matrix multiplication is performed between the kernel entries and those parts of the input matrix I , over which K currently lays. Afterward, the multiplication results are summed up and stored in a single entry of the output matrix s . In the next step, K slides to the next position, and the process is repeated until the whole input matrix I has been scanned (Li et al., 2021a). This process is applied in every convolutional layer where different levels of abstraction are learned each. As a result, the kernels of the different layers usually contain different sets of learned parameters (Goodfellow, Bengio, & Courville, 2016). For example, the first layer could detect horizontal edges in an image, while the second layer detects vertical edges (Mutasa, Sun, & Ha, 2021; Sarvamangala & Kulkarni, 2021).

After normalizing the output of convolutional layers to make the training more robust to changes in the distribution of parameters from the previous layers, a non-linear activation function is applied to adjust the current layer's output and control the activation of the units of the next layer. Today, ReLU is one of the most commonly used activation functions, as defined in Section 2.4.2. It thresholds all negative values to zero, generating several benefits, like cheap computation or tackling the problem of vanishing gradients (Albawi, Mohammed, & Al-Zawi, 2017), further addressed in the following subsection.

In the pooling layer, a filter with a fixed size is used for downsampling to reduce the number of parameters and prevent overfitting (Soffer et al., 2019). Here, a function is used to summarize the output of neighboring output units, for example, by returning the maximum or average value of neighboring units (Goodfellow, Bengio, & Courville, 2016).

The architecture of CNNs leads to beneficial properties for usage in image analysis. Firstly, due to its design for grid-like inputs, CNNs are predestined for usage in image analysis (Goodfellow, Bengio, & Courville, 2016).

Secondly, CNNs are more memory efficient and have lower computational cost than fully connected neural networks due to sparse interactions and parameter sharing (Goodfellow, Bengio, & Courville, 2016). In this context, computational cost refers to the complexity of tasks needed to perform certain operations within the network. Sparse interaction is achieved by using a kernel with size k , much smaller than the size of the input layer. In a convolution, only k units of the output layer are affected by an input unit in contrast to the interaction between each input unit with every output layer unit, as in the case in fully connected layers (Goodfellow, Bengio, & Courville, 2016). Thus, fewer parameters must be stored, and fewer operations are required to compute each step, reducing memory usage and runtime, meaning the time required for training. As the same kernel is used for matrix multiplication in every part of the image, only one set of parameters needs to be learned instead of individual sets of parameters for every location of the input layer. This is called parameter sharing and further contributes to efficient memory usage and runtime optimization (Goodfellow, Bengio, & Courville, 2016).

Thirdly, the ways of sharing parameters in convolutional operations make the convolutional layer equivariant to translations. As a result, the same function can be used in every part of the input, even if the input is shifted translationally. For example, a kernel detecting edges in an image can still detect edges in a motive, even if the object has been shifted in the input image (Goodfellow, Bengio, & Courville, 2016).

Furthermore, equivariance to translation is achieved by the pooling layer contained in a convolutional block. As the pooling functions summarize the output of several neighboring output units, the output of the pooling layer does not change if the input is translated by a small step. This is an important feature for

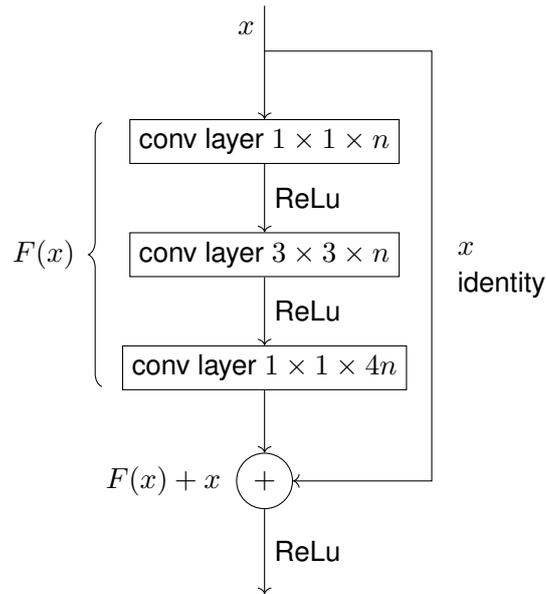


Figure 2.6 Building of a residual block in a ResNet 50 with input x and layer-specific dimension n of the convolutional layers. A skip connection performing identity mapping of x is added to the output $F(x)$ of the convolutional layers.

tasks like image classification, where the location of the identified object in the image should not influence the performance of the neural network trained (Goodfellow, Bengio, & Courville, 2016). It must be noted, however, that convolution is not equivariant to other input transformations, such as rotation or scaling (Goodfellow, Bengio, & Courville, 2016).

Lastly, in contrast to traditional neural network architectures, CNNs can be used to analyze inputs with varying sizes as the number of times a kernel is applied to the input layer can be varied. Depending on the restrictions imposed on the network's output, architectural changes of the network, like inserting a pooling layer, may be needed to retain a fixed output size (Goodfellow, Bengio, & Courville, 2016).

To summarize, advantageous properties of CNNs for image analysis tasks include memory efficiency, low computational cost, the design for grid-like inputs, and equivariance to translations.

ResNet 50

Due to the advantageous properties of CNNs explained in Section 2.4.2, this type of deep learning architecture is the most commonly used type of neural network in image analysis tasks today (He et al., 2019b). In P-II (Jungmann et al., 2022), on which this thesis is partially based, a residual neural network (ResNet) as one type of CNN was used.

In brief, a ResNet uses skip connections to connect non-neighboring layers as shown in Figure 2.6. As a result, information from a layer l can be passed directly to a layer $l + i$ with $i \geq 2$, where they are added element-wise to the regular input of this layer (He et al., 2016). The same holds true for backpropagation, where the gradients are passed directly from layer $l + i$ to layer l .

This architecture poses some advantages over traditional neural network architectures. Firstly, ResNets with hundreds of hidden layers can be trained while having a lower complexity than much shallower CNNs as the skip connection adds no relevant additional computational cost (He et al., 2016).

Secondly, the skip connections were shown to facilitate network optimization during training. This can be attributed to the mitigation of the vanishing gradients (Veit, Wilber, & Belongie, 2016) and to stable backpropagation due to the maintenance of the gradient norm (Zaeemzadeh, Rahnavard, & Shah, 2021). Vanishing gradients refers to the problem of exponentially decaying gradients during backpropagation as a result of computing gradients using the partial derivative of the loss function, especially in deep neural networks (Hochreiter, 1998; LeCun, Bengio, & Hinton, 2015). As the magnitudes of gradients are used to update the parameter of a network's layers, vanishing gradients lead to minimal or no update of early layer parameters. This, in turn, can impede effective training (Basodi et al., 2020). Better preservation of gradients due to backpropagation of gradients through skip connections can alleviate the vanishing gradient problem and thus lead to improved network performance and faster convergence (Borawar & Kaur, 2023). Another advantageous property of ResNets is the usage of batch normalization in the residual blocks, which leads to improved network performance (Borawar & Kaur, 2023; Ioffe & Szegedy, 2015). Additionally, batch normalization contributes to tackling the problem of vanishing gradients (Basodi et al., 2020).

In the second paper P-II (Jungmann et al., 2022), on which this thesis is partially based, a ResNet 50 was used. A scheme of the model's architecture is shown in Figure 2.7. It consists of multiple residual blocks of different sizes, repeated a certain number of times, as shown. The residual blocks include three convolutional layers, each followed by batch normalization (not depicted) and ReLU as an activation function. The combination of 1×1 , 3×3 , and 1×1 convolutional layers of each residual block, as shown in Figure 2.6, acts as a bottleneck to reduce the number of matrix multiplications (He et al., 2016). If the size of the output feature map of a residual block is halved, the number of filters is doubled to maintain the information complexity of a layer (He et al., 2016). As described above, a skip connection is added to every residual block to enable information flow between non-neighboring layers. At the bottom of the residual blocks, average pooling is performed, and a single fully connected layer with 1000 nodes and a softmax output is added (He et al., 2016).

2.5 Pancreatic ductal adenocarcinoma

The first publication P-I (Kaissis et al., 2020a), on which this thesis is based, explored quantitative image analysis methods with regard to survival prediction in PDAC. This section provides general information on this tumor entity as well as the diagnostic and therapeutic options used in clinical routine.

2.5.1 Epidemiology

'Pancreatic ductal adenocarcinoma (PDAC), despite its relative rarity, remains among the deadliest tumor entities in the developed world. For instance, PDAC is the 4th leading cause of cancer-related death, while representing only 3 % of newly diagnosed cancer cases in the United States (Siegel, Miller, & Jemal, 2019)' (Kaissis et al., 2020a, p. 1). It is a malignancy usually observed in older patients. For example, the mean age of diagnosis in Germany is 76 years in females and 72 years in males (Zentrum für Krebsregisterdaten im Robert Koch-Institut, 2016a).

Despite extensive research into therapeutic targets and newly introduced therapy regimens, the five-year overall survival rate of patients with PDAC does not exceed 10 % (Howlader et al., 2017). Due to its

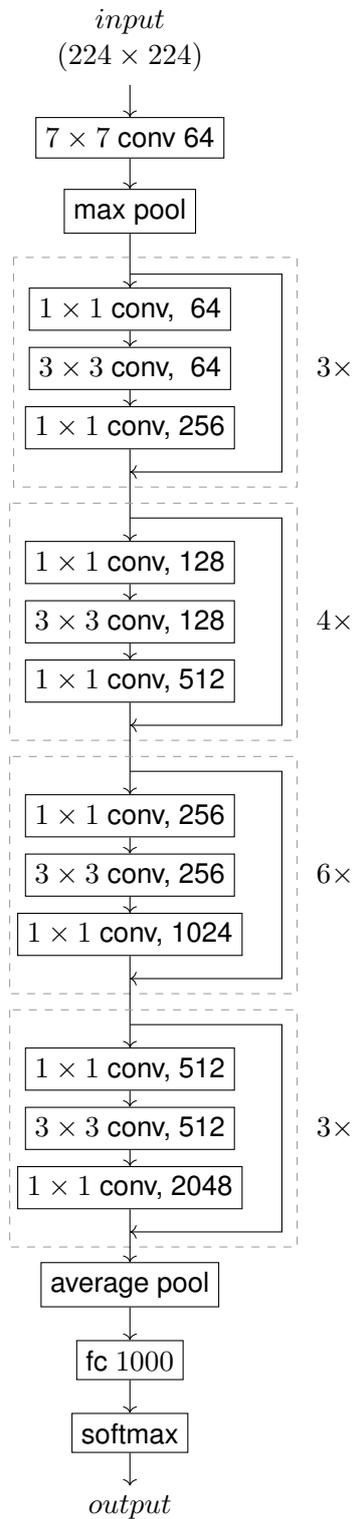


Figure 2.7 Schematic architecture of a ResNet 50, as used in publication P-II. The model takes input of the shape 224×224 and outputs the predictions generated by a softmax layer. Each convolutional layer (conv) is of the shape $x \times x \times n$, with x and n being layer-specific parameters. A residual block consists of three convolutional layers and a skip connection, as shown detailed in Figure 2.6. Each residual block (delimited by dashed lines) repeated a specific number of times, as shown. After the last residual block, average pooling is performed and a single fully connected layer with 1000 nodes (fc 1000) and a softmax output is added.

poor prognosis and rising incidence in developed countries, PDAC is predicted to be the second leading cause of cancer-related death by 2030 (Rahib et al., 2014). Despite standard-of-care treatment, patients often develop therapy-resistant tumor recurrence or metachronous metastasis. Furthermore, generally low response rates and different intra-individual responses to the common chemotherapy regimes have been observed (Aung et al., 2018; Muckenhuber et al., 2017). This is likely due to the heterogeneity of the tumor based on a genetic and histo-morphological level, as described in Section 2.5.2.

On a genetic level, four driver genes associated with different patient survival have been identified (Schlitter et al., 2017; Waddell et al., 2015). On a histo-morphological level, different subtypes based upon RNA (ribonucleic acid) expression and deoxyribonucleic acid (DNA) translocation profiling have been shown to have different response rates to the common chemotherapy regimes (Aung et al., 2018; Collisson et al., 2011) and a correlation to overall patient survival (Kaissis et al., 2020a; Muckenhuber et al., 2017).

2.5.2 Pathology

PDAC is a solid carcinoma originating from the epithelial cells of the pancreatic ducts and accounts for more than 90 % of pancreatic cancers (Fitzgerald et al., 2008) alongside other malignancies such as acinus cell carcinoma or malignant neuroendocrine tumors. It is most commonly located in the pancreatic head and consists of tubular glands and desmoplastic stroma (Schlitter et al., 2019).

Driver genes

Most PDAC arise from precursor lesions with an accumulation of multiple mutations in the sense of an adenoma-carcinoma-sequence (Ottenhof et al., 2011). Four driving genes for tumor genesis of PDAC have been identified (Biankin et al., 2012; Waddell et al., 2015):

- cyclin-dependent kinase inhibitor 2A (p16) (CDKN2A (p16))
- kirsten rat sarcoma virus (K-RAS)
- mothers against decapentaplegic homolog 4 (SMAD4)
- tumor protein p53 (TP53)

K-RAS is a proto-oncogene located at chromosome 12, which encodes a GTPase transductor protein. It represents an essential mediator in the epidermal growth factor receptor (EGFR) signaling pathway, which influences the differentiation and proliferation of cells (Jancik et al., 2010). Mutations of K-RAS occur in approximately 90 % of patients with pancreatic cancer (Biankin et al., 2012) and lead to permanent activity of K-RAS with resulting high cell proliferation rates. Many studies found the mutation to be correlated with reduced response to radiotherapy (Rengan, Cengel, & Hahn, 2008) and EGFR-targeting chemotherapies in various tumor entities (Jancik et al., 2010) and K-RAS mutation was associated with a reduced survival rate in patients with PDAC (Schlitter et al., 2017; Sinn et al., 2014).

The inactivation of the tumor suppressor gene CDKN2A (p16) plays a vital role in the carcinogenesis of multiple tumor entities (Liggett & Sidransky, 1998) and occurs in approximately 95 % of patients with PDAC (Ottenhof et al., 2011). CDKN2A (p16) is a cyclin-dependent kinase inhibitor protein that regulates the cell cycle and inhibits G1/S-transition (Biankin et al., 2012; Liggett & Sidransky, 1998). P16-mutation leads to

unregulated cell proliferation (Ottenhof et al., 2011) and can induce the overexpression of K-RAS (Jancik et al., 2010; Ju et al., 2017), whose function and effects of mutation were described above. Accordingly, studies found p16-mutations to be correlated with poor prognosis in patients with pancreatic cancer (Luo et al., 2012; Schlitter et al., 2017), higher levels of lymphatic invasion, and postoperative extensive metastasis (Oshima et al., 2013).

DNA damages in cells lead to activation of TP53, which prevents the progression of the cell cycle into the S-phase and ultimately induces apoptosis if the damages cannot be repaired. Thus, TP53 is classified as a tumor suppressor gene. Loss-of-function mutation of TP53 is one of the most frequent mutations in the genesis of multiple tumor entities and appears in 75 % of patients with PDAC (Bailey et al., 2016; Ottenhof et al., 2011). In pancreatic cancer, mutant TP53 was found to be associated with increased tumor fibrosis, reduced CD8⁺ t cell infiltration (Maddalena et al., 2021), promotion of metastasis (Morton et al., 2010) and resistance to gemcitabine-based chemotherapies (Fiorini et al., 2015).

SMAD4 is a transcription factor protein and part of the transforming growth factor beta (TGF- β) signaling pathway, which amongst others, induces cell growth and proliferation (Ottenhof et al., 2011). Furthermore, the loss-of-function mutation of SMAD4 is known to promote angiogenesis and immune suppression in tumors and occurs in more than 50 % of patients with PDAC (Ahmed et al., 2017). The inactivation of SMAD4 as a tumor suppressor gene is associated with widespread metastasis and shorter survival in patients with PDAC (Blackford et al., 2009; Iacobuzio-Donahue et al., 2009).

Molecular subtypes

On a histo-morphological level, prognostic relevant subtypes based upon immunohistochemistry (IHC) screening and transcriptome profiling have been identified: Classical, exocrine, and quasi-mesenchymal (QM) subtypes (Collisson et al., 2011; Muckenhuber et al., 2017). The IHC markers used for distinction of the different subtypes are hepatocyte nuclear factor 1A (HNF1A), a transcription factor involved in glucose metabolism and mediator of b cell differentiation (Lin Baochuan, 1998; Muckenhuber et al., 2017), and cytokeratin 81 (KRT81), a cytokeratin associated with different malignancies like non-small cell lung cancer and Epstein-Barr virus-associated lymphoma (Campayo et al., 2011).

The QM subtype is defined by high expression levels of mesenchyme-associated genes (Collisson et al., 2011) and expression of KRT81 (Muckenhuber et al., 2017). Patients with PDAC of the QM subtype were found to have a poorer prognosis than patients with other subtypes of PDAC in both adjuvant and palliative settings and no benefit from intensive chemotherapy regimes like FOLFIRINOX (Collisson et al., 2011; Muckenhuber et al., 2017; Noll et al., 2016). In contrast, a better response to gemcitabine-based adjuvant chemotherapy was observed in patients with QM subtype PDAC (Kaissis & Braren, 2019).

HNF1A expression is characteristic for the exocrine subtype, which expresses high levels of tumor cell-derived digestive enzyme genes (Collisson et al., 2011; Muckenhuber et al., 2017). This subtype was found to be associated with resistance to paclitaxel and tyrosine kinase inhibitors (Noll et al., 2016), increased response rates to FOLFIRINOX treatment, and an overall better prognosis in patients after tumor resection (Muckenhuber et al., 2017). However, the absence of HNF1A positive samples in different datasets has raised doubt on the reliability of this subtype (Collisson et al., 2011). Thus, and since survival advantage is lost in cohorts of patients with PDAC in palliative settings (Muckenhuber et al., 2017), the exocrine and classical subtype are often combined and referred to as non-QM subtype.

The classical subtype of PDAC is defined by high levels of adhesion-associated and epithelial genes (Collisson et al., 2011) and characterized by the absence of HNF1A and KRT81 expression in IHC (Noll et al., 2016). Studies found patients with classical subtypes to have a survival rate between those with QM or exocrine subtype (Muckenhuber et al., 2017). It is hypothesized that the classical subtype might be a mixture of different biological characterizations, and studies started further differentiation into multiple subgroups (Bailey et al., 2016).

2.5.3 Diagnosis

At diagnosis, patients with pancreatic cancer often have an extensive disease with metastasis or locally irresectable tumors due to rapid tumor growth and late clinical symptoms. Common symptoms are non-specific and include abdominal pain, loss of body weight, or night sweat (Abbassi & Algül, 2019; Kelsen et al., 1997). More specific symptoms, like newly diagnosed diabetes or painless icterus, may not always be present or detectable only in advanced tumor stages, thus impeding early tumor detection (Abbassi & Algül, 2019; Keane et al., 2014).

If pancreatic cancer is suspected, a step-wise diagnostic workup is performed. According to the guidelines of the German cancer institute, an ultrasound of the abdomen is followed by contrast-enhanced CT or MRI to enable local assessment and metastasis detection. Further evaluation of the local tumor can be achieved using endosonography (Deutsche Krebsgesellschaft, 2013). Other diagnostic parameters include the tumor markers carcinoembryonic antigen (CEA) and carbohydrate antigen 19-9 (CA19-9), which were shown to be both predictive for the extent of the tumor mass and patient survival after tumor resection (Distler et al., 2013; van Manen et al., 2020).

The different prognostic histopathologic characteristics of PDAC as described in Section 2.5.2 underline the urge for pre-therapeutic patient stratification to enable individual treatment. However, this has not proven easy in clinical routine for several reasons. Firstly, tumor biopsy always bears the risk of branch canal metastasis, which might hinder complete tumor resection. Thus, biopsy via fine-needle aspiration is primarily performed in palliative settings only. Secondly, pathologic workup relies on the high quality of tumor tissue, which is both time-consuming and expensive (Kaissis et al., 2019). Thirdly, the lack of transcriptomic assays, which are both robust and clinically suitable, results in sampling errors (Kaissis et al., 2020a).

Lastly, PDAC exhibits a high tumor heterogeneity which is marked as a mix of tumor cell clusters with different levels of cellularity or even different subtypes, as well as desmoplasia, which consists of cancer-associated fibroblasts and extracellular matrix components (Jungmann et al., 2021; Kaissis et al., 2019). Thus, a biopsy is prone to undersampling and inefficient capturing of tumor heterogeneity (Qi et al., 2018). Recent studies evaluated the possibility of non-invasive tumor characterization and patient stratification in various tumor entities based upon quantitative image analysis (Kaissis et al., 2020a; Rathore et al., 2018; Zhou et al., 2018). These findings form the methodical foundation for this work.

2.5.4 Therapy

Therapy regimes for patients with pancreatic cancer differ between patients eligible for resection and those with irresectable disease. Thus, the differentiation of these two groups prior to therapy onset is a crucial

part of the diagnosis. For this, imaging modalities such as CT or MRI are of great importance in determining the extent of the disease.

Surgical Management

Complete tumor resection combined with adjuvant or perioperative chemotherapy remains the only curative therapy for patients with PDAC (Kaissis et al., 2019). However, only 10% to 20% of patients have a resectable tumor at the time of diagnosis (Abbassi & Algül, 2019). Criteria for resectability determine the likelihood for complete tumor resection and include vascular infiltration, local tumor spreading, lack of metastasis as well as comorbidities, and health condition of the patient (Tempero et al., 2017).

Different surgical techniques are performed depending on the tumor's localization in the pancreatic head, neck, or tail. The most common ones include (pylorus-preserving) partial pancreaticoduodenectomy ((pp-) Whipple) for cancer in the pancreatic head, distal pancreatectomy for tumors located in the pancreatic tail or total pancreatectomy for diffuse tumor spread (Hackert, Werner, & Büchler, 2012; Tempero et al., 2017). Recent studies have shown increased complete resectability rates and prolonged disease-free survival for neoadjuvant chemo- and radiochemotherapy in patients with borderline-resectable and locally advanced pancreatic cancer (Kunzmann et al., 2019; Scheufele, Hartmann, & Friess, 2019; Versteijne et al., 2020), which translated into adapted guidelines for the treatment of pancreatic cancer by the German Cancer Institute in 2021 (Deutsche Krebsgesellschaft, 2021b).

In patients who underwent tumor resection, a histopathologic workup of the resection specimen is performed to confirm the diagnosis of PDAC and to determine other critical pathologic characteristics of the tumor, like T (primary tumor size), N (involved (regional) lymph nodes), M (distant metastasis), G (grading of tumor cells) and R (resection boundaries). These characteristics are incorporated in the pathologic staging of PDAC according to the Union for International Cancer Control (UICC), which correlates with patient survival and serves as a foundation for the clinical decision of further patient therapy (Adsay et al., 2012; Cong et al., 2018).

Chemotherapy

Around 80% to 90% of patients have extensive disease with either widespread metastasis or locally irresectable cancer and thus are not eligible for primary surgery. Hence, chemotherapy is essential in treating patients with PDAC.

Chemotherapy is applied in three main settings: Neoadjuvant, adjuvant, or palliative. In patients with borderline-resectable or locally advanced, irresectable cancer, chemotherapy is given with a neoadjuvant intent to enable tumor resectability. If patients can undergo primary surgery, adjuvant chemotherapy starting within 12 weeks after surgery is recommended to prolong patient survival. Multiple studies showed a prolonged five-year patient survival of $\geq 20\%$ in patients who received adjuvant chemotherapy compared to 10% in patients who underwent tumor resection only (Neoptolemos et al., 2004; Oettle et al., 2013). Recently, prolonged patient survival was found in studies investigating possible benefits of perioperative chemotherapy, meaning chemotherapy is applied before and after surgery compared to adjuvant chemotherapy alone (Ettrich et al., 2022; Kunzmann et al., 2019). In case of metastasis or poor patient condition, chemotherapy is applied in a palliative rather than curative setting.

The standard chemotherapy regimes can be divided into two central schemes: 5-Fluorouracil (5-FU)- and gemcitabine-based. The gemcitabine-based chemotherapy regimes comprise gemcitabine monotherapy, gemcitabine/nab-paclitaxel, gemcitabine/erlotinib, and gemcitabine/capecitabin. As monotherapy with gemcitabine is the most tolerable of these three regimes, it is most often used in patients with poor health conditions (Abbassi & Algül, 2019). In patients with metastasized PDAC, gemcitabine/nab-paclitaxel increased both progression-free and overall survival of patients (Goldstein et al., 2015; Hoff et al., 2013).

First-line chemotherapy regimes based upon 5-FU encompass monotherapy with 5-FU, FOLFIRINOX (5-FU, folinic acid, irinotecan, and oxaliplatin), and mFOLFIRINOX, which is a modified version of FOLFIRINOX. FOLFIRINOX has greatly increased the progression-free and overall survival time of patients in both adjuvant and palliative settings, however on the costs of higher toxicity compared to gemcitabine monotherapy (Conroy et al., 2018; Romero, 2019). To reduce toxicity to enable therapy for patients with lower health conditions, mFOLFIRINOX was introduced and showed superiority to gemcitabine monotherapy concerning both overall and progression-free survival (Abbassi & Algül, 2019; Seufferlein & Ettrich, 2019). For example, one study showed a prolonged overall survival (OS) of patients treated with mFOLFIRINOX of 20 months and increased progression free survival (PFS) of 9 months compared to patients treated with gemcitabine monotherapy in adjuvant settings (Conroy et al., 2018). Monotherapy with 5-FU is not recommended, as it was shown to be inferior to gemcitabine monotherapy (Burris et al., 1997).

Targeted therapy

Numerous studies investigated the effect of other molecular agents, like tyrosine kinase inhibitors or epidermal growth factor receptor inhibitors, in the treatment of patients with pancreatic cancer, as targeted therapies have shown remarkable advances in the treatment of other malignancies like non-small cell lung cancer or malignant melanoma (Kumarakulasinghe, van Zanwijk, & Soo, 2015; Lorentzen, 2019). Most studies investigated the effect of such substances in patients with metastasized PDAC, mostly in combination with gemcitabine-based chemotherapy. However, no significant advantage in patient survival could be achieved (Ciliberto et al., 2016; Tong et al., 2019).

A subgroup of patients, which benefit from targeted therapy, are patients with BRCA-mutated PDAC. BRCA-1/-2 mutation is associated with an increased risk of pancreatic cancer and is one of the primary mutations observed in familial PDAC (Kowalewski et al., 2018). Studies found patients with such mutations to significantly benefit from platinum-based chemotherapy (O'Reilly et al., 2020). Some studies further found maintenance therapy, including the poly-ADP-ribose polymerase inhibitor olatinib, to have a possible advantage in PFS (Golan et al., 2019). However, no advantage of olatinib in addition to platinum-based chemotherapy was found (O'Reilly et al., 2020).

As a result, targeted therapy is not recommended as first-line therapy for patients in clinical routine (Deutsche Krebsgesellschaft, 2021b).

Radiochemotherapy

The role of radiochemotherapy is still subject to extensive research. Until 2023, multiple studies investigated the possible advantages of radiochemotherapy in patients with locally irresectable pancreatic cancer. However, no consensus has been found so far.

Some studies found radiochemotherapy to be advantageous for local tumor control and OS (de Geus et al., 2017; Loehrer et al., 2011). In contrast, other studies found no significant differences between patients receiving radiochemotherapy compared with chemotherapy only (Ambe et al., 2015; Hammel et al., 2016; Stocken et al., 2005). In some studies, patients treated with radiochemotherapy even had a poorer OS than the control group (Chauffert et al., 2008). However, increased toxicity was reported for patients receiving radiochemotherapy, with exceptionally high rates of gastrointestinal side effects (Chauffert et al., 2008; Loehrer et al., 2011).

Further investigation is needed to define possible advantages and patient groups profiting from radiochemotherapy.

2.6 Breast Cancer

The second publication P-II (Jungmann et al., 2022), on which this thesis is based, explored possible use cases of quantitative image analysis methods in clinical routine by analyzing different types of AI-based assistance for lesion classification in mammography images. This section provides general information on breast cancer and its diagnosis, focusing on mammography screening.

2.6.1 Epidemiology

Breast cancer is the most commonly diagnosed type of cancer in women, with about 2.3 million new cases yearly. It has recently surpassed lung cancer as the most common type of cancer overall (Sung et al., 2021). It accounts for about a quarter of all cancer cases in women (Ferlay et al., 2014), and the lifetime risk for women to develop breast cancer is about 12 % (Zentrum für Krebsregisterdaten im Robert Koch-Institut, 2016b). While breast cancer is commonly diagnosed in women, about 1 % of all cases occur in men.

Although overall increasing worldwide, the incidence of breast cancer varies by geographic region and is highest in developed countries (Ginsburg et al., 2017). However, over the last years, the incidence has especially been increasing in countries of the global south while being stable or even decreasing in developed countries (Torre et al., 2017). This can be partly attributed to the higher prevalence of risk factors for breast cancer in developed countries, which will be further explained below, and partly due to improved screening programs (Sung et al., 2021). This increases the number of breast cancers diagnosed and especially leads to a higher number of cancers identified in the early stages. This and breakthroughs in treatment options have led to a significant decrease in breast cancer mortality in developed countries over the last decades. For example, in the United States, the 5-year survival of patients with breast cancer was 91 % in 2019 compared to 75 % in 1978 (Siegel et al., 2023), which corresponds to a decline of 43 % in mortality rate between 1989 and 2020 (American Cancer Society, 2023b).

Breast cancer is more commonly found in older women, with most cases diagnosed in the age group 55 to 64 years (Rojas & Stuckey, 2016). In Germany, the mean age of diagnosis is 65 years (Zentrum für Krebsregisterdaten im Robert Koch-Institut, 2016b).

Until today, several risk factors for breast cancer have been identified, which can be divided into modifiable and non-modifiable factors. Modifiable risk factors include smoking and alcohol consumption, obesity,

menopausal hormone use, oral contraceptives, childbearing, and breastfeeding. Non-modifiable risk factors encompass a longer reproductive life span, family history, genetic risk factors, female sex, and higher age (Key, Verkasalo, & Banks, 2001; Winters et al., 2017). Furthermore, breast cancer incidence and mortality rates vary among racial and ethnic groups. For example, in the United States, breast cancer is more common among white women than among Black, Hispanic, or Asian women (Rojas & Stuckey, 2016).

2.6.2 Pathology

Breast cancer is a heterogeneous disease with multiple driver genes and subtypes with distinct clinical, pathological, and molecular characteristics.

Main Subtypes

Several molecular and histological subtypes of breast cancer have been identified until today. Ductal carcinoma in situ (DCIS), invasive ductal carcinoma (IDC), and invasive lobular carcinoma (ILC) are three of the main subtypes of breast cancer, each with its unique features (Sims et al., 2007).

DCIS is a non-invasive precursor lesion of breast cancer that arises from the epithelial cells lining the breast's milk ducts. The cancerogenic cells in DCIS are confined to the ductal system and have not invaded the surrounding breast tissue (Ryan et al., 2017). DCIS is characterized by the presence of abnormal, proliferating cells within the ducts and is considered a pre-cancerous lesion that has the potential to progress to invasive breast cancer (ductal or lobular) (Martínez-Pérez et al., 2017; Wiechmann & Kuerer, 2008). The cells of DCIS are typically arranged in a linear, branching pattern and may contain calcifications or necrotic debris (Giuseppetti, Boria, & Baldassarre, 2018; Thomson et al., 2001). The nuclei of the neoplastic cells in DCIS are typically uniform and have a low mitotic rate, indicating slow cellular division (Pinder, 2010). DCIS can be further classified into different subtypes based on its genetic markers, such as luminal A, luminal B, or triple-negative (Clark et al., 2010), or based on its architecture, such as solid, cribriform, or micropapillary subtypes (Ryan et al., 2017).

IDC is the most common type of breast cancer, constituting about 80% of all cases (Watkins, 2019). It arises from the epithelial cells lining the milk ducts, but unlike DCIS, the malignant cells have invaded the surrounding breast tissue (Winters et al., 2017). These cells form irregular nests or cords of cells surrounded by fibrous stroma. The nuclei of the cancer cells in IDC are often pleomorphic, meaning they vary in size and shape, and have a high mitotic rate, indicating rapid cellular division (Mallon, 2000).

The molecular and genetic characteristics of IDC can vary, and different subtypes of IDC have been identified based on their molecular profiles, such as luminal A, luminal B, human epidermal growth factor receptor 2 (Her2-neu)2-positive, and basal-like subtypes, also called triple-negative. Luminal A IDC is characterized by the expression of both hormone receptors (estrogen receptor (ER) and progesterone receptor (PR)) and low levels of Ki67, a marker of cellular proliferation. The luminal B subtype is defined by expression of ER, Her2-neu negative and either higher levels of Ki67, missing or low expression of PR. In contrast, the Her2-neu-positive subtype is characterized by the overexpression of Her2-neu protein and ER. IDC of the triple-negative/basal-like subtype lack hormone receptors and a Her2-neu protein expression while having a high proliferation rate (Gao & Swain, 2018; Goldhirsch et al., 2013). Luminal A IDCs tend to

have a better prognosis than other subtypes, while triple-negative/basal-like tumors are associated with more aggressive disease and lower patient survival (Foulkes, Smith, & Reis-Filho, 2010; Gao & Swain, 2018). Due to the correlation of these subtypes with patient prognosis, the hormone receptor and Her2-neu status are usually determined in clinical routine to enable personalized treatment.

ILC is the second most common type of breast cancer with around ten percent of all cases (Mallon, 2000). It arises from the epithelial cells of the lobular units of the breast, which produce milk. Unlike DCIS and IDC, ILC is characterized by the infiltration of cancer cells into the stroma, or supportive tissue of the breast, rather than the ductal system. Cells with a distinct 'signet ring' appearance, in which the nucleus is pushed to the periphery of the cell by large cytoplasm vacuoles, can often be seen. The nuclei of the cancer cells in ILC are typically small and uniform, with a low mitotic rate (Mallon, 2000).

The diffuse growth pattern with the neoplastic cells arranged in a linear, single-file pattern, or 'Indian file' pattern, with minimal surrounding fibrous stroma can make ILC challenging to detect by imaging studies (Brem et al., 2009; Mallon, 2000). Thus, ILC is often diagnosed at a later stage than IDC and is generally less responsive to chemotherapy and radiation therapy (Katz et al., 2007).

The molecular and genetic characteristics of ILC are distinct from IDC and other breast cancer subtypes. They can include gene alterations such as CDH1, which is involved in cell adhesion and migration. Furthermore, ILC is characterized by the loss of the cell adhesion protein E-cadherin (Katz et al., 2007), which is responsible for maintaining the integrity of the mammary gland, and the activation of the PI3K/AKT/mTOR pathway, which regulates cell growth and survival (Brouxhon et al., 2013). ILC can also be classified into different subtypes based on their morphology and molecular characteristics, such as classic, pleomorphic, and solid subtypes (Katz et al., 2007).

Driver genes

As explained above, breast cancer is a heterogeneous disease with diverse genetic and molecular subtypes (Sims et al., 2007). Until today, several key genetic mutations and driver genes for the development and progression of breast cancer have been identified.

Among the most well-known genetic mutations associated with breast cancer are the breast cancer 1 (BRCA1) and breast cancer 2 (BRCA2) genes. These genes are involved in DNA repair and maintenance by repairing DNA double-strand breaks through homologous recombination, and mutations in either gene can increase the risk of developing breast and ovarian cancer. Mutations in BRCA1 and BRCA2 lead to genomic instability and are associated with a significantly increased risk of developing different types of cancers, such as ovarian and pancreatic cancer, and especially hereditary breast cancer, which accounts for approximately 5% to 10% of all breast cancer cases.

According to estimates from the National Cancer Institute, women with a BRCA1 mutation have a lifetime risk of developing breast cancer of up to 72%, compared to a lifetime risk of about 12% in the general population Kuchenbaecker et al., 2017; Zentrum für Krebsregisterdaten im Robert Koch-Institut, 2016b. Women with a BRCA2 mutation have a lifetime risk of up to 69% for developing breast cancer, which is slightly less than in women with BRCA1 mutations but still significantly higher than the risk of the general population Kuchenbaecker et al., 2017. The age of onset of breast cancer is typically younger in women with mutations in BRCA1 or BRCA2 than in women without mutations Kuchenbaecker et al., 2017. Furthermore,

especially women with mutations in BRCA1 have a higher risk of developing triple-negative breast cancer, which tends to be more aggressive and more challenging to treat than other types.

Her2-neu encodes a protein involved in cell growth and division, and mutations occur in approximately 20 % of breast cancers, leading to increased proliferation and survival of cancer cells (Loibl & Gianni, 2017; Yarden, 2001). Her2-neu modification or amplification can provoke the activation of downstream signaling pathways, which promote cell growth and survival and may result in uncontrolled cell proliferation (Yarden, 2001). Targeted therapies such as trastuzumab and pertuzumab have been developed to specifically target Her2-neu-positive breast cancer, leading to improved outcomes for patients with this subtype of breast cancer (Loibl & Gianni, 2017; Oh & Bang, 2019).

Another important genetic mutation associated with breast cancer is the PIK3CA gene (Kadota et al., 2009). It is a proto-oncogene that encodes the catalytic subunit of the PI3K enzyme, which regulates cell growth and survival through activation of the PI3K/AKT/mTOR pathway (Paplomata & O'Regan, 2014). Mutations lead to constitutive activation of the PI3K pathway, promoting cell growth and survival (Gymnopoulos, Elsliger, & Vogt, 2007; Kalinsky et al., 2009). Mutations in PIK3CA are found in approximately 30 % to 40 % of all breast cancer cases and are particularly common in hormone receptor-positive breast cancers (Campbell et al., 2004; Zardavas, Phillips, & Loi, 2014).

The TP53 gene is another well-known driver gene associated with breast cancer. TP53 is a tumor suppressor gene that plays a crucial role in regulating the cell cycle and preventing cancer development by regulating cell cycle arrest, apoptosis, and genomic repair in response to DNA damage (Shahbandi, Nguyen, & Jackson, 2020). Mutations in TP53 are found in up to 30 % of all breast cancer cases (The Cancer Genome Atlas Network, 2012). They are associated with more aggressive tumors and poorer prognosis, as these mutations can lead to impaired DNA repair and genomic instability, promoting the development and progression of cancer (Mathias et al., 2020). In addition, TP53 mutations can confer resistance to chemotherapy and other treatments, making these cancers more challenging to treat (Gao et al., 2020).

The CHEK2 gene encodes a protein called checkpoint kinase 2, which is involved in DNA damage repair, apoptosis, and cell cycle regulation. Mutations in the CHEK2 gene have been associated with an increased risk of several types of cancer, including breast cancer, prostate cancer, and colorectal cancer (Apostolou & Papatotiriou, 2017; Nevanlinna & Bartek, 2006). Women who carry the 1100delC mutation have a lifetime risk of breast cancer that is estimated to be between 20 % and 25 %, which is higher than the general population but lower than women with BRCA1 or BRCA2 mutations (Boonen, Vreeswijk, & van Attikum, 2022; Schmidt et al., 2016).

In addition to these well-known mutations and driver genes, recent advances in genetic sequencing technology have allowed the identification of additional genetic mutations frequently mutated in breast cancer (Kadota et al., 2009; The Cancer Genome Atlas Network, 2012). These include the ESR1 gene, which encodes the estrogen receptor alpha and is commonly mutated in hormone receptor-positive breast cancers (Brett et al., 2021), CDH1, which encodes E-cadherin, a protein involved in cell adhesion and migration, and is frequently mutated in ILC (Grabenstetter et al., 2020) and the PTEN gene, which plays a role in cell signaling pathways for migration and cell proliferation and is commonly mutated in triple-negative breast cancer (Chai et al., 2022). Numerous other mutations associated with breast cancer have been identified until today. However, the explanation of them all is beyond the scope of this thesis.

2.6.3 Diagnosis

Breast cancer can present with various symptoms, although not all women with breast cancer will experience the same signs. The most common symptom is a palpable lump or mass in the breast tissue, which may be painless or tender to touch (Watkins, 2019). Other symptoms include changes in breast size or shape, skin dimpling (peau d'orange), nipple changes such as inversion or discharge, erythema, and breast pain (American Cancer Society, 2023b; Chen et al., 2011; Hester, Hortobagyi, & Lim, 2021; Sharma et al., 2010; Yamauchi et al., 2012). These local symptoms may be accompanied by systemic symptoms such as fatigue, weight loss, and swelling of lymph nodes in the armpit or collarbone area if cancer has spread to other parts of the body (American Cancer Society, 2023b; Bardwell & Ancoli-Israel, 2008; National Cancer Institute, 2020b).

However, the symptoms of breast cancer are not specific to the disease and can be caused by a variety of other conditions (National Cancer Institute, 2020b), and many women with breast cancer may not have any symptoms in the early stages. Therefore, regular screening is essential to enable early detection and treatment of breast cancer, which can improve outcomes and reduce the risk of complications (Heywang-Köbrunner, Hacker, & Sedlacek, 2011; Starker, Kraywinkel, & Kuhnert, 2017). Mammography screening is a standard tool used for the early detection of breast cancer, which will be further explained in Section 2.6.4.

The first step in diagnosing breast cancer is a clinical breast examination, where a doctor palpates the breast tissue, looking for lumps or other abnormalities. The examination also includes the examination of the lymph nodes in the armpit (Deutsche Krebsgesellschaft, 2021a). Women are generally advised to regularly perform a self-examination of breast tissue and armpit in order to detect new abnormalities like palpable masses (American Cancer Society, 2023a).

Imaging studies are an essential part of the diagnostic workup of breast cancer (American Cancer Society, 2023a; Deutsche Krebsgesellschaft, 2021a). Mammography is the most commonly used imaging modality for breast cancer screening and will be explained in Section 2.6.4. In addition to mammography, breast ultrasound and MRI are other imaging modalities used in diagnosing breast cancer (Deutsche Krebsgesellschaft, 2021a).

Ultrasound is a non-invasive medical imaging technique that uses the absorption and reflection of high-frequency sound waves to create real-time images of internal organs and tissues (Chan & Perlas, 2011). In the context of breast cancer diagnosis, it can be used to evaluate suspicious breast lumps or abnormalities detected in the clinical examination or additional to mammography screening in high-risk women who are unable to undergo MRI (Kuhl et al., 2005) or in women with dense breast tissue (Nothacker et al., 2009). Ultrasound is helpful in evaluating breast lumps because it can help determine whether a lump is solid or fluid-filled (cystic) and provide information about the size, shape, and location of the lump (Candelaria et al., 2013). In addition, ultrasound can guide a needle for a biopsy procedure, which involves taking a small tissue sample from the breast for further analysis (Candelaria et al., 2013). One advantage of ultrasound is that it is non-invasive and does not expose patients to ionizing radiation, unlike mammography, and that it is widely available at low costs (American Cancer Society, 2023a; Sanches, Laine, & Suri, 2012). However, ultrasound is time-consuming and less effective than mammography for detecting small or early-stage breast cancers. Furthermore, it is more operator-dependent, meaning that the quality of the images can vary depending on the skill of the technician performing the exam (Berg et al., 2008; Duric et al., 2012; Kuhl et al., 2005).

Another imaging technique, which can be used to detect breast cancer, is MRI. Here, a strong magnetic field and radio waves are used to produce detailed images of the body's internal structures based upon differences in proton spin (Beitzel et al., 2017). Like ultrasound, MRI can be used to evaluate breast abnormalities that are detected on other imaging modalities or to screen high-risk women who may have an increased risk of developing breast cancer (American Cancer Society, 2023a; Lee et al., 2010). MRI is particularly useful for evaluating breast abnormalities in women who have dense breast tissue, as well as for detecting small breast cancers that may not be visible on mammography or ultrasound (Morrow, Waters, & Morris, 2011; Sardanelli et al., 2004). It can also be used to evaluate the extent of disease in women who have already been diagnosed with breast cancer and to monitor response to treatment (Morrow, Waters, & Morris, 2011). However, it is important to note that MRI is not a replacement for mammography or ultrasound and is generally used as a complementary imaging modality (Deutsche Krebsgesellschaft, 2021a). While MRI is highly sensitive for detecting breast cancer, it can produce false-positive results, leading to additional testing and unnecessary biopsies (American Cancer Society, 2023a).

After identifying suspect lesions in the imaging techniques explained above, an image-guided biopsy can be performed. During a biopsy, a small amount of tissue is removed from the breast and examined under a microscope to identify malignant cells and to determine other cancer characteristics, such as hormone receptor status or tumor subtype (Deutsche Krebsgesellschaft, 2021a). The most common type of biopsy used to diagnose breast cancer is the core needle biopsy. However, other types of biopsies are possible, including fine-needle aspiration biopsy, vacuum-assisted biopsy, or surgical biopsy (American Cancer Society, 2023a; Park & Hong, 2014).

