

SCIENTIFIC REPORTS



Correction: Author Correction

OPEN

Deep Learning and Radiomics predict complete response after neo-adjuvant chemoradiation for locally advanced rectal cancer

Jean-Emmanuel Bibault^{1,2}, Philippe Giraud¹, Martin Housset¹, Catherine Durdux¹, Julien Taieb³, Anne Berger⁴, Romain Coriat^{5,6}, Stanislas Chaussade^{5,6}, Bertrand Dousset⁷, Bernard Nordlinger⁸ & Anita Burgun^{2,9}

Treatment of locally advanced rectal cancer involves chemoradiation, followed by total mesorectum excision. Complete response after chemoradiation is an accurate surrogate for long-term local control. Predicting complete response from pre-treatment features could represent a major step towards conservative treatment. Patients with a T2-4 N0-1 rectal adenocarcinoma treated between June 2010 and October 2016 with neo-adjuvant chemoradiation from three academic institutions were included. All clinical and treatment data was integrated in our clinical data warehouse, from which we extracted the features. Radiomics features were extracted from the tumor volume from the treatment planning CT Scan. A Deep Neural Network (DNN) was created to predict complete response, as a methodological proof-of-principle. The results were compared to a baseline Linear Regression model using only the TNM stage as a predictor and a second model created with Support Vector Machine on the same features used in the DNN. Ninety-five patients were included in the final analysis. There were 49 males (52%) and 46 females (48%). Median tumour size was 48 mm (15–130). Twenty-two patients (23%) had pathologic complete response after chemoradiation. One thousand six hundred eighty-three radiomics features were extracted. The DNN predicted complete response with an 80% accuracy, which was better than the Linear Regression model (69.5%) and the SVM model (71.58%). Our model correctly predicted complete response after neo-adjuvant rectal chemoradiotherapy in 80% of the patients of this multicenter cohort. Our results may help to identify patients who would benefit from a conservative treatment, rather than a radical resection.

The standard treatment for locally advanced rectal cancer is neoadjuvant chemoradiation followed by total mesorectal excision (TME) after a 6 to 10 week interval¹. In 20% to 30% of the patients, no residual tumor is found at histopathology². In this selected group of patients, it has been suggested that surgery could be omitted, since it did not improve outcome and was associated with a high rate of morbidity³. This strategy has initially been

¹Radiation Oncology Department, Georges Pompidou European Hospital, Assistance Publique – Hôpitaux de Paris, Paris Descartes University, Paris Sorbonne Cité, Paris, France. ²INSERM UMR 1138 Team 22: Information Sciences to support Personalized Medicine, Paris Descartes University, Sorbonne Paris Cité, Paris, France. ³Department of gastroenterology and digestive oncology, Georges Pompidou European Hospital, Assistance Publique – Hôpitaux de Paris, Paris Descartes University, Paris Sorbonne Cité, Paris, France. ⁴Department of General Surgery and Surgical Oncology, Georges Pompidou European Hospital, Assistance Publique – Hôpitaux de Paris, Paris Descartes University, Paris Sorbonne Cité, Paris, France. ⁵Gastroenterology and digestive oncology unit, Cochin University Hospital, Assistance Publique – Hôpitaux de Paris, Paris, France. ⁶INSERM 1016, Université Paris Descartes, Paris, France. ⁷Department of Digestive, Hepato-biliary and Endocrine Surgery, Cochin Hospital, Assistance Publique – Hôpitaux de Paris, Paris, France. ⁸Department of General Surgery and Surgical Oncology, Hôpital Ambroise Paré, Assistance Publique – Hôpitaux de Paris, Boulogne-Billancourt, France. ⁹Biomedical Informatics and Public Health Department, Georges Pompidou European Hospital, Assistance Publique – Hôpitaux de Paris, Paris Descartes University, Paris Sorbonne Cité, Paris, France. Correspondence and requests for materials should be addressed to J.-E.B. (email: jean-emmanuel.bibault@aphp.fr)

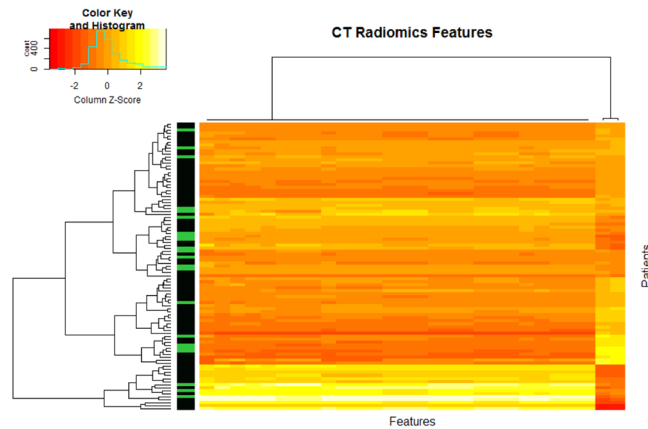


Figure 1. Heatmap of radiomics features correlated to complete response.

met with significant skepticism, but organ preservation strategies with watch-and-wait or transanal endoscopic microsurgery have since shown good long-term results^{2,4–8}.

