Quantitative imaging and artificial intelligence in oncology

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

Quantitative Imaging and Artificial Intelligence in Oncology

by
Ashish Kumar Jha

Quantitative Imaging and Artificial Intelligence in Oncology

Dissertation

to obtain the degree of Doctor at the Maastricht University,
on the authority of the Rector Magnificus,
Prof. dr. Pamela Habibović,
in accordance with the decision of the Board of Deans,
to be defended in public
on Tuesday, 22nd of November 2022, at 13.00 hours

by
Ashish Kumar Jha

Supervisor

Prof.dr.ir. A.L.A.J. Dekker

Co-supervisors

Dr. L.Y.L. Wee

Dr. Alberto Traverso

Assessment Committee

Prof. dr. ir. H.J.W.L. Aerts (chair)

Prof. dr. R. Boellaard

Dr. S. Vasudeva Rao

Prof. dr. ir. J-J. Sonke

Dr. B.F.M. Slangen

To my Parents and my Gurus, Ambika Prasad Jha and Baidehi Devi, who served as wonderful role models and imparted the importance of plain living and high thinking to us.

Index

Chapter 1: Introduction and Outline of Thesis

GLOBOCAN 2018 showed that around 18.1 million cancer cases are diagnosed every year globally [5]. The total estimated new cancer cases globally rose to 19.3 million and 10 million cancer deaths in 2020. This increment in new cancer cases in the small span of two years is around 1.2 million globally, which is alarming [6]. Out of 19.3 million around 9.5 (49.2%), million new cases are from Asia and 1.32 (6.83%) million new cases are from India alone. Globally and in India breast cancer is the most common cancer for both sexes together (Globally: 11.71%; India: 13.5%) and in the female population (Globally: 24.5%; India: 26.3%) [6]. Lung cancer is the most common cancer in males globally (Globally: 14.3%) and Lip oral cavity cancer is most common in India (16.2%) [6]. In this thesis, I have reviewed the role of the prediction model in cervical cancer.

Cancer treatment & Role of imaging:

Treatment of cancer is not straightforward; it is a complex and long process [7-9]. The cancer treatment process starts with the diagnosis of disease and in some cases never ends while the patient is alive. The process of cancer treatment and the role of imaging is summarized in figure 1.
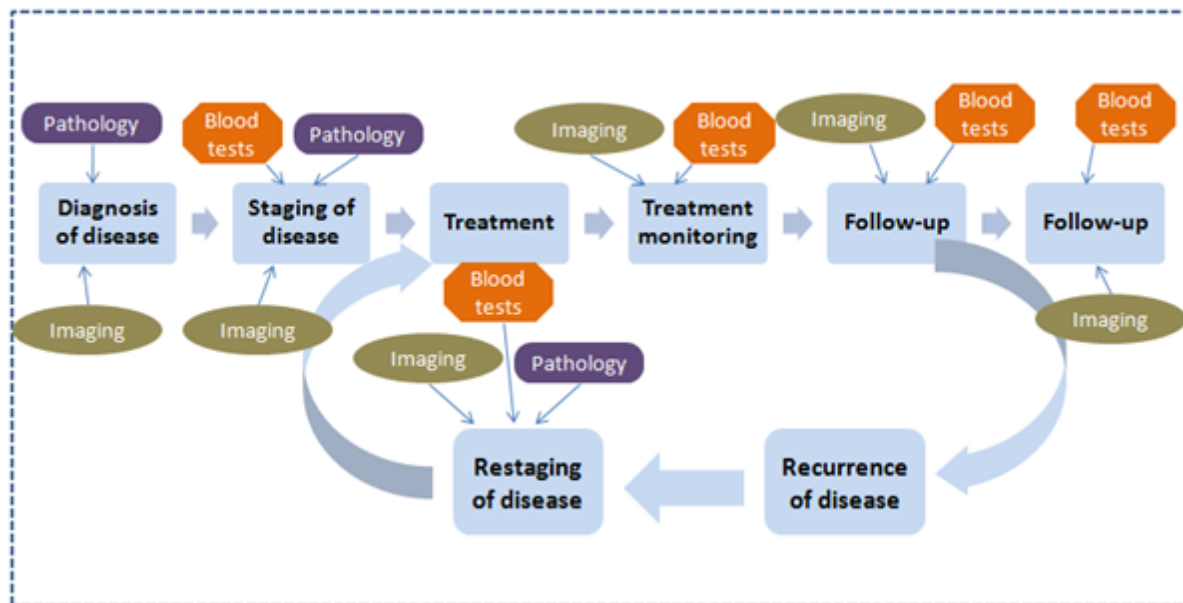
Figure-1: Various stages of cancer treatment process: Diagnosis of disease -> staging of disease -> treatment -> treatment monitoring -> follow-up

Although pathology is the confirmatory diagnosis of cancer, imaging plays a significant role in the diagnosis of cancer [7-9]. Simple imaging modalities like an x-ray to complex modalities like sonography, computed tomography (CT), single-photon emission tomography (SPECT), positron emission tomography (PET) and magnetic resonance imaging (MRI) are used for the diagnosis of cancer [7-15]. These modalities are used for the localization and characterization of the disease. For instance, a chest x-ray is often

performed for the initial identification and localization of lung cancer. Mammography x-ray is performed for diagnosis of breast cancer. Ultrasound is widely used for (additional) diagnosis of breast cancer, prostate cancer and gynecological cancers [1-12]. MRI and CT are helpful in the diagnosis of head and neck cancer and bone and soft tissue and brain tumours [12-15]. PET/CT is widely used for the diagnosis of lymphomas and many other malignant conditions [16-18]. Often in parallel with diagnosis, imaging is being used to stage disease. This complex process includes consideration of the primary tumour, nodal metastasis and distance metastasis. Based on imaging and pathology findings the tumour (T), nodal (N) and distance metastasis (M) i.e. the TNM staging is performed which results in a final overall stage of the disease which to a large extent determines the treatment [7-9], which typically requires a multidisciplinary approach with surgery, radiotherapy or chemotherapy or a combination of these often considered the best treatment[5-9]. In diagnosis, staging and treatments of cancers, the role of imaging is extensive [7-24].

Treatment monitoring and follow-up are performed to assess the efficacy of the treatment and related side effects of the treatment. Monitoring of patients is performed by checking the vitals of the patients, blood parameters and changes in imaging parameters. Most imaging parameters are used to monitor the regression or progression of disease during and after treatment [5-24].

Post-treatment follow-up is performed to identify the general status of the patient post-treatment, condition of the disease and recurrence or progression of the disease. Various blood parameters are used to check the general condition and progression of the disease. Imaging parameters are used to identify the disease condition i.e., stable disease, local recurrence, nodal and distant metastasis [7-9].

Many times during follow-up we encounter the *recurrence of the disease*. Treatment may be warranted in case of recurrence of the disease. In this scenario restaging of disease is performed based on imaging and pathological finding and subsequently suitable treatment is offered [5-9].

Imaging plays a significant role in the management of cancer. Imaging is one of the important diagnostic modalities which indicates the onset of cancer, and recurrence during the follow-up.

Various types of imaging can be used for the management of cancer i.e., X-ray, Ultrasound, computed tomography (CT), magnetic resonance imaging (MRI) single-photon emission tomography (SPECT) and positron emission tomography (PET) [5-24]. X-ray and ultrasound studies are commonly performed during the diagnosis and follow-up of various malignancies. CT, MRI, PET and SPECT are performed during staging, restaging and follow-ups [10-24]. Multi-modality imaging like PET/CT, SPECT/CT and PET/MRI is considered more and more in cancer management. In this thesis, I have reviewed the role of artificial intelligence in imaging.

Quantitative Imaging:

Image findings can be categorized into qualitative and quantitative parameters that are helpful in diagnosis and cancer management. Qualitative parameters are subject to interpreter variability whereas quantitative parameters are independent of interpreter variability [27]. In current care, imaging data is often analyzed qualitatively and semi-quantitatively to diagnose and stage the pathological condition. With visual perception and semi-quantitative data, only limited information is extracted from the image.

But information stored in medical images may not amenable to the visual interpretation by expert human eyes. This information can be extracted using mathematical and statistical formulas as various quantitative parameters which may be helpful in disease stratification and prognostication in cancer [25]. Quantitative imaging is the conversion of images into quantitative features that may be associated with important outcomes such as tumour response to the treatment in cancer management [26].

Conventional quantitative parameters like the difference in Hounsfield unit (HU) tumour size measurement in CT, the difference in standardized uptake value (SUV) value or uptake value in PET and difference in proton density, diffusion coefficient and spectral peak in MRI have demonstrated the capabilities to differentiate between the responders and non-responders in cancer therapies [26-28].

For example, response evaluation criteria in solid tumours (RECIST) [29, 30] is used to assess the treatment response on CT and similarly, Positron Emission Tomography Response Criteria in Solid Tumours (PERCIST) [31, 32] is used to assess the treatment response on PET. RECIST assesses the response of treatment based on size regression or tumour which amounts to an anatomical response whereas PERCIST is based on the change in SUV which is considered as a physiological response [31, 33-35].

Radiomics moves beyond these conventional parameters and applies quantitative imaging in a much broader context, e.g., by extracting many more features from the image. This has shown the potential to access the treatment response and outcomes in initial studies [36]. Radiomics, in other words, is a process to extract high throughput quantitative parameters from medical images to unearth various pathological conditions [36-40] that are associated with outcomes.

Radiomic features extracted from medical images can be classified as shape-based features, first-order features, higher-order features, textural features, LOG features and Wavelet features [37-41]. The entire radiomic process involves image extraction, tumour segmentation, pre-processing of images i.e., conversion of images in the required format, voxel normalization, image masking, filtering of image, image transformation, radiomic extraction involves extraction of radiomic features from original, filtered and transformed images [40, 41]. This process leads to data explosion which can be managed by employing feature selection or reduction methods. Finally extracted/selected

features are used for radiomic prediction model development. In this thesis, I reviewed the literature to find the role of radiomics in precision oncology.

Artificial Intelligence and big data

AI can be defined as the process to develop intelligent machines which can replace or outperform the natural intelligence of human beings [42]. Machine learning (ML) and Big Data are two main components of AI. Big Data can be defined by 5V i.e.; Volume: the data in large volume, Variety: data from various sources, Velocity: data grows very fast, Veracity: quality and integrity of data, and Value: the richness of data [43]. ML is the method by which a machine learns from past events or data without being explicitly programmed for that. There are three main types of ML methods known as (i) supervised learning; (ii) unsupervised learning and (iii) semi-supervised [44]. Deep learning is a subtype of machine learning in which the algorithm learns a composition of features that are represented in a hierarchy of structures in the data. Complex representations of data are expressed in terms of simpler representation of data in deep learning [45]. There are various deep learning algorithms that propose an end-to-end approach to predict outcomes by learning simple features in a hierarchical manner as components of complex features.

There are various kinds of AI algorithms e.g., Linear regression, Logistic regression, K-Nearest Neighbour (KNN), Random Forest (RF), Support Vector Machine (SVM), Bayesian Network (BN), and deep learning algorithms such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Artificial Neural Networks (ANN). [44-50]. Typically selecting the "right" algorithm depends on the task at hand. In this thesis, I have discussed the various algorithms used in the development of prediction models in oncology.

Precision oncology

Cancer treatment has advanced significantly over the last several decades in terms of technical developments in diagnostic and therapeutic equipment, development of various new drugs and newer and more precise surgical techniques. Often clinicians decide the treatment meticulously considering all factors i.e., histology of the tumour, stage of the disease and general condition of the patients [54, 55]. Despite tremendous development in cancer treatment and expertise, treatment success has lots of variabilities and on several occasions, these treatments fail miserably [56. This led to the evolution of personalized medicine in oncology [57]. Personalized oncology works on the principle of identification of subgroups of patients in particular disease types and uses specific treatment to the subgroups [57, 58]. Many biomarkers and gene mutations have been investigated to identify the subgroups of the patients in various cancers and targeted drugs for those subgroups [58-59]. In recent years several imaging biomarkers

are also being investigated for identification of such subgroups and predict the outcome of the treatment.

Limitations of Radiomics

An inherent problem of repeatability and reproducibility exists with radiomics which is caused due to the difference in the scanner from different vendors, different acquisition protocols and intra scanner variations. In our earlier repeatability and reproducibility study, we found that only 10% of CT radiomic features had good repeatability and reproducibility in clinical cohort and phantom [87]. Traverso A. et al. in a systematic literature review have also concluded that there are stability issues with the majority of radiomic features [88]. To harmonize radiomic extraction tools, features and imaging standards several initiatives are started by various agencies, like The Quantitative Imaging Network (QIN) [89], the Quantitative Imaging Biomarkers Alliance (QIBA) [95], and Quantitative Imaging in Cancer: Connecting Cellular Processes with Therapy (QuIC-ConCePT) [90]. These agencies are working continuously to standardize imaging and imaging biomarkers. The Image Biomarker Standardization Initiative (IBSI) is another consortium that works towards the harmonization of radiomic features across the globe by minimizing the deviation in imaging and standardizing the radiomic extraction process [91, 92]. QIN is an initiative to harmonize the imaging parameters and hardware across the vendor. QIB has initiated the scanning of phantoms to trace the standard differences between the equipment. The main objective of such a phantom study is to find the errors related to data collection and establish the procedures to harmonize the performance of imaging equipment among different makes and models with the goal to reduce the bias and variance across the equipment. QIN initiative can be an important step towards stratifying patients through accurate measurements of imaging biomarkers [90]. The radiomics quality score (RQS) is another such initiative proposed by Lambin et al. to address the issues related to radiomic study reporting [37]. Most of these initiatives will assist in advancing the standardization process of imaging biomarkers.

As repeatability and reproducibility of radiomic features are questionable there is a requirement for detailed investigation of radiomic features for their repeatability and reproducibility. In this thesis, I have performed a detailed study on the repeatability and reproducibility of radiomic features and presented the most robust features for further evaluation.

AI infrastructure development using the FAIR data concept

A prerequisite for the implementation of AI in hospitals is the transformation of the different data elements (DICOM, radiomic features, and clinical data, diagnostic, pathology, genetic) in a FAIR format (Findable, Accessible, Interoperable, and

Reusable). Data should be organized and archived as such; it should be findable, accessible, interoperable, and reusable for both humans and machines. Hence, the data should be anonymized and assigned with global and persistent identifiers. Furthermore, the accessibility aspect includes a standardized protocol (authorization and identification) so that the authorized users can easily access the data. The connection of the different data concepts with domain ontologies enables the interoperability between the different researchers interested in using the data, while the reusability is succeeded with the inclusion of rich documentation to support data interpretation for the different users. These data principles are the guidelines to improve the quality of data holdings and focus on enhancing the automation to find the data and reuse it.

Oncology treatment produces a huge amount of data that satisfies all 5Vs of big data. But those data are either stored in various files in the medical record section of the hospital or in the form of free text or tables in hospital HIS which are not amenable to machine learning. Harvesting and transforming those data in machine-understandable form is a challenging task for a data scientist. There is a need to develop an automated system to transform and store these medical records in a format amenable to machine learning. In this thesis, we have tried to develop an automated system to transform medical records in a machine-readable format and store them in a form amenable to local and distributed machine learning adhering to the FAIR data principle. We have also developed the pyradiomics based GUI software for radiomic extraction.

Radiomics and Precision oncology

Various studies have been performed to demonstrate the utility of radiomics in cancer management [59]. Radiomics based studies have witnessed a rapid growth in the last decade; several studies have been published that show the potential of radiomics in the diagnosis and treatment of cancer. Many radiomics based AI decision support systems have been developed in oncology and reported in the literature. In the last few years new aspects of radiomics such as delta radiomics are being researched [60]. Delta radiomics comprises extraction and comparison of quantitative features from sequential scans acquired over the course of treatment, which provides information on the efficacy of treatment [60, 61]. The utility of radiomic based prediction modelling has been tested widely in the diagnosis and treatment of all varieties of solid tumours. Literature is suggestive of the utilization of radiomic features as quantitative biomarkers for several oncological conditions like; Brain Tumours, Head-and-neck cancer, Breast cancer, Lung cancer, Colorectal cancer, Prostate cancer, Gastrointestinal (GI), Liver cancer and Cervical cancer [61-87].

As radiomics features have been identified as a potential digital marker in precision oncology, the next goal should be to identify the robust radiomic features as a potential digital biomarker. In this thesis, I have also evaluated the role of robust radiomic features in the prediction of outcomes in cervical and lung cancer.

Aim and outline of the thesis

Given the above challenges, it is the central hypothesis of this thesis that radiomics can be combined with other data sources and by artificial intelligence applications to predict outcomes in oncology and that these predictions are clinically relevant. It is the aim of the thesis to investigate the robustness to radiomic features and the role of those features and artificial intelligence in the prediction of clinical endpoints in oncology in Indian scenarios. The secondary aim of this thesis was to develop an automated system for data harvesting, transformation and storage to perform artificial intelligence-based research in oncology in an Indian scenario.

This thesis contains six sections described in Table-1. Each section of this thesis addresses the issues related to radiomic implementation in precision oncology.

Table 1: Summary of the topics and characteristics of the studies presented in each chapter of this thesis.

Section	Chapter	Learning Objective	Title
Introduction	**Chapter 1**	Introduction of thesis	Introduction and Outline of Thesis
Current status	**Chapter 2**	Review of literature	Emerging role of artificial intelligence in Imaging
	Chapter 3		Systematic Review and Meta-analysis of Prediction Models used in cervical cancer
	Chapter 4		Radiomics: A Quantitative Imaging Biomarker in Precision Oncology
Stability Study	**Chapter 5**	Stability study of Imaging biomarker	Repeatability and reproducibility study of

			radiomic features on a phantom and human cohort
Infrastructure development for Machine learning	**Chapter 6** **Chapter 7**	Infrastructure development for Radiomic Extraction Infrastructure development for Machine learning	Development and validation of GUI radiomics feature extractor software (PyRadGUI) using PyRadiomic package Implementation of Big Imaging Data Pipeline Adhering to FAIR Data Principles for Distributed Machine Learning in Oncology
Machine learning	**Chapter 8**	Machine learning example	Radiomics Signature: A Potential Imaging Biomarker for the Prediction of Overall Survival in cervical cancer: An India experience
Discussion	**Chapter 9**	Discussion of thesis	Discussion

References

[1] What Is Cancer? https://www.cancer.gov/about-cancer/understanding/ what-is-cancer accessed on 30-05-2021

[2] The Genetics of Cancer, https://www.cancer.net/navigating-cancer-car e/cancer-basics/genetics/genetics-cancer accessed on 30-05-2021

[3] Lodish H, Berk A, Zipursky SL, et al. Molecular Cell Biology. 4th edition. New York: W. H. Freeman; 2000. Section 24.1, Tumor Cells and the Onset of Cancer. Available from: https://www.ncbi.nlm.nih.gov/books/NBK21590/

[4] Cooper GM. The Cell: A Molecular Approach. 2nd edition. Sunderland (MA): Sinauer Associates; 2000. The Development and Causes of Cancer. Available from: https://www.ncbi.nlm.nih.gov/books/NBK9963/

[5] Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries [published correction appears in CA Cancer J Clin. 2020 Jul;70(4):313]. CA Cancer J Clin. 2018;68(6):394-424. doi:10.3322/caac.21492

[6] Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021;71(3):209-249. doi:10.3322/caac.21660

[7] Cancer Treatment, National Cancer Institute, https://www.cancer.gov/ about-cancer/treatment accessed on 15/06/2021

[8] Principles of cancer management, Clinical oncology for students, https://wiki.cancer.org.au/oncologyformedicalstudents/Principles_of_cancer_management, accessed on 15/06/2021

[9] Types of Cancer Treatment, National Cancer Institute, https://www.cancer.gov/about-cancer/treatment/types accessed on 15/06/2021

[10] Burnside, E. , 2007. Differentiating benign from malignant solid breast masses with US strain imaging. Radiology 245,

[11] Sood R, Rositch AF, Shakoor D, et al. Ultrasound for Breast Cancer Detection Globally: A Systematic Review and Meta-Analysis. J Glob Oncol. 2019;5:1-17. doi:10.1200/JGO.19.00127

[12] Lu J, Feng F, Jin Z. Cancer diagnosis and treatment guidance: role of MRI and MRI probes in the era of molecular imaging. Curr Pharm Biotechnol. 2013;14(8):714-722. doi:10.2174/1389201014666131226105916

[13] Sim AJ, Kaza E, Singer L, Rosenberg SA. A review of the role of MRI in diagnosis and treatment of early stage lung cancer. Clin Transl Radiat Oncol. 2020;24:16-22. Published 2020 Jun 6. doi:10.1016/j.ctro.2020.06.002

[14] Heitzman ER. The role of computed tomography in the diagnosis and management of lung cancer. An overview. Chest. 1986;89(4 Suppl):237S-241S.

[15] Computed Tomography (CT) Scans and Cancer, National Cancer Institute, https://www.cancer.gov/about-cancer/diagnosis-staging/ct-scans-fact-sheet accessed on 15/06/2021

[16] Francis IR, Brown RK, Avram AM. The clinical role of CT/PET in oncology: an update. Cancer Imaging. 2005;5 Spec No A(Spec No A):S68-S75. Published 2005 Nov 23. doi:10.1102/1470-7330.2005.0024

[17] Chen, W. , 2005. Imaging proliferation in brain tumors with 18F-FLT PET: comparison with 18F-FDG. Journal of Nuclear Medicine 46, (6) 945–952.

[18] Even-Sapir, E. , 2003. Lymphoscintigraphy for sentinel node mapping using a hybrid SPECT/CT system. Journal of Nuclear Medicine 44, (9) 1413–1420.

[19] Dugdale, P.E. , Miles, K.A. , Bunce, I. , Kelley, B.B. , Leggett, D.A.C. , 1999. CT measurement of perfusion and permeability within lymphoma masses and its ability to assess grade, activity, and chemotherapeutic response. Journal of Computer Assisted Tomography 23, (4) 540–547.

[20] Buijs, M. , 2007. Assessment of metastatic breast cancer response to chemoembolization with contrast agent enhanced and diffusion-weighted MR imaging. Journal of Vascular and Interventional Radiology 18, (8) 957–963.

[21] Evans, S.F. , 2006. 18F EF5 PET imaging with immunohistochemical validation in patients with brain lesions. International Journal of Radiation Oncology Biology Physics 66, (3) S248

[22] de Geus-Oei, L.-F. , 2007. Chemotherapy response evaluation with 18F-FDG PET in patients with non-small cell lung cancer. Journal of Nuclear Medicine 48, (10) 1592–1598.

[23] Aboagye, E.O. , Bhujwalla, Z.M. , Shungu, D.C. , Glickson, J.D. , 1998. Detection of tumour response to chemotherapy by 1H nuclear magnetic resonance spectroscopy: effect of 5-fluorouracil on lactate levels in radiation-induced fibrosarcoma I tumours. Cancer Research 58, 1063–1067.

[24] Czernin, J. , Weber, W.A. , Herschman, H.R. , 2006. Molecular imaging in the development of cancer therapeutics. Annual Reviews of Medicine 57, 99–118.

[25] Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. Radiology. 2016;278(2):563-577. doi:10.1148/radiol.2015151169

[26] Clarke LP, Nordstrom RJ, Zhang H, Tandon P, Zhang Y, Redmond G, Farahani K, Kelloff G, Henderson L, Shankar L, Deye J, Capala J, Jacobs P. The quantitative imaging network: NCI's historical perspective and planned goals. Trans Oncol. 2014;7(1):1–4.

[27] Clarke LP, Croft BS, Nordstrom R, Zhang H, Kelloff G, Tatum J. Quantitative imaging for evaluation of response to cancer therapy. Trans Oncol. 2009;2(4):195–197.

[28] Nordstrom RJ. The Quantitative Imaging Network in Precision Medicine. Tomography. 2016;2(4):239-241. doi:10.18383/j.tom.2016.00190

[29] Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer 2009;45(2):248-260.

[30] Bogaerts J, Ford R, Sargent D, Individual patient data analysis to assess modifications to the RECIST criteria. et al. Eur J Cancer 2009; 45:248–260

[31] Choi H, Charnsangavej C, Faria SC, et al., Correlation of computed tomography and positron emission tomography in patients with metastatic gastrointestinal stromal tumor treated at a single institution with imatinib mesylate: proposal of new computed tomography response criteria. J. Clin Oncol 2007;25(13):1753-1759.

[32] Aras M, Erdil TY, Dane F, Gungor S, Ones T, Dede F, Inanir S, Turoglu HT. Comparison of WHO, RECIST 1. 1, EORTC, and PERCIST criteria in the evaluation of treatment response in malignant solid tumors. Nucl Med Commun. 2016;37:9–15.

[33] Yanagawa M, Tatsumi M, Miyata H, Morii E, Tomiyama N, Watabe T, Isohashi K, Kato H, Shimosegawa E, Yamasaki M, Mori M, Doki Y, Hatazawa J. Evaluation of response to neoadjuvant chemotherapy for esophageal cancer: PET response criteria in solid tumors versus response evaluation criteria in solid tumors. J Nucl Med. 2012;53:872–80

[34] Min SJ, Jang HJ, Kim JH. Comparison of the RECIST and PERCIST criteria in solid tumors: a pooled analysis and review. Oncotarget. 2016;7(19):27848-27854. doi:10.18632/oncotarget.8425

[35] Agrawal A, Purandare N, Shah S, Puranik A, Banavali S, Rangarajan V. Response assessment in metronomic chemotherapy: RECIST or PERCIST? Indian J Nucl Med. 2014;29:74–80.

[36] Park JE, Kim HS. Radiomics as a Quantitative Imaging Biomarker: Practical Considerations and the Current Standpoint in Neuro-oncologic Studies. Nucl Med Mol Imaging. 2018;52(2):99-108. doi:10.1007/s13139-017-0512-7

[37] Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol. 2017;14(12):749-762. doi:10.1038/nrclinonc.2017.141

[38] Kumar V, Gu Y, Basu S, et al. Radiomics: the process and the challenges. Magn Reson Imaging. 2012;30(9):1234-1248. doi:10.1016/j.mri.2012.06.010

[39] O'Connor JP. Rethinking the role of clinical imaging. Elife. 2017;6:e30563. Published 2017 Sep 6. doi:10.7554/eLife.30563

[40] Griethuysen, J. J. M., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., Beets-Tan, R. G. H., Fillon-Robin, J. C., Pieper, S., Aerts, H. J. W. L. (2017).

[41] Computational Radiomics System to Decode the Radiographic Phenotype. Cancer Research, 77(21), e104–e107. `https://doi.org/10.1158/0008-5472.CAN-17-0339

[42] McCarthy J, Minsky ML, Rochester N, Shannon CE. A proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955. AI Mag. 2006;27:12–14.

[43] Wang, Lidong & Alexander, Cheryl. (2016). Machine Learning in Big Data. International Journal of Mathematical, Engineering and Management Sciences. 1. 52-61. 10.33889/IJMEMS.2016.1.2-006.

[44] Hussain, Mubashir & Manhas, Dr. (2016). ARTIFICIAL INTELLIGENCE FOR BIG DATA: POTENTIAL AND RELEVANCE. International Acadmey of Engineering and Medical Research, 2016. Volume-1, ISSUE-1.

[45] Goodfellow I, Bengio Y, Courville A. Deep learning. Cambridge, Mass: MIT Press, 2016; 1–26.

[46] De Raedt, L., Kersting, K., Natarajan, S., and Poole, D. (2016). Statistical Relational Artificial Intelligence: Logic, Probability, and Computation. Synthesis Lectures on Artificial Intelligence and Machine Learning. San Rafael, CA: Morgan & Claypool Publishers.

[47] Goodfellow, I. J., Bengio, Y., and Courville, A. C. (2016). Deep Learning. Adaptive Computation and Machine Learning. Boston, MA: MIT Press.

[48] Jordan, M. I. and Mitchell, T. M. (2015). Machine learning: trends, perspectives, and prospects. Science 349, 255–260. doi: 10.1126/science.aaa8415

[49] Pons E, Braun LM, Hunink MG, Kors JA. Natural Language Processing in radiology: a systematic review. Radiology 2016; 279:329–343.

[50] Hadley TD, Pettit RW, Malik T, Khoei AA, Salihu HM. Artificial Intelligence in Global Health -A Framework and Strategy for Adoption and Sustainability. Int J MCH AIDS. 2020;9(1):121-127. doi:10.21106/ijma.296

[51] Schwalbe N, Wahl B. Artificial intelligence and the future of global health. Lancet. 2020;395(10236):1579-1586. doi:10.1016/S0140-6736(20)30226-9

[52] Alami H, Rivard L, Lehoux P, et al. Artificial intelligence in health care: laying the Foundation for Responsible, sustainable, and inclusive innovation in low- and middle-income countries. Global Health. 2020;16(1):52. Published 2020 Jun 24. doi:10.1186/s12992-020-00584-1

[53] Fernandez-Luque L, Imran M. Humanitarian health computing using artificial intelligence and social media: A narrative literature review. Int J Med Inform. 2018;114:136-142. doi:10.1016/j.ijmedinf.2018.01.015

[54] Lodish H. Section 24.1, Tumor Cells and the Onset of Cancer, In: Lodish H, Berk A, Zipursky SL, et al. Molecular Cell Biology. 4th edition. New York: W. H. Freeman; 2000. Available from: https://www.ncbi.nlm.nih.gov/books/NBK21590/

[55] Rai KR, Keating MJ. Treatment. In: Kufe DW, Pollock RE, Weichselbaum RR, et al., editors. Holland-Frei Cancer Medicine. 6th edition. Hamilton (ON): BC Decker; 2003. Available from: https://www.ncbi.nlm.nih.gov/books/NBK12735/

[56] Maeda H, Khatami M. Analyses of repeated failures in cancer therapy for solid tumors: poor tumor-selective drug delivery, low therapeutic efficacy and unsustainable costs. Clin Transl Med. 2018 Mar 1;7(1):11. doi: 10.1186/s40169-018-0185-6.

[57] Schwartzberg L, Kim ES, Liu D, Schrag D. Precision Oncology: Who, How, What, When, and When Not?. Am Soc Clin Oncol Educ Book. 2017;37:160-169. doi:10.1200/EDBK_174176

[58] Takeuchi S, Okuda S. Knowledge base toward understanding actionable alterations and realizing precision oncology. Int J Clin Oncol. 2019;24(2):123-130. doi:10.1007/s10147-018-1378-0

[59] Napel S, Mu W, Jardim-Perassi BV, Aerts HJWL, Gillies RJ. Quantitative imaging of cancer in the postgenomic era: Radio(geno)mics, deep learning, and habitats. Cancer. 2018;124(24):4633-4649. doi:10.1002/cncr.31630

[60] Fave X, Zhang L, Yang J, et al. Delta-radiomics features for the prediction of patient outcomes in non-small cell lung cancer. Sci Rep. 2017;7(1):588. Published 2017 Apr 3.

[61] Bai Y, Lin YS, Tian J, Shi DP, Cheng JL, Haacke EM. et al. Grading of Gliomas by Using Monoexponential, Biexponential, and Stretched Exponential Diffusion-weighted MR Imaging and Diffusion Kurtosis MR Imaging. Radiology. 2016;278:496–504.

[62] Shim KY, Chung SW, Jeong JH, et al. Radiomics-based neural network predicts recurrence patterns in glioblastoma using dynamic susceptibility contrast-enhanced MRI. Sci Rep. 2021;11(1):9974.

[63] Papp L, Potsch N, Grahovac M, Schmidbauer V, Woehrer A, Preusser M. et al. Glioma Survival Prediction with Combined Analysis of In Vivo C-11-MET PET Features, Ex Vivo Features, and Patient Features by Supervised Machine Learning. J Nucl Med. 2018;59:892–9.

[64] Perez-Beteta J, Molina-Garcia D, Ortiz-Alhambra JA, Fernandez-Romero A, Luque B, Arregui E. et al. Tumor Surface Regularity at MR Imaging Predicts Survival and Response to Surgery in Patients with Glioblastoma. Radiology. 2018;288:218–25

[65] Ren J, Tian J, Yuan Y, Dong D, Li X, Shi Y. et al. Magnetic resonance imaging based radiomics signature for the preoperative discrimination of stage I-II and III-IV head and neck squamous cell carcinoma. European journal of radiology. 2018;106:1–6.

[66] Chen L, Wang H, Zeng H, Zhang Y, Ma X. Evaluation of CT-based radiomics signature and nomogram as prognostic markers in patients with laryngeal squamous cell carcinoma. Cancer Imaging. 2020 Apr 22;20(1):28.

[67] Zhang B, He X, Ouyang F, Gu D, Dong Y, Zhang L. et al. Radiomic machine-learning classifiers for prognostic biomarkers of advanced nasopharyngeal carcinoma. Cancer Letters. 2017;403:21.

[68] Wang G, He L, Yuan C, Huang Y, Liu Z, Liang C. Pretreatment MR imaging radiomics signatures for response prediction to induction chemotherapy in patients with nasopharyngeal carcinoma. European Journal of Radiology. 2018;98:100–6.

[69] Sanduleanu S, Jochems A, Upadhaya T, et al. Non-invasive imaging prediction of tumor hypoxia: A novel developed and externally validated CT and FDG-PET-based radiomic signatures. Radiother Oncol. 2020;153:97-105. doi:10.1016/j.radonc.2020.10.016

[70] Tran WT, Gangeh MJ, Sannachi L, Chin L, Watkins E, Bruni SG. et al. Predicting breast cancer response to neoadjuvant chemotherapy using pretreatment diffuse optical spectroscopic texture analysis. British Journal of Cancer. 2017;116:1329–39.

[71] Park H, Lim Y, Ko ES, Cho H-h, Lee JE, Han B-K, Radiomics Signature on Magnetic Resonance Imaging: Association with Disease-Free Survival in Patients with Invasive Breast Cancer. Clinical Cancer Research; 2018. clincanres. 3783.2017.

[72] Zhang L, Chen B, Liu X, Song J, Fang M, Hu C. et al. Quantitative Biomarkers for Prediction of Epidermal Growth Factor Receptor Mutation in Non-Small Cell Lung Cancer. Translational Oncology. 2018;11:94–101.

[73] Zhu X, Dong D, Chen Z, Fang M, Zhang L, Song J, Radiomic signature as a diagnostic factor for histologic subtype classification of non-small cell lung cancer. European Radiology; 2018. pp. 1–7.

[74] Aerts HJWL, Patrick G, Tan Y, Oxnard GG, Naiyer R, Schwartz LH. et al. Defining a Radiomic Response Phenotype: A Pilot Study using targeted therapy in NSCLC. Scientific Reports. 2016;6:33860.

[75] Huang Y, Liu Z, He L, Chen X, Pan D, Ma Z. et al. Radiomics Signature: A Potential Biomarker for the Prediction of Disease-Free Survival in Early-Stage (I or II) Non-Small Cell Lung Cancer. Radiology. 2016;281:947.

[76] Liu ZY, Zhang XY, Shi YJ, Wang L, Zhu HT, Tang ZC. et al. Radiomics Analysis for Evaluation of Pathological Complete Response to Neoadjuvant Chemoradiotherapy in Locally Advanced Rectal Cancer. Clin Cancer Res. 2017;23:7253–62

[77] Toiyama Y, Inoue Y, Shimura T, Fujikawa H, Saigusa S, Hiro J. et al. Serum Angiopoietin-like Protein 2 Improves Preoperative Detection of Lymph Node Metastasis in Colorectal Cancer. Anticancer Research. 2015;35:2849–56.

[80] Chaddad A, Kucharczyk MJ, Niazi T. Multimodal Radiomic Features for the Predicting Gleason Score of Prostate Cancer. Cancers; 2018. p. 10

[81] Shiradkar R, Ghose S, Jambor I, Taimen P, Ettala O, Purysko AS. et al. Radiomic features from pretreatment biparametric MRI predict prostate cancer biochemical recurrence: Preliminary findings. J Magn Reson Imaging. 2018;48:1626–36

[82] Staal FCR, van der Reijd DJ, Taghavi M, Lambregts DMJ, Beets-Tan RGH, Maas M. Radiomics for the Prediction of Treatment Outcome and Survival in Patients With Colorectal Cancer: A Systematic Review. Clin Colorectal Cancer. 2021;20(1):52-71. doi:10.1016/j.clcc.2020.11.001

[83] Bakr S, Echegaray S, Shah R, Kamaya A, Louie J, Napel S. et al. Noninvasive radiomics signature based on quantitative analysis of computed tomography images as a surrogate for microvascular invasion in hepatocellular carcinoma: a pilot study. Journal of Medical Imaging. 2017;4:041303

[84] Liu S, Zhang Y, Xia J, Chen L, Guan W, Guan Y. et al. Predicting the nodal status in gastric cancers: The role of apparent diffusion coefficient histogram characteristic analysis. Magnetic Resonance Imaging. 2017;42:144.

[85] Altazi BA, Fernandez DC, Zhang GG, Hawkins S, Naqvi SM, Kim Y, Hunt D, Latifi K, Biagioli M, Venkat P, Moros EG. Investigating multi-radiomic models for enhancing prediction power of cervical cancer treatment outcomes. Phys Med. 2018 Feb;46:180-188. doi: 10.1016/j.ejmp.2017.10.009. Epub 2018 Feb 21.

[86] Reuzé S, Orlhac F, Chargari C, Nioche C, Limkin E, Riet F, Escande A, Haie-Meder C, Dercle L, Gouy S, Buvat I, Deutsch E, Robert C. Prediction of cervical cancer recurrence using textural features extracted from 18F-FDG PET images acquired with different scanners. Oncotarget. 2017 Jun 27;8(26):43169-43179. doi: 10.18632/oncotarget.17856.

[88] Jha AK, Mithun S, Jaiswar V, et al. Repeatability and reproducibility study of radiomic features on a phantom and human cohort. Sci Rep. 2021;11(1):2055. Published 2021 Jan 21. doi:10.1038/s41598-021-81526-8

[89] Traverso A, Wee L, Dekker A, Gillies R. Repeatability and Reproducibility of Radiomic Features: A Systematic Review. Int J Radiat Oncol Biol Phys. 2018;102(4):1143-1158. doi:10.1016/j.ijrobp.2018.05.053

[90] National Cancer Institute, Division of Cancer Treatment & Diagnosis. Quantitative Imaging Network (QIN) [online], https://imaging.cancer.gov/programs_resources/specialized_initiatives/qin.htm . last accessed on 15-05-2021

[91] Radiological Society of North America. Quantitative Imaging Biomarkers Alliance® (QIBA®). rsna.org https://www.rsna.org/research/quantitative-imaging-biomarkers-alliance/. last accessed on 15-05-2021

[92] QUantitative Imaging in Cancer: CONnecting CEllular Processes with Therapy, https://cordis.europa.eu/project/id/115151 last accessed on 15-05-2021

[93] Zwanenburg A, Vallières M, Abdalah MA, et al. The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. Radiology. 2020;295(2):328-338. doi:10.1148/radiol.2020191145

[94] The image biomarker standardisation initiative, https://ibsi.readthedocs.io/en/latest/index.html, Last accessed on 15-05-2021

Chapter 2: Emerging role of artificial intelligence in nuclear medicine

Abstract

The role of artificial intelligence is increasing in all branches of medicine. The emerging role of artificial intelligence applications in nuclear medicine is going to improve the nuclear medicine clinical workflow in the coming years. Initial research outcomes are suggestive of increasing role of artificial intelligence in nuclear medicine workflow, particularly where selective automation tasks are of concern. Artificial intelligence-assisted planning, dosimetry and procedure execution appear to be areas for rapid and significant development. The role of artificial intelligence in more directly imaging-related tasks, such as dose optimization, image corrections and image reconstruction, have been particularly strong points of artificial intelligence research in nuclear medicine. Natural Language Processing (NLP)-based text processing

Introduction

Artificial intelligence applied to clinical workflows and personalized cancer care has attracted immense attention in recent years [1–3]. The strong appeal of artificial intelligence is due in large part to its potential for significantly speeding up mundane clinical tasks through selective automation. In various domains of cancer care, artificial intelligence tools have shown potential to assist with automated diagnosis, segmentations of normal organs and tumor volumes, complex image transformation and interpretation, automated processing of textural reports, as well as other tasks that were previously thought to be the exclusive preserve of human experts. The other strong appeal of artificial intelligence is its potential to 'mine' for diagnostic or prognostic patterns (i.e., a 'signature') among a very large number of potential variables in order to make a reliable prediction of outcome.

In the scope of this article, artificial intelligence shall be defined as a scientific study of mathematical processes (i.e., algorithms) that are able to approximate a tightly confined aspect of human cognitive actions, without requiring constant and continuous control by a human operator. The mark of artificial intelligence is therefore computer applications that appear to demonstrate some degree of autonomy, adaptability and agnosticism towards the completion of a narrowly-defined function. Machine learning is a subfield of study in artificial intelligence involving training of mathematical and statistical algorithms, in order to generate the desired output when provided with a given set of inputs. It does so by being exposed to an extremely large number of repeated training episodes whereby the machine is conceptually 'penalized' for every incorrect output and 'rewarded' for every acceptable output. This dependency on a large number of training instances necessarily makes this type of artificial intelligence exceptionally sensitive to the volume, variety, velocity and veracity of the data on which it has to train on; these four V's are immediately recognizable as the signature hallmarks of 'big data' [4,5] (Fig. 1).
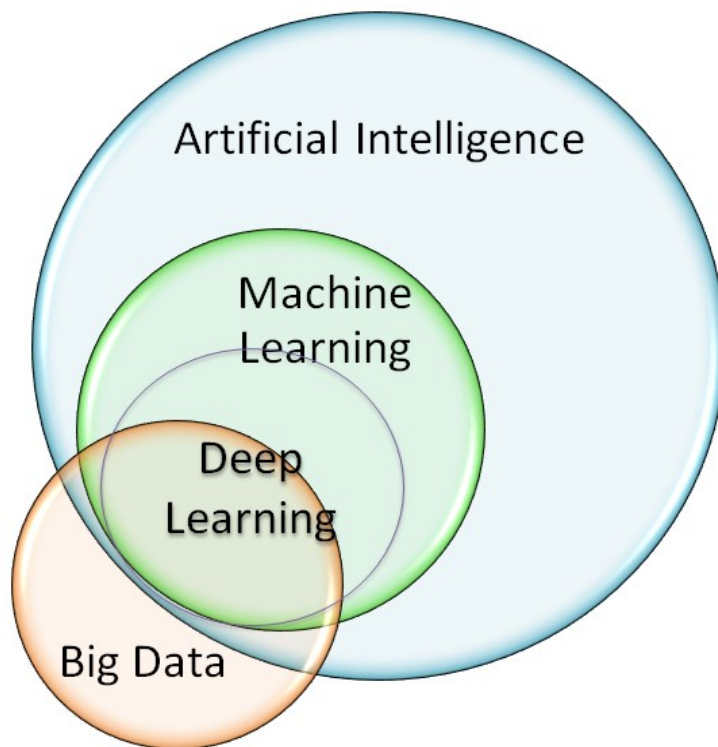
Figure 1: Interdependency of Artificial intelligence (AI), Machine Learning (ML) and Big data

Diagnostic and interventional nuclear medicine departments routinely generate huge amounts of data in the form of medical images, text reports and interventional data. With current generation of hybrid scanners such as PET/computed tomography (PET/CT), PET/MRI and single-photon emission tomography (SPECT)/CT scanners, every patient encounter could generate several gigabytes of data in the form of images [6]. These images are interpreted by an imaging expert and textual data are also generated in the form of diagnostic reports. A significant amount of intervention data is also generated in nuclear medicine every day in the form of clinical reports and predosimetry/postdosimetry calculations. Recently, several mathematical algorithms have been developed to extract a vast number of quantitative tumor metrics from medical images in the form of 'radiomic features' [7]. Nuclear medicine data generation is therefore vast, fast, diversified and also highly variable in interpretation, due to human expert subjectivity and variety of scanners from several vendors. Hence, nuclear medicine data clearly qualifies as big data, that satisfies all of the so-called '4 V's' [8].

With particular attention to the big imaging data that is found in nuclear medicine, a subfield of machine learning, known as deep learning, has become an extremely powerful artificial intelligence tool for image processing and image analysis. Deep learning algorithms learn composition of data that is represented in a hierarchy of structures of simpler features as a representation of the complex data [5,9]. Deep learning neural networks exploit a vast number of extremely simple computational

units, called artificial neurons, which are organized in deep stacks of interconnecting layers [9–11]. Specific deep learning architectures known as convolutional neural networks (CNNs) have been shown to be extremely adaptable to general image-based tasks such as segmentation, object detection and object classification [9–11].

Radiomics

Radiomics refers to the automated extraction and analysis of quantitative features of an image, such as a PET or a CT, in order to recognize potentially useful diagnostic or prognostic signatures [7]. Therefore, radiomics draws very heavily from technology in both machine learning and deep learning domains, to derive statistical prediction models from ultra-high dimensionality data using regression, decision trees, principal components analysis and ever increasing in recent years, deep learning. The radiomics hypothesis posits that medical images are not simply pictures for qualitative interpretation [7], but can be directly converted into minable data that could be used for personalized cancer care [12]. A typical radiomics analysis cycle is schematically summarized in Fig. 2. This involves the extraction of the region of interest on the image, preprocessing (such as digital image filters and resampling, among others), feature extraction and finally model development with validation [13].
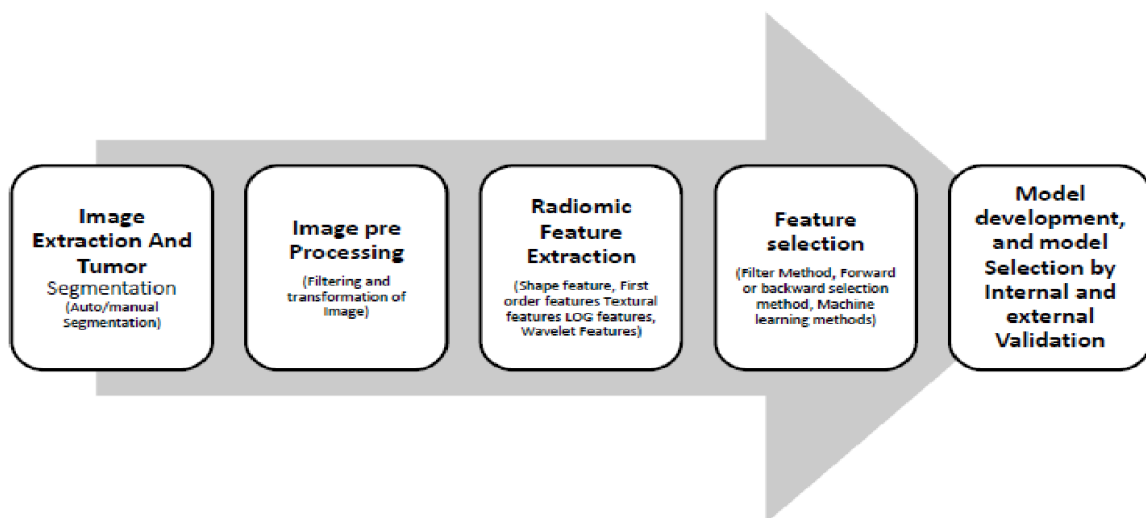
Figure 4: figure describes radiomic process for radiomic feature extraction and feature selection

Natural Language Processing

Natural Language Processing (NLP) employs machine learning and deep learning tools to aid in the extraction of structured data from natural language sources, such as recorded speech or written text reports or both. There are various well-established applications of NLP in diagnostic and clinical text processing in nuclear medicine such as report classification, report interpretation, sentiment analysis and text generation, etc. [14]. The NLP workflow consists of various steps and is summarized in Fig. 3. The text is first preprocessed by various processes like segmentation in which the text is segmented for sections and sentences. Then certain preposition and conjunction words are removed (stop-word removal). The remaining words are tokenized and normalized by linking to the root word (tokenization and word normalization). The text is then analyzed for grammar and part of speech (syntactic analysis). This is then subject to named entity recognition, concept recognition and relation extraction using relevant ontologies (semantic analysis). Relevant concepts maybe checked for negation using negation detection algorithms. Text features are then extracted and vectorized. The extracted features are used in model development. Models maybe rule-based, machine learning-based or a hybrid of both. The models are developed for defined outcomes and require annotated corpus for training in case of supervised learning. The trained model then undergoes validation with internal data as well as external data. These models maybe developed and applied for the classification of problems, as a part of decision support systems, treatment outcome deduction, summarization or text/report generation.
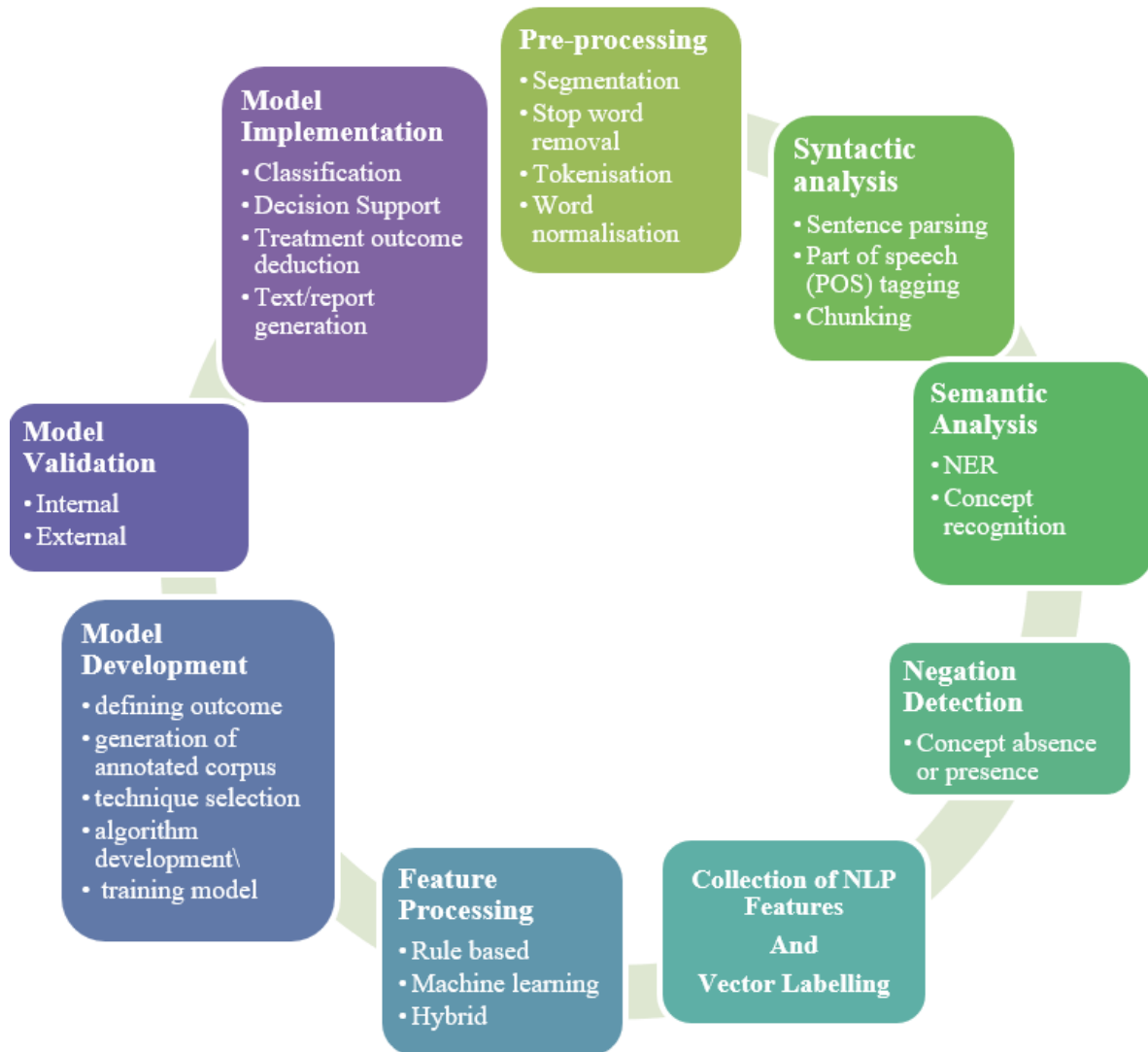
Figure 5: Natural Language processing workflow

Applications of artificial intelligence in nuclear medicine

Emergence of big data in cancer care

Over the past several years, continuous efforts have been expended to improve quality of patient care and accelerate the pace of cancer research through the ever-increasing use of artificial intelligence and medical big data informatics [15]. Large volumes of routine (standard of care) data related to cancer diagnosis, treatment planning and outcomes are stored as either structured or unstructured form within the confines of the hospital or clinic in the form of medical records [16]. The conversion of such highly granular patient's data into electronic formats has opened up a plethora of opportunities for the cancer research community. However, realization of clinical and societal value from routine medical big data comes with attendant risks. First and foremost, patient confidentiality must be protected at all times [17,18]. Anonymization, pseudonymization

and binding legal data-sharing contracts all go some way towards protecting confidentiality, and increasingly 'privacy-by-design' paradigms such as distributed learning and federated machine learning [19–21] are being used, where a diverse range of artificial intelligence algorithms can be successfully trained without exposing private individuals' identifiable information outside of the data providers' respective communications network firewalls [22–24]. Second, data quality and provenance issues due to missing data, inexact measurements and undocumented deviations from clinical protocol can lead to serious biases, lack of generalizability and potential patient harm, when developing artificial intelligence models on such data [25]. Last but not least, big data is plagued by interoperability issues such as different formatting standards, idiosyncratic data coding, language barriers and ambiguity in clinical meaning, thus limiting the utilization of this data. The growing impetus behind adherence to Findable-Accessible-Interoperable-Reusable data stewardship standards [26] are beginning to address such problems, thus gradually rendering more big data useable by automated artificial intelligence algorithms.

Potential benefits of applied artificial intelligence in nuclear medicine:

Several retrospective and prospective studies describe different approaches that could support earlier diagnosis and more accurate prognosis of cancer [27–29]. Various aspects of the clinical problem have been examined utilizing genomics, proteomics, radiomics and pathomics signatures [30,31]. These studies identify the potential as well as the limitations of these signatures for the prediction of cancer outcome. The vast majorities of these publications make use of one or more artificial intelligence algorithms, and further integrate data from heterogeneous sources for generating a prediction. The application of artificial intelligence to quantitative medical imaging has shown promising results in cancer diagnosis, prognosis of disease and treatment outcome prediction. In the following sections, we will discuss the role of artificial intelligence in several parts of the nuclear medicine clinical workflow. The key area of artificial intelligence implementation in nuclear medicine may be grouped into the planning of procedure, execution of procedure, image reconstruction, image interpretation, report generation and implementation of clinical decision support system. These potential domains of application in nuclear medicine are summarized in Fig. 4 and are discussed in detail in the review.
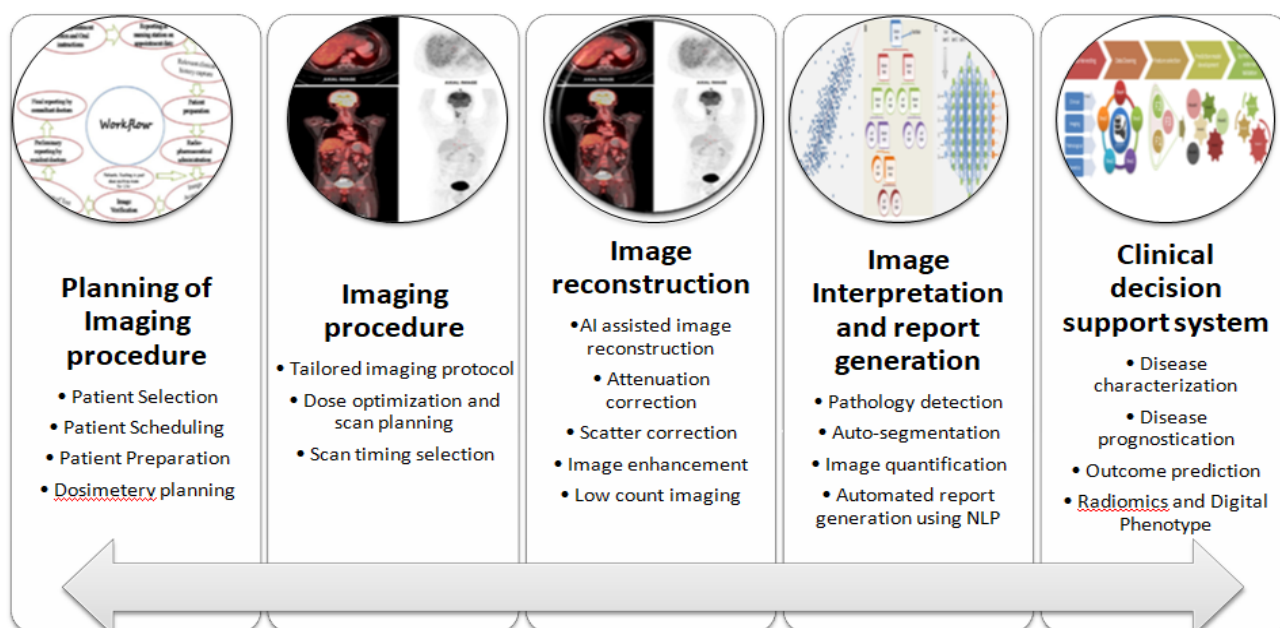
Figure 4: Role of AI in medical imaging and diagnostic workflow

Planning of nuclear medicine procedure:

Nuclear medicine procedures are associated with radiation exposure to the patient and are quite expensive. These procedures require patient specific-preparation (such as fasting prior to scan, cessation of insulin if diabetic, etc.) which has been challenging and required highly skilled human resources to juggle a large number of these requirements [32]. Diagnostic and therapeutic dosimetry planning requires imaging at multiple time points, image processing and mathematical calculations, all of which are technically demanding and widely seen as a challenging clinical task for nuclear medicine professionals [33]. Artificial intelligence algorithms can play a significant role in optimizing patient scheduling, procedure selection, patient pre-scan preparation and preimaging/treatment dosimetry planning. Ansart et al. [34] showed that artificial intelligence-based screening of patients led to an increased number of recruitments as well as reducing the number of expensive PET scans in a clinical trial. Massachusetts General Hospital developed a prediction model using multivariate logistic regression to predict the show or no-show of patient for PET/CT scan [35]. Shi et al. [36] have shown that a trained artificial neural network can reduce the individual radiation dosimetry prediction error by between factor 4 and factor 10, in comparison with population-based dosimetry approach in Lu-177-based dosimetry assessment of tumor and various organs. Xie et al. [37] have utilized 3D deep convolutional network algorithm for automated segmentation of normal organ and fetus with Dice similarity coefficient between 0.92 and 0.98, which resulted in less than 1% difference in the radiation dose calculation for normal organ and fetus for various radiopharmaceuticals by conventional and artificial intelligence-based segmentation model.

Optimization of nuclear medicine procedure:

Modern generation scanners have started implementing various artificial intelligence approaches [38] that assist with optimization of image quality, scan time and radiation exposure during diagnostic nuclear medicine procedures. During therapeutic procedures, artificial intelligence-based automation can assist with pre-therapeutic and post-therapeutic dosimetry by using artificial intelligence-assisted organ and tumor segmentation, and prediction of either organ and/or tumor dose using single-point imaging, etc. [39,40]. Utilization of artificial intelligence in individualized dose assessment can assist a nuclear medicine physician to implement pre-therapeutic and post-therapeutic personalized dosimetry-based assessment, which is currently not done because the process is very lengthy, cumbersome and requires high expertise to perform the task. During PET and SPECT examinations, we often require additional imaging or delayed imaging. For example, if a patient is scheduled for a PET scan but there is a lingering suspicion of brain metastasis, he/she may require an extended study that includes scans of the brain. This is currently a decision taken on the spot by a doctor or technologist, and image sets of such scenarios are available in the nuclear medicine department. These image sets can be used to train a deep learning artificial intelligence model to assist in decision making by predicting or scoring the likelihood of a brain metastasis actually being present. In modern multidetector CT scanners, the tube current, and hence patient exposure, is being modulated in real time, to minimize investigational radiation dose, but this is not patient-specific because the minimum and maximum tube currents are pre-defined such that the scanner modulates between the specified values [41]. The disadvantage of this is the dose delivered can only be known after the completion of the scan. An artificial intelligence-assisted system may allow us to perform real-time and patient-specific modulation of the tube current value to obtain the best image quality while optimally reducing radiation exposure to the patient [42]. The amount of radioisotope administered to a patient is currently calculated based on body weight, and there is ample margin for empirical (subjective) judgment. This could potentially lead to excess radiation (more than what is needed for the clinical purpose) or else to excessive noise and suboptimal image quality at lower radiation dose. An artificial intelligence-based radioisotope dosing system can be trained with the help of patient demographic data which can suggest the patient-specific required activity in a much more precise manner. This will reduce unnecessary radiation burden to the patient or avoid having to re-do suboptimal scans due to insufficient dose of activity administered.

Image reconstruction:

Nuclear medicine departments have been dealing with the balancing act between radiation doses, longer imaging times, low count density in images, low signal-to-noise ratio, high partial volume effect and low spatial resolution since its inception [43]. Advancements in instrumentation technology and reconstruction technology have

addressed these concerns to some degree, in multidetector helical CT and high field-strength MRI [44]. Image reconstruction in nuclear medicine imaging is still unable to catch up to its other radiological imaging counterparts. Nuclear medicine imaging involves various computationally intensive steps including attenuation correction, scatter correction, noise correction and partial volume correction. Reconstruction algorithms in use today range from filtered back projection to newer techniques such as 3D Ordered Subset Expectation Maximization, row-action maximum likelihood algorithm, often in combination with advanced signal processing techniques like partial volume correction and time-of-flight (TOF) correction [45]. These techniques have substantially improved image quality in nuclear medicine without necessarily leading to higher dose in the patient. Recent literature demonstrates a paradigm shift in image reconstruction in nuclear medicine imaging from analytic predetermined approaches towards a more adaptable, heuristic and potentially more patient-specific approach. Several researchers have examined artificial intelligence-based attenuation correction, scatter correction in PET and SPECT image reconstruction and extensive research is being performed to improve image quality of low count PET scan [46–50]. Hong et al. [46] used a CNN to enhance the image resolution of PET scans. Xiang et al. [47] have shown promising results with deep learning to improve the image quality with low dose and reduced scan time reconstruction during PET imaging. Wang et al. [48] demonstrated the utility of 3D deep convolutional network to predict high-dose PET scan using low administered activity. Kim et al. [49] used a de-noising deep learning network with local linear fitting to improve image quality of PET in an iterative reconstruction algorithm. Shiri et al. [50] demonstrated the utility of a deep learning residual network to synthetically generate a fulldose myocardial SPECT image from low-dose SPECT data as well as he also demonstrated the utility of artificial intelligence in prediction of full-time acquisition SPECT image from half time acquisition SPECT imaging data. Multimodality fusion imaging with hybrid scanners has proven to be a significant breakthrough in medical imaging; modalities such as combined-PET/CT have now been established as the modality of choice for cancer imaging [51]. Combined-PET/MRI is another hybrid medical scanner that is rapidly growing its importance in cancer imaging, but it needed to overcome the challenges related to MRI-based attenuation correction in PET for accurate PET imaging and quantification [52]. This can be achieved in number of ways, including the use of deep learning to synthetically generate a radiation attenuation map (in effect, a synthetic CT) directly from MRI. Several analytical techniques have been tested to improve PET attenuation correction in PET/ MRI scanners [52]. However, recent literature points towards a growing role of artificial intelligence [53–57] with a particularly active line of study in image corrections. Arabi et al. [53] have further shown the feasibility of deep learning to estimate the attenuation correction based solely on TOF PET emission data. Arabi et al. [54] also estimated the attenuation and scatter corrections simultaneously in a multi-tracer neuroimaging study. Shiri et al. [55], in a similar kind of study, used deep learning to

estimate the joint scatter and attenuation correction in PET image. Hwang et al. used a deep learning neural network to apply a Maximum Likelihood reconstruction of Attenuation and Activity algorithm to jointly generate activity and attenuation maps using emission data [56]. Liu et al. [57] in another study have used the deep learning technique to generate the attenuation map for PET reconstruction using MRI images.

Image interpretation and report generation:

Image interpretation and reporting of scan are amongst the most important tasks in clinical nuclear medicine as it is in radiology. The entire process involves various steps as image reading, extraction of quantitative data, comparison of a scan with earlier scans (quantitatively and qualitatively), then reporting on the findings with (in general) a clinical interpretation of the significance of the findings. The potential of artificial intelligence here lies in the valuable time saving of nuclear medicine physicians by automatic execution of more mundane jobs in the preparation of the report. Like reading the patient's history from the electronic medical record, the consolidation of the history in the report, quantitative data extraction and comparison often consumes lots of time and leads to reduced output from an expert nuclear medicine physician or radiologist [58–60]. Artificial intelligence tools may be utilized to reduce this burden so that the expert can do more meaningful tasks like interpretation of finding, providing expert opinion and contributing to multimodality treatment groups [58–60].

NLP is an evolving technology that can assist with nuclear medicine report processing in many ways. Several studies have shown the use of NLP in text processing in radiology. Pons et al. [14], in their systematic review of literature of NLP in radiology reporting, have identified the potential use of NLP in diagnostic surveillance, cohort building for epidemiologic studies, query-based case retrieval, and quality assessment of radiology practice and clinical support services with level of evidence. These same clinical needs exist in nuclear medicine. Pinto et al. [61], in their narrative review, have also emphasized the potential importance of artificial intelligence in processing radiology reporting requirements as well as the importance of enforcing structured reporting to increase the use of artificial intelligence for radiology data-mining.