2.6.4 Mammography Screening

Women are offered to join mammography screening programs in many countries to detect breast cancer in early, curative stages (Lehman et al., 2017). Those programs differ in aspects such as the age of women offered to participate or the interval in which the inspection is undertaken. For example, women aged 50 to 70 years in the United Kingdom are offered a mammography every three years, while in Germany, these women are invited every two years and in the United States, the American Cancer Society recommends annual mammography screening already starting between age 40 and 50 (Gemeinsamer Bundesausschuss, 2017; Lee et al., 2010; Marmot et al., 2013; Oeffinger et al., 2015). With widely offered mammography screening programs and improved mammography techniques, breast cancer mortality has been reduced by around 15% to 30% in many Western countries since 1990 (Heywang-Köbrunner, Hacker, & Sedlacek, 2011; Lee et al., 2010; Torre et al., 2017).

Mammography

The first systematic x-ray examinations of breast tissue were performed by Albert Soloman in 1913, who x-rayed mastectomy specimen to demonstrate the infiltrating growth of a tumor and its spreading to axillary lymph nodes (Soloman, 1913). During the 1930s, the first in-vivo mammographies were performed where the possibility to diagnose breast cancer and the possible distinction between benign and malignant lesions was shown (Vogel, 1932; Warren, 1930). Since then, intensive research has been performed to

improve image quality, reduce radiation exposure and develop systematic diagnostic criteria for interpreting mammograms (Gold, Bassett, & Widoff, 1990).

Full-field digital mammography is the standard-of-care method used in mammography screening nowadays. Here, the radiation is recorded by an electronic detector after passing through the breast tissue and processed digitally (Pisano & Yaffe, 2005). X-rays with low energies between 25–35kV are used to maximize the contrast in the breast's soft tissue to detect both benign and malignant lesions. For this, anodes comprised of molybdenum or tungsten are typically used. A moving grid is used to enhance radiographic sharpness, and a small x-tube focus, as well as a sufficiently large distance between the object and detector, is necessary. The breast is further compressed to reduce radiation dose and to enhance radiographic sharpness. Usually, an image is obtained from two directions: Craniocaudal (CC) view and mediolateral oblique (MLO) view, where the x-rays are parallel to the pectoralis muscle (Beitzel et al., 2017).

BI-RADS classification

The Breast Imaging-Reporting and Data System (BI-RADS) is a system to standardize the reporting of lesions in mammal tissue and is applied in mammography, MRI, and ultrasound images. It was initially introduced in 1993 by the American College of Radiology (ACR) and comprised a lexicon for describing anomalies, recommendations for the structure of a radiographic report as well as a framework for data collection (Balleyguier et al., 2007; Spak et al., 2017). Radiographic aspects described in the BI-RADS include breast density, shape, margin and density of masses, architectural distortions, and microcalcifications (American College of Radiology, 2013).

Usually, malignant lesions are dense structures compared to the breast's soft tissue. As a result, lesions are better detectable in fatty tissue than in fibroglandular one, as the contrast between dense tumor and fat is more enhanced compared to glandular tissue, which is more dense than fat. Fat content in breast tissue depends on the breast size and the age of a woman due to fatty involution of breast tissue in older age (Beitzel et al., 2017).

The density of breast tissue is described using the ACR classification as a subgroup of the BI-RADS assessment, which describes the relation between fat and fibroglandular tissue (American College of Radiology, 2013; Spak et al., 2017) and is taken into account to assess the validity of a mammogram report. Here, the breast density increases with higher ACR classes, where *ACR a* corresponds to predominantly fatty breast tissue, where mamograms usually produce good results. In contrast, *ACR d* describes highly dense breast tissue, which reduces the sensitivity and impairs the validity of findings (Beitzel et al., 2017; Spak et al., 2017; Weigel et al., 2016).

The final BI-RADS category of a mammogram is based upon a standardized description of all important image findings according to the lexicon provided and reflects the probability of malignancy of these findings (Obenauer, Hermann, & Grabbe, 2005). Furthermore, recommendations for further clinical actions are linked with the BI-RADS categories (Balleyguier et al., 2007). Table 2.1 provides an overview over the different BI-RADS categories (American College of Radiology, 2013).

In concordance with the ACR breast density classes, the likelihood of malignancy increases within the BI-RADS categories, where *BI-RADS 1* corresponds to breast tissue without any specific findings and *BI-RADS 5* refers to findings with a very high probability of malignancy (Mercado, 2014). In mamograms

Table 2.1 Definition of BI-RADS categories and linked management recommendations by the American College of Radiology, 2013.

Category	Likelihood of malignancy (%)	Assessment	Management recommendation
0	–	Incomplete	Additional imaging or comparison with prior examinations needed
1	≈ 0	Negative	Routine follow-up
2	≈ 0	Benign	Routine follow-up
3	0 - 2	Probably benign	Short interval follow-up
4	2 - 95	Suspicious	Tissue diagnostic (e.g. biopsy)
-A	2 - 10	Low suspicion for malignancy	
-B	10 - 50	Moderate suspicion for malignancy	
-C	50 - 95	High suspicion for malignancy	
5	≥ 95	Highly suspicious	Tissue diagnostic
6	100	Known biopsy-proven malignancy	Surgical excision when clinically appropriate

classified as *BI-RADS 2*, benign lesions are described which do not require any additional action, including calcified fibroadenomas, lipomas, or skin calcifications (American College of Radiology, 2013). Lesions classified as *BI-RADS 3* are most likely benign, but the stability of the lesion over time should be assessed to confirm benignity. Here, follow-up after a shorter interval is recommended (American College of Radiology, 2013).

As shown in Table 2.1, the *BI-RADS 4* category is divided into three subclasses *A - C*, defined by increasing the risk of malignancy for the described lesion. *BI-RADS 4* covers lesions that neither fulfill all malignant or benign findings criteria and thus require further tissue diagnosis, like fine-needle aspiration or biopsy. As this category covers a wide range of estimated likelihood of malignancy, it was divided into three subclasses with specific cut-off values in the fifth edition of BI-RADS classification. The intention for this was to enable a more informed decision on further clinical actions for both patients and clinicians (American College of Radiology, 2013; Mercado, 2014).

There are two additional categories, *BI-RADS 0* and *BI-RADS 6*, which represent exceptional cases. A mammography examination is classified as *BI-RADS 0* if further imaging evaluation is required to reach a final assessment. Further imaging evaluation includes additional imaging, like MRI or ultrasound, or comparison with prior examinations, which are not available at the time of the assessment. *BI-RADS 6* is used to describe lesions with biopsy-proven malignancy prior to surgical excision (American College of Radiology, 2013).

As the *BI-RADS 4* category includes a wide range of possible pathologies and poses a challenging task of estimating the likelihood of malignancy for radiologists, AI-based assistance for classifying *BI-RADS 4 A - C* lesions only was tested in the second publication P-II (Jungmann et al., 2022), on which this thesis is based.

2.6.5 Therapy

Breast cancer treatment is based on the disease stage, the tumor's characteristics, its subtype, and the individual patient's health status and preferences (American Cancer Society, 2023c). The primary modalities of breast cancer treatment include surgery, radiation therapy, and systemic therapies such as chemotherapy, targeted therapy, immunotherapy, and endocrine therapy, which will be explained in the following section.

Surgery

Surgery is usually the first treatment for limited-disease breast cancer, and it aims to remove all cancerous cells from the breast. The type of surgery used depends on the size and location of the tumor, as well as whether or not the cancer has spread to nearby lymph nodes. The type of surgery depends on the size and location of the tumor as well as the patient's preferences. There are two main types of surgery: Breast-conserving surgery and mastectomy.

In breast-conserving surgery, also known as a lumpectomy, only the tumor and a surrounding margin of healthy breast tissue are removed. It should be followed by radiation therapy to reduce the risk of local recurrence. Breast-conserving surgery may be recommended for women with DCIS or small carcinoma compared to their breast size. In these cases, it achieves similar survival rates as mastectomy.

Mastectomy, meaning the removal of the entire breast, can be recommended for several reasons, including a high risk of recurrence, inflammatory or multicentric breast cancer, prior radiotherapy, and cases where the tumor is too large for breast-conserving therapy or the patient declines adjuvant radiotherapy. Furthermore, a sentinel lymph node biopsy should be performed, and additional axillary lymphonodectomy can be necessary, depending on the tumor spread (American Cancer Society, 2023c; Deutsche Krebsgesellschaft, 2021a).

Radiation therapy

Radiation therapy is a treatment option for breast cancer that uses particle or photon radiation with high energy levels of 6 MV and higher to help destroy any cancer cells by inducing cell death through DNA breaks, inhibition of mitotic cell cycle or changes in the tumor microenvironment (Beitzel et al., 2017; Dutt et al., 2020). Adjuvant radiation therapy is the most effective treatment option to reduce the risk of intramammary cancer relapse, which in turn reduces the cancer-related mortality (Early Breast Cancer Trialists' Collaborative Group (EBCTCG), 2011a).

It is recommended after every breast-conserving surgery or may be part of the treatment plan after mastectomy (Deutsche Krebsgesellschaft, 2021a). External beam radiation therapy is the most common type of radiation therapy and is typically given daily for several weeks (Dutt et al., 2020). For example, in Germany, hypofractionated radiotherapy with 40 Gy or a fractured radiotherapy with 50 Gy, followed by a boost of the former tumor site, is recommended (Deutsche Krebsgesellschaft, 2021a). Another type of radiation therapy, known as brachytherapy, involves the placement of radioactive sources directly into the breast tissue near the tumor. This approach delivers a high dose focally while limiting the exposure of healthy tissue and its resulting complications. Brachytherapy is typically given in a single treatment or over

a few days (Chargari et al., 2019). Overall, the duration and intensity of radiation therapy depend on the patient's age, overall health, characteristics, and tumor spread (American Cancer Society, 2023c).

Chemotherapy

Chemotherapy is a systemic treatment for breast cancer. It may be used to shrink tumors before surgery (neoadjuvant), to reduce the risk of cancer recurrence after surgery (adjuvant), or to treat cancer that has spread to other parts of the body (palliative). Here, drugs are given to destroy tumor cells. The mechanisms of chemotherapeutic drugs include the inhibition of DNA synthesis, induction of DNA damage, or inhibition of the cell cycle (Chabner & Roberts, 2005). The field of different chemotherapy regimes and settings is complex and still subject to ongoing research. As therapeutic options for breast cancer were not part of the publications on which this thesis is based, only a short overview will be given in the following.

Adjuvant chemotherapy aims to destroy any cancer cells which may have remained after surgery and can lower the risk of cancer relapse (American Cancer Society, 2023c). The recommendation for adjuvant chemotherapy includes factors such as the size of the tumor, nodal status, grading of tumor cells, ER, PR or Her2-neu expression, and age of the patient (Early Breast Cancer Trialists' Collaborative Group (EBCTCG), 2005; Goldhirsch et al., 2011). Based upon multiple studies, the German Cancer Society recommends using a taxane and an anthracycline, which should be given over 18-24 months (Deutsche Krebsgesellschaft, 2021a).

In case of a locally advanced tumor, inflammatory breast cancer, or tumors which can not be resected primarily, neoadjuvant chemotherapy is recommended in order to enable resectability or even breast-conserving surgery due to reduced tumor size (Deutsche Krebsgesellschaft, 2021a). Furthermore, especially patients with triple-negative breast cancer seem to benefit from neoadjuvant chemotherapy (Cortazar et al., 2014; von Minckwitz et al., 2010). The chemotherapy regime should contain a taxane and an anthracycline (Deutsche Krebsgesellschaft, 2021a) like in adjuvant settings.

In case of distant metastasis, for example, into the lung or bone, an individual therapy concept is developed based on the patient's age, fitness, and wishes, as well as factors such as hormone status or tumor burden. Here, the aim is to improve the quality of life, so less toxic single-drug chemotherapy regimes are preferred (Deutsche Krebsgesellschaft, 2021a).

2.6.6 Endocrine Therapy

Hormone receptor-positive breast cancers (luminal A or B subtypes) express estrogen and/or progesterone receptors. Especially the binding of estrogen to these cells has been shown to promote tumor growth, mainly by stimulation of cellular proliferation (Kulendran, Salhab, & Mokbel, 2009). Thus, patients with hormone receptor-positive breast cancer should be treated with an endocrine therapy (Deutsche Krebsgesellschaft, 2021a; Early Breast Cancer Trialists' Collaborative Group (EBCTCG), 1998).

Endocrine therapy comprises antiestrogens, gonadotropin-releasing hormone (GnRH) agonists, and aromatase inhibitors (Lumachi et al., 2011). GnRH agonists and aromatase inhibitors prevent estrogen production, while estrogen antagonists impede the action of estrogen at a cellular level (Buzdar & Hortobagyi, 1998). Multiple studies have shown that antiestrogens, like Tamoxifen and aromatase inhibitors,

reduce mortality and recurrence rates (Early Breast Cancer Trialists' Collaborative Group (EBCTCG), 2005, 2011b). Tamoxifen, a selective ER modulator of the group of antiestrogens, is considered the first line endocrine therapy (Buzdar & Hortobagyi, 1998). It should be given after chemotherapeutic treatment for at least five years (Deutsche Krebsgesellschaft, 2021a).

2.6.7 Targeted Therapy

Targeted therapy of cancer refers to applying molecular agents, specifically targeting distinct enzymes, receptors, or signal transducers to obstruct oncogenetic cellular processes (Tsimberidou, 2015). Inhibitors of multiple pathway targets, including PI3K-AKT-mTOR pathway, CDK4/6, and EGFR inhibitors, have shown promising results in clinical trials (Mohamed et al., 2013).

Apart from endocrine therapeutic agents discussed above, Her2-neu targeting agents are most commonly used in clinical routine. The first approved Her2-neu targeting agent is Trastuzumab, which is a monoclonal antibody directed against an extracellular domain of Her2-neu and recommended for patients with Her2-neu positive breast cancer (Claret & Vu, 2012; Deutsche Krebsgesellschaft, 2021a). Multiple studies have shown an adjuvant therapy of Trastuzumab combined with chemotherapy to improve patient survival and reduce recurrence rates (Dahabreh et al., 2008; Joensuu et al., 2006; Petrelli & Barni, 2011).

Beside Trastuzumab, multiple other agents are targeting Her2-neu, and other targeted therapies exist, which are mainly used in hormone receptor-positive breast cancer subtypes. These agents include CDK4/6 inhibitors like palbociclib, mTOR inhibitors like everolimus, or PI3K inhibitors like alpelisib (American Cancer Society, 2023c). Furthermore, PARP inhibitors like olaparib and talazoparib can be used in patients with a BRCA1 or BRCA2 mutation and impede DNA repair, especially in mutated cells (American Cancer Society, 2023c).

There are no established targeted therapies for triple-negative breast cancer, as neither hormone receptor nor Her2-neu targeting agents is effective in these tumors (Mohamed et al., 2013). However, immune checkpoint inhibitors like pembrolizumab or atezolizumab, have been shown to be beneficial for patients with elevated levels of programmed death ligand 1 (PD-L1) survival in some studies, and atezolizumab in combination with nab-paclitaxel was approved by the US Food and Drug Administration in 2020 (Kwapisz, 2020). Multiple studies are currently investigating other possible targeting therapies (Lau, Tan, & Shi, 2022; Lyons, 2019). However, a more detailed description of these is beyond the scope of this thesis.

3 Methodology

This chapter describes the methodology used in the two publications, P-I and P-II, on which this thesis is partially based. In Section 3.1, the study design and procedures used for data curation and analysis in the first publication P-I (Kaissis et al., 2020a) are illustrated. Section 3.2 discloses the techniques applied for data pre-processing, network training, and in reader study as well as the statistics used in the second publication P-II (Jungmann et al., 2022).

3.1 Methodology of the first publication P-I

Sections 3.1.1 - 3.1.5 are adapted from P-I (Kaissis et al., 2020a), which is co-authored by the author of this thesis.

This retrospective cohort study 'was approved by the institutional ethics review board (Ethics Committee of the Technical University of Munich Faculty of Medicine, Protocol Number 180 / 17S; date of approval: 9 May 2017), and the requirement for individual written consent was waived' (Kaissis et al., 2020a, p. 4).

3.1.1 Study design

The study aimed to analyze the predictive value of both established clinical and histopathological data and quantitative imaging features obtained using a radiomics approach for assessing overall patient survival in a cohort of primary resected PDAC patients.

In brief, 177 patients with histopathologic confirmed PDAC were screened for eligibility. Seventy-four patients without sufficient data were excluded from the study, as shown in Figure 3.1. This yielded 103 patients to be analyzed. Exclusion criteria were: Insufficient image quality or incomplete imaging ($n = 35$), incomplete clinical data ($n = 23$), pre-existing malignant condition ($n = 5$), or a overall survival (OS) of shorter than six weeks after surgery ($n = 3$). All patients underwent curative intended surgery at the Klinikum rechts der Isar. Follow-up intervals lasted from October 2006 to April 2019.

3.1.2 Clinical and histopathological data

Clinical data were obtained from the hospital's information system and the national cancer registry. Histopathological data were gathered through genetic and histopathologic analyses of the resection specimen by the Department of Pathology of the Technical University of Munich. The molecular subtype was determined as quasi-mesenchymal (QM) or Non-QM, and tumor morphology and genetic alterations of

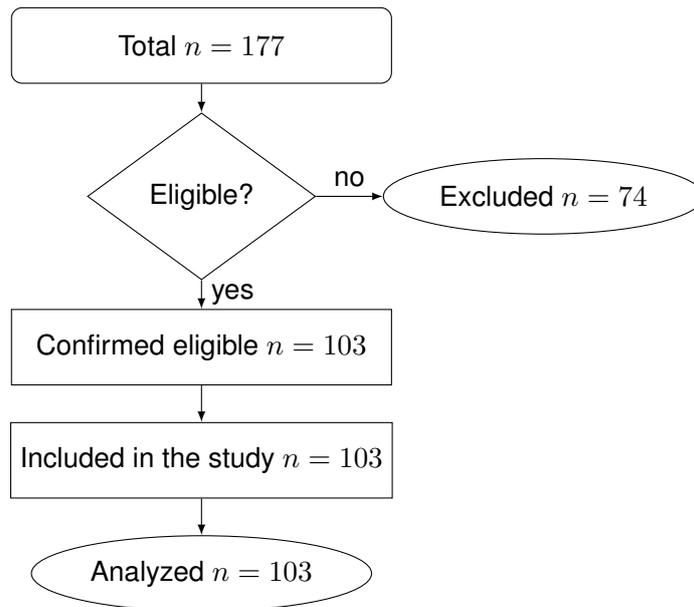


Figure 3.1 Patient inclusion flowchart of P-I, adapted from (Kaissis et al., 2020a).

TP53, KRAS, P16, and SMAD4 were analyzed according to previous literature (Muckenhuber et al., 2017; Schlitter et al., 2017). Table 3.1 displays the derived patient characteristics.

Fifty-one of the patients included in the study experienced progressive disease during the follow-up interval with a median PFS of 7.3 months. Fourteen patients (90%) encountered local recurrence, 29 patients (57%) developed liver metastasis, and eight patients (16%) both. 22 (43%) of patients with progressive disease received gemcitabine as first-line chemotherapy after progression, 13 (25%) FOLFIRINOX, and 16 (31%) no chemotherapy as part of a best supportive care regime. As the occurrence of progressive disease during follow-up and the type of treatment received afterward were not significantly associated with OS, they were not included in the final analysis.

3.1.3 Imaging Data

All patients received CT imaging before surgery containing a portal venous phase (acquired 70 s after injection of contrast agent). Imaging data were exported from the Picture Archiving Communication System (PACS) in a pseudonymized form to a segmentation workstation running ITK-SNAP version 3.6.0 (Yushkevich et al., 2006). Fully-automated pancreas segmentation was performed using in-house developed software, and all segmentations were corrected manually by an expert radiologist.

'Image-derived features were extracted using PyRadiomics version 2.2.0 with standard settings (van Griethuysen et al., 2017)' (Kaissis et al., 2020a, p. 3). In brief, intensity discretization was performed to a fixed bin number of 25 bins without applying normalization. Images were spatially resampled to $1 \times 1 \times 1$ mm using the BSpline interpolator. First-order statistics, shape-based, gray level run length matrix (GLRLM), gray level size zone matrix (GLSZM), neighbouring gray tone difference matrix (NGTDM), gray level dependence matrix (GLDM) and gray level co-occurrence matrix (GLCM) features were extracted from the images as described in Section 2.3.2. Feature extraction was performed on the original images as well as Laplacian of Gaussian-filtered (with Sigma values 1.0), wavelet-decomposition-based, square, square-root, exponential, gradient, and logarithm filtered versions of the images.

Table 3.1 Clinical and histopathological characteristics of patients included in P-I, adapted from (Kaissis et al., 2020a).

Variable	Classes	Frequency (%)
Sex	male	59 (57.2)
	female	44 (42.8)
Age	mean (years)	67.3
	range (years)	32 - 88
Adjuvant chemotherapy	gemcitabine	55 (53.3)
	none	48 (46.7)
Treatment response during follow-up interval	progressive disease	51 (49.5)
	stable disease	52 (50.5)
Type of progression	local recurrence	14 (27.4)
	liver metastasis	19 (56.9)
	both	8 (15.7)
Tumor location in pancreas	head	71 (68.9)
	body	19 (18.4)
	tail	13 (12.7)
Tumor size	pT 1/2	11 (10.6)
	pT 3/4	92 (89.4)
Lymph node status	pN 0	30 (30.1)
	pN 1	73 (70.9)
Grading	G 1/2	49 (47.7)
	G 3	54 (52.3)
Resection status	R 0	53 (51.4)
	R 1	50 (48.6)
Subtype	QM	16 (15.5)
	Non-QM	87 (84.5)
Tumor morphology	conventional	55 (51.4)
	combined	48 (46.6)
TP53	wild type	21 (20.3)
	mutated	82 (79.2)
KRAS	wild type	9 (8.8)
	mutated	94 (91.2)
CDKN2A/p16	intact	19 (18.5)
	altered	84 (81.5)
SMAD4	positive	41 (39.2)
	loss	62 (60.8)

In total, 1409 features were extracted from the images. Features yielding nil, constant, or missing values and those with a variance below 1×10^{-5} were excluded, resulting in 1384 features analyzed in the study. All features used in the analysis were normalized to the unit interval $[0, 1]$.

3.1.4 Survival Analysis

For multi-parametric evaluation of clinical, histopathological and image-based data, the following groups were analyzed in the study: First, clinical and standard histopathological data were obtained in clinical routine. This group comprises the age and sex of the patients, type of adjuvant chemotherapy received, size of the tumor, lymph node status, grading of the tumor, and resection status. Secondly, morpho-molecular and genetic data were analyzed, including tumor morphology, molecular subtype, and genetic alterations of TP53, K-RAS, SMAD4, and P16. Thirdly, imaging features derived from the baseline CT images using PyRadiomics were evaluated. Last, the combination of all parameter groups was assessed for additional benefit.

Several GLSZM features were associated with overall patient survival. As described in Section 2.3.2, these features describe the heterogeneity of the image texture. However, due to the large number of imaging features compared to the size of the study cohort, unsupervised machine learning-based dimensionality reduction was performed prior to survival analysis. For this, linear principle component analysis (PCA) was applied to the radiomic features to capture 99% of the parameter variance, resulting in 71 unique feature groups. Furthermore, univariate pre-selection using a log-rank-test statistic with a cut-off of $p < 0.1$ was applied, yielding eight imaging feature groups included in the final analysis.

To assess the predictive value of the feature groups described above, a 'Cox Proportional Hazards model was fit to each parameter group separately at first to assess the model concordance index (CI) after asserting that the proportional hazards assumptions were met. Parameters which did not meet the proportional hazards assumptions were excluded from the final analysis of the combined features' (Kaissis et al., 2020a, p. 3).

3.1.5 Statistics

Statistical analyses were performed using Python 3.7.6. A significance level of $\alpha = 0.05$ was chosen for all analyses, excluding the univariate pre-selection of features using the log-rank-test statistic (Rousseaux & Gad, 2013), in which $p < 0.1$ was considered significant (Emura, Matsui, & Chen, 2019). 'Five-fold hold-out cross-validation was employed in all analyses, and the concordance index (CI) across the five folds were averaged. 95% confidence intervals were calculated using 100-fold bootstrap resampling' (Kaissis et al., 2020a, p. 3).

3.2 Methodology of the second publication P-II

Sections 3.2.1 - 3.2.6 are adapted from P-II (Jungmann et al., 2022), which is co-authored by the author of this thesis.

This retrospective study was approved by the institutional ethics review board (Ethics Committee of the Technical University of Munich, Faculty of Medicine, Protocol Number 87/18S), and the requirement for individual written consent was waived.

3.2.1 Study Design

The study comprised two main parts: Network training and generation of AI assistance, followed by a reader study, in which the influence of the different methods of AI assistance on radiologists' performance was tested.

For the reader study, four radiologists from the Department of Diagnostic and Interventional Radiology of the Technical University of Munich were asked to classify 200 BI-RADS 4 mammograms into benign or malignant lesions. Furthermore, a ResNet 50 was trained to classify mammography images on a publicly available data set and validated on the independent test set containing 200 mammography images from our institution.

Based on the test set, different methods of AI-based assistance were created for the reader study: First, a heatmap highlighting image regions according to their contribution to the network's prediction. Secondly, an additional display of the network's prediction along with its certainty. After six weeks, the radiologists were asked to again classify all 200 images. Afterward, they were consecutively provided the two AI-based support methods. Each time, they could either remain with their original classification or change it.

In the last step, the readers' attitude towards AI applications was assessed based upon a questionnaire, and the association between the performance change and personality traits of the readers was analyzed. The study design is depicted in Figure 3.3.

3.2.2 Training, validation, and independent test datasets

For this study, a dataset of 2620 mammography images with was collected from the open database Curated Breast Imaging Subset of the Digital Database for Screening Mammography (CBIS-DDSM) (Lee et al., 2016). Only images containing non-calcified lesions were included, yielding a training set of 1696 images. The dataset provides histopathologically determined malignancy for each lesion, which was used as a target label. To extend the training set and provide more robust results, data augmentation was applied to the training set by flipping or rotating the images up to 90°, horizontal and vertical shifting of up to 20%, or scaling up to 20%. The data were randomly split into a set for training the ResNet and a validation set containing 20% of the images.

An independent test set of 200 full-field digital mammograms (CC and MLO views) from the Department of Diagnostic and Interventional Radiology of the Technical University of Munich acquired between February 2009 and February 2018 was collected to assess the model's generalization performance. The mammographies were selected to match the breast density and age of the patients from the CBIS-DDSM dataset and contained one non-calcified lesion originally classified as BI-RADS 4 each. Histopathologic labels obtained by biopsy within one month after the mammography served as ground truth. This process yielded in 200 mammographies included in the test set, thereof 101 showing malignant and 99 benign lesions. A bounding box was placed around all lesions to match the training set, and the images were adjusted to the size of the CBIS-DDSM images. 'By doing so, the cropped images simulate the output of an object detection algorithm applied to the mammogram, as common in AI applications'(Jungmann et al., 2022, p. 3).

3.2.3 Network training

A ResNet 50 with weights pre-trained on ImageNet (Russakovsky et al., 2015) was modified to classify the training set into benign or malignant lesions. Due to the relatively small cohort size, the neural network to classify the images was not trained from scratch, but transfer learning was applied. Transfer learning is a machine learning method that aims to use representations learned in one task in another similar one without the need for extensive new datasets (Lu et al., 2015). Here, the imperative of both training and test set coming from the same distribution or feature space is avoided (Pan & Yang, 2010). If the two tasks are related to some extent, representations from low-level features learned in the early layers of a network can be transferred to the new task and thus improve and accelerate learning. In our example, the rationale is to use the knowledge of low-level image features learned by the pre-trained ResNet 50, like edges, corners, or image texture, as a starting point for the task of mammography classification. For this, we only trained the fully-connected layer of the ResNet 50 on our training set.

The ResNet 50 was trained on batches of 64 images, each using binary cross-entropy as a loss. Nadam (Dozat, 2016) was used to optimize the loss function. It is a combination of the Adam optimizer and NAG, each described in Section 2.4, which leads to accelerated minimization of the cost function with improved convergence (Dozat, 2016). For this, the exponential decay of moving averages of the previous and current gradients (momentum) is added to the gradient descent function. The update rule for a given parameter θ of the network at each time step t is given as follows:

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_t} + \epsilon} * \hat{m}_t$$

for a given learning rate η and smoothing term ϵ to avoid division by zero. \hat{v}_t and \hat{m}_t represent the bias-corrected first and second momentum estimates, meaning the mean and uncentered variance of the gradients (Ruder, 2016).

The initial learning rate of 1×10^{-4} was chosen by visual observation of plotting the learning rate against the model loss after training the model for a few epochs with a linear increasing learning rate, a commonly used method based on a proposal by Smith et al. (Smith, 2017). Here, the value leading to the fastest loss decrease is chosen as the learning rate. The training was performed with a triangular learning rate policy, with the learning rate varying cycling between set boundaries. This method was shown to achieve good training results without needing to tune the learning rate as a hyperparameter (Smith, 2017). After five training epochs without an improvement in validation loss, early stopping was applied. The algorithm was trained to convergence, and the final model was selected based on the validation loss.

3.2.4 AI based assistance

In the reader study, different methods of AI-based assistance were tested concerning their influence on radiologists' performance in diagnosing mammograms.

'A heatmap highlighting the image's most important areas for classification was created by computing the occlusion-based sensitivity map for each image (Zeiler & Fergus, 2014). Here, pixels were colorized according to their contribution to the prediction' (Jungmann et al., 2022, p. 3) by monitoring the loss increase, corresponding to a decline in probability score for the predicted class. The magnitude of the probability

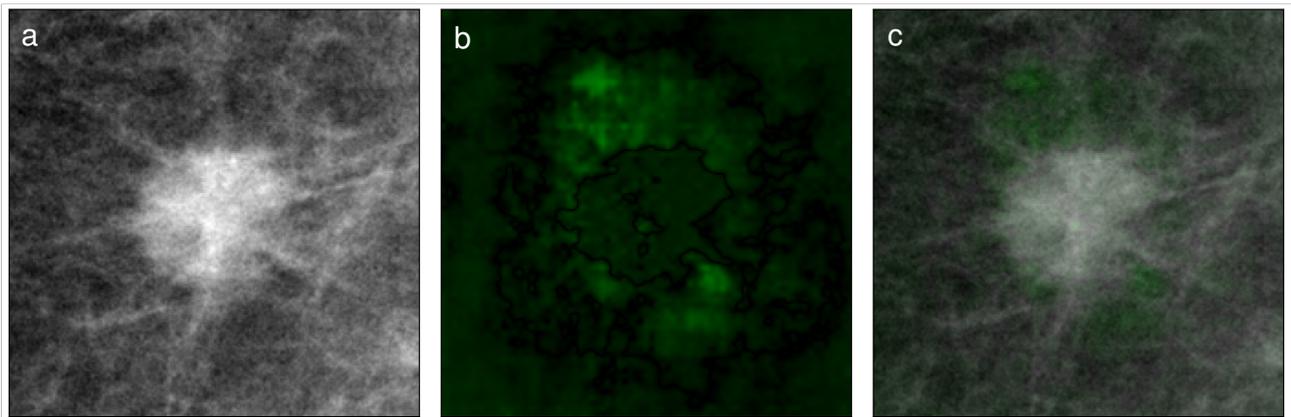


Figure 3.2 From (Jungmann et al., 2022). Heatmap assistance used in publication P-II on the example of a malignant lesion: a) Finding in mammogram, b) heatmap generated by the ResNet 50 and c) overlay of heatmap and mammogram.

decline correlates with the area's importance for the prediction. The importance of image regions was determined by shifting a 7×7 pixel occlusion map above the image and calculating the probability score for the predicted class in each step. In the study, the heatmaps were displayed both independently and overlaid with the cropped image from the original mammogram, as displayed in Figure 3.2.

'Furthermore, the calculated certainty of the ResNet 50's prediction using the classification scores was displayed. As the study excluded out-of-sample images, the certainty of the prediction was estimated by calculating the probability $[(P)]$ for the class 'malignant' and displayed thus or [as the complement $(1 - P)$] 1 - probability for benign classifications. In the study, a display of the probability of the predicted class ranging between 0% and 100% [along with the algorithm's classification] was added to the heatmap presentation' (Jungmann et al., 2022, p. 3).

3.2.5 Reader study

'Four radiologists from our institute were recruited for the study, which took place between June and September 2020. For the study, diagnostic monitors with a resolution of 2048×2560 pixels calibrated to the DICOM GDFS were used, and the ambient light was set below 50 lx. None of the participants had evaluated the mammograms used for the study during the last 27 months. To minimize decision bias from additional information sources, the radiologists were not given any clinical meta-information about the patients. The radiologists represented different stages of proficiency' (Jungmann et al., 2022, p. 3): One attending with a focus on mammography (20 years of experience), one attending with a focus on abdominal imaging (7 years of experience), one consultant radiologist (6 years of experience) and one radiology resident (2.5 years of experience).

The study design is shown in Figure 3.3. 'At first, the readers were asked to classify the cropped images simulating the output of an object detection algorithm applied to the mammograms of all 200 images of the test set into benign or malignant findings. They were given no additional clinical data but were told that all images contained lesions classified as BI-RADS 4 before. In a second step, after 6 weeks, the readers were asked to classify the images again to assess intra-reader reproducibility. Directly afterwards, the

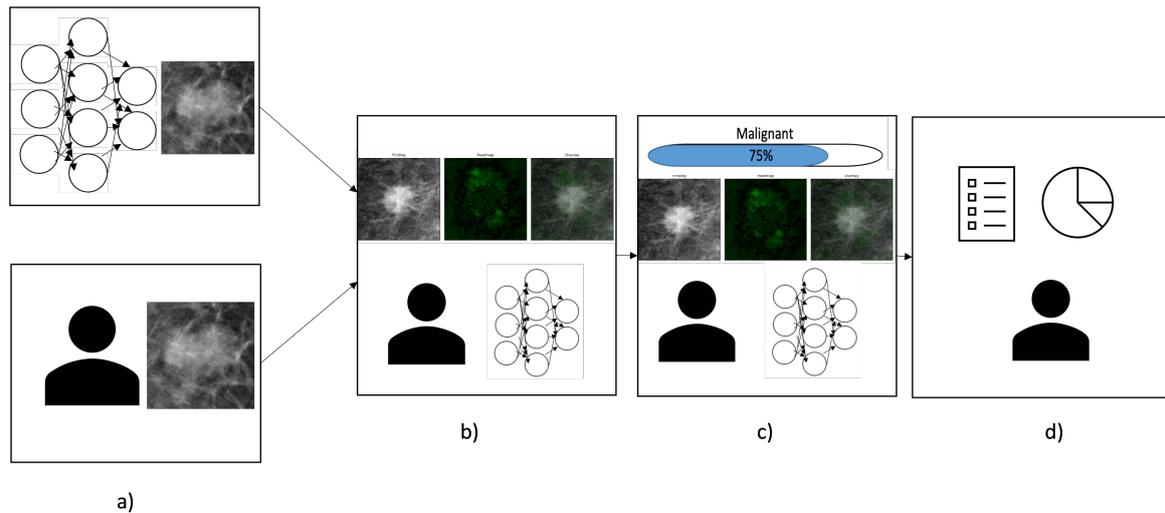


Figure 3.3 Study design of P-II. a) classification by human reader and ResNet 50 separately. b) Combined classification using heatmap. c) Combined classification with added prediction and certainty. d) Questionnaire and personality test. From (Jungmann et al., 2022)

radiologists were incrementally and additively provided with the different types of output display generated by the ResNet 50 (heatmap, prediction, and certainty), and human-AI- interaction was observed using a questionnaire and by an observing person during the session of the study. After each output presentation, the readers could either adhere to their original classification or change it' (Jungmann et al., 2022, p. 3).

'After performing the classification task, each reader was interviewed in a structured fashion and answered a questionnaire regarding their radiological experience and attitude towards AI applications in the medical field (...) We analyzed the possible influence of participants' personalities on human-AI-interaction and the resulting performance using a publicly available personality test (Enge & Gassoden, 2020) based upon the Big-five personality model (Goldberg, 1992; John & Soto., 2008; John & Srivastava, 1999) and items from the International Personality Item Pool (Goldberg et al., 2006). As these personality traits were found to be mostly stable in adults (Cobb-Clark & Schurer, 2012), we chose the model as being presumably independent from working experience and age' (Jungmann et al., 2022, p. 3).

3.2.6 Statistics

'Statistical analyses were performed using Python 3.8.2. A two-sided significance level of $\alpha = 0.05$ was chosen for all tests. For comparison of the radiologists' performance with and without AI assistance, the arithmetic mean over the readers' accuracy on each task was calculated, and McNemar's test (McNemar, 1947) with Yates correction (Yates, 1934) was calculated based on the classification changes made by the radiologists. In concordance with the International Personality Item Pool recommendation, analysis of personality traits and their inference with human-AI-interaction using Pearson's correlation was performed based on the absolute scores in each Big-five personality model category (Goldberg, 1992; Goldberg et al., 2006; Johnson, 2014)' (Jungmann et al., 2022, p. 4).

4 Journal Publications

4.1 Journal Publication P I: Multiparametric Modelling of Survival in Pancreatic Ductal Adenocarcinoma Using Clinical, Histomorphological, Genetic and Image-Derived Parameters

The publication *Multiparametric Modelling of Survival in Pancreatic Ductal Adenocarcinoma Using Clinical, Histomorphological, Genetic and Image-Derived Parameters* was published in the Journal of Clinical Medicine. It was authored by Georgios A. Kaissis, Friederike Jungmann, Sebastian Ziegelmayr, Fabian K. Lohöfer, Felix N. Harder, Anna M. Schlitter, Alexander Muckenhuber, Katja Steiger, Rebekka Schirren, Helmut Friess, Roland Schmid, Wilko Weichert, Marcus R. Makowski and Rickmer F. Braren.

4.1.1 Short summary

This section is adapted from P-I (Kaissis et al., 2020a), which is co-authored by the author of this thesis.

Purpose

This publication assesses the predictive value of established clinical and histopathological data and the added value of imaging features for the overall survival of patients with PDAC.

Material and Methods

This retrospective cohort study analyzed the overall survival of patients with PDAC resected at the Klinikum rechts der Isar of the Technical University of Munich. Clinical and survival data were retrieved from the hospital's information system and the national cancer registry, all histopathological analyses were performed by the Department of Pathology of the Technical University of Munich, and CT images were retrieved from the PACS. CT images were segmented fully automated and manually corrected afterward. Quantitative image features were derived using *PyRadiomics* version 2.2.0 and normalized to unity interval. To reduce the large number of radiomic features compared to the cohort size, PCA was employed to the data to describe 99 % of the parameter variance. The following parameter groups were analyzed for their predictive value: (1) clinical and standard histopathological data, (2) morpho-molecular and genetic data, (3) image-derived features, and (4) all features combined. All groups were tested for their prognostic strength using multivariate Cox proportional hazards survival modeling.

Results

The average CIs of the out-of-sample data were: (1) clinical and standard histopathological data: CI = 0.63 (95 % confidence interval = 0.60–0.66), (2) morpho-molecular and genetic features: CI = 0.53 (95 %

confidence interval = 0.47–0.59), (3) quantitative image features: CI = 0.65 (95% confidence interval = 0.62–0.69) and (4) combination of all features: CI = 0.65 (95% confidence interval = 0.60–0.69).

Conclusion

Quantitative image features were shown to be an independent predictor of postoperative overall survival in patients with PDAC. Although they do not outperform broadly available clinical data, they provide a non-invasive way of assessing patient survival. This may especially be interesting in patient cohorts with a neoadjuvant or palliative setting, as much histopathological data is often unavailable in this group. Here, quantitative image features could constitute a non-invasive, and therefore relatively risk-free, method of assessing the overall survival of patients.

4.1.2 Author's Contributions

The doctoral candidate contributed to the following aspects of the publication: Data curation, formal analysis, investigation, methodology development, project administration, software implementation, visualization, and manuscript writing.

4.2 Journal Publication P II: Algorithmic transparency and interpretability measures improve radiologists' performance in BI-RADS 4 classification

The publication *Algorithmic transparency and interpretability measures improve radiologists' performance in BI-RADS 4 classification* was published in *European Radiology*. It was authored by Friederike Jungmann, Sebastian Ziegelmayr, Fabian K. Lohöfer, Stephan Metz, Christina Müller-Leisse, Maximilian Englmaier, Marcus R. Makowski, Georgios A. Kaissis and Rickmer F. Braren.

4.2.1 Short summary

This section is adapted from P-II (Jungmann et al., 2022), which is co-authored by the author of this thesis.

Purpose

Recent developments in deep learning-based medical image analysis present new ways of including AI-based applications in clinical practice, which ultimately requires human-AI interaction. This publication evaluates 'the perception of different types of AI-based assistance and the interaction of radiologists with the algorithm's predictions and certainty measures' (Jungmann et al., 2022, p. 1).

Material and Methods

For this study, a dataset of 1696 mammograms from the open database CBIS-DDSM containing non-calcified lesions was collected. A ResNet 50 pre-trained on ImageNet was trained to classify the lesions into benign and malignant findings. An independent test set containing 200 mammograms from the Department of Diagnostic and Interventional Radiology of the Technical University of Munich acquired between 2009 and 2018 was collected. The mammograms were selected to match the breast density and age of the patients from the training set and contained one lesion classified as BI-RADS 4 each. A bounding box was placed around all lesions to match the training data, and the images were adjusted to those of the training set. Histopathologic labels obtained within a month after the mammography served as ground truth. Different types of AI assistance were generated: (1) An occlusion sensitivity map highlighting the image areas important for classification as well as (2) the prediction along with its certainty.

In a reader study, four radiologists with different levels of expertise from the Department of Diagnostic and Interventional Radiology of the Technical University of Munich were asked to classify the cropped images of the mammograms contained in the independent test set into benign or malignant findings. The radiologists were asked to classify the images again after six weeks to assess the intra-reader reproducibility. Directly afterward, they were stepwise provided with the different types of AI assistance created, first the heatmap, then, in addition, the algorithm's prediction and its certainty. 'After each output presentation, the readers could either adhere to their original classification or change it. After performing the classification task, each reader was interviewed in a structured fashion and answered a questionnaire regarding their radiological experience and attitude towards AI applications in the medical field' (Jungmann et al., 2022, p. 3), and took a publicly available personality test based on the Big Five model. Both the effect of different types of AI-based assistance on the radiologists' performance and the influence of the Big Five personality traits were analyzed.

Results

'Diagnostic accuracy was significantly improved by AI-based assistance (an increase of $2.8\% \pm 2.3\%$, 95%-CI 1.5% to 4.0%, $p = 0.045$) and trust in the algorithm was generated primarily by the certainty of the prediction (100 % of participants). Different human-AI interactions were observed, ranging from nearly no interaction to humanization of the algorithm. High scores in *neuroticism* were correlated with higher persuasibility (Pearson's $r = 0.98$, $p = 0.02$), while higher *consciousness* and change of accuracy showed an inverse correlation (Pearson's $r = -0.96$, $p = 0.04$)' (Jungmann et al., 2022, p. 1).

Conclusion

'In concordance with previous research, we found the performance of combined human and artificial intelligence superior to both individual groups and that displaying aspects of the classification process of the algorithm to readers leads to increased trust in the algorithm's performance. This highlights the importance of improved transparency and explainable AI systems. Furthermore, our findings indicate an association between different personality traits and human-AI interaction' (Jungmann et al., 2022, p. 7), a fact that still awaits deeper investigation.

4.2.2 Author's Contributions

The doctoral candidate contributed to the following aspects of the publication: Conceptualization, data curation, formal analysis, investigation, methodology development, project administration, software implementation, visualization, and manuscript writing.

5 Discussion

The primary objective of this thesis was to develop different methods of machine learning-based assistance to enhance diagnostic radiological performance and thus ultimately patient care. For this, two quantitative image analysis methods were used, representing current radiological research topics: Radiomics analysis and a deep learning approach using a neural network.

In the first publication, P-I (Kaissis et al., 2020a), radiomics-based quantitative imaging features were identified as an independent predictor for postoperative overall survival in patients with PDAC. The study underscored the potential benefit of quantitative imaging analysis in patient care enabling the non-invasive gathering of additional information and shows the importance of radiologists' contributions to personalized medicine.

