

Uniwersytet Medyczny w Łodzi

Wydział Lekarski

Rozprawa na stopień doktora nauk medycznych

Ocena wartości predykcyjnej obrazów RM

***w skojarzonym leczeniu chorych na miejscowo
zaawansowanego raka odbytnicy.***

Michał Masłowski

Lekarz

Promotor: prof. dr hab. n. med. Jacek Fijuth

**Pracę wykonano w
Zakładzie Radioterapii
Uniwersytetu Medycznego w Łodzi
Kierownik: prof. dr hab. n. med. Jacek Fijuth**

Łódź 2022

8 Streszczenie

Wstęp: Rak jelita grubego stanowi na świecie trzecią pod względem częstości grupę nowotworów złośliwych. Zachorowalność w Polsce stale wzrasta, stanowi szóstą wśród mężczyzn i dziewiątą wśród kobiet przyczynę zachorowań na nowotwory złośliwe. W ostatnim czasie strategie leczenia raka odbytnicy ulegają znaczącym modyfikacjom. Zarówno radioterapia, leczenie systemowe jak i procedury chirurgiczne wykazują coraz wyższą skuteczność. U pacjentów pierwotnie resekcyjnych podejmowane są próby ograniczania wskazań do radioterapii, natomiast u chorych u których uzyskano całkowitą odpowiedź na leczenie neoadjuwantowe coraz częściej rozważa się odstępowanie od leczenia operacyjnego. Tak wiele możliwości terapeutycznych, wymaga odpowiedniej selekcji pacjentów oraz decyzji podejmowanych przez wielodyscyplinarne konsylia onkologiczne. Radioterapia lub radiochemioterapia jest opcją terapeutyczną w leczeniu raka gruczołowego odbytnicy bez względu na zaawansowanie. Podkreśla się konieczność wnikliwej oceny czynników zwiększających ryzyko wznowy/przerzutów i potencjalnej resekcyjności guza na podstawie badania RM. Coraz większa skuteczność leczenia neoadjuwantowego poprzez dodatkową jego intensyfikację w postaci całkowitego leczenia neoadjuwantowego - TNT, intencja poprawy jakości życia poprzez zachowanie zwieraczy i brak konieczności wykonania zabiegu operacyjnego stały się podstawą do wdrożenia koncepcji *“watch-and-wait”/“watch-and-see”*. Różnorodność opcji leczenia implikuje konieczność szczegółowej diagnostyki i precyzyjnego stopniowania zaawansowania. Badaniem z wyboru w ocenie miejscowego zaawansowania nowotworu jest RM miednicy. Badanie RM, oprócz podstawowej informacji dotyczącej zaawansowania, może być poddane analizie komputerowej mającej na celu uzyskanie większej ilości danych, dziedzinę wiedzy zajmującą się analizą obrazów nazywa się radiomiką.

Radiomika w medycynie jest metodą polegającą na wyodrębnianiu dużej liczby cech z obrazów medycznych za pomocą algorytmów charakteryzujących dane. Cechy te, określane mianem cech radiomicznych, mogą potencjalnie ujawnić wzorce i cechy nowotworów, których nie można dostrzec gołym okiem. Hipotezą radiomiki jest to, że charakterystyczne cechy obrazowania pomiędzy postaciami choroby mogą być przydatne w przewidywaniu rokowania i odpowiedzi terapeutycznej dla różnych typów nowotworów, dostarczając w ten sposób cennych informacji dla spersonalizowanej terapii. Dla ekstrakcji zmiennych radiomicznych kluczowe jest określenie obszaru zainteresowania (VOI) lub, popularnie, wysegmentowanej (w żargonie radioterapeutycznym „zakonturowanej”) struktury. W dziedzinie radioterapii onkologicznej lekarz onkolog-radioterapeuta dokonuje manualnie „konturingu” obszarów