Clinical decision support system Nuclear medicine imaging plays an important role in personalized medicine. A theranostics concept is being used in radionuclide therapy to predict the efficacy of radionuclide therapy [62]. Various parameters like standardized uptake value (SUV), total lesion glycolysis (TLG) and PET Response Criteria in Solid Tumors (PERCIST), etc. have been utilized for disease outcome prognostication [63–65]. Mahadevaiah et al. [66] described the use of artificial intelligence in decision support systems in oncology, but very similar concepts will also hold for nuclear medicine. The potential role of artificial intelligence is being explored by several researchers working on the development of various kinds of prediction models for

outcome predictions in oncology, such as overall survival (OS), progression-free survival (PFS), loco-regional recurrence, distant metastasis, treatment outcome, toxicity and treatment selection. These studies utilize a range of qualitative, semiquantitative (SUV, TLG and PERCIST) and radiomics features extracted from nuclear medicine images [67–74].

Radiomic features derived from nuclear medicine images have been proposed as a potential digital imaging phenotype for disease-specific personalized predictions of individual outcomes [13]. Sanduleanu et al. [67] used a univariate Cox model to show that [F-18]-HX4 PET/CT hypoxia uptake was able to prognose OS and local PFS in head and neck cancer patient treated by chemotherapy [Spearman's correlation coefficient $(\rho S)=0.77$] (P<0.001). In a multicenter study, Betancur et al. [68] showed that a CNN model trained using polar maps of myocardial perfusion imaging (MPI) from SPECT scanners outperformed the human expert diagnosis [area under the curve (AUC): 0.81 vs. 0.78] for detecting obstructive myocardial disease. In an extended study, Betancur et al. also developed a prediction model combining clinical information and MPI data to predict the 3-year probability of major cardiac events. They demonstrated that the machine learning technique outperformed clinical decision-making by medical experts (AUC: 0.81 vs. 0.65) for predicting major cardiac events in a cohort of 2619 consecutive patients [69]. A similar study by Rios et al. [70] also suggested an improvement of radiomics biomarkers over conventional risk factors (AUC: 0.793 vs. 0.698) when predicting major adverse cardiac events. Shiri et al. [71] used a combination of radiomics with genomic sequencing to predict presence of epidermal growth factor receptor (AUC: 0.82) and V-KI-RAS2 Kirsten rat sarcoma viral oncogene homolog (KRAS) (AUC: 0.83) mutations from images of non-small cell lung cancer (NSCLC) patients. Kidd et al. [72] developed an FDG PET-based nomogram (SUVmaximum value and lymph node status) to predict recurrence-free survival (C-stat: 0.741), disease-specific survival (C-stat: 0.739) and OS (C-stat: 0.658) for locally advanced cervical cancer. However, some studies have also reported the poor validation performance of radiomic-based prognostic models. Foley et al. [73] performed an external validation of a prognostic model for esophageal cancer using quantitative PET features and radiomic features but found that these features were unable to discriminate between patient groups with different OS in an independent external validation cohort. van Timmeren et al. [74] also performed a multicenter radiomic-based study for treatment response assessment in NSCLC patients; they were able to demonstrate significance of FDG-PET/CT radiomic model in their individual-institution cohort but failed to reproduce this with a multicenter study.

Summary of review

The emergence of artificial intelligence applications in nuclear medicine appears to be on the verge of contributing significantly throughout the clinical workflow atomization. Initial research suggests an increasing role of artificial intelligence in nuclear medicine imaging, particularly where selective automation of tasks are concerned. Artificial intelligence-assisted planning, dosimetry and procedure execution appear to be areas for rapid and significant development. The role of artificial intelligence in more directly imaging-related tasks, such as dose optimization, image corrections and image reconstruction, have been particularly strong points of artificial intelligence research in nuclear medicine. A further aspect of artificial intelligence research and implementation in nuclear medicine will be disease characterization, prognostication and potentially treatment outcome prediction. The new study of radiomics has been especially active in this field, leading towards the potential for identifying image-based digital signatures of oncological disease that can inform clinical decision-making, and therefore a stepping stone towards individualized cancer care. Artificial intelligence-assisted image interpretation based on deep learning neural networks have the potential to open up a new horizon in quantitative nuclear medicine, and could become a powerful future tool in the hands of nuclear medicine physicians and radiologists. Lastly, an aspect of artificial intelligence that still needs to develop strongly and could be of immense utility in nuclear medicine is domain-specific NLP. Complex NLP tools may help to automate the report generation and clinical interpretation tasks, and likely enrich the clinical interpretation with the aid of quantitative features from radiomics and deep learning analysis. As a counterpoint to the promising potential future of artificial intelligence in nuclear medicine, one needs to be cognizant of the potential challenges and barriers to widespread artificial intelligence adoption in nuclear medicine clinics. Data, in vast quantities and with large dimensional variety, are the raw material for developing and testing artificial intelligence applications for nuclear medicine. While some data may be found in great abundance (e.g., PET and SPECT images), other types of data may be dispersed across other clinical departments (e.g. chemotherapy treatment details, or radiotherapy delineations, or long-term clinical follow-up) or have become disconnected completely from the radioisotope imaging due to suboptimal data management procedures. It is not simply the image data that needs to be curated and maintained with a view to long-term data sustainability, but it is also the metadata about the investigation that needs to be captured. For example, imaging settings, equipment and scanner software are likely to change rapidly over a short period of time, and we know that differences among vendors' equipment and variation in imaging parameters are a major concern that directly limits the wider applicability of radiomics models [75] and also of deep learning-based image analysis models [76]. A second major barrier to artificial intelligence applicability in nuclear medicine is one of clinical acceptability. This acceptability has two distinct dimensions; clinical

acceptability and psychological acceptability. The former requires a detailed and practical process of selecting, commissioning and validating an artificial intelligence tool with realistic and as locally-specific clinical conditions as reasonably possible [66]. This necessarily entails staff training, ongoing quality assurance and careful integration of emerging artificial intelligence tools into the clinical environment with oversight for safety and ethical usage. The latter degree of acceptability is much harder to quantify and define; however, it relates directly to another major topic of research in artificial intelligence, that is, the degree of 'explainability' (or lack thereof) in an artificial intelligence-based decision support system [77]. While a clinician can readily and eloquently supply a rationale for his/her clinical decisions to a qualified peer, to a presiding judge or to an inquiring patient, certain types of artificial intelligence are still lacking in this fundamental rationalizing power of an 'explanation'. For instance, why exactly does an imaging-related signature lead an artificial intelligence-based system to recommend a particular course of action or to suggest a particularly unpleasant prognosis? The risk for the clinician is, either to act upon or else to over-ride such recommendations from an artificial intelligence system (particularly in a question of diagnosis, prognosis or treatment selection). Either of these feels potentially psychologically unsafe in the absence of suitable explanation or justification from an artificial intelligence-based system, leading to deep mistrust or lack of application of artificial intelligence-based nuclear medicine tools.

Conclusion

The role of artificial intelligence is increasing in prominence in the clinical nuclear medicine workflow, starting from patient preparation to the development of decision support system in medicine. There appears to be ample evidence of development and advancement of artificial intelligence addressing nuclear medicine needs. This is pointing to an ever-increasing role of artificial intelligence in nuclear medicine. In the coming years, the computer science, informatics and clinical specialist fields need to collaborate closely to address some major hurdles related to the long-term sustainability of high-quality data (including imaging metadata) and clinical barriers to the adoption of artificial intelligence.

References

1. Bohr A, Memarzadeh K. The rise of artificial intelligence in healthcare applications. Artif Intell Healthcare 2020:25–60.
2. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol 2017; 2:230–243.
3. Wiens J, Shenoy ES. Machine learning for healthcare: on the verge of a major shift in healthcare epidemiology. Clin Infect Dis 2018; 66:149–153.
4. Lee CH, Yoon HJ. Medical big data: promise and challenges. Kidney Res Clin Pract 2017; 36:3–11.
5. Najafabadi MM, Villanustre F, Khoshgoftaar TM, Seliya N, Wald R, Muharemagic E. Deep learning applications and challenges in big data analytics. J Big Data 2015; 2:1.
6. Kharat AT, Singhal S. A peek into the future of radiology using big data applications. Indian J Radiol Imaging 2017; 27:241–248.
7. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. Radiology 2016; 278:563–577.
8. Morris MA, Saboury B, Burkett B, Gao J, Siegel EL. Reinventing radiology: big data and the future of medical imaging. J Thorac Imaging 2018; 33:4–16.
9. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015; 521:436–444.
10. Lee JG, Jun S, Cho YW, Lee H, Kim GB, Seo JB, Kim N. Deep learning in medical imaging: general overview. Korean J Radiol 2017; 18:570–584.
11. Choi H. Deep learning in nuclear medicine and molecular imaging: current perspectives and future directions. Nucl Med Mol Imaging 2018; 52:109–118.
12. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol 2017; 14:749–762.
13. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational radiomics system to decode the radiographic phenotype. Cancer Res 2017; 77:e104–e107.
14. Pons E, Braun LM, Hunink MG, Kors JA. Natural Language Processing in radiology: a systematic review. Radiology 2016; 279:329–343.
15. Liang G, Fan W, Luo H, Zhu X. The emerging roles of artificial intelligence in cancer drug development and precision therapy. Biomed Pharmacother 2020; 128:110255.
16. Dekker AL, Gulliford SL, Ebert MA, Orton CG. Point/Counterpoint. Future radiotherapy practice will be based on evidence from retrospective interrogation of linked clinical data sources rather than prospective randomized controlled clinical trials. Med Phys 2014; 41:030601.

17. Ienca M, Ferretti A, Hurst S, Puhan M, Lovis C, Vayena E. Considerations for ethics review of big data health research: a scoping review. PLoS One 2018; 13:e0204937.

18. Institute of Medicine (US) Committee on Regional Health Data Networks; Donaldson MS, Lohr KN, editors. Health Data in the Information Age: Use, Disclosure, and Privacy. Washington (DC): National Academies Press (US); 1994. p. 4. Confidentiality and Privacy of Personal Data. https://www.ncbi.nlm.nih.gov/books/NBK236546/.

19. Sun C, Ippel L, van Soest J, Wouters B, Malic A, Adekunle O, et al. A privacy-preserving infrastructure for analyzing personal health data in a vertically partitioned scenario. Stud Health Technol Inform 2019; 264:373–377.

20. Kazmierska J, Hope A, Spezi E, Beddar S, Nailon WH, Osong B, et al. From multisource data to clinical decision aids in radiation oncology: the need for a clinical data science community. Radiother Oncol 2020; 153:43–54.

21. Beyan O, Choudhury A, van Soest J, Kohlbacher O, Zimmermann L, Stenzhorn H, et al. Distributed analytics on sensitive medical data: the personal health train. Data Intell 2019; 2:96–107.

22. Tegbaru D, Braverman L, Zietman AL, Yom SS, Lee WR, Miller RC, et al. ASTRO journals' data sharing policy and recommended best practices. Adv Radiat Oncol 2019; 4:551–558.

23. Deist TM, Dankers FJWM, Ojha P, Scott Marshall M, Janssen T, FaivreFinn C, et al. Distributed learning on 20 000+ lung cancer patients – the personal health train. Radiother Oncol 2020; 144:189–200.

24. van Soest J, Sun C, Mussmann O, Puts M, van den Berg B, Malic A, et al. Using the personal health train for automated and privacy-preserving analytics on vertically partitioned data. Stud Health Technol Inform 2018; 247:581–585.

25. Hamzah FB, Hamzah FM, Razali SFM, Jaafar O, Jamil NA. Imputation methods for recovering streamflow observation: a methodological review. Cogent Environ Sci 2020; 6:1745133.

26. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR guiding principles for scientific data management and stewardship. Sci Data 2016; 3:160018.

27. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018; 68:394–424.

28. Bruni L, Albero G, Serrano B, Mena M, Gómez D, Muñoz J, Bosch FX, de Sanjosé S. ICO/IARC Information Centre on HPV and Cancer (HPV Information Centre). HumanPapillomavirus and Related Diseases in India. Summary Report. 2019. https://hpvcentre.net/statistics/reports/IND.pdf. [Accessed 17 March, 2020].

29. Rose PG, Bundy BN, Watkins EB, Thigpen JT, Deppe G, Maiman MA, et al. Concurrent cisplatin-based radiotherapy and chemotherapy for locally advanced cervical cancer. N Engl J Med 1999; 340:1144–1153.

30. Thomas GM. Improved treatment for cervical cancer–concurrent chemotherapy and radiotherapy. N Engl J Med 1999; 340:1198–1200.

31. Yu KH, Snyder M. Omics profiling in precision oncology. Mol Cell Proteomics 2016; 15:2525–2536.

32. Surasi DS, Bhambhvani P, Baldwin JA, Almodovar SE, O'Malley JP. ^{18}FFDG PET and PET/CT patient preparation: a review of the literature. J Nucl Med Technol 2014; 42:5–13.

33. Sgouros G, Hobbs RF. Dosimetry for radiopharmaceutical therapy. Semin Nucl Med 2014; 44:172–178.

34. Ansart M, Epelbaum S, Gagliardi G, Colliot O, Dormont D, Dubois B, et al.; Alzheimer's Disease Neuroimaging Initiative* and the INSIGHT-preAD study. Reduction of recruitment costs in preclinical AD trials: validation of automatic pre-screening algorithm for brain amyloidosis. Stat Methods Med Res 2020; 29:151–164.

35. Harvey HB, Liu C, Ai J, Jaworsky C, Guerrier CE, Flores E, Pianykh O. Predicting no-shows in radiology using regression modeling of data available in the electronic medical record. J Am Coll Radiol 2017; 14:1303–1309.

36. Shi K, Dong C, Gafita A, Zhao Y. Giles T. Artificial neural network for prediction of post-therapy dosimetry for 177Lu-PSMA I&T therapy. J Nucl Med 2019; 60 (Suppl 1):1185.

37. Xie T, Zaidi H. Estimation of the radiation dose in pregnancy: an automated patient-specific model using convolutional neural networks. Eur Radiol 2019; 29:6805–6815.

38. Zhu B, Liu JZ, Cauley SF, Rosen BR, Rosen MS. Image reconstruction by domain-transform manifold learning. Nature 2018; 555:487–492.

39. Xie T, Akhavanallaf A, Zaidi H. Construction of patient-specific computational models for organ dose estimation in radiological imaging. Med Phys 2019; 46:2403–2411.

40. Xue S, Gafita A, Afshar-Oromieh A, Eiber M, Rominger A, Shi K. Voxel-wise prediction of post-therapy dosimetry for 177Lu-PSMA I&T therapy using deep learning. J Nucl Med 2020; 61 (Suppl 1):1424.

41. Lee CH, Goo JM, Ye HJ, Ye SJ, Park CM, Chun EJ, Im JG. Radiation dose modulation techniques in the multidetector CT era: from basics to practice. Radiographics 2008; 28:1451–1459.

42. Chen H, Zhang Y, Kalra MK, Lin F, Chen Y, Liao P, et al. Low-dose CT with a residual encoder-decoder convolutional neural network. IEEE Trans Med Imaging 2017; 36:2524–2535.

43. IAEA Human Health Series. Quantitative Nuclear Medicine Imaging: concepts, Requirements And Methods. Vienna: International Atomic Energy Agency; 2014. p. 24. https://www-pub.iaea.org/MTCD/Publications/PDF/ Pub1605_web.pdf. [Accessed December 09, 2020].

44. Pichler BJ, Wehrl HF, Judenhofer MS. Latest advances in molecular imaging instrumentation. J Nucl Med 2008; 49:5S–23S.

45. Ozsahin DU, Isa NA, Uzun B, Ozsahin I. Effective analysis of image reconstruction algorithms in nuclear medicine using fuzzy PROMETHEE. 2018 Advances in Science and Engineering Technology International Conferences (ASET); Abu Dhabi; 2018. pp. 1–5.

46. Hong X, Zan Y, Weng F, Tao W, Peng Q, Huang Q. Enhancing the image quality via transferred deep residual learning of coarse PET sinograms. IEEE Trans Med Imaging 2018; 37:2322–2332.

47. Xiang L, Qiao Y, Nie D, An L, Wang Q, Shen D. Deep auto-context convolutional neural networks for standard-dose PET image estimation from low-dose PET/MRI. Neurocomputing 2017; 267:406–416.

48. Wang Y, Yu B, Wang L, Zu C, Lalush DS, Lin W, et al. 3D conditional generative adversarial networks for high-quality PET image estimation at low dose. Neuroimage 2018; 174:550–562.

49. Kim K, Wu D, Gong K, Dutta J, Kim JH, Son YD, et al. Penalized PET reconstruction using deep learning prior and local linear fitting. IEEE Trans Med Imaging 2018; 37:1478–1487.

50. Shiri I, Amir Mozafari Sabet K, Arabi H, Pourkeshavarz M, Teimourian B, Ay MR, Zaidi H. Standard SPECT myocardial perfusion estimation from halftime acquisitions using deep convolutional residual neural networks. J Nucl Cardiol 2020. doi: 10.1007/s12350-020-02119-y

51. Pinilla I, Rodríguez-Vigil B, Gómez-León N. Integrated FDG PET/CT: utility and applications in clinical oncology. Clin Med Oncol 2008; 2:181– 198.

52. Chen Y, An H. Attenuation correction of PET/MR imaging. Magn Reson Imaging Clin N Am 2017; 25:245–255.

53. Arabi H, Zaidi H. Deep learning-guided estimation of attenuation correction factors from time-of-flight PET emission data. Med Image Anal 2020; 64:101718.

54. Arabi H, Bortolin K, Ginovart N, Garibotto V, Zaidi H. Deep learning-guided joint attenuation and scatter correction in multitracer neuroimaging studies. Hum Brain Mapp 2020; 41:3667–3679.

55. Shiri I, Arabi H, Geramifar P, Hajianfar G, Ghafarian P, Rahmim A, et al. Deep-JASC: joint attenuation and scatter correction in whole-body 18F-FDG PET using a deep residual network. Eur J Nucl Med Mol Imaging 2020; 47:2533–2548.

56. Hwang D, Kang SK, Kim KY, Seo S, Paeng JC, Lee DS, Lee JS. Generation of PET attenuation map for whole-body time-of-flight 18F-FDG PET/MRI using a deep

neural network trained with simultaneously reconstructed activity and attenuation maps. J Nucl Med 2019; 60:1183–1189.

57. Liu F, Jang H, Kijowski R, Bradshaw T, McMillan AB. Deep learning MR imaging-based attenuation correction for PET/MR imaging. Radiology 2018; 286:676–684.

58. Pesapane F, Codari M, Sardanelli F. Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine. Eur Radiol Exp 2018; 2:35.

59. Nensa F, Demircioglu A, Rischpler C. Artificial intelligence in nuclear medicine. J Nucl Med 2019; 60 (Suppl 2):29S–37S.

60. Seifert R, Weber M, Kocakavuk E, et al. AI and machine learning in nuclear medicine: future perspectives. Semin Nucl Med 2020. In press. doi: 10.1053/j.semnuclmed.2020.08.003

61. Pinto Dos Santos D, Baeßler B. Big data, artificial intelligence, and structured reporting. Eur Radiol Exp 2018; 2:42.

62. Yordanova A, Eppard E, Kürpig S, Bundschuh RA, Schönberger S, Gonzalez-Carmona M, et al. Theranostics in nuclear medicine practice. Onco Targets Ther 2017; 10:4821–4828.

63. Pankowska V, Malkowski B, Wedrowski M, Wedrowska E, Roszkowski K. FDG PET/CT as a survival prognostic factor in patients with advanced renal cell carcinoma. Clin Exp Med 2019; 19:143–148.

64. Chen HH, Chiu NT, Su WC, Guo HR, Lee BF. Prognostic value of wholebody total lesion glycolysis at pretreatment FDG PET/CT in non-small cell lung cancer. Radiology 2012; 264:559–566.

65. O JH, Lodge MA, Wahl RL. Practical PERCIST: a simplified guide to PET response criteria in solid tumors 1.0. Radiology 2016; 280:576–584.

66. Mahadevaiah G, Rv P, Bermejo I, Jaffray D, Dekker A, Wee L. Artificial intelligence-based clinical decision support in modern medical physics: selection, acceptance, commissioning, and quality assurance. Med Phys 2020; 47:e228–e235.

67. Sanduleanu S, Hamming-Vrieze O, Wesseling FWR, Even AJG, Hoebers FJ, Hoeben A, et al. [18F]-HX4 PET/CT hypoxia in patients with squamous cell carcinoma of the head and neck treated with chemoradiotherapy: prognostic results from two prospective trials. Clin Transl Radiat Oncol 2020; 23:9–15.

68. Betancur J, Hu LH, Commandeur F, Sharir T, Einstein AJ, Fish MB, et al. Deep learning analysis of upright-supine high-efficiency SPECT myocardial perfusion imaging for prediction of obstructive coronary artery disease: a multicenter study. J Nucl Med 2019; 60:664–670.

69. Betancur J, Otaki Y, Motwani M, Fish MB, Lemley M, Dey D, et al. Prognostic value of combined clinical and myocardial perfusion imaging data using machine learning. JACC Cardiovasc Imaging 2018; 11:1000–1009.

70. Rios R, Hu LH, Einstein A, Sharir T. Prediction of major adverse cardiac events using reduced clinical features in machine learning: results from REFINE SPECT registry. J Nucl Med 2020; 61 (Suppl 1): 665.

71. Shiri I, Maleki H, Hajianfar G, Abdollahi H, Ashrafinia S, Hatt M, et al. Nextgeneration radiogenomics sequencing for prediction of EGFR and KRAS mutation status in NSCLC patients using multimodal imaging and machine learning algorithms. Mol Imaging Biol 2020; 22:1132–1148.

72. Kidd EA, El Naqa I, Siegel BA, Dehdashti F, Grigsby PW. FDG-PET-based prognostic nomograms for locally advanced cervical cancer. Gynecol Oncol 2012; 127:136–140.

73. Foley KG, Shi Z, Whybra P, Kalendralis P, Larue R, Berbee M, et al. External validation of a prognostic model incorporating quantitative PET image features in oesophageal cancer. Radiother Oncol 2019; 133:205–212.

74. van Timmeren JE, Carvalho S, Leijenaar RTH, Troost EGC, van Elmpt W, de Ruysscher D, et al. Challenges and caveats of a multi-center retrospective radiomics study: an example of early treatment response assessment for NSCLC patients using FDG-PET/CT radiomics. PLoS One 2019; 14:e0217536.

75. van Timmeren JE, Leijenaar RTH, van Elmpt W, Wang J, Zhang Z, Dekker A, Lambin P. Test-retest data for radiomics feature stability analysis: generalizable or study-specific? Tomography 2016; 2:361– 365.

76. Ker J, Wang L, Rao J, Lim T. Deep learning applications in medical image analysis. IEEE Access 2018; 6:9375–9389.

77. Yang Z, Zhang A, Sudjianto A. Enhancing explainability of neural networks through architecture constraints. IEEE Trans Neural Netw Learn Syst 2020; 27:1–12

Chapter 3: Systematic Review and Meta-analysis of Prediction Models used in cervical cancer

Adapted from Jha, A. K., Mithun, S., Rangarajan, V., Wee, L., & Dekker, A. Systematic Review and Meta-analysis of Prediction Models used in cervical cancer, Submission is under review

Abstract

Background: Cervical cancer is one of the most common cancers in women with an incidence of around 6.5 percent of all the cancer in women worldwide. Early detection and adequate treatment according to staging improve the life expectancy of the patient. Outcome prediction models might aid treatment decisions, but a systematic review on prediction models for cervical cancer patients is not available.

Design: We performed a systematic review for prediction models in cervical cancer following PRISMA guidelines. Key features that were used for model training and validation, the endpoints were extracted from the article and data were analyzed. Selected articles were grouped based on prediction endpoints i.e. **Group1:** Overall survival, **Group2:** progression-free survival; **Group3:** recurrence or distant metastasis; **Group4:** treatment response; **Group5:** toxicity or quality of life. We developed scoring system to evaluate the manuscript. As per our criteria manuscript scored more than 50% were considered significant. A meta-analysis was performed for all the groups separately.

Results: The first line of search selected 1358 articles and finally 39 articles were selected as eligible for inclusion in review. As per our assessment criteria 33articles were found to be significant and 6 were less significant. The intra-group pooled correlation coefficient for Group1, Group2, Group3, Group4, and Group5 were 0.76 [0.72, 0.79], 0.80 [0.73, 0.86], 0.87 [0.83, 0.90], 0.85 [0.77, 0.90], 0.88 [0.85, 0.90] respectively. All the models found to be good (prediction accuracy [c-index/AUC/R^2] >0.6) in endpoint prediction.

Conclusions: Prediction models of cervical cancer toxicity, local or distant recurrence and survival prediction show promising results with reasonable prediction accuracy [c-index/AUC/R^2>0.6]. These models should also be validated on external data and evaluated in prospective clinical studies.

Keywords: Systematic review, cervical cancer, prediction models, Radiomics

Introduction

As per GLOBOCAN2020, cervical cancer is the 4th most common cancer and contributes around 6.5% of all malignancies in women worldwide.[1] The role of Human papillomavirus (HPV) infection is a well-known cause of cervical cancer in women.[1-5] The newer technologies, awareness, and screening have reduced the growth of cervical cancer in the last decade.[1] The Catalan Institute of Oncology (ICO) / the International Agency for Research on Cancer (IARC) Information Centre on HPV-related cancers states that 2,784 million women aged 15 years and above are at risk of developing cancer worldwide.[1] Every year 604,127 are diagnosed and 341,831 die from the disease worldwide.[1]

Current research in cervical cancer focuses on two major issues; 1) cervical cancer screening for early detection of disease and 2) personalized treatment to increase overall survival (OS), progression-free survival (PFS), quality of life(QOL), and reduce the toxicity and loc-regional recurrence (LRR).[6,7] Practically, the therapeutic choice in the treatment of cervical cancer depends on many factors: the tumor size, parametrial involvement, the spread of the tumor into the blood or lymph vessels (lymphovascular space invasion), deep cervical stromal invasion.[8-13] According to the International Federation of Gynecology and Obstetrics (FIGO) guidelines, the best-known standard of care for early-stage cervical cancer (up to stage IB) is surgical resection. Neo-adjuvant external beam radiotherapy (with concomitant chemotherapy) followed by intracavitary brachytherapy boost irradiation is required for locally advanced disease (stage IB3 and above). For metastatic disease (stage IVB), clinical management is unclear but often palliative care is offered consisting of either radiotherapy, chemotherapy, or a combination.[8] As per the American Society of Cancer, 5-year relative survival rates in cervical cancer are 99%, 56%, 17%, and 66% for localized, regional, distant disease and across all the stages respectively .[13]

With the growing technologies and medical informatics in the last few decades, a large volume of cancer treatment data has been generated and stored in (un)structured electronic formats worldwide.[15,16] This helped machine learning in oncology, gradually taking the forefront in cancer treatment and management due to the insightful and actionable information they can provide on a patient's treatment and outcome .[17-24] These generated insights could help caregivers select the appropriate treatment for a patient or modify their treatment options to improve the outcome. This is very important particularly for high-risk patients responding very poorly to standard treatment protocols.

One of the sources for machine learning in oncology is imaging data. Routine clinical images particularly Positron Emission Tomography/Computed Tomography (PET/CT) and Magnetic Resonance Imaging (MRI) are used for diagnosis, pre-treatment staging, intervention planning, and subsequent follow-up examination and have in recent years

led into a machine learning gold mine .[9-13, 17-19] Machine learning technologies such as computer-aided diagnosis (CAD) and auto-contouring have been explored by researchers to assist in the interpretation of medical images and diagnosis.[20-25] Also, in personalized medicine, where patients are grouped into sub-populations based on their disease susceptibility or treatment response, machine learning technologies can identify a patient's specific tumor information within the pixels of the medical image which can be extracted using radiomic and deep learning technique for an optimal therapeutic intervention.

Machine learning methods have the potential to automate the treatment process, improve the quality of care provided to patients and optimize treatment outcomes in cervical cancer. The application of machine learning methods to medical images and omics data has been discussed extensively in the literature .[11, 12, 17-18]

Several predictive models have been developed, tested and utilized in the last decade in the screening and treatment of cervical cancer for different outcomes of interest .[20-28] In our study we perform a systematic review of articles on prediction models used in cervical cancer. Specifically, the evaluation of accuracy of prediction is the primary and a technological assessment of the development of prediction models is the secondary objective of this review.

Methods

Prospective protocol registration

A prospective review protocol has not been registered for this study.

Eligibility criteria

For this review, we sought original research articles relating to (1) gynecological malignancies of the cervix of all FIGO stages and all histology groups; (2) prognostic or predictive pre-treatment computer modeling of treatment outcomes i.e. overall survival (OS) (5 or more year) disease-free survival (DFS) (2 or more year), distance metastasis (DM), local-regional recurrence (LRR), treatment-related toxicities and quality of life ; 3) full-text available, and published solely in the English language; 4) Articles published between January 2010 to June 2020;

Exclusion criteria:

Additionally, we discarded articles based on the following exclusion criteria 1) study does not address human clinical subjects (phantom, in vitro, or animal studies); 2) study does not develop and test an outcome prognosis and prediction model; 3) study only predicts a screening outcome of cervical cancer, or 4) study that only uses gene sequencing data for prediction model development

Information Source and Search Method

We searched for eligible publications in one electronic database (PubMed). The base strategy was a high-sensitive search for diagnostic and prognostic models, using a combination of the broad Haynes[29] and Ingui[30] filters, with the additional modification proposed by Geersing[31], all combined with the "AND" logical operator.

A PubMed search[("Cc") OR ("Ccx") OR ("Cancer of the cervix") OR ("Cervix cancer") OR ("cervix Tumors")] AND [("Risk Model") OR ("Risk assessment model") OR ("Risk prediction model") OR ("Assessment tool") OR ("Prediction score")] retrieved (19,146) articles. A second PubMed search with modified keywords [("Cancer of the cervix") OR ("Cervix cancer") OR ("cervix Tumors")] AND [("Risk prediction model") OR ("Prediction model")] reduced the number of retrieved articles to (11,346 results). Finally, a PubMed search with limited keywords(("Cervix cancer"[All Fields] OR "Cervical cancer"[All Fields]) OR "Cancer Cervix"[All Fields]) AND (("Prediction model"[All Fields] OR "Prognostic model"[All Fields] OR "radiomic model"[All Fields] OR "radiomics model"[All Fields] OR "risk prediction"[All Fields] OR "risk prediction model"[All Fields] OR "Prediction"[All Fields] OR "Risk assessment"[All Fields] OR "nomogram"[All Fields] OR "nomograms"[All Fields])) reduced the number to 1358 articles.

Study selection

1354 abstracts selected in the final search were stored in tabular format and reviewed independently by the first and second authors (AKJ and SM) and the two senior authors (VR and LW). Finally, 39 articles were selected based on the aforementioned criteria. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flowchart shows the selection of articles for this review (Figure 1). [32] The full manuscripts of the selected article were downloaded from a university library or other sources.

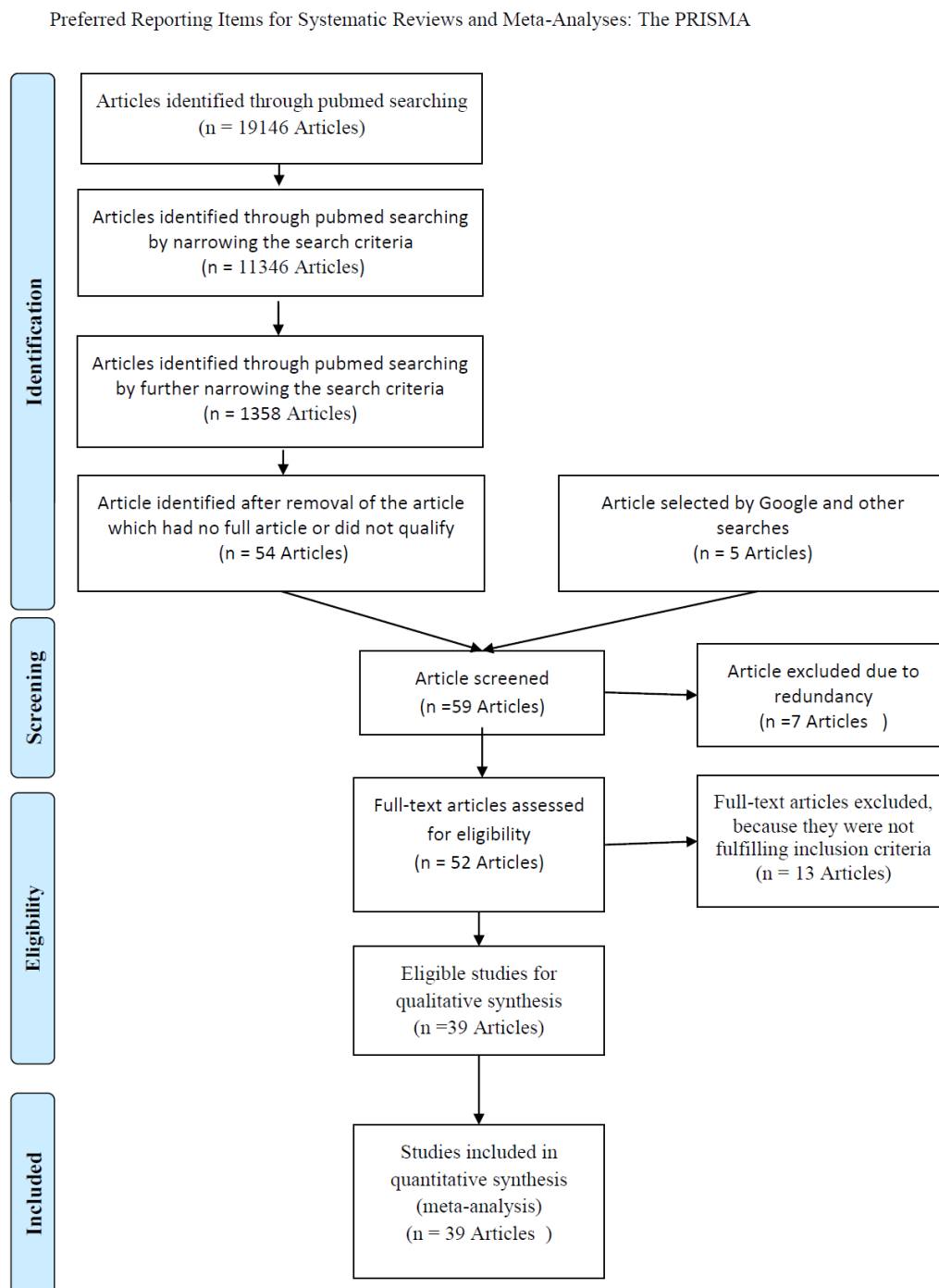Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA

Figure 1: Meta-Analyses (PRISMA) workflow to select articles for review

Data extraction

Data extraction included the study type, sample size, characteristics of the patient cohort, FIGO staging, histology, treatment provided, features used, feature selection methods used, prediction endpoints, prediction model used, and performance assessment of the prediction model i.e., internal or external validation.

Data Analysis:

The included studies were grouped by clinical endpoint used to build the model. Studies were classified into five groups based on prediction endpoint, **Group1:** overall/disease-specific survival, **Group2:** disease/progression-free survival; **Group3:** recurrence (local/nodal) or distant metastasis; **Group4:** treatment response; **Group5:** treatment-related toxicity or quality of life. Model performance was evaluated in their respective groups based on their performance in that group.

Evaluation of methodological quality of reviewed studies

We systematically evaluated the studies in this review based on the following criteria: 1) sample size and event rate 2) disease stage 3) histology stage 4) mention of treatment 5) feature used in study 6) feature selection method used 7) prediction model used 8) statistical tests used to evaluate the performance of prediction model 9) internal or external validation and 10) single or multi canter study. Data were extracted in tabular form under these heading for qualitative analysis.

Prediction Model Quality Score (PMQS):

Transparent Reporting of studies on prediction models for Individual Prognosis Or Diagnosis (TRIPOD) guideline is used for extracting the relevant information and calculating summary scores to determine adherence of scientific articles on primary prediction model reports[33]. This is an exhaustive tool that assesses the overall reporting of the prediction model development and validation. Whereas Radiomic quality score (RQS) is a guideline and scoring system to evaluate the radiomic based prediction model development, validation, and reporting[34]. In this study, we have reviewed the scientific literature on the prediction model and quality of prediction model used in cervical cancer. As TRIPOD evaluates the overall adherence of prediction model development and reporting and RQS only evaluate radiomic based prediction model, none of the existing scoring systems is suitable to evaluate the quality of prediction modeling studies only based on the development and validation of the prediction model hence here we propose a scoring system to evaluate the study based on development and validation prediction mode. A scoring system was developed and discussed among all the authors as mentioned in table 2. Articles with score >50% are considered as *significant* and score *< 50% less significant.*

Meta analysis of model performance:

We also performed a Meta-analysis using "metafor package" in R 4.1.0 open-source software.[35]A Meta-analysis was performed using a Random-Effects Model for specific prediction endpoints i.e., Overall survival (OS), Progression-free survival (PFS), Recurrence/Distant metastasis, Treatment response, and Toxicity/Quality of life as a separate analysis. The prediction score of a prediction model (AUC/C-index) in a validation set was considered for Meta-analysis and the AUC/C-index was pooled on the logit scale and the standard errors of the logit transformed AUC/C-Index were derived. Subsequently, the AUC/C-index was summarized using the inverse variance method random-effects model, estimated with restricted maximum likelihood and Knapp and Hartung adjustment to generate 95% CIs. The statistical heterogeneity was explored by estimating I^2 statistics.[35]An AUC/C-index higher than 0.6 was considered as a good prediction capability. The forest plots for each prediction endpoints were generated separately.

Results

Selection of studies

Out of the 39 selected articles[36-74] in this study, 25 studies had overall survival (OS)/Disease-specific survival (DSS) as an outcome, 10 studies had midterm disease-free survival (DSF) or progression-free survival (PFS) as an outcome, 11studies had recurrence or metastasis as the outcome, 3 studies had treatment response as an outcome, 2 studies had toxicity as the outcome and 1 study has the quality of life (QOL) as an outcome (table) (supplementary table 1).

Sample size:

All articles in this review were retrospective studies with a sample size ranging from 42 to 4200 (table 1).

First Author/Year of publication	Sample size (patients)	Feature used	Event Rate reported	Multi-centre study	Cancer/ FIGO stage/ Histology	Treatment	End point to predict	Model used	Maximum accuracy of best model in Validation Set
Lee WK et. al.36/2020	270	Age, TS, stage, SSCCA, and HPV Nodal SUVmax, primary tumor SUVmax,	No	No	LACC /IIB and ≥IIB/SCC,AC,ASC	RT, and CT	2-year DFS, 5-year OS and	NORM-mCox	2-year DFS: AUC; 0.75 (95% CI, 0.69–0.81) 5-year OS: AUC; 0.78 (95% CI, 0.71–0.85)
Paik ES et. al.37/2020	1,441	Age, BMI, TS, SSCCA, Hb, and platelet level, FIGO Stage, Histology, LVSI, Invasion depth, LNM, PMI, Resection margin	No	No	ESCC/ IB – IIA/SCC, AC	CCRT	2 years and 5 year DFS and OS	mCPHM	AUC at 2 and 5 years were 0.799/0.723 for DFS, and 0.844/0.806 for OS
Tian X et. al.38/2020	277	TD, LNM, FIGO Stage Radiomic feature: (glcm_correlation, LLH_glcm_entropy, HLL_glrlm_GLN, LHH_glcm_IDMN, glcm_homogeneity, "HHL_glrlm_SRHGLE	No	Yes (2-centre)	LACC /IB2, IIA2, IIB-III/SCC	NACT followed by SY/RT	Response of chemotherapy	RF	Raadiomic model: AUC: 0.816 (95% CI, 0.690-0.942), Combined model: AUC 0.821 (95% CI, 0.697-0.946)
Fang M et. al.39/2020	120	Age, Pregnancy Num, Parturition Num, Abortion Num, First age of sexual intercourse, Family history of cancer	No	No	LACC / IB2-IIB/NA	CCRT	Treatment response	RF	AUC of 0.798 (95% CI: 0.678-0.917)

		three habitat signatures from sagittal T2 image, axial T1 enhanced-MRI image, and ADC image (X_GLRLM_LRHGLE, X_GLRLM_RPXLL_GLRLM_SRLGLE, X_GLRLM_LRHGLE, X_GLRLM_LRE, X_GLRLM_SRLGLE, X_GLCM_variance, XLL_GLRLM_RLN XHH_GLRLM_RP, Surface_area, XLL_H_skewness, X_GLCM_dissimilarity, X_GLCM_homogeneity)							
Dong T et. al.40/2020	226	Five radiomic features: wavelet-LHH_firstorder_Mean, original_glszm_GrayLevelVariance, log-sigma-3-0-mm-3D_firstorder_Skewness, original_gldm_Small DependenceHighGrayLevelEmp	No	Yes (2-centre)	ESCC /IA-IIB/NA	SY	Lymph Node Status	LRM, SVM, DNN	Internal Validation: AUC of 0.99 and AC of 97.16% External Validation: AUC = 0.90

		hasis, or iginal_glcm_JointAverage Two clinicopathological features: tumor histology, and grade							
Fang J et. al.41/2020	248	Age, histology, grade, HPV, SSCCA, LNM, LVI, W_HH_FO_MED, O_S_FLT, W_HL_FO_MIN, W_HL_GLCM_SUMENTR, W_LH_GLDM_LDHGLE, W_LH_GLSZM_SALGLE, W_HL_GLCM_AUTOCOR, W_LH_GLDM_LRLGLE, T2_W_LH_GLSZM_SZNU,T2_O_GLRLM_LRLGLE,T2_LG5_GLRLM_LRLGLE, T2_O_FO_10PER,T2_O_GLRLM_LGLRE, T2_LG5_FO_SKEW, T2_LG5_GLSZMO_SZNUN, T2_LG4_FO_MAX	No	Yes (2-centre)	ESCC / IB-IIA/ SCC,AC,ASC, SMCC	SY	3 year DFS	NORM-mCox	Radiomics features: DFS (C-index, 0.753; 95% CI: 0.696-0.805) Clinicopathological features: DFS (C-index, 0.632; 95% CI: 0.567-0.700). Combined model : DFS (C-index, 0.714; 95%CI: 0.642-0.784)
Wang T et. al.42/2019	137	Age, grade, Radiomics feature mean(SD)	No	No	ESCC;/IB-IIA/NA	SY with PLN dissection surgery	PLNM	MLR	Radiomics nomograms C-index: 0.922(95% CI, 0.825–1)

		extracted from T2WI and DWI						Clinical nomograms C-index: 0.799(95% CI, 0.599-0.998)
Wang W et. al.43/2019	833	histology, FIGOstage, LNM, number and diameter of PLNM	Yes	No	Cervical cancer /I-IV/ SCC,AC ,ASC, others	DRT or CCRT	3- and 5-year OS, DFS, LC, DMF	NORM-mCOX	C-Index of 0.73 for predicting OS, 0.71 for DFS, 0.73 for LC and 0.67 for DMF.
Shen WC et. al.44/2019	142	18F-FDG PET/CT	No	No	LACC/ IB1-IIIB/SCC,AD	SY	LRR, DM	Deep learning model	Local recurrence: AC=0.89 Distant metastasis : AC= 0.87
Zhang S et. al.45/2019	1252	Age, Race, Pathological type, Metastasis numbers radiotherapy, and chemotherapy	No	No	MCC/I-IV/ SCC,AC	RT and CT	1- and 3-year OS	NORM-mCOX	1 year OS C-index=0.779, 3 year OS C-index=0.787
Obrzut B et. al.46/2019	102	Age, BMI, HS, POCD, FIGO stage, histologic type, grade, TS, LNM, NLND, NPLN, LNR, LVSI, SMS, PMI, deep SI, PORT, ST, MBL, PIOC, POC	No	No	ESCC/I A2-IIB2 / SCC,AC	SY	10 year OS	PNN, MLR	AUC ROC= 0.809
Yang J et. al.47/2019	389 9	Age, Marital status, Race, Histology Grade, FIGO, Tumour size, Lymph node density (LND)	No	No	Cervical cancer / I-IV/ SCC,AC ,ASC, others	RT/CT /CCRT	3- and 5-year OS and DSS	NORM-mCox	OS: Nomogram 3-year, AUC = 0.673 5-year, AUC = 0.675 DSS:Nomogram 3-year, AUC = 0.693 5-year, AUC = 0.693

Lucia F et. al.48/2019	190	ADC EntropyGLCM, PET GLNUGLRLM, Tumor size FIGO, Nodal stage, Volume	No	Yes (3-centre 2-countries)	LACC/IA2-IV-B/ SCC,AC ,ASC, CCC	CT and RT	2-year DFS and LRC	mCP HM	DFS model: AC= 90% (95% CI [79-98%]) LRC model: AC= 98% (95% CI [90-99%]) in the French cohort AC= 96% (95% CI [80-99%]) in the Canadian cohort.
Matsuo K et. al.49/2019	768	Age, BMI, Race, Blood urea, nitrogen, Creatinine, Albumin, Bicarbonate, Hemoglobin, stage, Histology, Treatment	Yes	No	cervical cancer / I-IV/ SCC,AC ,ASC, others	CT and RT	PFS and OS	Deep-learning model CPH	PFS: C-Index 0.795 ±0.066 OS: C-Index 0.616 ±0.041
Sun C et. al.50/2019	275	Age, FIGOStage, GLSZM-LGLZE on T1WI, NGTDM_busyness of T2WI, and NGTDM_complexity feature in the peritumoural zone of T2WI	No	No	LACC/IB2-IIIB/NA	NACT	Clinical response to neoadjuvant chemotherapy	SVM	AUC 0.999
Wang C et. al.51/2018	8202	Age, race, marital status, grade, FIGO stage, histology, TS and LODDS	No	No	cervical cancer / I-IV/ SCC,AC ,ASC, others	SY, CCRT, CT and RT, BT	3year and 5 year OS	NORM-mCox	5 year OS BSCCI =0.786 (95% CI, 0.764 to 0.808)
Marchetti C et. al.52/2018	245	TS, grade, and PMI	Yes	No	LACC/IB2-IIIB/ SCC,AC ,ASC	NACT and SY	5-year OS rates and the 2-year	NORM-mCox	5-year OS of 85% 2-year PFS of 98%

							PFS rates		
Altazi AB et. al.53/2018	80	Radiomics features: Difference Entropy and 2nd order mean SUV Max and SUV peak, andTumor Vol. and Surface Area	No	No	Cervical cancer /I to IV/SCC	RT	DM and LRR	MLR	LRR:(AUC: 0.89 95%CI: 0.78–1.00) DM: (AUC: 0.82. 95%CI: 0.69–0.94).
Chen J et. al.54/2018	42	dose geometric parameters (DGPs), dose volume parameters (DVPs), Radiomics features	Yes	No	Cervical cancer	RT and BT	Rectum toxicity	SVM	AUC=0.91 (95% CI : 0.87–0.93)
Zhen X et. al.55/2017	42	Dose distribution in rectum Computer vision 43 Radiomics texture features from the RSDM were used	Yes	No	Cervical cancer/ I to IV/SCC	RT and BT	Rectum toxicity	Pre-trained CNN (VGG-16)	AUC: 0.89
Rose P G et. al.56/2017	2,042	Histology, Race/ethnicity, performance status, TS, FIGO Stage, Grade, PLN, Treatment	No	No	Cervical cancer/ I to IV/ SCC, AC, ASC	RT and BT	2-year PFS, 5-year OS,	NORM-mCox	The 2-year PFS BCCI of 0.62. 5-year OS BCCI of 0.64.
Zheng RR et. al.67/2017	795	BMI, albumin, platelet, leukocyte, tumor differentiation, and status of PLN	No	No	ESCC/ IA1-IIA/SCC and non-SCC	SY	3- and 5-year OS	NORM-mCox	C-index 0.74 (95% CI, 0.68-0.80).
ReuzéS et. al.58/2017	118	PET radiomic features (LGZE, HGZE, Entropy), and	No	No	Cervical cancer/ I - IV/SCC,	CCRT followed by BT	Local tumor recurrence	MLR	AUC = 0.86 (95% C.I.: 0.75-0.97)

		SUVmean, SUVmax, SUVpeak, TLG, MV			AC, ASC				
Lucia F et. al.59/2017	102	Radiomic features PET (GLNUGLRLM-QE) and 4 from MRI (Inverse varianceGLCM-QF and EntropyGLCM-QF in ADC maps from DWI, RLVARGLRLM-QL in CE-MRI and LZLGEGLSZM-QF in T2)	Yes	No	Cervical cancer/ I – IV/ SCC, AC, ASC	RT and BT	3years LRC and DFS	NORM-mCox	3-year LRC AUC= 0.94 3-year DFS AUC = 0.95
Obrzut B et. al.60/2017	102	Age, BMI, HS, POCD, FIGO stage, histologic type, grade, TS, LNM, NLND, NPLN, LNR, LVSI, SMS, PMI, deep SI, PORT, ST, MBL, PIOC, POC	No	No	ESCC/I A2-IIB2/SCC and non-SCC	SY and pelvic lymphadenectomy, Some of the patients received adjuvant RT and BT	5-year OS	Six Models: PNN, MLP, GEP, SVM, RBF NN and k–Means method	AUC ROC= 0.818 and p < 0.001
Shim SH et. al.61/2016	245	TS on MRI and PALNM on PET/CT	No	No	Cervical cancer/ I to IV/ SCC, AC, ASC	RT, BT, CT	PALNM	MLR	BSCCI =0.886 (95% CI, 0.825-0.947)
Kong TW et. al.62/2016	298	DBTV and DCSR on MRI, serum SCC-Ag, MPS	No	No	ESCC/ IB1 ,IB2/ SCC,	SY	PMI	NORM-mCox	C index: 0.940 (95% CI, 0.908-0.967)

					AC, ASC			
Zhou H et. al.63/2015	1,563	LNM, LVSI, SI, PMI, TD, histology	Yes	No	ESCC/ IA-IIB/SCC, non-SCC	SY	5-year OS	NORM-mCox	The C-index : 0.71 (95% CI, 0.65 to 0.77),
Kim DY et. al.64/2014	493	Age, TS on MRI, LNM on PET/CT	No	No	ESCC/ IA-IIB/ SCC, AC, ASC, SMCC	SY and pelvic/para-aortic lymphadenectomy	LNM	NORM-mLR	C-index : 0.825 (95% CI, 0.736-0.895).
Kumar S et. al.65/2014	198	Education, Marital Status, Menopausal Status, Parity, Abdominal Mass, Staging(FIGO), Cell Type, Treatment Modality	No	No	Cervical cancer; I to IV; SCC, AC	SY, and CCRT	predicting post treatment HR QoL	PrediQt-Cx, (SVM (Linear); SVM (RBF); MLR; ANN	AUC: 0.866 (0.85-0.90)
Je HU et. al.66/2014	1069	Histology, PLNM, depth of SI, PMI	No	No	ESCC/ IA-IIA/ SCC, others	SY followed by adjuvant RT	5 years DMFS	NORM-mCox	C index: 0.71 (95% CI. 0.655-0.768) Externally validated c-index: 0.65
Shim SH et. al.67/2013	209	Histology, TS on MRI and PALNM on MRI	No	No	Cervical cancer/ I – IV/SCC, AC, ASC	RT, BT, CT	3year and 5year OS	NORM-mCox	BACC I= 0.69 (95% CI, 0.62-0.81)
Lee HJ et. al.68 /2013	1702	Age, no of PLN, LVI , CCRT, PMI	Yes	Yes (10-centre)	ESCC/ IA, IB, IIA/SCC and others	CCRT	3 year and 5-year OS	NORM-mCox	CI index: 0.69
Kang S et. al.69/2012	549	Histology SSCCA,	Yes	Yes	Cervical cancer/ I to IV/	CCRT, with or without	DM	NORM-	BACCI: 0.73 (95% CI, 0.65 to 0.81)

		PLNM on PET PALNM on PET		(3-centre)	SCC, AC, ASC	extended-field radiation		mCox	
Kidd EA et. al.70/2012	234	cervix tumor SUVmax , PTTV, PLNS, FIGO	No	No	Cervical cancer/ I - IV/SCC, AC, ASC	DRT or CCRT	3 year RFS, DSS and OS	NORM-mCox	Validation set CI index (PET nomograms) : RFS : 0.740±0.011, DSS: 0.739±0.153, OS: 0.658±0.105 CI index (FIGO nomograms): RFS : 0.605±0.086, DSS: 0.600±0.132, OS: 0.559±0.152
Polterauer S et. al.71/2012	528	The FIGO stage, TS, age, histology, LNR, and PMI	Yes	Yes (2-centre)	Cervical cancer/ I to IV/SCC, AC/ASC	SYor trachelectomy plus systematic pelvic and/or paraaortic Lymphadenectomy. or CTRT	5 year-OS	NORM-mCox	BSCCI: 0.723 (CI , 0.701 - 0.743)
Seo Y et. al.72/2011	549	Age, hemoglobin level before RT, FIGO stage, MTD, LN status, and RT dose at Point A	Yes	No	Cervical cancer/ I - IV/ SCC, AC, ASC	RT	5-year OS	NORM-mCox	BSCCI: 0.67
Biewenga P et. al.73/2011	710	PMI,TD, AC, ASC, LNM, DI, and LVSI	Yes	No	ESCC; IA2-IIA; SCC and others	SY	5-year DSS	NORM-mCox	AUC= 0.85 (95% CI, 0.79-0.92).

Tseng JY et. al.74/2010	251	Age, SSCCA, TS, PMI, HDN, B/RI and LNM	Yes	No	Cervical cancer;/ I – IV/ SCC, and others	CCRT	5 year OS	NORM-mCox	C-Indux: 0.69

AC=adenocarcinoma ; ASC= adenosquamous cell carcinoma ; EORTC= The European Organisation for Research and Treatment of Cancer; PLNM= Pelvic lymph node metastasis; PALNM= Pearaaiortic lymph node metastasis; LC= local control; LRC= locoregional control; DM=distant metastases; DI=depth of invasion ; LRR=local-regional recurrent disease; PMI= Parametrial infiltration; SI=stromal invasion, TD= tumor diameter, TV= tumor volume; TS=tumor size; LNM= lymph node metastasis; LRC=loco-regional control; DBTV=diameter-based tumor volume; DCSR=disruption of the cervical stromal ring; DMFS=distant metastasis-free survival; MRI= magnetic resonance imaging; PET/CT= positron emission tomography/computed tomography; SCC-Ag= squamous cell carcinoma antigen level; MPS=menopausal status; LVI=Lymphovascular invasion; B/RI=bladder/rectum invasion; HDT= hydronephrosis; SSCCA= serum squamous cell carcinoma antigen; PTTV=PET tumor volume; PLNS= PET LN status; HS=hormonal status; POCD=presence of concomitant diseases; NLND=number of lymph nodes dissected; NPLN=number of positive lymph nodes, LNR= lymph node ratio (ratio of positive to totally removed lymph nodes); LVSI=lymph-vascular space invasion; SMS=surgical margins status; PORT=postoperative radiotherapy; ST=surgery time; MBL=median blood lost; PIOC=presence of intraoperative complications; POC=presence of postoperative complication; GLSZM-LGLZE= low grey-level zone emphasis of the grey-level size zone matrix; NGTDM_busyness = neighborhood greytonedifferencematrix (NGTDM)_busyness; HGRE= High Graylevel Run Emphasis; LGZE=Low Gray-level Zone Emphasis; MV= Metabolic volume; W_HH_FO_MED = CET1w_wavelet-HH_firstorder_Media; O_S_FLT = CET1w_original_shape_Flatness; W_HL_FO_MIN = CET1w_wavelet-HL_firstorder_Minimum; W_HL_GLCM_SUMENTR = CET1w_wavelet-HL_glcm_SumEntropy, W_LH_GLDM_LDHGLE= CET1w_wavelet-LH_gldm_LargeDependenceHighGrayLevelEmphasis; W_LH_GLSZM_SALGLE= CET1w_wavelet-LH_glszm_SmallAreaLowGrayLevelEmphasis; W_HL_GLCM_AUTOCOR= CET1w_wavelet-HL_glcm_Autocorrelation, CET1w_wavelet-LL_glcm_Imc2, CET1w_log-sigma-4-o-mm-3D_firstorder_90Percentile; W_LH_GLDM_LRLGLE = CET1w_wavelet-LH_gldm_SmallDependenceLowGrayLevelEmphasis; T2_W_LH_GLSZM_SZNU = T2w_wavelet-LH_glszm_SizeZoneNonUniformity; T2_O_GLRLM_LRLGLE = T2w_original_glrlm_LongRunLowGrayLevelEmphasis;T2_LG5_GLRLM_LRLGLE = T2w_log-sigma-5-o-mm-3D_glrlm_LongRunLowGrayLevelEmphasis; T2_O_FO_10PER= T2w_original_firstorder_10Percentile; T2_O_GLRLM_LGLRE = T2w_original_glrlm_LowGrayLevelRunEmphasis; T2_LG5_FO_SKEW = T2w_log-sigma-5-o-mm-3D_firstorder_Skewness; T2_LG5_GLSZMO_SZNUN = T2w_log-sigma-5-o-mm3D_glszm_SizeZoneNonUniformityNormalized; T2_LG4_FO_MAX =T2w_log-sigma-4-o-mm-3D_firstorder_Maximum;CCRT= Concurrent chemotherapy ; ESCC= Early-stage cervical cancer; LACC :=Locally advanced cervical cancer; SCC = squamous cell carcinoma; AC = adenocarcinoma; ASC = adenosquamous cell carcinoma; MCC=Metastatic cervical cancer; SMCC= small cell carcinoma; CCC=Clear cell carcinoma; NACT=neoadjuvant chemotherapy ; CCRT=Concurrent chemo-radiation; RT=Radiotherapy; DRT= Definitive radiotherapy ; BT= brachytherapy ; CT=chemotherapy; SY= surgery; OS: Over All survival; PFS: Progression Free Survival; DSS: Disease Specific Survival; MLR: Multivariate Logistic regression; SVM: Support vector machine; PNN : Probable Neural Network; ANN: Artificial Neural Network; DNN : Deep Neural Network; MLP: Multilayer Perceptron; GEP: Gene Expression Programming; RBFNN: Radial basis function networks; mCPHM: multivariate Cox Proportional Hazard

Model; NORM-mLR: Nomogram Multivariate Logistic Regression; NORM-mCox: Nomogram Multivariate Cox Model

Table 1: List of papers selected in this review and important parameters extracted from the manuscripts

Disease:

Nineteen studies included early-stage cervical cancer (ESCC)/Locally advanced cervical cancer (LACC) i.e., FIGO Stage IA-IIIB, and others included all stages i.e., FIGO stage I-IV (table 1).

Feature selection:

The studies selected for this review have used clinical, pathology, imaging (like tumor related imaging finding, tumor size, SUV on PET image) or radiomics features, or a combination of these to develop the prediction model. In seven studies the feature selection process was not mentioned and in the rest of the studies feature selection was performed using univariate logistic regression, univariate Cox regression, SVM-RFE, LASSO, SBS, LOOCV and other techniques details provided in Supplementary table 2.

Prediction algorithm used:

Several prediction algorithms were used by various authors and all of the algorithms showed good prediction (AUC/C-index>0.6) in the validation set. The commonly used algorithms were multivariate logistic regression (MLR), multivariate Cox proportional hazard model (Cox-mPH), Random Forest (RF), Support Vector Machines (SVM), Convoluted Neural Network (CNN) VGG-16, Probabilistic neural network (PNN), Artificial neural network (ANN), multilayer perceptron network (MLP), gene expression programming classifier (GEP), radial basis function neural network (RBFNN) (table 1).

Logistic Regression:

Seven studies included in this review have used MLR as a prediction algorithm. Six studies have predicted recurrence or DM except Kumar S et.al.[65] who has predicted QOL. Dong T et.al.[40], Obrzut B et.al.[46], and Kumar S et.al.[65] have compared MLR with SVM, PNN, or ANN.

Probabilistic neural network (PNN):

Obrzut B. et.al.[60]selected twenty-three demographic, tumor-related, and preoperative parameters to train a PNN, MLP, GEP, SVM, RBFNN, and k–Means method for predicting 5-year overall survival. PNN outperformed all the models.

Convolutional neural network (CNN)

Zhen X. et.al.[55] used a pre-trained VGG-16[75] CNN with transfer learning technique to predict toxicity in rectal toxicity. Given the very small sample size (42 patients) in their study with a relatively small number of toxicity cases, an adaptive synthetic sampling approach (ADASYN) (He et.al. 2008)[76]was used to generate synthetic minority toxicity data to balance the training cohort. In this study VGG-16 model outperformed a radionics based multivariate regression model.

Support Vector Machine (SVM)

In five studies SVM algorithm was used. Chen J. et.al.[54] has developed an SVM model to predict rectal toxicity by using dose geometric parameters (DGPs), dose-volume parameters (DVPs) and Radiomics features.

Decision support:

22 studies included in this review have represented their models as a nomogram for predicting various endpoints. Twenty have used multivariate Cox proportional hazard regression and other two have used multivariate logistic regression as their predictive algorithm. All the prediction models have shown good results (c-index/AUC/R^2> 0.6) in predicting endpoints.

Methodological assessment:

The majority of studies have used feature selection or elimination techniques to reduce the redundancy and dimensionality of the data. In 6 studies the feature selection technique has not been described properly. 12 studies have used radiomic features along with clinicopathological features in different prediction algorithms, and the majority of them have shown the superiority of radiomic features over clinicopathological features alone. The complete synthesis with score has been shown in table 2 and shown in figure 2. As per our scoring criterion, 33 articles were considered as significant and 6 were considered as less significant. Thirteen studies have demonstrated a better association of radiomics features with prediction endpoints when compared with conventional image/clinic-pathological features.