Complete pathologic response (pCR) to neo-adjuvant chemoradiation is assessed during pathological examination after surgery. Identifying patients in pCR with a high rate of accuracy could lead to improved clinical outcome. Computational Imaging, also known as Radiomics, is the use of imaging data from routine clinical work-up to assess the tumor characteristics, such as spatial heterogeneity, texture or shape. This approach is transforming imaging into a high-throughput data mine that can be leveraged and analyzed with other clinical features for precision medicine and decision support. Its potential is currently being explored in several clinical setups⁹, including rectal cancer¹⁰.

Deep Learning (DL) is a subfield of machine learning and artificial intelligence that is increasingly used in medicine^{11–14} for diagnosis¹⁵, classification¹⁶, or prediction^{17,18}. In this study, we present a novel approach combining Deep Learning with clinical and radiomics features to build a model predicting pCR in a multicenter cohort of patients with locally-advanced rectal cancer treated with neo-adjuvant chemoradiation, followed by surgery.

Results

Radiomics features. One thousand six hundred eighty-three features were extracted from the two segmentations of the tumor volume, for each patient (319770 features in total). One hundred and twenty-four features (7.3%) had an Intraclass Correlation Coefficient higher than 0.8 in the following categories: Texture (Grey-Level Co-Occurrence Matrix in 2D and 3D, Grey-Level Run Length Matrix, Intensity Direct and Intensity Histogram) and Shape. Out of these, 28 features (22%) were filtered on the basis that they were significantly correlated to pathological Complete Response (Wilcoxon Test $p < 0.05$) in three categories of features: Gray-Level Co-occurrence Matrix 2D and 3D and IntensityDirect. The heatmap showed clustering of these features in two groups of patients (Fig. 1).

Deep Learning network training and testing. As a baseline, another model created with Linear Regression, using only the TNM stage, showed a lower accuracy of 69,5% (95% CI = 59,2% to 78,51%). Sensitivity was 34,78% (95% CI = 16,38% to 57,27%), specificity was 80,56% (95% CI = 69,53% to 88,94%). Mean area under the curve (AUC) for the LR model was 0.59 (95% CI = 0.46 to 0.69). In the DNN, 29 variables (T stage and the robust, filtered 28 radiomics features) were included. The DNNClassifier predicted pCR with an 80% accuracy (95% CI = 70,54% to 87,51%). Sensitivity was 68,2% (95% CI = 45,13% to 86,14%), specificity was 83,56% (95% CI = 73,05% to 91,21%). Mean AUC for the DNN model was 0.72 (95% CI = 0.65 to 0.87). A comparison of the main metrics of the model (accuracy, AUC, false and true positives and negatives rates) did not reveal any significant differences in the results between the folds (Chi-Squared test, $p > 0.05$). Increasing the number of hidden layers or the neurons in each layer did not improve performance: accuracy dropped to 70% when ten hidden layers were used with 100 neurons in each layer. A ten-fold increase of the number of learning steps ($n = 30,000$) or epochs ($n = 10$) for network training did not improve accuracy (80%).

As a comparison, a Support Vector Machine (SVM) model was created. The accuracy of this model, trained on the same features with a 5-fold cross validation, was 71.58% (95% CI = 61.40% to 80.36%). Sensitivity was 45.45% (95% CI = 24.39% to 67.79%), specificity was 79.45% (95% CI = 68.38% to 88.02%). Mean AUC for the SVM model was 0.62 (95% CI = 0.51 to 0.74). Confusion matrices for the LR, DNN and SVM models are shown in Table 1. There was no statistical difference for the metrics of the model between the fold (Chi-Squared test, $p > 0.05$).

There was no statistical correlation between pCR and overall survival (log-rank test, $p = 0.258$), but none of the patients in pCR died during the study (Fig. 2).

Discussion

This is the first study that used Deep Learning to predict pCR after neo-adjuvant chemoradiation in locally advanced rectal cancer. Routine care data were extracted from our CDW to create a profile of the patients. T stage was the only variable significantly correlated to pCR. It was combined with a high-dimensional radiomics phenotype, extracted from CT scans, in a Deep Neural Network to predict pCR after neoadjuvant chemoradiation. The

		Predicted class		
		pCR: n (% of actual class)	Non pCR: n (% of actual class)	Total
Actual class	pCR	Baseline LR: 8 (36%)	Baseline LR: 14 (64%)	22
		DNN: 15 (68%)	DNN: 7 (32%)	
		SVM: 10 (45,4%)	SVM: 12 (54,5%)	
	Non pCR	Baseline LR: 15 (21%)	Baseline LR: 58 (79%)	73
		DNN: 12 (16%)	DNN: 61 (84%)	
		SVM: 15 20%)	SVM: 62 (80%)	
	Total	LR: 23	LR: 72	95
		DNN: 27	DNN: 68	
		SVM: 25	SVM: 70	

Table 1. Confusion matrices: Baseline LR: Linear Regression model with T stage; DNN: Deep Neural Network model with 29 variables); SVM: Support Vector Machine Model with the same 29 variables.