In the second publication, P-II (Jungmann et al., 2022), a neural network was developed to classify BI-RADS 4 mammogram lesions into benign and malignant findings. The study showed a significant increase in radiologists' performance when provided with the AI-based assistance and that combined human and artificial intelligence outperformed both groups individually. Furthermore, different explainability methods increasing the transparency of the model's decisions lead to an increased trust of radiologists in the algorithm. This study highlights the potential benefit of integrating transparent, explainable AI-systems into future radiological workflows, especially as a second reader.

5.1 Findings and limitations of the publications

This section summarizes the findings of the two publications, which constitute the groundwork of this thesis. Furthermore, the limitations of the methods and findings of the publications are discussed.

5.1.1 First publication P-I

Section 5.1.1 is adapted from (Kaissis et al., 2020a), which is co-authored by the author of this thesis.

Findings

The first paper presents 'the results of a comprehensive multiparametric analysis for the prediction of overall patient survival in a group of resected PDAC patients' (Kaissis et al., 2020a, p. 7). Here, quantitative imaging features of the tumor were found to be an independent predictor of postoperative overall patient survival.

Quantitative imaging features were obtained using a radiomics approach as one of the two quantitative image analysis techniques explored in this thesis. For this, the regions of interest were segmented, and imaging features were extracted using PyRadiomics. Dimensionality reduction was performed using PCA to reduce the large number of features. At publication, this work was 'the first study to systematically evaluate a broad range of currently available PDAC risk biomarkers' (Kaissis et al., 2020a, p. 8), including quantitative imaging characteristics, clinical, histopathological, and morpho-molecular features.

'Imaging-derived features represent an independent survival predictor in PDAC and enable the multiparametric, machine learning-assisted modelling of postoperative overall survival with high performance compared to clinical and morpho-molecular/genetic parameters' (Kaissis et al., 2020a, p. 1). Although imaging features did not outperform broadly available clinical data, they underscored the potential of image-derived as additional, non-invasive biomarkers for clinical risk stratification in patients with PDAC.

Limitations

The study was subject to several limitations. First, the analyses were performed on a relatively small cohort size due to 'the high financial and personnel cost of the extensive morpho-molecular and genetic workup' (Kaissis et al., 2020a, p. 8) of PDAC specimen. This highlights the possible benefits of quantitative imaging biomarkers, which can be obtained non-invasively and based on routine patient diagnosis and follow-up imaging. Second, the high financial and personnel cost of specimen workup, to the extent that it is not part of the routine clinical workflow, also inhibited the collection of an external validation cohort so that the generalizability of the study's findings could not be assessed.

Third, due to the large number of quantitative imaging features compared to the cohort size, PCA analysis was performed. Although this approach reduces the risk of model overfitting, it prevents the identification of single, well-documented PyRadiomics features as described in section 2.3.2. Instead, imaging feature compositions were found to be an independent predictor of patient survival, which are more difficult to reassess in the following studies.

'Nevertheless, our results should be tested prospectively to ascertain their external validity. (...) A side-by-side direct comparison of whole tumor tissue and imaging-data-derived features will be required to adequately compare the true predictive performance of imaging against morpho-molecular tumor constituents. Furthermore, for multiparametric data integration (...) to reach clinical maturity, multi-centric cooperation will be required to attain sufficient sample sizes' (Kaissis et al., 2020a, p. 8).

5.1.2 Second publication P-II

Section 5.1.2 is modified from (Jungmann et al., 2022), co-authored by the author of this thesis.

Findings

The second publication investigated the perception of and interaction with AI-based assistance in radiologic workflows as an example of integrating deep learning-based quantitative image analysis techniques into

clinic procedures. It showed that the diagnostic accuracy of radiologist's performance on BI-RADS 4 lesion classification was significantly improved by combined AI-based assistance comprising an attention heatmap, binary malignancy prediction, and algorithm certainty indication. Trust in the algorithm's performance was dependent primarily on the certainty of the algorithm's prediction in combination with a reasonable heatmap, and 'the willingness to change a diagnosis upon AI-based assistance was most dependent on the AI certainty level. (. . .) In our study, the sole heatmap did not improve diagnostic accuracy and was deemed least useful' (Jungmann et al., 2022, p. 6) by radiologists. However, it was considered a 'good measure of the algorithm's prediction quality, indicating a positive effect on transparency and understandability' (Jungmann et al., 2022, p. 6).

Furthermore, different personality traits seemed to affect human-AI-collaboration. Radiologists with high neuroticism scores were more likely to change their classification according to the algorithm's prediction, which 'might be explained by the fact, that individuals scoring highly in this personality dimension tend to be more self-conscious. Furthermore, we observed a negative correlation between consciousness and performance improvement, which could indicate that more conscientious radiologists did not trust the AI support enough, instead relying on their own skills. [However,] this correlation may be confounded by more conscientious individuals already performing better without the model's assistance' (Jungmann et al., 2022, p. 6). Overall, 'the diagnostic performance of readers increased using AI-based assistance, and all participants stated that AI could act as a second reader in the possible future radiological workflow. In concordance with previous literature (Han et al., 2020; McKinney et al., 2020), our study thus supports AI as a second reader in radiological settings' (Jungmann et al., 2022, p. 6).

Limitations

Due to the exploratory character of the study, there are several limitations to be considered in this publication, which emphasize the necessity of further studies investigating human-AI-collaboration. As the study excluded patients with no detectable mass, its findings are not outright generalizable to mammography screening routine. As the study focused on assessing the impact of different types of AI-based assistance rather than object detection, only the cropped images were used in the reader study limiting the findings with respect to real-world applications.

'Furthermore, as we did not test another algorithm with different accuracy, we cannot evaluate the influence the accuracy of AI has on readers' decisions. As studies found faulty AI decreasing human performance in collaborative settings (Tschandl et al., 2020), further investigation of this aspect is required. Discrepancies between radiologists and the algorithm were not analyzed. To detect systematic failure or to identify cases in which the algorithm was most useful, further inquiry is needed' (Jungmann et al., 2022, p. 6).

'Moreover, the primary effect of the certainty display cannot be inferred as we evaluated it in combination with the heatmap only. As the analyzed types of algorithmic assistance were always presented in the same order, the participating radiologists may have withheld their change of decision based upon the heatmap awaiting the additional information from the prediction and its certainty, which could impose a significant confounding of the results. To eliminate this bias, future studies should vary the order in which radiologists are presented with different types of AI assistance. Our notion of uncertainty is limited because it does not cover out-of-distribution input and does not represent calibrated uncertainty (Nixon et al., 2019). Techniques based on Bayesian deep learning would have been capable to offer more robust uncertainty estimates

(Hüllermeier & Waegeman, 2021). Furthermore, we did not investigate the utilization of visual interpretability methods Grad-CAM (Selvaraju et al., 2017) and LIME (Ribeiro, Singh, & Guestrin, 2016), which have a similar scope' (Jungmann et al., 2022, pp. 6–7).

'Also, the communication with the algorithm, or lack of it, is likely to be influenced by radiologists knowing they were being observed and might not reflect their behavior in clinical practice. As all participants evaluated the images first without and later with AI assistance, the recall of images could be an influence on their performance. Lastly, due to the small number of readers, we cannot rule out confounding, especially regarding the influence of personality traits on the readers' improvement with AI assistance. To overcome these limitations and establish recommendations for improved human-AI interaction, studies with a larger number of participants and different set-ups are required. As the assessment of personality traits and interaction represents an interface to psychological research, the collaboration between these fields of research would be highly beneficial' (Jungmann et al., 2022, p. 7).

5.2 Advantages and current challenges of quantitative image analysis

Nowadays, image analysis in the clinic-radiologic workflow is performed primarily qualitatively by describing specific image details such as the size or texture of pathologic findings, like necrosis in or shape of a tumor (Aerts, 2016). Although this approach is widely used due to its intuitiveness of pictorial description, there are several disadvantages, which are especially impairing in research, like non-standardized vocabulary, low inter- and intra-reader reproducibility, and often time-consuming procedures, as described in more detail in section 2.2 (Kumar et al., 2012; Limkin et al., 2017; Tixier et al., 2014; van Griethuysen et al., 2017).

The field of quantitative image analysis has emerged to address these issues. It is an image analysis technique where quantitative features of a medical image are extracted using numerical or statistical methods (Aerts, 2016; Lambin et al., 2012). This can be achieved by computing handcrafted features, as in radiomics, or by analyzing underlying characteristics of image features using machine learning algorithms and neural networks. Quantitative image analysis has developed rapidly over the last decade, starting with radiomics analyses (Kumar et al., 2012; Segal et al., 2007). Methods of accelerating the training process, like the broad usage of GPUs, the availability of software frameworks such as PyTorch, Keras, and TensorFlow as well as the growing amount of labeled medical data have lead to an increased implementation of quantitative image analysis using deep learning techniques like neural networks (Chartrand et al., 2017; Hwang, 2018).

Many studies have shown AI algorithms to achieve results on-par or even superior to human readers in a variety of image analysis tasks, including segmentation, object detection, and classification (Ardila et al., 2019; Lotter et al., 2021; McKinney et al., 2020; Tschandl et al., 2019), encouraging the integration of quantitative image analysis techniques into clinical workflows. For example, AI-based applications could be used to monitor diseases and treatment efficacy or to identify new imaging biomarkers (Afshar et al., 2019; Huang et al., 2016; Lee et al., 2017a; Prescott, 2012). Despite its advantages, quantitative image analysis still faces some significant challenges, which must be overcome to enable integration into clinical workflows. The following section will discuss both the main advantages and challenges of quantitative image analysis.

5.2.1 Benefits

Quantitative image analysis is emerging as an essential tool in various fields, including biology (Moen et al., 2019), medicine, and medical image analysis in particular (Akkus et al., 2019; Greenspan, van Ginneken, & Summers, 2016). This is due to several advantages compared to traditional image assessment techniques, which will be discussed in the following.

Enabling precise analysis

First, quantitative image analysis can measure features within an image more precisely than humans (Aerts et al., 2014), using mathematical or statistical methods designed to extract information such as the shape or size of an image region, like a tumor (Aerts, 2016).

Furthermore, quantitative image analysis can measure features that are difficult to capture using qualitative image analysis methods, such as the texture of a tumor (Lambin et al., 2012). For example, surface irregularity or complexity can be measured using fractal analysis, or quantification of intratumor heterogeneity can be described by statistic analysis of HU distribution within the region. Both surface irregularity and heterogeneity which can provide insights into the underlying pathophysiological characteristics of a tumor, like aggressiveness or tumor cellularity (Aerts, 2016; Jungmann et al., 2021; Parekh & Jacobs, 2016). This was also shown in the underlying publication P-I, where radiomic features predictive for overall survival in patients with PDAC were extracted from the primary tumor using qualitative image analysis techniques (Kaissis et al., 2020a).

The precision of quantitative image analysis methods is especially advantageous in situations where small feature changes could have significant implications on patient outcome, for example, assessing the artery infiltration of pancreatic cancer, which (among other features) determines the resectability. Especially in radiomics, there are several distinct features describing different aspects of heterogeneity, which have been shown to correlate with labels such as overall patient survival of patients with PDAC (Khalvati et al., 2019) or the EGFR mutation status of non-small cell lung cancer (Mei et al., 2018). Nevertheless, deep learning models have also achieved good or even better results on the same tasks (Wang et al., 2022; Wang et al., 2017; Zhang et al., 2020b). Despite not offering mathematical defined, concrete image features, AI has outperformed radiomics analysis in multiple applications, with the combination of both methods achieving better results than both approaches alone (Sun et al., 2020; Wang et al., 2022; Zhang et al., 2019a).

Increasing objectivity

Second, quantitative image analysis constitutes an objective way of analyzing medical images (Buckler et al., 2011), which can improve the transparency and credibility of findings (Lee et al., 2017a). Objectivity is defined as being 'uninfluenced by emotion, surmise, personal prejudice or, belief' (Halpin, 1989) and is essential for valid, reliable findings in research and many other fields (Halpin, 1989). Quantitative image analysis can promote objectivity in medical research by providing reader-independent, quantitative measurements of imaging features (Chen et al., 2020a). Unlike qualitative image analysis techniques based on visual inspection and human interpretation, quantitative image analysis applies mathematical algorithms to obtain objective and reproducible imaging feature measurements (Lambin et al., 2012; van Griethuysen

et al., 2017). This reduces the potential for measurement variations and human error, which are a potential source of bias (Tixier et al., 2014).

In addition, the use of quantitative image analysis techniques can reduce inter- and intra-observer variability, which is often observed in qualitative image analysis (Limkin et al., 2017). In contrast, quantitative image analysis algorithms are designed to provide consistent, observer-independent, and accurate results. This is especially true for AI applications, as no or few handcrafted features or manual pre-processing, like segmentation of the ROI or VOI, are needed compared to radiomics approaches (Li et al., 2021a).

While quantitative image analysis offers significant advantages in terms of objectivity and observer-independence, the publications forming the basis of this thesis did not explicitly assess this aspect by directly comparing quantitative and qualitative image analysis for the same task. However, it is widely acknowledged that quantitative image analysis enhances objectivity by generating quantitative measurements of imaging features through mathematical analyses, thereby reducing observers' influence. (Buckler et al., 2011; Chen et al., 2020b; Lambin et al., 2012).

Enhancing reproducibility and reliability

Furthermore, quantitative image analysis might contribute to the standardization of procedures based upon medical image analysis in the future (Attanasio et al., 2020), for instance, by establishing consistent image acquisition and pre-processing protocols (Buckler et al., 2011). This can improve the reproducibility and comparability of results across different studies or human readers, ensuring reliable, trustworthy findings (Guimaraes, 2021; Lee et al., 2017a). This is especially important in medical fields, where accurate and consistent measurements are critical for patient care. Given that P-I (Kaissis et al., 2020a) and P-II (Jungmann et al., 2022) were the first publications on their respective topics, it was not feasible to evaluate the reproducibility of their findings in comparison to previous studies.

Nowadays, both qualitative and quantitative image analysis methods lack consistent procedures, such as a standardized description vocabulary in qualitative analysis techniques (van Griethuysen et al., 2017), or standardized image pre-processing and mathematical definitions of imaging features in quantitative approaches (Foy et al., 2018). However, multiple efforts to identify and mitigate sources of variability and improve standardization of quantitative image analysis as well as to enhance the reproducibility of its findings are currently undertaken, which represents a promising outlook (Attanasio et al., 2020; Hagiwara et al., 2020; Morin et al., 2018; Zwanenburg et al., 2020).

The accuracy and reliability of quantitative image analysis findings can further be improved by incorporating AI applications, as they have been shown to perform on-par or even superior to human readers in diverse medical image analysis tasks (Ardila et al., 2019; Lotter et al., 2021; McKinney et al., 2020). By integrating AI applications, such as machine learning algorithms or neural networks, into the analytic process, the accuracy, reliability and reproducibility of measurements can be improved. Furthermore, this process can reduce both inter-reader and inter-study variability in quantitative image analyses.

AI applications can be trained on multicentric datasets containing thousands of images to perform a variety of tasks, such as detecting specific diseases in screening procedures, like lung (Ardila et al., 2019) or breast cancer (McKinney et al., 2020), segmentation of areas of interest, like cardiac segmentation (Chen et al., 2020b) or peripancreatic arteries (Dima et al., 2021), or outcome prediction such as overall

survival of patients with PDAC (Muhammad et al., 2018) or disease-free survival of patients with pulmonary adenocarcinoma (Kim et al., 2020). These predictions can help to identify new hypotheses (Miotto et al., 2018; Tai et al., 2019) and could eventually be integrated into clinical decision-making (Jungmann et al., 2022; McKinney et al., 2020).

Decreasing human workload

Using quantitative image analysis techniques can furthermore contribute to enhance the efficiency of image analysis tasks. Qualitative image analysis methods can be time-consuming and require significant human resources to complete, like manual segmentation and feature measurement (Bai et al., 2018). Especially manual segmentation, is known to be prone to inter-observer variability (Sapate et al., 2018), which can pose a significant challenge both in research and clinical care. By automating parts of the image analysis process, like (semi-) automated segmentation (Aerts, 2016), high-throughput feature extraction (Lambin et al., 2012) or end-to-end learning of relevant features for image classification (Le et al., 2019; Wang et al., 2017) while achieving equivalent or even superior performance compared to human readers (Chen et al., 2020b), large amounts of data can quickly and accurately be processed. Taken together, this can both save time and reduce human workload (Aerts, 2016; Lakhani et al., 2018; Le et al., 2019; Ravanbakhsh et al., 2020) as well as improve the reproducibility of results (Attanasio et al., 2020; Chen et al., 2020a).

This is especially beneficial for large or longitudinal studies (Bai et al., 2018) or high-throughput screening, like mammography screening (McKinney et al., 2020). By reducing the variability of measurements and increasing efficiency, (semi-) automated quantitative image analysis procedures can both lead to more reliable conclusions (Bai et al., 2018) and accelerated diagnostics and disease monitoring, allowing for faster and more effective treatment decisions. Thereby, quantitative image analysis can contribute to improved patient care (O'Connor et al., 2016).

The improved efficiency due to process automation can especially be achieved by using deep learning applications, where little manual labor is needed in the process (Wang et al., 2017), allowing for cost- and time-efficient processing of large amounts of data (Bai et al., 2018). This contrasts with most radiomics approaches, where manual segmentation and feature engineering is often needed (Li et al., 2021a). However, machine learning algorithms for data analysis or image segmentation can be combined with radiomics analysis, leading to reduced human effort and improved efficiency (Aerts, 2016; Kumar et al., 2012).

Both publications, particularly P-II (Jungmann et al., 2022), demonstrated the advantage of reduced human workload by efficiently analyzing large datasets. Here, the algorithm was trained on images directly without requiring manual feature extraction or segmentation. Moreover, the AI-based assistance methods tested in the reader study were automatically generated by the algorithm, eliminating the requirement for additional human intervention. Consequently, more images could be included in the study compared to P-I. In publication P-I (Kaissis et al., 2020a), the manual workload, particularly regarding the histopathologic examination of specimens, posed a significant constraint on including a larger patient cohort. Automatic segmentation of the region of interest was implemented for radiomic feature extraction to mitigate further human workload. However, the necessitated subsequent verification and potential manual correction of segmentations still proved time-consuming, impeding the analysis of a larger cohort.

Increasing cost efficiency

In addition, the efficiency of quantitative image analysis can lead to improved cost-effectiveness compared to traditional image analysis techniques (Bai et al., 2018). As explained above, qualitative image analysis involves tasks such as manual measurement or segmentation, which can be labor-intensive, prone to human error, and require specialized expertise (Bai et al., 2018; Chen et al., 2020b; McKinney et al., 2020; Zhou et al., 2021). In contrast, quantitative image analysis can provide fast and accurate measurements with less human involvement.

As discussed in the subsection above, quantitative image analysis can significantly reduce the time and effort required to analyze large datasets (Mazurowski et al., 2018), which in turn can decrease the associated cost in research. Especially deep learning approaches using automated segmentation (Comelli et al., 2020), semi-supervised learning (Ravanbakhsh et al., 2020) and end-to-end pipelines (Le et al., 2019), can contribute to this (Bai et al., 2018; Xu et al., 2014; Zhuang et al., 2021). Furthermore, standardized, objective and reproducible procedures of quantitative image analysis can facilitate multi-center scientific cooperation, enhancing efficiency and reducing redundancy and, thereof, overall cost (Begley & Ioannidis, 2015; Buckler et al., 2011).

With the availability of free or low-cost image analysis software, researchers can perform radiomics analyses without the need for expensive soft- or hardware (Avanzo et al., 2020; van Griethuysen et al., 2017; Xu et al., 2014). In contrast, AI applications usually require more data storage capacities for the necessary larger datasets and more computational power for algorithm training, like GPU cluster (Singh et al., 2020), which can be very costly compared to qualitative image analysis or radiomics studies using statistic analysis techniques (Le et al., 2019). However, by significantly reducing the human workload and working time (Comelli et al., 2020), quantitative image analysis techniques involving advanced AI applications still represent a cost-effective alternative to qualitative image analyses.

Automation of processes through quantitative image analysis techniques can also reduce costs of clinical applications (He et al., 2019a), such as breast cancer screening (McKinney et al., 2020) or disease monitoring of cancer patients (Aerts et al., 2014). As quantitative image analysis techniques are designed to be obtained from routinely acquired images, no additional procedures are required to enhance disease monitoring and enable personalized patient treatment (Aerts et al., 2014). Thus, quantitative imaging biomarkers represent a cost-effective way for disease monitoring (O'Connor et al., 2016).

Especially in screening procedures, AI applications can be deployed as a second reader to improve human performance through human-AI-collaboration while reducing time and cost for a second human reader (McKinney et al., 2020; Tschandl et al., 2020). These findings were confirmed in the underlying publication P-II (Jungmann et al., 2022), which showed enhanced diagnostic performance of radiologists with AI assistance in a mammography screening-like setting. Most participants in the study anticipated time savings and enhanced diagnostic accuracy when incorporating AI-based assistance into radiologic workflows. Moreover, all participants viewed the integration of AI as a second reader as a potentially beneficial application of AI in future radiological routine.

The overall results of this study support the adoption of AI as a second reader in radiological settings while acknowledging the need for further exploration of human-AI-interaction and additional research into potential factors such as personality traits or interpretability measures that may impact its implementation.

Identifying predictive imaging biomarkers

By analyzing large, multicentric datasets of medical images and corresponding patient outcomes, current research often uses quantitative image analysis to identify patterns and image features that are predictive of disease progression or treatment outcomes yet not directly visible to the human eye (Aerts et al., 2014; Parekh & Jacobs, 2016). Here, predictive AI applications or statistical models are developed to eventually improve diagnostic accuracy and guide clinical decision-making to improve patient care (Gillies, Kinahan, & Hricak, 2016; Sevakula et al., 2020). For example, studies have developed AI applications to identify breast cancer (McKinney et al., 2020), predict the likelihood of cancer recurrence (Zhang et al., 2019b), or assess overall patient survival (Kaissis et al., 2020a).

One main advantage of using AI applications in quantitative image analysis, like machine learning algorithms or neural networks, is the possibility to include other types of data, such as clinical and genomic data or patient history (Kaissis et al., 2020a; Liu et al., 2016). This can enhance the predictive power of quantitative image analysis models (Gillies, Kinahan, & Hricak, 2016). Furthermore, the combination of imaging data with other types of data enables the development of more comprehensive models, which may more accurately approach the complex biological patterns of the underlying disease and thus be more accurate than traditional, qualitative image analysis alone (Gillies, Kinahan, & Hricak, 2016).

The first publication, P-I (Kaissis et al., 2020a), which serves as the foundation for this thesis, exemplifies the identification of predictive quantitative imaging biomarkers using multivariate survival models. In this study, radiomics imaging features outperformed both broadly available clinical parameters and novel molecular biomarkers individually. This finding aligns with other studies that demonstrate the superior performance of imaging characteristics over clinical features, such as in assessing the EGFR mutation status in patients with lung adenocarcinomas (Liu et al., 2016) or in detecting prostate cancer (Woznicki et al., 2023).

However, the combination of imaging features with other data, including histomorphologic or clinic patient characteristics, did not enhance the model's predictive power in P-I. This contrasts with other studies where integrating data from both modalities resulted in improved model performance compared to using clinical or imaging features alone (Lv et al., 2019; Woznicki et al., 2023). Possible explanations for this discrepancy include the relatively small cohort size, known limitations of broadly available clinical parameters, and reduced representativeness of histomorphologic parameters due to partial tumor analysis rather than whole tumor analysis.

Enabling personalized medicine to improve patient care

Lastly, quantitative image analysis has the potential to play an important role in personalized medicine by providing objective and accurate measurements of biological processes within the patient, allowing for personalized diagnosis, treatment, and monitoring of disease (Gillies, Kinahan, & Hricak, 2016; Lambin et al., 2017; Limkin et al., 2017). Personalized medicine involves adapting medical treatment to a patient's individual characteristics, considering their genetics, environment, and lifestyle (Miotto et al., 2018). The National Cancer Institute defines personalized medicine in the context of oncology as follows: "In cancer, personalized medicine uses specific information about a person's tumor to help make a diagnosis, plan treatment, find out how well treatment is working, or make a prognosis (National Cancer Institute, 2020a, Dictionary of Cancer Terms).

In this context, quantitative image analysis can help identify specific biomarkers within tumors, allowing for more targeted and effective treatments (Aerts et al., 2014; Lambin et al., 2012). By analyzing radiologic or pathologic images of oncologic patients obtained in standard-of-care procedures (Kumar et al., 2012), quantitative image analysis can identify imaging features that may provide relevant information about tumor biology, such as its heterogeneity (Gillies, Kinahan, & Hricak, 2016; Lambin et al., 2017), gene expressions (Coudray et al., 2018; Mei et al., 2018) or histomorphologic subtype (Kaissis et al., 2019). This information can help clinicians to adapt treatment for each patient based on their individual characteristics (Morin et al., 2018). Quantitative image analysis can also be used to monitor patient's response to their treatment (Jin et al., 2021), allowing for personalized adjustments to treatment plans.

Furthermore, quantitative image analysis constitutes a non-invasive way of monitoring complex biological processes in real-time without the need for invasive procedures, like biopsies or surgeries (Aerts et al., 2014; Jahanshahi et al., 2023; Lambin et al., 2017; Limkin et al., 2017). This advantage was demonstrated in publication P-I (Kaissis et al., 2020a), which identified radiomic features as a predictive biomarker for overall survival in PDAC patients. Notably, the derived quantitative imaging features demonstrated superior performance compared to broadly available clinical parameters as well as novel genetic and histomorphologic markers.

This finding confirms the advantage of comprehensive, non-invasive, whole tumor analysis using quantitative image analysis techniques, particularly when considering the potential limitations of traditional sampling methods that may yield unrepresentative tissue samples during specimen workup. The non-invasive way of disease monitoring further reduces patient discomfort and improves patient safety, as potential complications associated with invasive procedures can be avoided (Shofty et al., 2020). Furthermore, a 'virtual biopsy' of the whole tumor and its surroundings eludes the issue of unrepresentative tissue samples due to tumor heterogeneity or temporal alterations (Limkin et al., 2017). In the case of indispensable tissue samples, quantitative image analysis may help identify the tumor's most significant and diagnostically conclusive regions, which poses a valuable precondition for minimal-invasive, image-guided biopsy (Martin-Gonzalez et al., 2020). This can especially be beneficial in palliative settings, as patient discomfort and potential complications can be reduced. Overall, quantitative image analysis can contribute to patient safety and has the potential to improve patient outcomes due to personalized medicine.

5.2.2 Current challenges and improvement efforts

Despite its many advantages, the quantitative image analysis still faces some severe challenges, which will have to be overcome to enable a broad usage in both research and clinical workflow and will be discussed in the following.

Lacking standardization

First, the current lacking standardization of quantitative imaging procedures constitutes a main challenge. These include image acquisition protocols and methods of data pre-processing, two aspects which can have a large influence on quantitative imaging features and thus reliability and reproducibility of findings (Meyer et al., 2019; Xue et al., 2021). As quantitative imaging findings are dependent on the quality and consistency of input data, differences in image resolution, image contrast or data methods of pre-processing can lead to

inconsistent or unreliable results, impacting the accuracy and reliability of studies (Gillies, Kinahan, & Hricak, 2016; Kumar et al., 2012; Park et al., 2020; Rizzo et al., 2018; Xue et al., 2021). Furthermore, it impedes the comparison and reproducibility of results across different datasets or studies, which constitutes a major obstacle in establishing imaging biomarkers for clinical use (Buckler et al., 2011; Lubner et al., 2017).

This holds true especially for radiomic features. Here, the manual or semi-automated segmentation of the ROI or VOI needed is known to be subject to high inter-reader variability, and thus influences quantitative features extracted from the segmentations (Avanzo et al., 2020; Kumar et al., 2012). Furthermore, inconsistent mathematical feature definitions and the still prevailing use of in-house software without public access further limit the reproducibility and fidelity of findings (Bagher-Ebadian & Chetty, 2021; van Griethuysen et al., 2017). For example, the results of three studies investigating radiomics features derived from FDG-PET/CT (positron emission tomography with computed tomography) for distinguishing pulmonary adenocarcinoma and squamous cell carcinoma cannot easily be compared, as they all used different softwares with distinct radiomic feature definitions and were inconsistent in extracting imaging features from either PET or CT data or both (Hyun et al., 2019; Ren et al., 2020; Tomori et al., 2020).

Several efforts to address the lacking standardization in quantitative imaging analysis have been made over the last years (Avanzo et al., 2020; Buckler et al., 2011; Limkin et al., 2017). For instance, `PyRadiomics` has been developed as an open-source, Python-based platform for radiomics analyses (van Griethuysen et al., 2017), which provides detailed documentation of all features extracted from the images. By this, `PyRadiomics` adheres to the guidelines set forward by the Image Biomarker Standardization Initiative (IBSI) to enhance the standardization of quantitative image analysis. Another example of radiomics software systems adhering to the IBSI guidelines is the MATLAB-based framework `ROdiomiX` (Bagher-Ebadian & Chetty, 2021).

The IBSI was formed by a collaboration of 25 research teams with the aim to standardize radiomics-based quantitative image analysis. For this, they provide a set of well-defined and validated radiomics features as well as guidelines for image processing, development, and performance evaluation of algorithms (Zwanenburg et al., 2020). Multiple other collaboratives also aim to enhance reproducibility and accuracy of quantitative image analysis by providing guidelines to standardize image acquisition and pre-processing (Kennedy & Taouli, 2021), like the Quantitative Imaging Biomarkers Alliance (QIBA) formed by the Radiological Society of North America (Buckler et al., 2011; Guimaraes, 2021), the J-QIBA established by the Japan Radiological Society (Fukukita et al., 2014) or the European Imaging Biomarkers Alliance (EIBALL) of the European Society of Radiology (Alberich-Bayarri et al., 2020).

In the first publication, P-I (Kaissis et al., 2020a), non-standardized procedures and the challenge to ensure the reproducibility of findings in future studies had to be addressed. An in-house developed algorithm for automated segmentation was employed to mitigate the dependency of radiomic features on manual segmentation. Nevertheless, the subsequent validation and potential manual corrections of the generated segmentations introduced a source of observer dependency. Radiomics features were extracted using `PyRadiomics` (van Griethuysen et al., 2017) to identify well-defined quantitative imaging features and ensure their reproducibility. However, due to the relatively small cohort size compared to the number of imaging features extracted, PCA was applied to reduce the number of quantitative imaging features. Consequently, the resulting radiomics signature identified as a biomarker in P-I poses a bigger challenge for reimplementation compared to the identification of individual radiomic features, which can be referenced in the `PyRadiomics` documentation (pyradiomics community, 2016). Given that the second publication P-II

(Jungmann et al., 2022) concentrated on analyzing human-AI-collaboration and evaluating the perception of different types of AI-based assistance, the lack of standardized procedures that could hinder reproducibility did not present a significant challenge for the project.

Risk of overfitting

Second, quantitative image analysis studies are prone to overfitting due to the large number of imaging features compared to the relatively small cohort size (Kumar et al., 2012; Lambin et al., 2017). If overfitting occurs, a model is fitting noise in the training data and memorizing the data instead of capturing the underlying patterns of information (Chartrand et al., 2017; Jabbar & Khan, 2015). As a result, the model performs poorly on former unseen data due to lacking generalizability while performing well on the training set (Ying, 2019). This leads to imprecise and unreliable results and impedes the broad clinical application of such models.

Overfitting especially occurs in machine learning models and other AI applications, which have a large number of trainable parameters and are capable of learning complex, non-linear relationships between the input data and their corresponding labels (Zhang, Zhang, & Jiang, 2019). The main reasons for overfitting include a dataset too small to capture all underlying imaging patterns, too noisy or unrepresentative for the task, or a model which is too complex for the given data (Hawkins, 2003; Ying, 2019). A model with too many parameters compared to the number of features is more likely to overfit the data, as it can learn complex functions which exactly fit the training data instead of approximating the underlying patterns within the data (Hawkins, 2003; Ying, 2019).

Several techniques can be applied to mitigate overfitting, including network regularization, dropout, or early stopping, the detailed explanation of which lies beyond the scope of this thesis (Jabbar & Khan, 2015; Shorten & Khoshgoftaar, 2019; Srivastava et al., 2014). Another approach in quantitative image analysis studies comprises the pre-selection of imaging features to reduce the number of features compared to the size of the dataset (Afshar et al., 2019; Li et al., 2021a). This can be done in an unsupervised manner, for example by using PCA or cluster analysis (Parekh & Jacobs, 2016; Rizzo et al., 2018). Furthermore, the generalizability of a model can be evaluated using cross-validation. For this, the available data are resampled into multiple subsets, on which the algorithm is trained and tested on in order to estimate the performance of the final model trained on the whole training dataset (Berrar, 2019).

The lack of training data poses a major challenge for model training, especially in the medical field where data contain sensitive patient information and are thus difficult to obtain or to share across different sites (Le et al., 2019). A common method used to extend the data available for training of AI applications is transfer learning, where the model is trained to learn basic features, such as edge detection, on a source set of images of a related domain with the transfer learning usually being more successful the more similar the two domains are (Zhou et al., 2021; Zhuang et al., 2021). A common dataset used for transfer learning is ImageNet (Russakovsky et al., 2015), which contains millions of labeled images containing non-medical objects, or open-source medical databases such as The Cancer Imaging Archive (2019).

Another prevailing technique to enrich the training data is data augmentation. Here, additional data are generated by producing slightly different versions of the existing training data, for example, by shifting, rescaling, or rotating the image or by injecting noise (Shorten & Khoshgoftaar, 2019). However, data

augmentation can not compensate for large, diverse, and representative datasets in the first place, as it provides data that are still biased towards the spatial characteristics of the original images (Shorten & Khoshgoftaar, 2019). As a result, limited data availability still represents one of the major challenges of quantitative image analysis and will be further discussed later in this section.

Both studies in the underlying publications, P-I and P-II, were susceptible to the risk of overfitting data, which could compromise their generalizability and reliability. In P-I (Kaissis et al., 2020a), PCA was employed to mitigate the risk of overfitting caused by a large number of radiomic features relative to the relatively small cohort size, as previously described. Additionally, the generalizability of the survival model was evaluated using five-fold cross-validation, as no external test cohort was available due to the extensive histopathologic specimen workup.

In the second publication P-II (Jungmann et al., 2022), a ResNet 50 pre-trained on ImageNet (Russakovsky et al., 2015) was adapted to classify BI-RADS 4 lesions into benign or malignant findings. The intention behind transfer learning rather than training the ResNet 50 from scratch on our data was for the model to have already learned basic imaging features such as shapes or edges from the ImageNet data, and thereby requiring less training data in our study. Different data augmentation techniques, such as rotation, flipping, shifting, and zooming, were applied to extend the training data and to enhance its diversity. Moreover, early stopping was implemented during training to prevent overfitting. An independent test set of in-house mammography images was collected to evaluate the generalization performance of the ResNet 50 and further used to create the different types of AI-based assistance evaluated in the reader study.

Low reproducibility and generalizability across different data sets

Another challenge in quantitative image analysis is the often lacking generalizability of models and reproducibility of findings across different datasets (Chen et al., 2020b). This can partly be attributed to differences in patient populations or different scanners and acquisition protocols between sites with consequently varying imaging features, such as spatial resolution, signal-to-noise ratio, or image contrast (Castiglioni et al., 2021; Jahanshahi et al., 2023; Park et al., 2020). Many quantitative imaging features depend on these imaging characteristics, resulting in significant feature variance between studies (Jahanshahi et al., 2023; Xue et al., 2021). As a result, models trained on one dataset may not generalize to another dataset acquired using a different imaging protocol, or significant radiomics features discovered in one study might not correlate with the output label in another study. For example, Berger et al. (2023) tried to reproduce the findings of a radiomics study to predict late irradiation-induced toxicity in patients with head and neck cancer (van Dijk et al., 2017) and found none of the imaging features to be significant in their cohort, thus demonstrating lacking generalizability.

Furthermore, many studies lack an external validation cohort or prospective validation of their findings (Berger et al., 2023). This limitation is also evident in the first publication P-I (Kaissis et al., 2020a), where the extensive histopathologic workup of tumor specimens prevented the collection of an external validation cohort. While cross-validation was performed to estimate the generalizability of the survival model, the lacking external validation remains a significant limitation of the study. In contrast, the neural network deployed in publication P-II (Jungmann et al., 2022) was trained using the publicly available CBIS-DDSM dataset and tested on an independent test set composed of in-house mammograms. The model successfully demonstrated the ability to generalize its findings across these different datasets, achieving promising

results on the test set in the reader study. However, it is important to note that no prospective validation of the model's performance was performed. This decision was deliberate, as the focus of the study was to evaluate the effectiveness and perception of AI-based assistance in radiologic settings rather than to develop a classification algorithm explicitly intended for clinical use.

Currently, many efforts are being made to mitigate the issue of lacking generalizability and reproducibility in quantitative image analysis research, such as the development of standardized imaging protocols to ensure consistent data across different sites or methods of quality control for quantitative imaging studies, as proposed by the IBSI or QIBA (Buckler et al., 2011; Zwanenburg et al., 2020). Other approaches to improve the generalizability of models comprise the incorporation of prior knowledge about the underlying pathology into models, domain adaptation and transfer learning techniques (Zhang et al., 2020a; Zhou et al., 2021).

Nevertheless, the poor generalizability of most quantitative image analysis models remains a significant challenge that impedes clinical usage of these models (Meyer et al., 2019; Wong, Fortino, & Abbott, 2020), as generalizability and consequent reliability and reproducibility of findings are essential for application in patient care.

Difficult integration into clinical workflows

Clinical utilization of quantitative image analysis techniques is further impeded by difficulties of integrating such applications into existing clinical workflows and structures (Jungmann et al., 2022; Maniatopoulos et al., 2015). This includes an interface to existing medical infrastructure, like a PACS (Shaikh et al., 2018; Zhovannik et al., 2021), or user-friendly interfaces (Pasini et al., 2022) to enable uncomplicated use in daily routine. Until today, most quantitative image analysis algorithms have not been tested or implemented in a realistic clinical setting, and integration into existing infrastructure and workflows remains a challenge (He et al., 2019a; Huang et al., 2022). Furthermore, data storage and privacy issues must be tackled to enable broad usage of developed algorithms (He et al., 2019a; Zhou et al., 2021).

The difficulties of integration into patient care are especially relevant for radiomics approaches, as few steps are automated, and many aspects require manual interaction, such as manual segmentation or explicit feature engineering (Gillies, Kinahan, & Hricak, 2016). In contrast, AI applications are more likely to achieve integration into existing radiological infrastructure, as most procedures can be automated, and significantly less human interaction is needed in image analysis (Rotalinti et al., 2023; Soin et al., 2022).

Besides, AI systems can experience a phenomenon known as 'drift', which refers to a gradual degradation in performance over time, potentially affecting the accuracy and reliability of these models. Drift can occur for various reasons, including shifts in input data distribution or changes in the operational environment (Rotalinti et al., 2023; Soin et al., 2022). To mitigate drift and maintain the model's performance, re-training becomes necessary. However, the dynamic nature of continuously re-trained AI applications complicates the establishment of uniform regulatory standards (Food and Drug Administration (United States), 2019; Harvey & Gowda, 2020). Unlike traditional medical devices with fixed designs, re-trained AI systems constantly evolve through exposure to new data, potentially leading to unforeseen behaviors, performance variations, or non-transparent models (Soin et al., 2022). Regulatory agencies must establish robust frameworks to address these challenges and to ensure that AI applications remain effective and safe in real-world medical applications (Harvey & Gowda, 2020).

Although not the primary focus of the underlying publications, the challenges associated with integrating quantitative image analysis techniques into the clinical routine can be comprehended based on both publications, P-I and P-II. In the first publication, P-I (Kaissis et al., 2020a), a radiomics approach, combined with extensive histopathologic tumor analysis, was employed to identify imaging biomarkers predictive of overall survival in PDAC patients. However, even without the histopathologic workup, implementing radiomic analysis, including exporting images to software disconnected from the in-house PACS system and performing segmentation of the region of interest, poses practical obstacles that hinder seamless integration into daily clinical routines.

The underlying study of publication P-II (Jungmann et al., 2022) aimed to simulate the integration of various AI-based assistance methods in mammography screening, for example, as a second reader. However, the study conditions were not entirely realistic. The images had to be exported, resized, and cropped to be compatible with the ResNet 50 to create the different methods of AI-based assistance. Additionally, the absence of a PACS interface prevented the display of original image data and clinical information, thereby limiting the ability to replicate an actual screening process. These challenges highlight the need for further development and optimization of integration strategies to ensure the practical and efficient utilization of different quantitative image analysis techniques in routine radiologic workflows.

Lacking interpretability and explainability

Besides the technical challenges described above, human factors can hinder the broad clinical application of quantitative image analysis models. These obstacles include information overload provided by CAD applications (Kohli & Jha, 2018), aversion against algorithms (Dietvorst, Simmons, & Massey, 2015), lack of trust (Ribeiro, Singh, & Guestrin, 2016) and face-to-face validation as applied with human colleagues (Maddox, Rumsfeld, & Payne, 2019).

Many of these issues are linked to a lack of interpretability and non-intuitiveness of using quantitative image analysis techniques (Cai et al., 2019). In this context, interpretability refers to an inherent property of models, which enables humans to understand the models and their process of decision-making (Lipton, 2018). Explainable AI is a research field which aims to reduce the 'black-box' nature of AI models using different techniques to explain how a model works or it reached a specific conclusion (Barredo Arrieta et al., 2020). In the second publication, P-II, different types of explainability techniques were employed and their perception was evaluated (Jungmann et al., 2022).

In contrast to more traditional, qualitative analysis methods, quantitative image analysis generates numerical results rather than descriptive evaluations. These may be more challenging to interpret by physicians who are unfamiliar with the underlying mathematic analyses (Morin et al., 2018). Especially for radiologists, trained in visual inspection of images, the often lacking possibilities of visual quality control can reduce trust in quantitative image analysis algorithms and lead to reluctance to rely on these (Jungmann et al., 2022).

Due to their complex structure and lacking interpretability, which often makes it impossible to reconstruct how or why a model made a distinct prediction, clinicians frequently consider AI applications as 'black boxes' (Wickramasinghe et al., 2020). To overcome the insufficient interpretability and transparency of AI applications, which is often a source of lacking trust in these algorithms and impedes their broad usage in critical infrastructures, the fields of explainable and trustworthy AI have emerged (Castiglioni et al.,

2021; Samek & Müller, 2019; Thiebes, Lins, & Sunyaev, 2020). Trustworthy machine learning is based on principles such as 'fairness, robustness, explainability, accountability, verifiability, transparency and sustainability' (Wickramasinghe et al., 2020, p. 135) and, like explainable AI, aims to provide explanations on the internal processes of algorithms in order to generate trust and enable their utilization in safety-critical applications, such as patient care (Gunning et al., 2019; Samek & Müller, 2019).

Techniques to elevate the transparency and trustworthiness of AI applications include uncertainty quantification of predictions or visualization of image regions important for a model's output like LIME or Grad-CAM (Liu et al., 2021; Ribeiro, Singh, & Guestrin, 2016; Selvaraju et al., 2017). Especially in radiology, visual interpretability methods have the potential to promote trust in AI algorithms, as they provide a method of quality control that is similar to traditional qualitative image analysis (Jungmann et al., 2022).

The second publication P-II (Jungmann et al., 2022), which partially forms the basis of this thesis, aimed to investigate the perception of AI-based assistance in radiologic settings. The study findings did not replicate an underlying negative attitude towards algorithmic implementation in existing workflows as described in literature (Dietvorst, Simmons, & Massey, 2015). Radiologists who participated in the study did not express aversion towards the algorithm, nor did they anticipate negative impacts of future AI integration, such as a decline in diagnostic accuracy or complete replacement of human involvement. On the contrary, they perceived the impact of integrating AI into clinical workflows as primarily positive.

These findings align with current studies that indicate a predominantly positive attitude among physicians and medical students towards integrating AI in medical settings (Mousavi Baigi et al., 2023; Sarwar et al., 2019). The evolving attitude towards AI in the medical field, with recent studies reflecting a predominantly positive outlook compared to earlier studies reporting negative sentiments and anxieties (Chockley & Emanuel, 2016; Dietvorst, Simmons, & Massey, 2015), indicates a notable shift over the past years. Moreover, most participants in the underlying study of P-II felt that they were collaborating with the AI system rather than competing against it. Consequently, all participating radiologists considered integrating AI-based programs as a second reader a potential future application.

In concordance with previous literature, providing interpretability measures, such as the heatmap and the neural network's prediction along with its certainty, enhanced the trust of radiologists in the AI-based assistance and influenced their willingness to change their original lesion classification. While the heatmap, generated by calculating an occlusion-based sensitivity map of the image, did not independently improve the diagnostic accuracy of radiologists, it was valued by most participants as a valuable tool for quality control of the algorithm. It facilitated a comparison between their own visual findings and the algorithm's regional importance, contributing to increased confidence. Notably, experienced radiologists at the attending level particularly appreciated this aspect, as it provided a means for plausibility and quality control, leading to higher confidence in the algorithm's predictions.