Author	Year	Sample Size	Disease Stage	Histology	Treatment	Feature used	Event Rate reported	Multi-center study	Feature selection technique	Model algorithm	Model Assessment	Validation	Overall score
Lee WK et. al.[36]	2020	3	1	1	1	3	0	0	1	2	3	1	16
Paik ES et. al.[37]	2020	5	1	1	1	3	0	0	2	2	2	4	21
Tian X et. al.[38]	2020	3	1	1	1	3	0	2	2	2	3	1	19
Fang M et. al.[39]	2020	2	1	0	1	3	0	0	2	2	3	1	15
Dong T et. al.[40]	2020	3	1	0	1	3	0	2	2	3	3	4	22
Fang J et. al.[41]	2020	3	1	1	1	3	0	2	2	2	3	1	19
Wang T et. al.[42]	2019	2	1	0	1	3	0	0	1	2	3	1	14
Wang W et. al.[43]	2019	4	1	1	1	2	1	2	1	2	2	1	18
Shen WC et. al.[44]	2019	2	1	1	1	3	0	0	0	2	2	1	13
Zhang S et. al.[45]	2019	5	1	1	1	3	0	0	1	2	2	4	20
Obrzut B et. al.[46]	2019	2	1	1	1	2	0	0	0	3	2	1	13
Yang J et. al.[47]	2019	5	1	1	1	2	0	0	1	2	2	4	19
Lucia F et. al.[48]	2019	2	1	1	1	2	0	2	1	2	3	4	19
Matsuo K et. al.[49]	2019	4	1	1	1	2	1	0	0	3	3	1	17
Sun C et. al.[50]	2019	3	1	0	1	2	0	0	1	2	2	1	13
Wang C et. al.[51]	2018	5	1	1	1	2	0	0	1	2	3	4	20
Marchetti C et. al.[52]	2018	3	1	1	1	2	1	0	2	2	2	1	16
Altazi AB et. al.[53]	2018	1	1	1	1	2	0	0	2	2	3	1	14
Chen J et. al.[54]	2018	1			1	2	1	0	1	2	3	1	12
Zhen X et. al.[55]	2017	1	1	1	1	3	1	0	0	2	2	1	13
Rose P G et. al.[56]	2017	5	1	1	1	2	0	0	2	2	2	2	18
Zheng RR et. al.[57]	2017	4	1	1	1	2	0	0	1	2	3	1	16
ReuzéS et. al.[58]	2017	2	1	1	1	2	0	0	1	2	3	1	14
Lucia F et. al.[59]	2017	2	1	1	1	2	1	0	1	2	2	4	17
Obrzut B et. al.[60]	2017	2	1	1	1	2	0	0	0	3	2	1	13
Shim SH et. al.[61]	2016	3	1	1	1	1	0	0	1	2	3	2	15
Kong TW et. al.[62]	2016	3	1	1	1	2	0	0	2	2	3	1	16
Zhou H et. al.[63]	2015	5	1	1	1	2	1	0	1	2	3	1	18
Kim DY et. al.[64]	2014	3	1	1	1	2	0	0	1	2	3	4	18
Kumar S et. al.[65]	2014	2	1	1	1	3	0	0	0	3	3	1	15
Je HU et. al.[66]	2014	5	1	1	1	2	0	0	1	2	3	4	20
Shim SH et. al.[67]	2013	3	1	1	1	2	0	0	1	2	3	1	15
Lee HJ et. al.[68]	2013	5	1	1	1	2	1	2	1	2	1	1	18
Kang S et. al.[69]	2012	4	1	1	1	2	1	2	1	2	3	4	22
Kidd EA et. al.[70]	2012	3	1	1	1	2	0	0	0	2	3	1	14
PolterauerS et. al.[71]	2012	4	1	1	1	2	1	2	1	2	3	2	20
Seo Y et. al.[72]	2011	4	1	1	1	2	1	0	1	2	2	2	17

Biewenga P et. al.[73]	2011	4	1	1	1	2	1	0	1	2	3	1	17
Tseng JY et. al.[74]	2010	3	1	1	1	2	1	0	1	2	2	1	15

Scoring system used:	Maximum possible score= 27		
Sample size (SS) SS<=100 =1 100<SS<=200 =2 200<SS<=500 =3 500<SS<=1000 =4 SS>1000 =5 **Disease Stage (Mentioned):** No=0 Yes=1 **Histology Stage (Mentioned):** No=0 Yes=1 **Treatment (Mentioned):** No=0 Yes=1	**Feature Used** Only clinical =1 Only pathological =1 Only imaging=1 Only radiomics=1 **Event rate reported** YES=1 No=0 **Multi-center study** Yes=2 (two centre), 3(three and more centre), 4 (three and more centre with different country) No=0 **Feature selection Technique** Not described =0 Only one method used =1 Multiple method used=2	**Prediction Algorithm** One algorithm one model =1 One algorithm based multiple models = 2 Multiple algorithms multiple models=3 **Model Assessment** Only validation accuracy is reported without confidence interval =1 Validation accuracy with confidence interval=2 Train and validation accuracy with confidence interval=3	**Model Validation:** Train-test model validation=1 Bootstrap validation/cross validation=2 External validation=4

Table 2: Overall performance score of all the included manuscripts in this review
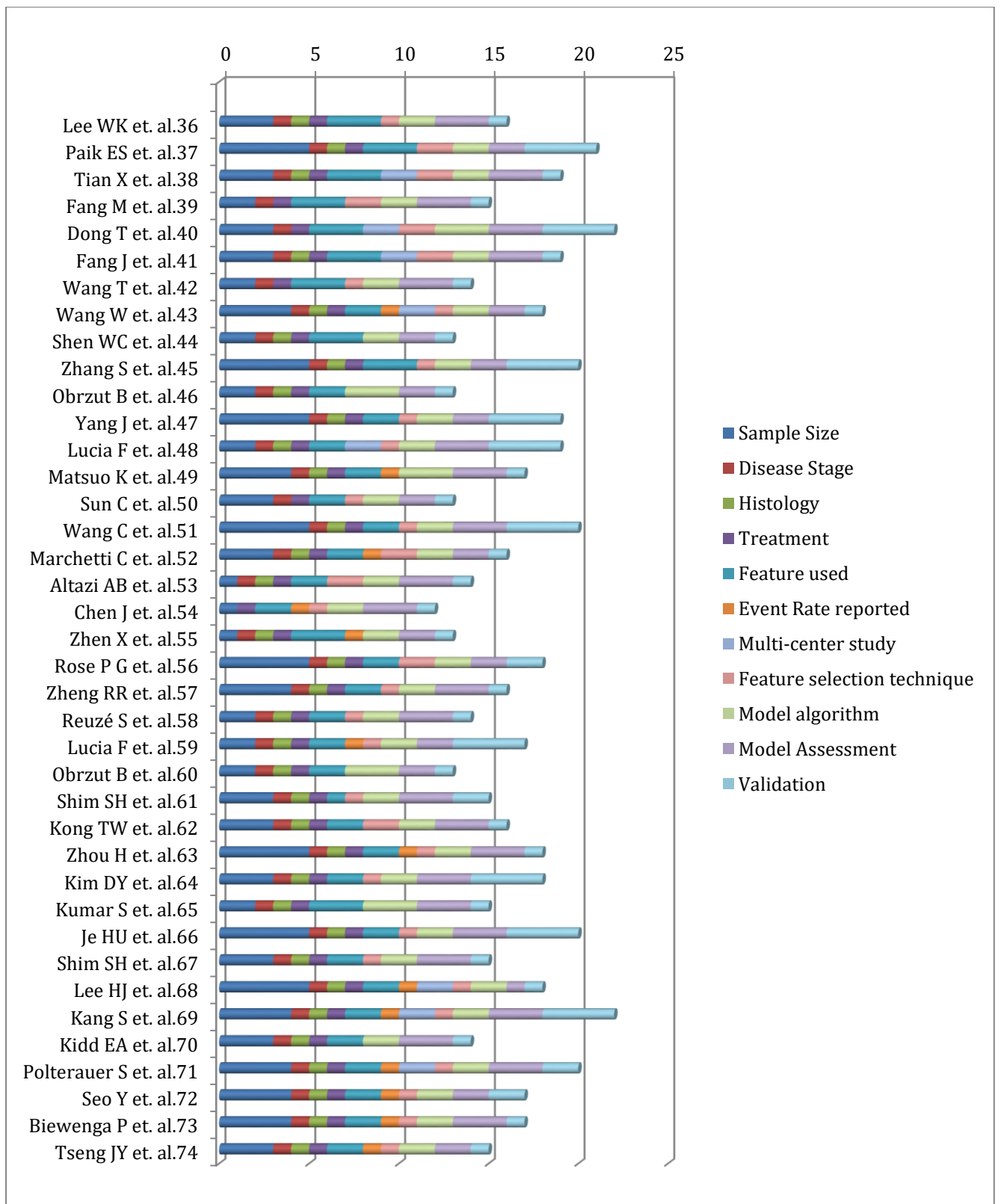
Figure 2: Synthesis score of studies reviewed

Meta Analysis:

The Meta-analysis in all the groups was performed using Random-Effects Model and the result has been shown in table 3. Pooled statistics in all five groups i.e., OS, PFS, Recurrence/DM, Treatment response, and Toxicity/QOL found to be significant in our

review. The heterogeneity was very high in all the groups except Toxicity /QOL. The pooled correlation coefficient was found to be significant in all the groups; the p-value was <0.0001. Forest plots of Meta-analysis for all the groups are shown in figure 3.

	k	tau^2	I^2	Pooled statistics	p-value
Group 1 Overall survival (OS)	25	0.0401 (SE = 0.0123)	97.87%	0.76 [95%CI: 0.72, 0.79]	<0.0001
Group 2 Progression-free survival (PFS)	10	0.0919(SE = 0.00449)	98.12%	0.80 [95%CI: 0.73, 0.86]	<0.0001
Group 3 Recurrence or Distant metastasis	11	0.0519 (SE = 0.0264)	91.38%	0.87 [95%CI: 0.83, 0.90]	<0.0001
Group 4 Treatment response	3	0.0352 (SE = 0.0404)	88.02%	0.85 [95%CI: 0.77, 0.90]	<0.0001
Group 5 Toxicity or Quality of life	3	0.0002 (SE = 0.0155)	1.04%	0.88 [95%CI: 0.85, 0.90]	<0.0001

Table 3: Result of Meta analysis in all the groups i.e., Overall survival, Progression-free survival, Recurrence or Distant metastasis, Treatment response, and Toxicity or Quality of life
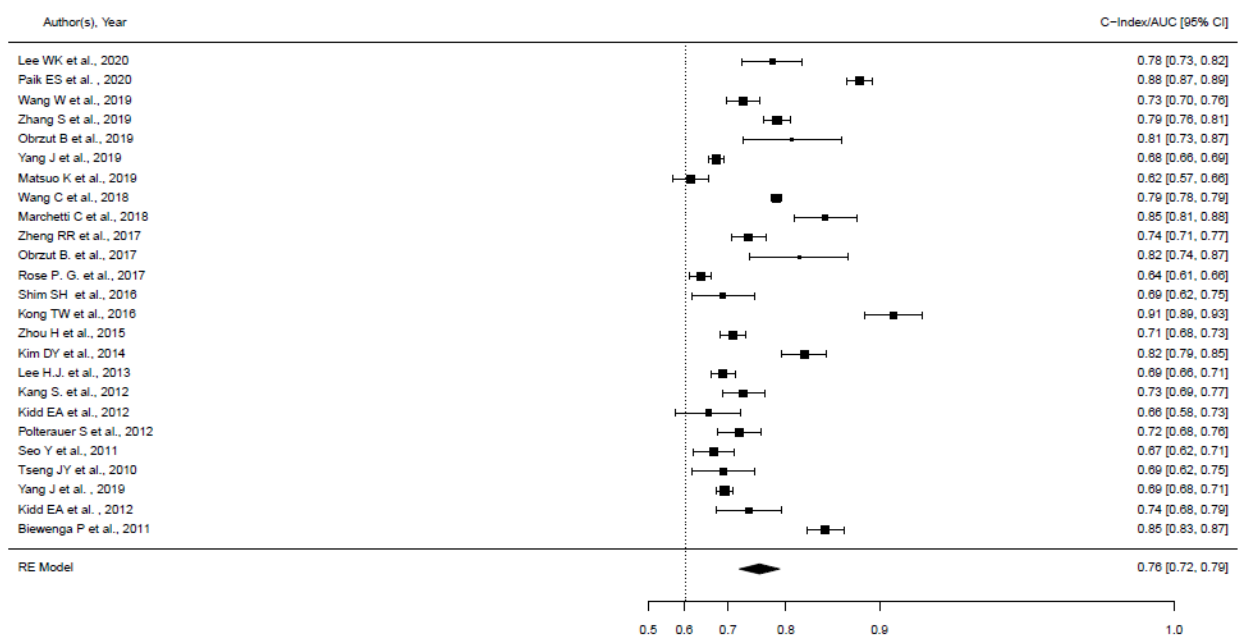

Figure 3a: Forest plots of Meta analysis of model predicting overall survival
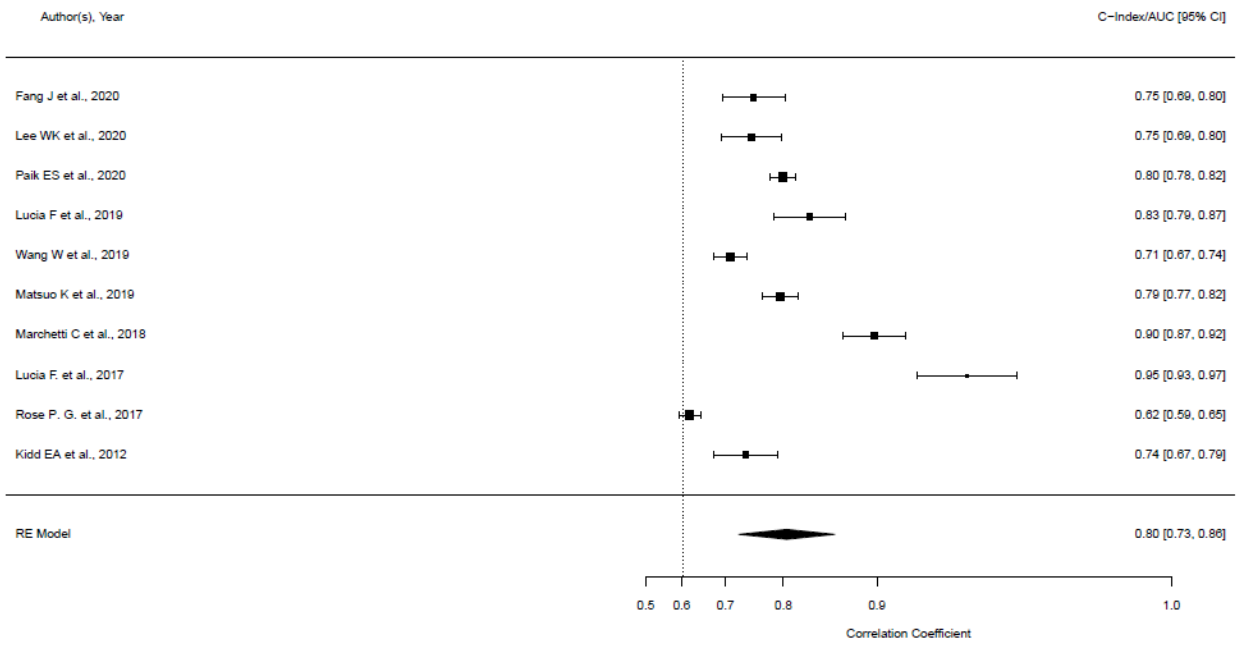
Figure 3b: Forest plots of Meta analysis of model predicting progression free survival
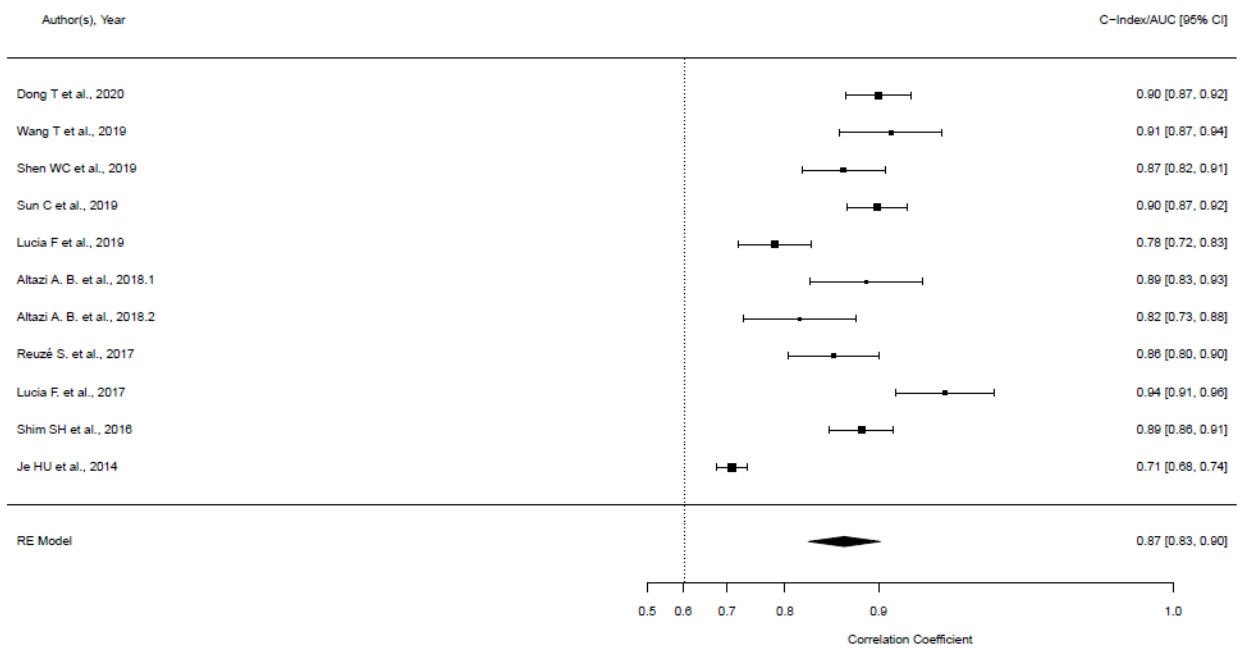

Figure 3c: Forest plots of Meta analysis of model predicting Distance metastasis /recurrence
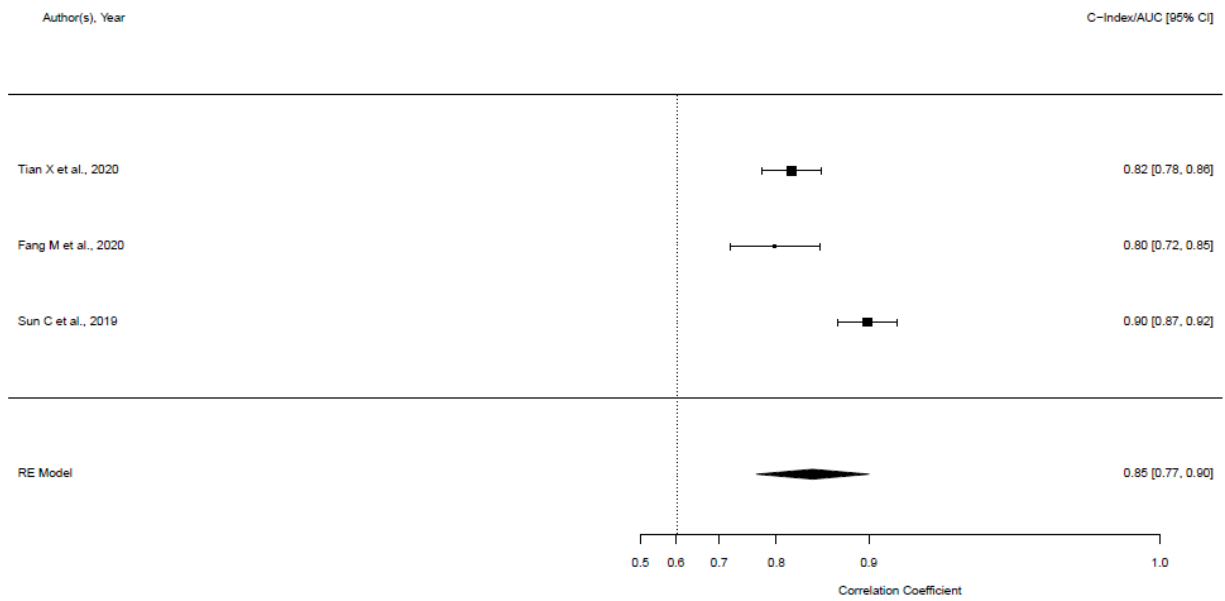
Tian X et al., 2020   0.82 [0.78, 0.86]

Fang M et al., 2020   0.80 [0.72, 0.85]

Sun C et al., 2019   0.90 [0.87, 0.92]

RE Model   0.85 [0.77, 0.90]

0.5   0.6   0.7   0.8   0.9   1.0

Correlation Coefficient

Figure 3d: Forest plots of Meta analysis of model predicting Treatment response

Author(s), Year   C−Index/AUC [95% CI]

Chen J. et al., 2018   0.91 [0.84, 0.95]

Zhen X. et al., 2017   0.89 [0.80, 0.94]

Kumar S et al., 2014   0.87 [0.83, 0.90]

RE Model   0.88 [0.85, 0.90]

0.5   0.6   0.7   0.8   0.9   1.0
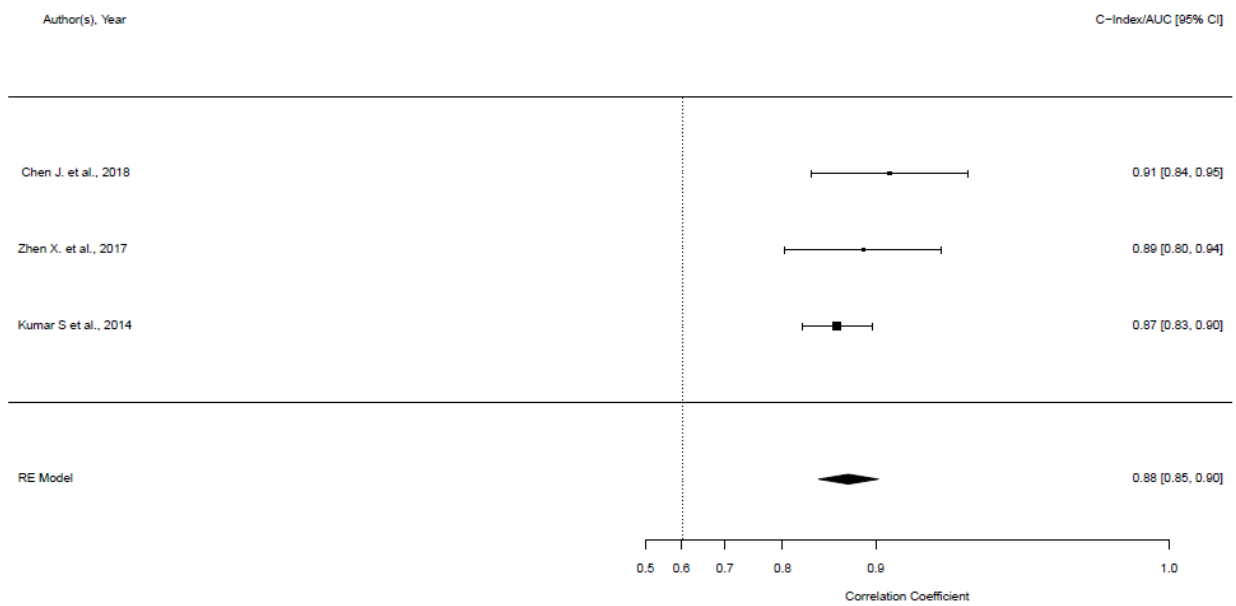
Correlation Coefficient

Figure 3e: Forest plots of Meta analysis of model predicting Toxicity/QOL

Discussion

Artificial intelligence and more specifically machine learning use advanced mathematical and computer algorithms to unearth underlying relations between features and the outcome variable[77-83]. The application of such algorithms in medicine holds promise to predict and improve an individual patient's response based on the available data. Hence, various researchers are working to utilize machine learning for personalized medicine to improve the detection of disease, selection of treatment, and treatment outcome[77,78]. The ultimate goal of artificial intelligence is to provide decision support for personalized treatment. Articles included in this review have used clinicopathological, radiological, and radionics features to develop a prediction model for various endpoints using different prediction algorithms.

Prediction model quality score (PMQS): Our proposed scoring system comprehensively evaluates all the aspects of prediction model development and validation. Our scoring system is based on 11 parameters and a 27-point scale. We have prioritized technical issues related to prediction model development. Out of eleven parameters, 4 parameters are related to the reporting of the findings i.e., disease stage, histology, treatment, and event rate have lesser score either 0 or 1. But more important parameters i.e., Sample size (Score: 1-5) Feature Used (Score: 1-3), multi-center study (Score: 0-4), Feature selection Technique (Score: 0-2), Prediction Algorithm (Score: 1-3), Model Assessment (Score: 1-3), Model Validation (Score: 1-4) have been assigned a wide range of score. According to our scoring system, 33 articles were significant (score≥50%) and 6 were less significant (score<50%). The overall score of PMQS is suggestive of the significance of the prediction model in cervical cancer. Details of our proposed scoring system PMQS are discussed below.

Sample size is a very important factor in the development of a prediction model. A model trained on a very small sample size often fails to demonstrate the underlying variability and diversity of the data which makes it difficult to validate the prediction model and often these models fail in the validation phase especially externally. Hence, we have provided a wider range of scores (1-5) for sample size. But few studies in this review have used advanced algorithms to circumvent the problem related to small sample size. Zhen X. et al.[55]have used CNN VGG-16 model to predict rectal toxicity in the sample size of 42 patients. In this study, they have used a pre-trainedVGG-16 network and applied the trained parameters to a new classification task in a process called transfer learning[75, 76]to overcome the small sample size issue. They have demonstrated the success of those advanced techniques, which may be beneficial for future studies in this domain.

Besides sample size, another very important factor is *event distribution (rate)* in the cohort and has been included in the scoring system. A majority of studies have not described the distribution of events in their study population except a few. He H et.al.

where they have described the highly imbalanced nature of their data. This study employed an adaptive synthetic sampling approach (ADASYN)[76] to generate synthetic minority toxicity data to balance the training data set. Two toxicity studies were included in our review, we have found the results encouraging. However, the reliability of these sophisticated algorithms like transfer learning and ADASYN can only be tested well in external or prospective validation.

The survival of the patient suffering from cervical cancer largely depends upon the *type of tumor, stage of the disease* as well as the *treatment offered* to them. These parameters have also been included in the PMQS. The majority of the included articles in this review have included patients of all histological types and stages. Also, they were able to demonstrate good predictability of several features both in univariate and multivariate analysis. However, none of the studies have tried to compare a model trained and validated across the FIGO stages to evaluate the robustness of the model across the various FIGO stages. For example; if a model is trained and validated in a cohort of FIGO stage I and II patients (early stage), whether that model can predict the outcome of patients with FIGO stage IV (advance stage) and vice versa. Whether the variables which have good predictive value for the patient cohort with FIGO stage I and II (early stage) have similar predictive value for the patient cohort with FIGO stage IV (advance stage) and vice versa. These are important questions to answer in future studies. The generalizability of models across the FIGO stages needs to be tested carefully.

The success of any model heavily depends on the *features used* to generate the prediction hence the selection of relevant clinical, diagnostic or radiomics features will have a huge impact on the outcome of the prediction model. A wide range of scores (1-3) has been assigned in PMQC considering the importance of inclusion of relevant features from various sources (Clinical, Pathological, radiomics, diagnostic). The initial inclusion of features among which best features are selected/extracted should be decided based on the evidence of merit of the feature. None of the papers have properly explained the inclusion of features at the initial stage. For example, Altazi AB et. al.[53] have extracted 72 radiomics features and Zhen X. et. al.[55] extracted 43 radiomics features to select the most relevant features but the reason for extraction of only 72 or 43 features respectively was not explained. We feel that the inclusion criteria for the features should be evidence-based, i.e., based on previous studies and or clinical experience. In the majority of the publications, very few features/variables are used to develop the prediction model and none of the studies included in our review found treatment variables as important features to be considered for model development.

Feature selection technique is the second most challenging task in prediction model development. Various feature selection techniques like univariate and multivariate analysis, etc. have been extensively explained in the literature [53, 57]. Studies included in this review have also applied feature selection methods wisely to eliminate less

important features and overcome issues related to overfitting and multi-co-linearity. Univariate logistic regression has been used by several authors to select the most important features for the development of multivariate models. Hence a wide range of scores (0-2) is assigned to feature selection in PMQC.

Besides selecting the most informative features, one also has to select an appropriate *prediction algorithm* to build a model. The development and validation of a single model based on a single prediction algorithm are often disadvantageous because a particular model may not be suitable for a particular prediction task .[83] Whereas the advantage of developing multiple prediction models for evaluation to select the best model is to increase the probability to correctly identify the best model that performs sufficiently well for a prediction task with the particular predicate.[83]Various algorithms are used for the best model selection for a prediction task. Westphal M et.al. have described the benefit of evaluation of multiple prediction models for the selection of bestthe model for the prediction task. The inclusion of multiple prediction models during the development process has been included in the PMQC and a wider range of scores has been assigned (1-3). Very few studies included in this review have tested multiple prediction algorithms and compared them to select the best model to predict the endpoint.

Model validation is one of the most important aspects of prediction model development. Mainly validation of prediction model is performed on internal data (internal validation) or along with external data (external validation). Various validation techniques like train-test validation, cross-validation, bootstrap validation, and nested cross-validation are used. Considering the importance of validation step wide range of scores (1-4) has been assigned in PMQC.

Multicenter studies are important for the generalizability of the prediction model. It is assumed that the prediction model developed on a single-centre study will have more chance to fail in external validation in comparison of multicenter studies. PMQC includes it as an important parameter and assigns a wide range of scores (0-4).

Meta-Analysis Systematically reviewing the predictive performance of one or more prediction models is crucial to examine a model's predictive ability across different study populations, settings, or locations. Quantitative synthesis of prediction model studies helps to better understand their potential generalizability and can be achieved by applying meta-analysis methods. The included studies in this review are typically differing in design, execution, and thus case-mix, variation between their results are unlikely to occur by chance only. For this reason, the meta-analysis should usually allow for (rather than ignore) the presence of heterogeneity and aim to produce a summary result (with its 95% confidence interval) that quantifies the average performance across studies. This can be achieved by implementing a random rather than a fixed-effects meta-analysis model. In our Meta-analysis, we found that pooled statistics for all

prediction endpoints is significantly high with a narrow 95% CI value which signifies the role of the prediction model in cervical cancer. However, high values of heterogeneity signify a lack of uniformity in the prediction model development and validation. In our study, except for toxicity/QOL heterogeneity is found to be very high because of various factors i.e., sample size, variability in features and prediction algorithms. Overall, the Meta-analysis of the review indicates the utility of these models in the prediction of various endpoints in cervical cancer.

This review has a few limitations like (1) non-inclusion of the paper published before 2010, (2)Meta-analysis was performed pooling to prediction endpoints but not for the type of prediction algorithm (3) all the studies included in this review are retrospective studies. This review has only included published literature hence positive result publication bias cannot be ruled out.

Conclusion:

Undoubtedly prediction models are going to help cancer management in future. Our review shows an increasing trend in the use of prediction models in cervical cancer research. The majority of studies included in this review were found to be significant on PMQS evaluation criteria. The Meta-analysis and overall quality score on PMQS together are also suggestive of the significant role of the prediction model in cervical cancer.

Key issues for the future will be to train and validate a model with a large amount of data, external validation, validation in prospective clinical trials, integration of these models into the electronic health record, and a more careful evaluation of models, particularly with respect to their effects on clinical outcomes.

Abbreviations:

ICO : InstitutCatalàd' Oncologia

IARC: International Agency for Research on Cancer

CONTEXT SUMMARY

Key objective

Cervical cancer is one of the most common cancers in women worldwide. This systematic review and Meta-analysis examined the significance of prediction models for the prediction of various clinical endpoints in cervical cancer. We have reviewed 39 articles on prediction modelling and cervical cancer and, to our knowledge; this is the first study of this kind in cervical cancer to date.

Knowledge Generated

The synthesis score pooled statistics of our analyses demonstrated there is strong evidence of the utility of the prediction model to predict various endpoints in cervical cancer.

Relevance

Based on our PMQS the highest-scoring prediction model may be validated prospectively and used in clinical decision making may improve patient care in cervical cancer management.

References

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021;71(3):209-249.

2. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2013. Ca-a Cancer Journal for Clinicians. 2013;63(1):11-30.

3. Waggoner SE. Cervical cancer. Lancet 2003;361:2217-25.

4. Cuzick J, Bergeron C, von Knebel DM, Gravitt P, Jeronimo J, Lorincz AT. Newtechnologies and procedures for cervical cancer screening. Vaccine.2012;30 Suppl 5:F107–16.

5. Bruni L, Albero G, Serrano B, Mena M, Gómez D, Muñoz J, Bosch FX, de Sanjosé S. ICO/IARC Information Centre on HPV and Cancer (HPV Information Centre). Human Papillomavirus and Related Diseases in India. Summary Report 17 June 2019. Available from: https://hpvcentre.net/statistics/reports/IND.pdf [Date Accessed: 17 March, 2021]

6. Rose PG, Bundy BN, Watkins EB, et al: Concurrent cisplatin-based radiotherapy and chemotherapy for locally advanced cervical cancer. N EnglJ Med 340:1144-1153, 1999

7. Thomas GM. Improved treatment for cervical cancer--concurrent chemotherapy and radiotherapy. N Engl J Med 1999;340:1198-200.

8. Neerja Bhatla , Jonathan S. Berek , Mauricio CuelloFredes , Lynette A. Denny , et.al. Revised FIGO staging for carcinoma of the cervix uteri, Int J GynecolObstet 2019; 145: 129–135

9. Choi J, Kim HJ, Jeong YH, Lee JH, Cho A, Yun M, et al. The role of (18) F-FDG PET/CT in assessing therapy response in cervix cancer after concurrent Chemoradiation therapy. Nucl Med Mol Imaging. 2014;48(2):130–6.

10. Czernin J, Allen-Auerbach M, Nathanson D, Herrmann K. PET/CT in Oncology: Current Status and Perspectives. CurrRadiol Rep. 2013;1:177–190.

11. Marcos Duarte Guimaraes, Alice Schuch, Bruno Hochhegger, Jefferson Luiz Gross, Rubens Chojniak, Edson Marchiori,Functional magnetic resonance imaging in oncology: state of the art, Radiol Bras. 2014 Mar-Apr; 47(2): 101–111.

12. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer.2012;48(4):441–6.VanCalster B, Steyerberg EW, Harrell FH. Risk prediction for individuals. JAMA 2015;314:1875.

13. American Cancer Society. Cancer Facts & Figures 2018. Atlanta: American Cancer Society; 2018.]. Available from:https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2018/cancer-facts-and-figures-2018.pdf   [Date Accessed: 17 March, 2021]

14. Machine Learning in MATLAB available from: https://in.mathworks.com/help/stats/machine-learning-in-matlab.html,[Date Accessed: 17 March, 2020]

15. Obermeyer Z, Emanuel EJ: Predicting the future—big data, machine learning, and clinical medicine. N Engl J Med 375:1216–12192016

16. W. Raghupathi, V. Raghupathi, Big data analytics in healthcare: promise and potential, Health Inf Sci Syst, 2 (1) (2014), p. 3, 10.1186/2047-2501-2-3

17. Chartrand, G. et al. Deep Learning : A Primer for Radiologists. RadioGraphics 37, 2113–2131 (2017).

18. Erickson, B. J., Korfiatis, P., Akkus, P. & Kline, T. L. Machine Learning for Medical Imaging. RadioGraphics (2017)37, 505–515

19. Yamashita R., Nishio M., Do R. K. G., Togashi K. Convolutional neural networks: an overview and application in radiology. Insights Imaging(2018). 10.1007/s13244-018-0639-9

20. Daniela XHEMALI, Christopher J. HINDE,Roger G. STONE, Naïve Bayes vs. Decision Trees vs. Neural Networks in the Classification of Training Web Pages, IJCSI International Journal of Computer Science Issues, Vol. 4, No. 1, 2009 1694-0814

21. Sperandei S., Understanding logistic regression analysis, Biochem Med (Zagreb). 2014 Feb; 24(1): 12–18.

22. Cruz JA, Wishart DS. Applications of machine learning in cancer prediction and prognosis. Cancer Informat2006;2:59.

23. Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. BMJ 2009;338:b606

24. Curioni-Fontecedro A., A new era of oncology through artificial intelligence ESMO Open 2017;2:e000198.

25. Simon G, DiNardo CD, Takahashi K, Cascone T, Powers C, Stevens R, Allen J, et. al., Applying Artificial Intelligence to Address the Knowledge Gaps in Cancer Care. Oncologist. 2018 Nov 16. pii: theoncologist.2018-0257.

26. Hingorani AD, Windt DA, Riley RD, Abrams K, Moons KG, Steyerberg EW, et al. Prognosis research strategy (PROGRESS) 4: stratified medicine research. BMJ 2013;346:e5793

27. Steyerberg EW. Clinical prediction models: a practical approach to development, validation, and updating [Internet]. New York (NY): Springer Science & Business Media; 2009 [cited 2019 Jan 31]. Available from: https://books.google.es/books?hl=es&lr=&id=kHGK58cLsMIC&oi=fnd&pg=PR2 &dq=Clinical+Prediction+Models:+A+Practical+Approach+to+Development,+V alidation,+and+Updating&ots=TMXdIoeLhi&sig=G1pavyIKowWc141D3kZlQEER lNw. [Date Accessed: 17 March, 2021]

28. Sakr S., Elshawi R., Ahmed A.M., Qureshi W.T., Brawner C.A., Keteyian S.J., Blaha M.J., Al-Mallah M.H. Comparison of machine learning techniques to predict all-cause mortality using fitness data: The Henry ford exercise testing (FIT) project. BMC Med. Inform. Decis. Mak. 2017;17:174.

29. Haynes RB, McKibbon KA, Wilczynski NL, et al. Optimal search strategies for retrieving scientifically strong studies of treatment from Medline: Analytical survey. BMJ 2005;330:1179.

30. Ingui BJ, Rogers MA. Searching for clinical prediction rules in MEDLINE. J Am Med Inform Assoc 2001;8:391–397.

31. Geersing GJ, Bouwmeester W, Zuithoff P, et al. Search filters for finding prognostic and diagnostic prediction studies in Medline to enhance systematic reviews. PloS One 2012;7:e32844.

32. Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097.

33. Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med. 2015;162(1):W1-W73. doi:10.7326/M14-0698

34. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol. 2017;14(12):749-762. doi:10.1038/nrclinonc.2017.141

35. Viechtbauer W (2010). "Conducting meta-analyses in R with the metafor package." Journal of Statistical Software, 36(3), 1–48. https://doi.org/10.18637/jss.v036.i03.

36. Lee WK, Chong GO, Jeong SY, Lee HJ, Park SH, Ryu JM, Choi YS, Kang S, Koo YJ, Lee DH, Kong E, Lee SW. Prognosis-Predicting Model Based on 18F fluorodeoxyglucose PET Metabolic Parameters in Locally Advanced Cervical Cancer Patients Treated with Concurrent Chemoradiotherapy: Multi-Center Retrospective Study. J Clin Med. 2020 Feb 5;9(2):427.

37. Paik ES, Lim MC, Kim MH, Kim YH, Song ES, Seong SJ, Suh DH, Lee JM, Lee C, Choi CH. Prognostic Model for Survival and Recurrence in Patients with Early-Stage Cervical Cancer: A Korean Gynecologic Oncology Group Study (KGOG 1028). Cancer Res Treat. 2020 Jan;52(1):320-333

38. Tian X, Sun C, Liu Z, Li W, Duan H, Wang L, Fan H, Li M, Li P, Wang L, Liu P, Tian J, Chen C. Prediction of Response to Preoperative Neoadjuvant Chemotherapy in Locally Advanced Cervical Cancer Using Multicenter CT-Based Radiomic Analysis. Front Oncol. 2020 Feb 4;10:77

39. Fang M, Kan Y, Dong D, Yu T, Zhao N, Jiang W, Zhong L, Hu C, Luo Y, Tian J. Multi-Habitat Based Radiomics for the Prediction of Treatment Response to Concurrent Chemotherapy and Radiation Therapy in Locally Advanced Cervical Cancer. Front Oncol. 2020 May 5;10:563

40. Dong T, Yang C, Cui B, Zhang T, Sun X, Song K, Wang L, Kong B, Yang X. Development and Validation of a Deep Learning Radiomics Model Predicting Lymph Node Status in Operable Cervical Cancer. Front Oncol. 2020 Apr 15;10:464.

41. Fang J, Zhang B, Wang S, Jin Y, Wang F, Ding Y, Chen Q, Chen L, Li Y, Li M, Chen Z, Liu L, Liu Z, Tian J, Zhang S. Association of MRI-derived radiomic biomarker with disease-free survival in patients with early-stage cervical cancer. Theranostics. 2020 Jan 16;10(5):2284-2292

42. Wang T, Gao T, Yang J, Yan X, Wang Y, Zhou X, Tian J, Huang L, Zhang M. Preoperative prediction of pelvic lymph nodes metastasis in early-stage cervical cancer using radiomics nomogram developed based on T2-weighted MRI and diffusion-weighted imaging. Eur J Radiol. 2019 May;114:128-135

43. Wang W, Liu X, Meng Q, Zhang F, Hu K. Nomograms predicting survival and patterns of failure in patients with cervical cancer treated with concurrent chemoradiotherapy: A special focus on lymph nodes metastases. PLoS One. 2019 Apr 15;14(4):e0214498

44. Shen WC, Chen SW, Wu KC, Hsieh TC, Liang JA, Hung YC, Yeh LS, Chang WC, Lin WC, Yen KY, Kao CH. Prediction of local relapse and distant metastasis in patients with definitive chemoradiotherapy-treated cervical cancer by deep

learning from 18F-fluorodeoxyglucose positron emission tomography/computed tomography. Eur Radiol. 2019 Dec;29(12):6741-6749

45. Zhang S, Wang X, Li Z, Wang W, Wang L. Score for the Overall Survival Probability of Patients With First-Diagnosed Distantly Metastatic Cervical Cancer: A Novel Nomogram-Based Risk Assessment System. Front Oncol. 2019 Nov 5;9:1106

46. Obrzut B, Kusy M, Semczuk A, Obrzut M, Kluska J. Prediction of 10-year Overall Survival in Patients with Operable Cervical Cancer using a Probabilistic Neural Network. J Cancer. 2019 Jul 10;10(18):4189-4195

47. Yang J, Tian G, Pan Z, Zhao F, Feng X, Liu Q, Lyu J. Nomograms for predicting the survival rate for cervical cancer patients who undergo radiation therapy: a SEER analysis. Future Oncol. 2019 Sep;15(26):3033-3045

48. Lucia F, Visvikis D, Vallières M, Desseroit MC, Miranda O, Robin P, Bonaffini PA, Alfieri J, Masson I, Mervoyer A, Reinhold C, Pradier O, Hatt M, Schick U. External validation of a combined PET and MRI radiomics model for prediction of recurrence in cervical cancer patients treated with chemoradiotherapy. Eur J Nucl Med Mol Imaging. 2019 Apr;46(4):864-877.

49. Matsuo K, Purushotham S, Jiang B, Mandelbaum RS, Takiuchi T, Liu Y, Roman LD. Survival outcome prediction in cervical cancer: Cox models vs deep-learning model. Am J Obstet Gynecol. 2019 Apr;220(4):381.e1-381.e14.

50. Sun C, Tian X, Liu Z, Li W, Li P, Chen J, Zhang W, Fang Z, Du P, Duan H, Liu P, Wang L, Chen C, Tian J. Radiomic analysis for pretreatment prediction of response to neoadjuvant chemotherapy in locally advanced cervical cancer: A multicentre study. EBioMedicine. 2019 Aug;46:160-169.

51. Wang C, Yang C, Wang W, Xia B, Li K, Sun F, Hou Y. A Prognostic Nomogram for Cervical Cancer after Surgery from SEER Database. J Cancer. 2018 Oct 10;9(21):3923-3928.

52. Marchetti C, De Felice F, Di Pinto A, Romito A, Musella A, Palaia I, Monti M, Tombolin V, Muzii L, Benedetti Panici P. Survival Nomograms after Curative Neoadjuvant Chemotherapy and Radical Surgery for Stage IB2-IIIB Cervical Cancer. Cancer Res Treat. 2018 Jul;50(3):768-776.

53. Altazi BA, Fernandez DC, Zhang GG, Hawkins S, Naqvi SM, Kim Y, Hunt D, Latifi K, Biagioli M, Venkat P, Moros EG. Investigating multi-radiomic models for enhancing prediction power of cervical cancer treatment outcomes. Phys Med. 2018 Feb;46:180-188.

54. Chen J, Chen H, Zhong Z, Wang Z, Hrycushko B, Zhou L, Jiang S, Albuquerque K, Gu X, Zhen X. Investigating rectal toxicity associated dosimetric features with deformable accumulated rectal surface dose maps for cervical cancer radiotherapy. Radiat Oncol. 2018 Jul 6;13(1):125

55. Zhen X, Chen J, Zhong Z, Hrycushko B, Zhou L, Jiang S, Albuquerque K, Gu X. Deep convolutional neural network with transfer learning for rectum toxicity prediction in cervical cancer radiotherapy: a feasibility study. Phys Med Biol. 2017 Oct 12;62(21):8246-8263

56. Rose PG, Java J, Whitney CW, Stehman FB, Lanciano R, Thomas GM, DiSilvestro PA. Nomograms Predicting Progression-Free Survival, Overall Survival, and Pelvic Recurrence in Locally Advanced Cervical Cancer Developed From an Analysis of Identifiable Prognostic Factors in Patients From NRG Oncology/Gynecologic Oncology Group Randomized Trials of Chemoradiotherapy. J Clin Oncol. 2015 Jul 1;33(19):2136-42

57. Zheng RR, Huang XW, Liu WY, Lin RR, Zheng FY, Lin F. Nomogram Predicting Overall Survival in Operable Cervical Cancer Patients. Int J Gynecol Cancer. 2017 Jun;27(5):987-993.

58. Reuzé S, Orlhac F, Chargari C, Nioche C, Limkin E, Riet F, Escande A, Haie-Meder C, Dercle L, Gouy S, Buvat I, Deutsch E, Robert C. Prediction of cervical cancer recurrence using textural features extracted from 18F-FDG PET images acquired with different scanners. Oncotarget. 2017 Jun 27;8(26):43169-43179.

59. Lucia F, Visvikis D, Desseroit MC, Miranda O, Malhaire JP, Robin P, Pradier O, Hatt M, Schick U. Prediction of outcome using pretreatment 18F-FDG PET/CT and MRI radiomics in locally advanced cervical cancer treated with chemoradiotherapy. Eur J Nucl Med Mol Imaging. 2018 May;45(5):768-786

60. Obrzut B, Kusy M, Semczuk A, Obrzut M, Kluska J. Prediction of 5-year overall survival in cervical cancer patients treated with radical hysterectomy using computational intelligence methods. BMC Cancer. 2017 Dec 12;17(1):840

61. Shim SH, Kim DY, Lee SJ, Kim SN, Kang SB, Lee SW, Park JY, Suh DS, Kim JH, Kim YM, Kim YT, Nam JH. Prediction model for para-aortic lymph node metastasis in patients with locally advanced cervical cancer. Gynecol Oncol. 2017 Jan;144(1):40-45.

62. Kong TW, Kim J, Son JH, Kang SW, Paek J, Chun M, Chang SJ, Ryu HS. Preoperative nomogram for prediction of microscopic parametrial infiltration in patients with FIGO stage IB cervical cancer treated with radical hysterectomy. Gynecol Oncol. 2016 Jul;142(1):109-114

63. Zhou H, Li X, Zhang Y, Jia Y, Hu T, Yang R, Huang KC, Chen ZL, Wang SS, Tang FX, Zhou J, Chen YL, Wu L, Han XB, Lin ZQ, Lu XM, Xing H, Qu PP, Cai HB, Song XJ, Tian XY, Zhang QH, Shen J, Liu D, Wang ZH, Xu HB, Wang CY, Xi L, Deng DR, Wang H, Lv WG, Shen K, Wang SX, Xie X, Cheng XD, Ma D, Li S. Establishing a Nomogram for Stage IA-IIB Cervical Cancer Patients after Complete Resection. Asian Pac J Cancer Prev. 2015;16(9):3773-7

64. Kim DY, Shim SH, Kim SO, Lee SW, Park JY, Suh DS, Kim JH, Kim YM, Kim YT, Nam JH. Preoperative nomogram for the identification of lymph node metastasis in early cervical cancer. Br J Cancer. 2014 Jan 7;110(1):34-41

65. Kumar S, Rana ML, Verma K, Singh N, Sharma AK, Maria AK, Dhaliwal GS, Khaira HK, Saini S. PrediQt-Cx: post treatment health related quality of life prediction model for cervical cancer patients. PLoS One. 2014 Feb 26;9(2):e89851

66. Je HU, Han S, Kim YS, Nam JH, Kim HJ, Kim JW, Park W, Bae DS, Kim JH, Shin SJ, Kim J, Lee KH, Yoon MS, Kim SM, Kim JY, Yoon WS, Lee NW, Choi JH, Park SY, Kim JY. A nomogram predicting the risks of distant metastasis following postoperative radiotherapy for uterine cervical carcinoma: a Korean radiation oncology group study (KROG 12-08). Radiother Oncol. 2014 Jun;111(3):437-41

67. Shim SH, Lee SW, Park JY, Kim YS, Kim DY, Kim JH, Kim YM, Kim YT, Nam JH. Risk assessment model for overall survival in patients with locally advanced cervical cancer treated with definitive concurrent chemoradiotherapy. Gynecol Oncol. 2013 Jan;128(1):54-59

68. Lee HJ, Han S, Kim YS, et al.., Individualized prediction of overall survival after postoperative radiation therapy in patients with early-stage cervical cancer: a Korean Radiation Oncology Group study (KROG 13-03)., Int J Radiat Oncol Biol Phys. 2013 Nov 15;87(4):659-64.

69. Kang S, Nam BH, Park JY, Seo SS, Ryu SY, Kim JW, Kim SC, Park SY, Nam JH. Risk assessment tool for distant recurrence after platinum-based concurrent chemoradiation in patients with locally advanced cervical cancer: a Korean gynecologic oncology group study. J Clin Oncol. 2012 Jul 1;30(19):2369-74

70. Kidd EA, El Naqa I, Siegel BA, Dehdashti F, Grigsby PW. FDG-PET-based prognostic nomograms for locally advanced cervical cancer. Gynecol Oncol. 2012 Oct;127(1):136-40

71. Polterauer S, Grimm C, Hofstetter G, Concin N, Natter C, Sturdza A, Pötter R, Marth C, Reinthaller A, Heinze G. Nomogram prediction for overall survival of patients diagnosed with cervical cancer. Br J Cancer. 2012 Sep 4;107(6):918-24

72. Seo Y, Yoo SY, Kim MS, Yang KM, Yoo HJ, Kim JH, Shin YJ, Kang JK, Lee KH, Lee ED, Rhu SY, Choi SC, Kim MH, Kim BJ, Kim MS, Cho CK. Nomogram prediction

of overall survival after curative irradiation for uterine cervical cancer. Int J Radiat Oncol Biol Phys. 2011 Mar 1;79(3):782-7

73. Biewenga P, van der Velden J, Mol BW, Stalpers LJ, Schilthuis MS, van der Steeg JW, Burger MP, Buist MR. Prognostic model for survival in patients with early stage cervical cancer. Cancer. 2011 Feb 15;117(4):768-76

74. Tseng JY, Yen MS, Twu NF, Lai CR, Horng HC, Tseng CC, Chao KC, Juang CM. Prognostic nomogram for overall survival in stage IIB-IVA cervical cancer patients treated with concurrent chemoradiotherapy. Am J Obstet Gynecol. 2010 Feb;202(2):174.e1-7

75. Tajbakhsh N, Shin J Y, Gurudu S R, Hurst R T, Kendall C B, Gotway M B and Liang JConvolutional neural networks for medical image analysis: full training or fine tuning?, 2016, IEEE Trans Med Imaging 35 1299–312

76. He H, Bai Y, Garcia E A and Li S 2008 ADASYN: adaptive synthetic sampling approach for imbalanced learning 2008 IEEE Int. Joint Conf. on Neural Networks (IEEE World Congress on Computational Intelligence) pp 1322–8

77. Kononenko I. Machine learning for medical diagnosis: history, state of the art and perspective. ArtifIntell Med 2001;23:89–109.

78. Niknejad A, Petrovic D. Introduction to computational intelligence techniques and areas of their applications in medicine. Med Appl ArtifIntell 2013;51.

79. D. Montana, "A weighted probabilistic neural network," in NIPS, 1991, pp. 1110–1117.

80. V.P. Balachandran, M. Gonen, J.J. Smith, R.P. Dematteo,Nomograms in oncology: More than meets the eye Lancet Oncol, 16 (4) (2015), pp. e173-e180

81. Noble WS, What is a support vector machine?, Nat biotechnol,2006, 24(12): 1565-1557.

82. Huang S, Cai N, Pacheco PP, Narrandes S, Wang Y, Xu W. Applications of Support Vector Machine (SVM) Learning in Cancer Genomics.Cancer Genomics Proteomics. 2018 Jan-Feb;15(1):41-51.

83. Westphal M, Brannath W. Evaluation of multiple prediction models: A novel view on model selection and performance assessment. Stat Methods Med Res. 2020;29(6):1728-1745. doi:10.1177/0962280219854487

Chapter 4: Radiomics: A Quantitative Imaging Biomarker in Precision Oncology

Abstract

Cancer treatment is heading towards precision medicine driven by genetic and biochemical markers. Various genetic and biochemical markers are utilized to render personalized treatment in cancer. In the last decade, non-invasive imaging biomarkers have also been developed to assist personalized decision support systems in oncology. High throughput quantitative imaging biomarkers i.e., radiomics is being researched to develop a specific digital phenotype of tumor in cancer. Radiomics is a process to extract high throughput data from medical images by using advanced mathematical and statistical algorithms. The Radiomics process involves various steps i.e. image generation, segmentation of the region of interest (e.g. a tumor), image preprocessing, radiomic feature extraction, feature analysis and selection and prediction model development. The radiomics process explores the heterogeneity, irregularity and size parameters of the tumor to calculate thousands of advanced features. Our study investigates the role of radiomics in prediction medicine. Radiomics research has witnessed rapid growth in the last decade with several studies published that show the potential of radiomics in the diagnosis and treatment of cancer. Several radiomics based prediction models have been developed and reported in the literature to predict various prediction endpoints i.e., overall survival, progression-free survival and recurrence in brain tumors, head and neck cancer, lung cancer and several other cancer types. Radiomics based digital phenotype markers have shown promising results in diagnosis and treatment in oncology. In the coming years, radiomics is going to play a significant role in precision oncology.

Introduction

Cancer is caused by genetic mutations leading to uncontrolled growth of tissue and cells of growing tissue can leave the tissue colony and metastasize in other parts of the body [1]. Conventionally, cancer is treated by surgery, chemotherapy, radiotherapy or combinations of these options [2]. Often the selection of treatment options depends upon the type of tumor, stage of the disease and general condition of the patient [2]. Although clinicians consider these factors to decide the course of treatment, on several occasions these treatments fail [3]. This led to the evolution of personalized medicine in oncology [4]. Personalized oncology works on the principle of identification of subgroups of patients in particular disease types [4-5]. Many biomarkers and gene mutations have been investigated to identify the subgroups of the patients in various cancers and targeted drugs for those subgroups [5-6]. For example, by sequencing and in situ hybridization (ISH) techniques a patient subgroup with epidermal growth factor receptor (EGFR) mutation can be identified in Non-Small Cell Lung Cancer (NSCLC) patients. These high-risk patients do not respond well to conventional treatment options but show good response with targeted therapies like Erlotinib, Gefitinib, Afatinib and similar drugs [6-9]. Precision oncology has the potential to personalize the screening, risk stratifications, treatment selection and response assessment [4-5]. Although most approaches towards precision oncology are centred on biomarkers and genetic mutation assessments [7], artificial intelligence (AI) driven technologies are also being explored to improve the accuracy of precision oncology [10, 12]. This technology-driven approach has also been tested in various fields in precision oncology i.e. screening, risk stratifications, treatment selection and response assessment [10, 12]. AI-based precision oncology has achieved success as witnessed in published literature in the last few years. Various imaging biomarkers are being developed and tested for their utility in precision oncology [13-15]. Those imaging biomarkers are of two types' i.e. qualitative (ex. spiculated margin of tumor, vascularity of tumor, position of tumor and contrast enhancement of the tumor etc.) and quantitative (Hounsfield unit (HU) in computed tomography (CT), standardized uptake value (SUV) and total lesion glycolysis (TLG) in positron emission tomography (PET)) [16]. In the last few years another kind of imaging biomarker, i.e., radiomic features are extracted from the medical images and being tested in precision oncology [16, 17]. The aim of this study is to review the radiomic process and its role in precision oncology and the secondary aim of this study was to investigate the growth of radiomics research in the last two decades.

Radiomics:

Radiomics as a word was first used by Lambin et al in 2012 to describe the quantification of medical imaging data [17]. Radiomics is a process to extract high throughput data from medical images like CT, PET, MRI or SPECT by using advanced mathematical and statistical analysis of images [16, 17]. The Radiomics process explores the heterogeneity,

irregularity and size parameters of the tumor to calculate thousands of advanced features [16-18]. There are mainly two types of radiomics i.e. handcrafted radiomics and deep learning-based radiomics. Here in this manuscript mainly we will discuss the first form of radiomics i.e. hand crafted radiomics and we will address these by the term radiomics itself [19].

Radiomics Process:

The Radiomics process involves various steps i.e. image generation, segmentation of the region of interest (e.g. a tumor), image preprocessing, radiomic feature extraction, feature analysis and selection and prediction model development [16-18]. The stepwise radiomic process is shown in Figure 1.
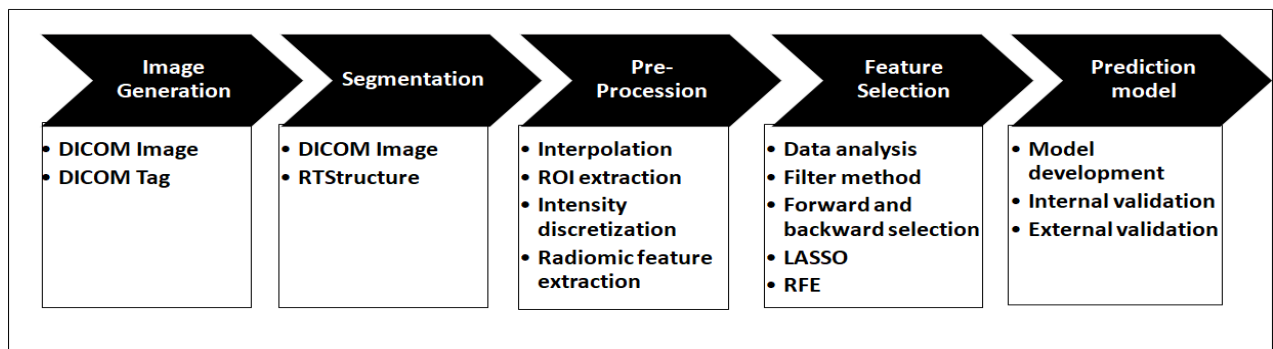

Figure 1: Radiomic process for radiomic feature extraction and feature selection

Image generation: Medical equipment like computed tomography (CT), Positron emission tomography (PET), magnetic resonance imaging (MRI) and single-photon emission tomography (SPECT) are used to image the patient and three-dimensional images are generated by sophisticated reconstruction techniques. These images are archived in the image repository i.e., picture archiving communication system (PACS) for future utilization.

Segmentation: The images are transferred to the workstations and the region of interest (ROI) is delineated surrounding the tumor, to extract radiomic features from that part of the image. The ROI is generated by medical experts or physicists and typically stored as DICOM RTstructure or Segmentation.

Preprocessing of image: Image preprocessing involves various steps performed on images and the ROI. As an example, the following steps are typically performed before radiomic extraction from the medical images [20].

Interpolation: Medical images are reconstructed and represented in three-dimensional matrices with one unit of the matrix called a voxel. Often voxels are not

isotropic and to extract textural radiomic features, the voxels are often re-sampled or interpolated into isotropic voxels.

Re-segmentation of the original ROI defined by an expert or by automated segmentation is utilized to generate a morphological mask and intensity mask. The morphological mask is the original mask. The intensity mask is re-segmented, which contains selected voxel inside or outside the morphological mask.

ROI extraction: Many features do not require voxels outside the ROI; hence the image volume is extracted for the image based on the ROI of the intensity mask.

Intensity discretization: Medical images contain noise and often quantization of image intensities is performed to suppress the noise inside the ROI to calculate the texture features. Two approaches are used for intensity discretization i.e. (1) fixed number of bins, and (2) fixed bin width.

Radiomic feature extraction: Automatic extraction of radiomic features is performed in this step. Up to multiple thousands of radiomic features are generated in this step which is further processed in the radiomic analysis step.

Radiomic analysis and feature selection: While sometimes 1000+ features are extracted from medical images; these are not all useful for phenotyping a particular disease or for the development of an outcome prediction. Many features are redundant and many have no association with the particular disease or outcome. Various statistical tests can be performed for feature reduction [21]. Hierarchical clustering, Spearman correlation, Pearson correlation paired t-test are performed to eliminate the redundancy of the feature; forward, backward feature selection, Least Absolute Shrinkage and Selection Operator (LASSO) or recursive feature elimination (RFE) techniques are used to reduce the dimensionality of the features. Finally, the most appropriate features are selected for disease prognostication or prediction model development for various endpoints like overall survival, recurrence, treatment selection or prediction of treatment outcomes.

Prediction model development: Finally, the prediction model is developed and validated by using the selected features. These features may also be combined with clinical features to develop prediction models. Various machine algorithms have been used to develop a prediction model depending upon the need i.e., regression algorithms, Linear and Logistic regression, K-Nearest Neighbor (KNN), decision trees algorithms, i.e., Random Forest (RF), Support Vector Machine (SVM), Bayesian Network (BN), and deep learning algorithms, i.e., Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Artificial Neural Networks (ANN). [22-24]

Radiomic features can be categorized into various groups [18]. Feature groups and the typical number of features extracted using Pyradiomics software [25] are shown in table 1.

Type of Feature	Feature descriptions	No. of Features
Shape-based Features	Shape features are the descriptors of the three-dimensional size and shape of the ROI and are independent from the grey level intensity distribution. These features are only calculated on the original image and mask.	13
First Order statistics	First-order statistics describe the distribution of voxel intensities within the ROI region of the image.	17
GLRLM	Gray Level Run Length Matrix (GLRLM) assesses the distribution of discretized grey levels in an image or in a stack of images assesses run lengths.	16
GLCM	Gray Level Co-Occurrence Matrix(GLCM) expresses how combinations of discretized intensities of neighbouring voxels in a 3D volume, are distributed along with one of the image directions.	22
GLSZM	Gray Level Size Zone Matrix (GLSZM) counts the number of groups/zones of linked voxels with identical discretized grey levels.	16
NGTDM	Neighboring Gray Tone Difference Matrix (NGTDM) contains the sum of grey level differences of voxels with discretized grey level and the average discretized grey level of neighbouring voxels within a Chebyshev distance δ.	5
GLDM	Gray Level Dependence Matrix (GLCM) quantifies grey level dependencies in an image in terms of the number of connected voxels within distance δ that are dependent on the central voxel.	14
LoG Features	A Laplacian of Gaussian (LoG) filter is applied to the original image and one set of derived images is generated for each sigma value specified. Usually, 1-5 sigma values are used, we use 3 sigma values 1, 2, 3 and three sets of derived images are produced. Subsequently, radiomic features are extracted from these image sets.	270
Wavelet Features	Wavelet transformation of image is performed using the three-dimensional wavelet decomposition and 8 sets of	720

| | images are generated from the original image set. Radiomic features are extracted for transformed image sets. | |

Table 1: Radiomic features can be extracted by using PyRadiomics software

Deep learning Radiomics workflow: Recently, an alternative to handcrafted radiomic workflow, a deep learning-based radiomics workflow [26-28] has emerged. A deep learning-based radiomics workflow extract features from medical images without predefined formulas. Images may be used with or without an ROI for this deep radiomic workflow. Usually, it is a two or three-step process. Step (1) Image data acquisition (2) Segmentation (may or may not be given) (3) development and validation of deep neural networks model. It is not possible in deep learning radiomics to describe features mathematically.

Radiomics and precision oncology: Radiomics has witnessed rapid growth in the last decade with several studies published that show the potential of radiomics in the diagnosis and treatment of cancer. Many radiomics based AI decision support systems have been developed in oncology and reported in the literature. Figure 2 shows the process of precision oncology leveraging radiomic and artificial intelligence.
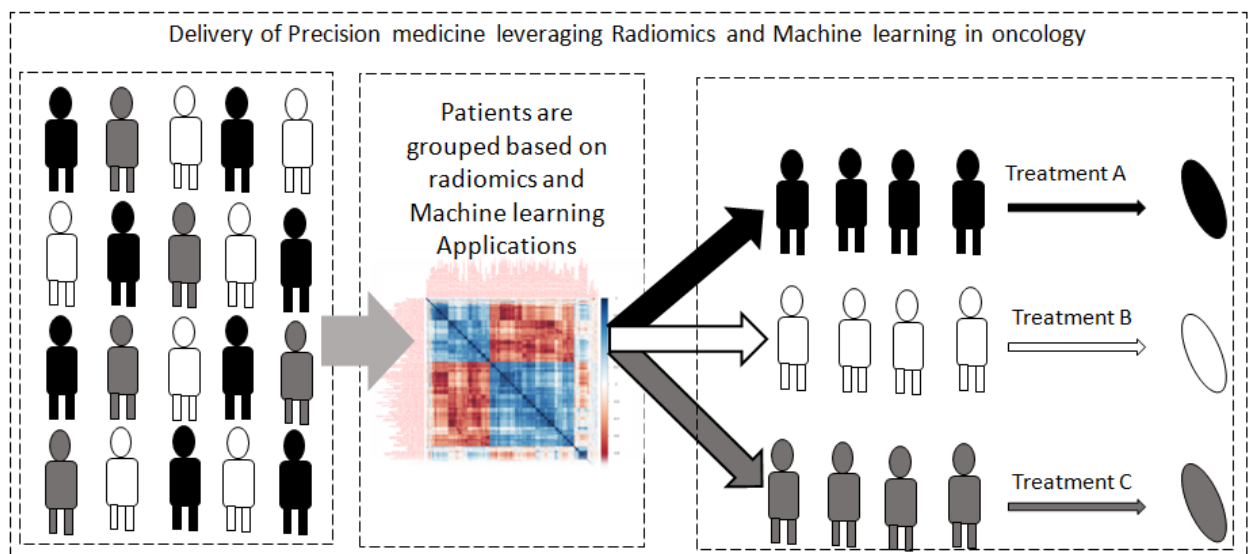

Figure 2: Mechanism to deliver personalized medicine leveraging the machine learning and artificial intelligence to decode digital signature of individual patient

In the last few years a new aspect of radiomics i.e., Delta Radiomics is being researched [29]. Delta radiomics comprises extraction and comparison of quantitative features from sequential scans acquired over the course of treatment, which provides information on the efficacy of treatment.

Methods

This study is approved by the institutional ethics committee (IEC) as a retrospective study. In this study, we have performed literature surveys to find the emerging trend of radiomics based publications in oncology. Our search criteria are optimized to search only those articles, which mention radiomics or related terms like texture analysis in their title. We further extended our search and added year of publication as a criterion to find the total number of publications available on radiomics on PubMed and year-wise distribution of those publications. Furthermore, we added disease and segregated articles based on disease type. To understand the trend of imaging modality used for radiomic study we further included keywords like CT or PET or MRI along with search criteria in all fields. The details of search criteria adopted in this study are mentioned in tables 2 and 3.

