Langenbeck's Archives of Surgery (2023) 408:321
https://doi.org/10.1007/s00423-023-03039-4

SYSTEMATIC REVIEW

Accuracy of deep neural learning models in the imaging prediction of pathological complete response after neoadjuvant chemoradiotherapy for locally advanced rectal cancer: a systematic review

Sowmya Prabhakaran[1] · Keith Wai Keong Choong[2] · Swetha Prabhakaran[3] · Kay Tai Choy[2] · Joseph CH Kong[3,4,5]

Received: 18 March 2023 / Accepted: 1 August 2023

Abstract

Purpose Up to 15–27% of patients achieve pathologic complete response (pCR) following neoadjuvant chemoradiotherapy (CRT) for locally advanced rectal cancer (LARC). Deep neural learning (DL) algorithms have been suggested to be a useful adjunct to allow accurate prediction of pCR and to identify patients who could potentially avoid surgery. This systematic review aims to interrogate the accuracy of DL algorithms at predicting pCR.

Methods Embase (PubMed, MEDLINE) databases and Google Scholar were searched to identify eligible English-language studies, with the search concluding in July 2022. Studies reporting on the accuracy of DL models in predicting pCR were selected for review and information pertaining to study characteristics and diagnostic measures was extracted from relevant studies. Risk of bias was evaluated using the Newcastle-Ottawa scale (NOS).

Results Our search yielded 85 potential publications. Nineteen full texts were reviewed, and a total of 12 articles were included in this systematic review. There were six retrospective and six prospective cohort studies. The most common DL algorithm used was the Convolutional Neural Network (CNN). Performance comparison was carried out via single modality comparison. The median performance for each best-performing algorithm was an AUC of 0.845 (range 0.71–0.99) and Accuracy of 0.85 (0.83–0.98).

Conclusions There is a promising role for DL models in the prediction of pCR following neoadjuvant-CRT for LARC. Further studies are needed to provide a standardised comparison in order to allow for large-scale clinical application.

PROPERO Registration PROSPERO 2021 CRD42021269904 Available from: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42021269904.

Keywords Rectal cancer · Deep neural learning · Artificial intelligence · Therapeutic response · Long course neoadjuvant chemoradiotherapy · Magnetic resonance imaging

Introduction

There has been a great effort to improve the diagnostic accuracy of treatment response in patients with locally advanced rectal cancer (LARC) after neoadjuvant chemoradiation (CRT). While the standard of care for LARC remains

Sowmya Prabhakaran and Keith Wai Keong Choong are co-first authors and contributed equally to this manuscript.

✉ Sowmya Prabhakaran
sowmyaprabhakaran@gmail.com

[1] Department of General Surgical Specialties, The Royal Melbourne Hospital, Melbourne, Victoria, Australia

[2] Department of Surgery, Austin Health, Melbourne, Victoria, Australia

[3] Department of Colorectal Surgery, Alfred Hospital, Melbourne, Victoria, Australia

[4] Central Clinical School, Monash University, Melbourne, Victoria, Australia

[5] Department of Surgical Oncology, Peter MacCallum Cancer Centre, Melbourne, Victoria, Australia

Published online: 18 August 2023

neoadjuvant CRT to the primary site and regional lymph nodes with subsequent total mesorectal excision, 15–27% of patients achieve pathologic complete response (pCR) with no residual tumour found at histopathology after resection [1–4], and another 60% experience tumour downstaging [5]. This is compounded by the increasing adoption of total neoadjuvant therapy (TNT), where additional systemic oxaliplatin-based chemotherapy is added to the neoadjuvant CRT regimen, with the pCR rate doubling up to 30% [6–8]. This suggests that instead of radical total mesorectal excision to resect the primary site and harvest all locoregional lymph nodes, there is a proportion of patients that may avoid surgery altogether. This subset of patients can be stratified to a regimen called "watch and wait" to avoid the morbidity of radical resection [9]. As a result, it becomes increasingly important to identify the group of patients that may achieve pCR.

At present, clinical complete response (cCR) is used as the marker of pCR in patients that do not undergo surgery. The assessment and stratification of this treatment pathway is based on a few criteria such as i) digital rectal exam that shows no palpable tumour and a soft mucosa at the previously irradiated tumour bed ii) endoscopic visualization of a flat, telangiectatic scar iii) MRI that shows complete tumour regression and iv) PET that shows complete metabolic response [10]. Patients need to meet all these criteria before they are deemed to have a clinical complete response (cCR).

An international multi-centre study of 1009 patients in the International Watch and Wait Database who had evidence of cCR after neoadjuvant CRT had a 5-year overall survival of 85% and a 5-year disease-free survival of 94% [11]. Although long-term data has been encouraging, showing an equivalent oncological outcome when comparing watch and wait against immediate surgery after neoadjuvant therapy, there is reservation to adopt it readily.