The results of publication P-II emphasize the potential benefits of integrating AI-based assistance in radiologic workflows, particularly when interpretability measures are provided, fostering trust, collaboration, and quality assurance. In contrast, publication P-I (Kaissis et al., 2020a) did not assess the collaboration of and interaction between humans and AI.

Complex human-AI collaboration

Consequently, successfully translating quantitative image analysis procedures into clinical practice depends on a sufficient interaction between radiologists and the CAD system (Khairat et al., 2018; Musen et al., 2021; Yang, Steinfeld, & Zimmerman, 2019). However, while many publications have shown on-par or even superior performance of algorithms in various medical imaging classification tasks (Ardila et al., 2019; Ghafoorian et al., 2017; Gulshan et al., 2016; Haenssle et al., 2018; Lakhani & Sundaram, 2017), the perception of and interaction with those algorithms still awaits deeper investigation (Jungmann et al., 2022).

Successful human-AI-interaction is fundamental for future integration of AI systems into existing clinical workflows and thereof for possible advances in diagnosis and treatment (Jungmann et al., 2022). Recent research has shown that the collaboration of humans with AI applications leads to improved performance compared to either of them alone (Hekler et al., 2019; Tschandl et al., 2020). In the cited studies, the algorithms outperformed the participants. However, combining the algorithm with human expertise leads to an additional increase in diagnostic accuracy. This observation was replicated in Publication P-II (Jungmann et al., 2022), on which this thesis is partially based. In this study, the inclusion of AI-based assistance resulted in a significant improvement in radiologists' performance compared to their initial classifications without any additional support and the algorithm's performance alone.

Nevertheless, most research today still focuses on developing more efficient or higher-performing algorithms which barely interact with human users instead of assessing the combined performance of humans with AI systems along with their possibilities and challenges (Jungmann et al., 2022). This can partly be explained by the predominant belief of humans proposing the biggest security issue in digital applications (Schneier, 2015) and constituting unpredictable behavior (Holzinger, 2016). However, the results of multiple studies show that including human-AI-collaboration into AI systems could be greatly beneficial, especially in safety-critical medical applications (Budd, Robinson, & Kainz, 2021; Holzinger, 2016; Tschandl et al., 2020). Here, transparency of AI-based decisions (Miller, 2019) is one crucial component for a successful human-AI-collaboration (He et al., 2019a; Molnar, 2020).

The second publication P-II (Jungmann et al., 2021), on which this work is partially based, confirms this thesis. In this study, the implementation of different methods to provide transparency in the classification process of the ResNet 50, such as indicating the certainty of its predictions or displaying heatmaps highlighting the most important image regions for classification, resulted in an increased level of trust in the algorithm. Consequently, the participating radiologists were more willing to change their initial assessments of mammogram images in favor of the algorithm's prediction, particularly when presented with a plausible heatmap alongside a high prediction certainty. The subsequent human-AI-collaboration significantly increased performance among radiologists. However, it is crucial to acknowledge that this aspect poses a significant risk, as a faulty AI system could thus potentially impair human performance (Tschandl et al., 2020). Since P-II did not evaluate other algorithms with varying or faulty performance levels, this risk could neither be confirmed nor ruled out in the specific study. Further investigation into this topic will be crucial before considering the integration of different AI algorithms into clinical care.

Recently, chatbots have had considerable success regarding natural interaction with humans (Jo et al., 2023; Liu et al., 2023). Integrating the large language models behind chatbots into quantitative image analysis procedures could be highly beneficial for a successful human-AI-collaboration for several reasons. First, using natural language enables more intuitive interaction with such models and enhances model

interpretability (Menon & Vondrick, 2022) in a way that might be more easily understandable than other explainability methods of AI models, such as heatmaps or certainty displays alone. Furthermore, large language models allow users to ask specific questions to better understand the reasoning behind a model's decision (Javaid, Haleem, & Singh, 2023), which in turn might lead to better recognition of cases in which the AI provides faulty results. Besides the advantage of enhancing human-AI-interaction, large language models could be integrated into medical workflows in several ways (Javaid, Haleem, & Singh, 2023). For more information on large language models and their possible applications on medical data, see section 5.3.

Human-in-the-loop promotes the combination of human and machine intelligence in order to optimize AI applications in various settings (Monarch, 2021), thus exploring the interaction and cooperation between humans and AI algorithms (Cai et al., 2019). Here, human input, as well as feedback loops and quality control, is used to improve the performance and certainty of machine learning algorithms (Monarch, 2021), be it for labeling-processes (Budd, Robinson, & Kainz, 2021) or for AI-based assistance in human tasks, like organ segmentation (Ravanbakhsh et al., 2020; Xue et al., 2021) or classifying findings in mammography screening (McKinney et al., 2020). As mentioned above, considering human-in-the-loop approaches offers several ways of addressing the challenges of transferring AI application into clinical use. However, it can be more time-consuming and costly to develop and deploy compared to fully autonomous AI applications, due to the necessary involvement of trained professionals to review and validate the results. Moreover, human-in-the-loop systems are vulnerable to user-dependant performance variations, as they depend on the expertise of the human involved in the loop. In contrast to AI systems without a human in the loop, this can increase the variability and subjectivity of models, potentially reducing the reliability of their findings. Besides these challenges, other aspects of human-AI-interaction have to be considered before implementation of such algorithms in patient care, such as possible drawbacks of faulty AI (Tschandl et al., 2020) or the tendency of humans to align their opinion with the decision of an algorithm (Le et al., 2019).

Human-in-the-loop approaches align with the Artificial Intelligence Act of the European Union, which was proposed in 2021 to ensure a safe and ethical usage of AI and to provide a legal framework for trustworthy AI implementations (European Commission, 2021). As of May 2023, it is currently under review by the European Parliament and Council and is expected to be adopted soon. The regulations pose a risk-based approach, where AI technologies are divided into four risk categories, each with its own requirements and regulations. High-risk applications, such as AI systems used in critical infrastructure, healthcare, or law enforcement, would be subject to strict requirements for documentation of performance and limitations, insurance of transparency and interpretability, data quality, as well as regular testing and contingency plans in case of system failure. Furthermore, high-risk applications would have to provide a human in the loop to monitor the AI system and to ensure the possibility of intervention, if necessary (Ebers, 2021).

Need for extensive data and privacy concerns

As previously stated, obtaining large, multicentric datasets constitutes one of the main challenges in quantitative image analysis (Mazurowski et al., 2018; Wong, Fortino, & Abbott, 2020). While the potential benefits of this quantitative imaging are significant, as discussed in section section 5.2.1, the accuracy and reliability of findings heavily depend on the availability of high-quality data (Castiglioni et al., 2021). The availability of a large dataset is essential for developing and validating robust and reliable algorithms as

a high number of cases can help to mitigate issues related to model overfitting (Le et al., 2019) and thus reduce the likelihood of false-positive results (Traverso et al., 2018).

Additionally, data from multiple sites are required to reduce the risk of bias and to ensure the generalizability of findings (Qiao et al., 2020), as these data increase the representativeness of a study population and more likely capture the heterogeneity of a patient population or the variability in imaging acquisition (Kumar et al., 2012; Le et al., 2019).

All these aspects allow for a more comprehensive evaluation of the performance and generalizability of a model, which is why multicenter, prospective studies will be crucial for ensuring a broad clinical implementation of quantitative image algorithms (Avanzo et al., 2020; Martin-Gonzalez et al., 2020). Especially deep learning-based quantitative image approaches require much training data, significantly more than is needed for radiomics approaches (Le et al., 2019). This can be explained by the complexity of the algorithm as well as more noisy training data, as the input is usually a full image without manual pre-processing, like segmentation of the ROI, which thus contains additional information besides the target lesion (Li et al., 2021a).

Both publications, which form the basis of this thesis, faced challenges in acquiring large, multicenter, and high-quality datasets. Consistent with previous research, the neural network utilized in P-II (Jungmann et al., 2022) necessitated a larger training dataset compared to the Cox Proportional Hazard model employed in the analysis of extracted radiomic features in study P-I (Kaissis et al., 2020a). To obtain sufficient training data, the open-source CBIS-DDSM dataset (Lee et al., 2017b) was used as a training cohort, and augmentation techniques were implemented. Additionally, transfer learning was applied using a ResNet 50 model pre-trained on ImageNet (Russakovsky et al., 2015). An independent dataset of in-house collected mammograms was used as the test set. However, including more data from multiple institutions would have greatly benefited the study. The same applies to publication P-I (Kaissis et al., 2020a), where the extensive histopathologic analysis required to obtain relevant features for the study hindered the collection of an independent external dataset or an expansion of the existing in-house data.

As demonstrated in the underlying publications, it can be difficult to collect a sufficient number of images from a single center, and it is challenging to obtain data from multiple centers due to several reasons, including varying data acquisition (Rizzo et al., 2018) and legal concerns in sharing sensitive information across different sites (Attanasio et al., 2020). A lacking standardization of image acquisition and pre-processing, as explained before, may lead to major differences in image quality, impacting the reliability of results (Xue et al., 2021) and thus pose a challenge to integrate data from multiple sites (Rizzo et al., 2018). Issues of collecting multi-institutional data are further aggravated by stringent regulation of patient data and requirements for its protection, for instance, as posed by the United States Health Insurance Portability and Accountability Act (HIPAA) of the Department of Health and Human Services (HHS) (2002) or the General Data Protection Regulation (GDPR) (2016). Although of utmost importance, data privacy and confidentiality regulations can make it challenging to share data across institutions (He et al., 2019a). This process can be further complicated by technical challenges related to data management, storage, and integration into existing infrastructure.

Despite these challenges, several large, multicenter and open-source databases for quantitative image analysis have already been collected, such as The Cancer Imaging Archive (TCIA) (The Cancer Imaging Archive, 2019), RadImageNet (Mei et al., 2022) or the UK Biobank (Sudlow et al., 2015), which provide

oncologic imaging data, general radiologic images, and genetic and health information of patients in the United Kingdom, respectively.

In the second publication P-II (Jungmann et al., 2022), the CBIS-DDSM dataset (Lee et al., 2017b) from TCIA was used for network training, enabling the usage of a large, independent dataset instead of relying solely on data from a single institution. In contrast, none of the accessible open-source datasets were suitable for usage in publication P-I (Kaissis et al., 2020a) due to the absence of histopathologic labels, which were essential for the study's objectives.

Despite the growing collection of medical databases, no medical imaging dataset yet reaches the dimension of publicly available, non-medical imaging data commonly used for AI transfer learning, like ImageNet (Russakovsky et al., 2015). As the success of transfer learning is dependant on the similarity of the domain used for transfer learning and the target domain, large open-source data containing medical images would be preferable to using non-medical datasets like ImageNet, which contains 2D, color images (Chartrand et al., 2017; Zhou et al., 2021; Zhuang et al., 2021).

Further attempts in tackling problems in sharing sensitive patient data include strategies and frameworks for secure and private sharing of patient records, like the Radiological Society of North America (RSNA) Image Sharing Network (Langer et al., 2015) for securely sharing imaging records between different sites. Other concepts aim to ensure personal data privacy by developing secure AI systems rather than gathering de-identified patient data to enable secure utilization of medical data for research and clinical routine. They are encompassed in secure and privacy-preserving AI, which nowadays poses a rapidly growing field of research (Kaissis et al., 2020b). The main methods include homomorphic encryption, secure multi-party computation, federated machine learning, and differential privacy, the last two of which will be explained shortly in the following (Kaissis et al., 2020b).

Federated learning is a machine learning technique, where multiple sites collaborate on AI development without sharing the individual data but rather the algorithm (Li et al., 2020). It relies on remote execution, where copies of a model are trained locally, and the updated model weights are then transferred to a common site (Zhang et al., 2021). Nowadays, federated learning is among the most commonly used privacy-preservation techniques in medical research and has considerable capability to enable future privacy-preserving cross-institutional research (Rieke et al., 2020). Although this approach poses a solution to data sovereignty and governance issues, federated learning itself does not guarantee data privacy and security unless combined with other privacy-preserving methods, like differential privacy (Kaissis et al., 2021).

In differential privacy techniques, recognizable information of a single patient is reduced while maintaining the global statistical distribution within the data (Ziller et al., 2021). This is achieved using different methods of data perturbation, like random shuffling, injection of noise to the data (Kaissis et al., 2020b) or by implementing differential private algorithms during training, like differentially private stochastic gradient descent (Rajkumar & Agarwal, 2012). Differential privacy techniques are especially beneficial to protect against re-identification attacks, as it is more difficult to determine whether a specific data point was used for obtaining the result, and individual data are more challenging to reconstruct (Kaissis et al., 2021).

Until a few years ago, only a small number of algorithms satisfying differential privacy had been deployed in quantitative image analysis due to more complicated implementation and often reduced performance compared to non-differentiable private algorithms (Kaissis et al., 2020b). However, first studies have shown

promising results in demonstrating the feasibility of using privacy-preserving AI systems in the medical field (Dou et al., 2021; Kaissis et al., 2021; Li et al., 2019). Due to the beneficial properties of differential privacy and ongoing research in this field, an increasing number of applications have very recently been proposed, some nearly matching the accuracy of non-private models already (Berrada et al., 2023). This could indicate that privacy-preserving model training might become a gold standard for future medical AI models.

5.3 Contemporary research topics and future trends of quantitative image analysis

As discussed in section section 5.2.2, various efforts are underway to address the existing challenges in quantitative image analysis. Most notable are advances concerning privacy guarantees of sensitive patient data, like differential privacy, and methods to enable more effective collaboration between algorithms and humans, especially human-in-the-loop approaches and explainable AI. For a more detailed explanation of these fields, see the corresponding paragraphs under section 5.2.2.

In addition to these focused research areas, other domains of study, especially regarding other network architectures, are currently emerging to enhance the field of quantitative medical image analysis. The subsequent section provides a concise overview of the present trends and research subjects that have the potential to significantly impact the future applications and capabilities of quantitative analysis in medical imaging.

5.3.1 Deep generative models

Deep generative models are currently being investigated for their potential to generate realistic and high-quality medical images (Celard et al., 2022). They are a class of AI models to generate new data, such as images or text, with similar characteristics as their training data. Two popular types of deep generative models are variational autoencoders (VAEs) and GANs (Celard et al., 2022). Furthermore, diffusion models, which can be based both on VAEs or GANs, have recently gained importance in image generation tasks (Kazerouni et al., 2022).

VAEs are generative models that learn a probabilistic latent space representation of their input data (Kingma & Welling, 2019). They consist of an encoder and a decoder network. The encoder reduces the dimensionality of latent space representation of the training data, from which the decoder tries to reconstruct the original data (Girin et al., 2021). VAEs are trained to maximize the likelihood of the decoder generating the original input data while enabling the latent space to follow a predefined probability distribution, typically a Gaussian distribution (Doersch, 2016). After training, this latent space can then be sampled to generate new data points resembling the distribution in training data. An advantageous property of VAEs is the ability to interpolate the latent space, allowing smooth transitions and generation of images with gradual changes in characteristics. This can be used to generate synthetic images with distinct variations, such as changing the size, shape, or appearance of structures in images (Cetin et al., 2023).

GANs were first proposed by Goodfellow et al. (2014) and, analogous to VAEs consist of two neural networks, called the generator and a discriminator. The aim of generator network is to generate realistic images from

a latent space or random noise input, while the discriminator network tries to distinguish between real and synthetic samples (Yi, Walia, & Babyn, 2019). The two networks are trained in an adversarial manner, which means that the generator aims to generate synthetic images indistinguishable from the corresponding original data, while the discriminator aims to achieve perfect accuracy in identifying both synthetic and real samples (Goodfellow et al., 2014). This enables GANs to generate highly realistic and high-quality samples, which is why they are more commonly used than VAEs in medical image applications nowadays (Celard et al., 2022). Thus, the following paragraphs will focus on GANs as representatives of deep generative models rather than VAEs.

Moreover, diffusion models have emerged as a powerful class of deep generative models in recent years. They are typically engineered as probabilistic networks and have shown good performance in capturing and propagating complex data distributions (Kazerouni et al., 2022; Kobzyev, Prince, & Brubaker, 2021). One of the main architectures in this domain is the Normalizing Flow model, which can transform a given, simple distribution, like a Gaussian or Laplacian distribution, into a more complex, data-driven distribution through a series of invertible transformations (Kobzyev, Prince, & Brubaker, 2021). Examples of Normalizing Flow models in the medical domain comprise RealNVP (Dinh, Sohl-Dickstein, & Bengio, 2016) and Glow (Kingma & Dhariwal, 2018), both of which have been used for image restoration and synthetic image generation. The PixelCNN family is another well-known class of diffusion models. These models are based on convolutional neural networks and can be used for various tasks, including synthetic medical image generation (van den Oord et al., 2016; van den Oord, Kalchbrenner, & Kavukcuoglu, 2016).

Besides, latent diffusion models, which are primarily based on VAEs, can operate directly in the autoencoder's latent space rather than in the pixel space of images (Rombach et al., 2022). The most notable representatives of these kinds of models are DALL-E (Ramesh et al., 2021, 2022) or Stable Diffusion (Rombach et al., 2022), which are mainly used for text-to-image generation (Gozalo-Brizuela & Garrido-Merchan, 2023).

Data augmentation

Deep generative models have shown promising results in various domains, including image synthesis (Han et al., 2018; Ledig et al., 2017), text generation (de Rosa & Papa, 2021; Shahriar, 2022) or creating images based upon textual descriptions (Frolov et al., 2021). In the context of quantitative medical image analysis, deep generative models can be used to augment limited datasets by creating diverse, high-quality medical images closely resembling actual patient data (Frid-Adar et al., 2018; Qin et al., 2020; Waheed et al., 2020). By increasing the diversity and size of the training dataset through additional synthetic samples, deep generative models can facilitate the development of quantitative image analysis algorithms with improved robustness and generalizability (Waheed et al., 2020). The implementation of deep generative model can be particularly valuable to augment limited or unbalanced data, such as when working with rare diseases (Kazemina et al., 2020; Li, 2022).

Furthermore, GANs can contribute to preserving patient privacy by generating synthetic images that do not contain identifiable patient information (Yi, Walia, & Babyn, 2019). This might enable the sharing and analyzing medical imaging data without breaching patient confidentiality in the future. Training on purely synthetic data or augmenting databases with it could solve privacy problems when working with sensitive patient data. However, most models today are trained without formal privacy techniques, such as differential

privacy. This procedure has been shown to lead to unsafe models, enabling the reconstruction of training data and consequent identification of individuals (Carlini et al., 2023). Meanwhile, privacy-preserving synthetic data generation for tabular data can already be achieved with good privacy guarantees (Ho et al., 2021; Yoon, Drumright, & van der Schaar, 2020). As this is not yet the case for large images, such as high-resolution radiographs, improving this task is an active research area (Ghalebikesabi et al., 2023; Wu, Guo, & Chaudhuri, 2023).

Image denoising and reconstruction

Another challenge faced by quantitative image analysis applications is that medical images acquired from various modalities, such as MRI, CT, or ultrasound, are known to be susceptible to noise and artifacts (Yi, Walia, & Babyn, 2019). This can significantly impact the accuracy and reliability of subsequent quantitative analysis, affecting both medical research and clinical practice.

GANs offer a solution as they can be employed for image denoising and reconstruction tasks. For this, they are trained mostly on paired datasets of noisy or degraded medical images and their corresponding high-quality version. The aim is for GANs to learn to generate images with reduced noise while preserving the essential diagnostic features (Yi & Babyn, 2018), for example by enhancing image quality, increasing spatial resolution, or removing artifacts. Thereby, GANs can aid to improve the accuracy of subsequent quantitative image analyses, such as segmentation, feature extraction, or anomaly detection. As of today, GAN-based denoising techniques have shown promising results in various medical imaging modalities, such as reducing noise in low-dose CT images (Wolterink et al., 2017) or enhancing the quality of MRI images (Kim, Do, & Park, 2018).

Nevertheless, generative models can create data samples with features or patterns not present in the training data and do not represent real-world data. This problem is called hallucination and refers to the generation of data samples that do not follow the underlying distribution of training data (Ji et al., 2023; Wang et al., 2020). Reasons for hallucination of GANs include limited training data and the complexity of capturing the underlying data distribution. For example, Hammernik et al., 2021 found that generative models trained for MRI reconstruction with only small pre-training datasets produce hallucination artefacts, with knee-like structures appearing in the brain (Hammernik et al., 2021). Overcoming hallucination is crucial for ensuring the reliability and safety of using images created by generative models, as realistic artifacts introduced by hallucination can facilitate misinterpretation and ultimately misdiagnosis. As hallucinated features are complicated to detect without a manual comparison of original and reconstructed images, medical images generated by GANs should not be used in real patient care (Cohen, Luck, & Honari, 2018).

Image super-resolution

Likewise, GANs can improve the resolution of medical images, a task called super-resolution. It refers to generating high-resolution images from low-resolution inputs, which are noisy, blurry, or contain artifacts due to metal or patient movement (Bulat, Yang, & Tzimiropoulos, 2018). This application of GANs can be especially beneficial if high-resolution images are unavailable or costly to acquire.

In a medical context, high-resolution images are crucial for accurate diagnosis and treatment planning. However, acquiring high-resolution medical images can be challenging due to factors such as imaging

hardware limitations (Zhu, Yang, & Lio, 2019), radiation exposure concerns (Yi & Babyn, 2018), motion artifacts (Shitrit & Riklin Raviv, 2017) or time constraints (Greenspan, 2008). GANs have shown promising results in addressing this challenge by enhancing image resolution in various medical imaging modalities, such as generating standard CT images from low-dose CT (Wolterink et al., 2017; Yi & Babyn, 2018) or by decreasing the slice thickness of MRI (Greenspan et al., 2002; Iglesias et al., 2021; Xia et al., 2021).

For this, GANs are trained on pairs of high and corresponding low resolution images. Here, the generator network of the GAN should learn the mapping necessary to convert a low-resolution image to its corresponding high quality counterpart and use this to create synthetic images that closely resemble the high-quality ground truth images (Zhu, Yang, & Lio, 2019). By aiming to distinguish between the original and generated images, the discriminator of the GAN provides feedback to the generator, which should produce images undistinguishable from the high-resolution ground truth data (Xia et al., 2021; Yi, Walia, & Babyn, 2019). After training, the generator network of the GANs can enhance the resolution of images which were not part of the training data. By this GANs might recover fine details, which would otherwise have been missed in the original low-resolution images and thus aiding radiologists in diagnostic image interpretation.

The generated super-resolution images can also be combined with other deep learning-based image analysis techniques, such as image registration, image fusion, segmentation, or classification tasks, to enhance overall model performance (Iglesias et al., 2021; Xia et al., 2021). GAN-generated super-resolution images can also be used to augment limited datasets and thereby overcome challenges such as a limited availability of high-resolution training data or imbalanced datasets. By generating high-resolution medical images from low-resolution inputs without loss of diagnostic information, GAN-based image-superresolution could mitigate the dependency on time-consuming imaging protocols or high-resolution scanners, thereby reducing cost in the healthcare system (Greenspan et al., 2002).

However, it is important to note that GAN-based image super-resolution techniques currently still face several limitations, as generating high-quality and clinically meaningful images is complex and crucial image details can be missed or hallucinated (Iglesias et al., 2021). Further research and careful validation of the generated high-resolution images is therefore crucial to ensure their clinical reliability.

Reduction of radiation dose and acquisition time

As shortly mentioned in the paragraph above, GAN-based denoising and reconstruction methods can contribute to reduced radiation dose (Yi & Babyn, 2018) or acquisition time (Greenspan et al., 2002) of medical imaging, both of which are highly beneficial. Reducing radiation dose of medical imaging is crucial to minimize patients' exposure to ionizing radiation (Nie et al., 2018) while shorter acquisition times lead to higher patient comfort and reduced healthcare costs (Shitrit & Riklin Raviv, 2017).

However, these benefits must be weighed against maintaining a high image quality for accurate patient care. When acquiring images with lower radiation dose or shorter acquisition time, the images are usually noisier (Sun et al., 2022), which can impede image quality and affect the accuracy of subsequent quantitative image analysis. On the other hand, imaging protocols requiring a prolonged scan time, like MRI, are known to be susceptible to motion artifacts, reducing image quality (Zaitsev, Maclaren, & Herbst, 2015).

By training a GAN on paired datasets of low-dose and corresponding high-dose medical images or pairs of images with high and low acquisition time, respectively, the generator network learns to capture the

underlying structures and features of high-quality images and to generate denoised versions that closely resemble the corresponding high-dose or fast-acquisition time images. It must be noted, however, that besides GANs also CNN-based approaches have achieved remarkable results in reconstructing high-quality images with reduced acquisition time, for example, the compressed sensing artificial intelligence framework (CSAI) introduced by Philips (Foreman et al., 2022; Pezzotti et al., 2020).

Reducing radiation dose or scan time through deep learning-based denoising methods offers several benefits. By reducing acquisition time, GAN-based denoising methods contribute to higher patient comfort, especially during otherwise lengthy scans, such as MRI (Qi et al., 2018; Shitrit & Riklin Raviv, 2017). Shorter acquisition times can also contribute to reduce costs of medical imaging, as it allows for more scans per day and could reduce the number of repeated acquisitions due to patient movement (Dar et al., 2019; Foreman et al., 2022). Last, GAN-based denoising methods can decrease patient exposure to ionizing radiation, which is particularly important for imaging procedures requiring repetitive scans or for patients who are more sensitive to radiation, such as children or pregnant women (Sun et al., 2022; Yu et al., 2020). Minimizing the radiation dose enhances patient safety and reduces the potential of long-term risks associated with cumulative radiation exposure (Nie et al., 2018; Sun et al., 2022).

Anomaly detection

In recent years, GANs have shown promising results in anomaly detection tasks (Kazeminia et al., 2020), which refers to identifying abnormal or anomalous characteristics in medical images that differ from physiological findings (Fernando et al., 2021). After learning the underlying patterns from unremarkable medical images, GANs can be used to detect deviations that may indicate the presence of disease (Meissen et al., 2023; Yi, Walia, & Babyn, 2019).

This approach involves training a GAN on a dataset containing only medical images from healthy subjects to capture the latent representation of these images and to generate realistic-looking images of healthy patients (Kazeminia et al., 2020). After training, the discriminator of the GAN can be used to identify images deviating from the learned representation of healthy subject, thereby classifying images as normal or containing anomalies (Fernando et al., 2021; Yi, Walia, & Babyn, 2019). Various techniques can be used to quantify the differences between images and identify anomalies, such as pixel-wise or feature-based comparisons (Meissen, Kaissis, & Rueckert, 2022; Schlegl et al., 2019).

The main advantages of GAN-based anomaly detection are that the training data have to contain images of healthy patients only and that no manual labeling of abnormal samples is needed, as the GAN learns patterns solely from the distribution of data containing medically unremarkable images in an unsupervised manner (Kazeminia et al., 2020; Schlegl et al., 2019). By identifying deviations from this learned pattern, GANs can assist physicians in detecting anomalies in medical images (Meissen et al., 2023), which could be especially beneficial in screening settings or to enable earlier diagnosis of patients who have only subtle image findings.

Image-to-image translation and style transfer

Image translation and style transfer are two other important research areas in the field of quantitative image analysis based upon GANs and other deep learning techniques.

Image translation refers to converting an image from one type into another (Li et al., 2021b; Zhu et al., 2017). In the context of medical imaging, this can involve converting MRI scans to CT-like images (Qian et al., 2020) and vice versa (Zhang, Yang, & Zheng, 2018) or transforming an image to simulate the appearance of a different imaging protocol (Dar et al., 2019). Image translation is usually based on GANs, who were trained to learn the mapping between different image modalities (Li et al., 2021b). After training, the GANs can be used to generate synthetic images mimicking the appearance of another modality or protocol while maintaining the underlying anatomic information or pathologic findings of the original image (Li et al., 2021b). This could be especially beneficial in low-resource settings, where some imaging modalities such as MRI may not be available, or to avoid imaging modalities associated with higher radiation exposure, such as CT (Li, 2022).

Besides, image translation techniques can be applied in quantitative image analysis research and medical imaging optimization in general. For example, researchers can simulate acquiring medical images under different imaging protocols or parameters using image translation applications (Olut et al., 2018), enabling them to explore the effects of different settings and optimize protocols without the need to acquire extensive additional images of patients. This not only mitigates the need of prolonged scan times but also reduces the costs associated with additional scans (Kazemini et al., 2020; Olut et al., 2018; Yi, Walia, & Babyn, 2019). In case of unavailable or missing data, image translation techniques can be used to generate synthetic images filling this gap and thereby providing more complete datasets for research (Dar et al., 2019; Olut et al., 2018).

Likewise, image translation can be used for data augmentation by creating additional, synthetic images with a different modality or with different parameters and thereby increasing the diversity and size of training data (Celard et al., 2022; Chartsias et al., 2017; Yi, Walia, & Babyn, 2019). This approach can be used to overcome challenges associated with a limited availability of certain imaging modalities or imbalanced datasets (Kazemini et al., 2020; Li et al., 2021b) and thus aid in improving the performance and generalization of machine learning algorithms trained on these data (Chartsias et al., 2017; Dar et al., 2019). Furthermore, generating synthetic images can mitigate privacy issues when collecting or publishing datasets (Guibas, Virdi, & Li, 2017). However, rigorous privacy guarantees and quality control mechanisms will have to be established for AI-based image synthesis techniques, such as image translation, to truly overcome issues concerning privacy and limited data availability, as explained in the previous section on image denoising and reconstruction.

Last, image translation allows for cross-modality analysis, where a deep learning model trained on images of one modality can be applied to those of a different one (Chen et al., 2021). For example, a model trained on CT images could be used to analyze MRI images after transforming them to a CT-like appearance. This approach can enable analysis across different imaging modalities, providing complementary information for better diagnostic accuracy (Nie et al., 2018).

Style transfer, on the other hand, refers to modifying the visual style of an image while preserving its content (Jing et al., 2020). This concept originates from the field of computer vision and has been successfully applied to various domains, including art, graphics and photography (Jing et al., 2020; Karras, Laine, & Aila, 2019).

In the context of medical imaging, style transfer techniques allow for the modification of certain imaging aspects such as colors, textures or protocol styles while preserving the anatomic and disease-specific

findings of the original image (Iqbal & Ali, 2018; Xu, Li, & Shin, 2020). This is achieved by separating the content of an image, such as underlying structures, from its style information, which includes visual characteristics of the image such as brightness, contrast or texture (Jing et al., 2020; Ma, Ji, & Gao, 2019; Zhu et al., 2017). By extracting and manipulating style and content components separately, style-transfer algorithms allow for the modification of style elements while maintaining the diagnostic content of medical images.

This offers several potential benefits, such as highlighting specific structures or features to enhance diagnostic image analysis or aid in the interpretation of findings (Qin et al., 2022). Furthermore, style transfer techniques can be used to adjust for differences in medical images obtained by different scanners or using different protocols, to improve the performance of quantitative image analysis algorithms when working with inhomogeneous or multi-center data (Ma, Ji, & Gao, 2019).

Both image translation and style transfer techniques rely on deep learning models, particularly on GANs (Celard et al., 2022), and offer opportunities both for medical imaging based research and patient care, as they enable the transformation of images between modalities or protocols (Singh & Raza, 2021), allow for simulation of different acquisition protocols and cross-modality analysis (Kazemina et al., 2020), facilitate the implementation of enhanced visualization techniques and can be used for data augmentation (Celard et al., 2022).

However, validating the results of image translation and style transfer techniques in the medical domain is crucial. While these techniques can enhance the visualization of medical images, they could introduce misleading artifacts and impair the diagnostic accuracy of human readers (Yi, Walia, & Babyn, 2019). Thus, both techniques should not be routinely used in patient care (Singh & Raza, 2021). In research, the generated images need to be carefully validated by medical experts to ensure that the essential diagnostic information is preserved (Kazemina et al., 2020).

5.3.2 Large language models

In recent years, large language models have emerged as powerful AI tools in natural language processing tasks, such as text generation (Wei et al., 2023), summarization (Zhang et al., 2023a), and document classification (Chan, Schweter, & Möller, 2020). They can further be employed in machine translation tasks (Zhang, Haddow, & Birch, 2023), where they have demonstrated significant improvements over traditional rule-based or statistical methods. Nowadays, large language models are part of many aspects of daily life, as they are applied in virtual assistants (Gondala et al., 2021) and chatbots (Jo et al., 2023), where they allow for interactive and context-aware conversations. One of the most notable advances in large language models was the introduction of OpenAI's GPT (Generative Pre-trained Transformer) model series (Floridi & Chiriatti, 2020), which have gained remarkable attention both in the computer science community and in the non-expert domain.

Such large language models typically have billions of trainable parameters and consist of very large transformers or ensembles of transformers (Shen et al., 2023). The transformers in turn comprise multiple layers of self-attention and feed-forward neural networks (Parmar et al., 2018; Vaswani et al., 2017), allowing the model to attend to different parts of the input and to capture long-range dependencies (Sukhbaatar et al., 2019), which ultimately leads to more coherent and contextually appropriate text generation.

During training, large language models are usually provided with a large corpus of text data (Wei et al., 2023), such as books, articles, or web pages, allowing them to learn semantic relationships, syntactic structures present in the training data. Based on this, the models are trained to predict the next word in a given sentence based on the preceding words (Wei et al., 2023) using a combination of supervised, semi-supervised and reinforcement learning (Radford et al., 2019).

Image description and metadata analysis

As of writing of this theses, employing large language models on medical data is a relatively new area of research, and literature on this topic is scarce. Possible applications include diagnostic assistance or patient education (Zhang et al., 2023b). Furthermore, large language models could be deployed in various ways in medical image analysis in the future (Shin, Lu, & Summers, 2017).

For example, large language models can be used to generate image captions (Vinyals et al., 2015) and to provide detailed textual descriptions for specific image regions or whole images (Karpathy & Fei-Fei, 2015; Vinyals et al., 2015). In medical image analysis, this can be used to generate preliminary reports or to standardize free-text radiology records into a structured report (Mallio et al., 2023). This could simplify data management processes, facilitate searching and retrieving specific images based on metadata criteria, or support research studies requiring metadata-driven analyses.

Image annotation and labeling

Furthermore, large language models can assist in medical image annotation and labeling tasks by providing automated annotations (Karpathy & Fei-Fei, 2015; Shin, Lu, & Summers, 2017). Especially combining large language models with other network architectures for image analysis (Radford et al., 2021), such as CNNs, and incorporating other sources of medical information, such as electronic health records, offers new ways of automated, multimodal medical images analysis. Given this range of information, large language models can annotate distinct regions of interest within images or detect the presence of specific pathologies (Shin et al., 2016). Furthermore, they can provide automated labeling of imaging datasets, thus supporting data organization and enabling the collecting large, annotated datasets for machine learning applications without the need of extensive manual labor. This might ultimately aid in developing automated pipelines and CAD systems.

Another advantage of large language models is the possibility to query the model about its predictions, like determining the certainty of a generated annotation. The model's reasoning behind its output can be inquired using natural language, which is often more intuitive to users than using other explainability methods to visualize attributions in the image space (Tursun et al., 2023).

Integration into patient care

In clinical routine, large language models could assist in various language-based tasks, including report generation (Javaid, Haleem, & Singh, 2023) or summarization of radiology reports (Cai et al., 2021). Based on prompts provided by a physician or a radiologist, these models could automatically generate drafts for medical reports and further assist in style and formatting (Biswas, 2023), thus reducing the time required for

humans to generate reports and improving the efficiency of bureaucratic processes (Klang, Cohen-Shelly, & Lopez-Jimenez, 2023).

Furthermore, large language models could assist in providing the reports in a structured fashion, thereby improving the consistency of records and lowering the risk of missing data or errors (Klang, Cohen-Shelly, & Lopez-Jimenez, 2023). Large language models could also integrate imaging data with their knowledge of medical literature, guidelines, and patient data from the hospital's information system or electronic health records to provide physicians with context-specific information and enhance clinical decision-making (Biswas, 2023).

Besides, large language models can generate patient-friendly explanations by transforming complex medical terminology into an easy-to-understand language for patients without a medical background (Jeblick et al., 2022; Klang, Cohen-Shelly, & Lopez-Jimenez, 2023). This can improve patient education, enhancing their compliance and facilitating shared-decision making. Furthermore, they can be used to create educational materials for medical students in an interactive way (Javaid, Haleem, & Singh, 2023).

Assisting organization tasks and research

Large language models could assist to improve the clinical documentation process by extracting relevant information from radiology reports or patient-written notes and automatically generate codes corresponding to a patient's diagnoses and received treatments needed for accounting, thus reducing administrative burdens (Maas et al., 2020).

Furthermore, by processing radiology reports, imaging metadata or other text-based patient information (Hu et al., 2023), large language models assist in searching, retrieving, and categorizing images based on specific criteria, such as modality, patient characteristics or disease. This could be especially beneficial in research, as it allows for more efficient retrieval of images based on search queries without manual curation of data.

Additionally, large language models could be used in standardized reporting frameworks, where they provide structured information based on recognized medical terminology in reports or coding systems (Klang, Cohen-Shelly, & Lopez-Jimenez, 2023).

Current challenges in the context of medical image analysis

Despite their remarkable achievements, large language models still face several challenges. One key limitation is the enormous computational resources required for training (Del Corro et al., 2023), including expensive hardware such as GPUs cluster, along with efficient parallel processing techniques to handle the vast amount of data and the complex model architecture (Zeng et al., 2021). Other challenges are related to AI safety, like the of biases and fairness, as large language models can inadvertently capture biases present in the training data (Wei et al., 2023).

Moreover, the ethical implications of large language models are a subject of concern (Klang, Cohen-Shelly, & Lopez-Jimenez, 2023). Like generative models working with imaging data, large language models are prone to hallucination, and might generate plausible-sounding but incorrect information (Ji et al., 2023). In medical applications, this could result in erroneous diagnoses or treatment recommendations that may be

harmful for patients. The potential for misuse, such as generating misleading or malicious content with possibly fatal consequences for patient care, highlights the need for responsible use and regulation.

Besides, the alignment of AI systems with human values (Gabriel & Ghazavi, 2021), ethics or rules (Gabriel, 2020) remains a challenging topic in developing large language models. In the medical context, it is crucial that AI models prioritize a patient's well-being and adhere to clinical guidelines to produce safe and substantiated recommendations. AI applications in the medical domain should be fair and non-maleficent (Jobin, Ienca, & Vayena, 2019), as ensuring patient safety is a fundamental concern. Furthermore, retaining patient privacy and ensuring data protection remains a major topic of concern and is subject to ongoing research (Jeblick et al., 2022; Jobin, Ienca, & Vayena, 2019; Shin, Lu, & Summers, 2017).

Continued research, including a realistic and critical evaluation, addressing biases, ensuring patient privacy and data security and continuous training and refinement of large language models, will be essential to enable their effective and responsible use in medical applications in the future.

6 Conclusion

The main goal of this thesis was to enhance diagnostic radiological performance and improve patient care through machine learning-based assistance. Two essential quantitative image analysis methods were employed in the underlying publications of this thesis: Radiomics analysis and deep learning with a neural network.

In the first publication P-I (Kaissis et al., 2020a), radiomics-based imaging features were identified as independent predictors for postoperative overall survival in patients with PDAC, emphasizing the role of radiologists in personalized medicine. In the second publication P-II (Jungmann et al., 2022), a neural network was developed to classify mammogram lesions, demonstrating significant performance improvements of radiologists with AI assistance. The study highlighted the potential of combining human and artificial intelligence, especially when AI systems are transparent and explainable, suggesting their integration into radiological routines as a second reader for enhanced patient outcomes.

Furthermore, this thesis includes a thorough research about the two main techniques, radiomics analyses and deep learning approaches (Chapter 2), as well as a detailed analysis of their advantages and challenges (Chapter 5). Both techniques were applied to medical problems in order to assess their application in clinical routine (Chapter 3).

This chapter summarizes the work done for this thesis and both publications P-I and P-II are outlined (Jungmann et al., 2022; Kaissis et al., 2020a). The main advantages and challenges of radiomics and deep learning as quantitative imaging techniques in medical settings are presented and an outlook for future applications of quantitative image analysis as well as possible improvements is given.

6.1 Summary of underlying publications

The first publication P-I explored radiomics-based quantitative image analysis and presents the results of a multiparametric analysis for predicting overall survival in a cohort of patients with PDAC, who underwent primary resection (Kaissis et al., 2020a). For this, radiomics features of the tumor were obtained using PyRadiomics based upon fully automated segmentation of pancreas and tumor, which were manually corrected. At publication, P-I was the first paper to systematically evaluate a wide range of currently available risk biomarkers for PDAC, including quantitative imaging characteristics, clinical, histopathological and morpho-molecular features.

Quantitative imaging features were found to be an independent predictor of overall survival in patients with PDAC and enabled 'the multiparametric, machine learning-assisted modelling of postoperative overall survival with a high performance compared to clinical and morpho-molecular/ genetic parameters' (Jungmann et al., 2022, p. 1). Although radiomics features did not outperform the clinical and biological risk biomarkers, they constitute a non-invasive and thus relatively risk-free way of assessing patient survival.

Nevertheless, the study was subject to some limitations, which were mainly due to high financial and personal cost of both morpho-molecular and genetic workup of PDAC specimen and manual work involved in radiomics feature extraction, like manual segmentation correction. This led to a relatively small cohort size of 103 patients and the necessity of feature reduction using PCA to mitigate the risk of model overfitting. As a result, the identification of single, well documented radiomics features was not possible. Instead, imaging feature compositions were found to be an independent predictor of patient survival, which are more difficult to reassess in following studies.

These limitations align with current challenges of radiomics analysis, where manual workload involved in feature extraction often impedes the analysis of large data and the creation of feature signatures or different softwares used for feature extraction reduces the reproducibility of findings. Nevertheless, the study confirms the possibility of extracting predictive biomarkers from medical images, which are not clearly visible to the human eye. These quantitative imaging biomarkers may contribute to better understanding of tumor biology and improved patient care in the course of personalized medicine in future.

The second publication, P-II, explored possible applications of quantitative image analysis based upon deep learning techniques. It evaluated the effect of different types of AI-based assistance on radiologists in classifying BI-RADS 4 lesions into benign and malignant findings and analyzed the human-AI-interaction (Jungmann et al., 2022). For this, a ResNet 50 pretrained on ImageNet was trained on the public CBIS-DDSM dataset and an independent test set of 200 in-house mammograms with histopathologic labels was collected. Different types of AI assistance were generated based upon the test set: An occlusion sensitivity map as well as the networks prediction along with its certainty. The effect of different types of AI-based assistance on radiologists' performance were measured in a retrospective reader study and the influence of the Big Five personality traits was analyzed.

Diagnostic accuracy of radiologists was significantly improved using AI-based assistance and the performance of combined human and artificial intelligence was superior to both individual groups. Furthermore, providing more insights into the algorithm's classification process to readers through the generated methods of AI assistance lead to increased trust in the algorithm's performance, which was 'mostly dependent on the certainty of the predictions in combination with a plausible heatmap. Human-AI interaction varied widely and was influenced by personality traits' (Jungmann et al., 2022, p. 1).

In concordance with previous research, this publication supports AI as a second reader in radiologic workflows to improve diagnostic accuracy and thereof patient outcome, to enable the efficient analysis of large data and to eventually reduce cost in the healthcare system. Furthermore, it highlights the importance of explainable AI in the context of medical settings to enable trust and thereby an improved human-AI-collaboration.

Still, several limitations of this publications have to be considered, including differences of study set up to routine workflows, which may influence the generalizability of findings, as well as the small number of participating radiologists. As a result, confounding cannot be ruled out, especially regarding the influence of personality traits on human-AI-collaboration. To establish recommendations for improved human-AI interaction, further studies are required, preferably in collaboration with psychological research.

6.2 Current state of quantitative image analysis

As discussed in Chapter 5, quantitative image analysis is a rapidly developing field of research, which offers many advantages compared to traditional medical image analysis techniques. However, there are still several challenges needed to be overcome to enable broad implementation of quantitative image analysis techniques in clinical routine.

Major benefits of quantitative image analysis include the potential to improve diagnostic accuracy, provide objective measurements, enhance diagnostic efficiency and reduce cost in both research healthcare system. Quantitative image analysis enables the extraction of quantitative features from routine medical images, which can be used to support clinical decision-making and improve patient outcomes, which is especially beneficial in personalized medicine. By extracting large amounts of information from medical images, these techniques allow for a more precise and individualised approach to diagnosis, prognosis, and treatment planning. As of writing of this thesis, multiple studies have shown different quantitative image analysis techniques to constitute a potential benefit in the course of personalized medicine, including development of imaging biomarkers (Cui et al., 2021), survival prediction (Kaissis et al., 2020a) or imaging features correlating with predictive, histomorphologic characteristics of cancer (Kaissis et al., 2019; Wang et al., 2022).