Modality	Search Criteria
Radiomics	"Texture Analysis"[Title] OR "Textural Analysis"[Title] OR "Imaging Biomarker" [Title] OR "Radiomic" [Title] OR "Radiomics" [Title] AND 2000/01/01: 2021/12/31[dp]
Computed Tomography	"Texture Analysis"[Title] OR "Textural Analysis"[Title] OR "Imaging Biomarker"[Title] OR "Radiomics"[Title] OR "Radiomic" [Title] AND ("CT"[ALL] OR "Computed Tomography"[ALL]) AND 2000/01/01: 2021/12/31[dp]
Positron Emission Tomography	"Texture Analysis"[Title] OR "Textural Analysis"[Title] OR "Imaging Biomarker" [Title] OR "Radiomics" [Title] OR "Radiomic" [Title] AND ("PET" [ALL] OR "Positron" [ALL]) AND 2000/01/01: 2021/12/31[dp]
Magnetic Resonance Imaging	"Texture Analysis"[Title] OR "Textural Analysis"[Title] OR "Imaging Biomarker" [Title] OR "Radiomics" [Title] OR "Radiomic" [Title] AND ("MRI" [ALL] OR "magnetic" [ALL]) AND 2000/01/01: 2021/12/31[dp]
Positron Emission Tomography and Computed Tomography	"Texture Analysis"[Title] OR "Textural Analysis"[Title] OR "Imaging Biomarker" [Title] OR "Radiomics" [Title] OR "Radiomic" [Title] AND ("PET" [ALL] OR "Positron" [ALL]) AND ("CT"[ALL] OR "Computed Tomography"[ALL]) AND 2000/01/01: 2021/12/31[dp]
Positron Emission Tomography and	"Texture Analysis"[Title] OR "Textural Analysis"[Title] OR "Imaging Biomarker" [Title] OR "Radiomics" [Title] OR "Radiomic" [Title] AND ("MRI" [ALL] OR "magnetic" [ALL])

Magnetic Resonance Imaging	AND ("PET" [ALL] OR "Positron" [ALL]) AND 2000/01/01: 2021/12/31[dp]
Magnetic Resonance Imaging and Computed Tomography	"Texture Analysis"[Title] OR "Textural Analysis"[Title] OR "Imaging Biomarker" [Title] OR "Radiomics" [Title] OR "Radiomic" [Title] AND ("MRI" [ALL] OR "magnetic" [ALL]) AND ("CT"[ALL] OR "Computed Tomography"[ALL]) AND 2000/01/01: 2021/12/31[dp]
Positron Emission Tomography and Computed Tomography and Magnetic Resonance Imaging	"Texture Analysis"[Title] OR "Textural Analysis"[Title] OR "Imaging Biomarker" [Title] OR "Radiomics" [Title] OR "Radiomic" [Title] AND ("MRI" [ALL] OR "magnetic" [ALL]) AND ("PET" [ALL] OR "Positron" [ALL]) AND ("CT"[ALL] OR "Computed Tomography"[ALL]) AND 2000/01/01: 2021/12/31[dp]
Radiomic Stability	"Texture Analysis"[Title] OR "Textural Analysis"[Title] OR "Imaging Biomarker" [Title] OR "Radiomic" [Title] OR "Radiomics" [Title] AND ("repeatability" [Title] OR "reproducibility" [Title] OR "stability" [Title]) AND 2000/01/01: 2021/12/31[dp]

Table 2: the table shows the term and search criteria used to select a study based on the above-mentioned criteria.

Disease site	*Search Criteria*
Brain Tumor	"Texture Analysis"[Title] OR "Textural Analysis"[Title] OR "Imaging Biomarker" [Title] OR "Radiomics" [Title] OR "Radiomic" [Title] AND ("Brain" [Title] OR "GBM" [Title] OR "glioblastoma" [Title] OR "glial" [Title]) AND 2000/01/01: 2021/12/31[dp]
Head & Neck cancer	"Texture analysis"[title] or "textural analysis"[title] or "imaging biomarker" [title] or "radiomics" [title] OR "Radiomic" [Title] and ("head- and-neck" [title] or "neck" [title] or "head" [title]) and 2000/01/01: 2021/12/31[dp]
Lung Cancer	"Texture Analysis"[Title] OR "Textural Analysis"[Title] OR "Imaging Biomarker" [Title] OR "Radiomics" [Title] OR "Radiomic" [Title] AND ("lung" [Title] OR "nsclc" [Title] OR "sclc" [Title]) AND 2000/01/01 : 2021/12/31[dp]

Breast Cancer	"Texture analysis"[title] or "textural analysis"[title] or "imaging biomarker" [title] or "radiomics" [title] OR "Radiomic" [Title] and "Breast" [title] and 2000/01/01: 2021/12/31[dp]
Gastrointestinal Cancer	"Texture Analysis"[Title] OR "Textural Analysis"[Title] OR "Imaging Biomarker" [Title] OR "Radiomics" [Title] OR "Radiomic" [Title] AND ("Gastrointestinal" [Title] OR "intestine" [Title] OR "intestinal" [Title] OR "Liver" [Title] OR "HCC" [Title] OR "hepatocellular" [Title] OR "pancreatic" [Title] OR "pancreas" [Title]) AND 2000/01/01: 2021/12/31[dp]
Cervical Cancer	"Texture Analysis"[Title] OR "Textural Analysis"[Title] OR "Imaging Biomarker" [Title] OR "Radiomics" [Title] OR "Radiomic" [Title] AND ("Cervical" [Title] OR "Cervix" [Title]) AND 2000/01/01: 2021/12/31[dp]
Prostate cancer	"Texture Analysis"[Title] OR "Textural Analysis"[Title] OR "Imaging Biomarker" [Title] OR "Radiomics" [Title] OR "Radiomic" [Title] AND ("Prostate" [Title] OR "Prostatic" [Title]) AND 2000/01/01: 2021/12/31[dp]
Colorectal cancer	"Texture analysis"[title] or "textural analysis"[title] or "imaging biomarker" [title] or "radiomics" [title] OR "Radiomic" [Title] and ("Colorectal" [title] or "rectal" [title] or "Colon" [title]) and 2000/01/01: 2021/12/31[dp]

Table 3: the table shows the term and search criteria used to select radiomic studies published on various cancer types.

Results

We found in total 5243 articles published on radiomics since the year 2000 and satisfied our search criteria. Out of the total articles published on radiomics, 624, 2234 and 2110 articles had mention of PET, CT and MRI respectively (table 4). The detailed distribution of the publications year wise in all categories are shown in table (table 4). There were 123 studies published on radioiomic stability study. Maximum 549 articles were published on lung cancer alone followed by 533 articles on GI cancer.

	Publications on radiomics							
Year	**2021**	**2020**	**2019**	**2018**	**2017**	**2016**	**2015**	**Total**
Total	1549	1277	798	535	298	175	132	5243
CT	733	592	343	236	132	70	38	2234
PET	169	151	105	71	43	31	14	624
MRI	645	527	345	205	108	74	43	2110
CT-PET	142	118	82	46	33	26	5	475
PET-MR	42	33	21	17	9	6	2	135
CT-MRI	91	58	42	15	11	8	2	234
CT-MRI-PET	32	19	16	7	6	3	0	85
Stability	47	25	29	9	4	5	0	123
Brain Tumor	80	74	45	36	24	14	8	307
Head & Neck Cancer	26	31	30	12	11	4	2	122
Lung Cancer	155	157	78	71	39	20	10	549
Breast Cancer	110	93	58	31	25	11	13	369
GI Cancer	182	156	85	51	17	10	11	533
Cervical Cancer	38	29	15	12	5	2	0	104
Prostate Cancer	57	42	38	20	8	10	3	251
Colorectal Cancer	82	69	39	22	10	11	9	187

Table 4: Table shows the total and year wise publications on radiomics in oncology

The percentage of radiomic articles published on CT and MRI are almost the same 45% and 42% respectively (figure 4A). Radiomics articles published on lung and GI cancers contribute approximately 20% of total publications on radiomics (Table 4B).

The publication trend on Radiomic has shown very steep growth in the last decade (figure 4A). The trend shows that the yearly publications have increased many folds in

the last five years (figure 4B). A similar growth trend has been witnessed in all imaging types (figure 4A, B) and all types of cancers (figure 5A, B). In our study, we found 85 articles that have utilized all three imaging modalities for radiomic study (figure 4C). Figure 4D shows the year wise publication of radiomic articles on the stability of radiomic features.
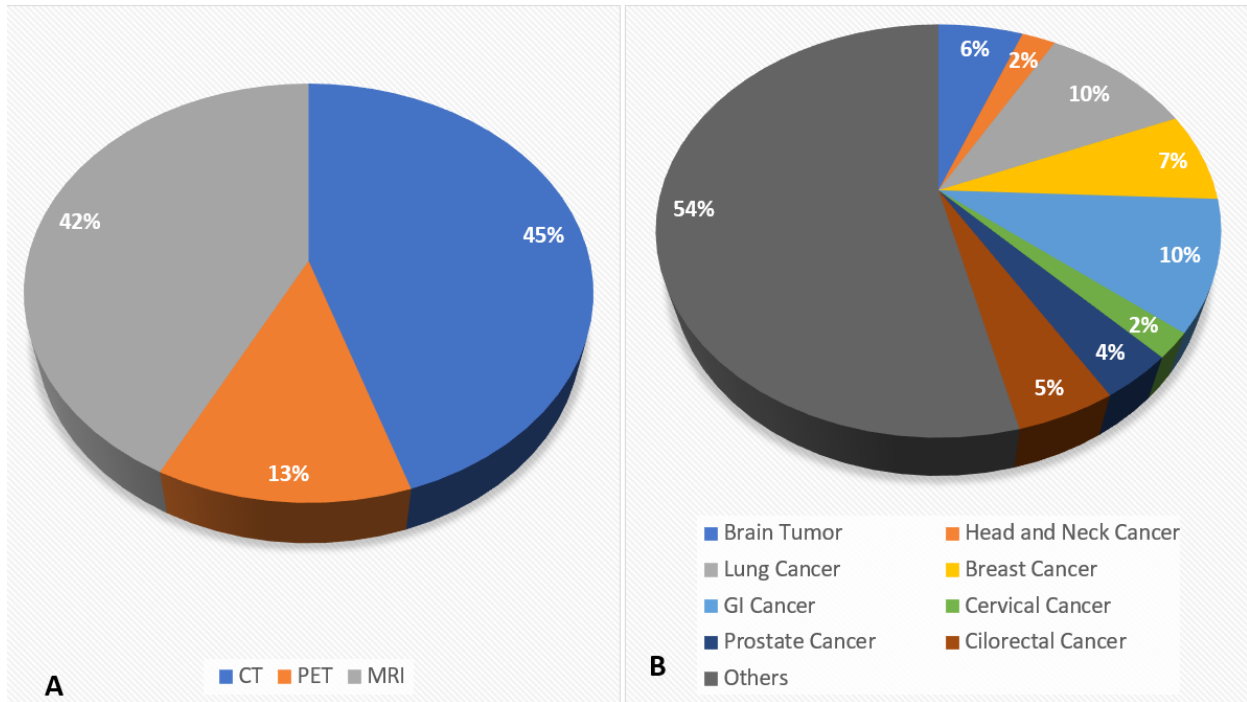

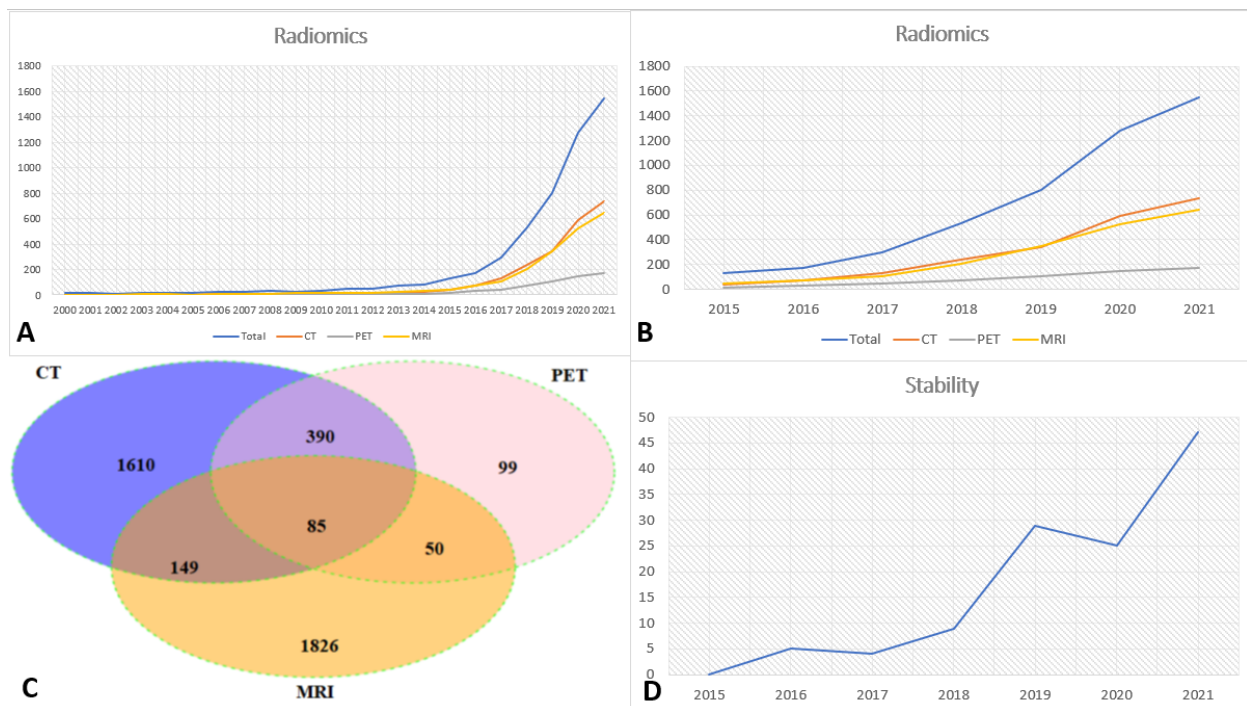Figure 3: Publications on Radiomics: A) Imaging the modality-based distribution of articles b) Disease wise distribution of articles

Figure 4: Figure shows (A) the trend of publications on PET, CT and MRI radiomics in oncology over the last two decades, (B) of publications on PET, CT and MRI radiomics in oncology since 2015. (C) Vann diagram shows the PET, CT and MRI imaging modality used for radiomic studies (D) shows the trend of published radiomic stability issues since 2015.
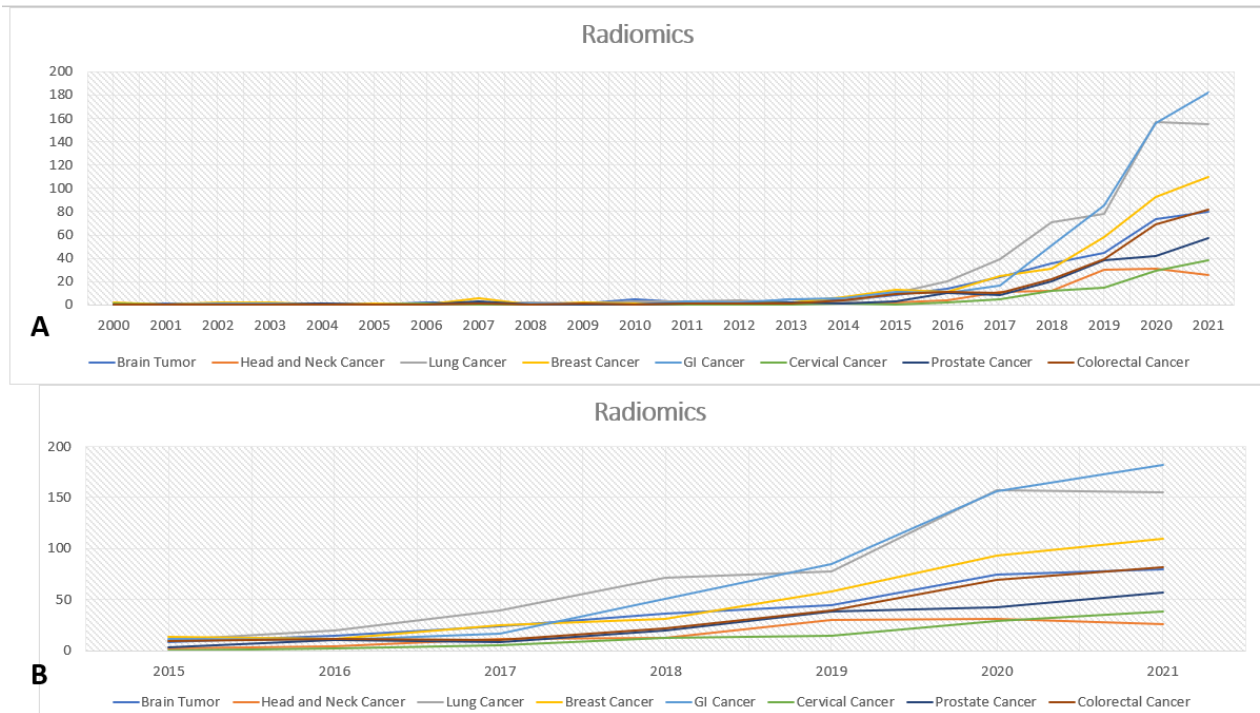

Figure 5: Figure shows (A) the trend of the number of publications on radiomics in various cancer types over the last two decades, the (B) trend of the number of publications on radiomics in various cancer types since 2015

Discussion

The utility of radiomic based prediction modelling has been tested widely in the diagnosis and treatment of all varieties of solid tumors. Several studies have been performed to differentiate high grade and low-grade gliomas and to develop various radiomic markers for treatment selection [26, 27]. Several studies have shown the association of radiomic features extracted from PET or MRI with survival in glioma [30-39]. Radiomics is widely used in the diagnosis and treatment of head-and-neck cancer [40]. A radiomic signature from PET, MRI and CT has been found to have a significant role in the prediction of stage of tumor, HPV status, hypoxia status and gene expression [41-51]. Studies have shown the role of radiomics in the characterization of sentinel lymph node metastasis in breast cancer non-invasively [52]. The role of radiomics has also been demonstrated by various researchers in breast cancer for response evaluation such as disease-free survival (DFS) [53-57]. The role of radiomics has been widely explored in lung cancer management [58]. Various studies have been performed to differentiate between benign and malignant tumor, pathology types (i.e., adenocarcinoma or squamous cell carcinoma), EGFR mutation status and various TMM stages [59-65]. Literature published in the last decade also suggests an increasing role of radiomic features in the prediction of OS, PFS, DFS, LRR, treatment response, toxicity and quality of life [66-72]. Radiomic features have been explored for the management of colorectal cancer. Various studies have demonstrated the role of radiomic features in the detection of lymph node metastasis, prediction of KRAS/NRAS/ BRAF mutation [73, 74]. The role of radiomic features has also been investigated for treatment selection, treatment modification and DFS prediction [75-78]. Radiomics has been investigated in prostate cancer management and radiomic features extracted from MRI and PET have shown promising results. Several studies have shown the utility of radiomic features in the differentiation between benign and malignant tumor, the aggressiveness of tumors and the Gleason Score [79-82]. Many researchers have also shown the utility of radiomic features extracted from MRI and PET to predict biochemical recurrence, PFS and OS [83-85]. GI and liver cancer is another area where the role of radiomics has been investigated in disease management. The role of radiomics has been successfully demonstrated in microvascular invasion detection of liver cancer and differentiation in various kinds of GI malignancies, histology type and TNM staging in GI cancer [86-92]. Various studies have demonstrated the role of radiomic features in the detection of lymph node metastasis, OS, PSF and toxicity prediction in cervical cancer [93-96].

Our study shows an increasing trend of radiomics in oncology in the last decade. The last five years witnessed the tremendous growth of radiomic studies in oncology. In all major disease types, the growth of radiomic studies has been witnessed. Several articles have been published on radiomic stability problems that show that the researchers have identified stability as a major issue in radiomic implementation.

Implementation or radiomics based workflow in the clinic: The future of radiomics lies in the clinical application and implementation of radiomics. A self-learning model may be developed and implemented in the clinic for participation in the decision support system. There will be requirements for a super-specialized model to address the specific clinical questions. As suggested by Lambin et al, the image archival system i.e., PACS has to be modified to Picture archiving and radiomics knowledge systems (PARKS) to store radiomic signatures [16]. The future implementation of the radiomic process may look like Figure 6.
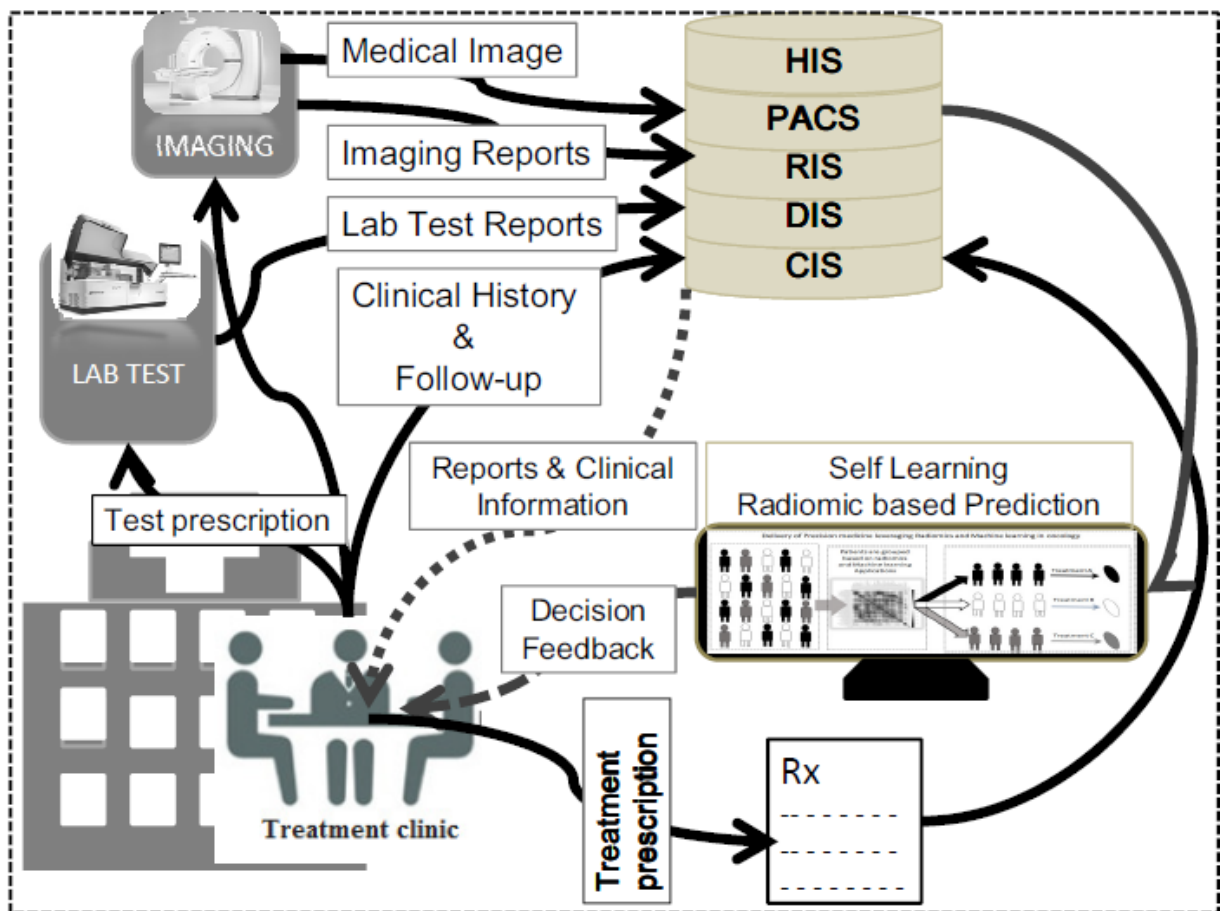

Figure 6: Clinical workflow of radiomics: Patient Arrival→ Patient examination → Test Prescription → Lab Test + Radiology Scan → Radiomic analysis (Self-learning Prediction model) → Decision support based on radiomic Analysis → Treatment prescription → follow-up data collection and entry

Limitations of Radiomic implementation: The main problem of Radiomics is its limited repeatability and reproducibility which is thought to be mainly caused by the difference in scanners from different vendors, different acquisition protocols and intra scanner variations. In our earlier repeatability and reproducibility study, we found that

only 10% of CT radiomic features had good repeatability and reproducibility in a clinical cohort and in phantoms [97]. Traverso et al. in a systematic literature review have also concluded that there are stability issues with the majority of radiomic features [98]. In order to harmonize radiomic extraction tools, features and imaging standards, several initiatives are started by various agencies, like The Quantitative Imaging Network (QIN) [99], the Quantitative Imaging Biomarkers Alliance (QIBA) [100], and Quantitative Imaging in Cancer: Connecting Cellular Processes with Therapy (QuIC-ConCePT) [101]. These initiatives are working to standardize imaging and imaging biomarkers. The Image Biomarker Standardization Initiative (IBSI) is another consortium that works towards the harmonization of radiomic features across the globe by minimizing the deviation in imaging and standardizing the radiomic extraction process [102, 103]. The radiomics quality score (RQS) is another such initiative proposed by Lambin et al. to address the issues related to radiomic study reporting [16]. Most of these initiatives will assist in advancing the standardization process of imaging biomarkers and are thus expected to address the repeatability and reproducibility challenges currently present in Radiomics.

Conclusions

Literature review is suggestive of the increasing role of radiomics in precision oncology. Publications on radiomics have increased many folds in the last 5 years. Initiatives like QIN, QIBA, QuIC-ConCePT, IBSI and RQS will be able to address repeatability and reproducibility of radiomic features. We envision that radiomics is going to play a pivotal role in phenotyping the cancer and guide cancer management to provide more precise treatments to patients in a true clinical environment soon.

References

[1] Lodish H. Section 24.1, Tumor Cells and the Onset of Cancer, In: Lodish H, Berk A, Zipursky SL, et al. Molecular Cell Biology. 4th edition. New York: W. H. Freeman; 2000. Available from:https://www.ncbi.nlm.nih.gov/books/NBK21590/

[2] Rai KR, Keating MJ. Treatment. In: Kufe DW, Pollock RE, Weichselbaum RR, et al., editors. Holland-Frei Cancer Medicine. 6th edition. Hamilton (ON): BC Decker; 2003. Available from:https://www.ncbi.nlm.nih.gov/books/NBK12735/

[3]Maeda H, Khatami M. Analyses of repeated failures in cancer therapy for solid tumors: poor tumor-selective drug delivery, low therapeutic efficacy and unsustainable costs. Clin Transl Med. 2018 Mar 1;7(1):11. doi: 10.1186/s40169-018-0185-6.

[4] Schwartzberg L, Kim ES, Liu D, Schrag D. Precision Oncology: Who, How, What, When, and When Not?. *Am Soc Clin Oncol Educ Book*. 2017;37:160-169. doi:10.1200/EDBK_174176

[5] Takeuchi S, Okuda S. Knowledge base toward understanding actionable alterations and realizing precision oncology. *Int J Clin Oncol*. 2019;24(2):123-130. doi:10.1007/s10147-018-1378-0

[6] Goossens N, Nakagawa S, Sun X, Hoshida Y. Cancer biomarker discovery and validation. Transl Cancer Res. 2015 Jun;4(3):256-269. doi: 10.3978/j.issn.2218-676X.2015.06.04.

[7] Biomarkers Definitions Working Group.. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin Pharmacol Ther. 2001;69(3):89-95. doi:10.1067/mcp.2001.113989

[8] Sasaki H, Endo K, Okuda K, et al. Epidermal growth factor receptor gene amplification and gefitinib sensitivity in patients with recurrent lung cancer. J Cancer Res Clin Oncol.

2008;134(5):569-577. doi:10.1007/s00432-007-0320-z

[9] Bethune G, Bethune D, Ridgway N, Xu Z. Epidermal growth factor receptor (EGFR) in lung cancer: an overview and update. J Thorac Dis. 2010 Mar;2(1):48-51.

[10] Shimizu H, Nakayama KI. Artificial intelligence in oncology. Cancer Sci. 2020;111(5):1452-1460. doi:10.1111/cas.14377

[11] Deig CR, Kanwar A, Thompson RF. Artificial Intelligence in Radiation Oncology. Hematol Oncol Clin North Am. 2019;33(6):1095-1104. doi:10.1016/j.hoc.2019.08.003

[12] Kann BH, Thompson R, Thomas CR Jr, Dicker A, Aneja S. Artificial Intelligence in Oncology: Current Applications and Future Directions. Oncology (Williston Park). 2019;33(2):46-53.

[13] O'Connor JP, Aboagye EO, Adams JE, et al. Imaging biomarker roadmap for cancer studies. Nat Rev Clin Oncol. 2017;14(3):169-186. doi:10.1038/nrclinonc.2016.162

[14] Skwarski M, Higgins GS. A new roadmap to improve translation of imaging biomarkers. Br J Cancer. 2016;115(12):1443-1444. doi:10.1038/bjc.2016.374

[15] European Society of Radiology (ESR). ESR statement on the stepwise development of imaging biomarkers. Insights Imaging. 2013;4(2):147-152. doi:10.1007/s13244-013-0220-5

[16] Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol. 2017;14(12):749-762. doi:10.1038/nrclinonc.2017.141

[17] Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer.* (2012) 48:441–6. doi: 10.1016/j.ejca.2011.11.036

[18] Jha AK, Mithun S, Rangarajan V, Wee L, Dekker A. Emerging role of artificial intelligence in nuclear medicine. Nucl Med Commun. 2021;42(6):592-601. doi:10.1097/MNM.0000000000001381

[19]Afshar, P., et al., From handcrafted to deep-learning-based cancer radiomics: challenges and opportunities. IEEE Signal Processing Magazine, 2019. 36(4): p. 132-160.

[20] Image processing, https://ibsi.readthedocs.io/en/latest/02_Image_processing.html, Last accessed on 15-05-2021

[21] Bermingham ML, Pong-Wong R, Spiliopoulou A, et al. Application of high-dimensional feature selection: evaluation for genomic prediction in man. Sci Rep. 2015;5:10312. Published 2015 May 19. doi:10.1038/srep10312

[22] Wang, Lidong& Alexander, Cheryl. (2016). Machine Learning in Big Data. International Journal of Mathematical, Engineering and Management Sciences. 1. 52-61. 10.33889/IJMEMS.2016.1.2-006.

[23] Goodfellow I, Bengio Y, Courville A. Deep learning. Cambridge, Mass: MIT Press, 2016; 1–26.

[24] Jordan, M. I., and Mitchell, T. M. (2015). Machine learning: trends, perspectives, and prospects. Science 349, 255–260. doi: 10.1126/science.aaa8415

[25] https://pyradiomics.readthedocs.io/en/latest/features.html accessed on 10/06/2021

[26]Li, Z., et al., Deep Learning based Radiomics (DLR) and its usage in noninvasive IDH1 prediction for low grade glioma. Scientific reports, 2017. 7(1): p. 1-11.

[27]Zhou, L., et al., A deep learning-based radiomics model for differentiating benign and malignant renal tumors. Translational oncology, 2019. 12(2): p. 292-300.

[28]Lao, J., et al., A deep learning-based radiomics model for prediction of survival in glioblastoma multiforme. Scientific reports, 2017. 7(1): p. 1-8.

[29] Fave X, Zhang L, Yang J, et al. Delta-radiomics features for the prediction of patient outcomes in non-small cell lung cancer. Sci Rep. 2017;7(1):588. Published 2017 Apr 3.

[30] Itakura H, Achrol AS, Mitchell LA, Loya JJ, Liu T, Westbroek EM, Magnetic resonance image features identify glioblastoma phenotypic subtypes with distinct molecular pathway activities. Sci Transl Med; 2015. p. 7.

[31] Bai Y, Lin YS, Tian J, Shi DP, Cheng JL, Haacke EM. et al. Grading of Gliomas by Using Monoexponential, Biexponential, and Stretched Exponential Diffusion-weighted MR Imaging and Diffusion Kurtosis MR Imaging. Radiology. 2016;278:496–504.

[32] Chen W, Liu B, Peng S, Sun J, Qiao X. Computer-Aided Grading of Gliomas Combining Automatic Segmentation and Radiomics. Int J Biomed Imaging. 2018;2018:2512037

[33] Kickingereder P, Gotz M, Muschelli J, Wick A, Neuberger U, Shinohara RT. et al. Large-scale Radiomic Profiling of Recurrent Glioblastoma Identifies an Imaging Predictor for Stratifying Anti-Angiogenic Treatment Response. Clinical Cancer Research. 2016;22:5765–71.

[34] Shim KY, Chung SW, Jeong JH, et al. Radiomics-based neural network predicts recurrence patterns in glioblastoma using dynamic susceptibility contrast-enhanced MRI. Sci Rep. 2021;11(1):9974.

[35] Kim JY, Park JE, Jo Y, et al. Incorporating diffusion- and perfusion-weighted MRI into a radiomics model improves diagnostic performance for pseudoprogression in glioblastoma patients. Neuro Oncol. 2019;21(3):404-414.

[36] Grossmann P, Narayan V, Chang K, Rahman R, Abrey L, Reardon DA. et al. Quantitative imaging biomarkers for risk stratification of patients with recurrent glioblastoma treated with bevacizumab. Neuro Oncol. 2017;19:1688–97.

[37] Papp L, Potsch N, Grahovac M, Schmidbauer V, Woehrer A, Preusser M. et al. Glioma Survival Prediction with Combined Analysis of In Vivo C-11-MET PET Features, Ex Vivo Features, and Patient Features by Supervised Machine Learning. J Nucl Med. 2018;59:892–9.

[38] Park JE, Kim HS, Jo Y, et al. Radiomics prognostication model in glioblastoma using diffusion- and perfusion-weighted MRI. Sci Rep. 2020;10(1):4250.

[39] Perez-Beteta J, Molina-Garcia D, Ortiz-Alhambra JA, Fernandez-Romero A, Luque B, Arregui E. et al. Tumor Surface Regularity at MR Imaging Predicts Survival and Response to Surgery in Patients with Glioblastoma. Radiology. 2018;288:218–25

[40] Jethanandani A, Lin TA, Volpe S, Elhalawani H, Mohamed ASR, Yang P, Exploring Applications of Radiomics in Magnetic Resonance Imaging of Head and Neck Cancer: A Systematic Review. Frontiers in Oncology; 2018. p. 8.

[41] Ren J, Tian J, Yuan Y, Dong D, Li X, Shi Y. et al. Magnetic resonance imaging based radiomics signature for the preoperative discrimination of stage I-II and III-IV head and neck squamous cell carcinoma. European journal of radiology. 2018;106:1–6.

[42] Rth L, Bogowicz M, Jochems A, Fjp H, Fwr F, Huang SH. et al. Development and validation of a radiomic signature to predict HPV (p16) status from standard CT imaging: a multicenter study. British Journal of Radiology. 2018;91:20170498.

[43] Zhou Z, Chen L, Sher D, Zhang Q, Shah J, Pham N-L, Predicting Lymph Node Metastasis in Head and Neck Cancer by Combining Many-objective Radiomics and 3-dimensioal Convolutional Neural Network through Evidential Reasoning. Conf Proc IEEE Eng Med Biol Soc; 2018. pp. 1–4

[44] Chen RY, Lin YC, Shen WC, Hsieh TC, Yen KY, Chen SW. et al. Associations of Tumor PD-1 Ligands, Immunohistochemical Studies, and Textural Features in 18 F-FDG PET in Squamous Cell Carcinoma of the Head and Neck. Sci Rep. 2018;8:105

[45] Chen L, Wang H, Zeng H, Zhang Y, Ma X. Evaluation of CT-based radiomics signature and nomogram as prognostic markers in patients with laryngeal squamous cell carcinoma. Cancer Imaging. 2020 Apr 22;20(1):28.

[46] Crispin-Ortuzar M, Apte A, Grkovski M, Oh JH, Lee NY, Schöder H. et al. Predicting hypoxia status using a combination of contrast-enhanced computed tomography and [18F]-Fluorodeoxyglucose positron emission tomography radiomics features. Radiother Oncol. 2018;127:36–42

[47] Zhang B, Tian J, Dong D, Gu D, Dong Y, Zhang L. et al. Radiomics Features of Multiparametric MRI as Novel Prognostic Factors in Advanced Nasopharyngeal Carcinoma. Clinical Cancer Research An Official Journal of the American Association for Cancer Research. 2017;23:4259.

[48] Zhang B, He X, Ouyang F, Gu D, Dong Y, Zhang L. et al. Radiomic machine-learning classifiers for prognostic biomarkers of advanced nasopharyngeal carcinoma. Cancer Letters. 2017;403:21.

[49] van Rossum PS, Fried DV, Zhang L, et al. The Incremental Value of Subjective and Quantitative Assessment of 18F-FDG PET for the Prediction of Pathologic Complete

Response to Preoperative Chemoradiotherapy in Esophageal Cancer. J Nucl Med. 2016;57(5):691-700.

[50] Wang G, He L, Yuan C, Huang Y, Liu Z, Liang C. Pretreatment MR imaging radiomics signatures for response prediction to induction chemotherapy in patients with nasopharyngeal carcinoma. European Journal of Radiology. 2018;98:100–6.

[51] Sanduleanu S, Jochems A, Upadhaya T, et al. Non-invasive imaging prediction of tumor hypoxia: A novel developed and externally validated CT and FDG-PET-based radiomic signatures. Radiother Oncol. 2020;153:97-105. doi:10.1016/j.radonc.2020.10.016

[52] Dong Y, Feng Q, Yang W, Lu Z, Deng C, Zhang L. et al. Preoperative prediction of sentinel lymph node metastasis in breast cancer based on radiomics of T2-weighted fat-suppression and diffusion-weighted MRI. European Radiology. 2018;28:582–91.

[53] Chan H, Bhm VDV, Loo CE, Kga G. Eigentumors for prediction of treatment failure in patients with early-stage breast cancer using dynamic contrast-enhanced MRI: a feasibility study. Physics in Medicine & Biology; 2017. p. 62.

[54] Braman NM, Etesami M, Prasanna P, Dubchuk C, Gilmore H, Tiwari P. et al. Intratumoral and peritumoral radiomics for the pretreatment prediction of pathological complete response to neoadjuvant chemotherapy based on breast DCE-MRI. Breast Cancer Research. 2017;19:57.

[55] Partridge SC, Zhang Z, Newitt DC, Gibbs JE, Chenevert TL, Rosen MA, Diffusion-weighted MRI findings predict pathologic response in neoadjuvant treatment of breast cancer: the ACRIN 6698 Multicenter Trial. Radiology. 2018; 1802. 73.

[56] Tran WT, Gangeh MJ, Sannachi L, Chin L, Watkins E, Bruni SG. et al. Predicting breast cancer response to neoadjuvant chemotherapy using pretreatment diffuse optical spectroscopic texture analysis. British Journal of Cancer. 2017;116:1329–39.

[57] Park H, Lim Y, Ko ES, Cho H-h, Lee JE, Han B-K, Radiomics Signature on Magnetic Resonance Imaging: Association with Disease-Free Survival in Patients with Invasive Breast Cancer. Clinical Cancer Research; 2018. clincanres. 3783.2017.

[58] Avanzo M, Stancanello J, Pirrone G, Sartor G. Radiomics and deep learning in lung cancer. Strahlenther Onkol. 2020;196(10):879-887.

[59] Maldonado F, Boland JM, Raghunath S, Aubry MC, Bartholmai BJ, Deandrade M. et al. Non-invasive Characterization of the Histopathologic Features of Pulmonary Nodules of the Lung Adenocarcinoma Spectrum using Computer Aided Nodule Assessment and Risk Yield (CANARY) - a Pilot Study. Journal of Thoracic Oncology. 2013;8:452–60.

[60] Liu Y, Balagurunathan Y, Atwater T, Antic S, Li Q, Walker RC. et al. Radiological Image traits Predictive of Cancer Status in Pulmonary Nodules. Clinical Cancer Research. 2016;23:1442–9.

[61] Zhang L, Chen B, Liu X, Song J, Fang M, Hu C. et al. Quantitative Biomarkers for Prediction of Epidermal Growth Factor Receptor Mutation in Non-Small Cell Lung Cancer. Translational Oncology. 2018;11:94–101.

[62] Wu W, Parmar C, Grossmann P, Quackenbush J, Lambin P, Bussink J. et al. Exploratory Study to Identify Radiomics Classifiers for Lung Cancer Histology. Frontiers in Oncology. 2016;6:71.

 [63]Bashir U, Kawa B, Siddique M, et al. Non-invasive classification of non-small cell lung cancer: a comparison between random forest models utilising radiomic and semantic features. Br J Radiol. 2019;92(1099):20190159. doi:10.1259/bjr.20190159

[64] Zhu X, Dong D, Chen Z, Fang M, Zhang L, Song J, Radiomic signature as a diagnostic factor for histologic subtype classification of non-small cell lung cancer. European Radiology; 2018. pp. 1–7. [PubMed] [Google Scholar]

[65] Fan L, Fang MJ, Li ZB, Tu WT, Wang SP, Chen WF, Radiomics signature: a biomarker for the preoperative discrimination of lung invasive adenocarcinoma manifesting as a ground-glass nodule. European Radiology; 2018. pp. 1–9. [Google Scholar]

[66] Cunliffe A, Armato SG 3rd, Castillo R, Pham N, Guerrero T, Al-Hallaq HA. Lung texture in serial thoracic computed tomography scans: correlation of radiomics-based features with radiation therapy dose and radiation pneumonitis development. Int J Radiat Oncol Biol Phys. 2015;91(5):1048-1056.

[67] Coroller TP, Grossmann P, Hou Y, et al. CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma. Radiother Oncol. 2015;114(3):345-350.

[68] Mattonen SA, Tetar S, Palma DA, Senan S, Ward AD. Automated Texture Analysis for Prediction of Recurrence After Stereotactic Ablative Radiation Therapy for Lung Cancer. International Journal of Radiation Oncology Biology Physics. 2015;93:S5–S6.

[69] Coroller TP, Agrawal V, Narayan V, Hou Y, Grossmann P, Lee SW. et al. Radiomic phenotype features predict pathological response in non-small cell lung cancer. Radiother Oncol. 2016;119:480–6.

[70] Aerts HJWL, Patrick G, Tan Y, Oxnard GG, Naiyer R, Schwartz LH. et al. Defining a Radiomic Response Phenotype: A Pilot Study using targeted therapy in NSCLC. Scientific Reports. 2016;6:33860.

[71] Cook GJ, O'Brien ME, Siddique M, Chicklore S, Loi HY, Sharma B. et al. Non-Small Cell Lung Cancer Treated with Erlotinib: Heterogeneity of (18)F-FDG Uptake at PET-Association with Treatment Response and Prognosis. Radiology. 2015;276:883–93.

[72] Huang Y, Liu Z, He L, Chen X, Pan D, Ma Z. et al. Radiomics Signature: A Potential Biomarker for the Prediction of Disease-Free Survival in Early-Stage (I or II) Non-Small Cell Lung Cancer. Radiology. 2016;281:947.

[73]Huang YQ, Liang CH, He L, Tian J, Liang CS, Chen X. et al. Development and Validation of a Radiomics Nomogram for Preoperative Prediction of Lymph Node Metastasis in Colorectal Cancer. Journal of Clinical Oncology. 2016;34:2157.

[74] Yang L, Dong D, Fang M, Zhu Y, Zang Y, Liu Z. et al. Can CT-based radiomics signature predict KRAS/NRAS/BRAF mutations in colorectal cancer? European radiology. 2018;28:2058–67. [PubMed] [Google Scholar]

[75] Liu ZY, Zhang XY, Shi YJ, Wang L, Zhu HT, Tang ZC. et al. Radiomics Analysis for Evaluation of Pathological Complete Response to Neoadjuvant Chemoradiotherapy in Locally Advanced Rectal Cancer. Clin Cancer Res. 2017;23:7253–62

[76] Chang GJ, Rodriguezbigas MA, Skibber JM, Moyer VA. Lymph node evaluation and survival after curative resection of colon cancer: systematic review. J Natl Cancer Inst. 2004;99:433–41.

[77] Toiyama Y, Inoue Y, Shimura T, Fujikawa H, Saigusa S, Hiro J. et al. Serum Angiopoietin-like Protein 2 Improves Preoperative Detection of Lymph Node Metastasis in Colorectal Cancer. Anticancer Research. 2015;35:2849–56.

[78] Huang YQ, Liang CH, He L, et al. Development and Validation of a Radiomics Nomogram for Preoperative Prediction of Lymph Node Metastasis in Colorectal Cancer [published correction appears in J Clin Oncol. 2016 Jul 10;34(20):2436]. J Clin Oncol. 2016;34(18):2157-2164.

[79] Chen T, Li M, Gu Y, Zhang Y, Yang S, Wei C, Prostate Cancer Differentiation and Aggressiveness: Assessment With a Radiomic-Based Model vs. PI-RADS v2. J Magn Reson Imaging; 2018.

[80] Wang J, Wu CJ, Bao ML, Zhang J, Wang XN, Zhang YD. Machine learning-based analysis of MR radiomics can help to improve the diagnostic performance of PI-RADS v2 in clinically relevant prostate cancer. European Radiology. 2017;27:4082–90.

[81] Algohary A, Viswanath S, Shiradkar R, Ghose S, Pahwa S, Moses D. et al. Radiomic features on MRI enable risk categorization of prostate cancer patients on active surveillance: Preliminary findings. J Magn Reson Imaging. 2018;48:818–28.

[82] Chaddad A, Kucharczyk MJ, Niazi T. Multimodal Radiomic Features for the Predicting Gleason Score of Prostate Cancer. Cancers; 2018. p. 10

[83] Lin YC, Lin G, Hong JH, Lin YP, Chen FH, Ng SH. et al. Diffusion radiomics analysis of intratumoral heterogeneity in a murine prostate cancer model following radiotherapy: Pixelwise correlation with histology. Journal of Magnetic Resonance Imaging. 2017;46:483–9

[84] Shiradkar R, Ghose S, Jambor I, Taimen P, Ettala O, Purysko AS. et al. Radiomic features from pretreatment biparametric MRI predict prostate cancer biochemical recurrence: Preliminary findings. J Magn Reson Imaging. 2018;48:1626–36

[85] Staal FCR, van der Reijd DJ, Taghavi M, Lambregts DMJ, Beets-Tan RGH, Maas M. Radiomics for the Prediction of Treatment Outcome and Survival in Patients With Colorectal Cancer: A Systematic Review. Clin Colorectal Cancer. 2021;20(1):52-71. doi:10.1016/j.clcc.2020.11.001

[86] Bakr S, Echegaray S, Shah R, Kamaya A, Louie J, Napel S. et al. Noninvasive radiomics signature based on quantitative analysis of computed tomography images as a surrogate for microvascular invasion in hepatocellular carcinoma: a pilot study. Journal of Medical Imaging. 2017;4:041303

[87] Ma Z, Fang M, Huang Y, He L, Chen X, Liang C. et al. CT-based radiomics signature for differentiating Borrmann type IV gastric cancer from primary gastric lymphoma. European Journal of Radiology. 2017;91:142.

[88] Ba-Ssalamah A, Muin D, Schernthaner R, Kulinna-Cosentini C, Bastati N, Stift J. et al. Texture-based classification of different gastric tumors at contrast-enhanced CT. European Journal of Radiology. 2013;82:E537–E43.

[89] Liu S, Song L, Ji C, Zheng H, Xia P, Zhang Y. et al. Application of CT texture analysis in predicting histopathological characteristics of gastric cancers. European Radiology. 2017;27:1–9.

[90] Liu S, Zhang Y, Chen L, Guan W, Guan Y, Ge Y. et al. Whole-lesion apparent diffusion coefficient histogram analysis: significance in T and N staging of gastric cancers. Bmc Cancer. 2017;17:665.

[91] Liu S, Zhang Y, Xia J, Chen L, Guan W, Guan Y. et al. Predicting the nodal status in gastric cancers: The role of apparent diffusion coefficient histogram characteristic analysis. Magnetic Resonance Imaging. 2017;42:144.

[92] Liu S, Zheng H, Zhang Y, Chen L, Guan W, Guan Y. et al. Whole-volume apparent diffusion coefficient-based entropy parameters for assessment of gastric cancer aggressiveness. Journal of Magnetic Resonance Imaging. 2017;47:168.

[93] Lucia F, Visvikis D, Vallières M, Desseroit MC, Miranda O, Robin P, Bonaffini PA, Alfieri J, Masson I, Mervoyer A, Reinhold C, Pradier O, Hatt M, Schick U. External validation of a combined PET and MRI radiomics model for prediction of recurrence in

cervical cancer patients treated with chemoradiotherapy. Eur J Nucl Med Mol Imaging. 2019 Apr;46(4):864-877. doi: 10.1007/s00259-018-4231-9.

[94] Sun C, Tian X, Liu Z, Li W, Li P, Chen J, Zhang W, Fang Z, Du P, Duan H, Liu P, Wang L, Chen C, Tian J. Radiomic analysis for pretreatment prediction of response to neoadjuvant chemotherapy in locally advanced cervical cancer: A multicentre study. EBioMedicine. 2019 Aug;46:160-169. doi: 10.1016/j.ebiom.2019.07.049. Epub 2019 Aug 6

[95] Altazi BA, Fernandez DC, Zhang GG, Hawkins S, Naqvi SM, Kim Y, Hunt D, Latifi K, Biagioli M, Venkat P, Moros EG. Investigating multi-radiomic models for enhancing prediction power of cervical cancer treatment outcomes. Phys Med. 2018 Feb;46:180-188. doi: 10.1016/j.ejmp.2017.10.009. Epub 2018 Feb 21.

[96] Reuzé S, Orlhac F, Chargari C, Nioche C, Limkin E, Riet F, Escande A, Haie-Meder C, Dercle L, Gouy S, Buvat I, Deutsch E, Robert C. Prediction of cervical cancer recurrence using textural features extracted from 18F-FDG PET images acquired with different scanners. Oncotarget. 2017 Jun 27;8(26):43169-43179. doi: 10.18632/oncotarget.17856.

[97]Jha AK, Mithun S, Jaiswar V, et al. Repeatability and reproducibility study of radiomic features on a phantom and human cohort. Sci Rep. 2021;11(1):2055. Published 2021 Jan 21. doi:10.1038/s41598-021-81526-8

[98]Traverso A, Wee L, Dekker A, Gillies R. Repeatability and Reproducibility of Radiomic Features: A Systematic Review. Int J Radiat Oncol Biol Phys. 2018;102(4):1143-1158. doi:10.1016/j.ijrobp.2018.05.053

[99] National Cancer Institute, Division of Cancer Treatment & Diagnosis. Quantitative Imaging Network (QIN) [online], https://imaging.cancer.gov/programs_resources/specialized_initiatives/qin.htm . last accessed on 15-05-2021

[100] Radiological Society of North America. Quantitative Imaging Biomarkers Alliance® (QIBA®). rsna.orghttps://www.rsna.org/research/quantitative-imaging-biomarkers-alliance/. last accessed on 15-05-2021

[101]QUantitative Imaging in Cancer: CONnecting CEllular Processes with Therapy, https://cordis.europa.eu/project/id/115151 last accessed on 15-05-2021

[102] Zwanenburg A, Vallières M, Abdalah MA, et al. The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. Radiology. 2020;295(2):328-338. doi:10.1148/radiol.2020191145

[103] The image biomarker standardisation initiative, https://ibsi.readthedocs.io/en/latest/index.html, Last accessed on 15-05-2021

Supplementary material:

We use open source pyRadiomics software for radiomic extraction. The radiomic features described below is based on pyRadiomic descriptions [1].

Following are the feature groups:

    A. First Order Features

    B. Shape Features (3D)

    C. Gray Level Co-occurrence Matrix (GLCM) Features

    D. Gray Level Size Zone Matrix (GLSZM) Features

    E. Gray Level Run Length Matrix (GLRLM) Features

    F. Neighbouring Gray Tone Difference Matrix (NGTDM) Features

    G. Gray Level Dependence Matrix (GLDM) Features

    H. Wavelet features

    I. Laplacian of Gaussian (LoG) Features

A. First Order Features:

It is the distribution of voxel intensities in the image within the Region of interest (ROI).

Let:

X be a set of N_p voxels included in the ROI

N_g is the number of non-zero bins, equally spaced from o with a width defined in the bin Width parameter.

$P(i)$ be the first order histogram with N_g discrete intensity levels,

$p(i) = P(i)/N_p$ (the normalized first order histogram)

    1. **Energy**

$$Energy = \sum_{i=1}^{N_p} (X(i)+c)^2$$

Here, c is optional value, defined by voxel Array Shift, which shifts the intensities to prevent negative values in X.

2. Total Energy

$$Total\ Energy = Vvoxel \sum_{i=1}^{N_p} (X(i)+c)^2$$

3. Entropy

$$Entropy = \sum_{i=1}^{N_p} p(i)\ log_2(p(i) + \epsilon)$$

Here, ϵ is an arbitrarily small positive number ($\approx 2.2 \times 10^{-16}$).

4. Minimum

Minimum = min(X)

The minimum gray level intensity within the ROI

5. 10th percentile

The 10th percentile of X

6. 90th percentile

The 90th percentile of X

7. Maximum

Maximum = max(X)

The maximum gray level intensity within the ROI

8. Mean

$$Mean = \frac{1}{N_p} \sum_{i=1}^{N_p} X(i)$$

The average gray level intensity within the ROI

9. Median

The median gray level intensity within the ROI

10. Interquartile Range

Interquartile range=P_{75}–P_{25}

Here, P_{25} and P_{75} are the 25th and 75th percentile of the image array, respectively.

11. Range

Range = max(X) – min(X)

12. Mean Absolute Deviation (MAD)

$$MAD = \frac{1}{N_p} \sum_{i=1}^{N_p} |X(i) - \overline{X}|$$

13. Robust Mean Absolute Deviation (rMAD)

$$rMAD = \frac{1}{N_p} \sum_{i=1}^{N_p} |X_{10-90}(i) - \overline{X}_{10-90}|$$

14. Root Mean Squared (RMS)

$$RMS = \sqrt{\frac{1}{N_p} \sum_{i=1}^{N_p} (X(i)+c)^2}$$

15. Standard Deviation

$$SD = \sqrt{\frac{1}{N_p} \sum_{i=1}^{N_p} (X(i) - \overline{X})^2}$$

16. Skewness

$$Skewness = \frac{\mu_3}{\sigma^3} = \frac{\frac{1}{N_p} \sum_{i=1}^{N_p} (X(i) - \overline{X})^3}{\left(\sqrt{\frac{1}{N_p} \sum_{i=1}^{N_p} (X(i) - \overline{X})^2}\right)^3}$$

17. Kurtosis

$$Kurtosis = \frac{\mu_4}{\sigma^4} = \frac{\frac{1}{N_p} \sum_{i=1}^{N_p} (X(i) - \overline{X})^4}{\left(\frac{1}{N_p} \sum_{i=1}^{N_p} (X(i) - \overline{X})^2\right)^4}$$

18. Variance

$$Variance = \frac{1}{N_p} \sum_{i=1}^{N_p} (X(i) - \overline{X})^2$$

19. Uniformity

$$Uniformity = \sum_{i=1}^{N_p} p(i)^2$$

B. Shape Features (3D)

Shape based features include the descriptors of the three-dimensional size and shape of the ROI. These features of this group are independent from the gray level intensity distribution in the ROI. Hence these features air only calculated on the original image and ROI.

Let:

V_f represent the number of volume elements included in the mesh

N_v represent the number of voxels included in the ROI

N_f represent the number of faces (triangles) defining the Mesh.

20. Mesh Volume

$$(V_i) = \frac{Oa_i \ . (Ob_i \ \times \ Oc_i)}{6} - - - -(1)$$

21. Voxel Volume

$$Mesh\ Volume\ (V) = \sum_{i=1}^{V_f} V_i - - - - - (2)$$

21. Voxel Volume

$$V_{voxel} = \sum_{k=1}^{N_v} V_k$$

22. Surface Area

$$A_i = \frac{1}{2} |a_i b_i \ \times \ a_i c_i| - - - -(1)$$

$$Surface\ Area\ (A) = \sum_{i=1}^{N_f} A_i \ ----(2)$$

Where,

$a_i b_i$ and $a_i c_i$ are edges of the ith triangle in the mesh, formed by vertices a_i, b_i and c_i.

23. Surface Area to Volume ratio

$$Surface\ Area\ to\ Volume\ ratio = \frac{A}{V}$$

24. Sphericity

$$Sphericity = \frac{\sqrt[3]{36\pi\ V^2}}{A}$$

25. Compactness 1

$$compactness\ 1 = \frac{V}{\sqrt{\pi\ A^3}}$$

26. Compactness 2

$$compactness\ 2 = 36\pi\ \frac{V^2}{A^3}$$

27. Spherical Disproportion

$$Spherical\ disproportion = \frac{A}{4\pi R^2} = \frac{A}{\sqrt[3]{36\pi\ V^2}}$$

28. Major Axis Length

$$major\ axis = 4\sqrt{\lambda_{major}}$$

29. Minor Axis Length

$$minor\ axis = 4\sqrt{\lambda_{minor}}$$

30. Least Axis Length

$$least\ axis = 4\sqrt{\lambda_{least}}$$

31. Elongation

$$elongation = \sqrt{\frac{\lambda_{minor}}{\lambda_{major}}}$$

32. Flatness

$$flatness = \sqrt{\frac{\lambda_{\text{least}}}{\lambda_{\text{major}}}}$$

C. Gray Level Co-occurrence Matrix (GLCM) Features

A GLCM of size $N_g{\times}N_g$ is defined as $P(i,j|\delta,\theta)$ describes the second-order joint probability function where $(i,j)th$ element represents the number of times the combination of intensity levels i and j occur in two pixels in the image separated by a distance of δ pixels along angle θ. Here N_g is the number of desecrate grey level intensity in the matrix. The co-occurrence features are calculated in 13 directions in 3 dimensions (3D) and each 3D GLCM feature is then calculated as mean of features calculated in all 13 directions.

Let:

- ϵ be an arbitrarily small positive number ($\approx 2.2{\times}10^{-16}$)

- $P(i,j)$ be the co-occurence matrix for an arbitrary δ and θ

- $p(i,j)$ be the normalized co-occurence matrix and equal to $\frac{P(i,j)}{\Sigma P(i,j)}$

- N_g be the number of discrete intensity levels in the image

- $p_x(i) = \sum_{j=1}^{Ng} P(i,j)$ be the marginal row probabilities

- $p_y(j) = \sum_{i=1}^{Ng} P(i,j)$ be the marginal column probabilities

- μ_x be the mean gray level intensity of p_x and defined $\mu_x = \sum_{i=1}^{Ng} p_x(i)i$

- μ_y be the mean gray level intensity of p_y and defined $\mu_y = \sum_{j=1}^{Ng} p_x(j)j$

- σ_x be the standard deviation of p_x

- σ_y be the standard deviation of p_y

- $p_{x+y}(k) = \sum_{i=1}^{Ng} \sum_{j=1}^{Ng} p(i,j)$ where $i+j=k$, and $k=2,3,...,2N_g$

- $p_{x-y}(k) = \sum_{i=1}^{Ng} \sum_{j=1}^{Ng} p(i,j)$ where $|i{-}j|=k$, and $k=0,1,...,N_g{-}1$

- $HX = -\sum_{i=1}^{Ng} p_x(i) \log_2(p_x(i)+\epsilon))$ be the entropy of p_x

- $HY = -\sum_{j=1}^{Ng} p_y(j) \log_2(p_y(j)+\epsilon))$ be the entropy of p_y

- $HXY = -\sum_{i=1}^{Ng} \sum_{j=1}^{Ng} p(i,j) \log_2(p_y(i,j)+\epsilon))$ be the entropy of $p(i,j)$

- $HXY1 = -\sum_{i=1}^{Ng} \sum_{j=1}^{Ng} p(i,j) \log_2(p_x(i)p_y(j) + \epsilon))$

- $HXY2 = -\sum_{i=1}^{Ng} \sum_{j=1}^{Ng} p_x(i) \, p_y(j) \log_2(p_x(i)p_y(j) + \epsilon))$

33. Autocorrelation

$$Autocorrelation = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j)ij$$

34. Joint Average

$$joint\ average = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j)i$$

35. Cluster Prominence

$$cluster\ prominence = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i + j - \mu_x - \mu_y)^4 p(i,j)$$

36. Cluster Shade

$$cluster\ shade = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i + j - \mu_x - \mu_y)^3 p(i,j)$$

37. Cluster Tendency

$$cluster\ tendency = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i + j - \mu_x - \mu_y)^2 p(i,j)$$

38. Contrast

$$contrast = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i - j)^2 p(i,j)$$

39. Correlation

$$correlation = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j)ij - \mu_x \mu_y}{\sigma_x(i)\sigma_y(j)}$$

40. Difference Average

$$difference\ average = \sum_{k=1}^{N_g-1} kp_{x-y}(k)$$

41. Difference Entropy

$$difference\ entropy = \sum_{k=1}^{N_g-1} p_{x-y}(k)\ \log_2(\ p_{x-y}(k) + \epsilon)$$

42. Difference Variance

$$difference\ Variance = \sum_{k=1}^{N_g-1} (k - DA)^2 p_{x-y}(k)$$

43. Joint Energy

$$joint\ energy = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (p(i,j))^2$$

44. Joint Entropy

$$joint\ entrpoy = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j) \log_2(p(i,j) + \epsilon)$$

45. Informational Measure of Correlation (IMC) 1

$$IMC1 = \frac{HXY - HXY1}{\max(HX, HY)}$$

46. Informational Measure of Correlation (IMC) 2

$$IMC2 = \sqrt{1 - e^{-2(HXY2-HXY)}}$$

47. Inverse Difference Moment (IDM)

$$IMD = \sum_{k=1}^{N_g-1} \frac{p_{x-y}(k)}{1 + k^2}$$

48. Maximal Correlation Coefficient (MCC)

$$MCC = \sqrt{\text{second largest eigenvalue of } Q}$$

$$Q(i,j) = \sum_{k=1}^{N_g-1} \frac{p(i,k)p(j,k)}{p_x(i)p_y(k)}$$

49. Inverse Difference Moment Normalized (IDMN)

$$IDMN = \sum_{k=1}^{N_g-1} \frac{p_{x-y}(k)}{1 + \left(\dfrac{k^2}{N_g^2}\right)}$$

50. Inverse Difference (ID)

$$ID = \sum_{k=1}^{N_g-1} \frac{p_{x-y}(k)}{1 + k}$$

51. Inverse Difference Normalized (IDN)

$$IDN = \sum_{k=1}^{N_g-1} \frac{p_{x-y}(k)}{1 + \dfrac{k}{N_g}}$$

52. Inverse Variance

$$inverse\ variance = \sum_{k=1}^{N_g-1} \frac{p_{x-y}(k)}{k^2}$$

53. Maximum Probability

$$maximum\ probability = \max(p(i,j))$$

54. Sum Average

$$sum\ average = \sum_{k=2}^{N_g-1} p_{x+y}(k)k$$

55. Sum Entropy

$$sum\ entropy = \sum_{k=2}^{N_g-1} p_{x+y}(k)\log_2(p_{x+y}(k) + \epsilon)$$

56. Sum of Squares

$$sum\ squares = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}(i - \mu_x)^2 p(i,j)$$

D. Gray Level Size Zone Matrix (GLSZM) Features

A GLSZM **P(i,j)** is the quantification of gray level zones in an image i.e., the number of connected voxels that share the same gray level intensity in the image, the **(i,j)ᵗʰ** element equals the number of zones with intensity level **i** and size **j** appear in image. A connected voxels can be defined as the voxel with the distance is 1 according to the

infinity norm in all 26 connections in 3D image. The GLSZM is rotation independent, with only one matrix calculated for all directions in the ROI.

Let:

N_g be the number of discreet intensity values in the image

N_s be the number of discreet zone sizes in the image

N_p be the number of voxels in the image

N_z be the number of zones in the ROI, which is equal to $\sum_{i=1}^{N_g} \sum_{j=1}^{N_s} P(i,j)$ and $1 \leq N_z \leq N_p$

P(i,j) be the size zone matrix

p(i,j) be the normalized size zone matrix, defined as $p(i,j) = \frac{P(i,j)}{N_z}$

57. Small Area Emphasis (SAE)

$$SAE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_s} \frac{P(i,j)}{j^2}}{N_z}$$

58. Large Area Emphasis (LAE)

$$LAE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_s} P(i,j)j^2}{N_z}$$

59. Gray Level Non-Uniformity (GLN)

$$GLN = \frac{\sum_{i=1}^{N_g} \left(\sum_{j=1}^{N_s} P(i,j) \right)^2}{N_z}$$

60. Gray Level Non-Uniformity Normalized (GLNN)

$$GLNN = \frac{\sum_{i=1}^{N_g} \left(\sum_{j=1}^{N_s} P(i,j) \right)^2}{N_z^2}$$

61. Size-Zone Non-Uniformity (SZN)

$$SZN = \frac{\sum_{j=1}^{N_s} \left(\sum_{i=1}^{N_g} P(i,j) \right)^2}{N_z}$$

62. Size-Zone Non-Uniformity Normalized (SZNN)

$$SZNN = \frac{\sum_{j=1}^{N_s} \left(\sum_{i=1}^{N_g} P(i,j) \right)^2}{N_z^2}$$

63. Zone Percentage (ZP)

$$ZP = \frac{N_z}{N_p}$$

64. Gray Level Variance (GLV)

$$GLV = \sum_{i=1}^{N_g} \sum_{j=1}^{N_s} p(i,j)(i - \mu)^2$$

Here

$$\mu = \sum_{i=1}^{N_g} \sum_{j=1}^{N_s} p(i,j)i$$

65. Zone Variance (ZV)

$$ZV = \sum_{i=1}^{N_g} \sum_{j=1}^{N_s} p(i,j)(j - \mu)^2$$

Here

$$\mu = \sum_{i=1}^{N_g} \sum_{j=1}^{N_s} p(i,j)j$$

66. Zone Entropy (ZE)

$$ZE = \sum_{i=1}^{N_g} \sum_{j=1}^{N_s} p(i,j)log_2(p(i,j) + \epsilon)$$

Here, ϵ is an arbitrarily small positive number ($\approx$2.2×10−16).