The reservation arises due to the risk of tumour regrowth, which is reported to be up to 25.2%. The majority of regrowths will occur within 2 years and 97% are within the rectal lumen. While the majority (88%) of patients with tumour regrowth could still undergo salvage surgery after local regrowth [12], there is a propensity for developing distant metastasis for those that experience tumour regrowth. [13] At the same time, the early identification of patients who would not respond well to neoadjuvant CRT alone would allow for individualized tailoring of the treatment regime, with early surgical intervention being considered in these patients.

Current assessment of cCR has its limitations, with concerns being raised about the ability to perform salvage surgery for local regrowth with a known increased risk of developing distant recurrence. One of the challenges is to accurately assess complete response to treatment by MRI and/or PET. This is partly due to the ongoing inflammation that still occurs at the tumour site after radiation therapy; giving a false positive signal whereas microscopic tumour (millimeters in size) is impossible to detect on any modality.

In the past decade, we have seen a meteoric increase in the use and applications of artificial intelligence (AI). Deep neural learning (DL) is a subfield of machine learning (ML) and AI that is increasingly being used in medicine [14–17] for diagnosis [18], classification [19] or prediction [20, 21]. DL allows pre-trained convolutional neural networks to extract information, and models can be constructed using clinical and radiological features to help predict pCR in patients with LARC treated with NA-CRT. This could represent a different way of predicting pCR.

A few centres have built models to study the role of DL in pCR with isolated studies being published on this topic. However, to date, no studies exist that have systematically analysed the data from these studies. The aim therefore of this systematic review was to interrogate published studies to establish the accuracy of DL algorithms in predicting pCR following NA-CRT in the management of LARC.

Methods

Search strategy & information sources

The protocol for this systematic review was registered with the International Prospective Register of Systematic Reviews (PROSPERO) with the unique identification number CRD42021269904. This review has been written in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [22].

A search of the electronic databases Embase (PubMed, MEDLINE) and Google Scholar to identify eligible published studies in English with human participants using the following terms: "deep neural learning" OR "machine learning" OR "artificial intelligence" and "rectal cancer".

There was no start date limit, and the search concluded in July 2022. Reference lists of reference publications were scanned to identify additional relevant studies. The search, selection and data collection were done independently by two authors (SP & KC), with discrepancies resolved by referral to senior author (JK).

Inclusion and exclusion criteria

Studies that reported on the accuracy of DL models in predicting pCR after neoadjuvant CRT in LARC were selected for review.

Full-text articles were then reviewed to confirm eligibility for inclusion using the following criteria: (1) patients were diagnosed with LARC, (2) patients underwent neoadjuvant CRT, (3) one or more DL models were developed or used to predict pCR and (4) accuracy of the model was

determined using histopathological confirmation. Conference abstracts, letters, comments, case reports and reviews were excluded. Publications were also excluded if (1) radiomics but not deep neural learning models were used to predict pCR, (2) DL models were developed but not used to predict pCR or (3) insufficient data was provided to interrogate accuracy.

Data collection and data items

We extracted basic study characteristics and diagnostic measures from each study. In particular, we compared the DL algorithm and software, as well as the imaging type under consideration. Specifics were noted regarding preoperative image processing details, personnel involved and blinding of the personnel involved. Thereafter, the performance of these models was assessed in terms of accuracy, sensitivity, specificity and/or area under the curve (AUC) for one or more deep neural learning models (including training and testing models). All of the above information and data was described and recorded by the same two authors independently.

Different studies have used varying terms to describe the histopathological response to neoadjuvant CRT, including pCR versus non-pCR, responders versus non-responders, as well as good versus poor responders. In our results section, we have listed the terms used by each study. This study dichotomises the patients into pCR vs. non-pCR groups.

The primary outcome measure was accuracy of the DL methods in the prediction of pCR, when compared to pCR findings at time of histopathology from surgery. Statistical analysis was performed using IBM SPSS Statistics (Version 27).

Quality assessment

Studies were judged based on the Newcastle-Ottawa Scale on domains comprising selection of study groups, comparability of the groups and the ascertainment of either the exposure or outcome interest, on a scale of 1–9 stars. Disagreements were resolved by consensus.

Results

Included papers and clinical research design

Our search identified 85 non-duplicate and relevant studies from electronic databases. After screening titles and abstracts, 19 articles were identified for full-text reading, and 12 of these articles were eventually included in our study cohort [23–34] (Fig. 1). All studies had at least six stars cumulatively in the Newcastle-Ottawa Scale domains of selection of study groups, comparability of the groups and outcome domains indicative of good quality to avoid publication bias [35].

Notably, there was an increase in the number of included studies year on year between 2016 and 2022. These 12 studies were equally split between retrospective and prospective cohort studies. The median number of patients in each paper was 96 (range 43–1033). All studies enlisted a multi-disciplinary team of pathologists and radiologists.