tarczowych i narządów krytycznych. Cechy radiomiczne można podzielić na kilka grup: cechy oparte na wielkości i kształcie, deskrytory histogramu intensywności obrazu, deskrytory relacji między wokselami obrazu (GLCM), cechy oparte o macierz długości przebiegu (RLM), cechy oparte o macierz stref wielkości (SZM) i macierz sąsiedztwa różnic tonów szarości (NGTDM), tekstury pochodne, tekstury wyodrębnione z przefiltrowanych obrazów oraz cechy fraktalne. Przy planowaniu i tworzeniu nowych narzędzi opartych o cechy radiomiczne, ze względu na ich ogromną różnorodność, konieczne jest wprowadzenie selekcji zmiennych w celu wyeliminowania zbędnych informacji. Ponadto należy wyeliminować cechy niestabilne i niepoddające się reprodukcji. Modelowanie predykcyjne, czyli tworzenie modeli matematycznych, które mają przewidzieć odpowiedź na leczenie czy rozwój powikłań, stanowi podstawę medycyny spersonalizowanej. Tworząc test predykcyjny, formułuje się model klasyfikacyjny, który ma za zadanie różnicować populacje np. osoby, które odpowiedziały na leczenie (klasa docelowa) i korzysta z poprzednio zebranej wiedzy, tzn. zbioru treningowego (wybranych zmiennych i przypadków). Jednym z kluczowych problemów tworzenia tego typu narzędzi jest zjawisko przeuczenia, czyli sytuacji, w której model wydaje się zbyt dobry, aby móc odzwierciedlać faktyczne zależności w środowisku. Ponieważ błąd jest losowy, takie modele nie mogą być generalizowane, tzn. nie sprawdzą się w walidacji na nowym zbiorze danych. W przypadku radiomiki problem ten jest dodatkowo zaostrzony ogromną ilością generowanych zmiennych. Proces modelowania predykcyjnego należy, zatem, kontrolować. Najprostszym sposobem takiej kontroli jest walidacją prostą, w której model sprawdza się na niezależnym zbiorze danych. Istnieje wiele metod stworzenia modeli klasyfikacyjnych, zwłaszcza w radiomice. Jedną z najprostszych jest tworzenie drzew decyzyjnych, takie modele oprócz funkcji przewidywania pokazują prosty algorytm, dlatego model wykazuje odpowiednią predykcję. Takie drzewo ma prostą i zrozumiałą klinicznie prezentację graficzną polegającą na sklasyfikowaniu członków populacji poprzez podzielenie jej na subpopulacje na podstawie kilku dychotomicznych zmiennych niezależnych.

Hipoteza badawcza: Celem pracy była weryfikacja czynników predykcyjnych odpowiedzi całkowitej na zastosowanie neoadjuwantowej radiochemioterapii u chorych na miejscowo zaawansowanego raka odbytnicy oraz określenie przydatności klinicznej ekstrakcji markerów radiomicznych obszaru zmiany nowotworowej. Celem drugorzędowym było utworzenie i walidacja modelu przewidującego odpowiedź całkowitą w oparciu o metody drzew decyzyjnych i model regresji logistycznej.