DNN predicted pCR correctly in 80% of the cases. Creating a profile of each patient was made possible through the use of our CDW, in which medical records of all our patients are prospectively and automatically stored. Data from 750,000 patients are stored in the HEGP CDW, including 14,000 cancer patients treated with radiation. A request can easily be created in i2b2 to identify cohorts of patients that can later be used to extract structured and unstructured data from the CDW, using custom-made software¹⁹. In the field of radiation oncology, data from treatment planning and delivery can also be easily extracted from the Treatment Planning and Record-and-Verify systems.

Computational Imaging consists in extracting quantitative features from CT scan, MRI or PET/CT²⁰. Medical images are no longer simple pictures to interpret visually; they are now treated as data. These analyses are intended to be conducted with routine care images and could be used like any other data for target volume delineation or decision-making²¹. However, there are several tools available and no standard way to extract radiomics features, meaning that reproducibility is a key challenge in this field²². We used an open and free tool that was designed for collaboration²³. We provide the parameters we used for feature extraction and selection in Supplementary File 1 and online²⁴. Robustness of the extracted features was assessed with ICC computation, and only features with a high ICC (>0.8) were kept.

Other studies have already been published, using MRI²⁵ or PET/CT²⁶ with logistic regression or simpler Artificial Neural Network (ANN) to predict treatment response, with correct accuracy (AUC = 0.71–0.79 for the MRI model, data not provided for the PET/CT model). The latest study included 222 patients to build a radiomics signature with 30 MRI features²⁵. The model was created with Support Vector Machine, another Machine Learning approach¹³, and found an area under the receiver operating characteristic curve of 0.9756 (95% confidence interval, 0.9185–0.9711) in the validation cohort. However, feature extraction from MRI is even more complex and less reproducible than CT Scan, meaning that these results cannot be easily reproduced on another cohort²⁷. Another frequent limit of radiomics studies is that they often explore the prognostic relevance of imaging features, without using any clinical, biological or treatment delivery data. This was not the case in our study since we used several other inputs beyond radiomics into our DNN. Our study has some limitations. First, the sample size is limited with a large number of predictors. In our cohort, patients were referred from three different hospitals for neoadjuvant chemoradiation and were then treated in their respective institution, which limits the risk of overfitting (i.e. when a neural network has been trained on a dataset and is unable to accurately predict the outcome on another unknown dataset). A 5-fold cross validation was performed. The feature selection performed before training could have caused some test-set leakage. To limit this, we selected features from 4 of the 5 partitioned datasets, leaving the validation set from the first validation. Since the accuracy and AUC of the model is consistent across all five cross-validations, we do not believe leakage had a significant effect on the model.

Physicians cannot intuitively understand the results given by the model we created: Deep Learning essentially remains a black box. With the use of a large number of variables and second order statistical data, such as radiomics, we believe this cannot be avoided. Visualization techniques such as a radiomics heatmap can only give a high-level representation of the data. These publications are hypothesis-generating studies and can help in identifying relevant prognostic or predictive factors, but their level of evidence is still low, no matter how innovative their approach is. It is considered that a human brain can only integrate up to 5 variables in order to make an adequate decision^{28,29}. Since oncology is relying on an increasing amount of data of different types, using computers as Clinical Decision Support Systems (CDSS) could become mandatory. Deep Learning, and Artificial Intelligence in a broader sense, will eventually disrupt the way we practice medicine in positive^{11,30,31} and negative³² ways. Among the disciplines poised to be radically changed, is medical imaging. Several studies have recently been published predicting longevity from routine CT-Scans¹⁸ or detecting pneumonia from chest X-Ray³³. The development of Deep Learning will transform the way we use imaging for diagnosis, treatment planning and decision making. It is not clear yet if these methods should be assessed as any other medical device in a randomized trial or if new approaches are needed.