Deep neural learning methodologies

Understandably a wide range of different DL methodologies were employed. Four studies used a custom deep learning architecture. Convolutional neural network (CNN) was the most popular established architecture used in five studies, followed by artificial neural network (ANN) and K-nearest neighbors (KNN) which were each used in two studies.

While these studies were selected for predicting pathological complete response, the heterogeneity continued with the algorithms incorporating different MRI phases such as T2, ADC, DWI or even CT findings. The radiological features handpicked to feed into these algorithms utilized different classifiers, ranging from Harlick's Gray level Co-occurrence Matrix (GLCM) texture extraction to Random Forest (RF) classifier strategies. Supervised learning was largely not reported as an explicit variable.

Performance

Of the 12 studies reviewed, nine utilized a direct comparison between different DL algorithms to present their best performing DL algorithm. The primary evaluation metric was area under curve (AUC). The median reported AUC was 0.845 (range 0.71–0.99). This resulted in positive predictive values (PPV) of between 0.43 to 0.87, and negative predictive values (NPV) of between 0.83 to 0.91.

Discussion

This systematic review on the accuracy of DL models in predicting pCR after neoadjuvant CRT for LARC highlights the promising role it has in predicting pCR. The median reported AUC was 0.845, showing that DL models have a role in predicting pCR. To the best of our knowledge, this is the first systematic review that examines the accuracy of DL models in predicting pCR, supplementing previous studies which have looked at the use of radiomics, or at the utility of DL models in diagnosing lymph node metastases [36, 37].

The paradigms of rectal cancer are shifting, with rectal preservation becoming increasingly considered in patients with cCR to neoadjuvant treatment, as this group of patients is likely to have pCR [12, 38]. "Watch and wait" is an