In the underlying publications of this thesis, two methods of quantitative image analysis techniques were used: Radiomics analysis and a deep learning approach. Both techniques have distinct benefits and challenges, and are thus suited for different kinds of applications.

Radiomics analyses have the advantage of defined, numerical features extracted from images, which can be well understood. Furthermore, by (semi-) automatic or manual segmentation of the ROI or VOI, features are extracted from a specific region of interest which leads to less noisy data. Thus, radiomics studies usually do not need a cohort the size of deep learning studies. However, human workload involved in the segmentation process impedes the analysis of large datasets and the radiomics workflow is not suited for integration into existing radiomic routine. Radiomics is thereof especially suited for clinical studies with smaller study populations and well curated data, which can be annotated or segmented by human experts. Common tasks include the development of imaging biomarkers for the prediction of therapeutic efficacy or histopathologic characteristics, such as subtypes or mutation status.

AI applications on the other hand have the advantage of being capable of analyzing whole images, which eliminates the dependency on manual segmentations and can lead to more holistic image analysis. This advantage however comes with the challenge of more noisy data due to additional imaging information besides the target lesion, which is why AI applications usually require significantly more training data than radiomics approaches. As large, multi-centric data are challenging to collect, especially regarding patient privacy and autonomy over their data, this requirement poses a major obstacle for quantitative image analysis studies based on deep learning algorithms.

In the first publication P-I, quantitative image features were shown to independently predict patient survival in a non-invasive way, whilst not outperforming broadly available clinical data (Kaissis et al., 2020a). In the second publication P-II, the combination of human and artificial intelligence lead to an increased diagnostic accuracy compared to human readers alone (Jungmann et al., 2022). In summary, two main quantitative imaging techniques were employed in the underlying publications of this thesis to enhance

diagnostic radiological performance and improve patient care through machine learning-based assistance. The publications emphasize the potential role of radiologists in personalized medicine and explore the integration of AI-based assistance into radiological routines for enhanced patient outcomes.

6.3 Future work

Despite the possible benefits of quantitative image analysis, there are still major issues hindering implementation into clinical workflows, which need to be overcome. Especially lacking standardization and generalizability of both radiomics and deep learning models, which impedes reproducibility of results so that false-positive discoveries cannot be eliminated, poses a major obstacle for broad implementation of quantitative image analysis techniques in patient care. Thus, the development of reliable and standardised methods for data acquisition and analysis is essential to ensure accurate, reproducible results and to enable the development of qualitative image analysis models that generalize across different patient populations and medical conditions.

Furthermore, the collection of well curated, large, multi-center data needed for quantitative image analysis while preserving patient privacy and autonomy is a challenging task. This holds especially true for deep learning approaches, as large amount of data are required to sufficiently train neural networks. Besides, preserving patient privacy both in data sharing and algorithm implementation needs to be addressed. Thus, the field of privacy preserving AI is expected to grow significantly over the coming years. Other challenges that need to be addressed include human obstacles related to lacking interpretability and non-intuitiveness of quantitative image analysis results and difficulties in including such techniques into clinical workflows. For this, more research on human-AI-collaboration and human-in-the-loop will be needed to enable a successful integration into clinical workflows and to avoid impairing of human performance.

Furthermore, due to the 'black-box' nature of AI, a lack of trust and lacking possibilities of quality control can hinder collaboration between radiologists and algorithmic assistance, which emphasizes the urge for more transparent and explainable AI systems. This aligns with the regulations set by the European Union Artificial Intelligence Act, which future deep-learning based quantitative image analysis models will have to comply with. However, due to less human involvement than radiomics analyses and the possibility of end-to-end applications, AI-based quantitative image analysis techniques have the potential to be integrated into radiologic workflow, including tasks like object detection, image segmentation or disease prediction, rather than being constrained to clinical studies.

Despite several challenges, which will need to be addressed, future perspectives of quantitative analysis of medical images, including radiomics and deep learning approaches, are promising. Continued efforts towards standardization and regulation of processes, further advances in AI and imaging technology, development of transparency and explainability methods as well as secure and privacy-preserving AI are expected to enhance the accuracy and clinical applicability of quantitative image analysis. Furthermore, deployment of generative models, like GANs or large language models, can improve medical image analysis in several ways, including image denoising, report standardization, data augmentation and implementation into patient care. The integration of quantitative image analysis techniques with other medical data, such as clinical, genetic or histomorphologic data, will enable a more comprehensive and personalized approach to

disease diagnosis and treatment, which is especially important in the course of personalized medicine and will eventually improve patient outcomes.

A References

- Abbassi, R., & Algül, H. (2019). Systemische Therapie des Pankreaskarzinoms - es hat sich viel getan. *InFo Hämatologie + Onkologie*, 22(11), 22–26. <https://doi.org/10.1007/s15004-019-6746-z>
- Abukmeil, M., Ferrari, S., Genovese, A., Piuri, V., & Scotti, F. (2021). A survey of unsupervised generative models for exploratory data analysis and representation learning. *ACM Comput. Surv.*, 54(5). <https://doi.org/10.1145/3450963>
- Adsay, N. V., Bagci, P., Tajiri, T., Oliva, I., Ohike, N., Balci, S., Gonzalez, R. S., Basturk, O., Jang, K.-T., & Roa, J. C. (2012). Pathologic staging of pancreatic, ampullary, biliary, and gallbladder cancers: pitfalls and practical limitations of the current AJCC/UICC TNM staging system and opportunities for improvement. *Seminars in Diagnostic Pathology*, 29(3), 127–141. <https://doi.org/10.1053/j.semdp.2012.08.010>
- Aerts, H. J. W. L. (2016). The Potential of Radiomic-Based Phenotyping in Precision Medicine. *JAMA Oncology*, 2(12), 1636. <https://doi.org/10.1001/jamaoncol.2016.2631>
- Aerts, H. J. W. L., Velazquez, E. R., Leijenaar, R. T. H., Parmar, C., Grossmann, P., Carvalho, S., Bussink, J., Monshouwer, R., Haibe-Kains, B., Rietveld, D., Hoebers, F., Rietbergen, M. M., Leemans, C. R., Dekker, A., Quackenbush, J., Gillies, R. J., & Lambin, P. (2014). Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nature Communications*, 5(1). <https://doi.org/10.1038/ncomms5006>
- Afshar, P., Mohammadi, A., Plataniotis, K. N., Oikonomou, A., & Benali, H. (2019). From Handcrafted to Deep-Learning-Based Cancer Radiomics: Challenges and Opportunities. *IEEE Signal Processing Magazine*, 36(4), 132–160. <https://doi.org/10.1109/msp.2019.2900993>
- Ahmed, S., Bradshaw, A.-D., Gera, S., Dewan, M. Z., & Xu, R. (2017). The TGF- β /smad4 signaling pathway in pancreatic carcinogenesis and its clinical significance. *Journal of Clinical Medicine*, 6(1). <https://doi.org/10.3390/jcm6010005>
- Akkus, Z., Cai, J., Boonrod, A., Zeinoddini, A., Weston, A. D., Philbrick, K. A., & Erickson, B. J. (2019). A survey of deep-learning applications in ultrasound: Artificial intelligence-powered ultrasound for improving clinical workflow [Special Issue: Quality and Data Science]. *Journal of the American College of Radiology*, 16(9, Part B), 1318–1328. <https://doi.org/10.1016/j.jacr.2019.06.004>
- Albawi, S., Mohammed, T. A., & Al-Zawi, S. (2017). Understanding of a convolutional neural network. *2017 International Conference on Engineering and Technology (ICET)*. <https://doi.org/10.1109/icengtechnol.2017.8308186>
- Alberich-Bayarri, A., Sourbron, S., Golay, X., deSouza, N., Smits, M., van der Lugt, A., & Boellard, R. (2020). ESR statement on the validation of imaging biomarkers. *Insights into Imaging*, 11(1). <https://doi.org/10.1186/s13244-020-00872-9>

- Alloghani, M., Al-Jumeily, D., Mustafina, J., Hussain, A., & Aljaaf, A. J. (2020). A Systematic Review on Supervised and Unsupervised Machine Learning Algorithms for Data Science. In M. W. Berry, A. Mohamed, & B. W. Yap (Eds.), *Supervised and unsupervised learning for data science* (pp. 3–21). Springer International Publishing. https://doi.org/10.1007/978-3-030-22475-2_1
- Ambe, C., Fulp, W., Springett, G., Hoffe, S., & Mahipal, A. (2015). A Meta-analysis of Randomized Clinical Trials of Chemoradiation Therapy in Locally Advanced Pancreatic Cancer. *Journal of Gastrointestinal Cancer*, *46*(3), 284–290. <https://doi.org/10.1007/s12029-015-9734-z>
- American Cancer Society. (2023a). Breast Cancer Early Detection and Diagnosis [last accessed on 12.04.2023]. <https://www.cancer.org/content/dam/CRC/PDF/Public/8579.00.pdf>
- American Cancer Society. (2023b). Cancer Facts & Figures 2023 [last accessed on 08.03.2023]. <https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2023/2023-cancer-facts-and-figures.pdf>
- American Cancer Society. (2023c). Treating Breast Cancer [last accessed on 12.04.2023]. <https://www.cancer.org/content/dam/CRC/PDF/Public/8581.00.pdf>
- American College of Radiology. (2013). BI-RADS Atlas - Reporting System [last accessed on 11.01.2023]. <https://www.acr.org/-/media/ACR/Files/RADS/BI-RADS/Mammography-Reporting.pdf>
- Apostolou, P., & Papatirou, I. (2017). Current perspectives on CHEK2 mutations in breast cancer [PMID: 28553140]. *Breast Cancer: Targets and Therapy*, *9*, 331–335. <https://doi.org/10.2147/BCTT.S111394>
- Ardila, D., Kiraly, A. P., Bharadwaj, S., Choi, B., Reicher, J. J., Peng, L., Tse, D., Etemadi, M., Ye, W., Corrado, G., Naidich, D. P., & Shetty, S. (2019). End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nature Medicine*, *25*(6), 954–961. <https://doi.org/10.1038/s41591-019-0447-x>
- Attanasio, S., Forte, S. M., Restante, G., Gabelloni, M., Guglielmi, G., & Neri, E. (2020). Artificial intelligence, radiomics and other horizons in body composition assessment. *Quantitative imaging in medicine and surgery*, *10*, 1650–1660. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7378090/>
- Aung, K. L., Fischer, S. E., Denroche, R. E., Jang, G.-H., Dodd, A., Creighton, S., Southwood, B., Liang, S.-B., Chadwick, D., Zhang, A., O’Kane, G. M., Albaba, H., Moura, S., Grant, R. C., Miller, J. K., Mbabaali, F., Pasternack, D., Lungu, I. M., Bartlett, J. M. S., . . . Knox, J. J. (2018). Genomics-Driven Precision Medicine for Advanced Pancreatic Cancer: Early Results from the COMPASS Trial. *Clinical Cancer Research*, *24*(6), 1344–1354. <https://doi.org/10.1158/1078-0432.ccr-17-2994>
- Avanzo, M., Wei, L., Stancanello, J., Vallières, M., Rao, A., Morin, O., Mattonen, S. A., & El Naqa, I. (2020). Machine and deep learning methods for radiomics. *Medical Physics*, *47*(5), e185–e202. <https://doi.org/10.1002/mp.13678>
- Bagher-Ebadian, H., & Chetty, I. J. (2021). Technical note: Radiomix: A validated software for radiomics analysis of medical images in radiation oncology. *Medical Physics*, *48*(1), 354–365. <https://doi.org/10.1002/mp.14590>
- Bai, W., Sinclair, M., Tarroni, G., Oktay, O., Rajchl, M., Vaillant, G., Lee, A. M., Aung, N., Lukaschuk, E., Sanghvi, M. M., Zemrak, F., Fung, K., Paiva, J. M., Carapella, V., Kim, Y. J., Suzuki, H., Kainz, B.,

- Matthews, P. M., Petersen, S. E., . . . Rueckert, D. (2018). Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. *Journal of Cardiovascular Magnetic Resonance*, 20(1). <https://doi.org/10.1186/s12968-018-0471-x>
- Bailey, P., Chang, D. K., Nones, K., Johns, A. L., Patch, A.-M., Gingras, M.-C., Miller, D. K., Christ, A. N., Bruxner, T. J. C., Quinn, M. C., Nourse, C., Murtaugh, L. C., Harliwong, I., Idrisoglu, S., Manning, S., Nourbakhsh, E., Wani, S., Fink, L., Holmes, O., . . . Grimmond, S. M. (2016). Genomic analyses identify molecular subtypes of pancreatic cancer. *Nature*, 531(7592), 47–52. <https://doi.org/10.1038/nature16965>
- Balleuguier, C., Ayadi, S., Van Nguyen, K., Vanel, D., Dromain, C., & Sigal, R. (2007). BIRADS classification in mammography [Breast Imaging: The BIRADS of the ACR]. *European Journal of Radiology*, 61(2), 192–194. <https://doi.org/10.1016/j.ejrad.2006.08.033>
- Bardwell, W. A., & Ancoli-Israel, S. (2008). Breast Cancer and Fatigue [Sleep and Disorders of Sleep in Women]. *Sleep Medicine Clinics*, 3(1), 61–71. <https://doi.org/10.1016/j.jsmc.2007.10.011>
- Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82–115. <https://doi.org/10.1016/j.inffus.2019.12.012>
- Basodi, S., Ji, C., Zhang, H., & Pan, Y. (2020). Gradient amplification: An efficient way to train deep neural networks. *Big Data Mining and Analytics*, 3(3), 196–207. <https://doi.org/10.26599/BDMA.2020.9020004>
- Begley, C. G., & Ioannidis, J. P. A. (2015). Reproducibility in science. *Circulation Research*, 116(1), 116–126. <https://doi.org/10.1161/CIRCRESAHA.114.303819>
- Beitzel, K., Brüning, R., Debus, J., Engelbrecht, V., Fritzsche, S., Grütznert, G., Haug, A., Helck, A., Hellerhoff, K., Helmberger, T., Hentschel, H., Herrmann, K. A., Hetterich, H., Hünerbein, R., Jäkel, O., Karger, C. P., Koerte, I. K., Krüger-Stollfuß, I., Kuhn, F.-P., . . . Zech, C. J. (2017, January). *Duale Reihe Radiologie* (M. Reiser, F.-P. Kuhn, & J. Debus, Eds.; 4th ed., Vol. 4). Georg Thieme Verlag. <https://doi.org/10.1055/b-004-132212>
- Berg, W. A., Blume, J. D., Cormack, J. B., Mendelson, E. B., Lehrer, D., Böhm-Vélez, M., Pisano, E. D., Jong, R. A., Evans, W. P., Morton, M. J., Mahoney, M. C., Hovanessian Larsen, L., Barr, R. G., Farria, D. M., Marques, H. S., Boparai, K., & ACRIN 6666 Investigators, f. t. (2008). Combined Screening With Ultrasound and Mammography vs Mammography Alone in Women at Elevated Risk of Breast Cancer. *JAMA*, 299(18), 2151–2163. <https://doi.org/10.1001/jama.299.18.2151>
- Berger, T., Noble, D. J., Yang, Z., Shelley, L. E. A., McMullan, T., Bates, A., Thomas, S., Carruthers, L. J., Beckett, G., Duffton, A., Paterson, C., Jena, R., McLaren, D. B., Burnet, N. G., & Nailon, W. H. (2023). Assessing the generalisability of radiomics features previously identified as predictive of radiation-induced sticky saliva and xerostomia. *Physics and Imaging in Radiation Oncology*, 25, 100404. <https://doi.org/10.1016/j.phro.2022.12.001>
- Berrada, L., De, S., Shen, J. H., Hayes, J., Stanforth, R., Stutz, D., Kohli, P., Smith, S. L., & Balle, B. (2023). Unlocking accuracy and fairness in differentially private image classification. <https://doi.org/10.48550/ARXIV.2308.10888>

- Berrar, D. (2019). Cross-validation. In S. Ranganathan, M. Gribskov, K. Nakai, & C. Schönbach (Eds.), *Encyclopedia of bioinformatics and computational biology* (pp. 542–545). Academic Press. <https://doi.org/10.1016/B978-0-12-809633-8.20349-X>
- Biankin, A. V., Waddell, N., Kassahn, K. S., Gingras, M.-C., Muthuswamy, L. B., Johns, A. L., Miller, D. K., Wilson, P. J., Patch, A.-M., Wu, J., Chang, D. K., Cowley, M. J., Gardiner, B. B., Song, S., Harliwong, I., Idrisoglu, S., Nourse, C., Nourbakhsh, E., Manning, S., . . . Grimmond, S. M. (2012). Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. *Nature*, *491*(7424), 399–405. <https://doi.org/10.1038/nature11547>
- Biswas, S. (2023). Chatgpt and the future of medical writing [PMID: 36728748]. *Radiology*, *307*(2), e223312. <https://doi.org/10.1148/radiol.223312>
- Blackford, A., Serrano, O. K., Wolfgang, C. L., Parmigiani, G., Jones, S., Zhang, X., Parsons, D. W., Lin, J. C.-H., Leary, R. J., Eshleman, J. R., Goggins, M., Jaffee, E. M., Iacobuzio-Donahue, C. A., Maitra, A., Cameron, J. L., Olino, K., Schlick, R., Winter, J., Herman, J. M., . . . Hruban, R. H. (2009). SMAD4 Gene Mutations Are Associated with Poor Prognosis in Pancreatic Cancer. *Clinical Cancer Research*, *15*(14), 4674–4679. <https://doi.org/10.1158/1078-0432.ccr-09-0227>
- Boonen, R. A. C. M., Vreeswijk, M. P. G., & van Attikum, H. (2022). CHEK2 variants: linking functional impact to cancer risk. *Trends in Cancer*, *8*(9), 759–770. <https://doi.org/10.1016/j.trecan.2022.04.009>
- Borawar, L., & Kaur, R. (2023). Resnet: Solving vanishing gradient in deep networks. In R. P. Mahapatra, S. K. Peddoju, S. Roy, & P. Parwekar (Eds.), *Proceedings of international conference on recent trends in computing. lecture notes in networks and systems* (pp. 235–247, Vol. 600). Springer Nature, Singapore. https://link.springer.com/chapter/10.1007/978-981-19-8825-7_21
- Brem, R. F., Ioffe, M., Rapelyea, J. A., Yost, K. G., Weigert, J. M., Bertrand, M. L., & Stern, L. H. (2009). Invasive Lobular Carcinoma: Detection with Mammography, Sonography, MRI, and Breast-Specific Gamma Imaging [PMID: 19155397]. *American Journal of Roentgenology*, *192*(2), 379–383. <https://doi.org/10.2214/AJR.07.3827>
- Brett, J. O., Spring, L. M., Bardia, A., & Wander, S. A. (2021). ESR1 mutation as an emerging clinical biomarker in metastatic hormone receptor-positive breast cancer. *Breast Cancer Research*, *23*(1). <https://doi.org/10.1186/s13058-021-01462-3>
- Brouxhon, S. M., Kyrkanides, S., Teng, X., Athar, M., Ghazizadeh, S., Simon, M., O'Banion, M. K., & Ma, L. (2013). Soluble E-cadherin: a critical oncogene modulating receptor tyrosine kinases, MAPK and PI3K/Akt/mTOR signaling. *Oncogene*, *33*(2), 225–235. <https://doi.org/10.1038/onc.2012.563>
- Buckler, A. J., Bresolin, L., Dunnick, N. R., & Sullivan, D. C. (2011). A Collaborative Enterprise for Multi-Stakeholder Participation in the Advancement of Quantitative Imaging. *Radiology*, *258*(3), 906–914. <https://doi.org/10.1148/radiol.10100799>
- Budd, S., Robinson, E. C., & Kainz, B. (2021). A survey on active learning and human-in-the-loop deep learning for medical image analysis. *Medical Image Analysis*, *71*, 102062. <https://doi.org/10.1016/j.media.2021.102062>
- Bulat, A., Yang, J., & Tzimiropoulos, G. (2018, September). To learn image super-resolution, use a gan to learn how to do image degradation first. In V. Ferrari, M. Hebert, C. Sminchisescu, & Y. Weiss

(Eds.), *Proceedings of the european conference on computer vision (eccv)* (pp. 185–200). https://openaccess.thecvf.com/content_ECCV_2018/html/Adrian_Bulat_To_learn_image_ECCV_2018_paper.html

- Burris, H. A. 3., Moore, M. J., Andersen, J., Green, M. R., Rothenberg, M. L., Modiano, M. R., Christine Cripps, M., Portenoy, R. K., Storniolo, A. M., Tarassoff, P., et al. (1997). Improvements in survival and clinical benefit with gemcitabine as first-line therapy for patients with advanced pancreas cancer: A randomized trial. *Journal of clinical oncology*, 15(6), 2403–2413.
- Burton, A. (2007). RECIST: right time to renovate? *The Lancet Oncology*, 6(8), 464–465.
- Buzdar, A. U., & Hortobagyi, G. (1998). Update on endocrine therapy for breast cancer. *Clinical Cancer Research*, 4(3), 527–534.
- Cai, C. J., Winter, S., Steiner, D., Wilcox, L., & Terry, M. (2019). "Hello AI": Uncovering the Onboarding Needs of Medical Practitioners for Human-AI Collaborative Decision-Making. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW), 1–24. <https://doi.org/10.1145/3359206>
- Cai, X., Liu, S., Han, J., Yang, L., Liu, Z., & Liu, T. (2021). ChestXRyBERT: A pretrained language model for chest radiology report summarization. *IEEE Transactions on Multimedia*, 25, 845–855. <https://doi.org/10.1109/tmm.2021.3132724>
- Campayo, M., Navarro, A., Viñolas, N., Tejero, R., Muñoz, C., Diaz, T., Marrades, R., Cabanas, M. L., Gimferrer, J. M., Gascon, P., Ramirez, J., & Monzo, M. (2011). A Dual Role for KRT81: A miR-SNP Associated with Recurrence in Non-Small-Cell Lung Cancer and a Novel Marker of Squamous Cell Lung Carcinoma. *PLOS ONE*, 6(7), 1–9. <https://doi.org/10.1371/journal.pone.0022509>
- Campbell, I. G., Russell, S. E., Choong, D. Y. H., Montgomery, K. G., Ciavarella, M. L., Hooi, C. S. F., Cristiano, B. E., Pearson, R. B., & Phillips, W. A. (2004). Mutation of the PIK3CAGene in Ovarian and Breast Cancer. *Cancer Research*, 64(21), 7678–7681. <https://doi.org/10.1158/0008-5472.can-04-2933>
- Candelaria, R. P., Hwang, L., Bouchard, R. R., & Whitman, G. J. (2013). Breast ultrasound: Current concepts [Small Parts Ultrasound]. *Seminars in Ultrasound, CT and MRI*, 34(3), 213–225. <https://doi.org/10.1053/j.sult.2012.11.013>
- Cao, X., Hou, J., An, Q., Assaraf, Y. G., & Wang, X. (2020). Towards the overcoming of anticancer drug resistance mediated by p53 mutations. *Drug Resistance Updates*, 49, 100671. <https://doi.org/10.1016/j.drug.2019.100671>
- Carlini, N., Hayes, J., Nasr, M., Jagielski, M., Sehwag, V., Tramèr, F., Balle, B., Ippolito, D., & Wallace, E. (2023). Extracting training data from diffusion models. *32nd USENIX Security Symposium (USENIX Security 23)*, 5253–5270. <https://www.usenix.org/conference/usenixsecurity23/presentation/carlini>
- Castiglioni, I., Rundo, L., Codari, M., Di Leo, G., Salvatore, C., Interlenghi, M., Gallivanone, F., Cozzi, A., D'Amico, N. C., & Sardanelli, F. (2021). Ai applications to medical images: From machine learning to deep learning. *Physica Medica*, 83, 9–24. <https://doi.org/10.1016/j.ejmp.2021.02.006>
- Celard, P., Iglesias, E. L., Sorribes-Fdez, J. M., Romero, R., Vieira, A. S., & Borrajo, L. (2022). A survey on deep learning applied to medical images: From simple artificial neural networks to generative

- models. *Neural Computing and Applications*, 35(3), 2291–2323. <https://doi.org/10.1007/s00521-022-07953-4>
- Cetin, I., Stephens, M., Camara, O., & González Ballester, M. A. (2023). Attri-vae: Attribute-based interpretable representations of medical images with variational autoencoders. *Computerized Medical Imaging and Graphics*, 104, 102158. <https://doi.org/10.1016/j.compmedimag.2022.102158>
- Chabner, B. A., & Roberts, T. G. (2005). Chemotherapy and the war on cancer. *Nature Reviews Cancer*, 5(1), 65–72. <https://doi.org/10.1038/nrc1529>
- Chai, C., Wu, H. H., Abuetabh, Y., Sergi, C., & Leng, R. (2022). Regulation of the tumor suppressor PTEN in triple-negative breast cancer. *Cancer Letters*, 527, 41–48. <https://doi.org/10.1016/j.canlet.2021.12.003>
- Chan, B., Schweter, S., & Möller, T. (2020). German's next language model. <https://doi.org/10.48550/ARXIV.2010.10906>
- Chan, V., & Perlas, A. (2011). Basics of ultrasound imaging. In S. N. Narouze (Ed.), *Atlas of ultrasound-guided procedures in interventional pain management* (pp. 13–19). Springer New York. https://doi.org/10.1007/978-1-4419-1681-5_2
- Chargari, C., Deutsch, E., Blanchard, P., Gouy, S., Martelli, H., Guérin, F., Dumas, I., Bossi, A., Morice, P., Viswanathan, A. N., & Haie-Meder, C. (2019). Brachytherapy: An overview for clinicians. *CA: A Cancer Journal for Clinicians*, 69(5), 386–401. <https://doi.org/10.3322/caac.21578>
- Chartrand, G., Cheng, P. M., Vorontsov, E., Drozdal, M., Turcotte, S., Pal, C. J., Kadoury, S., & Tang, A. (2017). Deep Learning: A Primer for Radiologists. *RadioGraphics*, 37(7), 2113–2131. <https://doi.org/10.1148/rg.2017170077>
- Chartsias, A., Joyce, T., Dharmakumar, R., & Tsiftaris, S. A. (2017, September). Adversarial image synthesis for unpaired multi-modal cardiac data. In S. A. Tsiftaris, A. Gooya, A. F. Frangi, & J. L. Prince (Eds.), *Simulation and synthesis in medical imaging* (pp. 3–13, Vol. 10557). Springer International Publishing. https://doi.org/10.1007/978-3-319-68127-6_1
- Chauffert, B., Mornex, F., Bonnetain, F., Rougier, P., Mariette, C., Bouché, O., Bosset, J. F., Aparicio, T., Mineur, L., Azzedine, A., Hammel, P., Butel, J., Stremsdoerfer, N., Maingon, P., & Bedenne, L. (2008). Phase III trial comparing intensive induction chemoradiotherapy (60 Gy, infusional 5-FU and intermittent cisplatin) followed by maintenance gemcitabine with gemcitabine alone for locally advanced unresectable pancreatic cancer. Definitive results of the 200001 FFCD/SFRO study. *Annals of Oncology*, 19(9), 1592–1599. <https://doi.org/10.1093/annonc/mdn281>
- Chen, A., Karwoski, R. A., Gierada, D. S., Bartholmai, B. J., & Koo, C. W. (2020a). Quantitative ct analysis of diffuse lung disease [PMID: 31782933]. *RadioGraphics*, 40(1), 28–43. <https://doi.org/10.1148/rg.2020190099>
- Chen, C., Qin, C., Qiu, H., Tarroni, G., Duan, J., Bai, W., & Rueckert, D. (2020b). Deep Learning for Cardiac Image Segmentation: A Review. *Frontiers in Cardiovascular Medicine*, 7. <https://doi.org/10.3389/fcvm.2020.00025>

- Chen, C.-H., Lin, P.-H., Hsieh, J.-G., Cheng, S.-L., & Jeng, J.-H. (2020c). Robust multi-class classification using linearly scored categorical cross-entropy. *2020 3rd IEEE International Conference on Knowledge Innovation and Invention (ICKII)*, 200–203. <https://doi.org/10.1109/ICKII50300.2020.9318835>
- Chen, L., Zhou, W.-B., Zhao, Y., Liu, X.-A., Ding, Q., Zha, X.-M., & Wang, S. (2011). Bloody nipple discharge is a predictor of breast cancer risk: A meta-analysis. *Breast Cancer Research and Treatment*, *132*(1), 9–14. <https://doi.org/10.1007/s10549-011-1787-5>
- Chen, X., Lian, C., Wang, L., Deng, H., Kuang, T., Fung, S., Gateno, J., Yap, P.-T., Xia, J. J., & Shen, D. (2021). Anatomy-regularized representation learning for cross-modality medical image segmentation. *IEEE Transactions on Medical Imaging*, *40*(1), 274–285. <https://doi.org/10.1109/tmi.2020.3025133>
- Chicco, D., & Jurman, G. (2020). The advantages of the matthews correlation coefficient (MCC) over f1 score and accuracy in binary classification evaluation. *BMC Genomics*, *21*(1). <https://doi.org/10.1186/s12864-019-6413-7>
- Chockley, K., & Emanuel, E. (2016). The end of radiology? three threats to the future practice of radiology. *Journal of the American College of Radiology*, *13*(12, Part A), 1415–1420. <https://doi.org/10.1016/j.jacr.2016.07.010>
- Ciliberto, D., Staropoli, N., Chiellino, S., Botta, C., Tassone, P., & Tagliaferri, P. (2016). Systematic review and meta-analysis on targeted therapy in advanced pancreatic cancer. *Pancreatology*, *16*(2), 249–258. <https://doi.org/10.1016/j.pan.2016.01.003>
- Claret, F., & Vu, T. (2012). Trastuzumab: Updated mechanisms of action and resistance in breast cancer. *Frontiers in Oncology*, *2*. <https://doi.org/10.3389/fonc.2012.00062>
- Clark, S. E., Warwick, J., Carpenter, R., Bowen, R. L., Duffy, S. W., & Jones, J. L. (2010). Molecular subtyping of DCIS: heterogeneity of breast cancer reflected in pre-invasive disease. *British Journal of Cancer*, *104*(1), 120–127. <https://doi.org/10.1038/sj.bjc.6606021>
- Cobb-Clark, D. A., & Schurer, S. (2012). The stability of big-five personality traits. *Economics Letters*, *115*(1), 11–15. <https://doi.org/10.1016/j.econlet.2011.11.015>
- Cohen, J. P., Luck, M., & Honari, S. (2018). Distribution matching losses can hallucinate features in medical image translation. In A. F. Frangi, J. A. Schnabel, C. Davatzikos, C. Alberola-López, & G. Fichtinger (Eds.), *Medical image computing and computer assisted intervention – miccai 2018* (pp. 529–536). Springer International Publishing. https://doi.org/10.1007/978-3-030-00928-1_60
- Cole, E. B., Zhang, Z., Marques, H. S., Hendrick, R. E., Yaffe, M. J., & Pisano, E. D. (2014). Impact of Computer-Aided Detection Systems on Radiologist Accuracy With Digital Mammography. *American Journal of Roentgenology*, *203*(4), 909–916. <https://doi.org/10.2214/ajr.12.10187>
- Collisson, E. A., Sadanandam, A., Olson, P., Gibb, W. J., Truitt, M., Gu, S., Cooc, J., Weinkle, J., Kim, G. E., Jakkula, L., Feiler, H. S., Ko, A. H., Olshen, A. B., Danenberg, K. L., Tempero, M. A., Spellman, P. T., Hanahan, D., & Gray, J. W. (2011). Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy. *Nature Medicine*, *17*(4), 500–503. <https://doi.org/10.1038/nm.2344>
- Comelli, A., Coronello, C., Dahiya, N., Benfante, V., Palmucci, S., Basile, A., Vancheri, C., Russo, G., Yezzi, A., & Stefano, A. (2020). Lung segmentation on high-resolution computerized tomography

- images using deep learning: A preliminary step for radiomics studies. *Journal of Imaging*, 6(11). <https://doi.org/10.3390/jimaging6110125>
- Cong, L., Liu, Q., Zhang, R., Cui, M., Zhang, X., Gao, X., Guo, J., Dai, M., Zhang, T., Liao, Q., & Zhao, Y. (2018). Tumor size classification of the 8th edition of TNM staging system is superior to that of the 7th edition in predicting the survival outcome of pancreatic cancer patients after radical resection and adjuvant chemotherapy. *Scientific Reports*, 8(1). <https://doi.org/10.1038/s41598-018-28193-4>
- Conroy, T., Hammel, P., Hebbar, M., Abdelghani, M. B., Wei, A. C.-c., Raoul, J.-L., Chone, L., Francois, E., Artru, P., Biagi, J. J., Lecomte, T., Assenat, E., Faroux, R., Ychou, M., Volet, J., Sauvanet, A., Jouffroy-Zeller, C., Patrick, R. A. T., Castan, F., & Bachet, J.-B. (2018). Unicancer GI PRODIGE 24/CCTG PA.6 trial: A multicenter international randomized phase III trial of adjuvant mFOLFIRINOX versus gemcitabine (gem) in patients with resected pancreatic ductal adenocarcinomas. *Journal of Clinical Oncology*, 36(18_suppl), LBA4001–LBA4001. https://doi.org/10.1200/jco.2018.36.18_suppl.lba4001
- Cortazar, P., Zhang, L., Untch, M., Mehta, K., Costantino, J. P., Wolmark, N., Bonnefoi, H., Cameron, D., Gianni, L., Valagussa, P., Swain, S. M., Prowell, T., Loibl, S., Wickerham, D. L., Bogaerts, J., Baselga, J., Perou, C., Blumenthal, G., Blohmer, J., . . . von Minckwitz, G. (2014). Pathological complete response and long-term clinical benefit in breast cancer: The CTNeoBC pooled analysis. *The Lancet*, 384(9938), 164–172. [https://doi.org/10.1016/s0140-6736\(13\)62422-8](https://doi.org/10.1016/s0140-6736(13)62422-8)
- Coudray, N., Ocampo, P. S., Sakellaropoulos, T., Narula, N., Snuderl, M., Fenyő, D., Moreira, A. L., Razavian, N., & Tsirigos, A. (2018). Classification and mutation prediction from nonsmall cell lung cancer histopathology images using deep learning. *Nature Medicine*, 24(10), 1559–1567. <https://doi.org/10.1038/s41591-018-0177-5>
- Cramer, A., Hecla, J., Wu, D., Lai, X., Boers, T., Yang, K., Moulton, T., Kenyon, S., Arzoumanian, Z., Krull, W., Gendreau, K., & Gupta, R. (2018). Stationary Computed Tomography for Space and other Resource-constrained Environments. *Scientific Reports*, 8(1). <https://doi.org/10.1038/s41598-018-32505-z>
- Cui, Y., Li, Y., Xing, D., Bai, T., Dong, J., & Zhu, J. (2021). Improving the Prediction of Benign or Malignant Breast Masses Using a Combination of Image Biomarkers and Clinical Parameters. *Frontiers Media SA*, 11. <https://doi.org/10.3389/fonc.2021.629321>
- Dahabreh, I. J., Linardou, H., Siannis, F., Fountzilas, G., & Murray, S. (2008). Trastuzumab in the Adjuvant Treatment of Early-Stage Breast Cancer: A Systematic Review and Meta-Analysis of Randomized Controlled Trials. *The Oncologist*, 13(6), 620–630. <https://doi.org/10.1634/theoncologist.2008-0001>
- Dar, S. U. H., Yurt, M., Karacan, L., Erdem, A., Erdem, E., & Cukur, T. (2019). Image synthesis in multi-contrast MRI with conditional generative adversarial networks. *IEEE Transactions on Medical Imaging*, 38(10), 2375–2388. <https://doi.org/10.1109/tmi.2019.2901750>
- de Rosa, G. H., & Papa, J. P. (2021). A survey on text generation using generative adversarial networks. *Pattern Recognition*, 119, 108098. <https://doi.org/10.1016/j.patcog.2021.108098>
- de Geus, S. W. L., Eskander, M. F., Kasumova, G. G., Ng, S. C., Kent, T. S., Mancias, J. D., Callery, M. P., Mahadevan, A., & Tseng, J. F. (2017). Stereotactic body radiotherapy for unresected pancreatic cancer: A nationwide review. *Cancer*, 123(21), 4158–4167. <https://doi.org/10.1002/cncr.30856>

- Del Corro, L., Del Giorno, A., Agarwal, S., Yu, B., Awadallah, A., & Mukherjee, S. (2023). Skipdecode: Autoregressive skip decoding with batching and caching for efficient llm inference. <https://doi.org/10.48550/ARXIV.2307.02628>
- Department of Health and Human Services (HHS). (2002). Standards for privacy of individually identifiable health information. final rule. [KIE: KIE Bib: confidentiality/legal aspects]. *Federal register*. 2000;65(250):82462-8282911503738, 67, 53181–273. <https://www.hhs.gov/hipaa/for-professionals/index.html>
- Deutsche Krebsgesellschaft. (2013, October). S3-Leitlinie zum exokrinen Pankreaskarzinom [Retrieved from: <https://www.leitlinienprogramm-onkologie.de/leitlinien/pankreaskarzinom/>]. <https://www.leitlinienprogramm-onkologie.de/leitlinien/pankreaskarzinom/>
- Deutsche Krebsgesellschaft. (2021a, June). Interdisziplinäre S3-Leitlinie für die Früherkennung, Diagnostik, Therapie und Nachsorge des Mammakarzinoms [Retrieved from: <https://www.leitlinienprogramm-onkologie.de/leitlinien/mammakarzinom/>]. <https://www.leitlinienprogramm-onkologie.de/leitlinien/mammakarzinom/>
- Deutsche Krebsgesellschaft. (2021b, December). S3-Leitlinie zum exokrinen Pankreaskarzinom [Retrieved from <https://www.leitlinienprogramm-onkologie.de/leitlinien/pankreaskarzinom/>]. <https://www.leitlinienprogramm-onkologie.de/leitlinien/pankreaskarzinom/>
- Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 144(1), 114–126. <https://doi.org/10.1037/xge0000033>
- Dima, A., Paetzold, J. C., Jungmann, F., Lemke, T., Raffler, P., Kaissis, G., Rueckert, D., & Braren, R. (2021, September). Segmentation of Peripancreatic Arteries in Multispectral Computed Tomography Imaging. In C. Lian, X. Cao, I. Rekik, X. Xu, & P. Yan (Eds.), *Machine Learning in Medical Imaging. MLMI 2021. Lecture Notes in Computer Science* (pp. 596–605, Vol. 12966). Springer International Publishing. https://doi.org/10.1007/978-3-030-87589-3_61
- Dinh, L., Sohl-Dickstein, J., & Bengio, S. (2016). Density estimation using real nvp. <https://doi.org/10.48550/ARXIV.1605.08803>
- Distler, M., Pilarsky, E., Kersting, S., & Grützmänn, R. (2013). Preoperative CEA and CA 19-9 are prognostic markers for survival after curative resection for ductal adenocarcinoma of the pancreas A retrospective tumor marker prognostic study. *International Journal of Surgery*, 11(10), 1067–1072. <https://doi.org/10.1016/j.ijssu.2013.10.005>
- Doersch, C. (2016). Tutorial on variational autoencoders. <https://doi.org/10.48550/ARXIV.1606.05908>
- Doi, K. (2007). Computer-aided diagnosis in medical imaging: Historical review, current status and future potential. *Computerized Medical Imaging and Graphics*, 31(4-5), 198–211. <https://doi.org/10.1016/j.compmedimag.2007.02.002>
- Dou, Q., So, T. Y., Jiang, M., Liu, Q., Vardhanabhuti, V., Kaissis, G., Li, Z., Si, W., Lee, H. H. C., Yu, K., Feng, Z., Dong, L., Burian, E., Jungmann, F., Braren, R., Makowski, M., Kainz, B., Rueckert, D., Glocker, B., ... Heng, P. A. (2021). Federated deep learning for detecting COVID-19 lung

- abnormalities in CT: A privacy-preserving multinational validation study. *npj Digital Medicine*, 4(1). <https://doi.org/10.1038/s41746-021-00431-6>
- Dozat, T. (2016, May). Incorporating Nesterov Momentum into Adam. In Y. Bengio & Y. LeCun (Eds.), *Proceedings of 4th international conference on learning representations (iclr), workshop track*. <https://openreview.net/pdf?id=OM0jvwB8jIp57ZJjtNEZ>
- Duric, N., Littrup, P., Li, C., Roy, O., Schmidt, S., Janer, R., Cheng, X., Goll, J., Rama, O., Bey-Knight, L., & Greenway, W. (2012). Breast ultrasound tomography: bridging the gap to clinical practice. In J. G. Bosch & M. M. Doyley (Eds.), *Medical imaging 2012: Ultrasonic imaging, tomography, and therapy* (83200O, Vol. 8320). SPIE. <https://doi.org/10.1117/12.910988>
- Dutt, S., Ahmed, M. M., Loo, B. W., & Strober, S. (2020). Novel radiation therapy paradigms and immunomodulation: Heresies and hope [Trials and Tribulations of Radio-Immuno-Oncology]. *Seminars in Radiation Oncology*, 30(2), 194–200. <https://doi.org/10.1016/j.semradonc.2019.12.006>
- Early Breast Cancer Trialists' Collaborative Group (EBCTCG). (1998). Polychemotherapy for early breast cancer: An overview of the randomised trials. *The Lancet*, 352(9132), 930–942. [https://doi.org/10.1016/S0140-6736\(98\)03301-7](https://doi.org/10.1016/S0140-6736(98)03301-7)
- Early Breast Cancer Trialists' Collaborative Group (EBCTCG). (2005). Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: An overview of the randomised trials. *The Lancet*, 365(9472), 1687–1717. [https://doi.org/10.1016/S0140-6736\(05\)66544-0](https://doi.org/10.1016/S0140-6736(05)66544-0)
- Early Breast Cancer Trialists' Collaborative Group (EBCTCG). (2011a). Effect of radiotherapy after breast-conserving surgery on 10-year recurrence and 15-year breast cancer death: Meta-analysis of individual patient data for 10 801 women in 17 randomised trials. *The Lancet*, 378(9804), 1707–1716. [https://doi.org/10.1016/S0140-6736\(11\)61629-2](https://doi.org/10.1016/S0140-6736(11)61629-2)
- Early Breast Cancer Trialists' Collaborative Group (EBCTCG). (2011b). Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: Patient-level meta-analysis of randomised trials. *The Lancet*, 378(9793), 771–784. [https://doi.org/10.1016/S0140-6736\(11\)60993-8](https://doi.org/10.1016/S0140-6736(11)60993-8)
- Ebers, M. (2021). Standardizing AI - the case of the european commission's proposal for an artificial intelligence act. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3900378>
- Emura, T., Matsui, S., & Chen, H.-Y. (2019). Compound.cox: Univariate feature selection and compound covariate for predicting survival. *Computer Methods and Programs in Biomedicine*, 168, 21–37. <https://doi.org/10.1016/j.cmpb.2018.10.020>
- Enge, J., & Gassoden, G. (2020, March). *Big Five Personality Test* [last accessed on 01.08.2022]. <https://bigfive-test.com>
- Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), 115–118. <https://doi.org/10.1038/nature21056>
- Ettrich, T. J., Uhl, W., Kornmann, M., Algül, H., Friess, H., Koenig, A., Gallmeier, E., Lutz, M. P., Wille, K., Schimanski, C. C., Kunzmann, V., Geissler, M., Waldschmidt, D., Daum, S., Blome, L., Tannapfel, A., Perkhofer, L., Tempero, M. A., Reinacher-Schick, A. C., & Seufferlein, T. (2022). Perioperative or

adjuvant nab-paclitaxel plus gemcitabine for resectable pancreatic cancer: Updated final results of the randomized phase II AIO-NEONAX trial. *Journal of Clinical Oncology*, 40(16_suppl), 4133–4133. https://doi.org/10.1200/jco.2022.40.16_suppl.4133