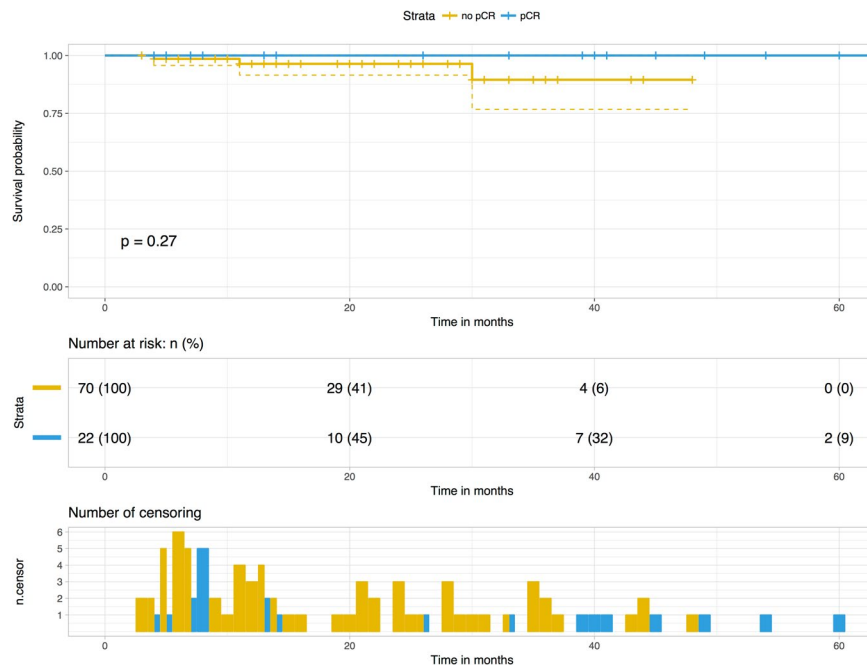


Figure 2. Kaplan-Meier curves for overall survival stratified on pathological complete response.

Conclusion

In this proof-of-concept study, we show that using a DNNClassifier on heterogeneous data combining clinical and radiomics features is feasible and can accurately predict patients who will have a complete pathological response after neo-adjuvant chemoradiotherapy for locally-advanced rectal cancer. For this subset of patients, conservative treatments could be a valid approach, with less long-term side effects. After careful prospective evaluation of this approach in a randomized clinical trial, this kind of methods could be directly implemented within the treatment planning systems used in radiation oncology to better personalize treatments.

Methods

Ethical statement. This study was approved by the IRB and ethics committee CPP Ile-de-France II: IRB Committee # 00001072, study reference # CDW_2015_0024. All experiments were carried out in accordance with relevant guidelines and regulations.

The study used only pre-existing medical data, therefore patient consent was not required by the ethics committee.

Dataset description. Patients with a T2-4 N0-1 rectal adenocarcinoma treated between June 2010 and October 2016 with neo-adjuvant chemoradiation (4 to 50.4 Gy) with Capecitabine (800 mg/m² twice a day) were included in the study. The patient recruitment originated from three academic institutions: Hôpital Européen Georges Pompidou (HEGP), Hôpital Cochin (HC) and Hôpital Ambroise Paré (HAP), belonging to the Assistance Publique – Hôpitaux de Paris. All patients had a pelvic MRI and PET-CT for staging. Chemoradiation was performed in our department and surgery was performed later 6 to 10 weeks by the surgery department in each recruiting institution. Ninety-five patients from three different institutions (HEGP: n = 35, 37%; HC: n = 23, 24%; HAP: n = 37, 39%) were included in the final analysis. Median follow-up was 16 months (range: 3–65). There were 49 males (52%) and 46 females (48%). Median age was 66 years old (32–84). Median tumor size was 48 mm (15–130). There were 9 T2 (9%), 75 T3 (79%) and 11 T4 (12%) tumors. Nineteen patients (20%) had no lymph node metastasis on pelvic MRI and there were 76 N+ patients (80%). Median baseline hemoglobin, neutrophils and lymphocytes counts were 13.6 g/dl (9.7–17.5), 1734/mm³ (336–3760) and 4050/mm³ (1100–11160) respectively. Delivered median doses were 50.4 Gy (45–50.4) to the GTVp and 45 Gy to the CTV. Median dose per fraction was 2 Gy (1.8–2.25 Gy). Median treatment length was 39 days (32–69). Average time between chemoradiation and surgery was 9 weeks (min = 4; max = 11). Twenty-two patients (23%) had pathologic complete response after chemoradiation. Forty-two patients received adjuvant chemotherapy. Patients' characteristics from each institution are shown in Table 2. Two patients had a local relapse (2.1%, none in the pCR group) and 7 a distant relapse (7.3%, none in the pCR group) during follow-up. Disease-free and overall survival rates at follow-up were 90.53% and 96.85% respectively. Among the variables extracted from our CDW, T stage was the only feature significantly correlated with pCR (Chi-Squared test, p = 0.036).