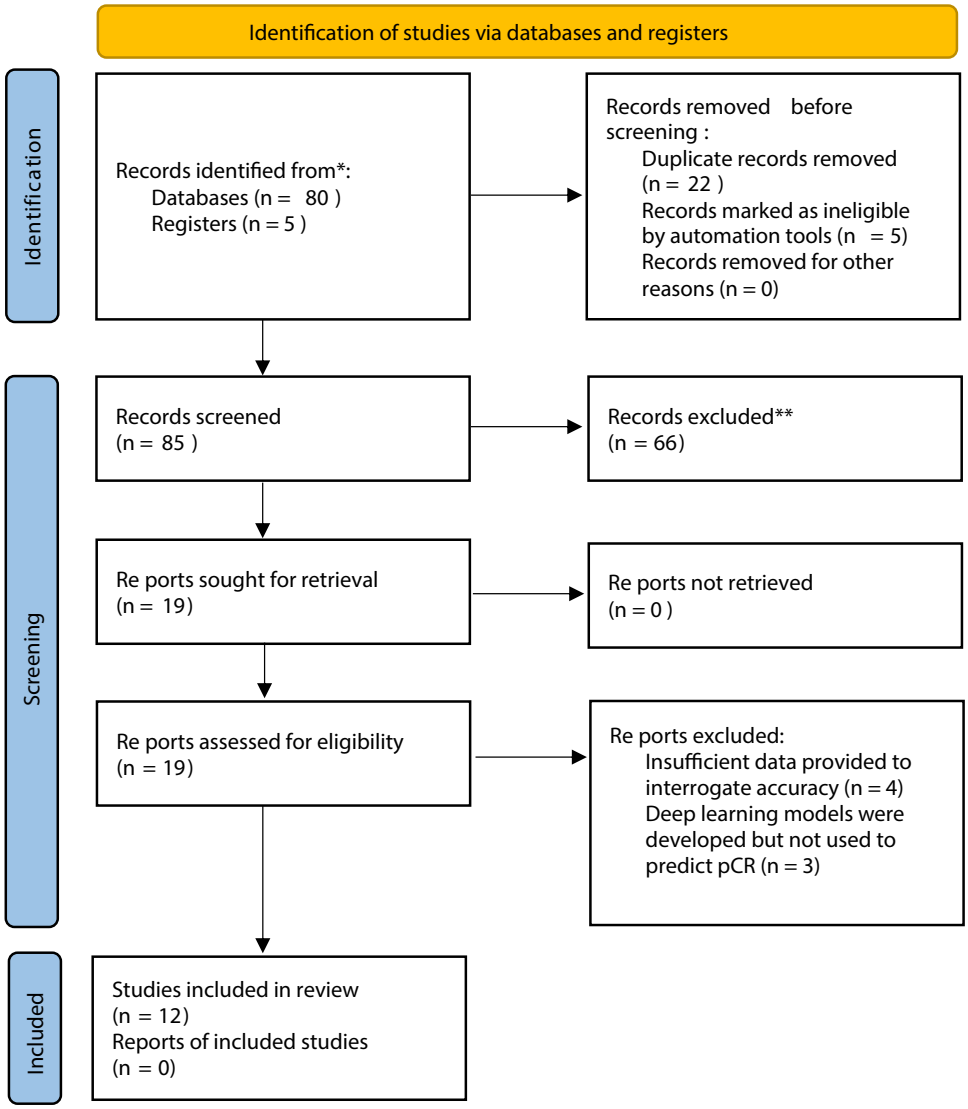

Fig. 1 PRISMA 2020 flow diagram

exciting approach of managing patients non-operatively if survival is similar compared to patients that have undergone surgery [39]. It is therefore of great clinical significance to be able to accurately predict the group of patients who will develop pCR (Tables 1, 2, and 3).

The role of imaging in medicine has grown exponentially, especially since the inclusion of radiomics. All the reviewed studies contained robust radiomics data with first-, second-, and higher-order statistics that can be combined with other patient data, and mined with increasingly sophisticated bioinformatics models to develop models that have clinical value in diagnosis, prediction, treatment and prognosis [40]. The validity of any diagnostic assessment is dependent on the quality of the ground truth, which in this case would refer to the datasets that were used for developing the models. This places great emphasis on the quality of imaging modalities and its resultant feature extraction [23]. While this has implications on equity in the provision of healthcare, it is likely that the conversion of clinical images to mineable data will also likely become routine, as the use of radiomics analyses on routine clinical images grows [40, 41].

Artificial intelligence (AI) research is being increasingly used in medicine. DL has shown great promise, with a recent paper by Seah et al in 2021, showing that a DL model can help chest x-ray interpretation [42]. In a preprint by Bizzo et al., the DL model developed for diffusion MRI to detect early acute infarcts on non-contrast head CT outperformed three expert neuroradiologists on a test CT of 150 scans.

Our review also showed good accuracy of the DL models in predicting pCR. This review has examined the many different DL models which have been successfully used to predict pCR in LARC. While the performance data is promising, the lack of a clear winner shows how broad clinical application of artificial intelligence in medicine can be hindered by transparency and reproducibility. The multitude of algorithms and DL models have generated

Table 1 Demographics

Author (Year)	Country	Number of centres	Study Design	prediction of:	Total	Total Male	Total Female	Total pCR	Total Good Responders (GR)	Total non-pCR	Total Non-Good Responders (nGR)
Feng (2022) [23]	China	1	prospective cohort	pCR	1033	706	327	244	NR	789	NR
Jin (2021) [24]	China	2	retrospective cohort	pCR (compared to resected sample)	321	267	54	57	NA	264	NA
Li (2021) [25]	China	1	retrospective cohort	pCR	80	60	20	15	29	65	51
Huang (2020) [26]	Taiwan	1	prospective cohort	pCR (compared to histopath)	270	176	94	63	-	207	-
Fu (2020) [27]	USA	1	prospective cohort	treatment response (of which there is good response vs. non good response, and pCR is a subset of good response - compared to histopath)	43	28	15	NA	22	-	21
Zhang (2020) [28]	China	1	prospective cohort	pCR (primary outcome) & tumour regression grade (secondary outcome) - based on histopath specimen	383	229	154	NA	216	74	167
Zhu (2020) [29]	China	1	retrospective cohort	pCR	700	NR	NR	NR	176	NR	524
Shayesteh (2019) [30]	Iran	1	prospective cohort	tumour response (of which pCR is a subset) (compared to histopathology)	98	66	35	17	45	19	53
Ferrari (2019) [31]	Italy	1	retrospective cohort	pCR (compared to histopath) / identifying non-responders after neoadjuvant CRT	55	31	24	16	16 (pCR)	27 (partial respond)	12 (non-responders

Table 1 (continued)

Author (Year)	Country	Number of centres	Study Design	prediction of:	Total	Total Male	Total Female	Total pCR	Total Good Responders (GR)	Total non-pCR	Total Non-Good Responders (nGR)
Shi (2019) [32]	Unknown	unknown	retrospective cohort	pCR (post surgery tumour response)	45	31	14	10 (pre-treatment) & 9 (mid-treatment)	31 (pre-treatment) & 27 (mid-treatment)	35 (pre-treatment) & 32 (mid-treatment)	14 (pre-treatment) & 14 (mid-treatment)
Bibault (2018) [33]	France	3	retrospective cohort	pCR (all patients had surgery - postop state)	95	49	46	22		73	
Nie (2016) [34]	China	1	prospective cohort	pCR (compared to histopath) / identifying non-responders after neoadjuvant CRT	48	30	18	11	31	37	17

considerable interest, both within and outside of medicine, but remain a work in progress. Nevertheless, its role in using clinical, biological or treatment inputs in order to predict further responses would no doubt enhance the capability of radiomics and serves to increase the role of imaging in clinical medicine [33].