Materiały i metody: Badanie wykonano w schemacie retrospektywno-prospektywnego badania obserwacyjnego. Pacjentów kwalifikowanych do przedoperacyjnej radiochemioterapii z miejscowo zaawansowanym rakiem odbytnicy leczono zgodnie z aktualnymi wytycznymi i obserwowano w czasie. Na podstawie zabranych przed leczeniem danych klinicznych, laboratoryjnych, obrazowych i radiomicznych podjęto próbę oceny przydatności, a potem także próbę stworzenia modelu predykcyjnego klinicznej odpowiedzi całkowitej na leczenie lub całkowitej odpowiedzi stwierdzanej histopatologicznie po zabiegu operacyjnym. U pacjentów zakwalifikowanych do leczenia neoadjuwantowego oznaczono objętości referencyjne: GTV; CTV; PTV; narządy krytyczne oraz określono dla nich limity dawek. GTV wyznaczany był na podstawie badania RM (w sekwencji T2 TSE) i (jeśli dostępne) PET-TK. Do szacowania liczebności próby wykorzystano ocenę opartą o prevalencje, wyliczono, że minimalna grupa potrzeba do zebrania w badaniu wynosi 64 pacjentów. W analizie opisowej użyto standardowych miar tendencji centralnej, w tym wartości średnich arytmetycznych, odchylenia standardowego, mediany, kwartyli i rozstępów międzykwartylowych. W analizie zmiennych nominalnych korzystano z testu Z dla proporcji oraz testu Chi2. Analizę zmiennych ciągłych przeprowadzono z wykorzystaniem narzędzi statystyki parametrycznej (test t dla grup zależnych i niezależnych). Ekstrakcji zmiennych radiomicznych dokonano z wykorzystaniem oprogramowania 3D Slicer. Esktrahowano wszystkie dostępne zmienne radiomiczne we wszystkich klasach z obszaru GTV. Dodatkowo, użyto opcji przepróbkowania wokseli (ang. *voxel resampling*). Ponadto, używano różnych wartości sigma dla laplacian filtru gaussowskiego (LoG), który uwypukla krawędzie. Ostatecznie sprawdzano także zmienne utworzone z wykorzystaniem filtrowania falkowego (ang. *wavelet*). Aby dokonać selekcji zmiennych i stworzenia najlepszego modelu predykcyjnego skorzystano z algorytmu OmicSelector. Aby utrzymać prostotę modeli przy małej grupie przypadków uczących wybrano ocenę tylko z wykorzystaniem regresji logistycznej, metod tworzenia drzew decyzyjnych oraz metody losowego lasu. Kluczowym było także przeciwdziałanie potencjalnemu efektowi serii. Zebraną grupę pacjentów podzielono tak, aby wszyscy pacjenci, których badanie rezonansu magnetycznego wykonane było poza naszym ośrodkiem (na innym sprzęcie) zostali przydzieleni do zbioru testowego. Pozostałych pacjentów przydzielono do zbioru treningowego. Zbiór treningowy balansowano z wykorzystaniem techniki SMOTE. Wydajność klasyfikacyjna modeli przedstawiono i analizowano na podstawie macierzy pomyłek, a także wartości dokładności, czułości i swoistości. Wybrany model drzewa decyzyjnego, ponadto, przedstawiono i omówiono.

Wyniki: Grupę badaną stanowiło 82 pacjentów, przeważali mężczyźni, a mediana wieku wynosiła 65.22 lat. U większości obserwowano zajęcie węzłów chłonnych, przeważał III stopień zaawansowania, u prawie połowy obserwowano EMVI. Większość chorych otrzymała dawkę 50.0 G z wykorzystaniem