Clinical, pathological, biological features extraction. All features were extracted from our clinical data warehouse (CDW), that relies on the Informatics for Integrating Biology and the Bedside (i2b2) model - an open source infrastructure developed by Harvard Medical School and adopted by more than 130 academic

Characteristics	HEGP (n = 35–37%)	HC (n = 23–24%)	HAP (n = 37–39%)	Chi-Squared Test (p)
Sex				
Male	18 (51.4%)	14 (61%)	17 (54%)	0.603
Female	17 (48.6%)	9 (39%)	20 (46%)	
Median age (range)	65 (34–79)	61 (37–84)	70 (32–84)	0.768
T stage				
2	2 (5.7%)	0 (0%)	7 (19%)	0.587
3	27 (77.2%)	20 (87%)	28 (75.6%)	
4	6 (17.1%)	3 (13%)	2 (5.4%)	
N stage				
0	9 (25.4%)	4 (17.4%)	6 (16%)	0.588
1	26 (74.6%)	19 (82.6%)	31 (84%)	
Median tumor size in mm (range)	50 (15–130)	48 (30–65)	45 (20–99)	0.954
Tumor differentiation				
Grade 1	0 (0%)	0 (0%)	0 (0%)	0.118
Grade 2	0 (0%)	2 (8.7%)	1 (2%)	
Grade 3	16	12 (52.2%)	16 (44%)	
Grade 4	19	9 (39.1%)	20 (54%)	
Median baseline hemoglobin in g/dl (range)	13.3 (9.7–15.5)	13.6 (11.3–17.5)	13.5 (10.3–17)	0.220
Median baseline neutrophils in /mm ³ (range)	3659 (1100–9380)	3879 (1820–10925)	3796 (1740–11160)	0.405
Median baseline lymphocytes in /mm ³ (range)	1638 (336–3504)	1740 (1159–3240)	1844 (520–3760)	0.327
Median dose to GTVp in Gy (range)	50.4 (45–50.4)	50 (45–50.4)	50 (45–50.4)	0.127
Median dose to CTV in Gy (range)	45 (45–46)	46 (45–49.5)	46 (45–46)	0.204
Median dose per fraction in Gy (range)	1.8 (1.8–2)	2 (1.8–2.25)	2 (1.8–2)	0.001
Median treatment length in days (range)	48 (32–54)	37 (32–69)	37 (32–113)	0.181
Pathological complete response rate (n - %)	9 (25.7%)	3 (12.5%)	10 (27%)	0.387

Table 2. Patient characteristics.

hospitals around the world^{34,35}. The i2b2 warehouse uses an Entity-Attribute-Value (EAV) data model for its adaptability and dynamic nature. Concepts are stored separately in a hierarchical data model. The following features were extracted: age, sex, smoking status, tumor differentiation and size, T and N stages, baseline hemoglobin, neutrophils and lymphocytes counts.

Treatment planning and delivery features extraction. Data on treatment planning and delivery in our institution (2001–2016) were extracted from the ARIA system using reverse engineering and the VARIAN ESAPI³⁶ for dose-volume histograms (DVH). Structures labels were sorted and filtered by number of occurrences. Each of them was then matched to the ROS ontology before integration into the CDW^{37–39}. Delivered total dose to the GTVp and CTV, dose per fraction and treatment length were then extracted for modeling.

Radiomics features extraction, quality control and filtering. All included patients' treatment planning CT-Scans were extracted from the institution PACS and reused. CT-Scan were performed on scanned on a General Electric Light Speed scanner (Boston, Massachusetts, USA) as follow: helical acquisition, contiguous slices of 1.25 mm, 512 × 512 × 512 matrix, 120 kV, mAs > 350, speed: 3.75, mode: 0.75, with contrast injection, software version 0.7MW11.10. The rectal Primary Gross Tumor Volume (GTVp) was manually segmented by two expert gastrointestinal radiation oncologists on Eclipse V.13 (Varian, Palo Alto, California, USA) before being exported for radiomics features extraction. All DICOM and RT-STRUCT data were imported into IBEX (MD Anderson, Texas, USA), an open infrastructure platform dedicated to radiomics features extraction²³. Each patient was segmented twice. Shape, Intensity Direct, Gray Level Co-Occurrence Matrix (GLCM) 25 (computed from all 2D image slices) and 3 (computed from all 3D image matrices), Neighbor Intensity Difference (NID) 3 and 25, Gray Level Run Length Matrix 25 were extracted without pre-processing on each segmentation. In all, 319 770 features were extracted (1683 features extracted for each segmentation on each 95 patients). All extraction categories are provided (Supplementary File 1). To estimate the robustness of the tumor features, the intra-class correlation coefficient (ICC) was calculated^{40,41}. ICC can be used when quantitative measurements are made on units that are organized into groups⁴². It ranges between 0 and 1, indicating null and perfect reproducibility. In order to determine the ICC (equation 1) for inter-observer segmentations, variance estimates were obtained from two-way mixed effect model of analysis of variance (ANOVA):