67. Low Gray Level Zone Emphasis (LGLZE)

$$LGLZE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_s} \frac{P(i,j)}{i^2}}{N_z}$$

68. High Gray Level Zone Emphasis (HGLZE)

$$HGLZE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_s} P(i,j)i^2}{N_z}$$

69. Small Area Low Gray Level Emphasis (SALGLE)

$$SALGLE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_s} \frac{P(i,j)}{i^2 j^2}}{N_z}$$

70. Small Area High Gray Level Emphasis (SAHGLE)

$$SAHGLE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_s} \frac{P(i,j)i^2}{j^2}}{N_z}$$

71. Large Area Low Gray Level Emphasis (LALGLE)

$$LALGLE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_s} \frac{P(i,j)j^2}{i^2}}{N_z}$$

72. Large Area High Gray Level Emphasis (LAHGLE)

$$LAHGLE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_s} P(i,j)i^2 j^2}{N_z}$$

E. Gray Level Run Length Matrix (GLRLM) Features

A GLRLM, $P(i,j|\theta)$ quantifies gray level runs, which are defined as the length in number of pixels, of consecutive pixels that have the same gray level value, the $(i,j)^{th}$ element describes the number of runs with gray level i and length j occur in the image (ROI) along angle θ. The gray level run length feature is calculated for each 13 direction in 3D image and 3D GLRLM feature is then calculated as mean of features calculated in all 13 directions.

Let:

N_g be the number of discreet intensity values in the image

N_r be the number of discreet run lengths in the image

N_p be the number of voxels in the image

$N_r(\theta)$ be the number of runs in the image along angle θ, which is equal to $\sum N_{gi=1} \sum N_{rj=1} P(i,j|\theta)$ and $1 \le N_r(\theta) \le N_p$

$P(i,j|\theta)$ be the run length matrix for an arbitrary direction θ

$p(i,j|\theta)$ be the normalized run length matrix, defined as $p(i,j|\theta)=P(i,j|\theta)N_r(\theta)$

73. Short Run Emphasis (SRE)

$$SRE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \frac{P(i,j|\Theta)}{j^2}}{N_r(\Theta)}$$

74. Long Run Emphasis (LRE)

$$LRE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} P(i,j|\Theta)j^2}{N_r(\Theta)}$$

75. Gray Level Non-Uniformity (GLN)

$$GLN = \frac{\sum_{i=1}^{N_g} \left(\sum_{j=1}^{N_r} P(i,j|\Theta)\right)^2}{N_r(\Theta)}$$

76. Gray Level Non-Uniformity Normalized (GLNN)

$$GLNN = \frac{\sum_{i=1}^{N_g} \left(\sum_{j=1}^{N_r} P(i,j|\Theta)\right)^2}{N_r(\Theta)^2}$$

77. Run Length Non-Uniformity (RLN)

$$RLN = \frac{\sum_{j=1}^{N_r} \left(\sum_{i=1}^{N_g} P(i,j|\Theta)\right)^2}{N_r(\Theta)}$$

78. Run Length Non-Uniformity Normalized (RLNN)

$$RLNN = \frac{\sum_{j=1}^{N_r} \left(\sum_{i=1}^{N_g} P(i,j|\Theta)\right)^2}{N_r(\Theta)^2}$$

79. Run Percentage (RP)

$$RP = \frac{N_r(\Theta)}{N_P}$$

80. Gray Level Variance (GLV)

$$GLV = \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} P(i,j|\Theta)(j - \mu)^2$$

Where,

$$\mu = \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} P(i,j|\Theta)j$$

81. Run Entropy (RE)

$$RE = \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} P(i,j|\Theta)log_2 \; (P(i,j|\Theta) + \epsilon)$$

Here, ϵ is an arbitrarily small positive number ($\approx$2.2×10^{-16}).

82. Low Gray Level Run Emphasis (LGLRE)

$$LGLRE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \frac{P(i,j|\Theta)}{i^2}}{N_r(\Theta)}$$

83. High Gray Level Run Emphasis (HGLRE)

$$HGLRE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} P(i,j|\Theta)i^2}{N_r(\Theta)}$$

84. Short Run Low Gray Level Emphasis (SRLGLE)

$$SRLGLE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \frac{P(i,j|\Theta)}{i^2 j^2}}{N_r(\Theta)}$$

85. Short Run High Gray Level Emphasis (SRHGLE)

$$SRHGLE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \frac{P(i,j|\Theta)i^2}{j^2}}{N_r(\Theta)}$$

86. Long Run Low Gray Level Emphasis (LRLGLE)

$$LRLGLE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \frac{P(i,j|\Theta)j^2}{i^2}}{N_r(\Theta)}$$

87. Long Run High Gray Level Emphasis (LRHGLE)

$$LRHGLE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} P(i,j|\Theta)i^2 j^2}{N_r(\Theta)}$$

F. Neighbouring Gray Tone Difference Matrix (NGTDM) Features

A NGTDM quantifies the difference between a gray value and the average gray value of its neighbours within distance δ. The sum of absolute differences for gray level i is stored in the matrix. Let X_{gl} be a set of segmented voxels and $x_{gl}(j_x, j_y, j_z) \in X_{gl}$ be the gray level of a voxel at position (j_x, j_y, j_z), then the average gray level of the neighbourhood is:

$$\bar{A}_i = \bar{A}\left(j_x, j_y, j_x\right) = \frac{1}{W} \sum_{k_x=-\delta}^{\delta} \sum_{k_y=-\delta}^{\delta} \sum_{k_z=-\delta}^{\delta} X_{gl}\left(j_x + k_x, j_y + k_y, j_z + k_z\right)$$

Where, $(k_x, k_y, k_z) \neq (0, 0, 0)$ and $x_{gl}(j_x+k_x, j_y+k_y, j_z+k_z) \in X_{gl}$

Let:

n_i be the number of voxels in X_{gl} with gray level i

$N_{v,p}$ be the total number of voxels in X_{gl} and equal to $\sum n_i$ (i.e. the number of voxels with a valid region; at least 1 neighbor). $N_{v,p} \leq Np$, where N_p is the total number of voxels in the ROI.

p_i be the gray level probability and equal to n_i/N_v

$$s_i = \begin{cases} \sum^{n_i} |i - \bar{A}_i| & for\ n_i \neq 0 \\ 0 & for\ n_i \neq 0 \end{cases} \text{ be the sum of absolute differences for gray level } i$$

N_g be the number of discreet gray levels

$N_{g,p}$ be the number of gray levels where $p_i \neq 0$

Where $p_i \neq 0,\ p_j \neq 0$

88. Coarseness

$$Coarseness = \frac{1}{\sum_{i=1}^{N_g} p_i s_i}$$

89. Contrast

$$Coarseness = \left(\frac{1}{N_{g,p}(N_{g,p} - 1)} \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p_i\, p_j (i - j)^2 \right) \left(\frac{1}{N_{v,p}} \sum_{i=1}^{N_g} s_i \right)$$

90. Busyness

$$Busyness = \frac{1}{N_{v,p}} \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} |i - j| \frac{p_i s_i + p_j s_j}{p_i + p_j}$$

91. Complexity

$$Complexity = \frac{\sum_{i=1}^{N_g} p_i s_i}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} |ip_i - jp_j|}$$

92. Strength

$$Strength = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (p_i + p_j)(i - j)^2}{\sum_{i=1}^{N_g} s_i}$$

G. Gray Level Dependence Matrix (GLDM) Features

A GLDM, quantifies gray level dependencies in an image and is defined as the number of connected voxels within distance δ that are dependent on the center voxel. In an GLDM, $P_{(i,j)}$ a neighbouring voxel with gray level j is considered dependent on center voxel with gray level i if $|i–j| \leq \alpha$. The $(i,j)^{th}$ element in GLDM, $P_{(i,j)}$ describes the number of times a voxel with gray level i with j dependent voxels in its neighbourhood appears in image within the ROI.

Let:

N_g be the number of discreet intensity values in the image

N_d be the number of discreet dependency sizes in the image

N_z be the number of dependency zones in the image, which is equal to $\sum_{i=1}^{N_g} \sum_{j=1}^{N_d} P(i,j)$

$P_{(i,j)}$ be the dependence matrix

$p_{(i,j)}$ be the normalized dependence matrix, defined as $p_{(i,j)} = P_{(i,j)} / N_z$

93. Small Dependence Emphasis (SDE)

$$SDE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_d} \frac{P(i,j)}{i^2}}{N_z}$$

94. Large Dependence Emphasis (LDE)

$$LDE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_d} P(i,j)i^2}{N_z}$$

95. Gray Level Non-Uniformity (GLN)

$$GLN = \frac{\sum_{i=1}^{N_g} \left(\sum_{j=1}^{N_d} P(i,j) \right)^2}{N_z}$$

96. Dependence Non-Uniformity (DN)

$$DN = \frac{\sum_{j=1}^{N_d} \left(\sum_{i=1}^{N_g} P(i,j) \right)^2}{N_z}$$

97. Dependence Non-Uniformity Normalized (DNN)

$$DNN = \frac{\sum_{j=1}^{N_d} \left(\sum_{i=1}^{N_g} P(i,j) \right)^2}{N_z{}^2}$$

98. Gray Level Variance (GLV)

$GLV = \sum_{i=1}^{N_g} \sum_{j=1}^{N_d} p(i,j)(i - \mu)^2$, **Where,** $\mu = \sum_{i=1}^{N_g} \sum_{j=1}^{N_d} ip(i,j)$

99. Dependence Variance (DV)

$$DV = \sum_{i=1}^{N_g} \sum_{j=1}^{N_d} p(i,j)(j-\mu)^2 \text{ , \textbf{Where,} } \mu = \sum_{i=1}^{N_g} \sum_{j=1}^{N_d} jp(i,j)$$

100. Dependence Entropy (DE)

$$DE = -\sum_{i=1}^{N_g} \sum_{j=1}^{N_d} p(i,j) \, log_2(p(i,j) + \epsilon)$$

101. Low Gray Level Emphasis (LGLE)

$$LGLE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_d} \frac{P(i,j)}{i^2}}{N_z}$$

102. High Gray Level Emphasis (HGLE)

$$HGLE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_d} P(i,j)i^2}{N_z}$$

103. Small Dependence Low Gray Level Emphasis (SDLGLE)

$$SDLGLE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_d} \frac{P(i,j)}{i^2 j^2}}{N_z}$$

104. Large Dependence Low Gray Level Emphasis (LDLGLE)

$$LDLGLE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_d} \frac{P(i,j)j^2}{i^2}}{N_z}$$

105. Large Dependence High Gray Level Emphasis (LDHGLE)

$$LDHGLE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_d} P(i,j)i^2 j^2}{N_z}$$

Wavelet features:

Wavelet features can be extracted from wavelet transformed images. A discrete, one-level and undecimated three dimensional wavelet transform is applied to each three dimensional medical mages (CT, PET, MRI, SPECT etc.), which decomposes the original image into 8 decompositions. The resultant decomposed images based of application of a combination of a low-pass and a high-pass filters functions in three dimension can be labeled as X_{LLL}, X_{LLH}, X_{LHL}, X_{LHH}, X_{HLL}, X_{HLH}, X_{HHL}, and X_{HHH}. For example, X_{LHH} is

filtering of image with a low-pass filter along x-direction, a high-pass filter along y-direction and a high-pass filter along z-direction and is constructed as:

$$X_{LHH}(i,j,k) = \sum_{p=1}^{N_L}\sum_{q=1}^{N_H}\sum_{r=1}^{N_H} L(p)H(q)H(r)X(i+p,j+q,k+r)$$

Where,

N_L is the length of filter L

N_H is the length of filter H

X is the original 3D image

The other decompositions can be constructed similarly by applying their respective ordering of low or high-pass filtering in x, y and z-direction. The schematic representation of Wavelet decomposition of the image is shown in Supplementary Figure 1.
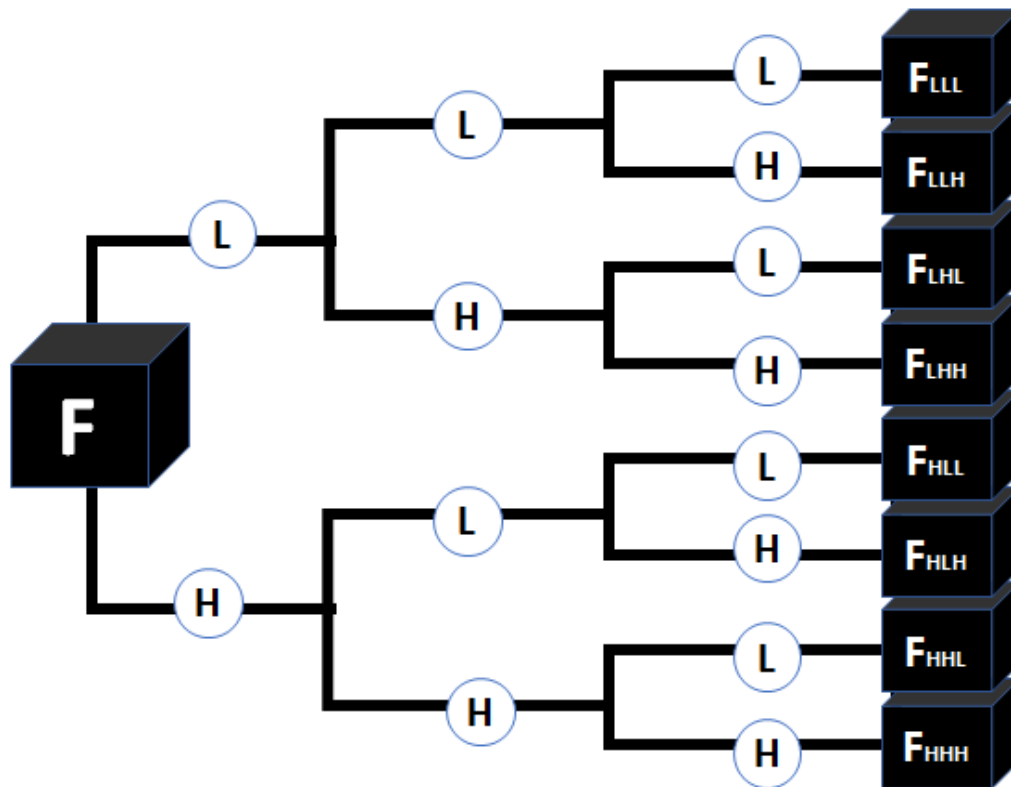
Supplementary Figure 1: Figure shows the schematic representation of eight Wavelet decomposition applied on image for radiomic extraction.

Laplacian of Gaussian (LoG) Features:

Laplacian filters are derivative filters edge detection filter used to find areas of rapid change of intensity in images. Since Laplacian filters are very sensitive to noise, it is

common to smooth the image using smoothing filters like a Gaussian filter before applying the Laplacian on the images. This two-step process of filtering the image is called the Laplacian of Gaussian (LoG) operation. Following equation can be the combined express of LoG filter.

$$LoG(i,j) = \frac{1}{\pi\sigma^4}\left[1 - \frac{i^2 + j^2}{2\sigma^2}\right]e^{-\frac{i^2+j^2}{2\sigma^2}}$$

We usually use three LOG filter using σ value *1, 2,* and *3* and generate *3* sets of *LoG* images and we extract all the above described features from these images except shape based features.

The pyRadiomic feature equivalence in Image biomarker standardisation initiative (IBSI) and IBSI feature code is given in supplementary table 1 [1,2].

no.	pyRadiomics features		IBSI radiomic features		
	class	feature name	class	index	feature name
1		10Percentile		QG58	10th intensity percentile
2		90Percentile		8DWT	90th intensity percentile
3		Energy		N8CA	Intensity-based energy
4		InterquartileRange		SAL0	Intensity interquartile range
5		Kurtosis		IPH6	Intensity Kurtosis
6		Maximum		84IY	Maximum intensity
7		MeanAbsoluteDeviation		4FUA	Intensity-based mean absolute deviation
8		Mean		Q4LE	Mean intensity
9		Median		Y12H	Median intensity
10		Minimum		1GSF	Minimum intensity
11	First Order	Range		20JQ	Intensity range
12		RobustMeanAbsoluteDeviation		1128	Intensity-based robust mean absolute deviation

13		RootMeanSquared		5ZWQ	Root mean square intensity
14		Skewness		KE2A	Intensity Skewness
15		TotalEnergy			*not present in IBSI feature definitions*
16		Variance		ECT3	Intensity variance
17		Entropy	Intensity histogram	TLU2	Discretised intensity entropy/ Intensity Histogram Entropy
18		Uniformity		BJ5W	Discretised intensity uniformity
1	Shape Features (3D)	MeshVolume	Morphological features	RNU0	Volume (mesh)
2		VoxelVolume		YEKZ	Volume (voxel counting)
3		SurfaceArea		CoJK	Surface area (mesh)
4		SurfaceVolumeRatio		2PR5	Surface to volume ratio
5		Compactness1		SKGS	Compactness 1
6		Compactness2		BQWJ	Compactness 2
7		SphericalDisproportion		KRCK	Spherical disproportion
8		Sphericity		QCFX	Sphericity
9		Maximum3DDiameter		L0JK	Maximum 3D diameter
10		MajorAxisLength		TDIC	Major axis length
11		MinorAxisLength		P9VJ	Minor axis length
12		LeastAxisLength		7J51	Least axis length
13		Elongation		Q3CK	Elongation
14		Flatness		N17B	Flatness
1	GLCM	Autocorrelation	GLCM	QWB0	Autocorrelation
2		ClusterProminence		AE86	Cluster Prominence
3		ClusterShade		7NFM	Cluster Shade

4		ClusterTendency		DG8W	Cluster Tendency
5		Contrast		ACUI	Contrast
6		Correlation		NI2N	Correlation
7		DifferenceAverage		TF7R	Difference Average
8		DifferenceEntropy		NTRS	Difference Entropy
9		DifferenceVariance		D3YU	Difference Variance
10		Inverse Difference(ID)		IB1Z	Inverse Difference
11		Inverse Difference Moment (IDM)		WF0Z	Inverse difference moment
12		Inverse Difference Moment Normalized (IDMN)		1QCO	Normalised Inverse difference moment
13		Inverse Difference Normalized (IDN)		NDRX	Normalised Inverse Difference
14		Informational Measure of Correlation (IMC) 1		R8DG	Information Correlation 1
15		Informational Measure of Correlation (IMC) 2		JN9H	Information Correlation 2
16		InverseVariance		E8JP	Inverse Variance
17		JointAverage		6oVM	Joint Average
18		JointEnergy		8ZQL	Angular Second Moment
19		JointEntropy		TU9B	Joint Entropy
20		Maximal Correlation Coefficient (MCC)			*not present in IBSI feature definitions*
21		MaximumProbability		GYBY	Joint Maximum
22		SumAverage		ZGXS	Sum Average
23		SumEntropy		P6QZ	Sum Entropy
24		SumSquares		UR99	Joint Variance
1	GLSZM	GrayLevelNonUniformity	GLSZM	JNSA	Grey Level Non-Uniformity

2		GrayLevelNonUniformityNormalized		Y1RO	Normalised Grey Level Non-Uniformity
3		GrayLevelVariance		BYLV	Grey Level Variance
4		HighGrayLevelZoneEmphasis		5GN9	High Grey Level Zone Emphasis
5		LargeAreaEmphasis		48P8	Large Zone Emphasis
6		LargeAreaHighGrayLevelEmphasis		J17V	Large Zone High Grey Level Emphasis
7		LargeAreaLowGrayLevelEmphasis		YH51	Large Zone Low Grey Level Emphasis
8		LowGrayLevelZoneEmphasis		XMSY	Low Grey Level Zone Emphasis
9		SizeZoneNonUniformity		4JP3	Zone size Non-Uniformity
10		SizeZoneNonUniformityNormalized		VB3A	Normalised Zone Size Non-Uniformity
11		SmallAreaEmphasis		5QRC	Small Zone Emphasis
12		SmallAreaHighGrayLevelEmphasis		HW1V	Small Zone High Grey Level Emphasis
13		SmallAreaLowGrayLevelEmphasis		5RAI	Small Zone Low Grey Level Emphasis
14		ZoneEntropy		GU8N	Zone Size Entropy
15		ZonePercentage		P3OP	Zone Percentage
16		ZoneVariance		3NSA	Zone Size Variance
1	GLRLM	GrayLevelNonUniformity	GLRLM	R5YN	Grey Level Non-uniformity
2		GrayLevelNonUniformityNormalized		OVBL	Normalised Grey Level Non-Uniformity
3		GrayLevelVariance		8CE5	Grey Level Variance
4		HighGrayLevelRunEmphasis		G3QZ	High Grey Level Run Emphasis
5		LongRunEmphasis		W4KF	Long Run Emphasis

6		LongRunHighGrayLevelEmphasis		3KUM	Long Run High Grey Level Emphasis
7		LongRunLowGrayLevelEmphasis		IVPO	Long Run Low Grey Level Emphasis
8		LowGrayLevelRunEmphasis		V3SW	Low Grey Level Run Emphasis
9		RunEntropy		HJ9O	Run Entropy
10		RunLengthNonUniformity		W92Y	Run Length Non-Uniformity
11		RunLengthNonUniformityNormalized		IC23	Normalised Run Length Non-Uniformity
12		RunPercentage		9ZK5	Run Percentage
13		RunVariance		SXLW	Run Length Variance
14		ShortRunEmphasis		22OV	Short Run Emphasis
15		ShortRunHighGrayLevelEmphasis		GD3A	Short Run High Grey Level Emphasis
16		ShortRunLowGrayLevelEmphasis		HTZT	Short Run Low Grey Level Emphasis
1	NGTDM	Busyness	NGTDM	NQ30	Busyness
2		Coarseness		QCDE	Coarseness
3		Complexity		HDEZ	Complexity
4		Contrast		65HE	Contrast
5		Strength		1X9X	Strength
1	GLDM	DependenceEntropy	NGLDM	FCBV	Dependence Count Entropy
2		DependenceNonUniformity		Z87G	Dependence Count Non-Uniformity
3		DependenceNonUniformityNormalized		OKJI	Normalised Dependence Count Non-Uniformity
4		DependenceVariance		DNX2	Dependence Count Variance
5		GrayLevelNonUniformity		FP8K	Grey Level Non-Uniformity
6		GrayLevelVariance		1PFV	Grey Level Variance

7		HighGrayLevelEmphasis		OAE7	High Grey Level Count Emphasis
8		LargeDependenceEmphasis		IMOQ	High Dependence Emphasis
9		LargeDependenceHighGrayLevelEmphasis		9QMG	High Dependence High Grey Level Emphasis
10		LargeDependenceLowGrayLevelEmphasis		NBZI	High Dependence Low Grey Level Emphasis
11		LowGrayLevelEmphasis		TL9H	Low Grey Level Count Emphasis
12		SmallDependenceEmphasis		SODN	Low Dependence Emphasis
13		SmallDependenceHighGrayLevelEmphasis		JA6D	Low Dependence High Grey Level Emphasis
14		SmallDependenceLowGrayLevelEmphasis		EQ3F	Low Dependence Low Grey Level Emphasis

GLCM= Gray Level Co-occurrence Matrix, GLDM: Gray Level Dependence Matrix, NGLDM: Neighboring Grey Level Dependence Matrix, GLRLM: Gray Level Run Length Matrix, GLSZM: Gray Level Size Zone Matrix, NGTDM: Neighborhood Gray Tone Difference Matrix

Supplementary Table 1: The list of 93 radiomics features with the IBSI standardization feature names.

References:

[1] Radiomic Features, https://pyradiomics.readthedocs.io/en/latest/features.html last accessed on 30-07-2021

[2] Zwanenburg, A., Leger, S., Vallières, M., and Löck, S. (2016). Image biomarker standardisation initiative - feature definitions. In eprint arXiv:1612.07003

Chapter 5: Repeatability and reproducibility study of radiomic features on a phantom and human cohort

Abstract

The repeatability and reproducibility of radiomic features extracted from CT scans need to be investigated to evaluate the temporal stability of imaging features with respect to a controlled scenario (test–retest), as well as their dependence on acquisition parameters such as slice thickness, or tube current. Only robust and stable features should be used in prognostication/prediction models to improve generalizability across multiple institutions. In this study, we investigated the repeatability and reproducibility of radiomic features with respect to three different scanners, variable slice thickness, tube current, and use of intravenous (IV) contrast medium, combining phantom studies and human subjects with non-small cell lung cancer. In all, half of the radiomic features showed good repeatability (ICC > 0.9) independent of scanner model. Within acquisition protocols, changes in slice thickness was associated with poorer reproducibility compared to the use of IV contrast. Broad feature classes exhibit different behaviors, with only few features appearing to be the most stable. 108 features presented both good repeatability and reproducibility in all the experiments, most of them being wavelet and Laplacian of Gaussian features.

Introduction

Copy (and adapt if necessary) from paper Medical images are routinely used for cancer staging, treatment planning and evaluation. Radiological findings are mainly evaluated in a qualitative or semi-qualitative fashion guided predominantly by visual inspection[1]. However, human interpretation of images is open to subjectivity and potentially misses some of the quantitative and objective information that could otherwise be retrieved from patients' scans through computer-assisted methods[2].

The field of "radiomics" aims to address the above-mentioned issues by objectively quantifying visual information in the images as a vast set of numerical metrics known as "features". Radiomics hypothesizes that a certain subset of features, analyzed with the aid of machine learning algorithms due to high dimensionality, may have some predictive/prognostic value. Such subsets of features denote a "signature", i.e. a digital image phenotype of the target disease, which opens the way towards personalized treatment in oncology[3].

One of the most challenging problems for translating radiomic studies into clinical decision support systems is to evaluate the robustness of radiomic-based models and hence their potential generalizability across multiple datasets from different institutions[4]. Different institutions commonly acquire scans with different settings (e.g., scanner manufacturers, slice thickness, signal-to-noise ratio) according to largely self-defined imaging protocols, which add unwanted variation in the resulting radiomic features that are not related to the disease phenotype. A feature that is useful on one dataset may therefore lose its value on another dataset, since the feature may be sensitive to different methods of acquisition[5].

When discussing robustness of radiomic studies two concepts need to be considered: "repeatability" and "reproducibility". Repeatability refers to features that remain the same when imaged multiple times in the same subject, be that a human or a suitable phantom, using the same image acquisition methods. Reproducibility refers to features that remain the same when extracted using different equipment, different software, different image acquisition settings, or different operators (e.g. other clinics), be that in the same subject or in different subjects[6]. Repeatability and reproducibility concerns have been raised as major source of uncertainties in radiomic models[7].

Most of the studies that investigated the reproducibility of radiomic features with respect to different image acquisition settings, demonstrate a strong dependence of radiomic features on such settings. Texture features appear to be more vulnerable to reproducibility/repeatability issues. There is a strong connection between reproducibility/repeatability and prognostic values[8]. In a study about time series classification, the investigators concluded that poorly reproducible/repeatable features were usually accompanied by poor discriminative performances[9].

Recent publications have also investigated the presence of correlations between radiomic features and tumor volume[9,10]. The latter has been shown to be one of the most generalizable features. Therefore, there is the need to investigate if the most reproducible features were also strongly correlated with tumor volume.

Several studies have investigated the repeatability/ reproducibility of radiomic features on phantom as well as well as clinical cohort[6,7,8,9,10,11,12,13,14,15,16,17]. Few publications have also investigated the disease specific dependency of radiomic feature repeatability/ reproducibility and presented the results. These studies have either performed repeatability or reproducibility study alone; or performed repeatability and reproducibility study only on phantoms[13,14] or clinical[15,16] cohorts, which (1) limits the possibility to isolate a subset of features that are both repeatable and reproducible, and (2) does not allow comparing differences in the results because of using only phantom or human data. There remains a need to evaluate reproducibility and repeatability of radiomic features, not only on phantoms datasets, but also on human cohorts in the same study. The risk is that phantom studies do not have sufficiently high complexity and heterogeneity within the synthetic "tumors" to be a fair test of feature robustness. In our study, stable feature refers to both repeatable and reproducible features at the same time. With our study, we provide an extension to currently available literature by performing a comprehensive evaluation of the reproducibility and repeatability of 1080 radiomic features considering not only different groups of features, but also features extracted using digital filtering both with phantoms and human data. In this study, we also investigated how the correlations between radiomic features and tumor volume impact the reproducibility and repeatability results.

Methods

This study was approved by the hospital Institutional Ethics Committee (Institutional Ethics Committee-I, Tata Memorial Centre [IEC, TMC], Mumbai, India) as a retrospective study, with waivers of informed consent from involved patients as per IEC policy of our hospital by the same Ethics Committee. All methods were carried out in accordance with relevant guidelines and regulations. This study comprises PET/CT images from a polymer phantom as well as from a clinical cohort. Our study has focused only on CT radiomic features stability. PET images were used to delineate the tumor (using SUV threshold of 40%) and this delineation was transferred to the corresponding CT images included in this study.

Phantom

The National Electrical Manufacturers Association (NEMA) Image Quality (IQ)PET/CT phantom (Data Spectrum Inc., NJ, USA) was used for this study[22]. The external dimensions of the phantom are 241 mm × 305 mm × 241 mm with interior length of 180 mm and volume of 9.7L. It has six fillable spheres and one central cylinder. The largest insert with a diameter of 37 mm was used for radiomic feature analysis study. The phantom was filled with distilled water containing 18F-FDG. The concentration of 18F-FDG was adjusted until a target to background signal ratio of 4:1 was created between the active sphere and water background.

Clinical cohort

Patients with non-small cell lung cancer (NSCLC) (n = 104) who underwent pre-treatment PET/CT scans in our department were included in this study. There were 85 males and 19 females. The median age was 66 (36–90) and 53 (35–72) years respectively for males and females. The median tumor volume was 92 (14–486) cm³ for men and 86 (22–432) cm³ for women. Population demographics and clinical information are provided in Supplementary table S2.

RIDER: The Reference Image Database to Evaluate Therapy Response (RIDER) data base was used in this study to perform repeatability study. All the 32 patients DICOM data (i.e. Images and RTSTRUCTs) of the RIDER data set were included in this study[23].

Scanners

Three different scanners were used in the study. Two scanners were from the same manufacturer (Philips Medical, Eindhoven, The Netherlands) but different models, and the last scanner was from another manufacturer (General Electric Medical System, Milwaukee, USA). For simplicity of reading we will refer to the scanners as follows: scanner 1 is the Philips Gemini TF16 PET/CT, scanner 2 is the Gemini TF64 PET/CT, and scanner 3 is the General Electric Discovery NM 670 pro SPECT/CT.

Scanning protocols

NEMA IQ phantom

The NEMA IQ phantom was scanned twice, 30 min apart ('coffee break') without repositioning, one the same scanner and within the same conditions. This procedure was performed for all the three scanners and considering six different acquisition protocols. They had the same tube voltage (120 kV for all three scanners), pitch (0.46 for scanner 1 and 2 and 2.5 for Scanner 3) and reconstruction kernel based on filtered back projection for scanner 1, 2 and adaptive statistical iterative reconstruction (ASiR) (40% ASiR setting and a noise index of 13.75) for scanner 3, but different tube currents (ranging from 100 to 300 mA) slice thicknesses (ranging from 2 to 5 mm for scanner 1& 2 and 2.5 to 5 for scanner 3). These protocols are listed in Table 2.

Table 2 Overview of the scanning protocols used to acquire images with the IQ phantom.

Protocol name	Tube current (mA)	Reconstruction slice thickness (mm) [voxel size (cubic millimeter)] (scanner 1&2)	Reconstruction slice thickness (mm) [voxel size (cubic millimeter)] (scanner 3)
Protocol 1	100	2 [0.86 × 0.86 × 2]	2.5 [0.9653 × 0.9653 × 2.5]
Protocol 2	100	5 [0.86 × 0.86 × 5]	5 [0.9653 × 0.9653 × 5]
Protocol 3	200	2 [0.86 × 0.86 × 2]	2.5 [0.9653 × 0.9653 × 2.5]
Protocol 4	200	5 [0.86 × 0.86 × 5]	5 [0.9653 × 0.9653 × 5]
Protocol 5	300	2 [0.86 × 0.86 × 2]	2.5 [0.9653 × 0.9653 × 2.5]
Protocol 6	300	5 [0.86 × 0.86 × 5]	5 [0.9653 × 0.9653 × 5]

Six scanning protocols, with same tube voltage (120 kV), pitch (Scanner 1&2: 0.46; Scanner 3: 2.5), and reconstruction kernel, but different tube currents and slice thicknesses were investigated. The phantom was scanned twice on scanners 1–2–3 without repositioning in a 30-min test–retest scenario. The total number of scans acquired with the IQ phantom is 6 protocols × 3 scanners × 2 (test–retest) = 36 scans.

Clinical cohort

Patients were scanned using three different clinical protocols on the Philips Gemini TF64 PET/CT (previously referred to as scanner 2). The three protocols had the same tube voltage (120 kV), pitch (0.46) and reconstruction kernel, but different slice thicknesses, tube current and presence or absence of an intravenous contrast medium, namely, one whole body contrast CT with 2 mm slice thickness (referred as WBCECT2), one whole body contrast CT with 5 mm slice thickness (referred as BLDCT5), and one non contrast thoracic CT with 2 mm slice thickness (referred as NCCTT2). Modulated tube current (between 100 and 200 mA) as per dose care automated system was used for BLDCT5 and WBCECT2. The protocols are listed in Table 3.

Table 3 Overview of the clinical protocols. Images were acquired on the Philips Gemini TF64 PET/CT (previously referred to as scanner 2) with three different protocols.

Clinical protocol name	Slice thickness (mm)	Intravenous contrast medium	Tube current (mA)	Voxel size (cubic millimeters)
BLDCT5	5	Yes—nonionic contrast	Modulated auto-mA (100–200)	$1.17 \times 1.17 \times 5$
WBCECT2	2	Yes—nonionic contrast	Modulated auto-mA (100–200)	$1.17 \times 1.17 \times 2$
NCCTT2	2	NO	Fixed mA 300	$0.87 \times 0.87 \times 2$

RIDER

The RIDER data set comprises of 32 NSCLC patient's test–retest CT imaging performed with a time lag of 15 min and two sets of delineations (RTSTRUCT) (i.e. tumor delineated by manual and automatic methods). Imaging parameters of RIDER database is summarized in Table 4. Radiomic extraction and statistical analysis was performed as per the study protocol.

Table 4 The imaging protocol of the RIDER data set.

Parameters	Rider data set
Manufacturer	GE healthcare
Acquisition type	Helical
Tube voltage	120 kVp
Tube current	Range 165–549 mAs
Slice thickness	1.25 mm
Pixels	512×512
Voxel size (cubic millimeter)	$0.66 \times 0.66 \times 1.25$

Study design

In this study we investigated both reproducibility and repeatability of radiomic features. The repeatability of radiomic features was evaluated using the test retest scans acquired with the IQ phantom on three different scanners and for all the 6 protocols listed in Table 2 and on the publicly available clinical cohort RIDER data set. The reproducibility of radiomic features with respect to different acquisition protocols but within the same scanner (intra-scanner variability) was evaluated comparing radiomic feature values using the test scans acquired with the IQ phantom across the 6 different protocols. This analysis was repeated for all the three scanners. The reproducibility of radiomic features with respect to different scanner models was evaluated comparing radiomic feature values extracted from the test scans acquired with the IQ phantom for each protocols on the three different scanners (inter-scanner variability). The reproducibility of radiomic features with respect to presence/absence of intravenous contrast medium

and slice thickness in clinical data was investigated comparing radiomic features using the images acquired with the NSCLC patients (clinical study). Figure 5 summarizes the overall study design.
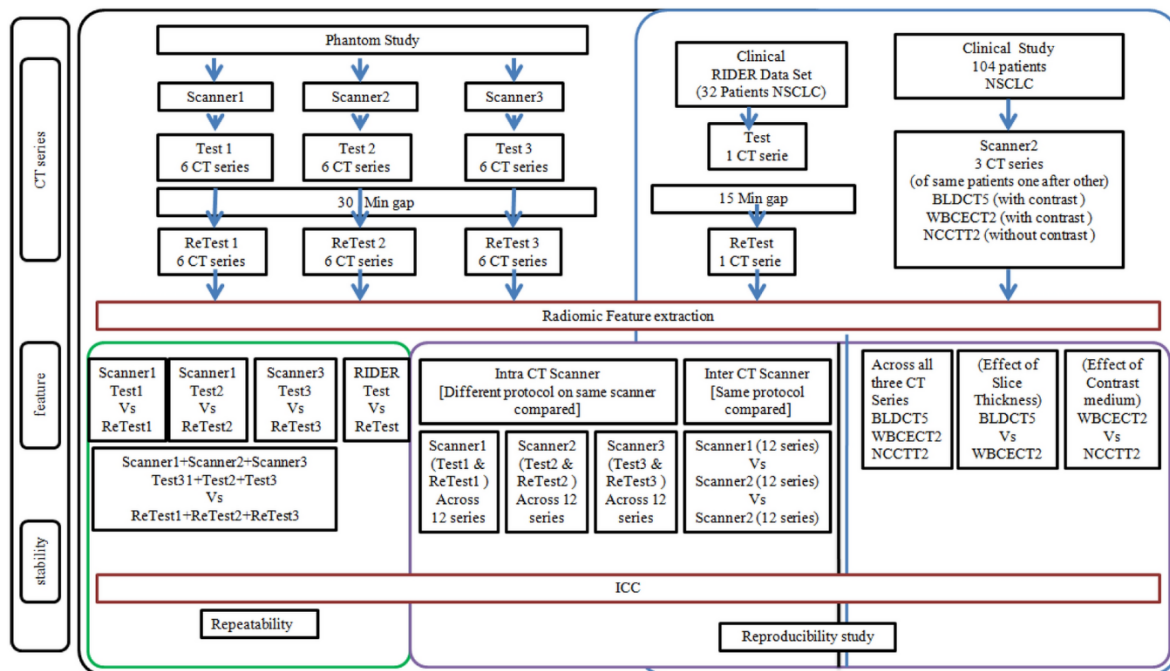
Figure 5: In this study we investigated both reproducibility and repeatability of radiomic features. The repeatability of radiomic features was evaluated using the test retest scans acquired with the IQ phantom on three different scanners and with 6 protocols and online available RIDER data set. The reproducibility of radiomic features with respect to different acquisition protocols but within the same scanner (intra-scanner variability) was evaluated comparing radiomic feature values using the test–retest scans acquired with the IQ phantom across the 6 different protocols. A clinical cohort of NSCLC patients was used to investigate the reproducibility of radiomic features with respect to 3 different clinical acquisition protocols, with a focus on the impact of slice thickness and IV contrast medium.

Results

Copy (and adapt if necessary) from paper Phantom—repeatability

The percentage of radiomic features presenting good repeatability (ICC ≥ 0.9) were 58% (624/1080) for scanner1 (Philips Gemini TF16), 43% (464/1080) for scanner2 (Philips Gemini TF64), 61% (661/1080) for scanner3 (GE Discovery NM 570) and 45% (488/1080) for the three scanners overall. Results are shown in Fig. 1 for each feature category.
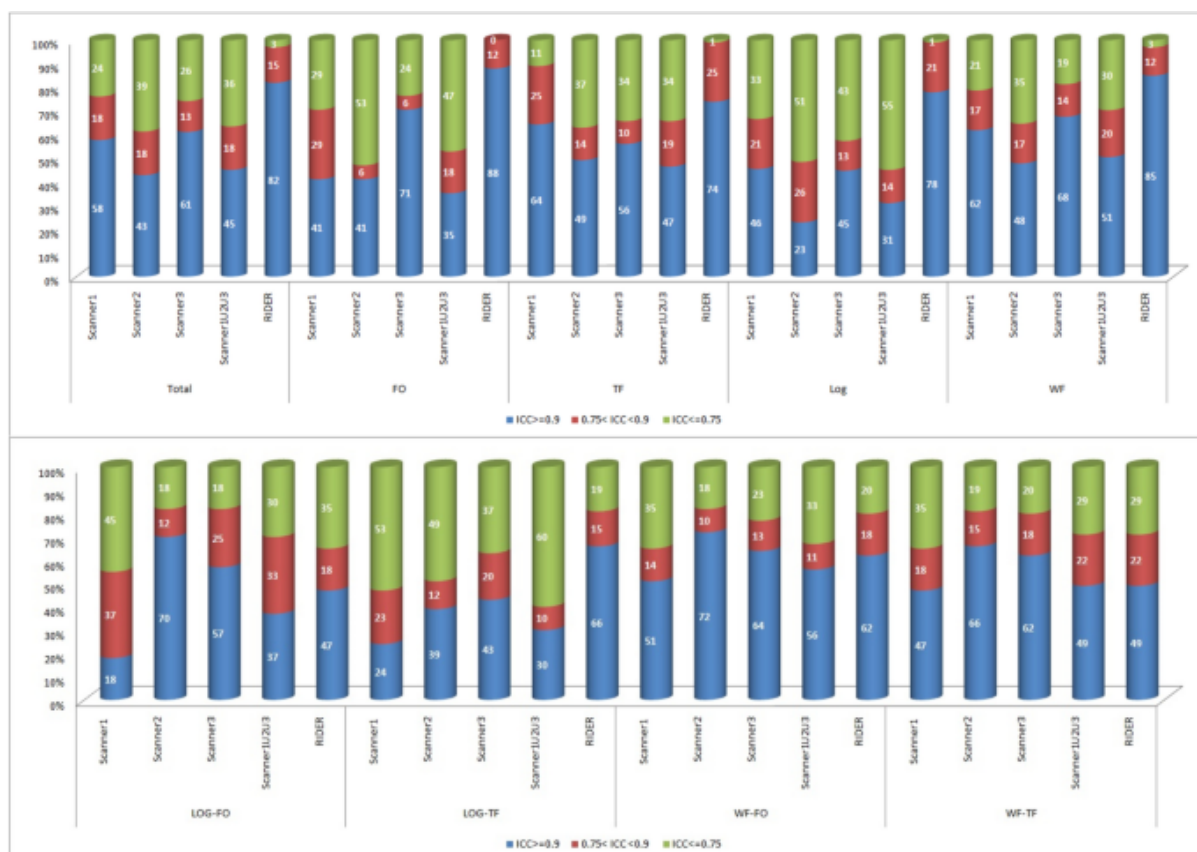
Figure 1: Repeatability analysis using repeated phantom scans for all the different radiomic feature classes. The median ICC values for all the 6 protocols is reported, separately for scanner1, scanner2, scanner3 and the union of the three. Repeatability analysis on RIDER (clinical cohort) was also performed. Three different levels of repeatability are defined: good (ICC ≥ 0.9), medium (0.75 < ICC < 0.9), and poor (ICC ≤ 0.75) (FO = First Order Feature; TF = Textural Feature; LOG = Total LoG Feature; WF = Total Wavelet Feature; LOG-FO = LoG First Order Feature; LOG-TF = LoG Textural Feature; WF-FO = Wavelet First Order Feature; WF-TF = Wavelet Textural Feature).

RIDER (clinical cohort)—repeatability

The percentage of radiomic features presenting good, moderate, and poor repeatability were 82% (888/1080), 15% (164/1080), and 3% (28/1080) respectively for the RIDER clinical cohort. The results per feature categories are shown in Fig. 1.

Phantom—reproducibility—intra and inter scanner variability

For the intra-scanner study, 30% (322/1080), 31% (332/1080) and 39% (426/1080) features presented good, moderate, and poor reproducibility (Fig. 2A) for all the scanners. For the inter-CT scanner study, 14% (154/1080), 19% (204/1080) and 67% (722/1080) features presented good, moderate, and poor reproducibility respectively (Fig. 2B) for all the 6 protocols. Reproducibility of the features individually for the six protocols are shown in the Supplementary material S1.
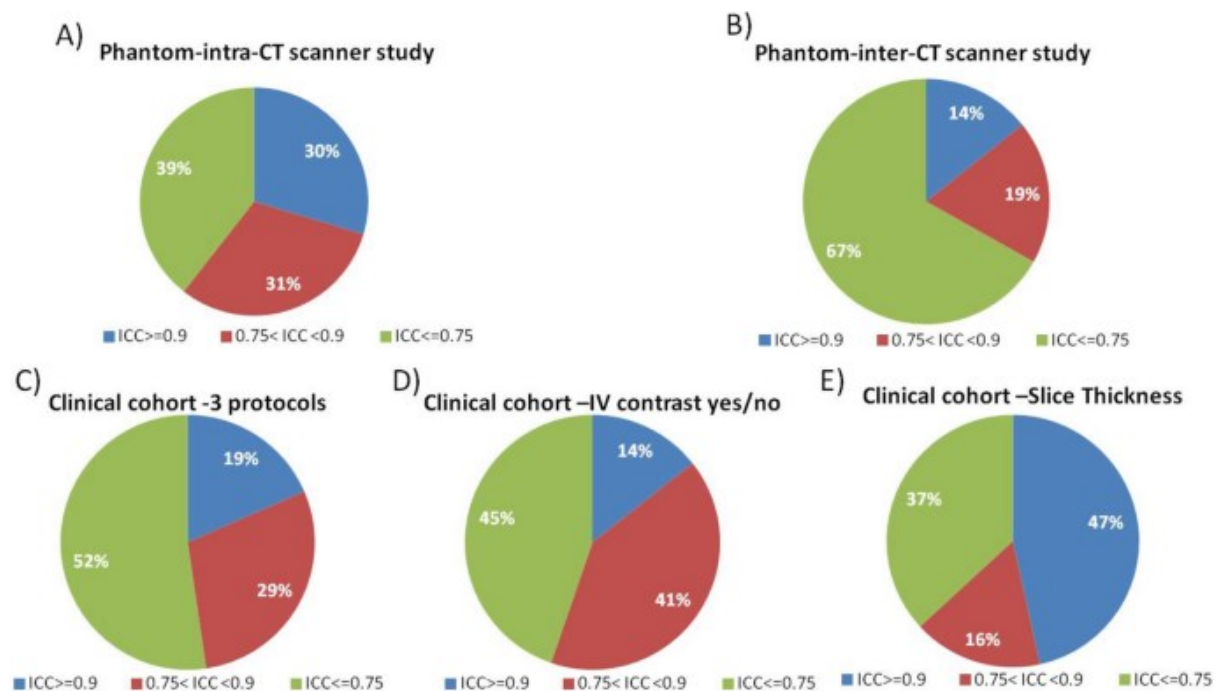
Figure 2: Results of the reproducibility experiments: (**A**) intra-scanner experiment using the phantom. By taking the median of all ICCs computed on the three scanners; (**B**) inter-scanner experiment using the phantom across all three scanners; (**C**) stability of radiomic features with respect to three different clinical protocols in the clinical study; (**D**) impact of IV (intravenous) contrast medium presence (WBCECT2)/ absence (NCCTT2) and difference in current (WBCECT2:Auto mA = 100–200; NCCTT2: fixed mA = 300) in the clinical study, and E) impact of slice thickness (2 vs 5 mm) in the human study. Three different levels of reproducibility are defined: good (ICC ≥ 0.9), medium (0.75 < ICC < 0.9), and poor (ICC ≤ 0.75).

Clinical cohort—reproducibility

Among the features tested, 19% (199/1080) good, 29% (315/1080), moderate and 52% (556/1080) had poor, reproducibility when comparing the 3 different imaging protocols on the Gemini TF16 scanner (Fig. 2C).

When comparing IV contrast (WBCECT2) versus non-contrast (NCCTT2) protocols, 45% (483/1080) of the features had poor, 41% (442/1080) moderate, and 14% (155/1080) good reproducibility (Fig. 2D). When comparing slice thickness, using the BLDCT5

protocol (slice thickness = 5 mm) versus the WBCECT2 protocol (slice thickness = 2 mm), 37% (398/1080) of the features had poor, 17% (179/1080) moderate, and 47% (503/1080) good reproducibility (Fig. 2E).

Volume correlations

In the clinical cohort, 7% (73/1080), 5% (57/1080) and 88% (950/1080) of the radiomic features had good ($\rho \geq 0.9$), moderate ($0.75 < \rho < 0.9$) and poor ($\rho \leq 0.75$) correlation with the GTV.

Overall summary

Median ICC was calculated for all the reproducibility studies performed using the phantom and clinical cohorts. A total of 22.5% (243/1080) features had good reproducibility (ICC > 0.9) in clinical cohort. When the median of ICC was calculated for repeatability study performed with phantom and clinical cohorts (RIDER); 46.1% (498/1080) of features had good repeatability (ICC > 0.9). For repeatability study on phantom and clinical cohort together 55% (599/1080) features had good stability (ICC > 0.9) (Fig. 3A). For reproducibility study on phantom and clinical cohort together 15% (164/1080) features had good stability (ICC > 0.9) (Fig. 3B). For repeatability and reproducibility study together on clinical cohort 18% (189/1080) features had good stability (Fig. 3C). For all the experiments, 13% (138/1080) of the features presented both high (median ICC > 0.9) repeatability and high reproducibility (Fig. 3D). Tumor volume was again confirmed to be the most repeatable and reproducible feature with a median ICC of 0.99. When considering volume collinearity, 21% of these stable features presented strong Spearman correlations ($\rho > 0.9$). If we removed the features with strong correlations with GTV, then the final number of repeatable and reproducible features was 108: 59 WF (Wavelet) (8% of total WF), 46 LOG (Laplacian of Gaussian) (17% of total LOG), and 3 TA (Texture Analysis) (3% of total TA) features (Table 1). Overall, TA had the largest median ICC (0.933 ± 0.024) followed by LOG (0.923 ± 0.017) and WF (0.917 ± 0.014) features ($p < 0.05$). The topmost robust feature per feature types were: GLRLM-Non-Uniformity (LOG-2 mm kernel); LLH-GLCM-JointEnergy (WF) and Gray Level Dependence Matrix (GLDM) Non-Uniformity (TA).
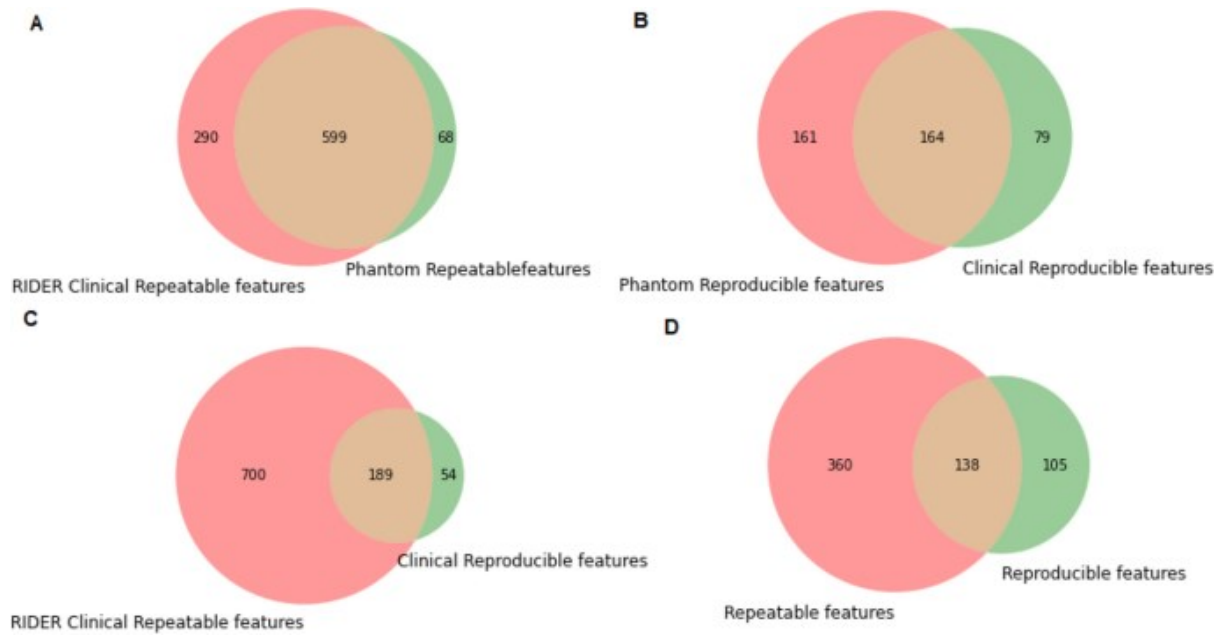
Figure 3: Common features in various studies showing good stability (ICC > 0.9): (**A**) Venn diagram shows the overlap of repeatability (RIDER) study and Phantom repeatability study. (**B**) Venn diagram shows the overlap of Phantom reproducibility study and reproducibility study in clinical cohort. (**C**) Venn diagram shows the overlap of repeatability (RIDER) study and reproducibility study in clinical cohort. (**D**) Overall summary of all the experiments. The Venn diagram shows the overlap of the repeatability experiment (phantom + clinical [RIDER] data) with the reproducibility experiments (phantom + clinical data).

Table 1 Overall summary of the 108 most repeatable and reproducible features for all the experiments and presenting correlations with tumour volume ρ < 0.9.

Overall summary TOP 108 features: 46 LOG, 59 WF, 3 TA					
Median ICC and STD per categories					
TA (0.933 ± 0.024)					
LOG (0.923 ± 0.017)					
WF (0.917 ± 0.014)					
Feature name	**Category**	**Median ICC**	**Feature name**	**Category**	**Median ICC**
RANK 1–50					
wavelet-LLH_glcm_JointEnergy	WF	0.9822	log-sigma-2-0-mm-3D_glrlm_ShortRunEmphasis	LOG	0.9295

log-sigma-2-0-mm-3D_glrlm_RunLength NonUniformity	LOG	0.9766	wavelet-LLH_glrlm_ShortRunE mphasis	WF	0.9293
log-sigma-3-0-mm-3D_glrlm_RunLength NonUniformity	LOG	0.9724	log-sigma-2-0-mm-3D_glrlm_RunLengthN onUniformityNormalize d	LOG	0.9284
original_gldm_Depen denceNonUniformity	TA	0.9571	log-sigma-2-0-mm-3D_firstorder_Entropy	LOG	0.9284
wavelet-LLH_glcm_Idm	WF	0.9542	wavelet-LHL_firstorder_10Perce ntile	WF	0.928
log-sigma-3-0-mm-3D_firstorder_10Perce ntile	LOG	0.9484	log-sigma-1-0-mm-3D_gldm_LargeDepend enceEmphasis	LOG	0.9274
log-sigma-3-0-mm-3D_glrlm_RunPercent age	LOG	0.9475	wavelet-LHL_firstorder_Interqu artileRange	WF	0.927
wavelet-LLH_glcm_Id	WF	0.9474	wavelet-LHL_firstorder_MeanA bsoluteDeviation	WF	0.9258
wavelet-LLH_glcm_SumEntro py	WF	0.9424	log-sigma-1-0-mm-3D_glrlm_ShortRunEm phasis	LOG	0.9257
log-sigma-3-0-mm-3D_firstorder_Mean	LOG	0.9409	wavelet-LHL_firstorder_Robust MeanAbsoluteDeviation	WF	0.9255
log-sigma-3-0-mm-3D_firstorder_Median	LOG	0.9386	wavelet-LHL_glcm_Imc2	WF	0.9252
log-sigma-2-0-mm-3D_glrlm_LongRunE mphasis	LOG	0.9379	wavelet-LHL_glcm_Idn	WF	0.9227
log-sigma-2-0-mm-3D_gldm_LargeDepen denceEmphasis	LOG	0.9374	wavelet-LHL_glszm_SizeZoneN onUniformity	WF	0.9222
log-sigma-2-0-mm-3D_firstorder_10Perce ntile	LOG	0.9372	log-sigma-3-0-mm-3D_firstorder_MeanAbs oluteDeviation	LOG	0.9219
wavelet-LLH_glrlm_GrayLevel NonUniformityNorma lized	WF	0.9367	wavelet-LHL_glszm_ZonePercen tage	WF	0.9216
log-sigma-2-0-mm-3D_glrlm_RunVarianc e	LOG	0.9364	log-sigma-2-0-mm-3D_firstorder_Kurtosis	LOG	0.9213

log-sigma-2-0-mm-3D_glrlm_RunPercentage	LOG	0.9361	wavelet-LHL_gldm_SmallDependenceEmphasis	WF	0.9209
log-sigma-3-0-mm-3D_glrlm_RunLengthNonUniformityNormalized	LOG	0.9352	log-sigma-2-0-mm-3D_glcm_SumEntropy	LOG	0.92
original_glrlm_RunLengthNonUniformity	TA	0.9333	log-sigma-2-0-mm-3D_firstorder_RobustMeanAbsoluteDeviation	LOG	0.9198
log-sigma-3-0-mm-3D_gldm_DependenceVariance	LOG	0.9326	log-sigma-3-0-mm-3D_firstorder_Kurtosis	LOG	0.9196
wavelet-LLH_glrlm_RunLengthNonUniformityNormalized	WF	0.9322	wavelet-LHL_ngtdm_Contrast	WF	0.9188
log-sigma-3-0-mm-3D_glrlm_ShortRunEmphasis	LOG	0.9317	wavelet-LHL_ngtdm_Strength	WF	0.9182
log-sigma-3-0-mm-3D_firstorder_RootMeanSquared	LOG	0.9314	wavelet-HLL_firstorder_10Percentile	WF	0.9174
wavelet-LLH_glrlm_RunPercentage	WF	0.9304	wavelet-HLL_firstorder_Entropy	WF	0.9169
log-sigma-2-0-mm-3D_firstorder_Mean	LOG	0.9297	wavelet-HLL_firstorder_InterquartileRange	WF	0.9169
RANK 51–100					
wavelet-HLL_firstorder_MeanAbsoluteDeviation	WF	0.9168	wavelet-HLL_gldm_LargeDependenceEmphasis	WF	0.9089
wavelet-HLL_firstorder_RobustMeanAbsoluteDeviation	WF	0.9167	wavelet-HLL_gldm_SmallDependenceEmphasis	WF	0.9089
log-sigma-2-0-mm-3D_glcm_JointEntropy	LOG	0.9163	original_gldm_SmallDependenceEmphasis	TA	0.9084
wavelet-HLL_firstorder_Uniformity	WF	0.9162	log-sigma-2-0-mm-3D_glcm_JointEnergy	LOG	0.9082
wavelet-HLL_glcm_DifferenceAverage	WF	0.9157	wavelet-HLL_ngtdm_Contrast	WF	0.9074
wavelet-HLL_glcm_DifferenceEntropy	WF	0.9154	wavelet-HHL_glrlm_RunPercentage	WF	0.9073

log-sigma-3-0-mm-3D_firstorder_Variance	LOG	0.9153	wavelet-HHL_gldm_LargeDependenceEmphasis	WF	0.9069
log-sigma-2-0-mm-3D_firstorder_RootMeanSquared	LOG	0.9142	log-sigma-3-0-mm-3D_firstorder_RobustMeanAbsoluteDeviation	LOG	0.9059
log-sigma-2-0-mm-3D_firstorder_Uniformity	LOG	0.9139	wavelet-LLL_firstorder_Entropy	WF	0.9056
log-sigma-2-0-mm-3D_glcm_Id	LOG	0.9138	log-sigma-1-0-mm-3D_firstorder_RobustMeanAbsoluteDeviation	LOG	0.9055
log-sigma-2-0-mm-3D_glcm_Idm	LOG	0.9137	wavelet-LLL_firstorder_RootMeanSquared	WF	0.9051
wavelet-HLL_glcm_JointEntropy	WF	0.9134	wavelet-LLL_glcm_Contrast	WF	0.9051
wavelet-HLL_glcm_Idm	WF	0.9132	wavelet-LLL_glcm_DifferenceAverage	WF	0.9044
wavelet-HLL_glcm_Idmn	WF	0.9131	log-sigma-3-0-mm-3D_firstorder_InterquartileRange	LOG	0.9042
wavelet-HLL_glcm_Id	WF	0.9131	log-sigma-2-0-mm-3D_glrlm_GrayLevelNonUniformityNormalized	LOG	0.9039
wavelet-HLL_glcm_Idn	WF	0.913	log-sigma-2-0-mm-3D_glcm_DifferenceAverage	LOG	0.9039
wavelet-HLL_glcm_MaximumProbability	WF	0.9123	wavelet-LLL_glcm_DifferenceEntropy	WF	0.9037
log-sigma-2-0-mm-3D_firstorder_MeanAbsoluteDeviation	LOG	0.912	wavelet-LLL_glcm_JointEntropy	WF	0.9036
wavelet-HLL_glcm_SumEntropy	WF	0.9119	log-sigma-2-0-mm-3D_glcm_DifferenceEntropy	LOG	0.9036

log-sigma-1-0-mm-3D_glrlm_RunPercentage	LOG	0.9115	log-sigma-1-0-mm-3D_glcm_Id	LOG	0.9035
wavelet-HLL_glrlm_GrayLevelNonUniformityNormalized	WF	0.9114	wavelet-LLL_glcm_Idm	WF	0.9032
wavelet-HLL_glrlm_RunLengthNonUniformityNormalized	WF	0.9112	log-sigma-1-0-mm-3D_glrlm_RunLengthNonUniformityNormalized	LOG	0.9031
log-sigma-2-0-mm-3D_firstorder_InterquartileRange	LOG	0.911	wavelet-LLL_glcm_Id	WF	0.9031
wavelet-HLL_glrlm_RunPercentage	WF	0.9109	log-sigma-1-0-mm-3D_firstorder_InterquartileRange	LOG	0.9026
wavelet-HLL_glrlm_RunVariance	WF	0.9107	wavelet-LLL_glcm_Idn	WF	0.9025
wavelet-HLL_glrlm_ShortRunEmphasis	WF	0.9105	log-sigma-1-0-mm-3D_firstorder_10Percentile	LOG	0.9015
wavelet-HLL_glszm_LargeAreaEmphasis	WF	0.9098	wavelet-LLL_gldm_DependenceNonUniformity	WF	0.9009
wavelet-HLL_glszm_ZonePercentage	WF	0.9098	wavelet-LLL_gldm_DependenceNonUniformityNormalized	WF	0.9005
wavelet-HLL_glszm_ZoneVariance	WF	0.9092	wavelet-LLL_gldm_SmallDependenceEmphasis	WF	0.9001

The features are ordered by decreasing median ICC values (computed on all the experiments). Most reproducible and repeatable features per categories were: GLRLM-Non-Uniformity (LOG-2 mm kernel); LLH-GLCM-JointEnergy (WF) and Gray Level Dependence Matrix (GLDM) Non-Uniformity (TA).

It is interesting to notice how the top 50 repeatable features presented strong inter Spearman correlations, with Wavelet and Laplacian of Gaussian features being strongly clustered together (heatmap on Fig. 4). Overall, the number of features with good repeatability was found to be significantly larger than the number of reproducible features. Reproducibility experiments using phantom data (IntraCT experiment) led to more features being found reproducible compared to experiments performed using the clinical cohort (30% vs 19% of features with ICC ≥ 0.9, $p < 0.05$). Around 57% (138/243) of the robust features overlapped with features from repeatability and reproducibility study. The remaining 67 features being 36% Wavelet and 74% Laplacian of Gaussian were reproducible, but not repeatable.
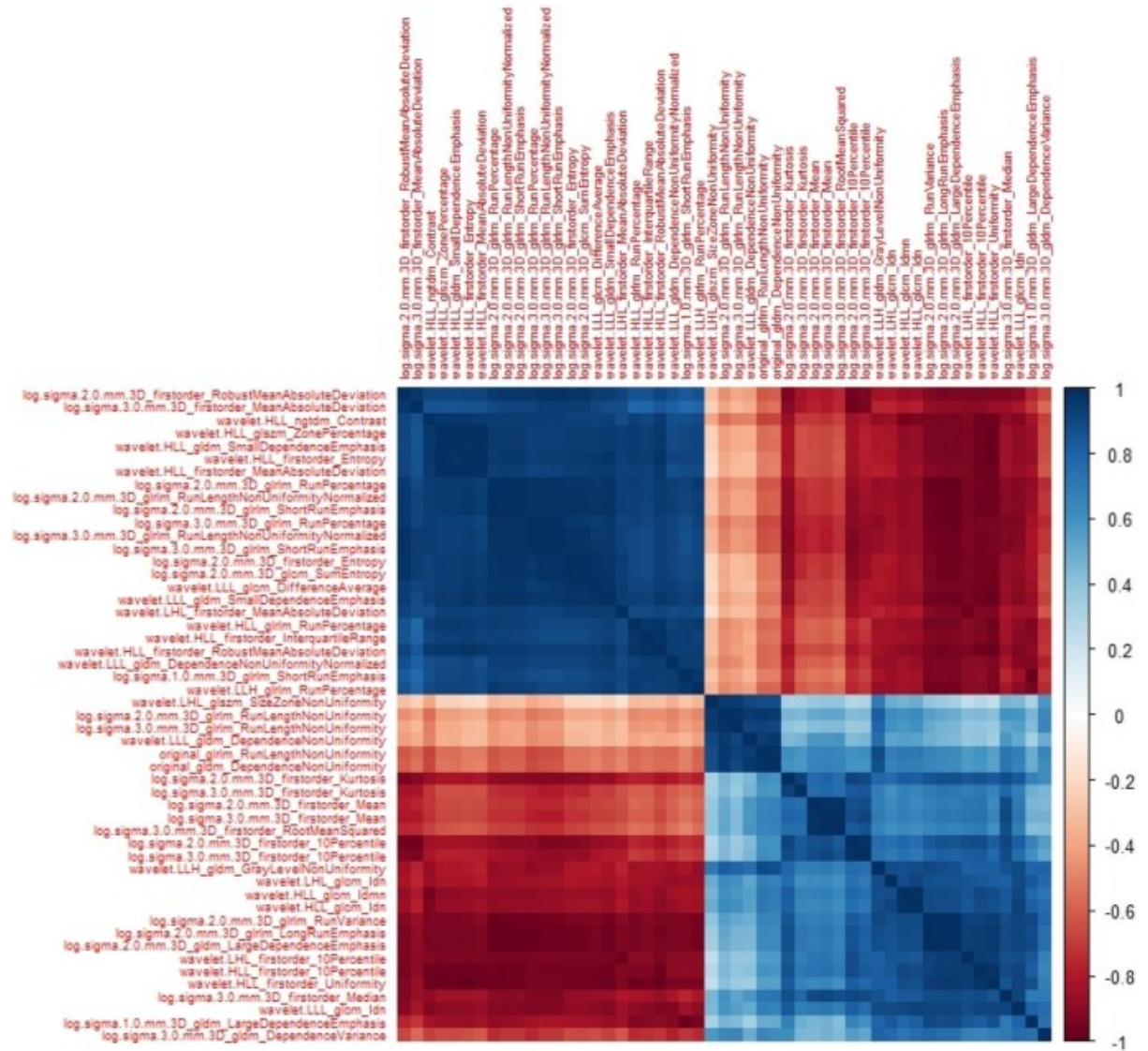
Figure 4: Heatmap showing Spearman correlations between the top 50 repeatable features.