techniki SIB. U 8% pacjentów zdecydowano się na strategię *watch-and-wait*, u 14% stwierdzono śródoperacyjnie nieoperacyjność nowotworu. Całkowitą regresję obserwowano u 15 pacjentów (n=15/82; 18,3%). Porównanie cech klinicznych u pacjentów, którzy uzyskali całkowitą odpowiedź na leczenie nie wykazało, aby pacjenci w tej podgrupie różnili się istotnie statystycznie. Leczenie prowadziło do istotnego statystycznie zmniejszania długości nacieku widocznego w badaniu rezonansu magnetycznego. Istotną statystycznie regresję zajętych węzłów nowotworowych obserwowano w obszarach M i PS, ale nie udało się udowodnić wpływu leczenia na zajęcie węzłów chłonnych w obszarach LLN i ON. Zastosowanie radiochemioterapii było istotnie statystycznie związane z “negatywizacją” statusu zajętych węzłów chłonnych. U 18 pacjentów (22%) nie obserwowano korzyści z zastosowanego leczenia. Żaden z czynników klinicznych nie był istotnym statystycznie samodzielnym czynnikiem predykcyjnym odpowiedzi całkowitej na leczenie. Wykorzystując wszystkie zmienne kliniczne i radiomiczne przypadki ze stwierdzoną odpowiedzią całkowitą nie ulegały samoistnej klasteryzacji. Podobnie, wyekstrahowane zmienne zdawały się nie zależeć od wyjściowego stopnia zaawansowania choroby. W całej grupie 12-miesięczne przeżycie wyniosło 86.6%, 2-letnie 76.3%, a 3-letnie 64.4%. W toku całej obserwacji nie udało udowodnić się, że pacjenci, u których wystąpiła odpowiedź całkowita cechują się istotnym statystycznie lepszym przeżyciem, ale porównując przeżycie 12-miesięczne i 2-letnie różnica była widoczna na korzyść pacjentów z całkowitą odpowiedzią. W analizie jednoczynnikowej wykazano, że samodzielnymi czynnikami, które skracały przeżycie był *grading* histopatologiczny, stężenie CA19-9 przed radiochemioterapią, obecność EMVI przed leczeniem, *grading* histopatologiczny G3 w porównaniu do G1 oraz wyższe stężenie CEA po radiochemioterapii. Uzyskanie jakiegokolwiek odpowiedzi na leczenie istotnie statystycznie wydłużało przeżycie. Podobnie, pooperacyjnie stwierdzana infiltracja węzłów chłonnych działała niekorzystnie rokowniczo. Co ciekawe, nie udowodniono prognostycznego charakteru skali TRG czy cechy pT. Selekcję zmiennych przeprowadzono z wykorzystaniem wszystkich metod dostępnych w oprogramowaniu OmicSelector. Sygnatury kliniczno-radiomiczne limitowano do 10 cech celem uniknięcia przeuczenia. Wybrano 33 zestawy cech, z których prawie wszystkie opierały się wyłącznie o zmienne radiomiczne. Wykorzystując zbiór treningowy indukowano wiele modeli klasyfikacyjnych i oceniano je pod kątem wydajności predykcyjnej na zbiorze treningowym i testowym. Najlepszymi właściwościami diagnostycznymi wykazało się drzewo decyzyjne utworzone z wykorzystaniem algorytmu „*ctree*” na sygnaturze wybranej metodą „AUC_MDL” na zbiorze treningowym zbalansowanym metodą SMOTE. Drzewo to opierało się o następujące cechy radiomiczne: *wavelet* HHL glcm IDN; log.sigma.2.0.mm.3D glcm *Difference Variance*; *wavelet* LHH ngtdm *Contrast*; log.sigma.1.0.mm.3D glcm IDMN; log.sigma.2.0.mm.3D glcm *Cluster Prominence*. Wybrany model osiągnął wysokie pole pod krzywą ROC

równe 0,95 (95%CI: 0,91-0,98), osiągnął dokładność 86% (95%CI: 78%-92%), czułość 82% i swoistość 91%. Na niezależnym zbiorze testowym, model nie utracił zdolności predykcyjnych wykazując dokładność 94%, czułość 67% i swoistość 100%.

Wnioski:

1. Klasyczne czynniki kliniczne może cechować niedostateczna wartość predykcyjna odpowiedzi całkowitej po skojarzonym leczeniu chorych na miejscowo zaawansowanego raka odbytnicy.
2. W analizie sygnatur kliniczno-radiomicznych, wśród wybranych 33 zestawów cech, prawie wszystkie opierały się wyłącznie o zmienne radiomiczne, co potwierdza wyższość predykcyjną cech radiomicznych nad klinicznymi.
3. Stworzony model drzewa decyzyjnego, oparty na 5 cechach radiomicznych wykazuje potencjalną przydatność kliniczną, osiągając wysokie pole pod krzywą ROC, nie tracąc swoich wartości predykcyjnych na zbiorze testowym.
4. Prostota modelu pozwala na jego szerokie zastosowanie w praktyce klinicznej.

9 Summary in English

Introduction: Colorectal cancer is the third most common group of malignancies worldwide. Its incidence index in Poland is steadily increasing and represents the sixth cause of malignant neoplasms among men and the ninth among women. Recently, treatment strategies for rectal cancer have been undergoing significant modifications. Both radiation therapy, systemic treatment, and surgical procedures are showing increasing efficacy. In initially resectable cases, attempts are being made to limit the indications for radiotherapy. Patients who have achieved a complete response to neoadjuvant treatment are increasingly considering postponing the surgical treatment. With so many therapeutic options, it requires appropriate patient selection and decisions made by multidisciplinary oncological boards. Radiation therapy or radiochemotherapy is a therapeutic option for rectal adenocarcinoma regardless of stage. The need for careful evaluation of factors that increase the risk of recurrence/metastasis and the potential resectability of the tumor based on RM examination is emphasized. The increasing efficacy of neoadjuvant treatment, through additional intensification called total neoadjuvant treatment - TNT, the intention to improve quality of life through sphincter preservation and the lack of need for surgery became the basis for the implementation of the "watch-and-wait"/"watch-and-see" concept. The variety of treatment options implies the need for detailed diagnostics and precise staging. The test of choice in assessing local tumor progression is pelvic RM. In addition to basic information on the staging, the RM examination can be subjected to computer analysis to obtain more data; the field of knowledge that deals with image analysis is called radiomics.