$$ICC = \frac{MSr - MSw}{MSr + (k - 1)MSw} \quad (1)$$

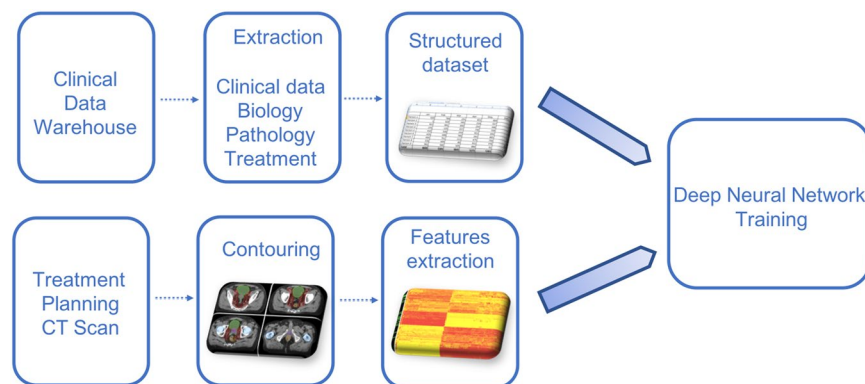


Figure 3. Global analysis pipeline.

where MSR = mean square for rows, MSW = mean square for residual source of variance, k = number of observers involved and n = number of subjects. R version 3.4.2 with the ICC package was used for computation⁴³.

Statistical analysis. Chi-squared tests were performed to evaluate differences in patients' characteristics between institutions and to select categorical variables correlated to pCR ($p < 0.05$). Wilcoxon tests were performed to filter features significantly correlated ($p < 0.05$) with pCR. As Parmar *et al.* showed, Wilcoxon test-based feature selection has the highest prognostic performance with high stability against data perturbation⁴⁴. A heatmap showing radiomics features clustering and correlations to pCR was generated. Survival rates were calculated from the date of surgery to create a Kaplan-Meier curve for overall survival. All statistical analysis were performed with R version 3.4.2⁴⁵ with the ggplot2⁴⁶, survival⁴⁷ and survminer⁴⁸ packages.

A 5-fold cross validation was performed: the original dataset was randomly partitioned into 5 equal sized subsamples. Of the 5 subsamples, a single subsample was retained as the validation data for testing the model, and the remaining 4 subsamples were used as training data. The cross-validation process was then repeated 5 times, with each of the 5 subsamples used once as the validation data. Values are reported as a mean of the 5 models. To limit test-set leakage, we calculated the ICC and the Wilcoxon correlation in 4 of the 5 partitioned datasets (that were created for the 5-fold cross-validation), leaving out the validation set from the first validation.

Deep learning training and validation. Robust features significantly correlated to pCR were used as inputs to a Deep Neural Network (DNN) created using the DNNClassifier Custom Estimator from the TensorFlow open-source framework (v1.3, Google, Mountain View, California, USA)⁴⁹. We explored a range of combinations of batch size, layer depth and layer size. We determined the optimal architecture for this deep learning model empirically, testing numerous variants. Changing the depth of the network reduced performance. We did not increase our model depth beyond 10,000 hidden units due to computational constraints. The resulting DNN was a compromise between performance and computational cost and included three hidden layers with 10, 20 and 10 neurons respectively. A Rectified Linear Unit (ReLU) function⁵⁰ was used for activation in the hidden layers because it resulted in a faster training: $F(x) = \max(x, 0)$, where x is the input to a neuron. Gradient descent was performed with the Adagrad Optimizer (equation 2)⁵¹:

$$\Theta_{t+1} = \Theta_t - \frac{\eta}{\sqrt{G_t + \epsilon}} * g_t \quad (2)$$

where Θ are parameters, t is the time-step, η is the learning rate, g_t is the gradient, G_t is a matrix of the sum of the squares of gradients up to time step t , and ϵ is a smoothing term that avoids division by zero. Adagrad adapts the learning rate to the parameters, performing larger updates for infrequent and smaller updates for frequent parameters. For this reason, it is well-suited for sparse data. The output of the network was binary (pCR or no pCR).

To avoid overfitting, a low number of epoch was chosen (3000 steps, 1 epoch). Training and validation was performed on a Linux Ubuntu 17.04 workstation with a Quad Core 2.8 Ghz Intel Core i7-770HQ and a GeForce GTX1060 Graphics Processing Unit (GPU). Results were visualized with the TensorBoard suite (Google, Mountain View, California, USA).

A logistic regression model was built from the same training and testing datasets using only the TNM stage as a baseline comparison, with the glmnet R package⁵². A Support Vector Machine model was created with the same variables as the DNNClassifier with Sci-Kit Learn⁵³. AUCs were calculated with the pROC R package^{39,54}.

The global analysis pipeline is shown in Fig. 3.

Data Availability. The datasets generated during and/or analyzed during the current study are not publicly available due to the clinical and confidential nature of the material but can be made available from the corresponding author on reasonable request.