Discussion

Copy (and adapt if necessary) from paper In this study, we investigated: (A) radiomic feature repeatability in a test–retest scenario using a NEMA IQ phantom; (B) radiomic feature reproducibility with respect to different tube currents, slice thickness as well as dependencies to different scanner models using an image quality phantom, and (C) radiomic feature reproducibility in a clinical cohort comparing three different acquisition protocols as well as the impact of slice thickness and the presence of IV contrast medium. We isolated a list of repeatable and reproducible features for all the experiments. Furthermore, we computed the correlations between radiomic features and tumor volume with the aim of investigating if the most repeatable and reproducible features also presented strong correlations. In fact, tumor volume was found to be the most robust feature and we wanted to assess if this could be a reason for a feature to present high reproducibility and repeatability. As shown in the results, only a relatively small percentage of radiomic features (around 13% of the total) presented both good repeatability and reproducibility across all the experiments. However, differences were found between repeatability and reproducibility. The number of features with good repeatability was larger than the number of reproducible features in the phantom experiment. Unfortunately, because we did not have any test–retest clinical data it was not possible to draw the same conclusion. Nevertheless, to obtain a fair comparison, we used the publicly available dataset RIDER to investigate the repeatability of radiomic features in NSCLC patients. The Venn Diagram in Fig. 3D shows that most of the repeatable and reproducible features in human data overlap with features from the phantom studies. This clearly shows that features computed on phantom are a superset of features computed on real human data. Our experiments also showed that there are some features extracted from human data that are robust but do not overlap with phantom results. Two main reasons could be associated with this: (A) statistical fluctuations because of the large number of computed features; (B) differences in the dynamic range of the features between phantom and human data. Point (B) is strictly related to the fact that the image quality phantom with spherical homogenous inserts are still not advanced enough to replicate tumor complexity seen in patients' data. Therefore, our study should be improved by including several types of imaging phantoms or considering new types of plugs that can better mimic tumor heterogeneity. In the last years, attention has been devoted to produce more realistic inserts by using 3D printing techniques[18,19]. The above-mentioned hypothesis seem to be confirmed by the fact that the features that did not overlap were only wavelet and Laplacian of Gaussian features, which might indicate that some real tumors' texture patterns are still difficult to be reproduced with imaging phantoms.

We found large variation of radiomic feature in repeatability study even within a short time gap of 30 min "coffee-break". Overall, less than 50% of features had a good repeatability (ICC > 0.9) using phantom scans, in agreement with previously published

literature[19,20,21]. When considering time-series analysis of radiomic features (e.g. for monitoring treatment response), temporal stability of radiomic features becomes imperative to be investigated. As mentioned in the introduction section, poor repeatability seems to be associated with poor prognostic/predictive power, while the reverse might not be equally true[9]. Therefore, our results can be taken by other radiomic studies to reduce the dimensionality of computed features by excluding poorly repeatable features.

When considering radiomic reproducibility, the presence or absence of IV contrast medium had a stronger impact than differences in slice thickness in the human study: 14% (155/1080) versus 47% (503/1080) ($p < 0.05$) of features with good reproducibility.

From the overall summary section in the results, it emerges that the different feature categories are sensitive with different degrees to reproducibility and repeatability. Our results are in line with the previous literature. The usage of image filtering could enhance the quality of the images even when acquired with different protocols and thus improve reproducibility. It is important to point out that this study did not investigate the robustness of shape metrics, since the contours were co-registered from PET to CT images and the same contour was used for all sets of CT series. However, shape metrics have been shown to be strongly affected by inter-observer variability in tumor delineations and this aspect was not investigated in this study.

We investigated how correlations between tumor volume and radiomic features could impact the repeatability and reproducibility. In line with other studies, not only tumor volume was the most repeatable and reproducible feature (median ICC = 0.99), but most of the top reproducible features showed strong Spearman correlations ($\rho > 0.9$) with tumor volume. This opens the debate whether their robustness could be an effect of an underlying "volume effect". However, more investigation is needed to isolate and further explain this effect. Therefore, in Table 1 we proposed the final list of most repeatable and reproducible features with lower correlations with tumor volume.

Finally, the list provided in Table 1 represents a starting point to isolate repeatable and robust features, but this is not enough to conclude about their prognostic predictive performance. Furthermore, as shown in Fig. 4, most of these features present strong intercorrelations and might produce redundant information if all are injected into a classifier for radiomic-based models. The results presented in this study needs to be validated in additional multi-institutional studies and considering additional parameters that can affect features' reproducibility and repeatability. First, in our analysis we only considered two different scanner manufacturers. We did not investigate the role of other acquisition parameters such as reconstruction kernels or tube voltage. These results are intended to be shared within the radiomic community for confirmation.

References

1. Beaton, L., Bandula, S., Gaze, M. N. & Sharma, R. A. How rapid advances in imaging are defining the future of precision radiation oncology. *Br. J. Cancer* **120**, 779–790 (2019).

2. Gillies, R. J., Kinahan, P. E. & Hricak, H. Radiomics: images are more than pictures, they are data. *Radiology* **278**, 563–577 (2016).

3. Gardin, I. *et al.* Radiomics: principles and radiotherapy applications. *Crit. Rev. Oncol. Hematol.* **138**, 44–50 (2019).

4. Kumar, V. *et al.* Radiomics: the process and the challenges. *Magn. Reson. Imaging* **30**, 1234–1248 (2012).

5. Yip, S. S. F. & Aerts, H. J. W. L. Applications and limitations of radiomics. *Phys. Med. Biol.* **61**, R150–R166 (2016).

6. O'Connor, J. P. B. *et al.* Imaging biomarker roadmap for cancer studies. *Nat. Rev. Clin. Oncol.* **14**, 169–186 (2017).

7. Traverso, A., Wee, L., Dekker, A. & Gillies, R. Repeatability and reproducibility of radiomic features: a systematic review. *Int. J. Radiat. Oncol. Biol. Phys.* **102**, 1143–1158 (2018).

8. Gudmundsson, S., Runarsson, T. P. & Sigurdsson, S. Test–retest reliability and feature selection in physiological time series classification. *Comput. Methods Programs Biomed.* **105**, 50–60 (2012).

9. Zhovannik, I. *et al.* Volume bias in textural radiomics. *Int. J. Radiat.* **105**(1), S118–S119 (2019).

10. van Griethuysen, J. J. M. *et al.* Computational radiomics system to decode the radiographic phenotype. *Can. Res.* **77**, e104–e107 (2017).

11. Larue, R. T. H. M. *et al.* Quantitative radiomics studies for tissue characterization: a review of technology and methodological procedures. *Br. J. Radiol.* **90**, 20160665 (2017).

12. Kim, H. G. *et al.* Quantitative analysis of the effect of iterative reconstruction using a phantom: determining the appropriate blending percentage. *Yonsei Med. J.* **56**, 253–261 (2015).

13. Shafiq-Ul-Hassan, M. *et al.* Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels. *Med. Phys.* **44**, 1050–1062 (2017).

14. van Timmeren, J. E. *et al.* Test-retest data for radiomics feature stability analysis: generalizable or study-specific?. *Tomography* **2**, 361–365 (2016).

15. Balagurunathan, Y. *et al.* Reproducibility and prognosis of quantitative features extracted from CT images. *Transl. Oncol.* **7**, 72–87 (2014).

16. Parmar, C. *et al.* Robust radiomics feature quantification using semiautomatic volumetric segmentation. *PLoS ONE* **9**, e102107 (2014).

17. Welcha, M. L. *et al.* Vulnerabilities of radiomic signature development: the need forsafeguards. *Radiother. Oncol.* **130**, 2–9 (2019).

18. Mitsouras, D. *et al.* Medical 3D printing for the radiologist. *Radiographics* **35**(7), 1965–1988 (2015).

19. Samei, E., Hoye, J., Zheng, Y., Solomon, J. B. & Marin, D. Design and fabrication of heterogeneous lung nodule phantoms for assessing the accuracy and variability of measured texture radiomics features in CT. *J. Med. Imaging (Bellingham)* **6**(2), 021606. https://doi.org/10.1117/1.JMI.6.2.021606 (2019).

20. Traverso, A. *et al.* Stability of radiomic features of apparent diffusion coefficient (ADC) maps for locally advanced rectal cancer in response to image pre-processing. *Phys. Med.* **61**, 44–51 (2019).

21. Sanduleanu, S. *et al.* Tracking tumor biology with radiomics: a systematic review utilizing a radiomics quality score. *Radiother. Oncol.* **127**, 349–360 (2018).

22. Jha, A. K. *et al.* Performance characteristic evaluation of a bismuth germanate-based high-sensitivity 5-ring discovery image quality positron emission tomography/computed tomography system as per National Electrical Manufacturers Association NU 2–2012. *World J. Nucl. Med.* **18**, 351–360 (2019).

23. Armato, S. G. 3rd. *et al.* The Reference Image Database to evaluate response to therapy in lung cancer (RIDER) project: a resource for the development of change-analysis software. *Clin. Pharmacol. Ther.* **84**(4), 448–456 (2008).

24. CT/PET FUSION DICOM CONFORMANCE STATEMENT for DICOM V3.0, Technical Publications Direction 2290660-100 Revision A, GE Medical Systems. https://www.gehealthcare.com/-/jssmedia/5337d686cfe442b2a75083038a877029.pdf?la=en-us. Accessed 17 July 2020.

25. Fedorov, A. *et al.* 3D slicer as an image computing platform for the quantitative imaging network. *Magn. Reson. Imaging* **30**, 1323–1341 (2012).

26. Lowekamp, B. C., Chen, D. T., Ibáñez, L. & Blezek, D. The design of simple ITK. *Front. Neuroinform.* **7**, 45 (2013).

27. Bartko, J. J. The intraclass correlation coefficient as a measure of reliability. *Psychol. Rep.* **19**, 3–11 (1966).

28. Koo, T. K. & Li, M. Y. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J. Chiropract. Med.* **15**, 155–163 (2016).

Chapter 6: Development and validation of GUI radiomics feature extractor software (PyRadGUI) using PyRadiomic package

Abstract

Radiomics is the method of extracting high throughput mathematical features from medical images. These features have potential to uncover disease characteristics inappreciable to a trained human eye. There are several open source and licenced tools to extract radiomic features such as pyradiomics, LIFEx, TexRAD and RaCat. Although pyradiomics is a widely used radiomics package by researchers, this software is not very user friendly and can be run using command line. We have developed and validated the GUI tool, PyRadGUI to make the radiomics software easy to operate. This software adheres to IBSI radiomic feature definition and implements the radiomic pipeline to in batch processing to extract radiomic features from multiple patient's data and stores it in comma separated value (csv). We validated PyRadGUI software with the existing pyradiomic pipeline.

Keywords— Graphical User Interface (GUI), machine learning, radiomics, pyradiomics, plastimatch, 3DSlicer

Introduction

As per the World health organization (WHO), cancer is the second leading cause of death worldwide. Globally one out of six deaths are caused by cancer alone which amounts to 9.6 million deaths in 2018.[1,2] Treatment of cancer has always remained a challenging task for the oncology community. Although earlier diagnosis and treatment is often associated with better outcome of treatment, at the same time selection of appropriate patients for appropriate treatment has been found to be important.[3] Considering the complexity of this disease and treatment, oncology is gradually moving towards personalized medicine.[4] A pathological test is considered a confirmatory test for cancer. Imaging tests like Computed Tomography (CT)/Positron Emission Tomography (PET)/Magnetic Resonance Imaging (MRI) also play an important role in diagnosis, treatment planning and follow-up of disease.[5] In last several years, the role of imaging and various advanced tests like immunohistochemistry (IHC), polymerase chain reaction (PCR) and gene sequencing have been increasing gradually in personalizing the treatment. Similarly, the role of artificial intelligence (AI) and radiomics has also witnessed a surge in oncology research over the last few years. Radiomics has been identified as an area of research to develop imaging biomarkers for personalized treatment of cancer.[6-8] Radiomics is a method to extract high throughput data from medical images. These features have the potential to uncover disease characteristics that are not appreciated by an expert radiologist or imaging personnel through visual interpretation.[9] As radiomic features are extracted directly from medical images it provides a non-invasive method for tumor characterization as demonstrated by various researchers in the past.[6-8] Researchers have shown the role of radiomics as a clinical predictor helping in advanced cancer care as personalized medicine in cancer.[10] Apart from its role in precision diagnoses and characterization of tumor, the role of radiomics has also been demonstrated in treatment planning.[11]

Several open-source and licensed software packages for radiomic extraction, like RaCaT, LifeX, IBEX, CaPTK or CGITA, Pyradiomics, TexRad have been developed and used in many published research.[12–17] The main challenge with these software packages are the complexity in radiomic extraction and lack of standardized feature extraction from these software. In order to standardize the radiomic extraction and feature definition, the image biomarker standardization initiative (IBSI) has provided mathematical feature definitions and phantom data sets with corresponding feature values.[18,19] Although this initiative is widely known, only few radiomic software have standardized the entire radiomic pipeline for feature extraction. Furthermore, the majority of these radiomic extractors have operability issues and not all defined features are extracted as defined by IBSI. Pyradiomics is a widely used open source radiomics package and adheres to the IBSI standards, but it is not user friendly. For instance, Pyradiomics is run on command prompt and customization is technically demanding. Hence, use of

this software is cumbersome and technically demanding for clinical doctors or scientists.

In order to make pyradiomics software more user friendly with the correct feature implementation of all features defined by the IBSI standard, we developed a graphic user interface (GUI) Radiomics extraction tool, PyRadGUI, that is easy to use and does not require any programming skills.

Methods

This work is part of Big Imaging data approach for Oncology in a Netherlands India Collaboration (BIONIC) and "personal health Train for radiation oncology in India and the Netherlands'' (TRAIN) project is approved by IEC of hospital as retrospective study. This software (PyRadGUI) is developed using Python open-source software on Windows systems.

Software Development

PyRadGUI front end was developed by using open-source software Python 3.6.5[20] and Python tk8.6[21] module was used for development of the GUI. Open source Plastimatch package 1.8.0[22] was used for DICOM to NRRD conversion of imaging data .and radiomic package in python; Pyradiomics 3.0[23, 24] was used for radiomic feature extraction. This software accesses DICOM image and ROI from a specific folder on the computer. First it converts the image and ROI in NRRD format using Plastimatch 1.8.0, subsequently it extracts radiomic features from the image and finally stores the output in comma separated values (csv) format in an output folder on the computer. The details of software are described below (figure 1).

To start the program, we run python3 GUI_batch_radiomics.py in the command line or terminal.

GUI, shown in figure 2, has been divided in two parts, i.e., left and right container. The left container contains 3 tabs which are used for customization and radiomic extraction and the right container displays the process and error if any. Once the GUI starts, the user must do the following before starting the radiomic extraction.
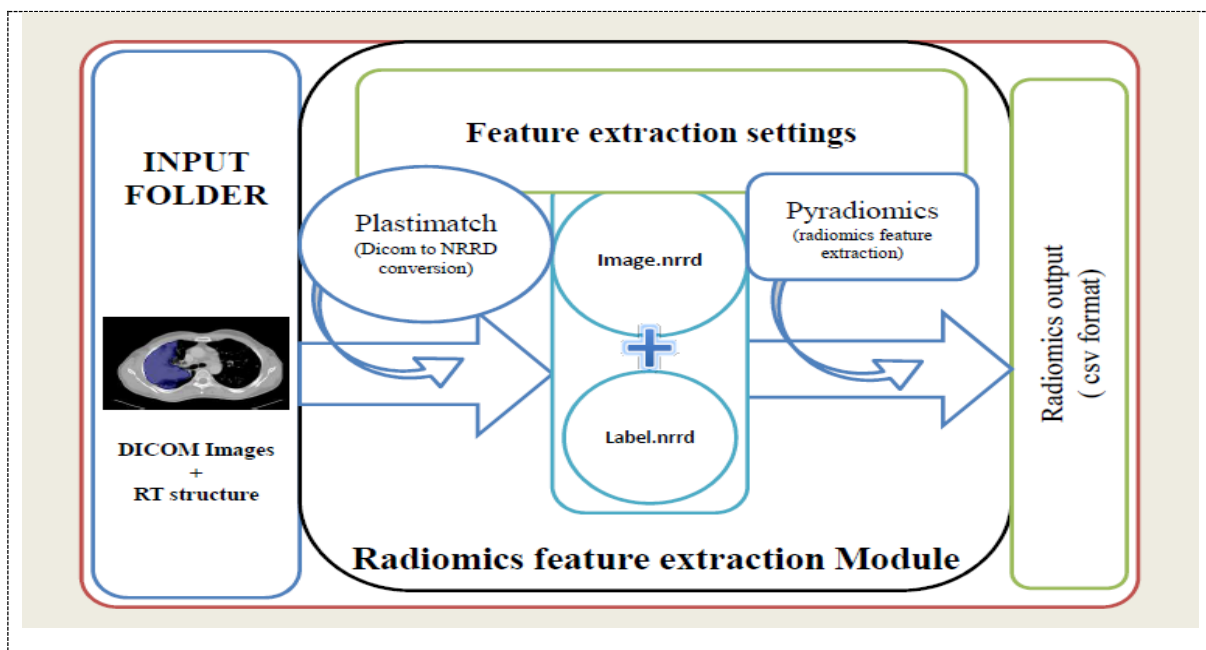
Figure 1. The batch process of radiomics features extraction from our tool. User has to select the input folder containing the DICOM Image and RT structure files. We can select the output folder and also customize the feature extraction process by changing values in the '.yaml' file.

1) Click on INPUT FOLDER BUTTON to provide path to input image folder containing multiple patients Image and RT structure files in DICOM format; 2) Click on OUTPUT FOLDER BUTTON to provide path to output folder; 3) Click on CUSTOMIZATION BUTTON to customize radiomic extraction.

After starting the program, the user must select an input folder containing multiple patient Image and RT structure files in DICOM format. Next, the user must select an output folder and then change the feature extraction settings from the settings tab.

Steps involved in customization settings:

1. Selection of Image type (Original, Laplace of Gaussian, Wavelet): To extract features on the original image, we can select Image type as original. If feature extraction has to be done on the transformed image, then we can select either LoG (Laplace of Gaussian) or Wavelet or both.

2. Selection of feature types: We have to specify the type of features we want to extract. The default includes all 1093 features from all image types.

3. Bin width selection: In the customization window one can select bin width as per the user requirement and default bin width is 25. We have used 25 bin-width for CT and MRI and 0.5 bin width for PET radiomic extraction.

4. Sigma value for LoG features: In the customization window one can select sigma value for LoG features. We have selected default 1-, 2-, and 3-mm sigma values for radiomic feature extraction.

5. Selection of Resampled pixel spacing: In the customization window one can select pixel spacing as desired by the user for radiomic extraction. We have used the default 2×2×2 cubic mm pixel spacing.

Figure 2. A) shows main GUI of radiomic extraction module console. B) is the customization window displayed on clicking on the customization button on the console.
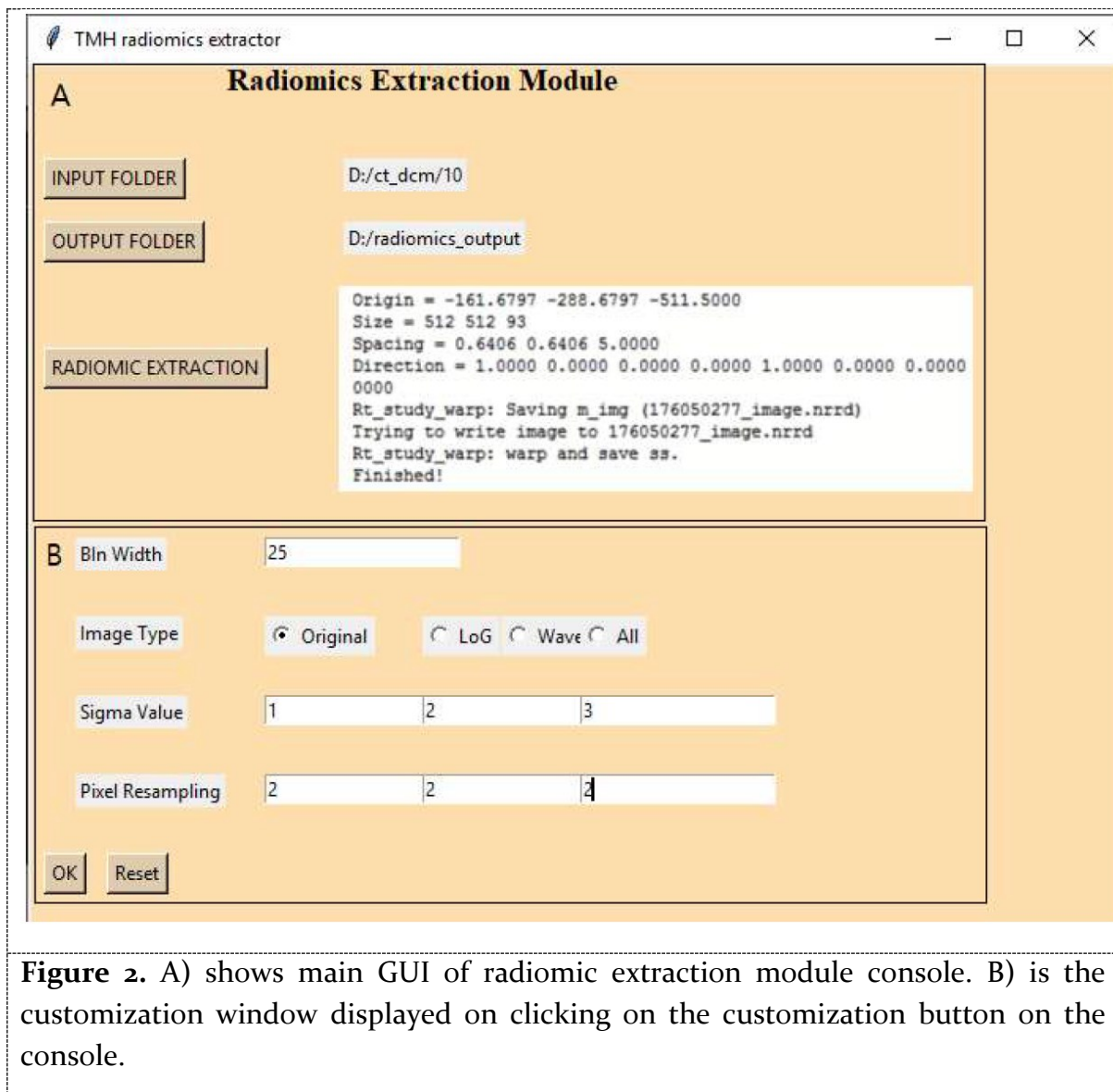
After the settings have been customised, radiomics extraction can be started by clicking on the radiomics extraction button. The batch extraction starts by loading the DICOM folder and calls plastimatch. Plastimatch takes the reference CT folder and converts the input DICOM to NRRD (nearly raw raster data) format. It converts image to image.nrrd and Mask to mask.nrrd. Pyradiomics takes the converted nrrd image, nrrd mask and the settings specified by the user and then extracts the radiomics features and writes the output in csv format in the selected output folder. In this output csv file columns represent radiomic features and rows represent individual patients. First column of the file is the patient identification number. This csv file can be used for various analyses. Status of running of the process, success, failures and error report are displayed in the right container in GUI and processing reports are stored in the output folder as log files.

System requirements:

We have developed and tested our software on two computer systems (Machines) using various software packages. The details of system information and software packages are shown in table 1.

	Machine 1	Machine 2
Processor	intel i7 10th generation, 3.6 GHz	intel Xeon E3-1220, 3.0 GHz
Operating system	Windows 10, 64 bits	Windows server 2008 R2, 64 bits
RAM	8 GB	8 GB
Software packages	Plastimatch 1.8.0 Pyradiomics 3.0 Python 3.6.5 Python tkinter 8.6	Plastimatch 1.8.0 Pyradiomics 3.0 Python 3.6.5 Python tkinter 8.6

Table 1: computer configuration and packages used for this study

Patient cohort:

In total 50 non-small cell lung carcinomas (NSCLC) patient's PET/CT data with delineation and 20 chondrosarcoma patient's MRI data with delineation who were imaged between 2014 to 2017 used for validation and performance testing of this software. Patient's demographic data is shown in table 2. 100 NSLC patients' images were used for testing the batch processing.

Disease	Gender	Total no.	Median Age	Image Type
NSCLC	Male	84	66	CT and PET Whole body
	Female	16	53	CT and PET Whole body
Chondrosarcoma	Male	17	46	MRI regional
	Female	3	42	MRI regional

Table 2: Demographic data of patient used in this study

Software Validation:

Software validation was performed by comparing Radiomic feature value extracted using PyRadGUI Workflow (PrGW) and the two reference workflows i.e.,Reference Workflow1(RW1) (3DSlicer +Pyradiomics) and Workflow2 (RW2) (Plastimatch +Pyradiomics). Same version of the Pyradiomic package and Plastimatch was used for all the workflows. 10 patients CT imaging data, 5 Patients PET imaging data and 5

Patients MRI imaging data were used for the comparison. PyRadGUI validation algorithm workflows are shown in figure 3.

Reference Workflow1 (RW1): Individual patients DICOM image and ROI were loaded in 3D slicer 3.0. First appropriateness of tumor delineation was checked by an experienced imaging physicist (15years experience). Image and ROI was converted in NRRD format and image saved as *image.nrrd* and ROI as *label.nrrd* in the patient's image folder. Again, NRRD image and label were loaded in 3D Slicer and appropriateness on tumor delineation was checked by the same physicist. Subsequently the radiomic feature was extracted and saved in CSV format using pyradiomic package on command prompt. Same process was repeated for all 20 patients' data.

Reference Workflow2 (RW2): DICOM image and RTStructure was converted in NRRD format using Plastimatch 1.8.0 and stored similarly as it is done in Workflow1. Subsequently radiomic features were extracted and stored in csv format similarly as it is done in workflow1. The algorithm used for manual extraction of radiomic features is shown in figure 3. All the patient's radiomic feature data was arranged similarly as it was arranged in automated extraction of radiomic feature data as described in automatic extraction section.
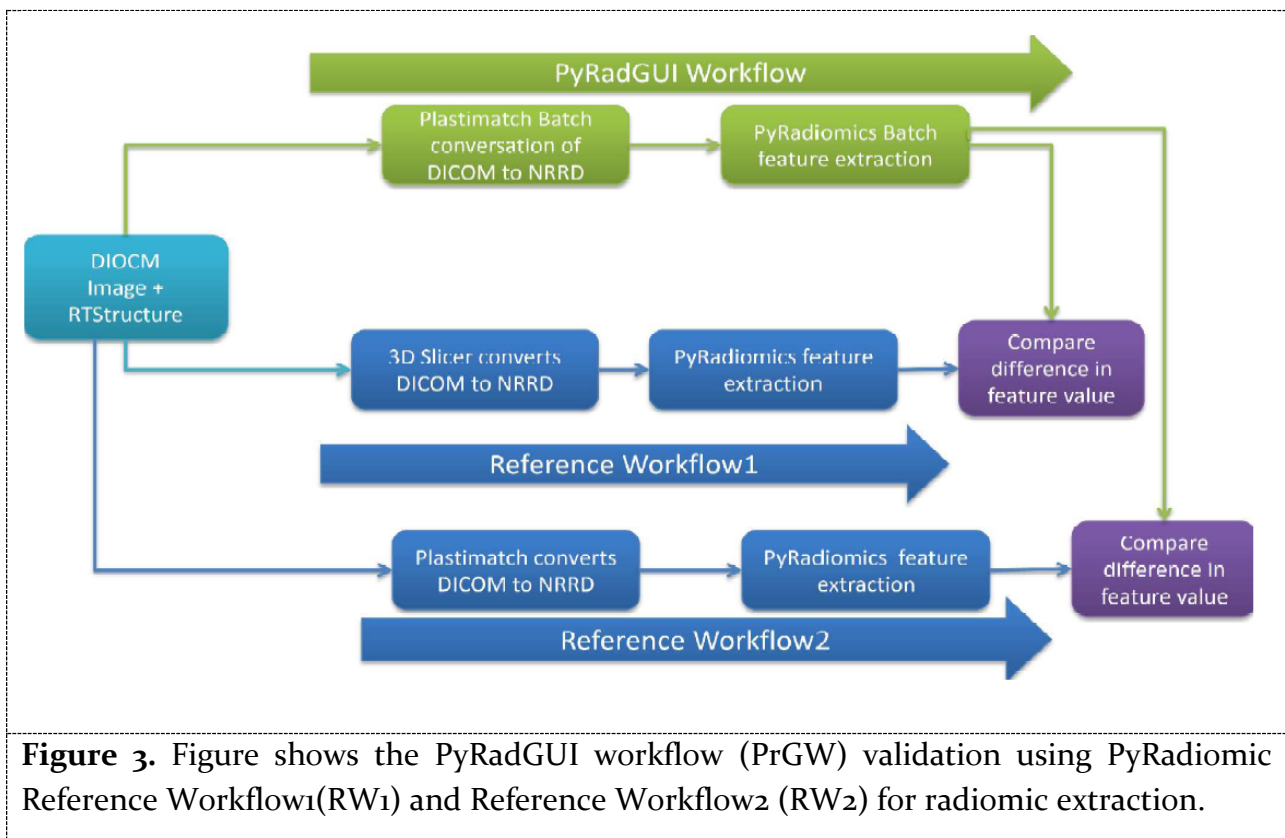
Figure 3. Figure shows the PyRadGUI workflow (PrGW) validation using PyRadiomic Reference Workflow1(RW1) and Reference Workflow2 (RW2) for radiomic extraction.

Statistical analysis:

Our PyRadGUI workflow was compared with manual Reference Workflow1 and Reference Workflow2 as shown in figure 3. Inter class correlation (ICC) was calculated for all 1093 radiomic features to compare PyRadGUI workflow and Reference Workflow1 using python code. As we were expecting the same result in PyRadGUI Workflow and Reference Workflow2, we used the EXACT function of Microsoft Excel 2007 software to compare both the data sets. The EXACT function compares two strings for all the characters if all the characters are the same in both the strings it gives true as output otherwise false.

Performance testing of software:

Performance of the software was tested for the batch processing on two machines mentioned in table 1. CT and PET DICOM data was used in the batches of 10, 20, 50 and MRI DICOM data was used in the batches of 10, and 20 The total time required for individual batch processing was recorded.

Results

We successfully installed the software on two computers mentioned in table 1. We successfully ran the software (GUI) several times on both the systems. Software tabs are successfully able to assign input and output folders on the computer. Software is also successfully able to assign values through the customization button on both the computers. We were able to perform batch processing of up to 100 patients CT data on both the computers.

1.1. Software validation:

The ICC value for the comparison of PrGW and RW1 was ICC =0.978(range: 0.9612-1.0) across all the modality. The ICC value for the comparison of PrGW and RW1 is shown in table 3. All the values (for 1093 radiomic features) in the validation steps for CT, PET and MRI features for PrGW and RW2 were TRUE it shows that we were able to extract same values for all the features for the same data across all the modalities by using this software as confirmed by the manual process.

	ICC value (Mean±SD)		
	CT (1093 features) [10patient]	PET (1093 features) [5patient]	MRI (1093 features) [5patient]
Comparison: PrGW vs. RW1	0.983±0.0613	0.72±0.31	0.84±0.23
PrGW vs. RW2	0.983±0.0613	0.72±0.31	0.84±0.23

Table 3: table shows ICC value between PrGW and RW1 for CT, PET and MRI radiomic feature extraction

Performance testing of software:

On both the machines the batch processing performance was found to be satisfactory. The details of run time are shown in table 4. The tool was tested with different modalities images.

Modality	Computer system	Runtime for 10 patients		Runtime for 20 patients		Runtime for 50 patients	
		Total time (min)	Total time for one patient's data (sec)	Total time (min)	Total time for one patient's data (sec)	Total time (min)	Total time for one patient's data (sec)
CT	Machine 1	13.3	81.2	17.4	53.4	62.0	75.2
	Machine 2	08.0	58.0	11.4	30.7	35.3	31.5
PT	Machine 1	24.0	180	63.4	190.2	150.3	180.3
	Machine 2	17.0	102.1	53.0	159.0	93.9	112.7
MRI	Machine 1	12.3	75.8	19.5	94.2		
	Machine 2	01.1	7.7	02.3	7.6		

Table 4: Shows processing time required by both Machines for batch processing of radiomic extraction

Discussion

Several licenced and open-source software are available for radiomic extraction, which have capability to extract radiomic features from two- or three-dimensional medical images.[25-29] Shi Z. et.al have developed Ontology-guided radiomics analysis workflow (O-RAW) using Pyradiomics package and SITK Python package, which is also able to extract radiomics feature from DICOM image and RTSTRUCT in batch processing and store it in resource description framework (RDF) triple store.[30] Another radiomic extraction software TexRAD has been used by many researchers is a GUI based system. TexRAD is a licensed package but unable to do the batch processing.[26] This software has capabilities to display and review the image and delineate tumor manually but another drawback of this software is its inability to use existing delineation (RTStructure).[31-32] The RaCaT is radiomic software implemented the IBSI defined feature but not available as GUI tool as well as unable to perform batch processing. The PyRadGUI software is a GUI based tool and it can implement batch processing of DICOM image and RT structure for radiomic extraction. As this software is an extension of the pyradiomic package it inherently implements IBSI feature definition. It can extract radiomic features from hundreds of patient's images and RTStructure in batch processing mode and store the result in csv format. Although we have not compared the delineation converted to NRRD using Plastimatch and 3D Slicer, radiomic feature values comparison show excellent agreement (ICC=0.998±0.012) between two methods. As our results show, this software calculates radiomics features accurately and reliably. Radiomic extraction from PET and CT images take much longer time compared to MRI images, as PET and CT have whole body images [contains more data] and MRI has regional images [contains less data]. We have used Pyradiomics, an open-source software for radiomic extraction in our research infrastructure because this infrastructure can easily be replicated in other research centres. This software works as a plug-in and has no dependencies on pyradiomic package version, it can be upgraded as and when new pyradiomic package is available. Customization is the unique feature of this software, which provides flexibility to the user to customize the parameters in 'yaml' file of the pyradiomics package. The ability of our software to customize and extract 1093 radiomic features from medical images in batch processing enables faster processing of radiomic extraction and storage of feature values in csv format. During the customization user can also select a specific group of features to be extracted. The advantage of this software is its GUI and GUI based customization of extraction, which allows performing the entire task from the GUI console by clicking the available buttons on the console. The csv format in which this software stores data where each column represents radiomic features and rows represent individual patient's data makes it easier to be utilized for machine learning. It can also be concatenated with clinical data if required. Log files can be used for identifying any error that occurs during the processing. Error log of the individual patient's data is stored which can allow us to

identify the specific data with error and we may take corrective action afterward. In our existing project we have also developed artificial intelligence (AI) infrastructure for AI based research in oncology and PyRadGUI is also an integral part of it. The PyRadGUI can be implemented as standalone as well as part of AI infrastructure for radiomic based research. The portability and easy installation of this software will encourage the radiomic community to use this software and this software can be a valuable addition to radiomic based research infrastructure.

There are also few limitations of this software like it is unable to display the image before or during the procedure. It requires DICOM image as well as RTStructure for radiomic processing. Future work will be to test this software on *Linux operating system*, add a statistical and prediction analytics module and image segmentation and display module to this tool.

Conclusion

We successfully implemented and validated, PyRadGUI, a GUI based easy to use Radiomic extraction software. This software can easily be implemented on Windows systems. The extracted features using this software are meeting the IBSI standards. We have found this software to perform batch processing of up-to 100 patients and extract radiomic features and store it in ready to use csv format for machine learning. A documentation including the description of how to install and use this software can be found on GitHub. (https://github.com/Bionic-TMH/PyRadGUI)

References

[1] Cancer, https://www.who.int/news-room/fact-sheets/detail/cancer

[2] Ferlay J, Colombet M, Soerjomataram I, Mathers C et.al. 2019 *Int J Cancer.* **144(8)**:1941-1953

[3] Stewart CL, Warner S, Ito K, Raoof M, Wu GX, et.al. 2018 *Curr Probl Surg* **55(9)** 330-379

[4] Agyeman AA, Ofori-Asenso R 2015 *J Pharm Bioallied Sci* **7(3)** 239-44

[5] Caers J, Withofs N, Hillengass J, Simoni P, et.al. 2014 *Haematologica* **99(4)** 629-37

[6] Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, et.al. 2012 *Eur J Cancer* **48(4)** 441-6

[7] Goodwin J, Neugent ML, Lee SY, Choe JH, Choi H, et.al. 2017 *Nat Commun.* **8** 15503

[8] Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C et.al. 2014 *Nat Commun* **5** 4006

[9] Yip SS, Liu Y, Parmar C, Li Q, Liu S, Qu F, et al. 2017 *Scientific Reports* **7 (1)** 3519

[10] Haider, S.P., Burtness, B., Yarbrough, W.G. et al. 2020 *Cancers Head Neck* **5** 6

[11] Limkin EJ, Sun R, Dercle L, Zacharaki EI, Robert C et al. 2017 *Ann Oncol* **28** 1191-206.

[12] Pfaehler E, Zwanenburg A, de Jong JR, Boellaard R 2019 *PLoS One* **14(2)** e0212223

[13] Nioche C, Orlhac F, Boughdad S, Reuzé S, Goya-Outi J et al. 2018. *Cancer Res* **78** 4786-4789.

[14] Zhang L, Fried D V, Fave XJ, Hunter LA, Yang J, Court LE. 2015 *Med Phys* **42** 1341-53

[15] Davatzikos C, Rathore S, Bakas S, Pati S, Bergman M, et al. 2018 *J Med Imaging* **5** 1

[16] Johnson, McCormick I. 2015 *The ITK Software Guide* https://itk.org/ItkSoftwareGuide.pdf

[17] Haralick RM, Shanmugam K, Dinstein I. 1973 *IEEE Trans Syst Man Cybern.* **3** 610–621

[18] Zwanenburg A, Vallières M, Abdalah MA, et al. 2020 *Radiology* **295(2)** 328-338

[19] Zwanenburg A. 2017 *Radiother Oncol* **123** S914-S915

[20] Python, https://www.python.org/downloads/release/python-365/

[21] Python, https://docs.python.org/3/library/tkinter.html

[22] Plastimatch, http://www.plastimatch.org/

[23] Pyradiomics, https://pyradiomics.readthedocs.io/en/latest/

[24] Griethuysen, J. J. M., Fedorov, A., Parmar, C., et.al. 2017 *Cancer Research* **77(21)** e104-e107

[25] Shi Z. 2017 PyRadiomics Extension (Py-rex) https://github.com/zhenweishi/Pyrex.

[26] Gotz M, Nolden M, Maier-Hein K 2019 *Radiother Oncol* **131** 108-111

[27] Szczypinski PM, Strzelecki M, Materka A, Klepaczko A 2009 *Comput Methods Programs Biomed* **94** 66-76.

[28] Zhang L, Fried DV, Fave XJ, Hunter LA, Yang J, Court LE 2015 *Med Phys.* **42** 1341-1353

[29] Apte AP, Iyer A, Crispin-Ortuzar M et al. 2018 *Med Phys* **45** 3713–3720.

[30] Shi Z, Traverso A, van Soest J, Dekker A, Wee L 2019 *Med Phys* **46(12)** 5677-5684

[31] Ganeshan B, Miles KA, Young RC, Chatwin CR 2008 *Invest Radiol* **43(6)** 382-94

[32] Davnall F, Yip CS, Ljungqvist G, Selmi M, Ng F, Sanghera B, Ganeshan B, Miles KA, Cook GJ, Goh V 2012 *Insights Imaging* **3(6)** 573-89

Chapter 7: Implementation of Big Imaging Data Pipeline Adhering to FAIR Principles for Federated Machine Learning in Oncology

Abstract

Cancer is a fatal disease and one of the leading causes of death worldwide. The cure rate in cancer treatment remains low; hence cancer treatment is gradually shifting towards personalized treatment. AI and radiomics have been recognized as one of the potential areas of research in personalized medicine in oncology. Several researchers have identified the capabilities of AI and radiomics to characterize the phenotype and thereby predict the outcome of treatment in oncology. Although AI and radiomics have shown promising initial results in diagnosis and treatment in oncology, these technologies are also facing challenges of standardization and scalability. In the last few years, researchers have been trying to develop a research infrastructure for federated machine learning that increases the usability of Big Data for clinical research. These research infrastructures are based on the Findable, Accessible, Interoperable and Reusable (i.e., FAIR) data principles. The India-Dutch "Big Imaging data approach for Oncology in a Netherlands India Collaboration" (BIONIC) is a jointly funded initiative by the Dutch Research Council (NWO) and the Indian Ministry of Electronics and Information Technology (MeitY), aiming to introduce radiomic-based research into clinical environments using federated machine learning on geographically dispersed collections of FAIR data. This paper described a prototype end-to-end research infrastructure implemented through the BIONIC partnership into a leading cancer care public hospital in India.

Introduction

Cancer is a fatal disease and is the second leading cause of deaths worldwide. According to GLOBOCAN, cancer accounted for about 10 million deaths in 2020 [1]. Cancer treatment is complex and requires multidisciplinary collaboration [2]. Gradually, the paradigm in cancer treatment is shifting towards more personalized medicine [3]. While the personalized approach is gaining ground in oncology, artificial intelligence (AI) is also being applied to support personalized treatment [4].

There is a vast spectrum of treatment-related data i.e., imaging, pathology, biochemical measurements and blood tests, genetics and clinical observation, being generated every day during cancer treatment and recorded in electronic format. The data are stored in various parts of a hospital information system (HIS) [5]. Medical imaging such as computed tomography (CT), positron emission tomography (PET), magnetic resonance imaging (MRI), and single-photon emission tomography (SPECT) play an integral role in cancer diagnosis and treatment [6]. These constitute the majority proportion by volume in a typical HIS. Medical images are stored in specialized high-throughput storage devices known as Picture Archival and Communication Systems (PACS) [7].

Imaging features, along with a wide spectrum of histological and clinical features, have been explored as prognostic indicators to predict the future outcomes of cancer treatment. "Radiomics" is a major research development in recent years that attempts to define a digital "fingerprint" of cancer [8]. Radiomics involves high-throughput mathematical extraction of structured quantitative data (i.e., features) from qualitative clinical imaging [9]. A radiomics signature, meaning a set of single features carrying either prognostic or predictive value, can augment conventional visual interpretation by human experts, and reveal deeper insights into structure, behaviour, and therapeutic response of cancer [10-14]. However, the vast volume of features extracted by radiomics can also be unwieldy to manage.

The implementation of AI research in oncology touches upon many problems; these include: a) data retrieval from a HIS and syntactic standardization of data, b) standardization of radiomics extraction, c) incorporation of radiomics models into clinical practice, d) maintaining persistent and inter-operable descriptions of the data, and e) archival of data in such a way that is findable and accessible to future researchers.

Some of the aforementioned issues can be ameliorated by following FAIR data principles. Kalendralis et al. [15] have shown how semantic ontologies can be used to apply syntactic standards to, and provide descriptive metadata about feature extraction methods on top of, existing medical data but without editing the original source data itself. Semantic Web standards based on the Resource Descriptor Framework (RDF) can be used to store data with persistent unique identifiers, in such a way that is completely agnostic to the underlying database schema in the original data source [16]. This

therefore allows powerful query, filter and joining commands in the SPARQL language that can link disparate data sources together.

Above and beyond the FAIR principles, scientific collaboration can be encouraged at the health systems and institutions level by homogenizing data collection in clinical routine procedures and allowing multi-institutional sharing of data [15].

However, actual sharing of patient-level data requires additional contractual procedures associated with data ownership, control, access, permitted usage and protection of patient confidentiality. Such legal needs vary immensely between jurisdictions, for example, The Netherlands and India [18]. The need to share individual patient data may be side-stepped via a privacy-preserving distributed learning approach using federated (decentralized) datasets for statistical modelling [19]. Researchers have demonstrated the feasibility of federated learning over a variety of open and non-open-source infrastructures, showing that models can be trained on large datasets, are equivalent to results on centralized data and can support radiomics model training and validation [19-24]. The RDF-based FAIR data representation forms the basis of distributed learning systems that were able to operate with multi-site geographically-dispersed data sources [25-28].

In this work, we explain how a Big Imaging Data processing pipeline has been implemented in Tata Memorial Hospital (TMH) in Mumbai, India, with the support of BIONIC partners. We discuss how FAIR has been used as a guiding principle, such that data is available for internal research as well as for privacy-preserving federated machine learning. While the solutions presented needed to be intrinsically adapted to TMH, the intention is to demonstrate how similar technologies could be implemented in other hospitals that make private data available for federated learning studies.

Methods

Overall organization of the implementation

The overall implementation scheme of the pipeline is shown in Figure 1. This pipeline serves as the template for preparing data in a FAIR manner, using RDF and ontologies in the data aggregation/integration layer. Prior to this, we show how clinical, imaging, and text report data need to be individually extracted from different sections of the HIS using data stream-specific workflows, and how standardization and meta-data need to be applied within each of these stream-specific workflows.
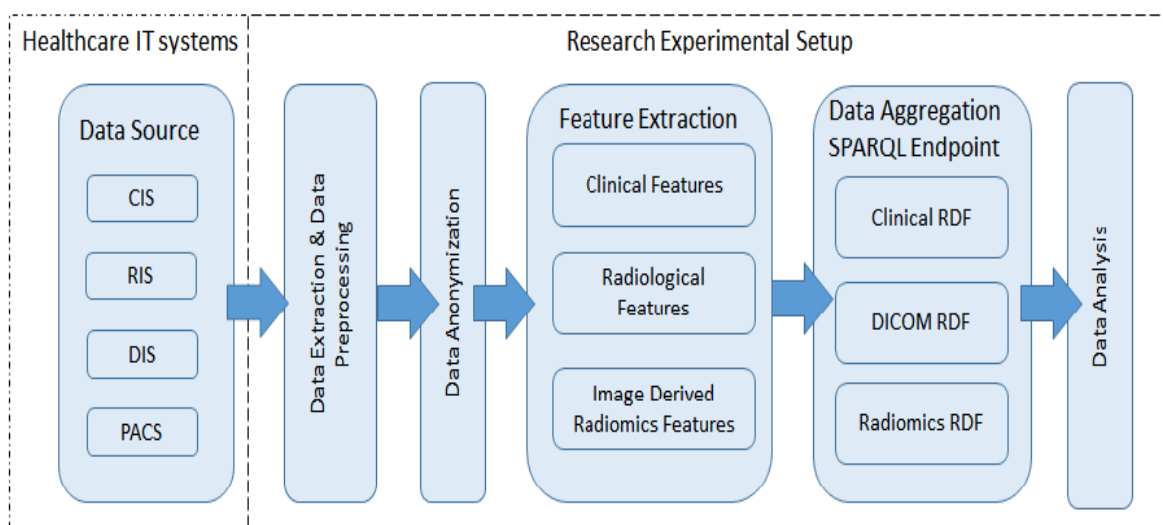
Figure 1: Schematic diagram providing overview of the BIONIC big imaging data processing pipeline.

Datasets

Health IT infrastructure comprises multiple software applications that are required to manage day-to-day clinical activities. The clinical data, non-image diagnostic data, radiologist reports and radiological images are stored in the HIS as sub-systems; a clinical information system (CIS), diagnostic information system (DIS), radiological information system (RIS), and PACS, respectively. Table 1 shows the types of data within each of the aforementioned subsystems. Imaging data and its associated metadata are stored in a PACS using the industry-standard Digital Imaging and Communications in Medicine (DICOM) format.

Patient Consent

For retrospective non-experimental studies, the institutional policy was to apply for a consent waiver through the Institutional Ethics Committee (IEC). The IEC reviews the research study and gives approval (including any necessary usage conditions) for private data of hospital subjects to be harvested from the HIS.

Information Systems	Nature of Data
Clinical Information Systems (CIS)	Demographic data, treatment data and follow-up data including clinical baseline factors such as TNM stage, pleural effusion etc.
Diagnostic Information Systems (DIS)	Pathology report, Immunohistochemistry (IHC) Report and Blood Report that are based on the blood and tissue samples
Radiological information systems (RIS)	Radiology Report and Nuclear Medicine report that includes Imaging related diagnostic findings like malignancy, non-malignancy, disease progression

Table 1: Health IT systems and attributes

Data Extraction

In-house scripts were developed at TMH in Python programming language to customize and control the data extraction from the HIS subsystems. The data extraction is covered in two major modules – A) non-imaging data for the CIS, RIS and DIS streams (see Figure 2), and B) imaging data from the PACS (see Figure 3).

A.   Non-image data

The non-image data extraction workflow shown in Figure 2 from CIS, DIS and RIS requires detailed understanding of the internal schema in each system. Our procedure also includes checking authorization to access the data elements (in accordance with the IEC approval). The customizable parts refer to mapping of the data elements to specific locations in the HIS and extraction routines to retrieve values from the HIS.

The data harvested from the HIS will rarely be complete in all aspects. There is generally a high prevalence of "missing values". The customization also covers how missing values are handled. For instance, the missing values can be initially filled in as 'NA' text strings. Next specific filtering rules set by the clinical user for a given study can be applied, such as:

a. If 20% or more of values in the data field is missing, then the data field may be omitted.

b.   Or if 20% of the data fields for a given subject is missing, the subject may be omitted.

c. Or if a value is missing in a strictly mandatory data field, such as "gender" or "survival status", then the subject may be omitted.

These filtering steps are performed to ensure a reasonably high degree of data completeness coming from the harvesting process, without entirely relying on complete case analysis. Additional filters, imputation and exclusion rules can be manually added into the workflow if required. Lastly, Protected Health Information (PHI) is obscured by using a lookup key file to replace identifiable information (supplementary material). The result of extraction of non-image data is a de-identified plain text file in comma-separated value (CSV) format.
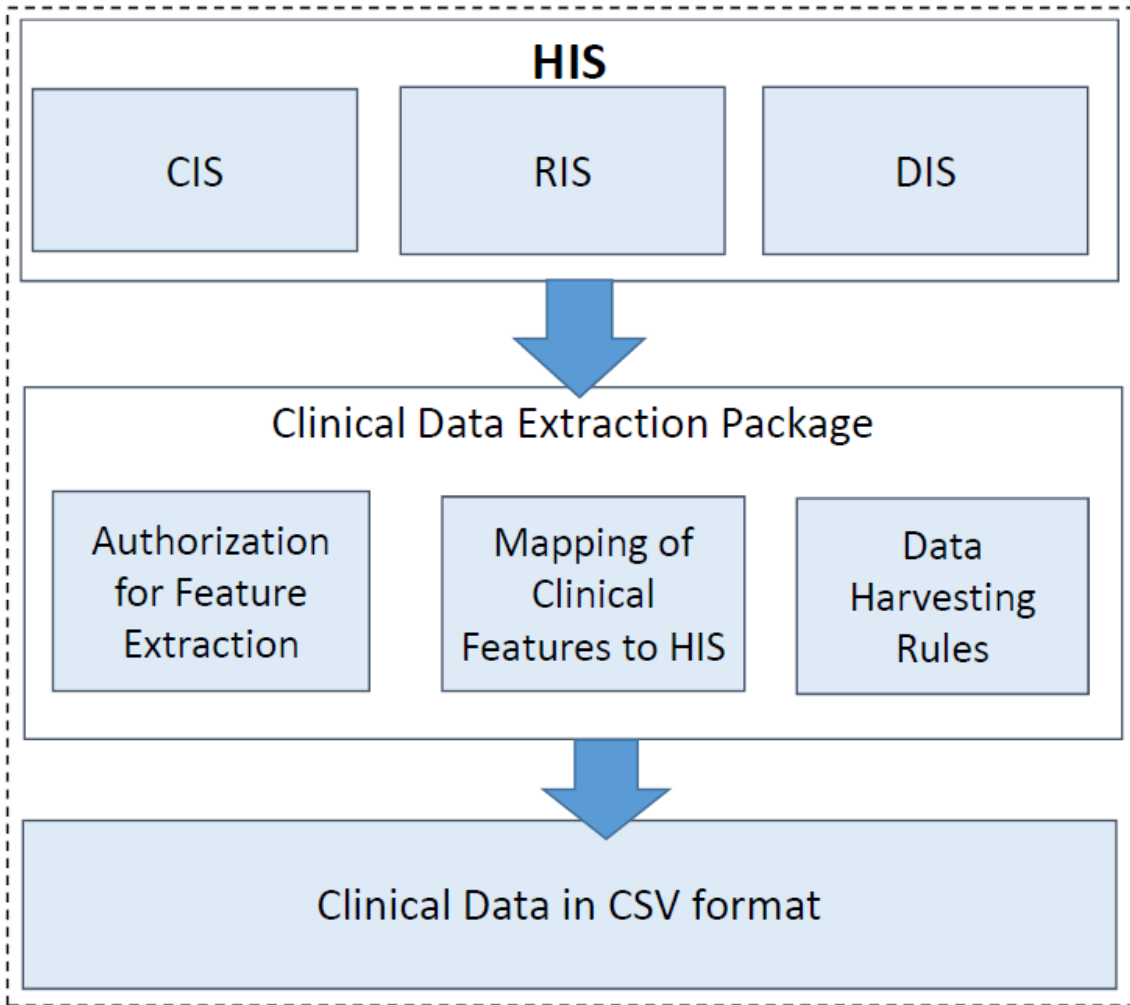
Figure 2: Overview of the non-image data extraction module.

B. Imaging data

The image data extraction workflow is shown in Figure 3. Individual subjects imaging studies comprising of CT, MRI and/or PET need to be manually "pushed" from the PACS or retrieved using one of the integrated image management workstations provided by a vendor, e.g. Philips Intellispace Discovery (research-only build; Philips Medical System, Eindhoven, The Netherlands) or Advantage 4.6 (General Electric, Waukesha WI, USA. The native format for medical images was retained as DICOM.

For BIONIC data preparation, we were specifically interested in the image data and the region-of-interest annotation file "RTSTRUCT". The RTSTRUCT files were generated in the vendor workstations using radiological annotation tools. The Philips Intellispace Discovery platform permits the option to connect with other tools, such as a radiomics extraction tool or a deep-learning automated segmentation algorithm. As before, PHI is obscured using the same lookup key file as for the non-imaging data.
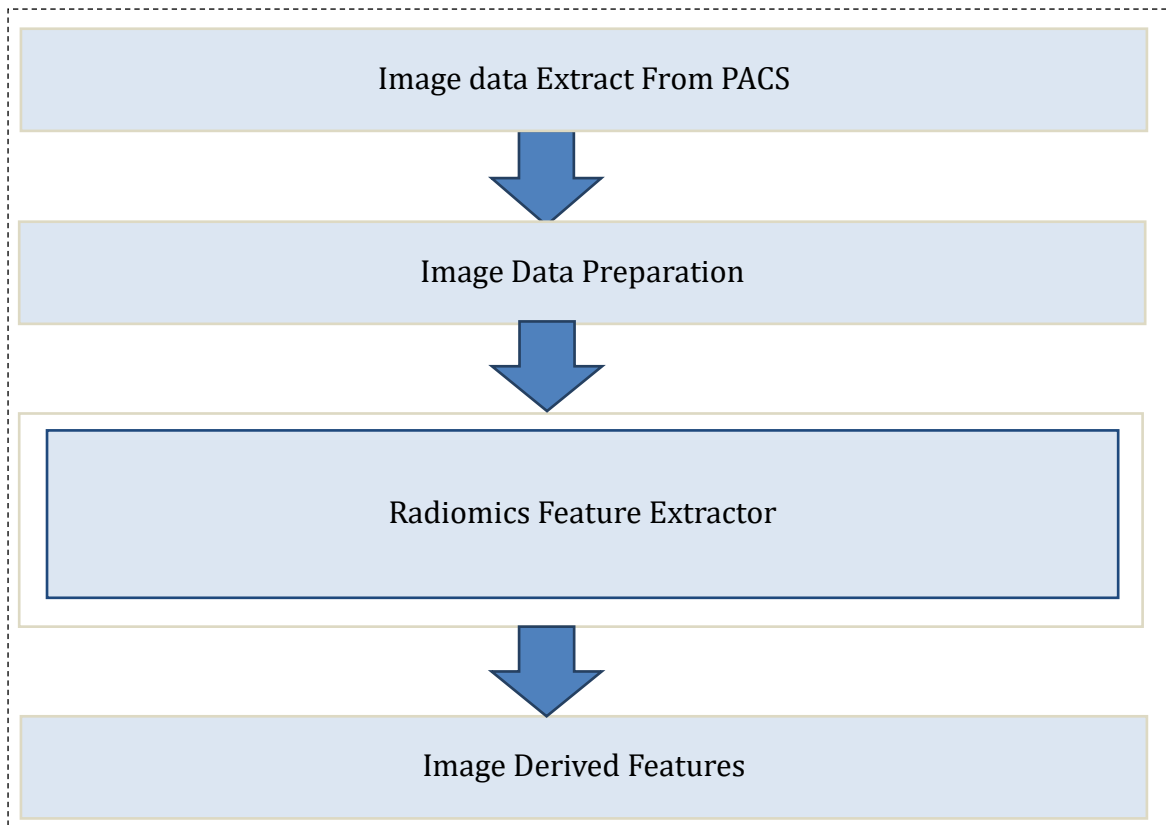
Figure 3: Overview of the image data extraction module.

We implemented a radiomics feature extractor in Python language based on ORAW [29] but we used the *Plastimatch*v1.9.0 [30] library to convert a gross tumor volume (GTV) region of interest from each RTSTRUCT file into a binary mask. The image and corresponding binary mask were passed to *Pyradiomics* v2.1.2 [31-33] to compute 1093 features. Features consisted of 13 shape, 17 intensity-histogram and 73 textural features. Intensity-histogram and textural features were re-computed after applying Laplacian of Gaussian (LoG) filters with three widths (total 270 features) and wavelet decomposition filters at 8 levels (total 720 features). Details about the features and filters are available on the online Pyradiomics documentation. Parameters of radiomics extraction were controlled via configuration (.yml) files for each of Plastimatch and Pyradiomics. The radiomics features were saved as CSV format (figure 4).
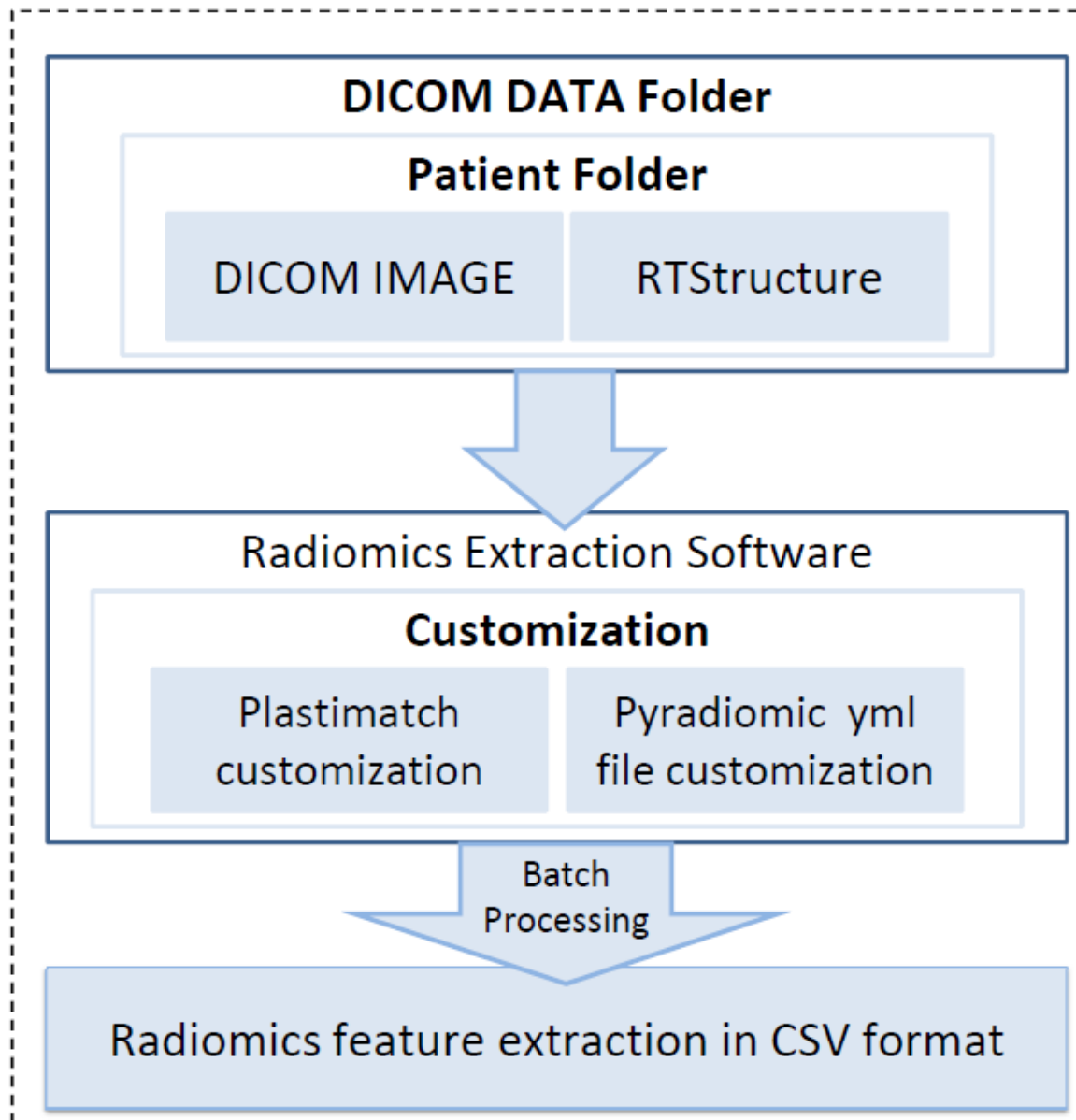
Figure 4: Radiomics extraction pipeline

Data Aggregation

The above steps have defined how clinical features and radiomics features have been separately extracted de-identified and then pre-processed as individual CSV objects. The intermediate step of saving these CSV objects allows additional quality assurance and inspection of the data for any errors or inconsistencies that had escaped interception, particularly for the clinical factors. At the moment, this is being done manually by domain experts and researchers but we leave the option open for future automation (see Discussion later). The CSV format made it amenable to be copied into SPSS or other data management tools for additional cleaning and filling of missing values (where possible).

Data integration, standardization and inter-operability are performed using an ontology-guided semantic mapping procedure. Figure 5 indicates the ontologies that are applied to the individual data streams; the Radiation Oncology Ontology (ROO) [34, 35] and the National Cancer Institute Thesaurus (NCIT) [36] for clinical and treatment-related features, an open-source DICOM ontology [37] for DICOM imaging metadata and the Radiomics Ontology (RO) [38, 39] for the radiomics features that follow the Image Biomarker Standardization Initiative recommendations [40, 41].