Radiomics in medicine is a method of extracting many features from medical images using algorithms that characterize the data. These features, referred to as radiomic features, have the potential to reveal patterns and characteristics of tumors that cannot be seen with the naked eye. The hypothesis of radiomics is that characteristic imaging features between disease forms can be helpful in predicting prognosis and therapeutic response for different types of cancer, thus providing valuable information for personalized therapy. For the extraction of radiomic variables, it is crucial to define the region of interest (VOI) or, more commonly, the segmented (in radiotherapy jargon, "contracted") structure. The radiation oncologist manually "contours" the target areas and critical organs in radiation oncology. Radiomic features can be divided into several groups: size- and shape-based features, image intensity histogram descriptors, image voxel relationship descriptors (GLCM), run length matrix (RLM) based features, size zone matrix (SZM) and gray tone difference neighborhood matrix (NGTDM) based features, derived textures, textures

extracted from filtered images, and fractal features. When planning and creating new tools based on radiomic features, due to their vast variety, it is necessary to introduce variable selection to eliminate redundant information. In addition, unstable and non-reproducible features must be eliminated. Predictive modeling, or the creation of mathematical models to predict the response to treatment or the development of complications, is the basis of personalized medicine. When creating a predictive test, a classification model is formulated to differentiate populations, e.g., those who have responded to treatment (the target class), and uses previously collected knowledge, i.e., the training set (selected variables and cases). One of the critical problems of creating such tools is the phenomenon of overfitting, i.e., a situation in which the model seems too good to reflect the actual relationships in the environment. Since the error is random, such models cannot be generalized, i.e. they will fail in validation on a new dataset. In the case of radiomics, this problem is further exacerbated by the significant number of variables generated. The predictive modeling process must, therefore, be controlled. The simplest way of such control is simple validation, in which the model is checked on an independent dataset. There are many methods for creating classification models, especially in radiomics. One of the simplest is the creation of decision trees; such models, in addition to the prediction function, show a simple algorithm for why the model shows an appropriate prediction. Such a tree has a simple and clinically understandable graphical presentation of classifying members of a population by dividing it into subpopulations based on several dichotomous independent variables.

Hypothesis: The purpose of this study was to validate the predictive factors of complete response to neoadjuvant radiochemotherapy in patients with locally advanced rectal cancer and to determine the clinical utility of extracting radiomic markers of the tumor lesion area. The secondary objective was to create and validate a model predicting complete response based on decision tree methods and logistic regression models.

Materials and Methods: This was a retrospective-prospective observational study. Patients qualified for preoperative radiochemotherapy with locally advanced rectal cancer were treated according to current guidelines and observed over time. Based on the clinical, laboratory, imaging, and radiomic data taken before treatment, an attempt was made to assess their usefulness and create a predictive model of the complete clinical response to treatment or complete response found histopathologically after surgery. In patients qualified for neoadjuvant treatment, reference volumes were determined: GTV; CTV; PTV; critical organs, and dose limits were determined for them. GTV was determined by RM (in T2 TSE sequence) and (if available) PET-CT. A prevalence-based assessment was used to estimate sample size, and it was calculated that the minimum group needed to be collected in the study was 64 patients. Descriptive analysis

used standard measures of central tendency, including arithmetic means, standard deviation, median, quartiles, and interquartile ranges. The analysis of nominal variables used the Z test for proportions and the Chi2 test. Analysis of continuous variables was carried out using the tools of parametric statistics (t-test for dependent and independent groups). Radiomic variables were extracted using 3D Slicer software. All available radiomic variables in all classes from the GTV area were extracted. In addition, a voxel resampling option was used. In addition, different sigma values were used for the laplacian Gaussian filter (LoG), which emphasizes edges. Finally, variables created using wavelet filtering were also checked. To select variables and create the best predictive model, the OmicSelector algorithm was used. In order to keep the models simple with a small group of learning cases, an evaluation using only logistic regression, decision tree creation methods, and a random forest method was chosen. It was also crucial to counteract the potential series effect. The collected group of patients was divided so that all patients whose MRI examination was performed outside our center (on other equipment) were assigned to the test set. The remaining patients were assigned to the training set. The training set was balanced using the SMOTE technique. The classification performance of the models was presented and analyzed based on the confusion matrix and accuracy, sensitivity, and specificity values. The selected decision tree model, moreover, was presented and discussed.