References

- National Comprehensive Cancer Network. NCCN Clinical Practice Guidelines in Oncology. Available at: https://www.nccn.org/professionals/physician_gls/default.aspx#site. (Accessed: 27th November 2017)
- Maas, M. *et al.* 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**, 835–844 (2010).
- Habr-Gama, A. *et al.* Operative versus nonoperative treatment for stage 0 distal rectal cancer following chemoradiation therapy: long-term results. *Ann. Surg.* **240**, 711–717; discussion 717–718 (2004).
- Bhangu, A. *et al.* Survival outcome of local excision versus radical resection of colon or rectal carcinoma: a Surveillance, Epidemiology, and End Results (SEER) population-based study. *Ann. Surg.* **258**, 563–569; discussion 569–571 (2013).
- Habr-Gama, A. *et al.* Local recurrence after complete clinical response and watch and wait in rectal cancer after neoadjuvant chemoradiation: impact of salvage therapy on local disease control. *Int. J. Radiat. Oncol. Biol. Phys.* **88**, 822–828 (2014).
- Appelt, A. L. *et al.* High-dose chemoradiotherapy and watchful waiting for distal rectal cancer: a prospective observational study. *Lancet Oncol.* **16**, 919–927 (2015).
- Rehnan, A. G. *et al.* Watch-and-wait approach versus surgical resection after chemoradiotherapy for patients with rectal cancer (the OnCoRe project): a propensity-score matched cohort analysis. *Lancet Oncol.* **17**, 174–183 (2016).
- Martens, M. H. *et al.* Long-term Outcome of an Organ Preservation Program After Neoadjuvant Treatment for Rectal Cancer. *J. Natl. Cancer Inst.* **108** (2016).
- Aerts, H. J. W. L. The Potential of Radiomic-Based Phenotyping in Precision Medicine: A Review. *JAMA Oncol.* **2**, 1636–1642 (2016).
- Nie, K. *et al.* Rectal Cancer: Assessment of Neoadjuvant Chemoradiation Outcome based on Radiomics of Multiparametric MRI. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **22**, 5256–5264 (2016).
- Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V. & Fotiadis, D. I. Machine learning applications in cancer prognosis and prediction. *Comput. Struct. Biotechnol. J.* **13**, 8–17 (2015).
- Kang, J., Schwartz, R., Flickinger, J. & Beriwal, S. Machine Learning Approaches for Predicting Radiation Therapy Outcomes: A Clinician's Perspective. *Int. J. Radiat. Oncol. Biol. Phys.* **93**, 1127–1135 (2015).
- Bibault, J.-E., Giraud, P. & Burgun, A. Big Data and machine learning in radiation oncology: State of the art and future prospects. *Cancer Lett.* **382**, 110–117 (2016).
- Machine Learning in Radiation Oncology - Theory and | Issam El Naqa | Springer.*
- Esteva, A. *et al.* Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **542**, 115–118 (2017).
- Gulshan, V. *et al.* Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA* **316**, 2402–2410 (2016).
- Miotto, R., Li, L., Kidd, B. A. & Dudley, J. T. Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records. *Sci. Rep.* **6**, 26094 (2016).
- Oakden-Rayner, L. *et al.* Precision Radiology: Predicting longevity using feature engineering and deep learning methods in a radiomics framework. *Sci. Rep.* **7**, 1648 (2017).
- Escudié, J.-B. *et al.* Reviewing 741 patients records in two hours with FASTVISU. *AMIA Annu. Symp. Proc. AMIA Symp.* **2015**, 553–559 (2015).
- Gillies, R. J., Kinahan, P. E. & Hricak, H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology* **278**, 563–577 (2016).
- Lambin, P. *et al.* Radiomics: the bridge between medical imaging and personalized medicine. *Nat. Rev. Clin. Oncol.* **14**, 749–762 (2017).
- Gan, J. *et al.* MO-DE-207B-09: A Consistent Test for Radiomics Softwares. *Med. Phys.* **43**, 3706 (2016).
- Zhang, L. *et al.* IBEX: an open infrastructure software platform to facilitate collaborative work in radiomics. *Med. Phys.* **42**, 1341–1353 (2015).
- Bibault, J.-E. *Deep Radiomics Rectal Cancer: Code, parameters and scripts used for our study predicting pathologic complete response after neo-adjuvant chemoradiation for locally advanced rectal cancer* (2017).
- Liu, Z. *et al.* Radiomics Analysis for Evaluation of Pathological Complete Response to Neoadjuvant Chemoradiotherapy in Locally Advanced Rectal Cancer. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* <https://doi.org/10.1158/1078-0432.CCR-17-1038> (2017).
- Lovinfosse, P. *et al.* FDG PET/CT radiomics for predicting the outcome of locally advanced rectal cancer. *Eur. J. Nucl. Med. Mol. Imaging.* <https://doi.org/10.1007/s00259-017-3855-5> (2017).
- Kumar, V. *et al.* Radiomics: the process and the challenges. *Magn. Reson. Imaging* **30**, 1234–1248 (2012).
- Abernethy, A. P. *et al.* Rapid-learning system for cancer care. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **28**, 4268–4274 (2010).
- Obermeyer, Z. & Lee, T. H. Lost in Thought - The Limits of the Human Mind and the Future of Medicine. *N. Engl. J. Med.* **377**, 1209–1211 (2017).
- Applying machine learning to radiotherapy planning for head & neck cancer. *DeepMind* Available at: <https://deepmind.com/blog/applying-machine-learning-radiotherapy-planning-head-neck-cancer/> (Accessed: 17th May 2017).
- Beam, A. L. & Kohane, I. S. Translating Artificial Intelligence Into Clinical Care. *JAMA* **316**, 2368–2369 (2016).
- Cabitza, F., Rasoini, R. & Ginsini, G. F. Unintended Consequences of Machine Learning in Medicine. *JAMA* **318**, 517–518 (2017).
- Rajpurkar, P. *et al.* CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. (2017).
- Zapletal, E., Rodon, N., Grabar, N. & Degoulet, P. Methodology of integration of a clinical data warehouse with a clinical information system: the HEGP case. *Stud. Health Technol. Inform.* **160**, 193–197 (2010).
- Rance, B., Canuel, V., Countouris, H., Laurent-Puig, P. & Burgun, A. Integrating Heterogeneous Biomedical Data for Cancer Research: the CARPEM infrastructure. *Appl. Clin. Inform.* **7**, 260–274 (2016).
- Varian Incorporated. Varian Developers Forum. *CodePlex* Available at: <https://variandevolver.codeplex.com/Wikipage?ProjectName=variandevolver>, (Accessed: 15th April 2017).
- Bibault, J.-E. Radiation Oncology Structures Ontology - Summary | NCBO BioPortal. Available at: <http://bioportal.bioontology.org/ontologies/ROS/?p=summary>. (Accessed: 15th April 2017).
- Bibault, J.-E., Zapletal, E., Rance, B., Giraud, P. & Burgun, A. Labeling for Big Data in radiation oncology: The Radiation Oncology Structures ontology. *PLoS One* **13**, e0191263 (2018).
- Zapletal, E., Bibault, J.-E., Giraud, P. & Burgun, A. Integrating Multimodal Radiation Therapy Data into i2b2. *Applied Clinical Informatics* **09**(02), 377–390 (2018).
- Grove, O. *et al.* Quantitative computed tomographic descriptors associate tumor shape complexity and intratumor heterogeneity with prognosis in lung adenocarcinoma. *PLoS One* **10**, e0118261 (2015).
- Parmar, C. *et al.* Robust Radiomics feature quantification using semiautomatic volumetric segmentation. *PLoS One* **9**, e102107 (2014).
- Shrout, P. E. & Fleiss, J. L. Intraclass correlations: uses in assessing rater reliability. *Psychol. Bull.* **86**, 420–428 (1979).
- Wolak, M. *ICC: Facilitating Estimation of the Intraclass Correlation Coefficient*, (2015).
- Parmar, C., Grossmann, P., Bussink, J., Lambin, P. & Aerts, H. J. W. L. Machine Learning methods for Quantitative Radiomic Biomarkers. *Sci. Rep.* **5**, 13087 (2015).
- R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/> (2013).
- Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (2009).
- Therneau, T. M. & until 2009 T. L. (original S. >R port and maintainer. *survival: Survival Analysis*. (2017).