This study has a few limitations. This is a small review with only 12 studies included. There was significant heterogeneity in the 12 studies as they included patients with various cancer stages, different radiomics features and different patient characteristics. We did not have access to individual patient data here, and could not account for all confounders, including the use of TNT which may increase the rate of pCR.

Additionally, there were small sample sizes in our study. In the Lancet paper by Seah et al., the model was trained on well over 800,000 images [42]. Great investment is required and future studies should validate these DL models using ground truth, which would include imaging data that has also been tampered with parameters such as clinical and endoscopic assessment.

Deep neural learning is also subject to the "black box" effect, which suggests that it is not easily or intuitively interpreted due to the large variety of variables, as well as the use of higher order data. This is a key issue as our conclusions here are therefore based on hypotheses, and are difficult to prove empirically using current scientific methods [33]. Still, it is worth considering that our ultimate aim is to consider if deep neural learning can augment or even supersede the thinking models used by the human brain, especially given that these models are able to analyse a much higher volume of variables at a faster speed. Thus, it is possible that these models will become necessary for us in the future to handle the volume of variables and higher-order data [33].

We would put forth that the early success and accuracy of these models needs to be interpreted with caution. Asimov's Three Laws of Robotics proposed a blueprint for integrating robots into society, constrained, and controlled to do good. Similarly, with safety in mind, it is important to emphasize that the responsibility for clinical decisions ultimately rests with the treating team. It has been suggested that these DL algorithms should have checks and balances that can use multidisciplinary decision-making with the fullest patient participation to arrive at any conclusion. Combining MRI-based radiomics classifications with a specialist radiologist's qualitative assessment may also improve the reliability of the prediction of which patients will be diagnosed with pCR after neoadjuvant treatment. Similarly, pathomics or the extraction of quantitative information from digital histopathological slides to reflect underlying genetic patterns or molecular characteristics has proved itself to be a useful adjunct to MRI-based radiomics.

Table 2 Deep neural algorithm data

Author (Year)	Deep neural algorithms	Software	Imaging type	MRI phase	Image Pre-processing	People involved
Feng (2022) [23]	VGG-19,CNN	Pyradiomics	MRI	T2/T1/DWI		radiologist, pathologist
Jin (2021) [24]	two Siamese subnetworks joined at multiple layers	Tensorflow	MRI - pre/post CRT	T2/DWI	Combination of 4-dimensional tensor images with last dimension being each of 4 MRI sequences	radiologist, pathologist
Li (2021) [25]	KNN	Python	MRI	T2		radiologist, pathologist
Huang (2020) [26]	ANN, KNN, SVM, NBC, MLR	Statistica 13.0	CT, MRI	NA	N/A	radiologist, pathologist
Fu (2020) [27]	CNN, VGG19	NR Matlab?	MRI	DWI	NR	surgeon, radiologist, pathologist
Zhang (2020) [28]	CNN	Python with Network architecture (neural network with 8 inputs) Python 3.6 -stochastic gradient descent algorithm with the adaptive moment estimation algorithm (known as ADAM) optimizer	MRI - pre/post CRT	T2	Freehand and manually with iTK-SNAP software	2 experienced pathologists, 2 experienced radiologists
Zhu (2020) [29]	CNN	Python 3.6	MRI	ADC / T 2		radiologist, pathologist
Shayesteh (2019) [30]	SVM, Bayesian network, ANN, KNN	SPSS modeler 18	MRI	T2	1. Collewet et al method 2. LOG filter	radiologist, radiation oncologist, pathologist
Ferrari (2019) [31]	deep neural network (unspecified)	NR	MRI - pre/mid/post CRT	T2	Manual segmentation by radiologisthen DICOM open source software used to extract textural features	radiologist, pathologist
Shi (2019) [32]	CNN	Tensorflow	MRI pre treatment and mid treatment 3-4 weeks post CRT	T2/DWI/DCE	Harlick's Gray level Co-occurence Matrix (GLCM) texture extraction; Tumour ROI analysis - MIM Meastro Software Radiomics - Matlab neural network Toolbox software version 7.2	radiation oncologist, pathologist
Bibault (2018) [33]	deep neural network (unspecified)	Tensorflow	CT	not stated	Eclipse V.13 (segmentation) IBEX (radiomics feature extraction)	radiation oncologist, pathologist
Nie (2016) [34]	ANN	Mattlab	MRI - pre/post CRT	T1/T2/DWI/DCE	MIMMaestro workstation by rectal MRI radiologist outlining region of interest (ROI) and transferred onto image maps	radiologist, pathologist