Each of the CSV objects were converted to RDF format using an in-house Python script using the rdflib 5.0.0 [42] library with different target ontologies as defined above. The DICOM headers were extracted directly into RDF (without passing through an intermediate CSV) via a plug-in provided on Philips Intellispace Discovery.
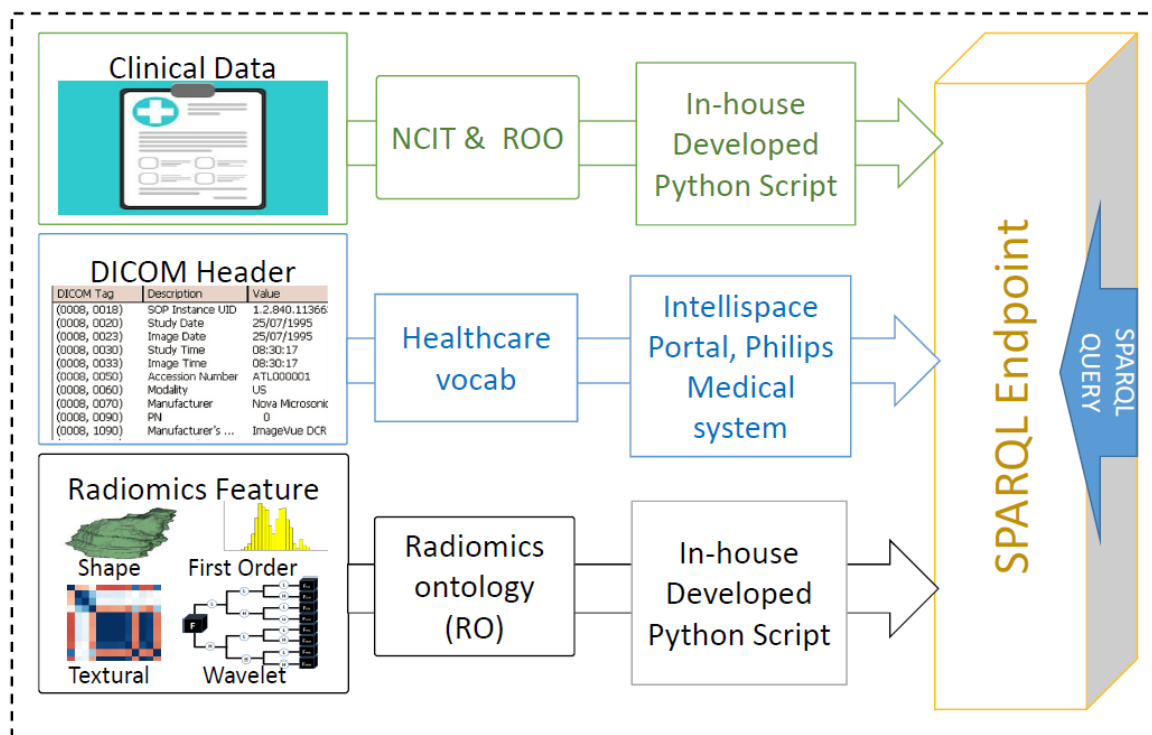
Figure 5: Illustrated that the clinical, DICOM metadata and radiomics data are converted to RDF format and integrated as linked data in a local SPARQL repository.

The final destination of the generated RDF data was an Apache Jena Fuseki server [43] installed as a SPARQL endpoint inside the hospital IT firewall. The SPARQL query language is then used to access the RDF triples that are archived in the SPARQL endpoint [44]. The RDF triples are maintained in a persistent online graph database through a TDB triple store client application [45], which also supplies a user interface through which remote query of RDF data is possible using a SPARQL v1.1 compliant engine called ARQ [46]. SPARQL queries are entered in ARQ to retrieve the data from the RDF store [46].

Results

Given a large number of Python scripts, management of the data pipelines can potentially become cumbersome. In collaboration with BIONIC partner C-DAC, a prototype workflow management system has been developed that integrated some of the scripting work behind a graphical web interface (dashboard). This was done to try to reduce the level of technical complexity for a typical clinically-minded researcher. Figure 6 provides a cursory overview of the functionality implemented to date in the prototype BIONIC dashboard at TMH. This includes views for managing data, for task logging and for script execution status reports. The dashboard provides interfaces to the user to inspect clinical and image derived features directly from CSV files or DICOM folders. Backend subroutines can call on internal scripts where needed to extract radiomics features and export the data directly into RDF for immediate querying using a built-in SPARQL interface. We hasten to add that such work is still a developmental prototype at the present time; however more functionality and scripting integration are planned in future.
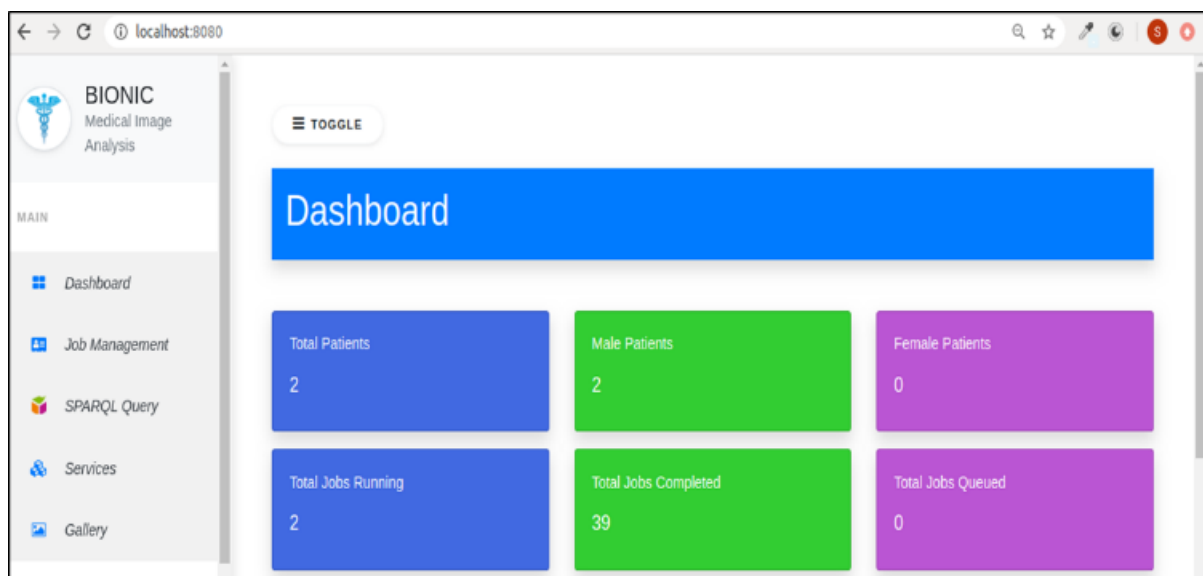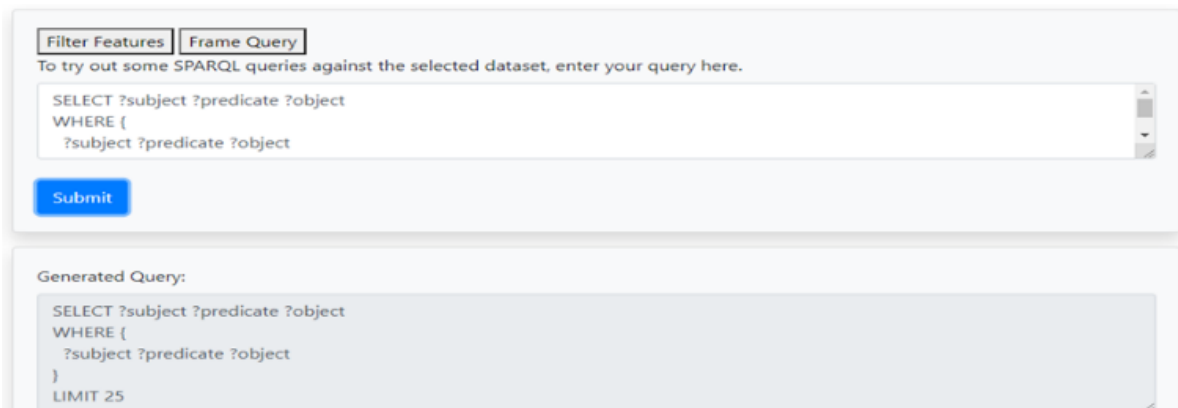
Figure 6: Home screen of a prototype web-based dashboard for the BIONIC project.

Figure 7 shows the user-friendly SPARQL query web interface that allows filtering and joining operations on RDF data residing in the local RDF data endpoint. As an example, that illustrates the power of SPARQL queries to retrieve and join data from disparate sources, we provide an example in Textbox S1 in the Supplementary Materials. In this demonstration query, we retrieved patients' age and biological sex from the clinical RDF, then linked it with two radiomics features in the Radiomics RDF. The join was achieved via the patients' ID and patients' CT scan identifier stored in the DICOM

metadata RDF. The result of such a query is shown in Figure S1 also in the supplemental material.

Figure 7: Close up view of the SPARQL querying interface, where a researcher may enter filtering and linking queries to locate data in the RDF database.

Discussion

The purpose of this paper is to describe a medical imaging data processing pipeline linking clinical features, imaging metadata and extracted radiomics features. The goal of the BIONIC collaboration is to help get data ready for federated machine learning and multi-centre collaboration at large scale, but without forcing anyone to share identifiable data. The lynchpin of this work is thus to integrate disparate sources of data in such a way that is agnostic to the internal schema of the databases, language, local coding, etc. Semantic ontologies and mapping scripts were central to linking hospital data together, guided by the FAIR management principles.

Though the procedures presented here need to be specifically adapted to TMH, our intention was to demonstrate how similar procedures might be implemented in other hospitals that can make vast amounts of private data more efficiently interoperable and re-usable in research.

Patients' demographics or clinical details were reasonably directly findable on the local HIS. We designed the data extraction module to be customizable to harvest project-dependent data from separate subsystems within the HIS; assuming appropriate regulatory permissions have been given. This level of automation could search thousands of patients' data fields within minutes. At the present time, quality assurance and searching for missing values still unavoidably needs expert human intervention, however the utilization of such automated data collection procedures had already significantly reduced time and effort in finding relevant data.

Several licensed and open-source software have been made widely available for radiomic extraction [9, 47-50], but close integration with clinical data pipelines has not been previously demonstrated in detail. This work demonstrates that is indeed feasible, with simple connector scripts, to integrate such radiomics software as "plug and play" modules into an imaging research workflow that can consume clinical real-world images close to its source.

While it is widely accepted that AI models require vast amounts of data for training, it has not been so well established as to how model training can be integrated with on-the-ground clinical record keeping systems such as the HIS and its multiple subsystems [51]. Interoperability of the data and richly descriptive metadata will become ever more critical, since AI research still faces significant challenges in independent validation; AI models developed in one setting with one set of data seem rarely able to inter-operate smoothly with data from an independent setting, and hence AI model performance is generally degraded during independent external validation [52-54]. The recommendation to pool data together or share data to centralized repositories does not adequately acknowledge the administrative and legal barriers that stand in the way with regards to sharing individual-level data. This is the foundation of our interest in rendering data in such a way as to be more inter-operable in federated machine learning projects.

However, there remain some major challenges that we have not yet addressed in the present work. First, there exists extreme heterogeneity in HIS schema and design, such that the low-level extraction scripts that interface with the HIS and its components are expected to be very difficult to clone from one hospital to another; in this scenario, one size certainly does not fit all. Great effort has been expended in TMH in order to customize the data harvesting procedures to the systems and schema that are available locally. It is likely such adaptation efforts need to be repeated in every hospital, unless vendors of HIS and related electronic systems step in with some universally and strictly applied standards. Clinical medical imaging already has the global DICOM standard, which aids significantly with information retrieval and sharing. While HIS vendors are certainly investing in query and data retrieval tools that are compliant with FHIR and HL7, progress and adoption remains slow.

Secondly, even assuming that data can be rapidly harvested from the HIS, it remains a major effort to impose syntactic and semantic interoperability on the data, so that it can be universally understood by any person and, more importantly, by any machine algorithm. In this work, we have used the power of semantic ontologies to harmonize terminology and attach metadata to data, such that these "unique identifiers" can be used to construct filtering and linking queries over heterogeneous data sources. This design task remains highly time consuming and requires high levels of expertise, as well as knowledge of the clinical domain. With rapid development of the Natural Language Processing (NLP) domain within AI and machine learning, it is expected that emerging NLP technologies can eventually be deployed to task of the semantic labelling of data.

This is related to another limitation of the present work that remains to be addressed in detail; much of clinical and diagnostic data exists as natural free-flowing human language texts (so-called "free text"). Such narrative descriptions of disease - its diagnosis, development and outcome – is recognized as a rich source of potentially clinically actionable information, but we have presently placed the major part of our initial efforts on structured information encoded into specific data fields in the HIS. We acknowledge therefore that our work is not yet complete, and better utilization of unstructured free-text information must be achieved in the near future by exploiting NLP.

With regard to FAIR principles, we need to consider how close we have managed to come to the FAIR principles, while also recognizing gaps in implementation that need to be closed in future [55].

In terms of findability, we have successfully transitioned from highly localized data retrieval from the HIS into an RDF database with persistent unique identifiers based on public domain ontologies, including identifiers for richly descriptive metadata for images and radiomics features. However, due to the privacy sensitive nature of the data, patient data and metadata of such detailed nature cannot be placed online, even in de-identified form. In its place, we have not yet developed an institutional protocol to make our data repository findable without betraying confidentiality.

With regards to accessibility, we have discussed our aforementioned strategy favoring privacy-preserving federated learning. Our present vision is that data accessibility matters need to be defined within formal research collaborations, each with focused clinical questions and contractual legal arrangements pertaining to use of real-world patient data.

In the matters pertaining to inter-operability and re-usability, we propose that our data integration with RDF already applies a high degree of general understandability to the data, which in turn allows us to re-use patient data garnered from routine encounters and provision of standard care. At this time, specifications of lexica and ontologies used, as well as conditions of using the data, are all governed within a formal multi-party collaboration structure.

Finally, the aforementioned medical imaging data pipeline was intended to make data FAIR for internal clinical research, in addition to being the necessary preparatory steps to making data FAIR for federated international collaboration. To this end, we felt the development of a prototype dashboard for management, process monitoring, and linked data queries, and presenting this through a user-friendly internal browser-based interface, is an important step towards usability and clinical deployment. Whereas web-based database SPARQL query is also possible for authorized collaborators using web-based query tools.

Several state-of-the-art ETL and research data warehouse are described in literature, which are designed for customized extraction, ontological mapping and the automated generation of SQL statements [56-59]. These data warehouses also allow the storage of heterogeneous medical data as ours but unlike uses relational structured database. The ETL tools of these data warehouse extract data from structured table of HIS, transforming them to a given target structure with the help of ontology and finally load the source system contents into a research data warehouse. Whereas our ETL module is able to extract data from EHR without knowing the relational database structure and table directly through web access and free text data mining provides more scalability to the module. Ontology mapped triples allows SPARQL query independent of local terminology.

Conclusion

We implemented a medical imaging data processing pipeline linking clinical factors, imaging metadata and extracted radiomics features. This was done for a comprehensive cancer hospital in India (TMH) as a demonstration of the BIONIC Indo-Dutch research collaboration. Data integration across HIS subsystems was guided by the FAIR data principles; specifically we relied on domain semantic ontologies for terminology standardization, knowledge representation and internal data linkage. We harvested and then stored clinical data, DICOM imaging metadata and extracted radiomics features as an RDF repository. A browser-based dashboard was also provided to facilitate usage and future deployment as a possible clinical research environment. This local RDF-based system is synergistic with, and readily connected to, a multi-centre federated machine learning infrastructure.

References

[1] Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021;71(3):209-249. doi:10.3322/caac.21660

[2] Chakraborty S, Rahman T. The difficulties in cancer treatment. Ecancermedicalscience. 2012 Nov 14;6:ed16.

[3] Agyeman AA, Ofori-Asenso R. Perspective: Does personalized medicine hold the future for medicine? J Pharm Bioallied Sci. 2015 Jul-Sep;7(3):239-44.

[4] Schork NJ. Artificial Intelligence and Personalized Medicine. Cancer Treat Res. 2019; 178:265-283.

[5] Mohamed Khalifa, Osama Alswailem, Hospital Information Systems (HIS) Acceptance and Satisfaction: A Case Study of a Tertiary Care Hospital, Procedia Computer Science, Volume 63, 2015, Pages 198-204, ISSN 1877-0509

[6] Fass L. Imaging and cancer: a review. Mol Oncol. 2008 Aug;2(2):115-52.

[7]Schultheiss TE, Coia LR, Martin EE, Lau HY, Hanks GE. Clinical Applications of Picture Archival and Communications Systems in Radiation Oncology. SeminRadiatOncol. 1997 Jan;7(1):39-48.

[8] Aerts, H., Velazquez, E., Leijenaar, R. et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun 5, 4006 (2014).

[9] Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, Zegers CM, Gillies R, Boellard R, Dekker A, Aerts HJ. Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer. 2012 Mar;48(4):441-6.

[10] Fornacon-Wood I, Faivre-Finn C, O'Connor JPB, Price GJ. Radiomics as a personalized medicine tool in lung cancer: Separating the hope from the hype. Lung Cancer. 2020 Aug;146:197-208.

[11] Liu Z, Wang S, Dong D, Wei J, Fang C, Zhou X, Sun K, Li L, Li B, Wang M, Tian J. The Applications of Radiomics in Precision Diagnosis and Treatment of Oncology: Opportunities and Challenges. Theranostics. 2019 Feb 12;9(5):1303-1322.

[12]Lambin P, RiosVelazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer. 2012;48:441–446.

[13] Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun. 2014;5:4006.

[14] Lambin P, Leijenaar RT, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol. 2017;14:749–762.

[15]Kalendralis P, Shi Z, Traverso A, Choudhury A, Sloep M, Zhovannik I, Starmans MPA, Grittner D, Feltens P, Monshouwer R, Klein S, Fijten R, Aerts H, Dekker A, van

Soest J, Wee L. FAIR-compliant clinical, radiomics and DICOM metadata of RIDER, interobserver, Lung1 and head-Neck1 TCIA collections. Med Phys. 2020 Jun 10.

[16]Resource Description Framework (RDF), https://www.w3.org/RDF/ (Accessed 30Apr. 2021).

[17]Zuiderwijk A, Shinde R, Jeng W. What drives and inhibits researchers to share and use open research data? A systematic literature review to analyze factors influencing open research data adoption. PLoS One. 2020 Sep 18;15(9):e0239283.

[18]Tucker K, Branson J, Dilleen M, Hollis S, Loughlin P, Nixon MJ, Williams Z. Protecting patient privacy when sharing patient-level data from clinical trials. BMC Med Res Methodol. 2016 Jul 8;16 Suppl 1(Suppl 1):77.

[19]Shi Z, Zhovannik I, Traverso A, Dankers FJWM, Deist TM, Kalendralis P, et al. Distributed radiomics as a signature validation study using the Personal Health Train infrastructure. Sci Data. (2019) 6:1–8.

[20]Shi Z, Foley KG, Pablo de Mey J, Spezi E, Whybra P, Crosby T, van Soest J, Dekker A, Wee L. External Validation of Radiation-Induced Dyspnea Models on Esophageal Cancer Radiotherapy Patients. Front Oncol. 2019 Dec 16;9:1411.

[21]Deist TM, Jochems A, van Soest J, Nalbantov G, Oberije C, Walsh S, et al. Infrastructure and distributed learning methodology for privacy-preserving multi-centric rapid learning health care: euroCAT. Clin TranslRadiat Oncol. (2017) 4:24–31.

[22] Wolfson, M. et al. DataSHIELD: resolving a conflict in contemporary bioscience—performing a pooled analysis of individual-level data without sharing the data. International journal of epidemiology 39, 1372–1382 (2010).

[23] Lu, C.-L. et al. WebDISCO: a web service for distributed cox model learning without patient-level data sharing. Journal of the American Medical Informatics Association 22, 1212–1219 (2015).

[24] Bogowicz, M., Jochems, A., Deist, T.M. et al. Privacy-preserving distributed learning of radiomics to predict overall survival and HPV status in head and neck cancer. Sci Rep 10, 4542 (2020). https://doi.org/10.1038/s41598-020-61297-4

[25] Jochems A, Deist TM, van Soest J, et al. Distributed learning: Developing a predictive model based on data from multiple hospitals without data leaving the hospital - A real life proof of concept. Radiother Oncol. 2016;121(3):459-467. doi:10.1016/j.radonc.2016.10.002

[26] Deist TM, Dankers FJWM, Ojha P, et al. Distributed learning on 20 000+ lung cancer patients - The Personal Health Train. Radiother Oncol. 2020;144:189-200. doi:10.1016/j.radonc.2019.11.019

[27]Dankers, F.. "Prediction modeling and distributed learning for radiotherapy outcomes in lung cancer patients." (2019).

[28]Shi Z, Zhovannik I, Traverso A, et al. Distributed radiomics as a signature validation study using the Personal Health Train infrastructure. Sci Data. 2019;6(1):218. Published 2019 Oct 22. doi:10.1038/s41597-019-0241-0

[29]Shi Z, Traverso A, van Soest J, Dekker A, Wee L. Technical Note: Ontology-guided radiomics analysis workflow (O-RAW). Med Phys. 2019 Dec;46(12):5677-5684.

[30]Plastimatch V1.9.0 software, https://sourceforge.net/projects/plastimatch/(Accessed 30Apr. 2021).

[31]Pyradiomics package V2.2.0, https://pyradiomics.readthedocs.io/en/latest/ (Accessed 30Apr. 2021).

[32] Griethuysen, J. J. M., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., Beets-Tan, R. G. H., Fillon-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.

[33] van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, Beets-Tan RGH, Fillion-Robin JC, Pieper S, Aerts HJWL. Computational Radiomics System to Decode the Radiographic Phenotype. Cancer Res. 2017 Nov 1;77(21):e104-e107.

[34]Radiation Oncology Ontology | NCBO BioPortal. https://bioportal.bioontology.org/ontologies/ROO. (Accessed 30Apr. 2021).

[35] Traverso A, van Soest J, Wee L, Dekker A. The radiation oncology ontology (ROO): Publishing linked data in radiation oncology using semantic web and ontology techniques. Med Phys. 2018 Oct;45(10):e854-e862.

[36]National Cancer Institute Thesaurus(NCIT)https://bioportal.bioontology.org/ontologies/NCIT (Accessed 30Apr. 2021)

[37]Semantic DICOM Ontology, https://bioportal.bioontology.org/ontologies/SEDI (Accessed 30Apr. 2021).

[38]Healthcare vocab, http://purl.org/healthcarevocab/v1 (Accessed 30Apr. 2021).

[39]Radiomics Ontology NCBOBioPortal. https://bioportal.bioontology.org/ontologies/RO. (Accessed 30Apr. 2021).

[40]The image biomarker standardisation initiative, https://ibsi.readthedocs.io/en/latest/index.html, (Accessed 30Apr. 2021).

[41]Radiological Society of North America. Quantitative Imaging Biomarkers Alliance® (QIBA®). rsna.org https://www.rsna.org/research/quantitative-imaging-biomarkers-alliance/. (Accessed 30Apr. 2021).

[42]rdflib 5.0.0, https://rdflib.readthedocs.io/en/stable/(Accessed 30Apr. 2021).

[43] Apache Jena - Apache Jena Fuseki. https://jena.apache.org/documentation/fuseki2/. (Accessed 30Apr. 2021).

[44] Apache Jena - SOH - SPARQL over HTTP. https://jena.apache.org/documentation/fuseki2/soh.html. (Accessed 30Apr. 2021).

[45] TDB, https://jena.apache.org/documentation/tdb/(Accessed 30Apr. 2021).

[46] ARQ- Extending Query Execution, https://jena.apache.org/documentation/query/arq-query-eval.html(Accessed 30Apr. 2021).

[47]Pfaehler E, Zwanenburg A, de Jong JR, Boellaard R. RaCaT: An open source and easy to use radiomics calculator tool. PLoS One. 2019 Feb 20;14(2):e0212223.

[48]Court, Laurence E., Xenia Fave, Dennis Mackin, Joonsang Lee, Jinzhong Yang, &Lifei Zhang. "Computational resources for radiomics." Translational Cancer Research [Online], 5.4 (2016): 340-348. Web. 26 Nov. 2020

[49]Doi K. Computer-aided diagnosis in medical imaging: historical review, current status and future potential. Comput Med Imaging Graph 2007;31:198-211.

[50]Rogers W, Ryack B, Moeller G. Computer-aided medical diagnosis: literature review. Int J Biomed Comput1979;10:267-89.

[51]Zodwa Dlamini, Flavia Zita Francies, Rodney Hull, RahabaMarima,Artificial intelligence (AI) and big data in cancer and precision oncology, Computational and Structural Biotechnology Journal,Volume 18,2020,Pages 2300-2311, ISSN 2001-0370,

[52]Boon IS, Au Yong TPT, Boon CS. Assessing the Role of Artificial Intelligence (AI) in Clinical Oncology: Utility of Machine Learning in Radiotherapy Target Volume Delineation. Medicines (Basel). 2018 Dec 11;5(4):131.

[53] Ahmed Z, Mohamed K, Zeeshan S, Dong X. Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. Database (Oxford). 2020 Jan 1;2020:baaa010.

[54]Shaikh F, Franc B, Allen E, Sala E, Awan O, Hendrata K, Halabi S, Mohiuddin S, Malik S, Hadley D, Shrestha R. Translational Radiomics: Defining the Strategy Pipeline and Considerations for Application-Part 2: From Clinical Implementation to Enterprise. J Am Coll Radiol. 2018 Mar;15(3 Pt B):543-549.

Supplementary Materials

Concept mapping with ontology

We have manually created semantic relationships between medical concepts with target ontology concept/id. The mapping of the source concept of our data was performed with the ontology connects of the target ontology. Supplementary table 1 shows the example of mapping of clinical and radiomic concepts with target ontology concept.

Local concept	Ontology concept	Ontology ID
Patient	Patient	ncit:#C16960
patient id	patient id	ncit:C164337
gender	gender	roo: P100018
age	age	roo: P100042
scan	scan	roo: P100284
small cell lung carcinoma (SCLC)	small cell lung carcinoma (SCLC)	roo: P100021
non small cell lung carcinoma (NSCLC)	non small cell lung carcinoma (NSCLC)	ncit:C2926
mesh volume	mesh volume	ro: RNU0

Supplementary table 1: table shows the mapping of local concept with ontology concept
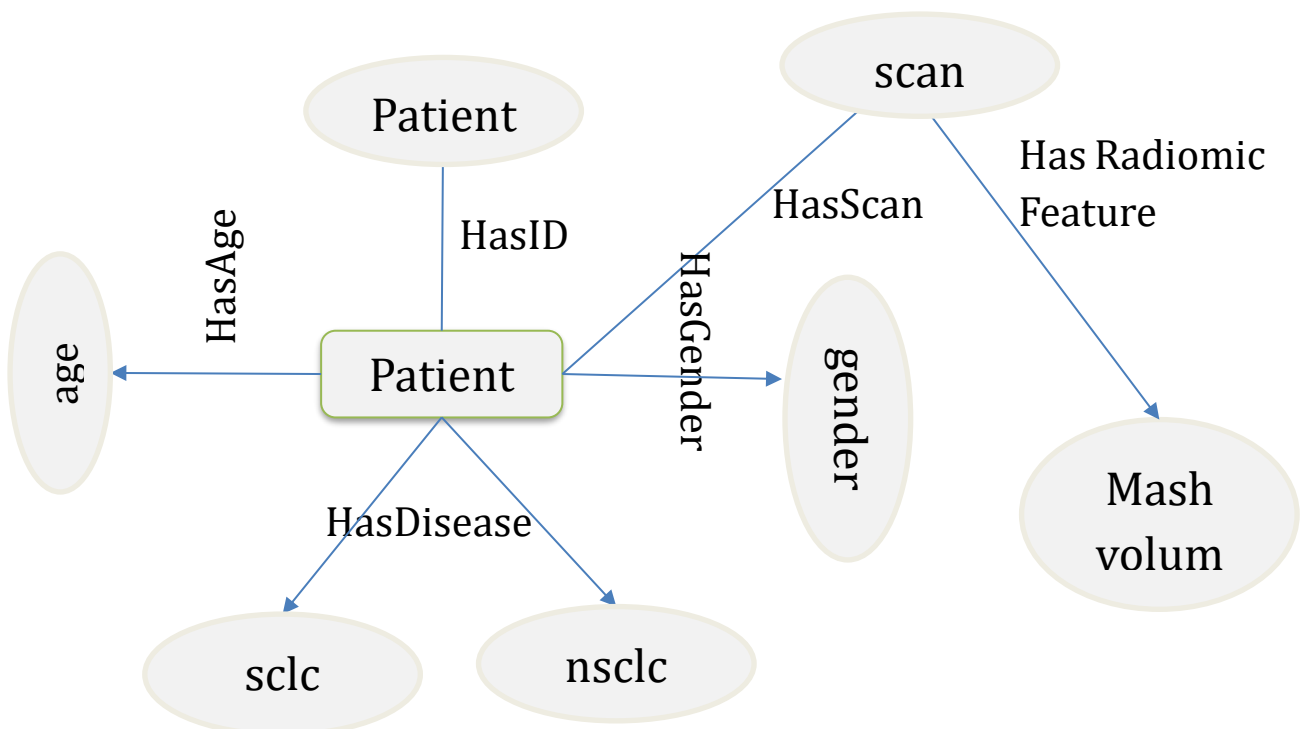
Supplementary figure S1: Overview of the mapping based on ROO/NCIT/RO id to the local concept.

Anonimization of patient personal information

We have followed the HIPAA Privacy Rule to deidentify protected *health information* (PHI). Following patient identifiers were deidentified using manual methods. Deidentification of imaging data was performed on image processing workstations.

Patient identifiers: Patient Name, Patient number, Address, Telephone/Mobile numbers, Email addresses, Social security numbers, Medical record numbers, Biometric identifiers, Health plan beneficiary numbers, Full-face photographs and any comparable images, Account numbers

Textbox S1. Example of a SPARQL query linking clinical factors (age and biological sex) with radiomics features (intensity histogram feature 10th percentile and shape feature mesh volume).

```
prefix ns1: http://www.radiomics.org/RO/
prefix rdf: http://www.w3.org/1999/02/22-rdf-syntax-ns#
prefix rdfs: http://www.w3.org/2000/01/rdf-schema#
prefix xml: http://www.w3.org/XML/1998/namespace
prefix xsd: http://www.w3.org/2001/XMLSchema#
PREFIX ncit: http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#
prefix roo: http://www.cancerdata.org/roo/

select ?patient_id ?rad_feature1 ?value1 ?rad_feature2 ?value2 ?age ?sex
where
{
?patient_idrdf:type ncit:C16960; # patient id is anncit type person
roo:P100000 ?ac1. # Linking patient's age
?ac1 roo:P100042 ?age.
?patient_id roo:P100018 ?sex1. # Linking patient's sex
?sex1 roo:local_value ?sex.
?patient_id roo:100284 ?scan.  # Linking patient's scan
?scanrdf:type ncit:C17999;
ns1:0310 ?imageVolume.

?imageVolumerdf:type ns1:0271;  # Linking radiomics image volume
ns1:0298 ?imageSpace.
?imageSpacerdf:type ns1:0225;
ns1:0296 ?rad_feature1.

?rad_feature1ns1:010191 ?value1.  #Retrieving associated feature value
?rad_feature1rdf:type ns1:GPMT. # Selecting feature set from class id "GPMT" i.e.
"firstorder_10Percentile"
filter( regex(str(?rad_feature1), "wavelet-LHH_firstorder_10Percentile" )) # Filtering
by the radiomics feature by name.
```

```
?rad_feature2ns1:010191 ?value2.#Retrieving associated feature value
?rad_feature2 rdf:type ns1:RNU0.#Selecting feature from class id "RNU0" i.e. "mesh volume"
filter( regex(str(?patient_id), "BW" ))      # Filtering by presence of text in case number
}

limit 5#arbitrarily fivesubjects purely as a demonstration in this example
```

	patient_id	rad_feature1	value1	rad_feature2	value2	age	sex
1	http://localhost/data/patient/BZ_05121	http://localhost/data/feature/BZ_05121/wavelet-LHH_firstorder_10Percentile	-49.739691003801944	http://localhost/data/feature/BZ_05121/original_shape_MeshVolume	15848.666666666666	"48"^^xsd:integer	Female
2	http://localhost/data/patient/BZ_05420	http://localhost/data/feature/BZ_05420/wavelet-LHH_firstorder_10Percentile	-66.0156059943281	http://localhost/data/feature/BZ_05420/original_shape_MeshVolume	1032.3333333333333	"65"^^xsd:integer	Male
3	http://localhost/data/patient/BZ_01059	http://localhost/data/feature/BZ_01059/wavelet-LHH_firstorder_10Percentile	-36.895799847527826	http://localhost/data/feature/BZ_01059/original_shape_MeshVolume	16305.0	"56"^^xsd:integer	Female
4	http://localhost/data/patient/BZ_09354	http://localhost/data/feature/BZ_09354/wavelet-LHH_firstorder_10Percentile	-33.37095489665515	http://localhost/data/feature/BZ_09354/original_shape_MeshVolume	15562.333333333334	"66"^^xsd:integer	Male
5	http://localhost/data/patient/BZ_00173	http://localhost/data/feature/BZ_00173/wavelet-LHH_firstorder_10Percentile	-41.990933805104675	http://localhost/data/feature/BZ_00173/original_shape_MeshVolume	3361.3333333333335	"54"^^xsd:integer	Male

Supplementary Figure S2. Example output of a data integration query based on the SPARQL language, on the internal RDF data storage ("endpoint"), illustrating the returned output from the graphical user query interface of the search command given in Textbox S1.

Chapter 8: Radiomics Signature: A Potential Imaging Biomarker for the Prediction of Overall Survival in cervical cancer

Abstract

Background: The role of artificial intelligence and radiomics in prediction model development in cancer has been increasing every passing day. Cervical cancer is the 4th most common cancer in women worldwide contributing 6.5% of all cancer types. The treatment outcome of cervical cancer patients varies and individualized prediction of disease outcome is of paramount importance.

Purpose: Purpose of this study is to develop and validate the digital signature for 5-years overall survival prediction in cervical cancer using robust CT radiomic and clinical features.

Materials and Methods: Pretreatment clinical features and CT radiomic features of 68 patients, who were treated with chemoradiation therapy in our hospital were used in this study. Radiomic features were extracted using an in-house developed python script and pyradiomic package. Clinical features were selected by the recursive feature elimination technique. Whereas radiomic feature selection was performed using a multi-step process i.e., step-1: only robust radiomic features were selected based on our previous study, step-2: a hierarchical clustering was performed to eliminate feature redundancy, and as final step-3: recursive feature elimination was performed to select the best features for prediction model development. Four machine algorithms i.e., Logistic regression (LR), Random Forest (RF), Support vector classifier (SVC), and Gradient boosting classifier (GBC) were used to develop 24 models (six models using each algorithm) using clinical, radiomic and combined features. Models were compared based on the prediction score in the internal validation.

Results: The average prediction accuracy was found to be 0.65±0.05, 0.72±0.09, 0.77±0.05 for clinical, radiomic, and combined models respectively. The average prediction accuracy was found to be 0.69±0.07, 0.79±0.07, 0.71±0.09, 0.72±0.06 for LR, RF, SVC and GBC models respectively.

Conclusion: Our study shows the strong correlation between robust radiomic signature and 5-year overall survival in cervical cancer patients.

Key words: cervices, cervical cancer, FIGO, prediction model, radiomics, deep learning

Introduction

Cancer is one of the fatal diseases and considered as the second most lethal disease across the world [1]. As per Global Cancer Statistics 2020 (GLOBOCAN 2020), cervical cancer is 4th commonest cancer worldwide, 6th commonest cancer in developed countries and 2nd commonest cancer in developing countries in female population [2, 3]. The cervical cancer related mortality rate among women varies across the globe and there is a distinct difference in developed and developing countries [2-4]. Breast and cervical cancer are the leading causes of cancer death in 103 and 42 countries, whereas lung cancer is leading causes of cancer death in 28 countries [1-4]. In developing countries, cervical cancer is the second leading cause of cancer related death, whereas in developed countries it is the sixth leading cause [1-7]. Cervical cancer management has been approached at two fronts i.e. prevention or early detection of cervical cancer by implementing screening programs and treatment of cervical cancer using evidence based medicine [8-14]. The incidence of cervical cancer in developed countries has reduced to half between 1972 to 2018 [8-10]. The reason for reduced incidence and mortality rate due to cervical cancer in developed countries can be attributed to effective implementation of cervical cancer screening and HPV vaccination programs. Due to availability of several new technologies or advancement in existing technology like CT, PET/CT, ultrasound and MRI has led to early diagnosis and better staging of the disease leading to improvement in overall survival and quality of life index [11-14]. The staging of cervical cancer is very complex and technology demanding. The staging system developed by the International Federation of Obstetrics and Gynecology (Fédération Internationale de Gynecologie et d'Obstetrique, or FIGO) is used for cervical cancer. Bhatla N. et.al. has published the recently revised FIGO staging of carcinoma of the cervix uteri to differentiate the various stages and substages of the disease [15]. Improvement in diagnostic accuracy due to the implementation of newer technologies like PET/CT, MRI, transvaginal ultrasound has improved cervical cancer staging and treatment in the last few years. As conventional treatment has a very low response rate of around 20 to 30 percent that proves; one size fits for all principle usually doesn't work in cancer management [15-17]. In the last few years, diagnostic modalities like immunohistochemistry (IHC), genetic profiling, tumor marker studies have established the fact that there are variations in disease in the same disease in different patients [18]. Hence, cancer treatment is gradually shifting towards personalized treatment or tailored treatment and replacing conventional treatment [19]. With the growing use of various computer-aided technologies in oncology in the last decade, the utilization of these technologies has taken the forefront in cancer management worldwide [20]. These technologies are being utilized for diagnosis, treatment planning, interim evaluation, and follow-up of the disease. In the last few years as the effort is being taken to provide personalized treatment to the patients, the ability of these technologies is being tested to predict treatment outcome, toxicity profile, and

treatment selection for patients. Utilization of available technologies like machine learning, radiomics, genomics, etc. for enabling the personalized treatment, especially those at high risk and are responding very poorly to standard treatment protocols, is of great interest for the clinicians [20]. Such a technological-driven system has shown promising results in the selection or modification of treatment plans, to improve the treatment outcome [ref needed]. Major types of ML techniques include Decision Tree (DT), Support Vector Machine (SVM), Artificial Neural Networks (ANN), Naïve Bayesian Classifier (BC), Bayesian Network (BN), K-Nearest Neighbor (KNN) and Random Forest (RF) have been used for nearly three decades in cancer detection [21-25]. In cancer prediction modeling, the main three predictive tasks are the prediction of cancer susceptibility, the prediction of cancer recurrence/ metastasis, and the prediction of survival. Several such technology-driven prediction models have been developed, tested, and utilized in the last decade in screening programs and the treatment of cervical cancer [26-39]. However, several prediction models have been developed using clinical and radiomics features predicting survival outcomes but the stability of radiomic features has been questioned by many researchers. In our earlier study, we have performed a detailed stability study of CT radiomic features and found around 100 robust radiomic features. In this study, we have tried to find the prediction capability of robust radiomic features with and without clinical features in predicting 5-year overall survival. This study is also the first of this kind from India.

Methods

Patient demographics:
The study was approved by the institutional ethics committee as a retrospective study with a waiver of consent. In total 68 patients were included in this study and had ages ranging 25–86 years (median: 50 years), at the time of diagnosis. All patients were diagnosed with cervical cancer between 2005 and 2009 and treated with definitive chemoradiotherapy or concomitant chemo and radiation therapy were included in this study. External beam radiation therapy (EBRT) dose range between 43.2 and 60.4 Gy (median=50Gy) was considered as radiotherapy procedures. Disease staging was performed according to the International Federation of Gynecology and Obstetrics (FIGO) classification. The numbers of patients in various FIGO stages in this cohort of patients are provided in table 1. The majority of the patients (85%) had squamous cell carcinoma and only a few patients (15%) had other histology. From diagnosis to the last follow-up, the meantime was 72 months. In our study, we have aimed to establish the correlation between radiomics/clinical features and overall survival. The initial characteristics of the study population are given in table 2.

Characteristics	Patients
Sample size	68
Age	56 (Range: 45-72)
Sex	
Male	0
Female	68/68(100%)
Tumor type	
Cervix cancer	69/69 (100%)
FIGO stages	
Stage 3	20
Stage 4	48
Pelvic Node	
Yes	42
No	26
Retroperitoneal Node	
Yes	58
No	10
Surgery	
Yes	15
No	53

Table 1: Demographic details of the study population.

20 clinical pathological and radiological features were extracted from electronic health records as approved by the hospital ethics committee. Pretreatment PET/CT scans were

also downloaded from the PACS for radiomic extraction. 1093 CT radiomics features were extracted from the CT series of PET/CT scans.

PET/CT imaging procedure

All of the baseline PET/CT scans were performed using Gemini TF16 or Gemini TF64 PET/CT scanners (Philips Medical Systems, Netherlands). F-18 FDG radiopharmaceutical was administered to the patient as per institutional protocol i.e. 4-5MBq/kg body weight after 6hrs of fasting. Scans were performed between 60 min to 100min after administration of the radiopharmaceutical.

Contrast-enhanced CT scans were performed after the injection of 60 to 80 ml of non-ionic contrast using the protocol mentioned in supplementary table 1. CT images were reconstructed using the Filtered back project (FBP) reconstruction algorithm.

Radiomic extraction:

DICOM images of PET/CT scan were downloaded on Philips Intellispace Discovery (research-only build; Philips Medical System, Eindhoven, The Netherlands) from PACS. The tumor was contoured using 3D contouring software installed on Intellispace Discovery by a 15year experienced medical physicist and checked & approved by a 30 year experienced nuclear medicine physician. The contours were saved as RTStructure by the name of GTV. Subsequently, the image and GTV were transferred to the research computer for radiomic extraction.

Images and GTV were converted into NRRD format using Plastmatch software [41]. Thereafter following pre-processing steps were applied using an in-house developed python script and the Pyradiomics package [42] for radiomic extraction. *Resampling:* Images were resampled using a 2 x 2 x 2 mm cube isotropic voxel. *Filtering and transformation of image:* From original images three sets of filtered images were produced applying Laplacian of Gaussian (LoG) filters with 1, 2, and 3mm sigma values. We also generated 8 sets of wavelet transformed images using eight combinations of high pass and low pass wavelet filters [42-44].

A total of 1093 radiomic features were extracted from 12 sets of images (1set original images, 3 sets LoG images, 8 sets of Wavelet Images) and corresponding GTVs [42].

Prediction algorithm used:

The commonly used machine learning algorithms for classification problems i.e., Logistic regression (LR), Random Forest classifier (RF), Gradient boosting classifier (GBC), and Support vector classifier (SVC) were used for prediction model development.

Feature selection:

The multi-step process was adopted for feature selection in this study. The following subsections describe the various methods adopted for feature selection. The steps utilized for feature selection are summarized in figure 1.

Clinical features selection:

Considering the completeness of data 13 clinical features were selected for further processing. Spearman correlation test was performed to find correlating features and reduce the redundancy among the features. The association of clinical features with outcome i.e., 5years overall survival (OS) was carried out using a t-test. Finally, recursive feature elimination (RFE) methods using logistic regression (RFE-LR) and random forest (RFE-RF) were applied to select two sets of features for prediction model development.

Radiomic Feature selection:

We opted for a two-steps process to select the best radiomic features for OS prediction out of 1093 radiomic features extracted from CT images. In the first step of feature selection, we included 121 stable radiomic features for the next step of feature selection based on our earlier radiomic stability study [53]. In the second step of feature selection, we performed a Spearman correlation test to identify redundant features followed by recursive feature elimination (RFE) methods using logistic regression (RFE-LR) and random forest (RFE-RF) were applied to select two sets of features for the prediction model development.

Combined (Clinical + Radiomic) features selection:

The top 7 clinical features and top 15 radiomic features which were identified in clinical and radiomic feature selection steps were used to select the best features for the combined model. Recursive features selection (RFE) methods using logistic regression (RFE-LR) and random forest (RFE-RF) were applied to select two sets of features for prediction model development.

Features selected using random forest model were used to develop models using random forest (RF) Support vector classifier (SVC) and Gradient Boosting and feature selected using logistic regression (LR) was used to develop logistic regression model.

Nested Cross-Validation:

Nested cross-validation was performed on the entire dataset using 7 outer and 6 inner loops were used for tuning the hyperparameters of the models [54]. Finally, a random train-test split (in 7:3 ratio) of data was performed and a prediction model was developed and validated.

Data balancing

After the train-test split, the training dataset was used to develop the prediction models with and without balancing the train data set for survival outcomes. Data balancing was performed by using minority oversampling. Validation was performed using the test data set without balancing the data.

Model development:

A total of 24 prediction models were developed using the aforementioned four prediction algorithms, three data sets with and without balancing the train data sets (table 1, figure 1).

ML algorithm	Feature used	Train data balancing	Prediction Model
Random Forest	Clinical	With	RF-Clinical-B
	Clinical	Without	RF-Clinical
	Radiomics	With	RF-Radiomics-B
	Radiomics	Without	RF-Radiomics
	Clinical + Radiomics	With	RF-Combined-B
	Clinical + Radiomics	Without	RF-Combined
Gradient Boosting	Clinical	With	GB-Clinical-B
	Clinical	Without	GB-Clinical
	Radiomics	With	GB-Radiomics-B
	Radiomics	Without	GB-Radiomics
	Clinical + Radiomics	With	GB-Combined-B
	Clinical + Radiomics	Without	GB-Combined
Support Vector Classifier	Clinical	With	SV-Clinical-B
	Clinical	Without	SV-Clinical
	Radiomics	With	SV-Radiomics-B
	Radiomics	Without	SV-Radiomics
	Clinical + Radiomics	With	SV-Combined-B
	Clinical + Radiomics	Without	SV-Combined
Logistic Regression	Clinical	With	LR-Clinical-B
	Clinical	Without	LR-Clinical
	Radiomics	With	LR-Radiomics-B
	Radiomics	Without	LR-Radiomics
	Clinical + Radiomics	With	LR-Combined-B
	Clinical + Radiomics	Without	LR-Combined

Table: 1: the table shows the prediction models developed in various combination
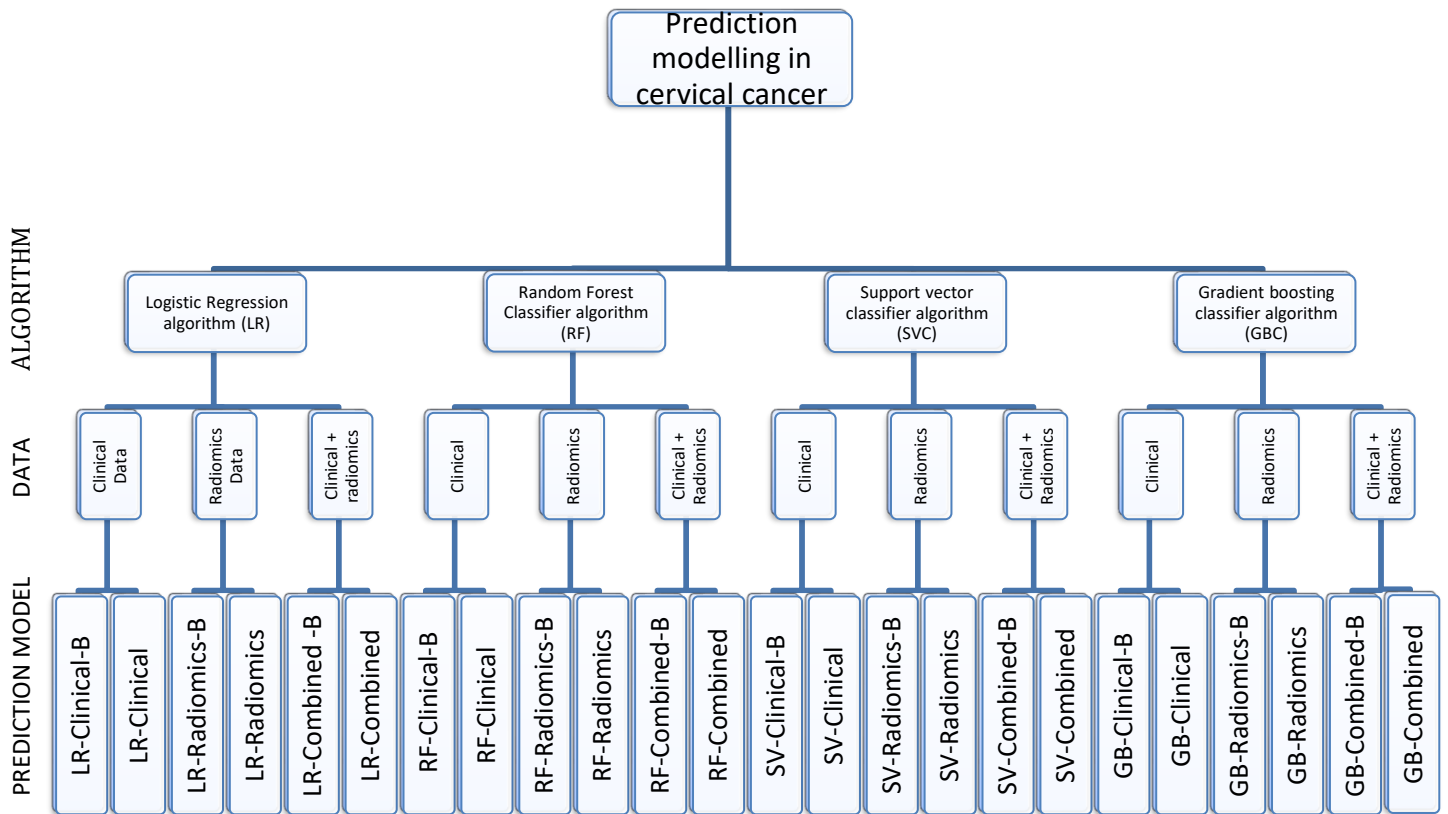
Figure 1: the figure shows our algorithm to develop 24 prediction models using various combinations. In the model's name "-B" indicates the model developed using balanced train data set.

Model evaluation and selection

All the developed models were evaluated by plotting the area under the receiver operator curve (AUC) to graphically represent the association between the features and the outcome i.e., 5-year overall survival in validation set. The best model was selected based on the performance score of each model in the validation set.

Statistical analysis

Statistical analyses were performed using R (v3.5.2, the R foundation for statistical computing, Vienna, Austria) or Python 3.2 software. Prediction model development and validation of model were performed using python 3.2 software.

Results

Given In total 68 patients were selected for this study who fulfilled the criteria of completeness of data sets. The details of data collection are provided in figure 2.
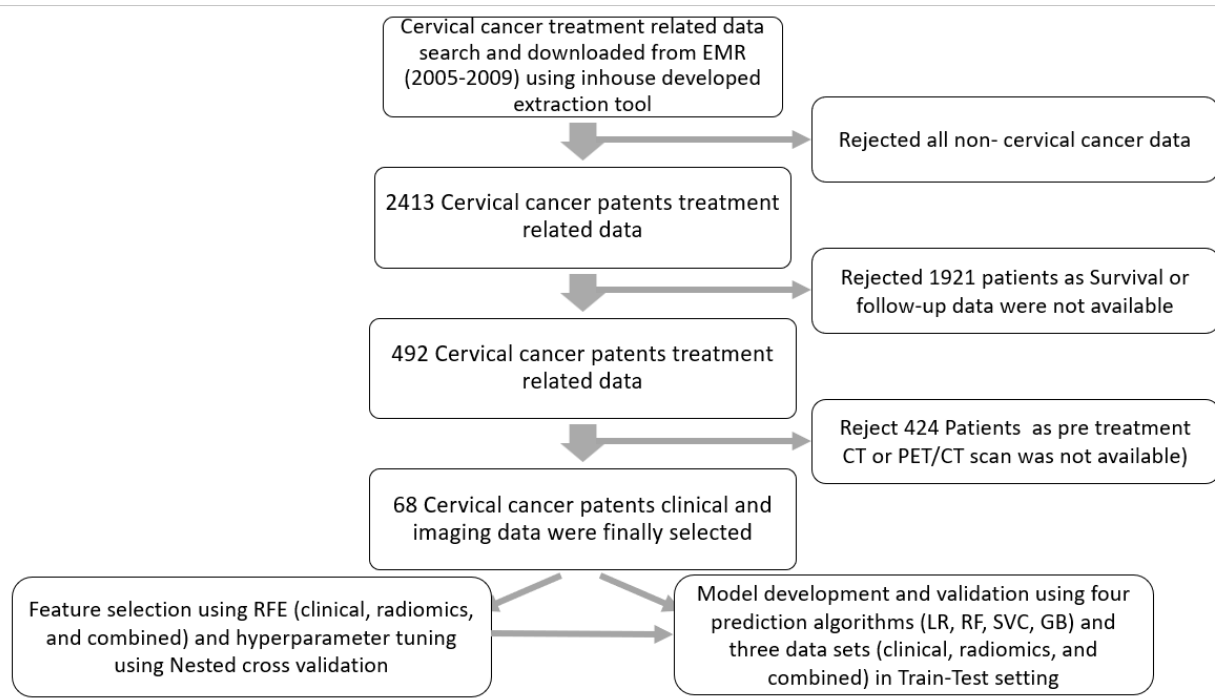
Figure 2: figure shows process of data collection and prediction model development

Feature selection
Clinical:

In total 13 clinical and radiological features were used for this study. The Spearman correlation test shows no strong correlation among the features. The maximum value of R2 was found to be 0.17 between age and retroperitoneal node. Spearman correlations among the features are shown in figure 2. Recursive feature elimination (RFE) was performed using logistic regression and random forest algorithms. Total 5 clinical features were found significant for each algorithm independently (table 3) (figure 4).
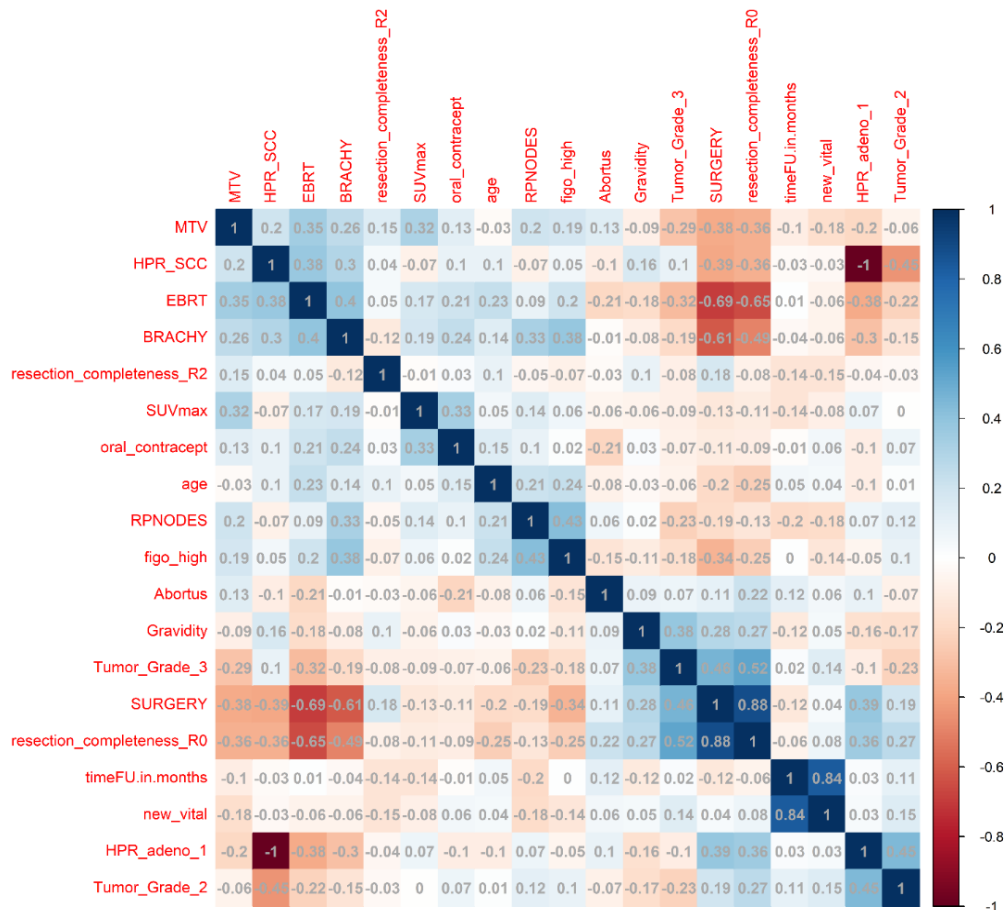
Figure 3: Figure shows the spearman correlation among the clinical features

Radiomics:

121 stable radiomics features based on our earlier study were included in this study [53]. Spearman correlation shows distinct 10 clusters (figure 3). Recursive feature elimination (RFE) was performed using logistic regression and random forest algorithms. In total 3 and 4 radiomic features were found to be significant for logistic regression and random forest algorithms respectively (table 3) (figure 4).

Figure 4: Figure shows the spearman correlation among the radiomic features showing clusters of features with positive and negative correlations

Combined (Clinical + Radiomics):

Among clinical and radiomic features selected independently, the most significant mixed features were selected using recursive feature elimination with logistic regression and random forest algorithms. Total 5 clinical+radiomics features were found significant for each of the algorithms separately (table 3) (figure 4).

Figure 5: figure shows feature importance in various combinations of algorithms and features. First row shows feature importance using random forest algorithm for clinical, radiomic and combined (clinical+radiomics) features; second row shows feature importance using logistic regression algorithm for clinical, radiomic and combined (clinical+radiomics) features.

(Abbreviations: OSFL= Original_shape_Flatness; WLNC: = Wavelet_LHL_ngtdm_Contrast; WLGI = Wavelet_LLL_glcm_Idn;OSM2DDR = Original_shape_Maximum2DDiameterRow; L3M3DFO10P = Log-sigma-3-0-mm-3D_firstorder_10percentile; L2M3DGRP= Log-sigma-2-0-mm-

3D_glrlm_RunPercentage; L2M3DGLFE = Log-sigma-2-0-mm-3D_glrlm_LongRunEmphasis)

Feature selection technique	Feature type	Number of features selected	Accuracy with selected features	Kappa value
Recursive Feature Elimination with Logistic regression	Clinical	5	0.69	0.31
	Radiomics	3	0.64	0.17
	Clinical + Radiomics	5	0.68	0.26
Recursive Feature Elimination with Random Forest	Clinical	5	0.68	0.32
	Radiomics	4	0.72	0.38
	Clinical + Radiomics	5	0.77	0.46

Table 3: Table shows the number of features selected, accuracy and Kappa value for various combination of data sets using multivariate recursive feature elimination with logistic regression and random forest.

Model development and validation:
Four algorithms i.e., Logistic regression (LR), Random Forest (RF), Support vector classifier (SVC) and gradient boost classifier (GBC) were used for prediction model development. total 24 prediction models using four prediction algorithms for clinical, radiomics and combined features.
Nested cross validation: nested cross validation performed for all the prediction algorithms for tuning there hyperparameters. The prediction algorithms along with best hyperparameters and validation scores are shown in table 4.

Algorithms	Features	Hyper parameters	Accuracy in nested cross validation
Logistic regression	Clinical	{c: '10', penalty: '12', solver: 'newton-cg'}	0.66(±0.17)
	Radiomics	{c: '10', penalty: '12', solver: 'liblinear'}	0.68(±0.09)
	Clinical + Radiomics	{c: '100', penalty: '12', solver: 'newton-cg'}	0.66(±0.06)
Random forest	Clinical	{bootstrap: 'true', criterion: 'gini', max_depth: '10', 'min_samples_leaf': 2,n_estimators: '80'}	0.66(±0.15)
	Radiomics	{'bootstrap': True, 'criterion': 'gini', 'max_depth': 25, 'min_samples_leaf': 2, 'n_estimators': 40}	0.79(±0.09)
	Clinical + Radiomics	{bootstrap: 'true', criterion: 'gini', max_depth: '10', 'min_samples_leaf': 2,n_estimators: '80'}	0.75(±0.07)
Support vector classifier	Clinical	{c:'1', gamma:'1', kernel:'linear'}	0.72(±0.11)
	Radiomics	{c:'0.1', gamma:'0.001', kernel:'rbf'}	0.74(±0.08)
	Clinical + Radiomics	{c:'100', gamma:'0.001', kernel:'rbf'}	0.67(±0.16)
Gradient boost classifier	Clinical	{learning_rate: '0.1', max_depth: '7', n_estimators: '60'}	0.75(±0.12)
	Radiomics	{'learning_rate': 0.1, 'max_depth': 7, 'n_estimators': 80}	0.75(±0.13)
	Clinical + Radiomics	{learning_rate: '1', max_depth: '3', n_estimators: '10'}	0.75(±0.09)

Table 4: This table shows the selected hyperparameters and nested cross validation scores of various models

All 24 models showed good prediction capability of 5year overall survival. The average accuracy and AUC in validation sets across all the 24-prediction models were found to be 0.73(±0.07) and 0.60(±0.11) respectively.