48. Kassambara, A., Kosinski, M., Biecek, P. & Fabian, S. *survminer: Drawing Survival Curves using 'ggplot2'* (2017).
49. Martin Abadi *et al.* *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. (2015).
50. Nair, V. & Hinton, G. Rectified Linear Units Improve Restricted Boltzmann Machines. *Proceedings of the 27th International Conference on Machine Learning* (2010).
51. Duchi, J., Hazan, E. & Singer, Y. Adaptive Subgradient Methods for Online Learning and Stochastic Optimization. *J. Mach. Learn. Res.* **12**, 2121–2159 (2011).
52. Friedman, J., Hastie, T., Simon, N., Qian, J. & Tibshirani, R. *glmnet: Lasso and Elastic-Net Regularized Generalized Linear Models*. (2017).
53. scikit-learn: machine learning in Python — scikit-learn 0.19.1 documentation. Available at: <http://scikit-learn.org/stable/> (Accessed: 27th April 2018).
54. Robin, X. *et al.* pROC: Display and Analyze ROC Curves. (2018).

Author Contributions

J.E.B. conceived and developed the research idea with the assistance of P.G. and A.B. The study design was performed by J.E.B, P.G. and A.B. J.E.B, M.H. and C.D performed the segmentation task. J.E.B., M.H., C.D., J.T., A.B., B.D., R.C., S.C., B.N. performed data collection. J.E.B. performed the computational experiments with guidance from P.G., A.B. The paper was written by J.E.B. with technical content from A.B., and extensive editorial input from all authors.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-30657-6>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018