Table 3 Accuracy

Author (Year)	Comparison	Comparison details	Most accurate DL model	Model accuracy	Model sensitivity	Model specificity	Model AUC	Model PPV	Model NPV
Feng (2022) [23]	RAPIDS vs rMRI vs Pathomics nucleus model vs Pathomics microenvironment model	single modality	RAPIDS (RAdiomics Pathomics Integrated preDiction System)	NR	0.82	0.816	0.868	0.637	0.909
Jin (2021) [24]	-	combined with blood-based tumour markers to achieve higher AUC	DL model	-	0.92	0.91	0.92	0.83	0.96
Li (2021) [25]	KNN vs others	single modality	KNN	NR	0.857	0.985	0.945	NR	NR
Huang (2020) [26]	AAN vs other deep learning models	single modality	AAN	0.88	0.94	-	0.84	89	0.87
Fu (2020) [27]	Handcrafted classifier vs deep-learning based classifier	single modality	DL model	NA	NA	NA	0.73		
Zhang (2020) [28]	compared with radiologists	single modality	CNN	0.98	1	0.97	0.99	0.9	1
Zhu (2020) [29]		- single modality	CNN with ADC phase	NR	0.943	0.68	0.85	0.87	0.837
Shayesteh (2019) [30]	Multiple - ensemble learning model	single modality + combined	4 classifier (SVM. NN.BN.KNN) & BN (without pre-processed MRI)	0.95 (in validation sets)	-	-	0.928	-	-
Ferrari (2019) [31]	training (2 AI models) / validation	single modality	DL model	0.83	0.86	0.74	0.83	-	-
Shi (2019) [32]		- combined both pre-treatment and mid-treatment MRI to achieve higher AUC	DL model	NR	NR	NR	0.83	-	-

Table 3 (continued)

Author (Year)	Comparison	Comparison details	Most accurate DL model	Model accuracy	Model sensitivity	Model specificity	Model AUC	Model PPV	Model NPV
Bibault (2018) [33]	model (DNN) / 2nd model (Support Vector Model) / baseline (LR) using only TNM stage	5-fold cross-validation with original dataset partitioned into 5 equal sized subsamples, with a single subsample retained as the validataion data for testing the model & remaining 4 subsamples used as training data. Cross-validation process was then repeated 5-times with each of the 5 subsamples used once as the validation data	DNN	0.8	0.68	0.836	0.72	-	-
Nie (2016) [34]	ANN conventional volume-averaged analysis vs ANN voxelized heterogeneity analysis	single modality	ANN voxelized heterogeneity analysis	NA	NA	NA	0.71	-	-

In order to translate this into practical application, it has been suggested that clinical protocols include measures to mitigate any inaccuracy in prediction. It might be prudent to preference intensive neoadjuvant CRT or even TNT in these patients to maximise the likelihood of a pCR. Close follow-up with confirmative examinations such as digital rectal examination or endoscopic ultrasound examination, as well as fine needle aspiration of any suspicious lesions should be undertaken with this organ-preserving strategy.

Conclusion

While DL models have shown promise in predicting the group of patients who will have pCR after neoadjuvant CRT for LARC, further large prospective studies in collaboration with an experienced AI team are needed to measure the utility of this novel approach to radiological assessment [42].

Authors' contributions Sowmya Prabhakaran - Study conception and design, acquisition of data, analysis and interpretation of data, drafting of manuscript, critical revision of manuscript; Keith Wai Keong Choong - Study conception and design, acquisition of data, analysis and interpretation of data, drafting of manuscript, critical revision of manuscript; Swetha Prabhakaran - Analysis and interpretation of data, drafting of manuscript, critical revision of manuscript; Kay Tai Choy - Analysis and interpretation of data, drafting of manuscript, critical revision of manuscript; Joseph CH Kong - Study conception and design, acquisition of data, analysis and interpretation of data, drafting of manuscript, critical revision of manuscript.

Data availability Datasets used and analysed during the current study can be made available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors declare no competing interests.