- European Commission. (2021, April 21). COM(2021) 206 final. Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52021PC0206>
- Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D. M., Forman, D., & Bray, F. (2014). Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *International Journal of Cancer*, 136(5), E359–E386. <https://doi.org/10.1002/ijc.29210>
- Fernando, T., Gammulle, H., Denman, S., Sridharan, S., & Fookes, C. (2021). Deep learning for medical anomaly detection - a survey. *ACM Comput. Surv.*, 54(7). <https://doi.org/10.1145/3464423>
- Fiorini, C., Cordani, M., Padroni, C., Blandino, G., Di Agostino, S., & Donadelli, M. (2015). Mutant p53 stimulates chemoresistance of pancreatic adenocarcinoma cells to gemcitabine. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1853(1), 89–100. <https://doi.org/10.1016/j.bbamcr.2014.10.003>
- Fitzgerald, T. L., Hickner, Z. J., Schmitz, M., & Kort, E. J. (2008). Changing Incidence of Pancreatic Neoplasms. *Pancreas*, 37(2), 134–138. <https://doi.org/10.1097/mpa.0b013e318163a329>
- Flohr, T., Petersilka, M., Henning, A., Ulzheimer, S., Ferda, J., & Schmidt, B. (2020). Photon-counting CT review [125 Years of X-Rays]. *Physica Medica*, 79, 126–136. <https://doi.org/10.1016/j.ejmp.2020.10.030>
- Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines*, 30(4), 681–694. <https://doi.org/10.1007/s11023-020-09548-1>
- Food and Drug Administration (United States). (2019, April). Proposed regulatory framework for modifications to artificial intelligence/machine learning (ai/ml)-based software as a medical device (samd). <https://apo.org.au/node/228371>
- Foreman, S. C., Neumann, J., Han, J., Harrasser, N., Weiss, K., Peeters, J. M., Karampinos, D. C., Makowski, M. R., Gersing, A. S., & Woertler, K. (2022). Deep learningbased acceleration of compressed sense MR imaging of the ankle. *European Radiology*, 32(12), 8376–8385. <https://doi.org/10.1007/s00330-022-08919-9>
- Foulkes, W. D., Smith, I. E., & Reis-Filho, J. S. (2010). Triple-Negative Breast Cancer [PMID: 21067385]. *New England Journal of Medicine*, 363(20), 1938–1948. <https://doi.org/10.1056/NEJMr1001389>
- Foy, J. J., Robinson, K. R., Li, H., Giger, M. L., Al-Hallaq, H., & Armato, S. G. (2018). Variation in algorithm implementation across radiomics software. *Journal of Medical Imaging*, 5(4), 044505. <https://doi.org/10.1117/1.JMI.5.4.044505>
- Frid-Adar, M., Klang, E., Amitai, M., Goldberger, J., & Greenspan, H. (2018). Synthetic data augmentation using gan for improved liver lesion classification. *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, 289–293. <https://doi.org/10.1109/ISBI.2018.8363576>

- Frolov, S., Hinz, T., Raue, F., Hees, J., & Dengel, A. (2021). Adversarial text-to-image synthesis: A review. *Neural Networks*, *144*, 187–209. <https://doi.org/10.1016/j.neunet.2021.07.019>
- Fukukita, H., Suzuki, K., Matsumoto, K., Terauchi, T., Daisaki, H., Ikari, Y., Shimada, N., & Senda, M. (2014). Japanese guideline for the oncology FDG-PET/CT data acquisition protocol: Synopsis of version 2.0. *Annals of Nuclear Medicine*, *28*(7), 693–705. <https://doi.org/10.1007/s12149-014-0849-2>
- Gabriel, I. (2020). Artificial intelligence, values, and alignment. *Minds and Machines*, *30*(3), 411–437. <https://doi.org/10.1007/s11023-020-09539-2>
- Gabriel, I., & Ghazavi, V. (2021). The challenge of value alignment: From fairer algorithms to ai safety. <https://doi.org/10.48550/ARXIV.2101.06060>
- Gao, J. J., & Swain, S. M. (2018). Luminal A Breast Cancer and Molecular Assays: A Review. *The Oncologist*, *23*(5), 556–565. <https://doi.org/10.1634/theoncologist.2017-0535>
- Gardner, M. W., & Dorling, S. R. (1998). Artificial neural networks (the multilayer perceptron) - a review of applications in the atmospheric sciences. *Atmospheric Environment*, *32*(14), 2627–2636. [https://doi.org/10.1016/S1352-2310\(97\)00447-0](https://doi.org/10.1016/S1352-2310(97)00447-0)
- Gemeinsamer Bundesausschuss. (2017, July). Richtlinie über die Früherkennung von Krebserkrankungen. https://www.g-ba.de/downloads/62-492-1461/KFE-RL_2017-07-20_iK-2017-11-08.pdf
- General Data Protection Regulation (GDPR). (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). *Official Journal of the European Union L 119/1*. <https://gdpr-info.eu>
- Ghafoorian, M., Karssemeijer, N., Heskes, T., van Uden, I. W. M., Sanchez, C. I., Litjens, G., de Leeuw, F.-E., van Ginneken, B., Marchiori, E., & Platel, B. (2017). Location Sensitive Deep Convolutional Neural Networks for Segmentation of White Matter Hyperintensities. *Scientific Reports*, *7*(1). <https://doi.org/10.1038/s41598-017-05300-5>
- Ghalebikesabi, S., Berrada, L., Goyal, S., Ktena, I., Stanforth, R., Hayes, J., De, S., Smith, S. L., Wiles, O., & Balle, B. (2023, February). Differentially private diffusion models generate useful synthetic images. <https://doi.org/10.48550/ARXIV.2302.13861>
- Gillies, R. J., Kinahan, P. E., & Hricak, H. (2016). Radiomics: Images Are More than Pictures, They Are Data. *Radiology*, *278*(2), 563–577. <https://doi.org/10.1148/radiol.2015151169>
- Ginsburg, O., Bray, F., Coleman, M. P., Vanderpuye, V., Eniu, A., Kotha, S. R., Sarker, M., Huong, T. T., Allemani, C., Dvaladze, A., Gralow, J., Yeates, K., Taylor, C., Oomman, N., Krishnan, S., Sullivan, R., Kombe, D., Blas, M. M., Parham, G., . . . Conteh, L. (2017). The global burden of women's cancers: A grand challenge in global health. *The Lancet*, *389*(10071), 847–860. [https://doi.org/10.1016/S0140-6736\(16\)31392-7](https://doi.org/10.1016/S0140-6736(16)31392-7)
- Girin, L., Leglaive, S., Bie, X., Diard, J., Hueber, T., & Alameda-Pineda, X. (2021). Dynamical variational autoencoders: A comprehensive review. *Foundations and Trends in Machine Learning*, *15*(1-2), 1–175. <https://doi.org/10.1561/22000000089>

- Giuseppetti, G. M., Boria, F., & Baldassarre, S. (2018, October). DCIS Imaging. In C. Mariotti (Ed.), *Ductal carcinoma in situ of the breast* (pp. 39–56). Springer International Publishing. https://doi.org/10.1007/978-3-319-57451-6_3
- Glorot, X., Bordes, A., & Bengio, Y. (2011, December). Deep Sparse Rectifier Neural Networks. In G. Gordon, D. Dunson, & M. Dudík (Eds.), *Proceedings of the fourteenth international conference on artificial intelligence and statistics* (pp. 315–323, Vol. 15). PMLR. <http://proceedings.mlr.press/v15/glorot11a/glorot11a.pdf>
- Golan, T., Hammel, P., Reni, M., Cutsem, E. V., Macarulla, T., Hall, M. J., Park, J.-O., Hochhauser, D., Arnold, D., Oh, D.-Y., Reinacher-Schick, A., Tortora, G., Algül, H., O'Reilly, E. M., McGuinness, D., Cui, K. Y., Schlienger, K., Locker, G. Y., & Kindler, H. L. (2019). Maintenance Olaparib for Germline BRCA-Mutated 10.1200/JCO.2014.56.2728Metastatic Pancreatic Cancer. *New England Journal of Medicine*, *381*(4), 317–327. <https://doi.org/10.1056/nejmoa1903387>
- Gold, R. H., Bassett, L. W., & Widoff, B. E. (1990). Highlights from the history of mammography. *Radiographics*, *10*(6), 1111–1131. <https://doi.org/10.1148/radiographics.10.6.2259767>
- Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment*, *4*(1), 26–42. <https://doi.org/10.1037/1040-3590.4.1.26>
- Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, *40*(1), 84–96. <https://doi.org/10.1016/j.jrp.2005.08.007>
- Goldhirsch, A., Winer, E. P., Coates, A. S., Gelber, R. D., Piccart-Gebhart, M., Thürlimann, B., Senn, H.-J., Albain, K. S., André, F., Bergh, J., Bonnefoi, H., Bretel-Morales, D., Burstein, H., Cardoso, F., Castiglione-Gertsch, M., Coates, A. S., Colleoni, M., Costa, A., Curigliano, G., . . . Wood, W. C. (2013). Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. *Annals of Oncology*, *24*(9), 2206–2223. <https://doi.org/10.1093/annonc/mdt303>
- Goldhirsch, A., Wood, W. C., Coates, A. S., Gelber, R. D., Thürlimann, B., & Senn, H.-J. (2011). Strategies for subtypes-dealing with the diversity of breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011 [In this issue: EU recognition for Medical Oncology Focus on translational research]. *Annals of Oncology*, *22*(8), 1736–1747. <https://doi.org/10.1093/annonc/mdr304>
- Goldman, L. W. (2007). Principles of CT and CT Technology. *Journal of Nuclear Medicine Technology*, *35*(3), 115–128. <https://doi.org/10.2967/jnmt.107.042978>
- Goldman, L. W. (2008). Principles of CT: Multislice CT. *Journal of Nuclear Medicine Technology*, *36*(2), 57–68. <https://doi.org/10.2967/jnmt.107.044826>
- Goldstein, D., El-Maraghi, R. H., Hammel, P., Heinemann, V., Kunzmann, V., Sastre, J., Scheithauer, W., Siena, S., Tabernero, J., Teixeira, L., Tortora, G., Laethem, J.-L. V., Young, R., Penenberg, D. N., Lu, B., Romano, A., & Hoff, D. D. V. (2015). nab-Paclitaxel Plus Gemcitabine for Metastatic Pancreatic Cancer: Long-Term Survival From a Phase III Trial. *JNCI Journal of the National Cancer Institute*, *107*(2), dju413–dju413. <https://doi.org/10.1093/jnci/dju413>

- Gondala, S., Verwimp, L., Pusateri, E., Tsagkias, M., & Gysel, C. V. (2021). Error-driven pruning of language models for virtual assistants. *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 7413–7417. <https://doi.org/10.1109/icassp39728.2021.9415035>
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning* (Vol. 1). MIT press. https://www.academia.edu/download/62266271/Deep_Learning20200303-80130-1s42zvt.pdf
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems* (Vol. 27). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf
- Goutte, C., & Gaussier, E. (2005). A probabilistic interpretation of precision, recall and f-score, with implication for evaluation. In D. E. Losada & J. M. Fernández-Luna (Eds.), *Advances in information retrieval* (pp. 345–359). Springer Berlin Heidelberg. https://link.springer.com/chapter/10.1007/978-3-540-31865-1_25
- Gozalo-Brizuela, R., & Garrido-Merchan, E. C. (2023). Chatgpt is not all you need. a state of the art review of large generative ai models. <https://doi.org/10.48550/ARXIV.2301.04655>
- Grabenstetter, A., Mohanty, A. S., Rana, S., Zehir, A., Brannon, A. R., D'Alfonso, T. M., DeLair, D. F., Tan, L. K., & Ross, D. S. (2020). E-cadherin immunohistochemical expression in invasive lobular carcinoma of the breast: correlation with morphology and CDH1 somatic alterations. *Human Pathology*, 102, 44–53. <https://doi.org/10.1016/j.humpath.2020.06.002>
- Greenspan, H. (2008). Super-resolution in medical imaging. *The Computer Journal*, 52(1), 43–63. <https://doi.org/10.1093/comjnl/bxm075>
- Greenspan, H., Oz, G., Kiryati, N., & Peled, S. (2002). Mri inter-slice reconstruction using super-resolution. *Magnetic Resonance Imaging*, 20(5), 437–446. [https://doi.org/10.1016/S0730-725X\(02\)00511-8](https://doi.org/10.1016/S0730-725X(02)00511-8)
- Greenspan, H., van Ginneken, B., & Summers, R. M. (2016). Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique. *IEEE Transactions on Medical Imaging*, 35(5), 1153–1159. <https://doi.org/10.1109/TMI.2016.2553401>
- Guibas, J. T., Virdi, T. S., & Li, P. S. (2017). Synthetic medical images from dual generative adversarial networks. <https://doi.org/10.48550/ARXIV.1709.01872>
- Guimaraes, A. R. (2021). Chapter 3: Quantitative Imaging Biomarker Alliance (QIBA): Protocols and Profiles. In R. J. Nordstrom (Ed.), *Quantitative Imaging in Medicine: Background and Basics* (pp. 3-1-3–22). AIP Publishing LLC. https://doi.org/10.1063/9780735423473_003
- Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J., Kim, R., Raman, R., Nelson, P. C., Mega, J. L., & Webster, D. R. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*, 316(22), 2402. <https://doi.org/10.1001/jama.2016.17216>
- Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., & Yang, G.-Z. (2019). Xai-explainable artificial intelligence. *Science Robotics*, 4(37), eaay7120. <https://doi.org/10.1126/scirobotics.aay7120>

- Gur, D., Sumkin, J. H., Rockette, H. E., Ganott, M., Hakim, C., Hardesty, L., Poller, W. R., Shah, R., & Wallace, L. (2004). Changes in Breast Cancer Detection and Mammography Recall Rates After the Introduction of a Computer-Aided Detection System. *JNCI: Journal of the National Cancer Institute*, *96*(3), 185–190. <https://doi.org/10.1093/jnci/djh067>
- Gymnopoulos, M., Elsliger, M.-A., & Vogt, P. K. (2007). Rare cancer-specific mutations in PIK3CA show gain of function. *Proceedings of the National Academy of Sciences*, *104*(13), 5569–5574. <https://doi.org/10.1073/pnas.0701005104>
- Hackert, T., Werner, J., & Büchler, M. W. (2012). Pankreas. In J. R. Siewert & H. J. Stein (Eds.), *Chirurgie. Springer-Lehrbuch* (9th, pp. 775–788). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-11331-4_7
- Hadsell, R., Sermanet, P., Ben, J., Erkan, A., Scoffier, M., Kavukcuoglu, K., Muller, U., & LeCun, Y. (2009). Learning long-range vision for autonomous off-road driving. *Journal of Field Robotics*, *26*(2), 120–144. <https://doi.org/10.1002/rob.20276>
- Hady, M. F. A., & Schwenker, F. (2013, January). Semi-supervised learning. In M. Bianchini, M. Maggini, & L. C. Jain (Eds.), *Handbook on neural information processing* (pp. 215–239). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-36657-4_7
- Haenssle, H. A., Fink, C., Schneiderbauer, R., Toberer, F., Buhl, T., Blum, A., Kalloo, A., Hassen, A. B. H., Thomas, L., Enk, A., Uhlmann, L., Alt, C., Arenbergerova, M., Bakos, R., Baltzer, A., Bertlich, I., Blum, A., Bokor-Billmann, T., Bowling, J., . . . Zalaudek, I. (2018). Man against machine: Diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Annals of Oncology*, *29*(8), 1836–1842. <https://doi.org/10.1093/annonc/mdy166>
- Hagiwara, A., Fujita, S., Ohno, Y., & Aoki, S. (2020). Variability and standardization of quantitative imaging: Monoparametric to multiparametric quantification, radiomics, and artificial intelligence. *Investigative radiology*, *55*, 601–616. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7413678/>
- Halpin, Z. T. (1989). Scientific objectivity and the concept of "the other" [Feminism and science: In memory of Ruth Bleier]. *Women's Studies International Forum*, *12*(3), 285–294. [https://doi.org/10.1016/S0277-5395\(89\)80006-8](https://doi.org/10.1016/S0277-5395(89)80006-8)
- Hammel, P., Huguet, F., van Laethem, J.-L., Goldstein, D., Glimelius, B., Artru, P., Borbath, I., Bouché, O., Shannon, J., André, T., Mineur, L., Chibaudel, B., Bonnetain, F., & Louvet, C. (2016). Effect of Chemoradiotherapy vs Chemotherapy on Survival in Patients With Locally Advanced Pancreatic Cancer Controlled After 4 Months of Gemcitabine With or Without Erlotinib. *JAMA*, *315*(17), 1844. <https://doi.org/10.1001/jama.2016.4324>
- Hammernik, K., Schlemper, J., Qin, C., Duan, J., Summers, R. M., & Rueckert, D. (2021). Systematic evaluation of iterative deep neural networks for fast parallel mri reconstruction with sensitivity-weighted coil combination. *Magnetic Resonance in Medicine*, *86*(4), 1859–1872. <https://doi.org/10.1002/mrm.28827>
- Han, C., Hayashi, H., Rundo, L., Araki, R., Shimoda, W., Muramatsu, S., Furukawa, Y., Mauri, G., & Nakayama, H. (2018). GAN-based synthetic brain MR image generation. *2018 IEEE 15th Interna-*

- tional Symposium on Biomedical Imaging (ISBI 2018)*, 734–738. <https://doi.org/10.1109/isbi.2018.8363678>
- Han, S. S., Park, I., Chang, S. E., Lim, W., Kim, M. S., Park, G. H., Chae, J. B., Huh, C. H., & Na, J.-I. (2020). Augmented Intelligence Dermatology: Deep Neural Networks Empower Medical Professionals in Diagnosing Skin Cancer and Predicting Treatment Options for 134 Skin Disorders. *Journal of Investigative Dermatology*. <https://doi.org/10.1016/j.jid.2020.01.019>
- Handels, H. (2009). *Medizinische Bildverarbeitung: Bildanalyse, Mustererkennung und Visualisierung für die computergestützte ärztliche Diagnostik und Therapie*. Springer-Verlag. <https://link.springer.com/book/10.1007/978-3-8348-9571-4>
- Harvey, H. B., & Gowda, V. (2020). How the fda regulates ai [Special Issue: Artificial Intelligence]. *Academic Radiology*, 27(1), 58–61. <https://doi.org/10.1016/j.acra.2019.09.017>
- Hawkins, D. M. (2003). The problem of overfitting. *Journal of Chemical Information and Computer Sciences*, 44(1), 1–12. <https://doi.org/10.1021/ci0342472>
- He, J., Baxter, S. L., Xu, J., Xu, J., Zhou, X., & Zhang, K. (2019a). The practical implementation of artificial intelligence technologies in medicine. *Nature Medicine*, 25(1), 30–36. <https://doi.org/10.1038/s41591-018-0307-0>
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778. https://openaccess.thecvf.com/content_cvpr_2016/papers/He_Deep_Residual_Learning_CVPR_2016_paper.pdf
- He, T., Zhang, Z., Zhang, H., Zhang, Z., Xie, J., & Li, M. (2019b). Bag of Tricks for Image Classification with Convolutional Neural Networks. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. https://openaccess.thecvf.com/content_CVPR_2019/papers/He_Bag_of_Tricks_for_Image_Classification_with_Convolutional_Neural_Networks_CVPR_2019_paper.pdf
- Hekler, A., Utikal, J. S., Enk, A. H., Hauschild, A., Weichenthal, M., Maron, R. C., Berking, C., Haferkamp, S., Klode, J., Schadendorf, D., Schilling, B., Holland-Letz, T., Izar, B., von Kalle, C., Fröhling, S., Brinker, T. J., Schmitt, L., Peitsch, W. K., Hoffmann, F., . . . Thiem, A. (2019). Superior skin cancer classification by the combination of human and artificial intelligence. *European Journal of Cancer*, 120, 114–121. <https://doi.org/10.1016/j.ejca.2019.07.019>
- Hester, R. H., Hortobagyi, G. N., & Lim, B. (2021). Inflammatory breast cancer: Early recognition and diagnosis is critical. *American Journal of Obstetrics and Gynecology*, 225(4), 392-3–96. <https://doi.org/10.1016/j.ajog.2021.04.217>
- Heywang-Köbrunner, S. H., Hacker, A., & Sedlacek, S. (2011). Advantages and disadvantages of mammography screening. *Breast Care*, 6(3), 2–2. <https://doi.org/10.1159/000329005>
- Ho, S., Qu, Y., Gu, B., Gao, L., Li, J., & Xiang, Y. (2021). Dp-gan: Differentially private consecutive data publishing using generative adversarial nets. *Journal of Network and Computer Applications*, 185, 103066. <https://doi.org/10.1016/j.jnca.2021.103066>

- Hochreiter, S. (1998). The vanishing gradient problem during learning recurrent neural nets and problem solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 06(02), 107–116. <https://doi.org/10.1142/S0218488598000094>
- Hoff, D. D. V., Ervin, T., Arena, F. P., Chiorean, E. G., Infante, J., Moore, M., Seay, T., Tjulandin, S. A., Ma, W. W., Saleh, M. N., Harris, M., Reni, M., Dowden, S., Laheru, D., Bahary, N., Ramanathan, R. K., Taberero, J., Hidalgo, M., Goldstein, D., . . . Renschler, M. F. (2013). Increased Survival in Pancreatic Cancer with nab-Paclitaxel plus Gemcitabine. *New England Journal of Medicine*, 369(18), 1691–1703. <https://doi.org/10.1056/nejmoa1304369>
- Hofmann, T. (2001). Unsupervised learning by probabilistic latent semantic analysis. *Machine learning*, 42(1-2), 177. <http://net.pku.edu.cn/~course/cs402/EM/hofmann.pdf>
- Holzinger, A. (2016). Interactive machine learning for health informatics: When do we need the human-in-the-loop? *Brain Informatics*, 3(2), 119–131. <https://doi.org/10.1007/s40708-016-0042-6>
- Hounsfield, G. N. (1980). Computed Medical Imaging. *Science*, 210(4465), 22–28. <https://doi.org/10.1126/science.6997993>
- Howlander, N., Noone, A. M., Krapcho, M., Miller, D., Bishop, K., Kosary, C. L., Yu, M., Ruhl, J., Tatalovich, Z., Mariotto, A., Lewis, D. R., Chen, H. S., Feuer, E. J., & Cronin, K. A. (2017, April). *SEER Cancer Statistics Review, 1975-2014*. [Last accessed on 19 June 2022.]. National Cancer Institute. https://seer.cancer.gov/csr/1975_2014/
- Hu, M., Pan, S., Li, Y., & Yang, X. (2023). Advancing medical imaging with language models: A journey from n-grams to chatgpt. <https://doi.org/10.48550/ARXIV.2304.04920>
- Huang, E. P., O'Connor, J. P. B., McShane, L. M., Giger, M. L., Lambin, P., Kinahan, P. E., Siegel, E. L., & Shankar, L. K. (2022). Criteria for the translation of radiomics into clinically useful tests. *Nature Reviews Clinical Oncology*, 20(2), 69–82. <https://doi.org/10.1038/s41571-022-00707-0>
- Huang, Y., Liu, Z., He, L., Chen, X., Pan, D., Ma, Z., Liang, C., Tian, J., & Liang, C. (2016). Radiomics Signature: A Potential Biomarker for the Prediction of Disease-Free Survival in Early-Stage (I or II) Non-Small Cell Lung Cancer. *Radiology*, 281(3), 947–957. <https://doi.org/10.1148/radiol.2016152234>
- Hüllermeier, E., & Waegeman, W. (2021). Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. *Machine Learning*, 110(3), 457–506. <https://doi.org/10.1007/s10994-021-05946-3>
- Hwang, T. (2018). Computational Power and the Social Impact of Artificial Intelligence. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3147971>
- Hyun, S. H., Ahn, M. S., Koh, Y. W., & Lee, S. J. (2019). A machine-learning approach using PET-based radiomics to predict the histological subtypes of lung cancer. *Clinical Nuclear Medicine*, 44(12), 956–960. <https://doi.org/10.1097/rlu.0000000000002810>
- Iacobuzio-Donahue, C. A., Fu, B., Yachida, S., Luo, M., Abe, H., Henderson, C. M., Vilardell, F., Wang, Z., Keller, J. W., Banerjee, P., Herman, J. M., Cameron, J. L., Yeo, C. J., Halushka, M. K., Eshleman, J. R., Raben, M., Klein, A. P., Hruban, R. H., Hidalgo, M., & Laheru, D. (2009). DPC4 Gene Status

- of the Primary Carcinoma Correlates With Patterns of Failure in Patients With Pancreatic Cancer. *Journal of Clinical Oncology*, 27(11), 1806–1813. <https://doi.org/10.1200/jco.2008.17.7188>
- Iglesias, J. E., Billot, B., Balbastre, Y., Tabari, A., Conklin, J., Gilberto González, R., Alexander, D. C., Golland, P., Edlow, B. L., & Fischl, B. (2021). Joint super-resolution and synthesis of 1.5mm isotropic mp-rage volumes from clinical mri exams with scans of different orientation, resolution and contrast. *NeuroImage*, 237, 118206. <https://doi.org/10.1016/j.neuroimage.2021.118206>
- Ioffe, S., & Szegedy, C. (2015, July). Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In F. Bach & D. Blei (Eds.), *Proceedings of the 32nd international conference on machine learning* (pp. 448–456, Vol. 37). PMLR. <https://proceedings.mlr.press/v37/loffel5.html>
- Iqbal, T., & Ali, H. (2018). Generative adversarial network for medical images (MI-GAN). *Journal of Medical Systems*, 42(11). <https://doi.org/10.1007/s10916-018-1072-9>
- Jabbar, H., & Khan, R. Z. (2015). Methods to avoid over-fitting and under-fitting in supervised machine learning (comparative study). *Computer Science, Communication and Instrumentation Devices*, 70, 163–172.
- Jafari, M., Wang, Y., Amiryousefi, A., & Tang, J. (2020). Unsupervised Learning and Multipartite Network Models: A Promising Approach for Understanding Traditional Medicine. *Frontiers in Pharmacology*, 11. <https://doi.org/10.3389/fphar.2020.01319>
- Jaffe, C. C. (2006). Measures of Response: RECIST, WHO, and New Alternatives. *Journal of Clinical Oncology*, 24(20), 3245–3251. <https://doi.org/10.1200/jco.2006.06.5599>
- Jahanshahi, A., Soleymani, Y., Ghaziani, M. F., & Khezerloo, D. (2023). Radiomics reproducibility challenge in computed tomography imaging as a nuisance to clinical generalization: A mini-review. *Egyptian Journal of Radiology and Nuclear Medicine*, 54(1). <https://doi.org/10.1186/s43055-023-01029-6>
- Jancik, S., Drabek, J., Radzioch, D., & Hajduch, M. (2010). Clinical Relevance of KRAS in Human Cancers. *Journal of Biomedicine and Biotechnology*, 2010, 1–13. <https://doi.org/10.1155/2010/150960>
- Javaid, M., Haleem, A., & Singh, R. P. (2023). Chatgpt for healthcare services: An emerging stage for an innovative perspective. *BenchCouncil Transactions on Benchmarks, Standards and Evaluations*, 3(1), 100105. <https://doi.org/10.1016/j.tbench.2023.100105>
- Jeblick, K., Schachtner, B., Dextl, J., Mittermeier, A., Stüber, A. T., Topalis, J., Weber, T., Wesp, P., Sabel, B., Ricke, J., & Ingrisich, M. (2022). Chatgpt makes medicine easy to swallow: An exploratory case study on simplified radiology reports. <https://doi.org/10.48550/ARXIV.2212.14882>
- Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., & Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12), 1–38. <https://doi.org/10.1145/3571730>
- Jin, C., Yu, H., Ke, J., Ding, P., Yi, Y., Jiang, X., Duan, X., Tang, J., Chang, D. T., Wu, X., Gao, F., & Li, R. (2021). Predicting treatment response from longitudinal images using multi-task deep learning. *Nature Communications*, 12(1). <https://doi.org/10.1038/s41467-021-22188-y>

- Jing, Y., Yang, Y., Feng, Z., Ye, J., Yu, Y., & Song, M. (2020). Neural style transfer: A review. *IEEE Transactions on Visualization and Computer Graphics*, 26(11), 3365–3385. <https://doi.org/10.1109/tvcg.2019.2921336>
- Jo, E., Epstein, D. A., Jung, H., & Kim, Y.-H. (2023). Understanding the benefits and challenges of deploying conversational ai leveraging large language models for public health intervention. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/3544548.3581503>
- Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389–399. <https://doi.org/10.1038/s42256-019-0088-2>
- Joensuu, H., Kellokumpu-Lehtinen, P.-L., Bono, P., Alanko, T., Kataja, V., Asola, R., Utriainen, T., Kokko, R., Hemminki, A., Tarkkanen, M., Turpeenniemi-Hujanen, T., Jyrkkiö, S., Flander, M., Helle, L., Ingalsuo, S., Johansson, K., Jääskeläinen, A.-S., Pajunen, M., Rauhala, M., . . . Isola, J. (2006). Adjuvant docetaxel or vinorelbine with or without trastuzumab for breast cancer [PMID: 16495393]. *New England Journal of Medicine*, 354(8), 809–820. <https://doi.org/10.1056/NEJMoa053028>
- John, L. P. N., Oliver P., & Soto., C. J. (2008). Paradigm shift to the integrative Big Five trait taxonomy: History, measurement, and conceptual issues. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (pp. 114–158). The Guilford Press. <http://www.elaborer.org/cours/psy7124/lectures/John2008.pdf>
- John, O. P., & Srivastava, S. (1999). The Big-Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (II, pp. 102–138). New York: Guilford Press. <http://www.personality-project.org/revell/syllabi/classreadings/john.pdf>
- Johnson, H. J., McCormick, M. M., Ibáñez, L., & the Insight Software Consortium. (2019, July). *The ITK Software Guide* (4th). <https://itk.org/ItkSoftwareGuide.pdf>
- Johnson, J. A. (2014). Measuring thirty facets of the Five Factor Model with a 120-item public domain inventory: Development of the IPIP-NEO-120. *Journal of Research in Personality*, 51, 78–89. <https://doi.org/10.1016/j.jrp.2014.05.003>
- Ju, H.-Q., Ying, H., Tian, T., Ling, J., Fu, J., Lu, Y., Wu, M., Yang, L., Achreja, A., Chen, G., Zhuang, Z., Wang, H., Nagrath, D., Yao, J., Hung, M.-C., DePinho, R. A., Huang, P., Xu, R.-H., & Chiao, P. J. (2017). Mutant Kras- and p16-regulated NOX4 activation overcomes metabolic checkpoints in development of pancreatic ductal adenocarcinoma. *Nature Communications*, 8(1). <https://doi.org/10.1038/ncomms14437>
- Jungmann, F., Kaissis, G., Ziegelmayer, S., Harder, F., Schilling, C., Yen, H.-Y., Steiger, K., Weichert, W., Schirren, R., Demir, I. E., Friess, H., Makowski, M. R., Braren, R. F., & Lohöfer, F. K. (2021). Prediction of Tumor Cellularity in Resectable PDAC from Preoperative Computed Tomography Imaging. *Cancers*, 13(9), 2069. <https://doi.org/10.3390/cancers13092069>
- Jungmann, F., Ziegelmayer, S., Lohoefer, F. K., Metz, S., Müller-Leisse, C., Englmaier, M., Makowski, M. R., Kaissis, G. A., & Braren, R. F. (2022). Algorithmic transparency and interpretability measures improve radiologists' performance in BI-RADS 4 classification. *European Radiology*. <https://doi.org/10.1007/s00330-022-09165-9>

- Kadota, M., Sato, M., Duncan, B., Ooshima, A., Yang, H. H., Diaz-Meyer, N., Gere, S., Kageyama, S.-I., Fukuoka, J., Nagata, T., Tsukada, K., Dunn, B. K., Wakefield, L. M., & Lee, M. P. (2009). Identification of Novel Gene Amplifications in Breast Cancer and Coexistence of Gene Amplification with an Activating Mutation of PIK3CA. *Cancer Research*, *69*(18), 7357–7365. <https://doi.org/10.1158/0008-5472.can-09-0064>
- Kaissis, G., & Braren, R. (2019). Pancreatic cancer detection and characterization state of the art cross-sectional imaging and imaging data analysis. *Proceedings of the ACM on Human-Computer Interaction*, *4*, 35–35. <https://doi.org/10.21037/tgh.2019.05.04>
- Kaissis, G., Jungmann, F., Ziegelmayer, S., Lohöfer, F., Harder, F., Schlitter, A. M., Muckenhuber, A., Steiger, K., Schirren, R., Friess, H., Schmid, R., Weichert, W., Makowski, M. R., & Braren, R. F. (2020a). Multiparametric Modelling of Survival in Pancreatic Ductal Adenocarcinoma Using Clinical, Histomorphological, Genetic and Image-Derived Parameters. *Journal of Clinical Medicine*, *9*(5), 1250. <https://doi.org/10.3390/jcm9051250>
- Kaissis, G., Makowski, M. R., Rückert, D., & Braren, R. F. (2020b). Secure, privacy-preserving and federated machine learning in medical imaging. *Nature Machine Intelligence*, *2*(6), 305–311. <https://doi.org/10.1038/s42256-020-0186-1>
- Kaissis, G., Ziegelmayer, S., Lohöfer, F., Steiger, K., Algül, H., Muckenhuber, A., Yen, H.-Y., Rummeny, E., Friess, H., Schmid, R., Weichert, W., Siveke, J. T., & Braren, R. (2019). A machine learning algorithm predicts molecular subtypes in pancreatic ductal adenocarcinoma with differential response to gemcitabine-based versus FOLFIRINOX chemotherapy (F. X. Real, Ed.). *PLOS ONE*, *14*(10). <https://doi.org/10.1371/journal.pone.0218642>
- Kaissis, G., Ziller, A., Passerat-Palmbach, J., Ryffel, T., Usynin, D., Trask, A., Lima, I., Mancuso, J., Jungmann, F., Steinborn, M.-M., Saleh, A., Makowski, M., Rueckert, D., & Braren, R. (2021). End-to-end privacy preserving deep learning on multi-institutional medical imaging. *Nature Machine Intelligence*. <https://doi.org/10.1038/s42256-021-00337-8>
- Kalinsky, K., Jacks, L. M., Heguy, A., Patil, S., Drobnjak, M., Bhanot, U. K., Hedvat, C. V., Traina, T. A., Solit, D., Gerald, W., & Moynahan, M. E. (2009). PIK3CA Mutation Associates with Improved Outcome in Breast Cancer. *Clinical Cancer Research*, *15*(16), 5049–5059. <https://doi.org/10.1158/1078-0432.ccr-09-0632>
- Karpathy, A., & Fei-Fei, L. (2015). Deep visual-semantic alignments for generating image descriptions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3128–3137. https://www.cv-foundation.org/openaccess/content_cvpr_2015/html/Karpathy_Deep_Visual-Semantic_Alignments_2015_CVPR_paper.html
- Karpfinger, C. (2014). *Höhere Mathematik in Rezepten*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-642-37866-9>
- Karras, T., Laine, S., & Aila, T. (2019). A style-based generator architecture for generative adversarial networks. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 4401–4410. https://openaccess.thecvf.com/content_CVPR_2019/html/Karras_A_Style-Based_Generator_Architecture_for_Generative_Adversarial_Networks_CVPR_2019_paper.html

- Katz, A., Saad, E. D., Porter, P., & Pusztai, L. (2007). Primary systemic chemotherapy of invasive lobular carcinoma of the breast. *The Lancet Oncology*, *8*(1), 55–62. [https://doi.org/10.1016/S1470-2045\(06\)71011-7](https://doi.org/10.1016/S1470-2045(06)71011-7)
- Kazeminia, S., Baur, C., Kuijper, A., van Ginneken, B., Navab, N., Albarqouni, S., & Mukhopadhyay, A. (2020). Gans for medical image analysis. *Artificial Intelligence in Medicine*, *109*, 101938. <https://doi.org/10.1016/j.artmed.2020.101938>
- Kazerouni, A., Aghdam, E. K., Heidari, M., Azad, R., Fayyaz, M., Hacıhaliloglu, I., & Merhof, D. (2022). Diffusion models for medical image analysis: A comprehensive survey. <https://doi.org/10.48550/ARXIV.2211.07804>
- Keane, M. G., Horsfall, L., Rait, G., & Pereira, S. P. (2014). A case-control study comparing the incidence of early symptoms in pancreatic and biliary tract cancer. *BMJ Open*, *4*(11). <https://doi.org/10.1136/bmjopen-2014-005720>
- Kelsen, D. P., Portenoy, R., Thaler, H., Tao, Y., & Brennan, M. (1997). Pain as a predictor of outcome in patients with operable pancreatic carcinoma. *Surgery*, *122*(1), 53–59. [https://doi.org/10.1016/S0039-6060\(97\)90264-6](https://doi.org/10.1016/S0039-6060(97)90264-6)
- Kennedy, P., & Taouli, B. (2021). How to implement quantitative imaging in your practice. *Abdominal Radiology*, *47*(9), 2970–2971. <https://doi.org/10.1007/s00261-021-03217-2>
- Kepuska, V., & Bohouta, G. (2018). Next-generation of virtual personal assistants (Microsoft Cortana, Apple Siri, Amazon Alexa and Google Home). *2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)*. <https://doi.org/10.1109/ccwc.2018.8301638>
- Key, T. J., Verkasalo, P. K., & Banks, E. (2001). Epidemiology of breast cancer. *The Lancet Oncology*, *2*(3), 133–140. [https://doi.org/10.1016/S1470-2045\(00\)00254-0](https://doi.org/10.1016/S1470-2045(00)00254-0)
- Khairat, S., Marc, D., Crosby, W., & Sanousi, A. A. (2018). Reasons For Physicians Not Adopting Clinical Decision Support Systems: Critical Analysis. *JMIR Medical Informatics*, *6*(2), e24. <https://doi.org/10.2196/medinform.8912>
- Khalvati, F., Zhang, Y., Baig, S., Lobo-Mueller, E. M., Karanicolas, P., Gallinger, S., & Haider, M. A. (2019). Prognostic Value of CT Radiomic Features in Resectable Pancreatic Ductal Adenocarcinoma. *Scientific Reports*, *9*(1). <https://doi.org/10.1038/s41598-019-41728-7>
- Kim, H., Goo, J. M., Lee, K. H., Kim, Y. T., & Park, C. M. (2020). Preoperative ct-based deep learning model for predicting disease-free survival in patients with lung adenocarcinomas [PMID: 32396042]. *Radiology*, *296*(1), 216–224. <https://doi.org/10.1148/radiol.2020192764>
- Kim, K. H., Do, W.-J., & Park, S.-H. (2018). Improving resolution of mr images with an adversarial network incorporating images with different contrast. *Medical Physics*, *45*(7), 3120–3131. <https://doi.org/10.1002/mp.12945>
- Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. <https://doi.org/10.48550/ARXIV.1412.6980>
- Kingma, D. P., & Welling, M. (2019). An introduction to variational autoencoders. *Foundations and TrendsSM in Machine Learning*, *12*(4), 307–392. <https://doi.org/10.1561/22000000056>

- Kingma, D. P., & Dhariwal, P. (2018). Glow: Generative flow with invertible 1x1 convolutions. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, & R. Garnett (Eds.), *Advances in neural information processing systems* (Vol. 31). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2018/file/d139db6a236200b21cc7f752979132d0-Paper.pdf
- Klang, E., Cohen-Shelly, M., & Lopez-Jimenez, F. (2023). Leveraging large language models to enhance digital health in cardiology: A preview of a cutting-edge language generation model. *Mayo Clinic Proceedings: Digital Health*, 1(2), 105–108. <https://doi.org/10.1016/j.mcpdig.2023.03.003>
- Kobyzev, I., Prince, S. J. D., & Brubaker, M. A. (2021). Normalizing flows: An introduction and review of current methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(11), 3964–3979. <https://doi.org/10.1109/tpami.2020.2992934>
- Kohli, A., & Jha, S. (2018). Why CAD Failed in Mammography. *Journal of the American College of Radiology*, 15(3), 535–537. <https://doi.org/10.1016/j.jacr.2017.12.029>
- Kowalewski, A., Szyllberg, L., Saganek, M., Napiótek, W., Antosik, P., & Grzanka, D. (2018). Emerging strategies in BRCA-positive pancreatic cancer. *Journal of Cancer Research and Clinical Oncology*, 144(8), 1503–1507. <https://doi.org/10.1007/s00432-018-2666-9>
- Kuchenbaecker, K. B., Hopper, J. L., Barnes, D. R., Phillips, K.-A., Mooij, T. M., Roos-Blom, M.-J., Jervis, S., van Leeuwen, F. E., Milne, R. L., Andrieu, N., Goldgar, D. E., Terry, M. B., Rookus, M. A., Easton, D. F., Antoniou, A. C., McGuffog, L., Evans, D. G., Barrowdale, D., Frost, D., . . . Olsson, H. (2017). Risks of Breast, Ovarian, and Contralateral Breast Cancer for BRCA and BRCA Mutation Carriers. *JAMA*, 317(23), 2402–2416. <https://doi.org/10.1001/jama.2017.7112>
- Kuhl, C. K., Schrading, S., Leutner, C. C., Morakkabati-Spitz, N., Wardelmann, E., Fimmers, R., Kuhn, W., & Schild, H. H. (2005). Mammography, breast ultrasound, and magnetic resonance imaging for surveillance of women at high familial risk for breast cancer. *Journal of Clinical Oncology*, 23(33), 8469–8476. <https://doi.org/10.1200/jco.2004.00.4960>
- Kulendran, M., Salhab, M., & Mokbel, K. (2009). Oestrogen-synthesising enzymes and breast cancer. *Anticancer Research*, 29(4), 1095–1109. <https://ar.iiarjournals.org/content/29/4/1095>
- Kumar, V., Gu, Y., Basu, S., Berglund, A., Eschrich, S. A., Schabath, M. B., Forster, K., Aerts, H. J. W. L., Dekker, A., Fenstermacher, D., Goldgof, D. B., Hall, L. O., Lambin, P., Balagurunathan, Y., Gatenby, R. A., & Gillies, R. J. (2012). Radiomics: the process and the challenges. *Magnetic Resonance Imaging*, 30(9), 1234–1248. <https://doi.org/10.1016/j.mri.2012.06.010>
- Kumarakulasinghe, N. B., van Zanwijk, N., & Soo, R. A. (2015). Molecular targeted therapy in the treatment of advanced stage non-small cell lung cancer (NSCLC). *Respirology*, 20(3), 370–378. <https://doi.org/10.1111/resp.12490>
- Kunzmann, V., Algül, H., Goekkurt, E., Siegler, G. M., Martens, U. M., Waldschmidt, D., Pelzer, U., Hennes, E., Fuchs, M., Siveke, J., Kullmann, F., Boeck, S., Ettrich, T. J., Ferenczy, P., Keller, R., Germer, C.-T., Stein, H., Hartlapp, I., Klein, I., & Heinemann, V. (2019). 671O - Conversion rate in locally advanced pancreatic cancer (LAPC) after nab-paclitaxel/gemcitabine- or FOLFIRINOX-based induction chemotherapy (NEOLAP): Final results of a multicenter randomised phase II AIO trial [Abstract Book of the 44th ESMO Congress (ESMO 2019) 27 September 1 October 2019, Barcelona, Spain]. *Annals of Oncology*, 30, v253. <https://doi.org/10.1093/annonc/mdz247>

- Kwapisz, D. (2020). Pembrolizumab and atezolizumab in triple-negative breast cancer. *Cancer Immunology, Immunotherapy*, 70(3), 607–617. <https://doi.org/10.1007/s00262-020-02736-z>
- Lakhani, P., Prater, A. B., Hutson, R. K., Andriole, K. P., Dreyer, K. J., Morey, J., Prevedello, L. M., Clark, T. J., Geis, J. R., Itri, J. N., & Hawkins, C. M. (2018). Machine Learning in Radiology: Applications Beyond Image Interpretation. *Journal of the American College of Radiology*, 15(2), 350–359. <https://doi.org/10.1016/j.jacr.2017.09.044>
- Lakhani, P., & Sundaram, B. (2017). Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks. *Radiology*, 284(2), 574–582. <https://doi.org/10.1148/radiol.2017162326>
- Lambin, P., Leijenaar, R. T. H., Deist, T. M., Peerlings, J., de Jong, E. E. C., van Timmeren, J., Sanduleanu, S., Larue, R. T. H. M., Even, A. J. G., Jochems, A., van Wijk, Y., Woodruff, H., van Soest, J., Lustberg, T., Roelofs, E., van Elmpt, W., Dekker, A., Mottaghy, F. M., Wildberger, J. E., & Walsh, S. (2017). Radiomics: The bridge between medical imaging and personalized medicine. *Nature Reviews Clinical Oncology*, 14(12), 749–762. <https://doi.org/10.1038/nrclinonc.2017.141>
- Lambin, P., Rios-Velazquez, E., Leijenaar, R., Carvalho, S., van Stiphout, R. G. P. M., Granton, P., Zegers, C. M. L., Gillies, R., Boellard, R., Dekker, A., & Aerts, H. J. W. L. (2012). Radiomics: Extracting more information from medical images using advanced feature analysis. *European Journal of Cancer*, 48(4), 441–446. <https://doi.org/10.1016/j.ejca.2011.11.036>
- Langer, S. G., Tellis, W., Carr, C., Daly, M., Erickson, B. J., Mendelson, D., Moore, S., Perry, J., Shastri, K., Warnock, M., & Zhu, W. (2015). The rsna image sharing network. *Journal of digital imaging*, 28, 53–61. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4305053/>
- Lau, K. H., Tan, A. M., & Shi, Y. (2022). New and emerging targeted therapies for advanced breast cancer. *International Journal of Molecular Sciences*, 23(4). <https://doi.org/10.3390/ijms23042288>
- Lawrence, S., Giles, C. L., Tsoi, A. C., & Back, A. D. (1997). Face recognition: A convolutional neural-network approach. *IEEE transactions on neural networks*, 8(1), 98–113. http://www.cs.cmu.edu/afs/cs/user/bhiksha/WWW/courses/deeplearning/Fall.2016/pdfs/Lawrence_et_al.pdf
- Le, E. P. V., Wang, Y., Huang, Y., Hickman, S., & Gilbert, F. J. (2019). Artificial intelligence in breast imaging. *Clinical Radiology*, 74(5), 357–366. <https://doi.org/10.1016/j.crad.2019.02.006>
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- LeCun, Y., Boser, B. E., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W. E., & Jackel, L. D. (1989, November). Handwritten Digit Recognition with a Back-Propagation Network. In D. Touretzky (Ed.), *Advances in neural information processing systems* (pp. 396–404, Vol. 2). Morgan-Kaufmann. <http://papers.nips.cc/paper/293-handwritten-digit-recognition-with-a-back-propagation-network.pdf>
- Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., & Shi, W. (2017). Photo-realistic single image super-resolution using a generative adversarial network. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*