Logistic regression modeling:

The average accuracy and AUC for logistic regression models across six models developed with various combinations were found to be 0.69(±0.07) and 0.60(±0.05) respectively. The detailed validation scores are shown in table 5. Area under the receiver operator curves (AUC) of all the logistic regression models are shown in figure 5. Radiomics [accuracy: 0.76 (LR-Radiomics-B); 0.71 (LR-Radiomics)] or combined prediction [accuracy: 0.71 (LR-Combined-B); 0.76 (LR-Combined)] model models had better prediction capabilities in comparison of clinical models [accuracy: 0.61 (LR-Clinical-B); 0.61 (LR-Clinical)] developed with logistic regression algorithm.
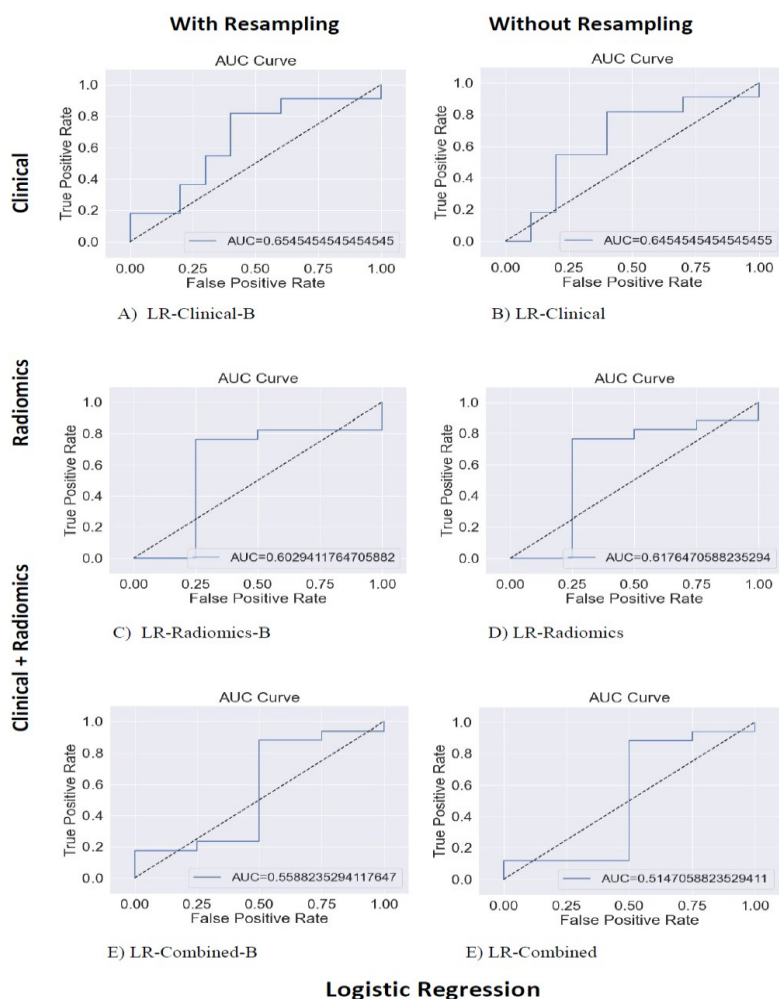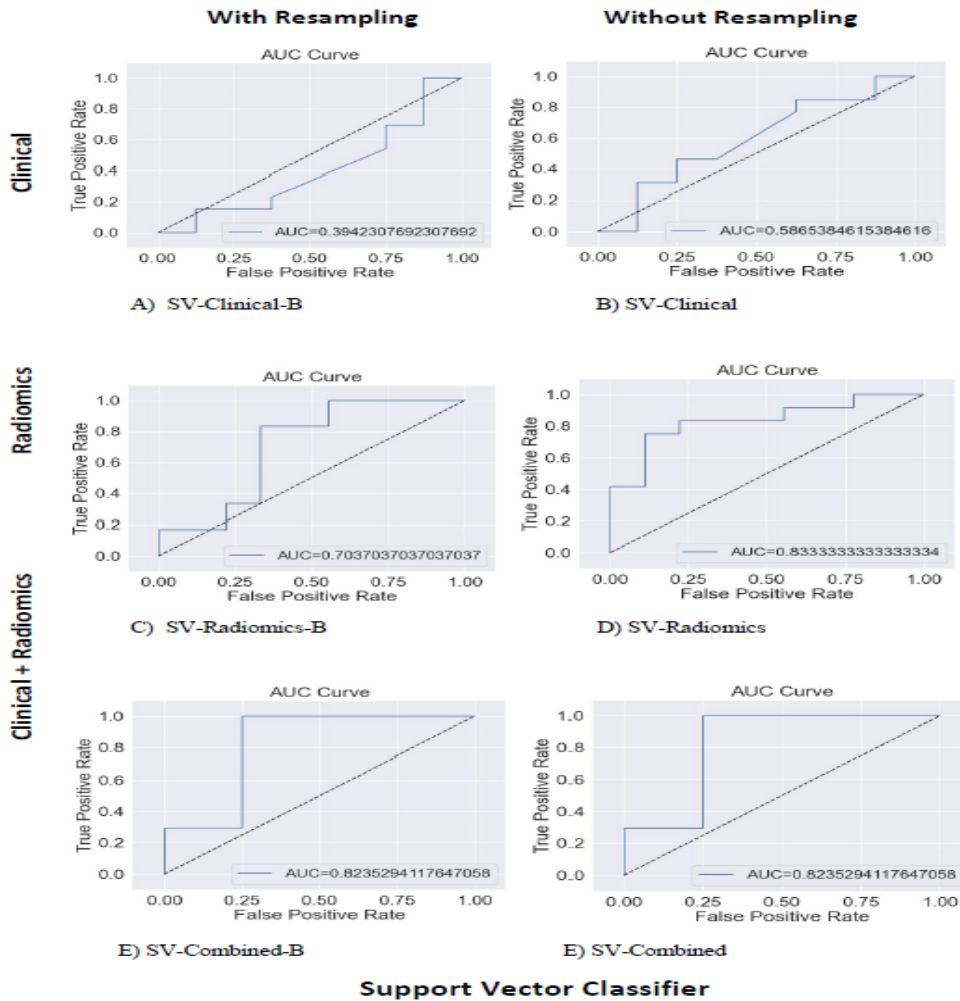

Figure 6: AUC curves of all the Logistic models are shown in this figure

Random forest modeling:

The average accuracy and AUC for random forest models were found to be 0.79±0.07 and 0.73±0.07 respectively. The detailed validation scores are shown in table 5. Area under the receiver operator curves (AUC) of all the Random Forest models is shown in figure 6. Radiomics [accuracy: 0.86 (RF-Radiomics-B); 0.81 (RF-Radiomics)] or combined prediction [accuracy: 0.81 (RF-Combined-B); 0.81 (RF-Combined)] model

models had better prediction capabilities in comparison of clinical models [accuracy: 0.67 (RF-Clinical-B); 0.76 (RF-Clinical)] developed with random forest algorithm.
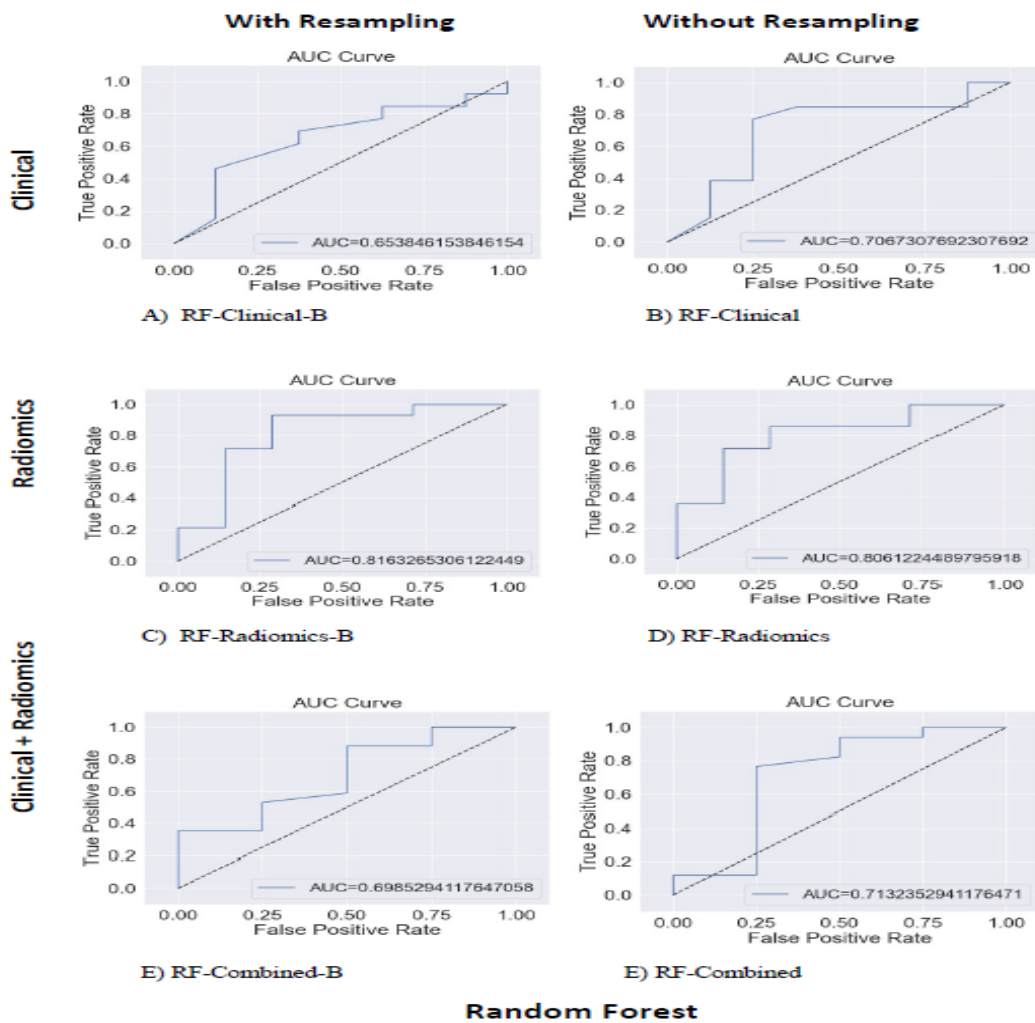

Figure 7: AUC curves of all the random forest models are shown in this figure

Support vector classifier (SVC) model:
The average accuracy and AUC for support vector models were found to be 0.71±0.08 and 0.69±0.18 respectively. The detailed validation scores are shown in table 5. Area under the receiver operator curves (AUC) of all the support vector classifier models is shown in figure 7. Radiomics [accuracy: 0.76 (SV-Radiomics-B); 0.71 (SV-Radiomics)] or combined prediction [accuracy: 0.76 (SV-Combined-B); 0.81 (SV-Combined)] model models had better prediction capabilities in comparison of clinical models [accuracy: 0.62 (SV-Clinical-B); 0.62 (SV-Clinical)] developed with support vector classifier algorithm.

Figure 8: AUC curves of all the SVC models are shown in this figure

Gradient boosting classifier (GBC) model:

The average accuracy and AUC for gradient boosting models were found to be 0.72±0.06 and 0.73±0.05 respectively. The detailed validation scores are shown in table 5. Area under the receiver operator curves (AUC) of all the Gradient busting classifier models are shown in figure 8. Radiomics [accuracy: 0.76 (GB-Radiomics-B); 0.76 (GB-Radiomics)] or combined prediction [accuracy: 0.76 (GB-Combined-B); 0.76 (GB-Combined)] model models had better prediction capabilities in comparison of clinical models [accuracy: 0.67 (GB-Clinical-B); 0.62 (GB-Clinical)] developed with gradient boosting algorithm.
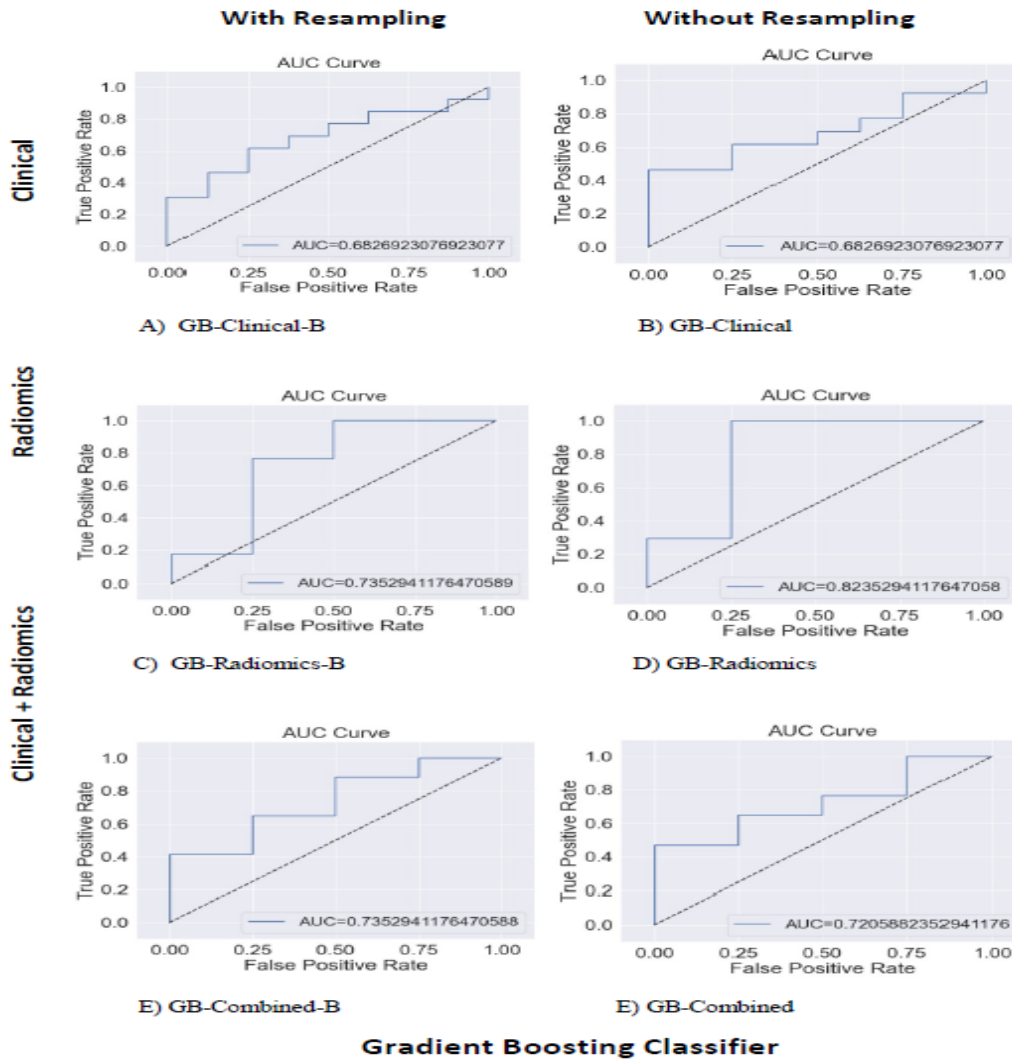
Figure 9: AUC curves of all the GBC models are shown in this figure

Feature Selection Function	ML algorithm	Prediction model	Accuracy	Precision	Recall	F1- Score	AUC
Logistic Regression	Logistic Regression	LR-Clinical-B	0.61	0.63	0.62	0.62	0.65
		LR-Clinical	0.61	0.63	0.62	0.6	0.65
		LR-Radiomics-B	0.76	0.83	0.76	0.78	0.60
		LR-Radiomics	0.71	0.77	0.71	0.73	0.62
		LR-Combined-B	0.71	0.77	0.71	0.73	0.56
		LR-Combined	0.76	0.78	0.76	0.77	0.51
Random Forest	Random Forest	RF-Clinical-B	0.67	0.7	0.67	0.67	0.65
		RF-Clinical	0.76	0.77	0.76	0.76	0.71
		RF-Radiomics-B	0.86	0.86	0.86	0.85	0.82
		RF-Radiomics	0.81	0.81	0.81	0.81	0.81
		RF-Combined-B	0.81	0.81	0.81	0.81	0.70
		RF-Combined	0.81	0.78	0.81	0.78	0.71

Support Vector Classifier	SV-Clinical-B	0.62	0.38	0.62	0.47	0.39
	SV-Clinical	0.62	0.38	0.62	0.47	0.59
	SV-Radiomics-B	0.76	0.76	0.76	0.76	0.70
	SV-Radiomics	0.71	0.74	0.71	0.69	0.83
	SV-Combined-B	0.76	0.83	0.76	0.78	0.82
	SV-Combined	0.81	0.85	0.81	0.82	0.82
Gradient Boosting	GB-Clinical-B	0.67	0.66	0.67	0.66	0.68
	GB-Clinical	0.62	0.6	0.62	0.61	0.68
	GB-Radiomics-B	0.76	0.83	0.76	0.78	0.74
	GB-Radiomics	0.76	0.83	0.76	0.78	0.82
	GB-Combined-B	0.76	0.78	0.76	0.77	0.74
	GB-Combined	0.76	0.74	0.76	0.75	0.72

Table 5: table shows accuracy, PPV, NPV, F1-score and AUC of all the models

Model Selection:

RF-Radiomics-B model had best prediction accuracy (C-index=0.86; AUC= 0.82) among all 24 models developed (figure 10). The average prediction accuracy (C-index) for clinical, radiomic, and combined models were found to be 0.65±0.05, 0.72±0.09, 0.77±0.05 respectively. The average prediction accuracy (C-index) for logistic regression, random forest, support vector classifier, and gradient boosting classifier models were found to be 0.69±0.07, 0.79±0.07, 0.71±0.09, 0.72±0.06 respectively.
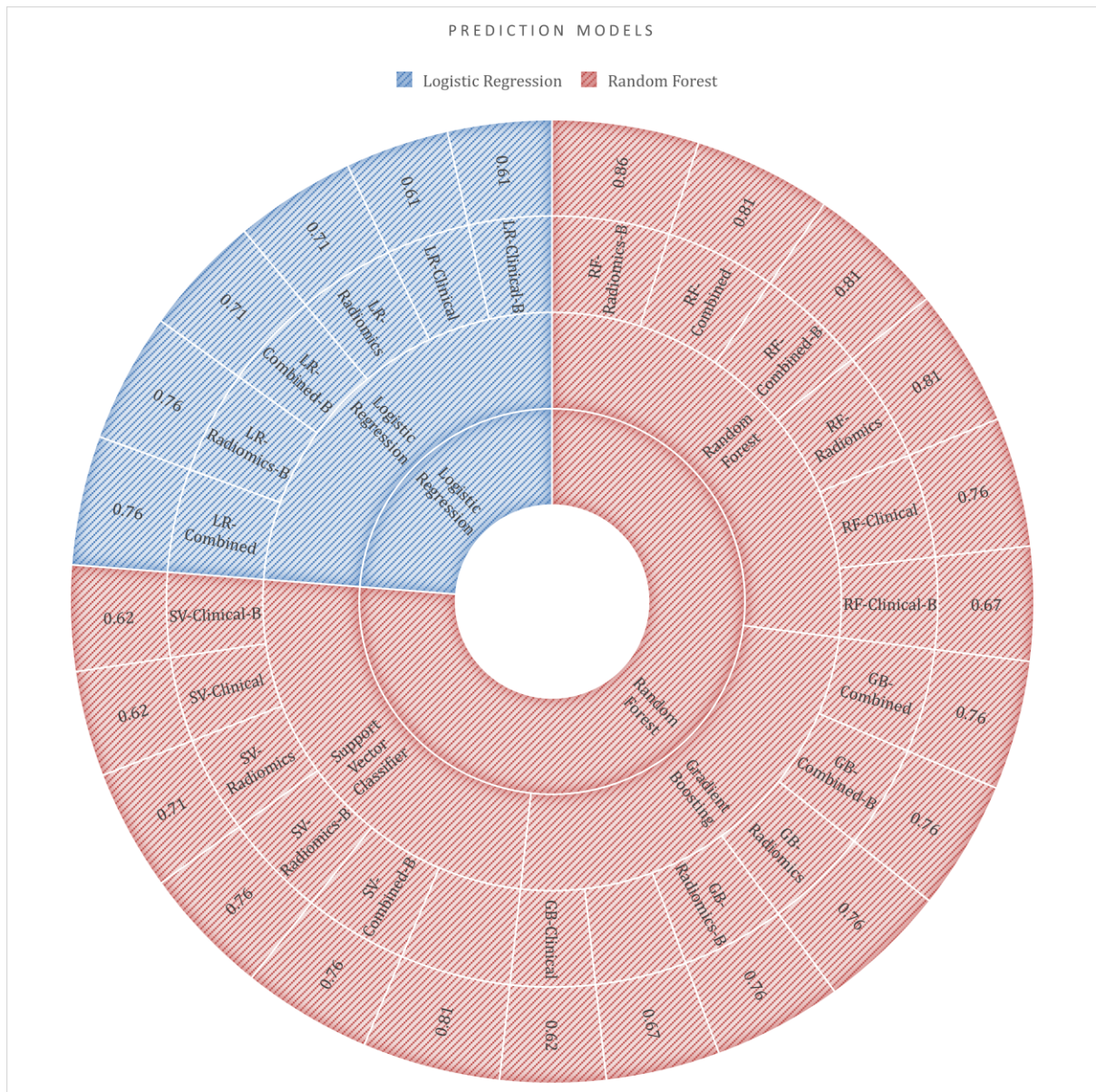
Figure 10: The figure shows (from inner circle to outer cycle) the algorithm used for feature election -> prediction algorithm -> prediction models -> corresponding prediction accuracy in the validation set.

Discussion

Our study shows the significance of radiomic features in generating statistical machine learning models for disease outcomes like 5year overall survival prediction in cervical cancer. With this study, we were able to identify the gap in the data archival system in our hospital related to medical image archives as well as other clinical data points as described in the result section. With this study, we were able to determine the most effective radiomic feature and their combination for the prediction of disease outcomes. A rigorous method of feature selection by applying various techniques has helped in this study to select the most efficient features which can become a digital signature for the stated disease outcome. We tested various prediction algorithms with radiomics and clinical features separately and in combination. In multivariate analysis with random forest, radiomic features were found to be better associated with disease outcome in our cohort. Our result was consistent with various other studies performed on cervical cancer outcome prediction. If we consider our study with other studies performed in this field our study design had similarities with others except, we tested several prediction algorithms to select best fits to our cohort. Our finding is consistent with other studies similar studies performed earlier [12, 14, 15, 26-39, 57-66]. Clinical features like age, presence or absence of retroperitoneal node, and peritoneal node FIGO stage at the time of diagnosis were also found to be prognostic markers in our study which was consistent with the published literature [12, 14, 15, 26-33, 57-60]. In univariate and multivariate analysis clinical features i.e., Age, FIGO stage, absence and presence of retroperitoneal node and peritoneal node, imaging features i.e., SUV MTV found an association with 5year overall survival, which was consistent with other published literature [12, 14, 29, 33, 34, 57, 61, 66]. Similarly in univariate and multivariate studies, radiomic features showed a significant association with 5year OS which is also consistent with published literature [28, 29, 34, 66]. As we had selected only stable radiomic features based on our earlier study [53] that shows the repeatable and reproducible radiomic features also show excellent prognostic and predictive value in cervical cancer. The effort of the radiomic community should be to identify the robust features and find out the predictive capabilities of those stable features in various disease groups for various prediction endpoints. Among various prediction models tested in our study, **RF-Radiomics-B** random forest model showed the best accuracy in nested cross-validation as well as the train-test final model outperforming all the prediction models used in our study. Whereas LR-Clinical-B and LR-Clinical the logistic regression models show the lowest accuracy in predicting overall survival in this study. When we compared the performance score of prediction models with radiomic, clinical and combined models, again random forest and gradient boosting models were at the top.

The average accuracy of clinical models with all the four prediction algorithms was less than that of radiomics and combined models which is similar to previously published

work [26-30]. The radiomic and combined model performance across all four prediction algorithms were found to be more or less similar. Our study also confirms the superiority of radiomic features over clinical features in predating overall survival in cervical cancer. Comparing the prediction algorithms, the random forest-based prediction models had better accuracy in comparison to the other three which affirms the findings of earlier published literature in cervical cancer [26, 27]. We did not find much difference between the models developed using with or without balanced train sets maybe because of event rate in our study was adequately balanced and did not require data balancing as an additional step. The radiomic community has been concerned about the stability of radiomic features. The radiomic community is skeptical about stable radiomic features' ability to predict outcomes [67]. This is probably the first study published on cancer prediction modeling using stable radiomic features independently or in combination with clinical features. In our study, we were able to show that radiomic features can be used for 5-year overall prediction in cervical cancer. This was also the first prediction modeling study to be conducted on cervical cancer patients in India. Other researchers in India will be motivated to conduct prediction modeling studies for evolving digital signatures of disease outcomes based on our study. Study type: single-center, small sample size, and no external or prospective validation are some of the limitations of this study. The future will involve repeating this study at our hospital with larger sample size, as well as initiating multicentric studies to develop a universally accepted model. It is the ultimate objective of this research to validate this model using prospective clinical trials and then implement decision support systems in clinics based on a validated predictive model with retrospective and prospective data.

Conclusion:

We have demonstrated in our study that robust radiomic features are predictive of overall survival for cervical cancer patients. According to this study, random forest prediction algorithms can predict better than other algorithms. The model's predictive ability is slightly improved by using data balancing. Although radiomic features are superior to clinical features in terms of prediction abilities, they are most effective when combined with clinical features. Overall, this study suggests the importance of radiomics and artificial intelligence in implementing decision support systems in the management of cervical cancer.

References

[1] Nagai H, Kim YH. Cancer prevention from the perspective of global cancer burden patterns. J Thorac Dis. 2017;9(3):448-451. doi:10.21037/jtd.2017.02.75

[2] Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021;71(3):209-249. doi:10.3322/caac.21660

[3] Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68(6):394-424.

[4] Bruni L, Albero G, Serrano B, Mena M, Gómez D, Muñoz J, Bosch FX, de Sanjosé S. ICO/IARC Information Centre on HPV and Cancer (HPV Information Centre). Human Papillomavirus and Related Diseases in India. Summary Report 17 June 2019., https://hpvcentre.net/statistics/reports/IND.pdf [Date Accessed]

[5] Siegel R, Naishadham D, Jemal A. Cancer statistics, 2013. Ca-a Cancer Journal for Clinicians. 2013;63(1):11-30.

[6] Cuzick J, Bergeron C, von Knebel DM, Gravitt P, Jeronimo J, Lorincz AT. Newtechnologies and procedures for cervical cancer screening. Vaccine.2012;30 Suppl 5:F107–16.

[7] Waggoner SE. Cervical cancer. Lancet 2003;361:2217-25.

[8] Catarino R, Petignat P, Dongui G, Vassilakos P. Cervical cancer screening in developing countries at a crossroad: Emerging technologies and policy choices. World J Clin Oncol. 2015;6(6):281-290. doi:10.5306/wjco.v6.i6.281

[9] Nagai H, Kim YH. Cancer prevention from the perspective of global cancer burden patterns. J Thorac Dis. 2017;9(3):448-451. doi:10.21037/jtd.2017.02.75

[10] R. Sankaranarayanan,Screening for Cancer in Low- and Middle-Income Countries,Annals of Global Health,Volume 80, Issue 5,2014,Pages 412-417,ISSN 2214-9996,https://doi.org/10.1016/j.aogh.2014.09.014.

[11] 13. American Cancer Society. Cancer Facts & Figures 2018. Atlanta: American Cancer Society; 2018.]. Available from:https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2018/cancer-facts-and-figures-2018.pdf[Date Accessed: 17 september, 2021]

[12] Choi J, Kim HJ, Jeong YH, Lee JH, Cho A, Yun M, et al. The role of (18) F-FDG PET/CT in assessing therapy response in cervix cancer after concurrent Chemoradiation therapy. Nucl Med Mol Imaging. 2014;48(2):130–6.

[13] Marcos Duarte Guimaraes, Alice Schuch, Bruno Hochhegger, Jefferson Luiz Gross, Rubens Chojniak, Edson Marchiori,Functional magnetic resonance imaging in oncology: state of the art, Radiol Bras. 2014 Mar-Apr; 47(2): 101–111.

[14] Czernin J, Allen-Auerbach M, Nathanson D, Herrmann K. PET/CT in Oncology: Current Status and Perspectives. CurrRadiol Rep. 2013;1:177–190.

[15] Neerja Bhatla , Jonathan S. Berek , Mauricio CuelloFredes , Lynette A. Denny , et.al. Revised FIGO staging for carcinoma of the cervix uteri, Int J GynecolObstet 2019; 145: 129–135

[16] Rose PG, Bundy BN, Watkins EB, et al: Concurrent cisplatin-based radiotherapy and chemotherapy for locally advanced cervical cancer. N EnglJ Med 340:1144-1153, 1999

[17] Thomas GM. Improved treatment for cervical cancer--concurrent chemotherapy and radiotherapy. N Engl J Med 1999;340:1198-200.

[18] Kori M, Yalcin Arga K. Potential biomarkers and therapeutic targets in cervical cancer: Insights from the meta-analysis of transcriptomics data within network biomedicine perspective. PLoS One. 2018;13(7):e0200717. Published 2018 Jul 18. doi:10.1371/journal.pone.0200717

[19] Cohen AC, Roane BM, Leath CA 3rd. Novel Therapeutics for Recurrent Cervical Cancer: Moving Towards Personalized Therapy. Drugs. 2020;80(3):217-227. doi:10.1007/s40265-019-01249-z

[20] Hamamoto R, Suvarna K, Yamada M, et al. Application of Artificial Intelligence Technology in Oncology: Towards the Establishment of Precision Medicine. Cancers (Basel). 2020;12(12):3532. Published 2020 Nov 26. doi:10.3390/cancers12123532

[21] Machine Learning in MATLAB available from: https://in.mathworks.com/help/stats/machine-learning-in-matlab.html,[Date Accessed: 17 March, 2020]

[22] Obermeyer Z, Emanuel EJ: Predicting the future—big data, machine learning, and clinical medicine. N Engl J Med 375:1216–12192016

[23] W. Raghupathi, V. Raghupathi, Big data analytics in healthcare: promise and potential, Health Inf Sci Syst, 2 (1) (2014), p. 3, 10.1186/2047-2501-2-3

[24] Chartrand, G. et al. Deep Learning : A Primer for Radiologists. RadioGraphics 37, 2113–2131 (2017).

[25] Erickson, B. J., Korfiatis, P., Akkus, P. & Kline, T. L. Machine Learning for Medical Imaging. RadioGraphics (2017)37, 505–515

[26] Tian X, Sun C, Liu Z, Li W, Duan H, Wang L, Fan H, Li M, Li P, Wang L, Liu P, Tian J, Chen C. Prediction of Response to Preoperative Neoadjuvant Chemotherapy in Locally Advanced Cervical Cancer Using Multicenter CT-Based Radiomic Analysis. Front Oncol. 2020 Feb 4;10:77

[27] Fang J, Zhang B, Wang S, Jin Y, Wang F, Ding Y, Chen Q, Chen L, Li Y, Li M, Chen Z, Liu L, Liu Z, Tian J, Zhang S. Association of MRI-derived radiomic biomarker with disease-free survival in patients with early-stage cervical cancer. Theranostics. 2020 Jan 16;10(5):2284-2292

[28] Wang T, Gao T, Yang J, Yan X, Wang Y, Zhou X, Tian J, Huang L, Zhang M. Preoperative prediction of pelvic lymph nodes metastasis in early-stage cervical cancer using radiomics nomogram developed based on T2-weighted MRI and diffusion-weighted imaging. Eur J Radiol. 2019 May;114:128-135

[29] Lucia F, Visvikis D, Vallières M, Desseroit MC, Miranda O, Robin P, Bonaffini PA, Alfieri J, Masson I, Mervoyer A, Reinhold C, Pradier O, Hatt M, Schick U. External validation of a combined PET and MRI radiomics model for prediction of recurrence in cervical cancer patients treated with chemoradiotherapy. Eur J Nucl Med Mol Imaging. 2019 Apr;46(4):864-877.

[30] Altazi BA, Fernandez DC, Zhang GG, Hawkins S, Naqvi SM, Kim Y, Hunt D, Latifi K, Biagioli M, Venkat P, Moros EG. Investigating multi-radiomic models for enhancing prediction power of cervical cancer treatment outcomes. Phys Med. 2018 Feb;46:180-188.

[31] Chen J, Chen H, Zhong Z, Wang Z, Hrycushko B, Zhou L, Jiang S, Albuquerque K, Gu X, Zhen X. Investigating rectal toxicity associated dosimetric features with deformable accumulated rectal surface dose maps for cervical cancer radiotherapy. Radiat Oncol. 2018 Jul 6;13(1):125

[32] Rose PG, Java J, Whitney CW, Stehman FB, Lanciano R, Thomas GM, DiSilvestro PA. Nomograms Predicting Progression-Free Survival, Overall Survival, and Pelvic Recurrence in Locally Advanced Cervical Cancer Developed From an Analysis of Identifiable Prognostic Factors in Patients From NRG Oncology/Gynecologic Oncology Group Randomized Trials of Chemoradiotherapy. J Clin Oncol. 2015 Jul 1;33(19):2136-42

[33] Reuzé S, Orlhac F, Chargari C, Nioche C, Limkin E, Riet F, Escande A, Haie-Meder C, Dercle L, Gouy S, Buvat I, Deutsch E, Robert C. Prediction of cervical cancer recurrence using textural features extracted from 18F-FDG PET images acquired with different scanners. Oncotarget. 2017 Jun 27;8(26):43169-43179.

[34] Lucia F, Visvikis D, Desseroit MC, Miranda O, Malhaire JP, Robin P, Pradier O, Hatt M, Schick U. Prediction of outcome using pretreatment 18F-FDG PET/CT and MRI radiomics in locally advanced cervical cancer treated with chemoradiotherapy. Eur J Nucl Med Mol Imaging. 2018 May;45(5):768-786

[35] Shim SH, Kim DY, Lee SJ, Kim SN, Kang SB, Lee SW, Park JY, Suh DS, Kim JH, Kim YM, Kim YT, Nam JH. Prediction model for para-aortic lymph node metastasis in patients with locally advanced cervical cancer. Gynecol Oncol. 2017 Jan;144(1):40-45.

[36] Kong TW, Kim J, Son JH, Kang SW, Paek J, Chun M, Chang SJ, Ryu HS. Preoperative nomogram for prediction of microscopic parametrial infiltration in patients with FIGO stage IB cervical cancer treated with radical hysterectomy. Gynecol Oncol. 2016 Jul;142(1):109-114

[37] Kim DY, Shim SH, Kim SO, Lee SW, Park JY, Suh DS, Kim JH, Kim YM, Kim YT, Nam JH. Preoperative nomogram for the identification of lymph node metastasis in early cervical cancer. Br J Cancer. 2014 Jan 7;110(1):34-41

[38] Kumar S, Rana ML, Verma K, Singh N, Sharma AK, Maria AK, Dhaliwal GS, Khaira HK, Saini S. PrediQt-Cx: post treatment health related quality of life prediction model for cervical cancer patients. PLoS One. 2014 Feb 26;9(2):e89851

[39] Kang S, Nam BH, Park JY, Seo SS, Ryu SY, Kim JW, Kim SC, Park SY, Nam JH. Risk assessment tool for distant recurrence after platinum-based concurrent

chemoradiation in patients with locally advanced cervical cancer: a Korean gynecologic oncology group study. J Clin Oncol. 2012 Jul 1;30(19):2369-74

[40] Jha AK, Mithun S, Singh AM, et al. 18-Month Performance Assessment of Gemini TF 16 PET/CT System in a High-Volume Department. J Nucl Med Technol. 2016;44(1):36-41. doi:10.2967/jnmt.115.168492

[41] Plastimatch documentation, https://plastimatch.org/contents.html [accessed on 23-09-2021]

[42] Pyradiomics, https://pyradiomics.readthedocs.io/en/latest/index.html[accessed on 23-09-2021]

[43] Griethuysen, J. J. M., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., Beets-Tan, R. G. H., Fillon-Robin, J. C., Pieper, S., Aerts, H. J. W. L. Computational Radiomics System to Decode the Radiographic Phenotype. Cancer Research, 2017 77(21), e104–e107. `https://doi.org/10.1158/0008-5472.CAN-17-0339 https://doi.org/10.1158/0008-5472.CAN-17-0339

[44] Zwanenburg, A., Leger, S., Vallières, M., and Löck, S. (2016). Image biomarker standardisation initiative - feature definitions. In eprint arXiv:1612.07003 [cs.CV]

[45] Nick TG, Campbell KM. Logistic regression. Methods Mol Biol. 2007;404:273-301. doi:10.1007/978-1-59745-530-5_14

[46] Stoltzfus JC. Logistic regression: a brief primer. Acad Emerg Med. 2011;18(10):1099-1104. doi:10.1111/j.1553-2712.2011.01185.x

[47] Rigatti SJ. Random Forest. J Insur Med. 2017;47(1):31-39. doi:10.17849/insm-47-01-31-39.1

[48] Breiman, L. Random Forests. Machine Learning 45, 5–32 (2001). https://doi.org/10.1023/A:1010933404324

[49] Nedaie A, Najafi AA. Support vector machine with Dirichlet feature mapping. Neural Netw. 2018;98:87-101. doi:10.1016/j.neunet.2017.11.006

[50] Chao CF, Horng MH. The construction of support vector machine classifier using the firefly algorithm. ComputIntellNeurosci. 2015;2015:212719. doi:10.1155/2015/212719

[51] Natekin A, Knoll A. Gradient boosting machines, a tutorial. Front Neurorobot. 2013;7:21. Published 2013 Dec 4. doi:10.3389/fnbot.2013.00021

[52] Friedman J. (2001). Greedy boosting approximation: a gradient boosting machine. Ann. Stat. 29, 1189–1232 10.1214/aos/1013203451

[53] Jha AK, Mithun S, Jaiswar V, et al. Repeatability and reproducibility study of radiomic features on a phantom and human cohort. Sci Rep. 2021;11(1):2055. Published 2021 Jan 21. doi:10.1038/s41598-021-81526-8

[54] Parvandeh S, Yeh HW, Paulus MP, McKinney BA. Consensus features nested cross-validation. Bioinformatics. 2020;36(10):3093-3098. doi:10.1093/bioinformatics/btaa046

[55] Yang J, Tian G, Pan Z, Zhao F, Feng X, Liu Q, Lyu J. Nomograms for predicting the survival rate for cervical cancer patients who undergo radiation therapy: a SEER analysis. Future Oncol. 2019 Sep;15(26):3033-3045

[56] Marchetti C, De Felice F, Di Pinto A, Romito A, Musella A, Palaia I, Monti M, Tombolin V, Muzii L, Benedetti Panici P. Survival Nomograms after Curative Neoadjuvant Chemotherapy and Radical Surgery for Stage IB2-IIIB Cervical Cancer. Cancer Res Treat. 2018 Jul;50(3):768-776.

[57] Lee WK, Chong GO, Jeong SY, Lee HJ, Park SH, Ryu JM, Choi YS, Kang S, Koo YJ, Lee DH, Kong E, Lee SW. Prognosis-Predicting Model Based on 18F fluorodeoxyglucose PET Metabolic Parameters in Locally Advanced Cervical Cancer Patients Treated with Concurrent Chemoradiotherapy: Multi-Center Retrospective Study. J Clin Med. 2020 Feb 5;9(2):427.

[58] Zheng RR, Huang XW, Liu WY, Lin RR, Zheng FY, Lin F. Nomogram Predicting Overall Survival in Operable Cervical Cancer Patients. Int J Gynecol Cancer. 2017 Jun;27(5):987-993.

[59] Shim SH, Lee SW, Park JY, Kim YS, Kim DY, Kim JH, Kim YM, Kim YT, Nam JH. Risk assessment model for overall survival in patients with locally advanced cervical cancer treated with definitive concurrent chemoradiotherapy. Gynecol Oncol. 2013 Jan;128(1):54-59

[60] Zhou H, Li X, Zhang Y, Jia Y, Hu T, Yang R, Huang KC, Chen ZL, Wang SS, Tang FX, Zhou J, Chen YL, Wu L, Han XB, Lin ZQ, Lu XM, Xing H, Qu PP, Cai HB, Song XJ, Tian XY, Zhang QH, Shen J, Liu D, Wang ZH, Xu HB, Wang CY, Xi L, Deng DR, Wang H, Lv WG, Shen K, Wang SX, Xie X, Cheng XD, Ma D, Li S. Establishing a Nomogram for Stage IA-IIB Cervical Cancer Patients after Complete Resection. Asian Pac J Cancer Prev. 2015;16(9):3773-7

[61] Kidd EA, El Naqa I, Siegel BA, Dehdashti F, Grigsby PW. FDG-PET-based prognostic nomograms for locally advanced cervical cancer. Gynecol Oncol. 2012 Oct;127(1):136-40

[62] Seo Y, Yoo SY, Kim MS, Yang KM, Yoo HJ, Kim JH, Shin YJ, Kang JK, Lee KH, Lee ED, Rhu SY, Choi SC, Kim MH, Kim BJ, Kim MS, Cho CK. Nomogram prediction of overall survival after curative irradiation for uterine cervical cancer. Int J Radiat Oncol Biol Phys. 2011 Mar 1;79(3):782-7

[63] Biewenga P, van der Velden J, Mol BW, Stalpers LJ, Schilthuis MS, van der Steeg JW, Burger MP, Buist MR. Prognostic model for survival in patients with early stage cervical cancer. Cancer. 2011 Feb 15;117(4):768-76

[64] Tseng JY, Yen MS, Twu NF, Lai CR, Horng HC, Tseng CC, Chao KC, Juang CM. Prognostic nomogram for overall survival in stage IIB-IVA cervical cancer patients treated with concurrent chemoradiotherapy. Am J Obstet Gynecol. 2010 Feb;202(2):174.e1-7

[65] Polterauer S, Grimm C, Hofstetter G, Concin N, Natter C, Sturdza A, Pötter R, Marth C, Reinthaller A, Heinze G. Nomogram prediction for overall survival of patients diagnosed with cervical cancer. Br J Cancer. 2012 Sep 4;107(6):918-24

[66] Lee HJ, Han S, Kim YS, et al.., Individualized prediction of overall survival after postoperative radiation therapy in patients with early-stage cervical cancer: a Korean

Radiation Oncology Group study (KROG 13-03)., Int J Radiat Oncol Biol Phys. 2013 Nov 15;87(4):659-64.

[67] Ibrahim A, Primakov S, Beuque M, et al. Radiomics for precision medicine: Current challenges, future prospects, and the proposal of a new framework. Methods. 2021;188:20-29. doi:10.1016/j.ymeth.2020.05.022

Tube Voltage (kVp)	Tube current (mA)	Slice thickness (millimeters)	Pitch (millimeters)	Voxel size (cubic millimeters)	Reconstruction Technique
120	100-200 Auto-mA	3.75	3.75	1.17 × 1.17 × 3.75	Filtered back project (FBP)

Supplementary Table 1: Overview of CT imaging protocol

Chapter 9: Discussion

9.1 Executive summary

In this thesis, we have investigated the role of imaging biomarkers (i.e., radiomics) in precision oncology by identifying and tackling the key issues that are limiting the implementation of radiomics based prediction models in the clinic. Four main issues have been identified: (a) lack of robustness of radiomic features related to imaging equipment, image acquisition and reconstruction protocols; (b) lack of standardization in radiomic extraction related to radiomic software; (c) lack of infrastructure for AI implementation in hospitals for radiomics implementation (d) lack of the adequate number of studies on robust radiomic features for disease prognostication using prediction modelling. The above issues were addressed in a systematic manner as stated below. Initially, we reviewed the role of artificial intelligence in imaging, which is described in **Chapter 2,** where we identified the state of artificial intelligence in imaging research and its translation into routine clinical use. We identified several areas of research like patient scheduling, natural language processing, image processing and imaging biomarker development where artificial intelligence is used. Radiomic research is identified as one of the major areas of research in imaging. In **Chapter 3** we have reviewed the existing literature investigating the role of artificial intelligence in cervical cancer. In this chapter, we have analyzed various aspects of prediction modelling used in cervical cancer to predict clinical endpoints. We have performed an audit of the studies based on a meta-analysis and a 27-point PMQS (Prediction Model Quality Score) scale. The two main issues we identified were a lack of feature selection details and details of model development and selection. Our review concludes that there is an increasing trend of the use of prediction models in cervical cancer research. The meta-analysis and overall quality score on PMQS together suggest a significant role for prediction models in cervical cancer. Key issues for the future will be to train and validate models with a large amount of data, external validation, validation on prospective clinical trials and integration of these models into the electronic health record, and a more careful evaluation of models, particularly with respect to their effects on clinical outcomes. **Chapter 4** describes the role of quantitative imaging biomarkers i.e. radiomics in precision oncology. We observed that radiomics based prediction models show promising results in various types of cancer research however their clinical implementation has not been widely reported owing to various aforementioned factors. In **Chapter 5** we have presented our study investigating repeatability and reproducibility of 1093 CT radiomic features associated with a) imaging equipment, b) test-retest, c) change in acquisition and reconstruction parameters, and d) presence or absence of contrast medium, using clinical cohort and phantom imaging. From this study, it emerges that different feature categories are sensitive to different degrees of reproducibility and repeatability. We identified 110 robust radiomic features which were later subjected to prediction modelling studies. In **Chapter 6** we have presented the

development of GUI based radiomics extractor. In **Chapter 7** we have presented the building blocks of an IT framework to perform centralized and federated machine learning. We have addressed three major issues i.e., data harvesting and processing, radiomic extraction, ontology-based RDF conversion and a data server for federated machine learning. We have developed the data harvesting and processing pipeline i.e., ETL (Extract, Transform, Load) tool and a GUI based radiomic extraction pipeline which is easy to implement and provides stability in radiomic extraction.

We have configured a data server repository in our hospital adhering to FAIR principles to store data for centralized and distributed machine learning. We have also implemented the pipelines for data conversion in RDF format using an ontology. On our data server, we were able to achieve complete integration of clinical and radiomic data in the form of RDF graphs. The ontology-based RDF provides flexibility for the universal query of our data repository server. **Chapters 8** demonstrate the predictive capabilities of radiomic features with or without clinical features. In chapter 8 we performed a prediction modelling study in cervical cancer using clinical and radiomics data and tested various prediction algorithms using various combinations and demonstrated the association of robust radiomic features with the endpoint (i.e., overall survival) and its ability to predict the clinical outcome with and without the clinical features.

9.2 Limitations of this work

There are several limitations to this work. We have focused on radiomics in CT images only. The main reasons behind this are A) CT was the first imaging modality investigated in radiomic studies, so we explored the same, and B) CT images have a higher resolution than PET images, which makes CT images more appropriate for quantifying textures. In our study, we neither investigated the robustness of PET and MRI radiomic features nor developed a prediction model using PET or MRI. Although, in the review of literature in chapter 4, we have analyzed the evidence of the role of radiomic features across imaging modalities and found significant evidence of the role of CT as well as PET and MRI radiomic features in precision oncology. Therefore, it remains open to debate whether the results available in this work can be extended to imaging modalities like PET or MRI.

Another limitation is that in the study of repeatability and reproducibility, a uniform phantom was used instead of a texture phantom. Texture phantom would have been a better option for this study. Also, in this thesis, we did not check the impact of a change of preprocessing parameters on the stability of radiomic features. Future work on the reproducibility of radiomics should focus on relevant phantoms and investigating all permutations of acquisition and reconstruction protocols to achieve a better understanding on how these affect radiomics. For PET and MRI similar studies should be conducted.

In this thesis, we limited our study to handcrafted radiomics whereas deep learning radiomics remain unexplored. Deep learning radiomics is a topic of many ongoing

research projects and both its reproducibility and its value compared to "classical" radiomics remains to be seen.

Although we were able to establish a FAIR principle-based AI infrastructure for centralized and distributed machine learning, our study was limited only to the development and validation of the infrastructure and not its real-world implementation which will be the topic of future work. We were able to demonstrate the importance of robust radiomic features in developing cancer prediction models in our thesis. But our study is limited to only one disease type (Cervical cancer), single outcome (overall survival), small sample size, single-centre study and retrospective study. Also, as said, the results of this thesis cannot be directly extrapolated to other imaging modalities. However, the framework proposed in chapters 5 to 8 can be extrapolated across other imaging modalities, cancer types and outcomes.

9.3 Directions for future research

In many ways, this thesis contributed to the implementation of imaging biomarkers in precision oncology. It has summarized the current status of radiomic based research in medical imaging and oncology; identify the key issues creating obstacles in translating AI-powered prediction models into decision support systems in oncology. We were also able to present our methodology to tackle a few key issues like stability of radiomic features, AI infrastructure requirement, data privacy and portability. We were in a position to demonstrate our strong belief in the critical role of radiomics in precision oncology. Although we were able to address several key issues faced by radiomic based prediction analytics in oncology, there are still several possibilities of future work in this field. The radiomic stability research needs to be further extended to various permutations and combinations in CT radiomics study as well as to other imaging modalities. As GTV (gross tumor volume) is one of the confounding factors in radiomic feature extraction, a detailed volume dependent radiomic stability study needs to be performed.

Furthermore, there is a requirement of development of an AI-driven ETL tool for extraction and management of meaningful data from hospital information systems. There is tremendous scope for research in prediction model development based on large datasets of multicentric retrospective and prospective studies. In spite of the fact that our hospital and hospitals using hospital information systems (HIS)/Electronic medical records (EMR) and picture archiving and communication systems (PACS) have treated a large number of cancer patients over the last 15 years, a disparity exists between the treatment of the patients and the available data on HIS/EMR or PACS. In the wake of this study, we felt the need for improved IT infrastructure to capture and archive patient data. We found in chapter 8 of our study that only 5 % of patients treated in our hospital between 2005 and 2009 had pretreatment imaging data stored on the hospital's PACS. As we investigated, it was found that the majority of patients had their imaging done outside the hospital and therefore, they weren't stored on the hospital's PACS. Following this revelation, we began archiving all imaging data of patients, regardless of

whether the imaging was done in or out of the hospital. The government and all stakeholders in the health care system need to establish strict guidelines for the archiving of treatment-related data in an electronic system. Furthermore, we recommend that the medical association create the standard operating procedure to standardize data archiving and imaging protocols across hospitals. Using this method, we can support the implementation of artificial intelligence in the Indian healthcare system.

Prediction model development based on deep learning radiomics may also be the major area of research in the future. Through deep learning algorithms, it is possible to improve low dose scans to full dose scan quality to reduce radiation exposure to patients undergoing CT and Nuclear Medicine studies. Moreover, deep learning can also be employed to convert images between different domains. An attenuation correction can be performed on PET/MRI images by means of this technique. A synthetic CT can be generated from MRI images and the same synthetic CT can also be used for CT planning which may improve the radiotherapy workflow. Also, AI-based image reporting using CNNs and RNNs may help physicians interpret medical images such as PET scans, CTs, ultrasounds, SPECT scans and MRIs.

Finally, utilization of our FAIR compliant AI infrastructure for distributed machine learning may be another area of research in future.

9.4 Research impact and utilization summary

9.4.1 Socio-cultural impact

The major challenge for precision oncology is to provide tailored personalized treatment to patients to improve the outcome of the treatment and reduce the associated risk of toxicity and relapse. The implementation of decision support systems in clinics will bring a benefit to clinicians and also to patients. The role of radiomics research is evident in precision oncology but currently it is facing several issues. These issues are strongly impacting the possibility to translate research prototypes as decision support systems in the clinic. Through our work we propose an AI-driven technique for automated image analysis and extraction of radiomic features and prediction of the outcome based on the information stored in radiomic features. Since very few studies have been performed in AI in oncology in our part of the world, this work will inspire more and more researchers in the field to perform similar kinds of studies. Our proposed AI infrastructure will be helpful for the research community in this country. Our initiative has already attracted researchers to explore AI based research in oncology in India. In our next project, we have been able to add a research partner in India and implemented our proposed AI infrastructure. Gradually the acceptance of AI-based research in oncology will grow the culture of using AI-based decision support systems in oncology.

9.4.2 Economic Impact

Cancer therapy is gradually evolving from conventional treatment to personalized treatment. In the last few years, treatment options have increased drastically. Besides the utilization of conventional surgery, radiotherapy, and chemotherapy, newer treatment options like robotic surgery, image-guided radiotherapy, proton therapy and targeted immunotherapy or molecular therapy have opened up new horizons for patients. These treatments offer better efficacy and fewer complications and toxicity. A physician's decision to select the best treatment is driven by clinical judgment based on various parameters derived from diagnostic tests which are often subjected to the clinical acumen and expertise of the treating physician. Sometimes a decision becomes very difficult for the doctors and patients if there is a lack of strong evidence or differential costs involved. Several biomarkers have been developed helping in the decision, however, there are certain issues with the use of biomarkers: a) biomarkers are only helpful to guide the decision making for targeted therapies and do not have much benefits in decision making to select conventional treatment like surgery, radiotherapy and chemotherapy and b) biomarker testing is expensive and time-consuming. Imaging biomarker-based decision support systems presented in this thesis may be a game-changer in this scenario because of multiple reasons: a) imaging biomarkers can be derived from images already performed so no additional cost is involved, b) imaging biomarkers can be developed based on retrospective data and applied prospectively so evidence building can be achieved quicker and with much less cost, c) imaging biomarkers can be developed for conventional as well as targeted treatments d) imaging biomarker parameters can be obtained instantaneously with no delay in treatment

9.4.3 Technological impact

Although technical development was not the main aim of this thesis, it addresses three major technical issues related to the implementation of AI in oncology. We developed a GUI consisting of a) a data harvesting pipeline (ETL module), b) a GUI based radiomics extraction pipeline, c) a pipeline for RDF conversion of clinical, radiomic and DICOM data, and d) an RDF data server adhering to FAIR data principles for distributed and centralized machine learning. Philips as a corporate partner of the project has deployed Intellispace Discovery (research-only build; Philips Medical System, Eindhoven, The Netherlands) for research and development for the BIONIC project. We are using this system for various activities like image visualization and manual and automatic contouring of the GTV. The pipelines developed in this thesis can be integrated into Intellispace discovery for research or clinical uses. Intellispace discovery will thus be able to provide an automated solution for radiomic feature extraction and conversion of data into RDF format by integrating pipelines developed in this thesis with existing DL based auto contouring.

Summary

Cancer is the second most fatal disease worldwide. Management of cancer is a complex process consisting of diagnosis and staging of the disease and planning and execution of treatment followed by post-treatment follow up. The conventional method of treatment often fails in many patients due to the variability of the disease process amongst a heterogeneous patient population. In the past few years, various biomarkers have been developed to identify the subtype of disease which leads to developing personalized treatment in oncology i.e., precision oncology. Medical imaging plays a key role in cancer management at various stages. Imaging modalities are used in diagnosis, staging, planning of treatment and follow up of disease. It is also used in the restaging of disease in case of progression or recurrence. The information stored in medical images is analysed by imaging experts either by qualitatively using visual interpretation or by semi-quantitative methods, which allows sub-optimal use of information stored in medical images. The huge amount of informative quantitative data stored in medical images remains unexplored. After intended use, these medical images are stored in the archival system (PACS) of the hospital. In the last decade, the medical images archived in hospital PACS have been identified for quantitative analysis and development of imaging biomarkers. The quantitative analysis of medical images (radiomics) has led to the data explosion which is the source of BIG data in oncology. Artificial intelligence (AI) algorithms like machine learning (ML) and deep learning (DL) have been applied to imaging Big data to develop decision support systems in precision oncology. Several imaging biomarkers (radiomic features) have been identified as digital phenotypes of the disease. Nevertheless, several radiomic features have shown potential to predict various endpoints in oncology, but the translation of these radiomics based prediction models as decision support systems (DSS) in the clinic will require addressing several key issues. The radiomic community needs to address the key issues related to the implementation of radiomics based DSS: (a) robustness of radiomic features, (b) development and implementation of AI infrastructure in hospitals, (c) multicentre and prospective radiomics studies, (d) creating awareness and faith among doctors and patients. Through this work, we have tried to address most of these issues to facilitate the implementation of radiomics based DSS in clinical practice.

Curriculum Vitae

Ashish Kumar Jha was born in India on 14th February 1977. He holds a Master's degree in Information Technology and postgraduate diplomas in Bioinformatics and Nuclear Medicine. He is a certified radiation safety officer in diagnostic and therapeutic nuclear medicine and Cyclotron. He has 17 years of experience as a Medical Physicist and Radiation Safety Officer in Nuclear medicine. He is working as Scientific Officer at Tata Memorial Hospital, Mumbai, India for the last thirteen years. He joined as a PhD student at Maastricht University at the GROW - School for Oncology and Reproduction, under the supervision of Prof. Andre Dekker in 2017. He has authored several publications in the field of medical imaging and clinical data science in peer-reviewed international journals. He has presented his research at several national and international conferences. Where he has a keen interest in scientific research at the same time he is also interested in leadership in scientific societies. He has served as Treasurer and Secretary of the Society of Nuclear Medicine in India (SNMI) for six years and a member of, the Executive committee of the Nuclear Medicine Physicist Association of India (NMPAI) for two years. Currently, he is Chairman, the scientific committee of Council of Nuclear Medicine Science and Technology (CNMST) and Joint Secretary of Indian College of nuclear medicine. Recently in 2020, he has been bestowed with the **Earnest O' Lawrence Oration award** by Nuclear Medicine Physicist Association in India. He was a Member of the Indian College of Nuclear Medicine and this year; he has been elevated as a **Fellow** of the college. He has received several awards including: IAEA travel grant, WARMTH travel grant, WMIS travel grant ICTP travel grant. Ashish's vision is that digital phenotyping of cancer based on radiomics and Artificial Intelligence will empower the clinicians to provide personalize cancer treatment to the patient.

List of Publications

Original articles:

1. **Jha AK**, Mithun S, Purandare NC, et al. Radiomics: a quantitative imaging biomarker in precision oncology, Nucl Med Commun. 2022;43(5):483-493. doi:10.1097/MNM.0000000000001543

2. **Jha AK**, Mithun S, Rangarajan V, Wee L, Dekker A. Emerging role of artificial intelligence in nuclear medicine. Nucl Med Commun. 2021;42(6):592-601. doi:10.1097/MNM.0000000000001381

3. **Jha AK**, Mithun S, Jaiswar V, et al. Repeatability and reproducibility study of radiomic features on a phantom and human cohort. Sci Rep. 2021;11(1):2055. Published 2021 Jan 21. doi:10.1038/s41598-021-81526-8

4. **Jha A. K.**, S. Mithun, U. B. Sherkhane et al., "Implementation of Big Imaging Data Pipeline Adhering to FAIR Principles for Federated Machine Learning in Oncology," in IEEE Transactions on Radiation and Plasma Medical Sciences, vol. 6, no. 2, pp. 207-213, Feb. 2022, doi: 10.1109/TRPMS.2021.3113860.

5. U. B. Sherkhane, **Jha AK**, Mithun S, Purandare NC, et al., Development and validation of GUI radiomics feature extractor software (PyRadGUI) using PyRadiomic package, IOP Journal of Physics conference series Mar 2022 (Accepted for publication)

6. **Jha AK**, Mithun S, Puranik A, et al. Software development for hepatopulmonary shunt estimation by gamma camera method in transarterial radioembolization. World J Nucl Med. 2019;18(4):366-372.

7. **Jha AK**, Mithun S, Puranik AD, et al. Performance characteristic evaluation of a bismuth germanate-based high-sensitivity 5-ring discovery image quality positron emission tomography/computed tomography system as per National Electrical Manufacturers Association NU 2-2012. World J Nucl Med. 2019;18(4):351-360. Published 2019 Dec 18. doi:10.4103/wjnm.WJNM_72_18

Book Chapters:

1. **Jha AK**, Mithun S, Rangarajan V, Wee L, and Dekker A, Comprehensive Utility of Artificial Intelligence in Medical Imaging, Evidence Based Management of Cancers in India; Radiology Beyond Imaging: Intertwining Imaging with Advanced Technology (Part E), Publisher: Tata Memorial Hospital, Mumbai; Homi Bhabha National Institute, Mumbai, Vol. XVIII ISBN: 978-93-82963-63-9

2. Mithun S, **Jha AK**, Rangarajan V, Wee L, and Dekker A, Natural Language Processing in Radiology Reports, Evidence Based Management of Cancers in India; Radiology Beyond Imaging: Intertwining Imaging with Advanced Technology

(Part E), Publisher: Tata Memorial Hospital, Mumbai; Homi Bhabha National Institute, Mumbai , Vol. XVIII ISBN: 978-93-82963-63-9

Abstracts:

1. **Jha AK**, Mithun S, Sherkhane UB, Jaiswar V, Mehta G, Panchal S, Nath B, Tripathi A, Purandare N, Rangarajan V, Wee L, and Dekker A, Development and validation of robust CT radiomics signature for predicting 5 year overall survival in Cervical Cancer, WMIC-2022, Miami, USA

2. **Jha AK**, Mithun S, Sherkhane UB, Jaiswar V, Mehta G, Panchal S, Nath B, Tripathi A, Purandare N, Rangarajan V, Wee L, and Dekker A, Development and validation of radiomic signature for classification of high and low-grade chondrosarcoma: A pilot study, WMIC-2022, Miami, USA

3. Mehta G, Panchal S, **Jha AK**, Mithun S, Sherkhane UB, Jaiswar V, Panchal S, Nath B, Tripathi A, Purandare N, Rangarajan V, Wee L, and Dekker A, Influence of proportional change of GTV over the PET radiomic features in non small cell lung carcinoma patients. Presenter: WMIC-2022, Miami, USA

4. Panchal S, **Jha AK**, Mithun S, Mehta G, Sherkhane UB, Jaiswar V, Nath B, Tripathi A, Purandare N, Rangarajan V, Wee L, and Dekker A, Reproducibility of PET radiomic features of non-small cell lung carcinoma by changing pre-processing parameters, WMIC-2022, Miami, USA

5. Panchal S, **Jha AK**, Mithun S, Mehta G, Sherkhane UB, Jaiswar V, Nath B, Tripathi A, Purandare N, Rangarajan V, Wee L, and Dekker A, Impact of change in preprocessing parameters on the reproducibility of CT radiomic feature in rectal cancer, WMIC-2022, Miami, USA

6. Nath B, **Jha AK**, Mithun S, Tripathi A, Sherkhane UB, Jaiswar V, Purandare N, Rangarajan V, Wee L, and Dekker A, Disease classification based on features extracted from Fourier spectrum of CT image: A pilot study, WMIC-2022, Miami, USA

7. Dwivedi P, Vajarkar V, Choudhary S, **Jha AK**, Rangarajan V, Analysis of PET Radiomic features with different reconstruction algorithms, Eur J Nucl Med Mol Imaging 49 (Suppl 1), 1–751 (2022). https://doi.org/10.1007/s00259-022-05924-4

8. S. Mithun, **Jha AK**, Sherkhane UB, Jaiswar V, Purandare N, Rangarajan V, Wee L, and Dekker A, A Comparison of Transfer learning with BERT and Clinical BERT for Classification of Radiology Reports, Eur J Nucl Med Mol Imaging 49 (Suppl 1), 1–751 (2022). https://doi.org/10.1007/s00259-022-05924-4

9. **Jha AK**, Mithun S, Sherkhane UB, Jaiswar V, Mehta G, Panchal S, Nath B, Tripathi A, Purandare N, Rangarajan V, Wee L, and Dekker A, Implementation of Information Technology Infrastructure for Federated Machine Learning in Tertiary Cancer Care in India, Eur J Nucl Med Mol Imaging 49 (Suppl 1), 1–751 (2022). https://doi.org/10.1007/s00259-022-05924-4

10. **Jha AK**, Mithun S, Sherkhane UB, Jaiswar V, Mehta G, Panchal S, Nath B, Tripathi A, Purandare N, Rangarajan V, Wee L, and Dekker A, Development and validation of Radiomic Signature for Non-small-cell Lung Cancer: A Proof-of-Concept Study, Eur J Nucl Med Mol Imaging 49 (Suppl 1), 1–751 (2022). https://doi.org/10.1007/s00259-022-05924-4

11. Jaiswar V, **Jha AK**, Mithun S, Sherkhane UB, Mehta G, Panchal S, Nath B, Tripathi A, Purandare N, Rangarajan V, Wee L, and Dekker A, Resectability prediction of Non-Small Cell Lung Carcinoma based on Computed Tomography Radiomic Features using Machine Learning algorithms, Eur J Nucl Med Mol Imaging 49 (Suppl 1), 1–751 (2022). https://doi.org/10.1007/s00259-022-05924-4

12. Nath B, **Jha AK**, Mithun S, Sherkhane UB, Jaiswar V, Mehta G, Panchal S, Nath B, Tripathi A, Purandare N, Rangarajan V, Wee L, and Dekker A, Can we rely on PET Segmentation for radiomic extraction for malignancy of hollow organs like the rectum?, Eur J Nucl Med Mol Imaging 49 (Suppl 1), 1–751 (2022). https://doi.org/10.1007/s00259-022-05924-4

13. Sherkhane UB, Jaiswar V, Mithun M, **Jha AK**, Rangarajan V, Wee L, Dekker A, Accessing the predictability of Epidermal growth factor receptor status from Computed Tomography radiomics using machine learning, Eur J Nucl Med Mol Imaging 49 (Suppl 1), 1–751 (2022). https://doi.org/10.1007/s00259-022-05924-4

14. **Jha AK**, Kulkarni C, Mithun S, Sherkhane UB, Jaiswar V, Mehta G, Panchal S, Nath B, Tripathi A, Purandare N, Rangarajan V, Wee L, and Dekker A, Performance assessment of a pre-trained CNN model for auto-delineation of primary tumor on CT scans in non-small cell lung carcinoma (NSCLC) patients, J Nucl Med. June 1, 2022, 63 (supplement 2) 3201

15. Mithun S, **Jha AK**, Sherkhane UB, Jaiswar V, Nautiyal A, Puts S, Bermejo I, Zegers CM, Purandare N, Rangarajan V, Dekker A, Wee L, Comparison of Machine Learning Models for Lung Carcinoma Clinical Concept Extraction from Radiology Reports, J Nucl Med. June 1, 2022, 63 (supplement 2) 3238

16. Mehta G, **Jha AK**, Mithun S, Sherkhane UB, Jaiswar V, Panchal S, Nautiyal A, Purandare N, Rangarajan V, Wee L, Dekker A, Effect of resampling of voxel size on CT Radiomic features extraction, J Nucl Med. June 1, 2022, 63 (supplement 2) 3241

17. **Jha AK**, Mithun S, Sherkhane UB, Jaiswar V, Purandare N, Rangarajan V, Wee L, and Dekker A, Development and validation of GUI radiomics feature extractor software (PyRadGUI) using PyRadiomic package, International Conference on Recent Trends in Electrical Electronics communication and Instrumentation (ICRTEECI 2021) 26-27 Dec, 2021 · Dec 26, 2021

18. Mehta G, **Jha AK**, Mithun S, Sherkhane UB, Jaiswar V, Mehta G, Nautiyal A, Purandare N, Rangarajan V, Wee L, and Dekker A, Effect of change of gross tumour volume on CT radiomic features Conference: Eur J Nucl Med Mol Imaging (2021)At: vienna Austria Volume: 48 (Suppl 1): S398

19. Mithun S, **Jha AK**, Sherkhane UB, Jaiswar V, Purandare N, Rangarajan V, Wee L, and Dekker A, Development of a Deep Learning Natural Language Processing Model for Classification of Lung Cancer Radiology Reports Conference: Eur J Nucl Med Mol Imaging (2021)At: Vienna, Austria Volume: 48 (Suppl 1): S330

20. **Jha AK**, Mithun S, Sherkhane UB, Jaiswar V, Purandare N, Rangarajan V, Wee L, and Dekker A, Prediction of CT radiomic features using PET radiomic features and vice versa Conference: Eur J Nucl Med Mol Imaging (2021)At: Vienna,AustriaVolume: 48 (Suppl 1): S509

21. **Jha AK**, Mithun S, Sherkhane UB, Jaiswar V, Purandare N, Rangarajan V, Wee L, and Dekker A, Is there any correlation between tumour volume and other radiomic features, Indian Journal Of Nuclear Medicine, November 2019, 34(5):23-109

22. U. B. Sherkhane, Mithun S, **Jha AK**, Sherkhane UB, Jaiswar V, Purandare N, Rangarajan V, Wee L, and Dekker A, Natural Language Processing tool for detecting lung cancer terms in real-world radiology reports, Indian Journal Of Nuclear Medicine, November 2019, 34(5):23-109

23. Mithun S, **Jha AK**, Sherkhane UB, Jaiswar V, Purandare N, Rangarajan V, Wee L, and Dekker A, Validation of an open source Natural Language Processing (NLP) and an in-house developed python script for named entity recognition from radiology reports of lung carcinoma cases European Journal of Nuclear Medicine and Molecular Imaging (2019) 46 (Suppl 1): S761–S761

24. **Jha AK**, Mithun S, Sherkhane UB, Jaiswar V, Purandare N, Rangarajan V, Wee L, and Dekker A, Is there any correlation between the radiomic features extracted from CT and PET Images, European Journal of Nuclear Medicine and Molecular Imaging (2019) 46 (Suppl 1): S762–S762 24.

25. Prasad RV, **Jha AK**, Mithun S, Sherkhane UB, Jaiswar V, Purandare N, Rangarajan V, Wee L, and Dekker A, Using Artificial Intelligence and Radiomics to support cancer care decisions. 19 th International Conference on the use of Computers in Radiation Therapy.

26. **Jha AK**, Mithun S, Sherkhane UB, Jaiswar V, Purandare N, Rangarajan V, Wee L, and Dekker A,, Effect of differences in tumour region of interest (ROI) delineation on radiomics features : comparison of PET SUV threshold versus region- growing methods, Indian Journal of Nuclear Medicine, November 2018, Volume 33 | Issue 5 (Supplement), Page Nos. 1-113

27. Mthun S, J**ha AK**, Sherkhane UB, Jaiswar V, Purandare N, Rangarajan V, Wee L, and Dekker A, Concept extraction from radiology reports of lung carcinoma cases using an open source medical terminology annotator: A feasibility study, Indian Journal of Nuclear Medicine, November 2018, Volume 33 | Issue 5 (Supplement) Page Nos. 1-113

28. **Jha AK**, Mithun S, Sherkhane UB, Jaiswar V, Purandare N, Rangarajan V, Wee L, and Dekker A, Influence of CT acquisition and reconstruction parameters on

radiomic feature extraction, European Journal of Nuclear Medicine and Molecular Imaging, Volume 45 Number 10, Sep 2018/ 1619-7070

Acknowledgements