Results: The study group comprised 82 patients, predominantly men, and the median age was 65.22 years. Lymph node involvement was observed in the majority, stage III was predominant, and EMVI was observed in almost half. Most patients received a dose of 50.0 G using the SIB technique. In 8% of patients, a watch-and-wait strategy was opted for; in 14%, the tumor was found to be inoperable intraoperatively. Complete regression was observed in 15 patients (n=15/82; 18.3%). A comparison of clinical features in patients who received a complete treatment response did not show that patients in this subgroup were statistically significantly different. Treatment led to a statistically significant reduction in the length of the infiltrate seen on MRI. Statistically significant regression of involved tumor nodes was observed in M and PS areas. However, the effect of treatment on lymph node involvement in LLN and ON areas could not be proven. The use of radiochemotherapy was statistically significantly associated with "negativity" of the status of the involved lymph nodes. In 18 patients (22%), no benefit was observed from the applied treatment. None of the clinical factors was a statistically significant independent predictor of complete response to treatment. Using all clinical and radiomic variables, the cases with an identified complete response did not self-cluster. Similarly, the extracted variables seemed not to depend on the baseline disease stage. In the entire group, 12-month survival was 86.6%, 2-year survival was 76.3%, and 3-year survival was 64.4%. During the entire follow-up, it was not possible to prove that a statistically significant better

survival characterized patients who had a complete response. However, when comparing 12-month and 2-year survival, the difference was apparent in favor of patients with a complete response. Univariate analysis showed that independent factors that shortened survival were histopathological grading, CA19-9 levels before radiochemotherapy, the presence of EMVI before treatment, G3 versus G1, and higher CEA levels after radiochemotherapy. Achieving any response to treatment statistically significantly prolonged survival. Similarly, postoperatively found lymph node infiltration acted as an unfavorable prognostic factor. Interestingly, the prognostic nature of the TRG scale or the pT feature was not proven.

Variable selection was performed using all methods available in OmicSelector software. Clinical-radiomic signatures were limited to 10 features to avoid overfitting. Thirty-three feature sets were selected, almost all based solely on radiomic variables. Multiple classification models were induced and evaluated using the training set for predictive performance on the training and test sets. The best diagnostic performance was demonstrated by the decision tree created using the "ctree" algorithm on the signature selected by the "AUC_MDL" method on the training set balanced by the SMOTE method. The tree was based on the following radiomic features: wavelet HHL glcm IDN; log.sigma 2.0.mm.3D glcm Difference Variance; wavelet LHH ngtdm Contrast; log.sigma.1.0.mm.3D glcm IDMN; log.sigma.2.0.mm.3D glcm Cluster Prominence. The selected model achieved a high area under the ROC curve of 0.95 (95%CI: 0.91-0.98), achieved an accuracy of 86% (95%CI: 78%-92%), a sensitivity of 82%, and a specificity of 91%. On the independent test set, the model did not lose the predictive ability, showing an accuracy of 94%, a sensitivity of 67%, and a specificity of 100%.

Conclusions:

- classic clinical factors can be insufficient predictor of complete response to neoadjuvant treatment of locally advanced rectal cancer.
- In the analysis of clinical-radiomic signatures, among the selected 33 sets of features, almost all of them were based exclusively on radiomic variables, confirming the predictive superiority of radiomic features over clinical ones.
- The developed decision tree model based on 5 radiomic features shows potential clinical usefulness, achieving a high area under the ROC curve of 0.95 (95%CI: 0.91-0.98), accuracy of 86% (95%CI: 78%-92%), sensitivity of 82% and specificity of 91%, without losing its predictive values on the independent test set.
- The simplicity of the model could allow it to be useful in routine practice.