

Łukasz Fuławka

Wykorzystanie metod komputerowej analizy obrazu do oceny indeksu proliferacyjnego w raku piersi

ROZPRAWA DOKTORSKA

Streszczenie

Rak piersi jako najczęstszy nowotwór złośliwy u kobiet stanowi ważne wyzwanie dla opieki zdrowotnej. Istnieje kilka klasyfikacji raka piersi, spośród których największe znaczenie mają aktualnie klasyfikacja histologiczna wg WHO oraz klasyfikacja molekularna. Druga z wymienionych klasyfikacji powstała w oparciu o wyniki badań na mikromacierzach ekspresyjnych. Z uwagi na ograniczoną dostępność tej metody, opracowano klasyfikację zastępczą, opartą na immunohistochemii. Tzw. kliniko-patologiczne podtypy zastępcze (clinicopathological surrogate subtypes) obejmują następujące kategorie: luminalny A (luminal A-like), luminalny B - HER2 ujemny (luminal B-like - HER2 negative), luminalny B - HER2 dodatni (luminal B-like - HER2 positive), HER2 dodatni - Nieluminaly (HER2 positive - non-luminal) oraz trójujemny (triple-negative). Aby zaklasyfikować do określonego podtypu, wykonywane jest badanie immunohistochemiczne pod kątem statusu receptorów estrogenowych (ER), progesteronowych (PgR), HER2 oraz białka jądrowego Ki-67. Ocena Ki-67 polega na wyliczeniu indeksu proliferacyjnego, który w praktyce jest szacowany w trakcie oceny mikroskopowej, a uzyskiwane wyniki cechują się znaczącymi różnicami pomiędzy oceniającymi patologami (interobserver variability). Jest to o tyle istotne, że może mieć wpływ na niewłaściwe klasyfikowanie do podtypów biologicznych, a w dalszej konsekwencji na błędne decyzje terapeutyczne.

W celu potwierdzenia tego zjawiska w lokalnym środowisku, poproszono czworo praktykujących patologów o ocenę PI w grupie 98 przypadków raka piersi. Analiza wyników wykazała współczynnik korelacji Pearsona na poziomie 0.77 - 0.84 ($P < 0.001$) oraz współczynnik korelacji rang Spearmana 0.68 - 0.83 ($P < 0.001$). Współczynnik kappa Fleissa dla wartości granicznej 14% wyniósł 0.58, dla 20% - 0.60. Natomiast współczynnik kappa Cohena 0.45 - 0.69 dla 14% oraz 0.53 - 0.67 przy 20% punktu odcięcia. Test rang Friedmana wykazał istotne statystycznie rozbieżności pomiędzy oceniającymi patologami ($P < 0.001$). Ponadto zaobserwowano znaczące różnice pomiędzy metodami ewaluacji PI stosowanymi przez patologów.

W tym samym czasie autor analizował możliwości zastosowania darmowej aplikacji internetowej ImmunoRatio w rutynowej diagnostyce histopatologicznej. Zgromadzono bazę obrazów cyfrowych pROIs preparatów Ki-67 tej samej grupy przypadków raka piersi. Wyniki oceny PI obliczone za pomocą ImmunoRatio porównano z wartościami referencyjnymi. Za złoty standard służył program ImageJ z wtyczką Cell Counter do manualnego zliczania komórek. Wykazano współczynnik korelacji Pearsona na poziomie 0.84 ($p < 0.05$), współczynnik korelacji rang Spearmana 0.83 ($p < 0.05$). Badanie ujawniło umiarkowaną zgodność pomiędzy ImmunoRatio a metodą referencyjną, przy współczynniku kappa Cohena wynoszącym 0.47 ($p < 0.05$). Warto zauważyć, że w zależności od zastosowanej metody, 26 pacjentek zostałyby zakwalifikowanych do różnych podtypów biologicznych (zgodnie z obowiązującymi wówczas wytycznymi St. Gallen 2013). Tym samym wykazano, iż ImmunoRatio wymaga dalszego dopracowania, aby mogło zostać wykorzystane w rutynowej diagnostyce.