- (CVPR), 4681–4690. https://openaccess.thecvf.com/content_cvpr_2017/html/Ledig_Photo-Realistic_Single_Image_CVPR_2017_paper.html
- Lee, C. H., Dershaw, D. D., Kopans, D., Evans, P., Monsees, B., Monticciolo, D., Brenner, R. J., Bassett, L., Berg, W., Feig, S., Hendrick, E., Mendelson, E., D’Orsi, C., Sickles, E., & Burhenne, L. W. (2010). Breast Cancer Screening With Imaging: Recommendations From the Society of Breast Imaging and the ACR on the Use of Mammography, Breast MRI, Breast Ultrasound, and Other Technologies for the Detection of Clinically Occult Breast Cancer. *Journal of the American College of Radiology*, 7(1), 18–27. <https://doi.org/10.1016/j.jacr.2009.09.022>
- Lee, G., Lee, H. Y., Park, H., Schiebler, M. L., van Beek, E. J. R., Ohno, Y., Seo, J. B., & Leung, A. (2017a). Radiomics and its emerging role in lung cancer research, imaging biomarkers and clinical management: State of the art. *European Journal of Radiology*, 86, 297–307. <https://doi.org/10.1016/j.ejrad.2016.09.005>
- Lee, R. S., Gimenez, F., Hoogi, A., & Rubin, D. (2016). Curated breast imaging subset of dds. <https://doi.org/10.7937/K9/TCIA.2016.7002S9CY>
- Lee, R. S., Gimenez, F., Hoogi, A., Miyake, K. K., Gorovoy, M., & Rubin, D. L. (2017b). A curated mammography data set for use in computer-aided detection and diagnosis research. *Scientific Data*, 4(1). <https://doi.org/10.1038/sdata.2017.177>
- Lehman, C. D., Arao, R. F., Sprague, B. L., Lee, J. M., Buist, D. S. M., Kerlikowske, K., Henderson, L. M., Onega, T., Tosteson, A. N. A., Rauscher, G. H., & Miglioretti, D. L. (2017). National Performance Benchmarks for Modern Screening Digital Mammography: Update from the Breast Cancer Surveillance Consortium. *Radiology*, 283(1), 49–58. <https://doi.org/10.1148/radiol.2016161174>
- Lehman, C. D., Wellman, R. D., Buist, D. S. M., Kerlikowske, K., Tosteson, A. N. A., & Miglioretti, D. L. (2015). Diagnostic Accuracy of Digital Screening Mammography With and Without Computer-Aided Detection. *JAMA Internal Medicine*, 175(11), 1828. <https://doi.org/10.1001/jamainternmed.2015.5231>
- Li, K. (2022). Basic gan models and the application in medical image field. *Journal of Physics: Conference Series*, 2386(1), 012040. <https://doi.org/10.1088/1742-6596/2386/1/012040>
- Li, L., Fan, Y., Tse, M., & Lin, K.-Y. (2020). A review of applications in federated learning. *Computers & Industrial Engineering*, 149, 106854. <https://doi.org/10.1016/j.cie.2020.106854>
- Li, S., Deng, Y.-Q., Zhu, Z.-L., Hua, H.-L., & Tao, Z.-Z. (2021a). A Comprehensive Review on Radiomics and Deep Learning for Nasopharyngeal Carcinoma Imaging. *Diagnostics*, 11(9), 1523. <https://doi.org/10.3390/diagnostics11091523>
- Li, W., Milletari, F., Xu, D., Rieke, N., Hancox, J., Zhu, W., Baust, M., Cheng, Y., Ourselin, S., Cardoso, M. J., & Feng, A. (2019, October). Privacy-Preserving Federated Brain Tumour Segmentation. In H.-I. Suk, M. Liu, P. Yan, & C. Lian (Eds.), *Machine learning in medical imaging* (pp. 133–141). Springer International Publishing. https://doi.org/10.1007/978-3-030-32692-0_16
- Li, X., Jiang, Y., Rodriguez-Andina, J. J., Luo, H., Yin, S., & Kaynak, O. (2021b). When medical images meet generative adversarial network: Recent development and research opportunities. *Discover Artificial Intelligence*, 1(1). <https://doi.org/10.1007/s44163-021-00006-0>

- Liggett, W. H., & Sidransky, D. (1998). Role of the p16 tumor suppressor gene in cancer. *Journal of Clinical Oncology*, *16*(3), 1197–1206. <https://doi.org/10.1200/jco.1998.16.3.1197>
- Limkin, E. J., Sun, R., Dercle, L., Zacharaki, E. I., Robert, C., Reuzé, S., Schernberg, A., Paragios, N., Deutsch, E., & Ferte, C. (2017). Promises and challenges for the implementation of computational medical imaging (radiomics) in oncology. *Annals of Oncology*, *28*(6), 1191–1206. <https://doi.org/10.1093/annonc/mdx034>
- Lin Baochuan, M. D. W. a. J. Y. (1998). Hepatocyte nuclear factor 1 *Alpha* is an accessory factor required for activation of glucose-6-phosphatase gene transcription by glucocorticoids. *DNA and Cell Biology*, *17*(11), 967–974. <https://doi.org/10.1089/dna.1998.17.967>
- Lipton, Z. C. (2018). The Mythos of Model Interpretability. *Queue*, *16*(3), 31–57. <https://doi.org/10.1145/3236386.3241340>
- Lipton, Z. C., Berkowitz, J., & Elkan, C. (2015). A critical review of recurrent neural networks for sequence learning. <https://doi.org/10.48550/ARXIV.1506.00019>
- Liu, H., Wang, Y., Fan, W., Liu, X., Li, Y., Jain, S., Jain, A. K., & Tang, J. (2021). Trustworthy ai: A computational perspective. *arXiv preprint arXiv:2107.06641*. <https://arxiv.org/abs/2107.06641>
- Liu, Y., Kim, J., Qu, F., Liu, S., Wang, H., Balagurunathan, Y., Ye, Z., & Gillies, R. J. (2016). Ct features associated with epidermal growth factor receptor mutation status in patients with lung adenocarcinoma [PMID: 26937803]. *Radiology*, *280*(1), 271–280. <https://doi.org/10.1148/radiol.2016151455>
- Liu, Z., Zhong, A., Li, Y., Yang, L., Ju, C., Wu, Z., Ma, C., Shu, P., Chen, C., Kim, S., Dai, H., Zhao, L., Zhu, D., Liu, J., Liu, W., Shen, D., Li, X., Li, Q., & Liu, T. (2023). Radiology-gpt: A large language model for radiology. <https://doi.org/10.48550/ARXIV.2306.08666>
- Loehrer, P. J., Feng, Y., Cardenes, H., Wagner, L., Brell, J. M., Cella, D., Flynn, P., Ramanathan, R. K., Crane, C. H., Alberts, S. R., & Benson, A. B. (2011). Gemcitabine Alone Versus Gemcitabine Plus Radiotherapy in Patients With Locally Advanced Pancreatic Cancer: An Eastern Cooperative Oncology Group Trial. *Journal of Clinical Oncology*, *29*(31), 4105–4112. <https://doi.org/10.1200/jco.2011.34.8904>
- Loibl, S., & Gianni, L. (2017). HER2-positive breast cancer. *The Lancet*, *389*(10087), 2415–2429. [https://doi.org/10.1016/S0140-6736\(16\)32417-5](https://doi.org/10.1016/S0140-6736(16)32417-5)
- Lorensen, W. E., & Cline, H. E. (1987). Marching cubes: A high resolution 3D surface construction algorithm. *ACM SIGGRAPH Computer Graphics*, *21*(4), 163–169. <https://doi.org/10.1145/37402.37422>
- Lorentzen, H. F. (2019). Targeted therapy for malignant melanoma [Respiratory * Dermatology]. *Current Opinion in Pharmacology*, *46*, 116–121. <https://doi.org/10.1016/j.coph.2019.05.010>
- Lotter, W., Diab, A. R., Haslam, B., Kim, J. G., Grisot, G., Wu, E., Wu, K., Onieva, J. O., Boyer, Y., Boxerman, J. L., Wang, M., Bandler, M., Vijayaraghavan, G. R., & Sorensen, A. G. (2021). Robust breast cancer detection in mammography and digital breast tomosynthesis using an annotation-efficient deep learning approach. *Nature Medicine*, *27*(2), 244–249. <https://doi.org/10.1038/s41591-020-01174-9>
- Lu, J., Behbood, V., Hao, P., Zuo, H., Xue, S., & Zhang, G. (2015). Transfer learning using computational intelligence: A survey. *Knowledge-Based Systems*, *80*, 14–23. <https://doi.org/10.1016/j.knsys.2015.01.010>

- Lubner, M. G., Smith, A. D., Sandrasegaran, K., Sahani, D. V., & Pickhardt, P. J. (2017). CT Texture Analysis: Definitions, Applications, Biologic Correlates, and Challenges. *RadioGraphics*, *37*(5), 1483–1503. <https://doi.org/10.1148/rg.2017170056>
- Lumachi, F., Luisetto, G., Basso, S. M. M., Basso, U., Brunello, A., & Camozzi, V. (2011). Endocrine therapy of breast cancer. *Current Medicinal Chemistry*, *18*(4), 513–522. <https://doi.org/10.2174/092986711794480177>
- Luo, Y., Tian, L., Feng, Y., Yi, M., Chen, X., & Huang, Q. (2012). The Predictive Role of p16 Deletion, p53 Deletion, and Polysomy 9 and 17 in Pancreatic Ductal Adenocarcinoma. *Pathology & Oncology Research*, *19*(1), 35–40. <https://doi.org/10.1007/s12253-012-9555-3>
- Lv, W., Yuan, Q., Wang, Q., Ma, J., Feng, Q., Chen, W., Rahmim, A., & Lu, L. (2019). Radiomics analysis of PET and CT components of PET/CT imaging integrated with clinical parameters: Application to prognosis for nasopharyngeal carcinoma. *Molecular Imaging and Biology*, *21*(5), 954–964. <https://doi.org/10.1007/s11307-018-01304-3>
- Lyons, T. G. (2019). Targeted therapies for triple-negative breast cancer. *Current Treatment Options in Oncology*, *20*(11). <https://doi.org/10.1007/s11864-019-0682-x>
- Ma, C., Ji, Z., & Gao, M. (2019). Neural style transfer improves 3d cardiovascular mr image segmentation on inconsistent data. In D. Shen, T. Liu, T. M. Peters, L. H. Staib, C. Essert, S. Zhou, P.-T. Yap, & A. Khan (Eds.), *Medical image computing and computer assisted intervention – miccai 2019* (pp. 128–136). Springer International Publishing. https://link.springer.com/chapter/10.1007/978-3-030-32245-8_15
- Maas, L., Geurtsen, M., Nouwt, F., Schouten, S., Van De Water, R., Van Dulmen, S., Dalpiaz, F., Van Deemter, K., & Brinkkemper, S. (2020). The care2report system: Automated medical reporting as an integrated solution to reduce administrative burden in healthcare. *HICSS*, 1–10. <https://core.ac.uk/reader/286030500>
- Maddalena, M., Mallel, G., Nataraj, N. B., Shreberk-Shaked, M., Hassin, O., Mukherjee, S., Arandkar, S., Rotkopf, R., Kapsack, A., Lambiase, G., Pellegrino, B., Ben-Isaac, E., Golani, O., Addadi, Y., Hajaj, E., Eilam, R., Straussman, R., Yarden, Y., Lotem, M., & Oren, M. (2021). TP53 missense mutations in PDAC are associated with enhanced fibrosis and an immunosuppressive microenvironment. *Proceedings of the National Academy of Sciences*, *118*(23), e2025631118. <https://doi.org/10.1073/pnas.2025631118>
- Maddox, T. M., Rumsfeld, J. S., & Payne, P. R. O. (2019). Questions for Artificial Intelligence in Health Care. *JAMA*, *321*(1), 31. <https://doi.org/10.1001/jama.2018.18932>
- Mallio, C. A., Sertorio, A. C., Bernetti, C., & Zobel, B. B. (2023). Large language models for structured reporting in radiology: Performance of GPT-4, ChatGPT-3.5, perplexity and bing. *La radiologia medica*, *128*(7), 808–812. <https://doi.org/10.1007/s11547-023-01651-4>
- Mallon, E. (2000). The Basic Pathology of Human Breast Cancer. *Journal of Mammary Gland Biology and Neoplasia*, *5*(2), 139–163. <https://doi.org/10.1023/a:1026439204849>

- Maniatopoulos, G., Procter, R., Llewellyn, S., Harvey, G., & Boyd, A. (2015). Moving beyond local practice: Reconfiguring the adoption of a breast cancer diagnostic technology. *Social Science & Medicine*, *131*, 98–106. <https://doi.org/10.1016/j.socscimed.2015.02.036>
- Marmot, M. G., Altman, D. G., Cameron, D. A., Dewar, J. A., Thompson, S. G., & Wilcox, M. (2013). The benefits and harms of breast cancer screening: An independent review. *British Journal of Cancer*, *108*(11), 2205–2240. <https://doi.org/10.1038/bjc.2013.177>
- Martínez-Pérez, C., Turnbull, A. K., Ekatah, G. E., Arthur, L. M., Sims, A. H., Thomas, J. S., & Dixon, J. M. (2017). Current treatment trends and the need for better predictive tools in the management of ductal carcinoma in situ of the breast. *Cancer Treatment Reviews*, *55*, 163–172. <https://doi.org/10.1016/j.ctrv.2017.03.009>
- Martin-Gonzalez, P., Crispin-Ortuzar, M., Rundo, L., Delgado-Ortet, M., Reinius, M., Beer, L., Woitek, R., Ursprung, S., Addley, H., Brenton, J. D., Markowitz, F., & Sala, E. (2020). Integrative radiogenomics for virtual biopsy and treatment monitoring in ovarian cancer. *Insights into Imaging*, *11*(1). <https://doi.org/10.1186/s13244-020-00895-2>
- Mathias, C., Bortoletto, S., Centa, A., Komechen, H., Lima, R. S., Fonseca, A. S., Sebastião, A. P., Urban, C. A., Soares, E. W. S., Prando, C., Figueiredo, B. C., Cavalli, I. J., Cavalli, L. R., & Ribeiro, E. M. F. S. (2020). Frequency of the TP53 R337H variant in sporadic breast cancer and its impact on genomic instability. *Scientific Reports*, *10*(1). <https://doi.org/10.1038/s41598-020-73282-y>
- Matlaga, B. R., Kawamoto, S., & Fishman, E. (2008). Dual Source Computed Tomography: A Novel Technique to Determine Stone Composition. *Urology*, *72*(5), 1164–1168. <https://doi.org/10.1016/j.urology.2008.03.051>
- Mazurowski, M. A., Buda, M., Saha, A., & Bashir, M. R. (2018). Deep learning in radiology: An overview of the concepts and a survey of the state of the art with focus on MRI. *Journal of Magnetic Resonance Imaging*, *49*(4), 939–954. <https://doi.org/10.1002/jmri.26534>
- McKinney, S. M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafiyan, H., Back, T., Chesus, M., Corrado, G. C., Darzi, A., Etemadi, M., Garcia-Vicente, F., Gilbert, F. J., Halling-Brown, M., Hassabis, D., Jansen, S., Karthikesalingam, A., Kelly, C. J., King, D., ... Shetty, S. (2020). International evaluation of an AI system for breast cancer screening. *Nature*, *577*(7788), 89–94. <https://doi.org/10.1038/s41586-019-1799-6>
- McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, *12*(2), 153–157. <https://doi.org/10.1007/bf02295996>
- Mei, D., Luo, Y., Wang, Y., & Gong, J. (2018). CT texture analysis of lung adenocarcinoma: can Radiomic features be surrogate biomarkers for EGFR mutation statuses. *Cancer Imaging*, *18*(1). <https://doi.org/10.1186/s40644-018-0184-2>
- Mei, X., Liu, Z., Robson, P. M., Marinelli, B., Huang, M., Doshi, A., Jacobi, A., Cao, C., Link, K. E., Yang, T., Wang, Y., Greenspan, H., Deyer, T., Fayad, Z. A., & Yang, Y. (2022). Radimagenet: An open radiologic deep learning research dataset for effective transfer learning. *Radiology: Artificial Intelligence*, *4*(5), e210315. <https://doi.org/10.1148/ryai.210315>

- Meissen, F., Kaissis, G., & Rueckert, D. (2022). Challenging current semi-supervised anomaly segmentation methods for brain mri. In A. Crimi & S. Bakas (Eds.), *Brainlesion: Glioma, multiple sclerosis, stroke and traumatic brain injuries* (pp. 63–74). Springer International Publishing. https://link.springer.com/chapter/10.1007/978-3-031-08999-2_5
- Meissen, F., Paetzold, J., Kaissis, G., & Rueckert, D. (2023). Unsupervised anomaly localization with structural feature-autoencoders. In S. Bakas, A. Crimi, U. Baid, S. Malec, M. Pytlarz, B. Baheti, M. Zenk, & R. Dorent (Eds.), *Brainlesion: Glioma, multiple sclerosis, stroke and traumatic brain injuries* (pp. 14–24). Springer Nature Switzerland. https://link.springer.com/chapter/10.1007/978-3-031-33842-7_2
- Menon, S., & Vondrick, C. (2022, December). Visual classification via description from large language models. <https://doi.org/10.48550/ARXIV.2210.07183>
- Mercado, C. L. (2014). BI-RADS Update. *Radiologic Clinics of North America*, 52(3), 481–487. <https://doi.org/10.1016/j.rcl.2014.02.008>
- Meyer, M., Ronald, J., Vernuccio, F., Nelson, R. C., Ramirez-Giraldo, J. C., Solomon, J., Patel, B. N., Samei, E., & Marin, D. (2019). Reproducibility of ct radiomic features within the same patient: Influence of radiation dose and ct reconstruction settings [PMID: 31573400]. *Radiology*, 293(3), 583–591. <https://doi.org/10.1148/radiol.2019190928>
- Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 267, 1–38. <https://doi.org/10.1016/j.artint.2018.07.007>
- Miotto, R., Wang, F., Wang, S., Jiang, X., & Dudley, J. T. (2018). Deep learning for healthcare: Review, opportunities and challenges. *Briefings in bioinformatics*, 19, 1236–1246. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6455466/>
- Moen, E., Bannon, D., Kudo, T., Graf, W., Covert, M., & Valen, D. V. (2019). Deep learning for cellular image analysis. *Nature Methods*, 16(12), 1233–1246. <https://doi.org/10.1038/s41592-019-0403-1>
- Mohamed, A., Krajewski, K., Cakar, B., & Ma, C. X. (2013). Targeted therapy for breast cancer. *The American Journal of Pathology*, 183(4), 1096–1112. <https://doi.org/10.1016/j.ajpath.2013.07.005>
- Molnar, C. (2020). *Interpretable Machine Learning*. Lulu. com. https://originalstatic.aminer.cn/misc/pdf/Molnar-interpretable-machine-learning_compressed.pdf%20https://christophm.github.io/interpretable-ml-book/
- Monarch, R. M. (2021). *Human-in-the-Loop Machine Learning: Active learning and annotation for human-centered AI*. Simon; Schuster.
- Morin, O., Vallières, M., Jochems, A., Woodruff, H. C., Valdes, G., Braunstein, S. E., Wildberger, J. E., Villanueva-Meyer, J. E., Kearney, V., Yom, S. S., Solberg, T. D., & Lambin, P. (2018). A deep look into the future of quantitative imaging in oncology: A statement of working principles and proposal for change [Imaging in Radiation Oncology]. *International Journal of Radiation Oncology*Biophysics*, 102(4), 1074–1082. <https://doi.org/10.1016/j.ijrobp.2018.08.032>
- Morrow, M., Waters, J., & Morris, E. (2011). Mri for breast cancer screening, diagnosis, and treatment. *The Lancet*, 378(9805), 1804–1811. [https://doi.org/10.1016/S0140-6736\(11\)61350-0](https://doi.org/10.1016/S0140-6736(11)61350-0)

- Morton, J. P., Timpson, P., Karim, S. A., Ridgway, R. A., Athineos, D., Doyle, B., Jamieson, N. B., Oien, K. A., Lowy, A. M., Brunton, V. G., Frame, M. C., Evans, T. R. J., & Sansom, O. J. (2010). Mutant p53 drives metastasis and overcomes growth arrest/senescence in pancreatic cancer. *Proceedings of the National Academy of Sciences*, *107*(1), 246–251. <https://doi.org/10.1073/pnas.0908428107>
- Mousavi Baigi, S. F., Sarbaz, M., Ghaddaripouri, K., Ghaddaripouri, M., Mousavi, A. S., & Kimiafar, K. (2023). Attitudes, knowledge, and skills towards artificial intelligence among healthcare students: A systematic review. *Health Science Reports*, *6*(3), e1138. <https://doi.org/10.1002/hsr2.1138>
- Muckenhuber, A., Berger, A. K., Schlitter, A. M., Steiger, K., Konukiewitz, B., Trumpp, A., Eils, R., Werner, J., Friess, H., Esposito, I., Klöppel, G., Ceyhan, G. O., Jesinghaus, M., Denkert, C., Bahra, M., Stenzinger, A., Sprick, M. R., Jäger, D., Springfeld, C., & Weichert, W. (2017). Pancreatic Ductal Adenocarcinoma Subtyping Using the Biomarkers Hepatocyte Nuclear Factor-1A and Cytokeratin-81 Correlates with Outcome and Treatment Response. *Clinical Cancer Research*, *24*(2), 351–359. <https://doi.org/10.1158/1078-0432.ccr-17-2180>
- Muhammad, H., Häggström, I., Klimstra, D. S., & Fuchs, T. J. (2018, September). Survival modeling of pancreatic cancer with radiology using convolutional neural networks. In D. Stoyanov, Z. Taylor, S. Aylward, J. M. R. Tavares, Y. Xiao, A. Simpson, A. Martel, L. Maier-Hein, S. Li, H. Rivaz, I. Reinertsen, M. Chabanas, & K. Farahani (Eds.), *Simulation, image processing, and ultrasound systems for assisted diagnosis and navigation* (pp. 187–192). Springer International Publishing. https://doi.org/10.1007/978-3-030-01045-4_23
- Musen, M. A., Middleton, Blackford, & Greenes, R. A. (2021, November). Clinical Decision-Support Systems. In E. H. Shortliffe & J. J. Cimino (Eds.), *Biomedical informatics: Computer applications in health care and biomedicine* (pp. 795–840). Springer International Publishing. https://doi.org/10.1007/978-3-030-58721-5_24
- Mutasa, S., Sun, S., & Ha, R. (2021). Understanding artificial intelligence based radiology studies: CNN architecture. *Clinical Imaging*, *80*, 72–76. <https://doi.org/10.1016/j.clinimag.2021.06.033>
- National Cancer Institute. (2020a). NCI Dictionary of Cancer Terms [Retrieved from: <https://www.cancer.gov/publications/dictionaries/cancer-terms/def/personalized-medicine>]. <https://www.cancer.gov/publications/dictionaries/cancer-terms/def/personalized-medicine>
- National Cancer Institute. (2020b). Symptoms of Cancer [last accessed on 12.04.2023]. Retrieved November 6, 2020, from <https://www.cancer.org/treatment/understanding-your-diagnosis/signs-and-symptoms-of-cancer.html>
- Neoptolemos, J. P., Stocken, D. D., Friess, H., Bassi, C., Dunn, J. A., Hickey, H., Beger, H., Fernandez-Cruz, L., Dervenis, C., Lacaine, F., Falconi, M., Pederzoli, P., Pap, A., Spooner, D., Kerr, D. J., & Büchler, M. W. (2004). A Randomized Trial of Chemoradiotherapy and Chemotherapy after Resection of Pancreatic Cancer. *New England Journal of Medicine*, *350*(12), 1200–1210. <https://doi.org/10.1056/nejmoa032295>
- Nevanlinna, H., & Bartek, J. (2006). The CHEK2 gene and inherited breast cancer susceptibility. *Oncogene*, *25*(43), 5912–5919. <https://doi.org/10.1038/sj.onc.1209877>

- Nie, D., Trullo, R., Lian, J., Wang, L., Petitjean, C., Ruan, S., Wang, Q., & Shen, D. (2018). Medical image synthesis with deep convolutional adversarial networks. *IEEE Transactions on Biomedical Engineering*, 65(12), 2720–2730. <https://doi.org/10.1109/tbme.2018.2814538>
- Nixon, J., Dusenberry, M., Zhang, L., Jerfel, G., & Tran, D. (2019). Measuring Calibration in Deep Learning. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 38–41. <http://arxiv.org/abs/1904.01685>
- Noll, E. M., Eisen, C., Stenzinger, A., Espinet, E., Muckenhuber, A., Klein, C., Vogel, V., Klaus, B., Nadler, W., Rösli, C., Lutz, C., Kulke, M., Engelhardt, J., Zickgraf, F. M., Espinosa, O., Schlesner, M., Jiang, X., Kopp-Schneider, A., Neuhaus, P., . . . Sprick, M. R. (2016). CYP3A5 mediates basal and acquired therapy resistance in different subtypes of pancreatic ductal adenocarcinoma. *Nature Medicine*, 22(3), 278–287. <https://doi.org/10.1038/nm.4038>
- Nothacker, M., Duda, V., Hahn, M., Warm, M., Degenhardt, F., Madjar, H., Weinbrenner, S., & Albert, U.-S. (2009). Early detection of breast cancer: Benefits and risks of supplemental breast ultrasound in asymptomatic women with mammographically dense breast tissue. a systematic review. *BMC Cancer*, 9(1). <https://doi.org/10.1186/1471-2407-9-335>
- Obenauer, S., Hermann, K. P., & Grabbe, E. (2005). Applications and literature review of the BI-RADS classification. *European Radiology*, 15(5), 1027–1036. <https://doi.org/10.1007/s00330-004-2593-9>
- O'Connor, J. P. B., Aboagye, E. O., Adams, J. E., Aerts, H. J. W. L., Barrington, S. F., Beer, A. J., Boellaard, R., Bohndiek, S. E., Brady, M., Brown, G., Buckley, D. L., Chenevert, T. L., Clarke, L. P., Collette, S., Cook, G. J., deSouza, N. M., Dickson, J. C., Dive, C., Evelhoch, J. L., . . . Waterton, J. C. (2016). Imaging biomarker roadmap for cancer studies. *Nature Reviews Clinical Oncology*, 14(3), 169–186. <https://doi.org/10.1038/nrclinonc.2016.162>
- Oeffinger, K. C., Fontham, E. T. H., Etzioni, R., Herzig, A., Michaelson, J. S., Shih, Y.-C. T., Walter, L. C., Church, T. R., Flowers, C. R., LaMonte, S. J., Wolf, A. M. D., DeSantis, C., Lortet-Tieulent, J., Andrews, K., Manassaram-Baptiste, D., Saslow, D., Smith, R. A., Brawley, O. W., & Wender, R. (2015). Breast Cancer Screening for Women at Average Risk. *JAMA*, 314(15), 1599. <https://doi.org/10.1001/jama.2015.12783>
- Oettle, H., Neuhaus, P., Hochhaus, A., Hartmann, J. T., Gellert, K., Ridwelski, K., Niedergethmann, M., Zülke, C., Fahlke, J., Arning, M. B., Sinn, M., Hinke, A., & Riess, H. (2013). Adjuvant Chemotherapy With Gemcitabine and Long-term Outcomes Among Patients With Resected Pancreatic Cancer. *JAMA*, 310(14), 1473. <https://doi.org/10.1001/jama.2013.279201>
- Oh, D.-Y., & Bang, Y.-J. (2019). HER2-targeted therapies - a role beyond breast cancer. *Nature Reviews Clinical Oncology*, 17(1), 33–48. <https://doi.org/10.1038/s41571-019-0268-3>
- Olut, S., Sahin, Y. H., Demir, U., & Unal, G. (2018, September). Generative adversarial training for mra image synthesis using multi-contrast mri. In I. Rekik, G. Unal, E. Adeli, & S. H. Park (Eds.), *Predictive intelligence in medicine* (pp. 147–154). Springer International Publishing. https://link.springer.com/chapter/10.1007/978-3-030-00320-3_18
- OpenAI. (2023, March). Gpt-4 technical report. <https://doi.org/10.48550/ARXIV.2303.08774>

- O'Reilly, E. M., Lee, J. W., Zalupski, M., Capanu, M., Park, J., Golan, T., Tahover, E., Lowery, M. A., Chou, J. F., Sahai, V., Brenner, R., Kindler, H. L., Yu, K. H., Zervoudakis, A., Vemuri, S., Stadler, Z. K., Do, R. K. G., Dhani, N., Chen, A. P., & Kelsen, D. P. (2020). Randomized, Multicenter, Phase II Trial of Gemcitabine and Cisplatin With or Without Veliparib in Patients With Pancreas Adenocarcinoma and a Germline BRCA/PALB2 Mutation. *Journal of Clinical Oncology*, *38*(13), 1378–1388. <https://doi.org/10.1200/jco.19.02931>
- Oshima, M., Okano, K., Muraki, S., Haba, R., Maeba, T., Suzuki, Y., & Yachida, S. (2013). Immunohistochemically Detected Expression of 3 Major Genes (CDKN2A/p16, TP53, and SMAD4/DPC4) Strongly Predicts Survival in Patients With Resectable Pancreatic Cancer. *Annals of Surgery*, *258*(2), 336–346. <https://doi.org/10.1097/sla.0b013e3182827a65>
- Ottendorf, N. A., de Wilde, R. F., Maitra, A., Hruban, R. H., & Offerhaus, G. J. A. (2011). Molecular Characteristics of Pancreatic Ductal Adenocarcinoma. *Pathology Research International*, *2011*, 1–16. <https://doi.org/10.4061/2011/620601>
- Ouali, Y., Hudelot, C., & Tami, M. (2020). An overview of deep semi-supervised learning. *arXiv: 2006.05278*. <https://doi.org/10.48550/ARXIV.2006.05278>
- Palacio-Niño, J.-O., & Berzal, F. (2019). Evaluation metrics for unsupervised learning algorithms. *arXiv preprint arXiv:1905.05667*. <https://doi.org/10.48550/ARXIV.1905.05667>
- Pan, S. J., & Yang, Q. (2010). A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, *22*(10), 1345–1359. <https://doi.org/10.1109/tkde.2009.191>
- Paplomata, E., & O'Regan, R. (2014). The PI3K/AKT/mTOR pathway in breast cancer: targets, trials and biomarkers. *Therapeutic Advances in Medical Oncology*, *6*(4), 154–166. <https://doi.org/10.1177/1758834014530023>
- Parekh, V., & Jacobs, M. A. (2016). Radiomics: A new application from established techniques [PMID: 28042608]. *Expert Review of Precision Medicine and Drug Development*, *1*(2), 207–226. <https://doi.org/10.1080/23808993.2016.1164013>
- Park, B. W., Kim, J. K., Heo, C., & Park, K. J. (2020). Reliability of CT radiomic features reflecting tumour heterogeneity according to image quality and image processing parameters. *Scientific Reports*, *10*(1). <https://doi.org/10.1038/s41598-020-60868-9>
- Park, H.-L., & Hong, J. (2014). Vacuum-assisted breast biopsy for breast cancer. *Gland surgery*, *3*, 120–7. <https://doi.org/10.3978/j.issn.2227-684X.2014.02.03>
- Parmar, N., Vaswani, A., Uszkoreit, J., Kaiser, L., Shazeer, N., Ku, A., & Tran, D. (2018). Image transformer. In J. Dy & A. Krause (Eds.), *Proceedings of the 35th international conference on machine learning* (pp. 4055–4064, Vol. 80). PMLR. <https://proceedings.mlr.press/v80/parmar18a.html>
- Pasini, G., Bini, F., Russo, G., Comelli, A., Marinozzi, F., & Stefano, A. (2022). Matradiomics: A novel and complete radiomics framework, from image visualization to predictive model. *Journal of Imaging*, *8*(8). <https://doi.org/10.3390/jimaging8080221>
- Penny, W., & Frost, D. (1996). Neural networks in clinical medicine [PMID: 8912300]. *Medical Decision Making*, *16*(4), 386–398. <https://doi.org/10.1177/0272989X9601600409>

- Petrelli, F., & Barni, S. (2011). Meta-analysis of concomitant compared to sequential adjuvant trastuzumab in breast cancer: The sooner the better. *Medical Oncology*, *29*(2), 503–510. <https://doi.org/10.1007/s12032-011-9897-9>
- Pezzotti, N., Yousefi, S., Elmahdy, M. S., Gemert, J. H. F. V., Schuelke, C., Doneva, M., Nielsen, T., Kastruyulin, S., Lelieveldt, B. P. F., Osch, M. J. P. V., Weerd, E. D., & Staring, M. (2020). An adaptive intelligence algorithm for undersampled knee MRI reconstruction. *IEEE Access*, *8*, 204825–204838. <https://doi.org/10.1109/access.2020.3034287>
- Pinder, S. E. (2010). Ductal carcinoma in situ (DCIS): pathological features, differential diagnosis, prognostic factors and specimen evaluation. *Modern Pathology*, *23*, S8–S13. <https://doi.org/10.1038/modpathol.2010.40>
- Pisano, E. D., & Yaffe, M. J. (2005). Digital Mammography. *Radiology*, *234*(2), 353–362. <https://doi.org/10.1148/radiol.2342030897>
- Prescott, J. W. (2012). Quantitative imaging biomarkers: The application of advanced image processing and analysis to clinical and preclinical decision making. *Journal of Digital Imaging*, *26*(1), 97–108. <https://doi.org/10.1007/s10278-012-9465-7>
- pyradiomics community. (2016). Pyradiomics Documentation. <https://pyradiomics.readthedocs.io>
- Qi, Z.-H., Xu, H.-X., Zhang, S.-R., Xu, J.-Z., Li, S., Gao, H.-L., Jin, W., Wang, W.-Q., Wu, C.-T., Ni, Q.-X., Yu, X.-J., & Liu, L. (2018). The Significance of Liquid Biopsy in Pancreatic Cancer. *Journal of Cancer*, *9*(18), 3417–3426. <https://doi.org/10.7150/jca.24591>
- Qian, P., Xu, K., Wang, T., Zheng, Q., Yang, H., Baydoun, A., Zhu, J., Traughber, B., & Muzic, R. F. (2020). Estimating CT from MR abdominal images using novel generative adversarial networks. *Journal of Grid Computing*, *18*(2), 211–226. <https://doi.org/10.1007/s10723-020-09513-3>
- Qiao, M., Li, C., Suo, S., Cheng, F., Hua, J., Xue, D., Guo, Y., Xu, J., & Wang, Y. (2020). Breast DCE-MRI radiomics: A robust computer-aided system based on reproducible BI-RADS features across the influence of datasets bias and segmentation methods. *International Journal of Computer Assisted Radiology and Surgery*, *15*(6), 921–930. <https://doi.org/10.1007/s11548-020-02177-0>
- Qin, Z., Liu, Z., Zhu, P., & Ling, W. (2022). Style transfer in conditional gans for cross-modality synthesis of brain magnetic resonance images. *Computers in Biology and Medicine*, *148*, 105928. <https://doi.org/10.1016/j.compbiomed.2022.105928>
- Qin, Z., Liu, Z., Zhu, P., & Xue, Y. (2020). A gan-based image synthesis method for skin lesion classification. *Computer Methods and Programs in Biomedicine*, *195*, 105568. <https://doi.org/10.1016/j.cmpb.2020.105568>
- Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). Learning transferable visual models from natural language supervision. In M. Meila & T. Zhang (Eds.), *Proceedings of the 38th international conference on machine learning* (pp. 8748–8763, Vol. 139). PMLR. <https://proceedings.mlr.press/v139/radford21a.html>

- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I., et al. (2019). Language models are unsupervised multitask learners. *OpenAI blog*, 1(8), 9. <https://insightcivic.s3.us-east-1.amazonaws.com/language-models.pdf>
- Rahib, L., Smith, B. D., Aizenberg, R., Rosenzweig, A. B., Fleshman, J. M., & Matrisian, L. M. (2014). Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. *Cancer research*, 74(11), 2913–2921. <https://doi.org/10.1158/0008-5472.CAN-14-0155>
- Rajkumar, A., & Agarwal, S. (2012). A differentially private stochastic gradient descent algorithm for multiparty classification. In N. D. Lawrence & M. Girolami (Eds.), *Proceedings of the fifteenth international conference on artificial intelligence and statistics* (pp. 933–941, Vol. 22). PMLR. <https://proceedings.mlr.press/v22/rajkumar12.html>
- Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., & Chen, M. (2022). Hierarchical text-conditional image generation with clip latents. *ArXiv preprint*, 1(2), 3. <https://3dvar.com/Ramesh2022Hierarchical.pdf>
- Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., Chen, M., & Sutskever, I. (2021). Zero-shot text-to-image generation. In M. Meila & T. Zhang (Eds.), *Proceedings of the 38th international conference on machine learning* (pp. 8821–8831, Vol. 139). PMLR. <https://proceedings.mlr.press/v139/ramesh21a.html>
- Rashidi, H. H., Tran, N. K., Betts, E. V., Howell, L. P., & Green, R. (2019). Artificial Intelligence and Machine Learning in Pathology: The Present Landscape of Supervised Methods. *Academic Pathology*, 6, 2374289519873088. <https://doi.org/10.1177/2374289519873088>
- Rathore, S., Akbari, H., Rozycki, M., Abdullah, K. G., Nasrallah, M. P., Binder, Z. A., Davuluri, R. V., Lustig, R. A., Dahmane, N., Bilello, M., O'Rourke, D. M., & Davatzikos, C. (2018). Radiomic MRI signature reveals three distinct subtypes of glioblastoma with different clinical and molecular characteristics, offering prognostic value beyond IDH1. *Scientific Reports*, 8(1), Article Number: 5087. <https://doi.org/10.1038/s41598-018-22739-2>
- Ravanbakhsh, M., Tschernetzki, V., Last, F., Klein, T., Batmanghelich, K., Tresp, V., & Nabi, M. (2020). Human-Machine Collaboration for Medical Image Segmentation. *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. <https://doi.org/10.1109/icassp40776.2020.9053555>
- Ren, C., Zhang, J., Qi, M., Zhang, J., Zhang, Y., Song, S., Sun, Y., & Cheng, J. (2020). Machine learning based on clinico-biological features integrated 18f-FDG PET/CT radiomics for distinguishing squamous cell carcinoma from adenocarcinoma of lung. *European Journal of Nuclear Medicine and Molecular Imaging*, 48(5), 1538–1549. <https://doi.org/10.1007/s00259-020-05065-6>
- Rengan, R., Cengel, K. A., & Hahn, S. M. (2008). Clinical target promiscuity: Lessons from ras molecular trials. *Cancer and Metastasis Reviews*, 27(3), 403–414. <https://doi.org/10.1007/s10555-008-9133-z>
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144. <https://doi.org/10.1145/2939672.2939778>

- Ribli, D., Horváth, A., Unger, Z., Pollner, P., & Csabai, I. (2018). Detecting and classifying lesions in mammograms with Deep Learning. *Scientific Reports*, *8*(1). <https://doi.org/10.1038/s41598-018-22437-z>
- Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H. R., Albarqouni, S., Bakas, S., Galtier, M. N., Landman, B. A., Maier-Hein, K., Ourselin, S., Sheller, M., Summers, R. M., Trask, A., Xu, D., Baust, M., & Cardoso, M. J. (2020). The future of digital health with federated learning. *npj Digital Medicine*, *3*(1). <https://doi.org/10.1038/s41746-020-00323-1>
- Rizzo, S., Botta, F., Raimondi, S., Origgi, D., Fanciullo, C., Morganti, A. G., & Bellomi, M. (2018). Radiomics: The facts and the challenges of image analysis. *European Radiology Experimental*, *2*(1). <https://doi.org/10.1186/s41747-018-0068-z>
- Rojas, K., & Stuckey, A. (2016). Breast Cancer Epidemiology and Risk Factors. *Clinical Obstetrics and Gynecology*, *59*(4), 651–672. <https://doi.org/10.1097/grf.0000000000000239>
- Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 10684–10695. https://openaccess.thecvf.com/content/CVPR2022/html/Rombach_High-Resolution_Image_Synthesis_With_Latent_Diffusion_Models_CVPR_2022_paper.html
- Romero, D. (2019). FOLFIRINOX goes adjuvant. *Nature Reviews Clinical Oncology*, *16*(3), 145–145. <https://doi.org/10.1038/s41571-019-0171-y>
- Rotalinti, Y., Tucker, A., Lonergan, M., Myles, P., & Branson, R. (2023). Detecting drift in healthcare ai models based on data availability. In I. Koprinska, P. Mignone, R. Guidotti, S. Jaroszewicz, H. Fröning, F. Gullo, P. M. Ferreira, D. Roqueiro, G. Ceddia, S. Nowaczyk, J. Gama, R. Ribeiro, R. Gavaldà, E. Masciari, Z. Ras, E. Ritacco, F. Naretto, A. Theissler, P. Biecek, . . . S. Pashami (Eds.), *Machine learning and principles and practice of knowledge discovery in databases* (pp. 243–258). Springer Nature Switzerland. https://link.springer.com/chapter/10.1007/978-3-031-23633-4_17
- Rousseaux, C. G., & Gad, S. C. (2013). Chapter 30 - statistical assessment of toxicologic pathology studies. In W. M. Haschek, C. G. Rousseaux, & M. A. Wallig (Eds.), *Haschek and rousseaux's handbook of toxicologic pathology (third edition)* (Third Edition, pp. 893–988). Academic Press. <https://doi.org/10.1016/B978-0-12-415759-0.00030-3>
- Roy, M., Wang, F., Vo, H., Teng, D., Teodoro, G., Farris, A. B., Castillo-Leon, E., Vos, M. B., & Kong, J. (2020). Deep-learning-based accurate hepatic steatosis quantification for histological assessment of liver biopsies. *Laboratory Investigation*. <https://doi.org/10.1038/s41374-020-0463-y>
- Ruder, S. (2016). An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*. <https://arxiv.org/pdf/1609.04747.pdf>
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., & Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, *115*(3), 211–252. <https://doi.org/10.1007/s11263-015-0816-y>