References

1. Maas M, Nelemans PJ, Valentini V, Das P, Rödel C, Kuo LJ et al (2010) Long-term outcome in patients with a pathological complete response after chemoradiation for rectal cancer: a pooled analysis of individual patient data. Lancet Oncol 11(9):835–844
2. Janjan NA, Khoo VS, Abbruzzese J, Pazdur R, Dubrow R, Cleary KR et al (1999) Tumor downstaging and sphincter preservation with preoperative chemoradiation in locally advanced rectal cancer: the M. D. Anderson Cancer Center experience. Int J Radiat Oncol Biol Phys 44(5):1027–1038
3. Pucciarelli S, Friso ML, Toppan P, Fornasiero A, Carnio S, Marchiori E et al (2000) Preoperative Combined Radiotherapy and Chemotherapy for Middle and Lower Rectal Cancer: Preliminary Results. Ann Surg Oncol 7(1):38–44
4. Moreno CC, Sullivan PS, Mittal PK (2017) MRI Evaluation of Rectal Cancer: Staging and Restaging. Curr Probl Diagn Radiol 46(3):234–241
5. Panagiotopoulou IG, Parashar D, Qasem E, Mezher-Sikafi R, Parmar J, Wells AD et al (2015) Neoadjuvant Long-Course Chemoradiotherapy for Rectal Cancer: Does Time to Surgery Matter? Int Surg 100(6):968–973
6. Conroy T, Lamfichekh N, Etienne PL, Rio E, Francois E, Mesgouez-Nebout N et al (2020) Total neoadjuvant therapy with mFOLFIRINOX versus preoperative chemoradiation in patients with locally advanced rectal cancer: Final results of PRODIGE 23 phase III trial, a UNICANCER GI trial. J Clin Oncol 38:4007. https://doi.org/10.1200/JCO.2020.38.15_suppl.4007
7. Giunta EF, Bregni G, Pretta A, Deleporte A, Liberale G, Bali AM et al (2021) Total neoadjuvant therapy for rectal cancer: Making sense of the results from the RAPIDO and PRODIGE 23 trials. Cancer Treat Rev 96:102177
8. Bahadoer RR, Dijkstra EA, van Etten B, Marijnen CAM, Putter H, Kranenbarg EMK et al (2021) Short-course radiotherapy followed by chemotherapy before total mesorectal excision (TME) versus preoperative chemoradiotherapy, TME, and optional adjuvant chemotherapy in locally advanced rectal cancer (RAPIDO): a randomised, open-label, phase 3 trial. Lancet Oncol 22(1):29–42
9. Habr-Gama A, Perez RO, Nadalin W, Sabbaga J, Ribeiro U, Silva e Sousa AH et al (2004) Operative Versus Nonoperative Treatment for Stage 0 Distal Rectal Cancer Following Chemoradiation Therapy: Long-term Results. Ann Surg 240(4):711–718
10. Sammour T, Price BA, Krause KJ, Chang GJ (2017) Non-operative management or "watch and wait" for rectal cancer with complete clinical response after neoadjuvant chemoradiotherapy – a critical appraisal. Ann Surg Oncol 24(7):1904–1915
11. van der Valk MJM, Hilling DE, Bastiaannet E, Kranenbarg EMK, Beets GL, Figueiredo NL et al (2018) Long-term outcomes of clinical complete responders after neoadjuvant treatment for rectal cancer in the International Watch & Wait Database (IWWD): an international multicentre registry study. Lancet 391(10139):2537–2545
12. Kong JC, Guerra GR, Warrier SK, Ramsay RG, Heriot AG (2017) Outcome and Salvage Surgery Following "Watch and Wait" for Rectal Cancer after Neoadjuvant Therapy: A Systematic Review. Dis Colon Rectum 60(3):335–345
13. Smith JJ, Strombom P, Chow OS, Roxburgh CS, Lynn P, Eaton A et al (2019) Assessment of a Watch-and-Wait Strategy for Rectal Cancer in Patients With a Complete Response After Neoadjuvant Therapy. JAMA Oncol 5(4):e185896
14. Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI (2014) Machine learning applications in cancer prognosis and prediction. Comput Struct Biotechnol J 13:8–17
15. Kang J, Schwartz R, Flickinger J, Beriwal S (2015) Machine Learning Approaches for Predicting Radiation Therapy Outcomes: A Clinician's Perspective. Int J Radiat Oncol Biol Phys 93(5):1127–1135
16. Bibault JE, Giraud P, Burgun A (2016) Big Data and machine learning in radiation oncology: State of the art and future prospects. Cancer Lett 382(1):110–117
17. Naqa IE, Li R, Murphy MJ (eds) (2015) Machine learning in radiation oncology: theory and applications [Internet]. Springer International Publishing. [cited 2021 Jul 31]. Available from: https://www.springer.com/gp/book/9783319183046
18. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM et al (2017) Dermatologist-level classification of skin cancer with deep neural networks. Nature 542(7639):115–118
19. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A et al (2016) Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. JAMA 316(22):2402