- Ryan, R., Tawfik, O., Jensen, R. A., & Anant, S. (2017). Chapter Two - Current Approaches to Diagnosis and Treatment of Ductal Carcinoma In Situ and Future Directions. In R. Lakshmanaswamy (Ed.), *Approaches to understanding breast cancer* (pp. 33–80, Vol. 151). Academic Press. <https://doi.org/10.1016/bs.pmbts.2017.08.001>
- Samek, W., & Müller, K.-R. (2019). Towards Explainable Artificial Intelligence. In W. Samek, G. Montavon, A. Vedaldi, L. K. Hansen, & K.-R. Müller (Eds.), *Explainable ai: Interpreting, explaining and visualizing deep learning* (pp. 5–22). Springer International Publishing. https://doi.org/10.1007/978-3-030-28954-6_1
- Sanches, J. M., Laine, A. F., & Suri, J. S. (Eds.). (2012). *Ultrasound imaging: Advances and applications* (1st ed.). Springer US. <https://doi.org/10.1007/978-1-4614-1180-2>
- Sapate, S. G., Mahajan, A., Talbar, S. N., Sable, N., Desai, S., & Thakur, M. (2018). Radiomics based detection and characterization of suspicious lesions on full field digital mammograms. *Computer Methods and Programs in Biomedicine*, *163*, 1–20. <https://doi.org/10.1016/j.cmpb.2018.05.017>
- Saravanan, R., & Sujatha, P. (2018). A state of art techniques on machine learning algorithms: A perspective of supervised learning approaches in data classification. *2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS)*, 945–949. <https://doi.org/10.1109/ICCONS.2018.8663155>
- Sardanelli, F., Giuseppetti, G. M., Panizza, P., Bazzocchi, M., Fausto, A., Simonetti, G., Lattanzio, V., & Del Maschio, A. (2004). Sensitivity of mri versus mammography for detecting foci of multifocal, multicentric breast cancer in fatty and dense breasts using the whole-breast pathologic examination as a gold standard [PMID: 15385322]. *American Journal of Roentgenology*, *183*(4), 1149–1157. <https://doi.org/10.2214/ajr.183.4.1831149>
- Sarvamangala, D. R., & Kulkarni, R. V. (2021). Convolutional neural networks in medical image understanding: A survey. *Evolutionary Intelligence*, *15*(1), 1–22. <https://doi.org/10.1007/s12065-020-00540-3>
- Sarwar, S., Dent, A., Faust, K., Richer, M., Djuric, U., Ommeren, R. V., & Diamandis, P. (2019). Physician perspectives on integration of artificial intelligence into diagnostic pathology. *npj Digital Medicine*, *2*(1). <https://doi.org/10.1038/s41746-019-0106-0>
- Scheufele, F., Hartmann, D., & Friess, H. (2019). Treatment of pancreatic cancer: adjuvant treatment in borderline resectable/locally advanced pancreatic cancer. *Translational Gastroenterology and Hepatology*, *4*, 32–32. <https://doi.org/10.21037/tgh.2019.04.09>
- Schlegl, T., Seeböck, P., Waldstein, S. M., Langs, G., & Schmidt-Erfurth, U. (2019). F-anogan: Fast unsupervised anomaly detection with generative adversarial networks. *Medical Image Analysis*, *54*, 30–44. <https://doi.org/10.1016/j.media.2019.01.010>
- Schlitter, A. M., Kasajima, A., GroSS, C., Konukiewitz, B., & Klöppel, G. (2019). Pathologie des PDAC: Übersicht und Neues. *InFo Hämatologie + Onkologie*, *22*(11), 10–15. <https://doi.org/10.1007/s15004-019-6744-1>
- Schlitter, A. M., Segler, A., Steiger, K., Michalski, C. W., Jäger, C., Konukiewitz, B., Pfarr, N., Endris, V., Bettstetter, M., Kong, B., Regel, I., Kleeff, J., Klöppel, G., & Esposito, I. (2017). Molecular,

- morphological and survival analysis of 177 resected pancreatic ductal adenocarcinomas (PDACs): Identification of prognostic subtypes. *Scientific Reports*, 7(1). <https://doi.org/10.1038/srep41064>
- Schmidt, M. K., Hogervorst, F., van Hien, R., Cornelissen, S., Broeks, A., Adank, M. A., Meijers, H., Waisfisz, Q., Hollestelle, A., Schutte, M., van den Ouweland, A., Hoening, M., Andrulis, I. L., Anton-Culver, H., Antonenkova, N. N., Antoniou, A. C., Arndt, V., Bermisheva, M., Bogdanova, N. V., . . . Easton, D. F. (2016). Age- and Tumor Subtype-Specific Breast Cancer Risk Estimates for CHEK2*1100delC Carriers. *Journal of Clinical Oncology*, 34(23), 2750–2760. <https://doi.org/10.1200/jco.2016.66.5844>
- Schneier, B. (2015). *Secrets and lies: Digital security in a networked world*. John Wiley & Sons. https://terrorgum.com/tfox/books/secrets_and_lies_digital_security_in_a_networked_world.pdf
- Segal, E., Sirlin, C. B., Ooi, C., Adler, A. S., Gollub, J., Chen, X., Chan, B. K., Matcuk, G. R., Barry, C. T., Chang, H. Y., & Kuo, M. D. (2007). Decoding global gene expression programs in liver cancer by noninvasive imaging. *Nature Biotechnology*, 25(6), 675–680. <https://doi.org/10.1038/nbt1306>
- Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual Explanations From Deep Networks via Gradient-Based Localization. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 618–626. https://openaccess.thecvf.com/content_iccv_2017/html/Selvaraju_Grad-CAM_Visual_Explanations_ICCV_2017_paper.html
- Seufferlein, T., & Ettrich, T. J. (2019). Treatment of pancreatic cancerneoadjuvant treatment in resectable pancreatic cancer (PDAC). *Translational Gastroenterology and Hepatology*, 4, 21–21. <https://doi.org/10.21037/tgh.2019.03.05>
- Sevakula, R. K., Au-Yeung, W.-T. M., Singh, J. P., Heist, E. K., Isselbacher, E. M., & Armoundas, A. A. (2020). State-of-the-art machine learning techniques aiming to improve patient outcomes pertaining to the cardiovascular system. *Journal of the American Heart Association*, 9(4), e013924. <https://doi.org/10.1161/JAHA.119.013924>
- Shahbandi, A., Nguyen, H. D., & Jackson, J. G. (2020). TP53 Mutations and Outcomes in Breast Cancer: Reading beyond the Headlines. *Trends in Cancer*, 6(2), 98–110. <https://doi.org/10.1016/j.trecan.2020.01.007>
- Shahriar, S. (2022). Gan computers generate arts? a survey on visual arts, music, and literary text generation using generative adversarial network. *Displays*, 73, 102237. <https://doi.org/10.1016/j.displa.2022.102237>
- Shaikh, F., Franc, B., Allen, E., Sala, E., Awan, O., Hendrata, K., Halabi, S., Mohiuddin, S., Malik, S., Hadley, D., & Shrestha, R. (2018). Translational radiomics: Defining the strategy pipeline and considerations for application-part 1: From methodology to clinical implementation [Data Science: Big Data Machine Learning and Artificial Intelligence]. *Journal of the American College of Radiology*, 15(3, Part B), 538–542. <https://doi.org/10.1016/j.jacr.2017.12.008>
- Sharma, G. N., Dave, R., Sanadya, J., Sharma, P., Sharma, K., et al. (2010). Various types and management of breast cancer: An overview. *Journal of advanced pharmaceutical technology & research*, 1(2), 109. https://japtr.org/citation.asp?issn=2231-4040;year=2010;volume=1;issue=2;spage=109;epage=126;aualast=Sharma;aid=JAdvPharmTechRes_2010_1_2_109_72251

- Sharma, S., Sharma, S., & Athaiya, A. (2020). Activation functions in neural networks. *International Journal of Engineering Applied Sciences and Technology*, 4(12), 310–316. <https://www.ijeast.com/papers/310-316,Tesma412,IJEAST.pdf>
- Sheller, M. J., Edwards, B., Reina, G. A., Martin, J., Pati, S., Kotrotsou, A., Milchenko, M., Xu, W., Marcus, D., Colen, R. R., & Bakas, S. (2020). Federated learning in medicine: Facilitating multi-institutional collaborations without sharing patient data. *Scientific Reports*, 10(1), 1–12. <https://doi.org/10.1038/s41598-020-69250-1>
- Shen, Y., Heacock, L., Elias, J., Hentel, K. D., Reig, B., Shih, G., & Moy, L. (2023). Chatgpt and other large language models are double-edged swords [PMID: 36700838]. *Radiology*, 307(2), e230163. <https://doi.org/10.1148/radiol.230163>
- Shin, H.-C., Lu, L., & Summers, R. M. (2017). Chapter 17 - natural language processing for large-scale medical image analysis using deep learning. In S. K. Zhou, H. Greenspan, & D. Shen (Eds.), *Deep learning for medical image analysis* (pp. 405–421). Academic Press. <https://doi.org/10.1016/B978-0-12-810408-8.00023-7>
- Shin, H.-C., Roberts, K., Lu, L., Demner-Fushman, D., Yao, J., & Summers, R. M. (2016). Learning to read chest x-rays: Recurrent neural cascade model for automated image annotation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2497–2506. https://openaccess.thecvf.com/content_cvpr_2016/html/Shin_Learning_to_Read_CVPR_2016_paper.html
- Shitrit, O., & Riklin Raviv, T. (2017, September). Accelerated magnetic resonance imaging by adversarial neural network. In M. J. Cardoso, T. Arbel, G. Carneiro, T. Syeda-Mahmood, J. M. R. S. Tavares, M. Moradi, A. Bradley, H. Greenspan, J. P. Papa, A. Madabhushi, J. C. Nascimento, J. S. Cardoso, V. Belagiannis, & Z. Lu (Eds.), *Deep learning in medical image analysis and multimodal learning for clinical decision support* (pp. 30–38, Vol. 10553). Springer International Publishing. https://link.springer.com/chapter/10.1007/978-3-319-67558-9_4
- Shofty, B., Artzi, M., Shtrozberg, S., Fanizzi, C., DiMeco, F., Haim, O., Hason, S. P., Ram, Z., Bashat, D. B., & Grossman, R. (2020). Virtual biopsy using MRI radiomics for prediction of BRAF status in melanoma brain metastasis. *Scientific Reports*, 10(1). <https://doi.org/10.1038/s41598-020-63821-y>
- Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, 6(1). <https://doi.org/10.1186/s40537-019-0197-0>
- Siegel, R. L., Miller, K. D., & Jemal, A. (2019). Cancer statistics, 2019. *CA: A Cancer Journal for Clinicians*, 69(1), 7–34. <https://doi.org/10.3322/caac.21551>
- Siegel, R. L., Miller, K. D., Wagle, N. S., & Jemal, A. (2023). Cancer statistics, 2023. *CA: A Cancer Journal for Clinicians*, 73(1), 17–48. <https://doi.org/10.3322/caac.21763>
- Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., Chen, Y., Lillicrap, T., Hui, F., Sifre, L., van den Driessche, G., Graepel, T., & Hassabis, D. (2017). Mastering the game of Go without human knowledge. *Nature*, 550(7676), 354–359. <https://doi.org/10.1038/nature24270>

- Simard, P. Y., Steinkraus, D., Platt, J. C., et al. (2003). Best practices for convolutional neural networks applied to visual document analysis. *Proceedings of the Seventh International Conference on Document Analysis and Recognition*, 3(2003), 958–963. <https://doi.org/10.1109/ICDAR.2003.1227801>
- Sims, A. H., Howell, A., Howell, S. J., & Clarke, R. B. (2007). Origins of breast cancer subtypes and therapeutic implications. *Nature Clinical Practice Oncology*, 4(9), 516–525. <https://doi.org/10.1038/ncponc0908>
- Singh, N. K., & Raza, K. (2021). Medical image generation using generative adversarial networks: A review. In R. Patgiri, A. Biswas, & P. Roy (Eds.), *Health informatics: A computational perspective in healthcare* (pp. 77–96). Springer Singapore. https://doi.org/10.1007/978-981-15-9735-0_5
- Singh, S. P., Wang, L., Gupta, S., Goli, H., Padmanabhan, P., & Gulyás, B. (2020). 3d deep learning on medical images: A review. *Sensors*, 20(18). <https://doi.org/10.3390/s20185097>
- Sinn, B. V., Striefler, J. K., Rudl, M. A., Lehmann, A., Bahra, M., Denkert, C., Sinn, M., Stieler, J., Klauschen, F., Budczies, J., Weichert, W., Stenzinger, A., Kamphues, C., Dietel, M., & Riess, H. (2014). KRAS Mutations in Codon 12 or 13 Are Associated With Worse Prognosis in Pancreatic Ductal Adenocarcinoma. *Pancreas*, 43(4), 578–583. <https://doi.org/10.1097/mpa.0000000000000077>
- Smith, L. N. (2017). Cyclical Learning Rates for Training Neural Networks. *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 464–472. <https://doi.org/10.1109/wacv.2017.58>
- Söder, M. (2019, October). Hightech Agenda Bayern [Retrieved from: <https://www.bayern.de/hightech-agenda-bayern> on 15th January 2020]. <https://www.bayern.de/hightech-agenda-bayern>
- Soffer, S., Ben-Cohen, A., Shimon, O., Amitai, M. M., Greenspan, H., & Klang, E. (2019). Convolutional Neural Networks for Radiologic Images: A Radiologist's Guide. *Radiology*, 290(3), 590–606. <https://doi.org/10.1148/radiol.2018180547>
- Soin, A., Merkow, J., Long, J., Cohen, J. P., Saligrama, S., Kaiser, S., Borg, S., Tarapov, I., & Lungren, M. P. (2022). Chexstray: Real-time multi-modal data concordance for drift detection in medical imaging ai. <https://doi.org/10.48550/ARXIV.2202.02833>
- Soloman, A. (1913). Beitrage zur Pathologie und Klinik des Mammakarzinoms. *Arch Klin Chir*, 101, 573–668.
- Spak, D. A., Plaxco, J. S., Santiago, L., Dryden, M. J., & Dogan, B. E. (2017). BI-RADS fifth edition: A summary of changes. *Diagnostic and Interventional Imaging*, 98(3), 179–190. <https://doi.org/10.1016/j.diii.2017.01.001>
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1), 1929–1958. http://www.jmlr.org/papers/volume15/srivastava14a/srivastava14a.pdf?utm_content=buffer79b43&utm_medium=social&utm_source=twitter.com&utm_campaign=buffer
- Starker, A., Kraywinkel, K., & Kuhnert, R. (2017). Früherkennung von Brustkrebs: Inanspruchnahme der Mammografie in Deutschland. In *Journal of health monitoring* (pp. 74–80, Vol. 2). Robert Koch-Institut, Epidemiologie und Gesundheitsberichterstattung. <https://doi.org/10.17886/RKI-GBE-2017-114>
- Stocken, D. D., Büchler, M. W., Dervenis, C., Bassi, C., Jeekel, H., Klinkenbijn, J. H. G., Bakkevold, K. E., Takada, T., Amano, H., & Neoptolemos, J. P. (2005). Meta-analysis of randomised adjuvant therapy

- trials for pancreatic cancer. *British Journal of Cancer*, 92(8), 1372–1381. <https://doi.org/10.1038/sj.bjc.6602513>
- Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., Liu, B., Matthews, P., Ong, G., Pell, J., Silman, A., Young, A., Sprosen, T., Peakman, T., & Collins, R. (2015). UK biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS medicine*, 12, e1001779. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4380465/>
- Sukhbaatar, S., Grave, E., Lample, G., Jegou, H., & Joulin, A. (2019). Augmenting self-attention with persistent memory. <https://doi.org/10.48550/ARXIV.1907.01470>
- Sultana, F., Sufian, A., & Dutta, P. (2018, November). Advancements in image classification using convolutional neural network. In S. Bhattacharyya, V. Snasel, D. De, I. Pan, S. Dhar, D. Bhattacharjee, & H. Bhaumik (Eds.), *2018 fourth IEEE international conference on research in computational intelligence and communication networks (icrcicn)* (pp. 122–129). IEEE. <https://doi.org/10.1109/ICRCICN.2018.8718718>
- Sun, H., Jiang, Y., Yuan, J., Wang, H., Liang, D., Fan, W., Hu, Z., & Zhang, N. (2022). High-quality PET image synthesis from ultra-low-dose PET/MRI using bi-task deep learning. *Quantitative Imaging in Medicine and Surgery*, 12, 5326–5342. <https://doi.org/10.21037/qims-22-116>
- Sun, Q., Lin, X., Zhao, Y., Li, L., Yan, K., Liang, D., Sun, D., & Li, Z.-C. (2020). Deep Learning vs. Radiomics for Predicting Axillary Lymph Node Metastasis of Breast Cancer Using Ultrasound Images: Don't Forget the Peritumoral Region. *Frontiers in Oncology*, 10. <https://doi.org/10.3389/fonc.2020.00053>
- Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, 71(3), 209–249. <https://doi.org/10.3322/caac.21660>
- Sutskever, I., Jozefowicz, R., Gregor, K., Rezaeifar, D., Lillicrap, T., & Vinyals, O. (2015). Towards principled unsupervised learning. <https://doi.org/10.48550/ARXIV.1511.06440>
- Tai, A. M. Y., Albuquerque, A., Carmona, N. E., Subramanieapillai, M., Cha, D. S., Sheko, M., Lee, Y., Mansur, R., & McIntyre, R. S. (2019). Machine learning and big data: Implications for disease modeling and therapeutic discovery in psychiatry. *Artificial Intelligence in Medicine*, 99, 101704. <https://doi.org/10.1016/j.artmed.2019.101704>
- Tempero, M. A., Malafa, M. P., Al-Hawary, M., Asbun, H., Bain, A., Behrman, S. W., Benson, A. B., Binder, E., Cardin, D. B., Cha, C., Chiorean, E. G., Chung, V., Czito, B., Dillhoff, M., Dotan, E., Ferrone, C. R., Hardacre, J., Hawkins, W. G., Herman, J., ... Darlow, S. (2017). Pancreatic Adenocarcinoma, Version 2.2017, NCCN Clinical Practice Guidelines in Oncology. *Journal of the National Comprehensive Cancer Network J Natl Compr Canc Netw*, 15(8), 1028–1061. <https://doi.org/10.6004/jnccn.2017.0131>
- The Cancer Genome Atlas Network. (2012). Comprehensive molecular portraits of human breast tumours. *Nature*, 490(7418), 61–70. <https://doi.org/10.1038/nature11412>

- The Cancer Imaging Archive. (2019). TCIA Collection [Retrieved from: <https://www.cancerimagingarchive.net>]. <https://www.cancerimagingarchive.net>
- Thiebes, S., Lins, S., & Sunyaev, A. (2020). Trustworthy artificial intelligence. *Electron Markets*, *31*(2), 447–464. <https://doi.org/10.1007/s12525-020-00441-4>
- Thomson, J. Z., Evans, A. J., Pinder, S. E., Burrell, H. C., Wilson, A. R. M., & Ellis, I. O. (2001). Growth pattern of ductal carcinoma in situ (DCIS): A retrospective analysis based on mammographic findings. *British Journal of Cancer*, *85*(2), 225–227. <https://doi.org/10.1054/bjoc.2001.1877>
- Tirkes, T., Hollar, M. A., Tann, M., Kohli, M. D., Akisik, F., & Sandrasegaran, K. (2013). Response criteria in oncologic imaging: Review of traditional and new criteria [PMID: 24025927]. *RadioGraphics*, *33*(5), 1323–1341. <https://doi.org/10.1148/rg.335125214>
- Tixier, F., Hatt, M., Valla, C., Fleury, V., Lamour, C., Ezzouhri, S., Ingrand, P., Perdrisot, R., Visvikis, D., & Rest, C. C. L. (2014). Visual Versus Quantitative Assessment of Intratumor 18F-FDG PET Uptake Heterogeneity: Prognostic Value in Non-Small Cell Lung Cancer. *Journal of Nuclear Medicine*, *55*(8), 1235–1241. <https://doi.org/10.2967/jnumed.113.133389>
- Tomori, Y., Yamashiro, T., Tomita, H., Tsubakimoto, M., Ishigami, K., Atsumi, E., & Murayama, S. (2020). Ct radiomics analysis of lung cancers: Differentiation of squamous cell carcinoma from adenocarcinoma, a correlative study with fdg uptake. *European Journal of Radiology*, *128*, 109032. <https://doi.org/10.1016/j.ejrad.2020.109032>
- Tong, M., Wang, J., Zhang, H., Xing, H., Wang, Y., Fang, Y., Pan, H., & Li, D. (2019). Efficacy and safety of gemcitabine plus anti-angiogenesis therapy for advanced pancreatic cancer: A systematic review and meta-analysis of clinical randomized phase III trials. *Journal of Cancer*, *10*(4), 968–978. <https://doi.org/10.7150/jca.26672>
- Torre, L. A., Islami, F., Siegel, R. L., Ward, E. M., & Jemal, A. (2017). Global Cancer in Women: Burden and Trends. *Cancer Epidemiology, Biomarkers & Prevention*, *26*(4), 444–457. <https://doi.org/10.1158/1055-9965.epi-16-0858>
- Traverso, A., Wee, L., Dekker, A., & Gillies, R. (2018). Repeatability and reproducibility of radiomic features: A systematic review. *International Journal of Radiation Oncology*Biology*Physics*, *102*(4), 1143–1158. <https://doi.org/10.1016/j.ijrobp.2018.05.053>
- Tschandl, P., Codella, N., Akay, B. N., Argenziano, G., Braun, R. P., Cabo, H., Gutman, D., Halpern, A., Helba, B., Hofmann-Wellenhof, R., Lallas, A., Lapins, J., Longo, C., Malvey, J., Marchetti, M. A., Marghoob, A., Menzies, S., Oakley, A., Paoli, J., . . . Kittler, H. (2019). Comparison of the accuracy of human readers versus machine-learning algorithms for pigmented skin lesion classification: An open, web-based, international, diagnostic study. *The Lancet Oncology*, *20*(7), 938–947. [https://doi.org/10.1016/s1470-2045\(19\)30333-x](https://doi.org/10.1016/s1470-2045(19)30333-x)
- Tschandl, P., Rinner, C., Apalla, Z., Argenziano, G., Codella, N., Halpern, A., Janda, M., Lallas, A., Longo, C., Malvey, J., Paoli, J., Puig, S., Rosendahl, C., Soyer, H. P., Zalaudek, I., & Kittler, H. (2020). Humancomputer collaboration for skin cancer recognition. *Nature Medicine*. <https://doi.org/10.1038/s41591-020-0942-0>

- Tsimberidou, A.-M. (2015). Targeted therapy in cancer. *Cancer Chemotherapy and Pharmacology*, 76(6), 1113–1132. <https://doi.org/10.1007/s00280-015-2861-1>
- Tursun, O., Denman, S., Sridharan, S., & Fookes, C. (2023, April). Towards self-explainability of deep neural networks with heatmap captioning and large-language models. <https://doi.org/10.48550/ARXIV.2304.02202>
- van den Oord, A., Kalchbrenner, N., Espeholt, L., Kavukcuoglu koray, k., Vinyals, O., & Graves, A. (2016). Conditional image generation with pixelcnn decoders. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, & R. Garnett (Eds.), *Advances in neural information processing systems* (Vol. 29). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2016/file/b1301141feffabac455e1f90a7de2054-Paper.pdf
- van den Oord, A., Kalchbrenner, N., & Kavukcuoglu, K. (2016). Pixel recurrent neural networks. In M. F. Balcan & K. Q. Weinberger (Eds.), *Proceedings of the 33rd international conference on machine learning* (pp. 1747–1756, Vol. 48). PMLR. <https://proceedings.mlr.press/v48/oord16.html>
- van Dijk, L. V., Brouwer, C. L., van der Schaaf, A., Burgerhof, J. G. M., Beukinga, R. J., Langendijk, J. A., Sijtsema, N. M., & Steenbakkers, R. J. H. M. (2017). Ct image biomarkers to improve patient-specific prediction of radiation-induced xerostomia and sticky saliva. *Radiotherapy and Oncology*, 122(2), 185–191. <https://doi.org/10.1016/j.radonc.2016.07.007>
- van Engelen, J. E., & Hoos, H. H. (2019). A survey on semi-supervised learning. *Machine Learning*, 109(2), 373–440. <https://doi.org/10.1007/s10994-019-05855-6>
- van Griethuysen, J. J. M., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., Beets-Tan, R. G. H., Fillion-Robin, J.-C., Pieper, S., & Aerts, H. J. W. L. (2017). Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Research*, 77(21), e104–e107. <https://doi.org/10.1158/0008-5472.can-17-0339>
- van Manen, L., Groen, J. V., Putter, H., Vahrmeijer, A. L., Swijnenburg, R.-J., Bonsing, B. A., & Mieog, J. S. D. (2020). Elevated CEA and CA19-9 serum levels independently predict advanced pancreatic cancer at diagnosis [PMID: 32009482]. *Biomarkers*, 25(2), 186–193. <https://doi.org/10.1080/1354750X.2020.1725786>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, . u., & Polosukhin, I. (2017). Attention is all you need. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in neural information processing systems* (Vol. 30). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf
- Veit, A., Wilber, M. J., & Belongie, S. (2016, December). Residual Networks Behave Like Ensembles of Relatively Shallow Networks. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, & R. Garnett (Eds.), *Advances in neural information processing systems* (pp. 550–558, Vol. 29). Curran Associates, Inc. <https://proceedings.neurips.cc/paper/2016/file/37bc2f75bf1bcfe8450a1a41c200364c-Paper.pdf>
- Versteijne, E., Suker, M., Groothuis, K., Akkermans-Vogelaar, J. M., Besselink, M. G., Bonsing, B. A., Buijsen, J., Busch, O. R., Creemers, G.-J. M., van Dam, R. M., Eskens, F. A. L. M., Festen, S., de Groot, J. W. B., Koerkamp, B. G., de Hingh, I. H., Homs, M. Y. V., van Hooft, J. E., Kerver, E. D., Luelmo, S. A. C., . . . van Tienhoven, G. (2020). Preoperative Chemoradiotherapy Versus

Immediate Surgery for Resectable and Borderline Resectable Pancreatic Cancer: Results of the Dutch Randomized Phase III PREOPANC Trial. *Journal of Clinical Oncology*, 38(16), 1763–1773. <https://doi.org/10.1200/jco.19.02274>

Vinyals, O., Toshev, A., Bengio, S., & Erhan, D. (2015). Show and tell: A neural image caption generator. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3156–3164. Retrieved June 7, 2015, from https://www.cv-foundation.org/openaccess/content_cvpr_2015/html/Vinyals_Show_and_Tell_2015_CVPR_paper.html

Vogel, W. (1932). Die Roentgendarstellung der Mammatumoren. *Arch Klin Chir*, 171, 618–626.

von Minckwitz, G., Untch, M., Nüesch, E., Loibl, S., Kaufmann, M., Kümmel, S., Fasching, P. A., Eiermann, W., Blohmer, J.-U., Costa, S. D., Mehta, K., Hilfrich, J., Jackisch, C., Gerber, B., du Bois, A., Huober, J., Hanusch, C., Konecny, G., Fett, W., . . . Jüni, P. (2010). Impact of treatment characteristics on response of different breast cancer phenotypes: Pooled analysis of the german neo-adjuvant chemotherapy trials. *Breast Cancer Research and Treatment*, 125(1), 145–156. <https://doi.org/10.1007/s10549-010-1228-x>

Waddell, N., Pajic, M., Patch, A.-M., Chang, D. K., Kassahn, K. S., Bailey, P., Johns, A. L., Miller, D., Nones, K., Quek, K., Quinn, M. C. J., Robertson, A. J., Fadlullah, M. Z. H., Bruxner, T. J. C., Christ, A. N., Harliwong, I., Idrisoglu, S., Manning, S., Nourse, C., . . . Grimmond, S. M. (2015). Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature*, 518(7540), 495–501. <https://doi.org/10.1038/nature14169>

Waheed, A., Goyal, M., Gupta, D., Khanna, A., Al-Turjman, F., & Pinheiro, P. R. (2020). Covidgan: Data augmentation using auxiliary classifier gan for improved covid-19 detection. *IEEE Access*, 8, 91916–91923. <https://doi.org/10.1109/ACCESS.2020.2994762>

Wang, C., Ma, J., Shao, J., Zhang, S., Liu, Z., Yu, Y., & Li, W. (2022). Predicting egfr and pd-11 status in nslc patients using multitask ai system based on ct images. *Frontiers in Immunology*, 13. <https://doi.org/10.3389/fimmu.2022.813072>

Wang, G., Gong, E., Banerjee, S., Martin, D., Tong, E., Choi, J., Chen, H., Wintermark, M., Pauly, J. M., & Zaharchuk, G. (2020). Synthesize high-quality multi-contrast magnetic resonance imaging from multi-echo acquisition using multi-task deep generative model. *IEEE Transactions on Medical Imaging*, 39(10), 3089–3099. <https://doi.org/10.1109/tmi.2020.2987026>

Wang, X., Yang, W., Weinreb, J., Han, J., Li, Q., Kong, X., Yan, Y., Ke, Z., Luo, B., Liu, T., & Wang, L. (2017). Searching for prostate cancer by fully automated magnetic resonance imaging classification: Deep learning versus non-deep learning. *Scientific Reports*, 7(1). <https://doi.org/10.1038/s41598-017-15720-y>

Warren, S. L. (1930). Roentgenologic study of the breast. *American Journal of Roentgenology*, 24, 113–124.

Watkins, E. J. (2019). Overview of breast cancer. *JAAPA*, 32(10), 13–17. <https://doi.org/10.1097/01.jaa.0000580524.95733.3d>

Wei, C., Wang, Y.-C., Wang, B., & Kuo, C. C. J. (2023, March). An overview on language models: Recent developments and outlook. <https://doi.org/10.48550/ARXIV.2303.05759>

- Weigel, S., Heindel, W., Heidrich, J., Hense, H.-W., & Heidinger, O. (2016). Digital mammography screening: Sensitivity of the programme dependent on breast density. *European Radiology*, *27*(7), 2744–2751. <https://doi.org/10.1007/s00330-016-4636-4>
- Wickramasinghe, C. S., Marino, D. L., Grandio, J., & Manic, M. (2020). Trustworthy AI Development Guidelines for Human System Interaction. *2020 13th International Conference on Human System Interaction (HSI)*, 130–136. <https://doi.org/10.1109/hsi49210.2020.9142644>
- Wiechmann, L., & Kuerer, H. M. (2008). The molecular journey from ductal carcinoma in situ to invasive breast cancer. *Cancer*, *112*(10), 2130–2142. <https://doi.org/10.1002/cncr.23430>
- Winters, S., Martin, C., Murphy, D., & Shokar, N. K. (2017). Breast Cancer Epidemiology, Prevention, and Screening. In R. Lakshmanaswamy (Ed.), *Approaches to understanding breast cancer* (pp. 1–32, Vol. 151). Academic Press. <https://doi.org/10.1016/bs.pmbts.2017.07.002>
- Wolterink, J. M., Leiner, T., Viergever, M. A., & Isgum, I. (2017). Generative adversarial networks for noise reduction in low-dose CT. *IEEE Transactions on Medical Imaging*, *36*(12), 2536–2545. <https://doi.org/10.1109/tmi.2017.2708987>
- Wong, K. K. L., Fortino, G., & Abbott, D. (2020). Deep learning-based cardiovascular image diagnosis: A promising challenge. *Future Generation Computer Systems*, *110*, 802–811. <https://doi.org/10.1016/j.future.2019.09.047>
- Woznicki, P., Westhoff, N., Huber, T., Riffel, P., Froelich, M. F., Gresser, E., von Hardenberg, J., Mühlberg, A., Michel, M. S., Schoenberg, S. O., & Nörenberg, D. (2023). Multiparametric mri for prostate cancer characterization: Combined use of radiomics model with pi-rads and clinical parameters. *Cancers*, *12*(7). <https://doi.org/10.3390/cancers12071767>
- Wu, R., Guo, C., & Chaudhuri, K. (2023, August). Large-scale public data improves differentially private image generation quality. <https://doi.org/10.48550/ARXIV.2309.00008>
- Xia, Y., Ravikumar, N., Greenwood, J. P., Neubauer, S., Petersen, S. E., & Frangi, A. F. (2021). Super-resolution of cardiac mr cine imaging using conditional gans and unsupervised transfer learning. *Medical Image Analysis*, *71*, 102037. <https://doi.org/10.1016/j.media.2021.102037>
- Xu, X., Chen, X., Zhang, J., Zheng, Y., Sun, G., Yu, X., & Xu, B. (2014). Comparison of the tada formula with software slicer. *Stroke*, *45*(11), 3433–3435. <https://doi.org/10.1161/STROKEAHA.114.007095>
- Xu, Y., Li, Y., & Shin, B.-S. (2020). Medical image processing with contextual style transfer. *Human-centric Computing and Information Sciences*, *10*(1). <https://doi.org/10.1186/s13673-020-00251-9>
- Xue, C., Yuan, J., Lo, G. G., Chang, A. T. Y., Poon, D. M. C., Wong, O. L., Zhou, Y., & Chu, W. C. W. (2021). Radiomics feature reliability assessed by intraclass correlation coefficient: A systematic review. *Quantitative Imaging in Medicine and Surgery*, *11*(10), 4431–4460. <https://doi.org/10.21037/qims-21-86>
- Yamauchi, H., Woodward, W. A., Valero, V., Alvarez, R. H., Lucci, A., Buchholz, T. A., Iwamoto, T., Krishnamurthy, S., Yang, W., Reuben, J. M., Hortobágyi, G. N., & Ueno, N. T. (2012). Inflammatory Breast Cancer: What We Know and What We Need to Learn. *The Oncologist*, *17*(7), 891–899. <https://doi.org/10.1634/theoncologist.2012-0039>

- Yang, Q., Steinfeld, A., & Zimmerman, J. (2019). Unremarkable AI. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*. <https://doi.org/10.1145/3290605.3300468>
- Yarden, Y. (2001). Biology of HER2 and Its Importance in Breast Cancer. *Oncology*, *61*(Suppl. 2), 1–13. <https://doi.org/10.1159/000055396>
- Yates, F. (1934). Contingency tables involving small numbers and the χ^2 test. *Supplement to the Journal of the Royal Statistical Society*, *1*(2), 217–235. Retrieved June 6, 2023, from <http://www.jstor.org/stable/2983604>
- Yi, X., & Babyn, P. (2018). Sharpness-aware low-dose CT denoising using conditional generative adversarial network. *Journal of Digital Imaging*, *31*(5), 655–669. <https://doi.org/10.1007/s10278-018-0056-0>
- Yi, X., Walia, E., & Babyn, P. (2019). Generative adversarial network in medical imaging: A review. *Medical Image Analysis*, *58*, 101552. <https://doi.org/10.1016/j.media.2019.101552>
- Ying, X. (2019). An overview of overfitting and its solutions. *Journal of Physics: Conference Series*, *1168*(2), 022022. <https://doi.org/10.1088/1742-6596/1168/2/022022>
- Yoon, J., Drumright, L. N., & van der Schaar, M. (2020). Anonymization through data synthesis using generative adversarial networks (ADS-GAN). *IEEE Journal of Biomedical and Health Informatics*, *24*(8), 2378–2388. <https://doi.org/10.1109/jbhi.2020.2980262>
- Yu, B., Wang, Y., Wang, L., Shen, D., & Zhou, L. (2020). Medical image synthesis via deep learning. In G. Lee & H. Fujita (Eds.), *Deep learning in medical image analysis : Challenges and applications* (pp. 23–44). Springer International Publishing. https://doi.org/10.1007/978-3-030-33128-3_2
- Yushkevich, P. A., Piven, J., Cody Hazlett, H., Gimpel Smith, R., Ho, S., Gee, J. C., & Gerig, G. (2006). User-Guided 3D Active Contour Segmentation of Anatomical Structures: Significantly Improved Efficiency and Reliability. *Neuroimage*, *31*(3), 1116–1128.
- Zaeemzadeh, A., Rahnavard, N., & Shah, M. (2021). Norm-Preservation: Why Residual Networks Can Become Extremely Deep? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *43*(11), 3980–3990. <https://doi.org/10.1109/tpami.2020.2990339>
- Zaitsev, M., Maclaren, J., & Herbst, M. (2015). Motion artifacts in mri: A complex problem with many partial solutions. *Journal of Magnetic Resonance Imaging*, *42*(4), 887–901. <https://doi.org/10.1002/jmri.24850>
- Zakaria, M., Al-Shebany, M., & Sarhan, S. (2014). Artificial Neural Network: A Brief Overview. *International Journal of Engineering Research and Applications*, *4*(2), 7–12.
- Zardavas, D., Phillips, W. A., & Loi, S. (2014). PIK3CA mutations in breast cancer: reconciling findings from preclinical and clinical data. *Breast Cancer Research*, *16*(1). <https://doi.org/10.1186/bcr3605>
- Zeiler, M. D., & Fergus, R. (2014). Visualizing and Understanding Convolutional Networks. In D. Fleet, T. Pajdla, B. Schiele, & T. Tuytelaars (Eds.), *Computer vision ECCV 2014* (pp. 818–833). Springer International Publishing. https://doi.org/10.1007/978-3-319-10590-1_53
- Zeng, W., Ren, X., Su, T., Wang, H., Liao, Y., Wang, Z., Jiang, X., Yang, Z., Wang, K., Zhang, X., Li, C., Gong, Z., Yao, Y., Huang, X., Wang, J., Yu, J., Guo, Q., Yu, Y., Zhang, Y., . . . Tian, Y. (2021). Pangu- α :

- Large-scale autoregressive pretrained chinese language models with auto-parallel computation. <https://doi.org/10.48550/ARXIV.2104.12369>
- Zentrum für Krebsregisterdaten im Robert Koch-Institut. (2016a). Bericht zum Krebsgeschehen in Deutschland 2016 [Retrieved from: <https://www.krebsdaten.de/Krebs/DE/Content/Krebsarten/Bauchspeicheldruesenkrebs/bauchspeicheldruesenkrebs.html>]. <https://www.krebsdaten.de/Krebs/DE/Content/Krebsarten/Bauchspeicheldruesenkrebs/bauchspeicheldruesenkrebs.html>
- Zentrum für Krebsregisterdaten im Robert Koch-Institut. (2016b). Bericht zum Krebsgeschehen in Deutschland 2016. https://www.krebsdaten.de/Krebs/DE/Content/Publikationen/Krebsgeschehen/Krebsgeschehen_node.html
- Zhang, B., Haddow, B., & Birch, A. (2023). Prompting large language model for machine translation: A case study. <https://doi.org/10.48550/ARXIV.2301.07069>
- Zhang, C., Xie, Y., Bai, H., Yu, B., Li, W., & Gao, Y. (2021). A survey on federated learning. *Knowledge-Based Systems*, 216, 106775. <https://doi.org/10.1016/j.knosys.2021.106775>
- Zhang, H., Zhang, L., & Jiang, Y. (2019). Overfitting and underfitting analysis for deep learning based end-to-end communication systems. *2019 11th International Conference on Wireless Communications and Signal Processing (WCSP)*, 1–6. <https://doi.org/10.1109/WCSP.2019.8927876>
- Zhang, L., Wang, X., Yang, D., Sanford, T., Harmon, S., Turkbey, B., Wood, B. J., Roth, H., Myronenko, A., Xu, D., & Xu, Z. (2020a). Generalizing Deep Learning for Medical Image Segmentation to Unseen Domains via Deep Stacked Transformation. *IEEE Transactions on Medical Imaging*, 39(7), 2531–2540. <https://doi.org/10.1109/tmi.2020.2973595>
- Zhang, S., He, L., Dragut, E., & Vucetic, S. (2019a). How to Invest my Time: Lessons from Human-in-the-Loop Entity Extraction. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. <https://doi.org/10.1145/3292500.3330773>
- Zhang, T., Ladhak, F., Durmus, E., Liang, P., McKeown, K., & Hashimoto, T. B. (2023a). Benchmarking large language models for news summarization. <https://doi.org/10.48550/ARXIV.2301.13848>
- Zhang, Y., Pei, H., Zhen, S., Li, Q., & Liang, F. (2023b). Chat generative pre-trained transformer (chatgpt) usage in healthcare. *Gastroenterology & Endoscopy*, 1(3), 139–143. <https://doi.org/10.1016/j.gande.2023.07.002>
- Zhang, Y., Lobo-Mueller, E. M., Karanicolas, P., Gallinger, S., Haider, M. A., & Khalvati, F. (2020b). CNN-based survival model for pancreatic ductal adenocarcinoma in medical imaging. *BMC Medical Imaging*, 20(1). <https://doi.org/10.1186/s12880-020-0418-1>
- Zhang, Z., Jiang, H., Chen, J., Wei, Y., Cao, L., Ye, Z., Li, X., Ma, L., & Song, B. (2019b). Hepatocellular carcinoma: Radiomics nomogram on gadoteric acid-enhanced MR imaging for early postoperative recurrence prediction. *Cancer Imaging*, 19(1). <https://doi.org/10.1186/s40644-019-0209-5>
- Zhang, Z., Yang, L., & Zheng, Y. (2018). Translating and segmenting multimodal medical volumes with cycle-and shape-consistency generative adversarial network. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 9242–9251.
- Zhou, M., Leung, A., Echegaray, S., Gentles, A., Shrager, J. B., Jensen, K. C., Berry, G. J., Plevritis, S. K., Rubin, D. L., Napel, S., & Gevaert, O. (2018). Non-Small Cell Lung Cancer Radiogenomics Map

- Identifies Relationships between Molecular and Imaging Phenotypes with Prognostic Implications. *Radiology*, 286(1), 307–315. <https://doi.org/10.1148/radiol.2017161845>
- Zhou, S. K., Greenspan, H., Davatzikos, C., Duncan, J. S., Ginneken, B. V., Madabhushi, A., Prince, J. L., Rueckert, D., & Summers, R. M. (2021). A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises. *Proceedings of the IEEE*, 109(5), 820–838. <https://doi.org/10.1109/jproc.2021.3054390>
- Zhou, X., & Belkin, M. (2014). Chapter 22 - semi-supervised learning. In P. S. Diniz, J. A. Suykens, R. Chellappa, & S. Theodoridis (Eds.), *Academic press library in signal processing: Volume 1* (pp. 1239–1269, Vol. 1). Elsevier. <https://doi.org/10.1016/B978-0-12-396502-8.00022-X>
- Zhovannik, I., Pai, S., da Silva Santos, T. A., van Driel, L. L. G., Dekker, A., Fijten, R., Traverso, A., Bussink, J., & Monshouwer, R. (2021). Radiomics integration into a picture archiving and communication system. *Physics and Imaging in Radiation Oncology*, 20, 30–33. <https://doi.org/10.1016/j.phro.2021.09.007>
- Zhu, J., Yang, G., & Lio, P. (2019). How can we make gan perform better in single medical image super-resolution? a lesion focused multi-scale approach. *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, 1669–1673. <https://doi.org/10.1109/isbi.2019.8759517>
- Zhu, J.-Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2223–2232. https://openaccess.thecvf.com/content_iccv_2017/html/Zhu_Unpaired_Image-To-Image_Translation_ICCV_2017_paper.html
- Zhu, X., Dong, D., Chen, Z., Fang, M., Zhang, L., Song, J., Yu, D., Zang, Y., Liu, Z., Shi, J., & Tian, J. (2018). Radiomic signature as a diagnostic factor for histologic subtype classification of non-small cell lung cancer. *European Radiology*, 28(7), 2772–2778. <https://doi.org/10.1007/s00330-017-5221-1>
- Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., & He, Q. (2021). A comprehensive survey on transfer learning. *Proceedings of the IEEE*, 109(1), 43–76. <https://doi.org/10.1109/jproc.2020.3004555>
- Ziller, A., Trask, A., Lopardo, A., Szymkow, B., Wagner, B., Bluemke, E., Nounahon, J.-M., Passerat-Palmbach, J., Prakash, K., Rose, N., Ryffel, T., Reza, Z. N., & Kaissis, G. (2021, June). Pysyft: A library for easy federated learning. In M. H. u. Rehman & M. M. Gaber (Eds.), *Federated learning systems: Towards next-generation ai* (pp. 111–139). Springer International Publishing. https://doi.org/10.1007/978-3-030-70604-3_5
- Zwanenburg, A., Vallières, M., Abdalah, M. A., Aerts, H. J. W. L., Andrearczyk, V., Apte, A., Ashrafinia, S., Bakas, S., Beukinga, R. J., Boellaard, R., Bogowicz, M., Boldrini, L., Buvat, I., Cook, G. J. R., Davatzikos, C., Depeursinge, A., Desseroit, M.-C., Dinapoli, N., Dinh, C. V., . . . Löck, S. (2020). The image biomarker standardization initiative: Standardized quantitative radiomics for high-throughput image-based phenotyping [PMID: 32154773]. *Radiology*, 295(2), 328–338. <https://doi.org/10.1148/radiol.2020191145>

B Lists

B.1 List of Figures

2.1	Example for a segmentation of a volume of interest on a computed tomography image . . .	8
2.2	Relationship between artificial intelligence, machine learning and deep learning with examples for each field.	12
2.3	Schematic architecture of a multilayer perceptron	15
2.4	Schematic architecture of a convolutional neural network	18
2.5	Illustration of a convolution from an input image to the output layer	18
2.6	Building of a residual block in a ResNet 50	20
2.7	Schematic architecture of a ResNet 50	22
3.1	Patient inclusion flowchart of P-I	40
3.2	From (Jungmann et al., 2022). Heatmap assistance used in publication P-II	45
3.3	Study design of P-II	46

B.2 List of Tables

2.1	Definition of BI-RADS categories and linked management recommendations	35
3.1	Characteristics of patients included in P-I	41