20. Miotto R, Li L, Kidd BA, Dudley JT (2016) Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records. Sci Rep 6:26094
21. Oakden-Rayner L, Carneiro G, Bessen T, Nascimento JC, Bradley AP, Palmer LJ (2017) Precision Radiology: Predicting longevity using feature engineering and deep learning methods in a radiomics framework. Sci Rep 7:1648
22. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD et al (2021) The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 372:n71
23. Feng L, Liu Z, Li C, Li Z, Lou X, Shao L et al (2022) Development and validation of a radiopathomics model to predict pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer: a multicentre observational study. The Lancet Digit Health 4(1):e8–e17
24. Jin C, Yu H, Ke J, Ding P, Yi Y, Jiang X et al (2021) Predicting treatment response from longitudinal images using multi-task deep learning. Nat Commun 12:1851
25. Li Z, Ma X, Shen F, Lu H, Xia Y, Lu J (2021) Evaluating treatment response to neoadjuvant chemoradiotherapy in rectal cancer using various MRI-based radiomics models. BMC Med Imaging 21(1):30
26. Huang CM, Huang MY, Huang CW, Tsai HL, Su WC, Chang WC et al (2020) Machine learning for predicting pathological complete response in patients with locally advanced rectal cancer after neoadjuvant chemoradiotherapy. Sci Rep 10(1):12555
27. Fu J, Zhong X, Li N, Van Dams R, Lewis J, Sung K et al (2020) Deep learning-based radiomic features for improving neoadjuvant chemoradiation response prediction in locally advanced rectal cancer. Phys Med Biol 65(7):075001
28. Zhang XY, Wang L, Zhu HT, Li ZW, Ye M, Li XT et al (2020) Predicting Rectal Cancer Response to Neoadjuvant Chemoradiotherapy Using Deep Learning of Diffusion Kurtosis MRI. Radiology 296(1):56–64
29. Zhu HT, Zhang XY, Shi YJ, Li XT, Sun YS (2020) A Deep Learning Model to Predict the Response to Neoadjuvant Chemoradiotherapy by the Pretreatment Apparent Diffusion Coefficient Images of Locally Advanced Rectal Cancer. Front Oncol 10:574337 Available from: https://www.frontiersin.org/articles/10.3389/fonc.2020.574337
30. Shayesteh SP, Alikhassi A, Fard Esfahani A, Miraie M, Geramifar P, Bitarafan-rajabi A et al (2019) Neo-adjuvant chemoradiotherapy response prediction using MRI based ensemble learning method in rectal cancer patients. Physica Medica 62:111–119
31. Ferrari R, Mancini-Terracciano C, Voena C, Rengo M, Zerunian M, Ciardiello A et al (2019) MR-based artificial intelligence model to assess response to therapy in locally advanced rectal cancer. Eur J Radiol 118:1–9
32. Shi L, Zhang Y, Nie K, Sun X, Niu T, Yue N et al (2019) Machine learning for prediction of chemoradiation therapy response in rectal cancer using pre-treatment and mid-radiation multi-parametric MRI. Magn Reson Imaging 61:33–40
33. Bibault JE, Giraud P, Housset M, Durdux C, Taieb J, Berger A et al (2018) Deep Learning and Radiomics predict complete response after neo-adjuvant chemoradiation for locally advanced rectal cancer. Sci Rep 8(1):12611
34. Nie K, Shi L, Chen Q, Hu X, Jabbour SK, Yue N et al (2016) Rectal Cancer: Assessment of Neoadjuvant Chemoradiation Outcome based on Radiomics of Multiparametric MRI. Clin Cancer Res 22(21):5256–5264
35. Wells G, Shea B, O'Connell D, Peterson J, Welch V, Losos M et al (2000) The Newcastle–Ottawa Scale (NOS) for assessing the quality of non-randomized studies in meta-analysis
36. Di Re AM, Sun Y, Sundaresan P, Hau E, Toh JWT, Gee H et al (2021) MRI radiomics in the prediction of therapeutic response to neoadjuvant therapy for locoregionally advanced rectal cancer: a systematic review. Expert Rev Anticancer Ther 21(4):425–449
37. Bedrikovetski S, Dudi-Venkata NN, Maicas G, Kroon HM, Seow W, Carneiro G et al (2021) Artificial intelligence for the diagnosis of lymph node metastases in patients with abdominopelvic malignancy: A systematic review and meta-analysis. Artif Intell Med 113:102022
38. Van Der Valk MJM, Hilling DE, Bastiaannet E, Meershoek-Klein Kranenbarg E, Beets GL, Figueiredo NL et al (2018) Long-term outcomes of clinical complete responders after neoadjuvant treatment for rectal cancer in the International Watch & Wait Database (IWWD): an international multicentre registry study. Lancet 391(10139):2537–2545
39. Prabhakaran S, Yang TWW, Johnson N, Bell S, Chin M, Simpson P et al (2022) Latest evidence on the management of early-stage and locally advanced rectal cancer: a narrative review. ANZ J Surg 92(3):365–372
40. Gillies RJ, Kinahan PE, Hricak H (2016) Radiomics: Images Are More than Pictures, they are data. Radiology 278(2):563–577
41. Gan J, Wang J, Zhong H, Luo R, Zhou Z, Hu P et al (2016) MO-DE-207B-09: A Consistent Test for Radiomics Softwares. Med Phys 43(6Part30):3706–3706
42. Seah JCY, Tang CHM, Buchlak QD, Holt XG, Wardman JB, Aimoldin A et al (2021) Effect of a comprehensive deep-learning model on the accuracy of chest x-ray interpretation by radiologists: a retrospective, multireader multicase study. The Lancet Digit Health 3(8):e496–e506