Na dalszym etapie prac, autor zgromadził nową kolekcję 95 wirtualnych preparatów mikroskopowych Ki-67 na bazie retrospektywnej grupy przypadków raka przewodowego in situ (DCIS). Dla każdego z nich wybrano pROI, z których utworzono referencyjną bazę obrazów i przyporządkowanych im wartości PI, wyliczonych za pomocą opisanej powyżej metodologii (ImageJ). Celem tego etapu badań było opracowanie algorytmu do precyzyjnego określania PI w obrazach mikroskopowych. Baza obrazów została losowo podzielona na 4 grupy: 3 zbiory uczące po 25 obrazów oraz jeden zbiór testowy zawierający 20 obrazów. Proponowana metoda wykorzystuje głębokie uczenie oraz teorię zbiorów rozmytych (fuzzy-set theory). Uzyskane ROI jest wykorzystywane do segmentacji właściwie rozpoznanych jąder komórek nowotworowych za pomocą klasycznych metod przetwarzania obrazów. Wartość PI jest szacowana na podstawie obliczenia stosunku powierzchni zajmowanej przez immunopoztywne jądra do całkowitej powierzchni jąder rozpoznanych komórek nowotworowych. Uzyskane wyniki porównano z wartościami referencyjnymi. W celu zwiększenia skuteczności algorytmu, modele były uczone w trzech turach, podczas których porównywano wpływ różnych hiperparametrów. Wykazano, że zastosowanie segmentacji komórek poprzez ROI oparte na sieciach neuronowych znacząco poprawia końcowe wyniki ewaluacji PI (zmniejszenie MAE z poziomu 0.17 do 0.024–0.11). Ponadto rozmiar okna ma istotny wpływ na poprawność oceny PI (MAE 0.02 vs. 0.11 dla modelu zespolonego). Dowiedziono, iż implementacja metody Otsu powoduje w większości przypadków zmniejszenie liczby niedopasowanych segmentacji tła, a zatem znacząco obniża błąd oceny PI z poziomu 0.058 na 0.034 (dla pojedynczego modelu z oknem 96 px). Podsumowując, opracowany algorytm przewyższa skutecznością inne najbardziej aktualne rozwiązania.

Summary

Breast cancer, being the most common malignancy in females, is a great challenge in public health. There are a few classifications of breast cancer, with the most commonly used WHO histological classification and molecular subtypes. The latter classification was initially based on gene expression arrays. Due to a poor access to its technique, the surrogate definitions, based on routine immunohistochemistry (IHC) were developed. The clinicopathological surrogate subtypes are: luminal A-like, luminal B-like (HER2 negative), luminal B-like (HER2 positive), HER2-positive (non-luminal) and triple-negative. To classify breast cancer cases into these types, the IHC assay of estrogen receptor (ER), progesterone receptor (PgR), HER2, and Ki-67 protein needs to be performed. The Ki-67 nuclear protein status is defined by proliferation index (PI). The method commonly applied to determine PI is visual scoring of the sample (VA, visual assessment) using the method called eyeballing (EB). However, this approach is prone to significant interobserver and intraobserver variability.

To investigate the phenomenon of interobserver variability in the local environment, the author asked four local non-scholar pathologists to evaluate PI in a group of 98 cases of breast carcinoma. Statistical analysis showed pairwise Pearson's coefficients ranging between 0.77 and 0.84 ($P < 0.001$) and Spearman's rank correlation coefficients ranging between 0.68 and 0.83 ($P < 0.001$). The Fleiss kappa value for the 14% cutoff point was 0.58 whereas for the 20% cutoff point was 0.60. The pairwise Cohen's kappa values ranged from 0.45 to 0.69 for the 14% cutoff point and 0.53 to 0.67 for the 20% cutoff point. Friedman's rank ANOVA test showed significant differences among the four pathologists ($P < 0.001$). Summarizing, this stage of research revealed substantial interobserver variability and difference in methods of PI evaluation between pathologists.

Meanwhile, the author evaluated applicability of ImmunoRatio web application in routine histopathological diagnostics. The digital images of pROIs were taken from above-mentioned group of breast carcinoma Ki-67 slides. The PI assessment results generated by ImmunoRatio were compared to reference values. A set of reference values was established using ImageJ, the open free image analysis software, with Cell Counter plug-in, which enables manual counting of single cells. The study revealed Pearson's correlation coefficient 0.84 ($p < 0.05$), Spearman's correlation coefficient 0.83 ($p < 0.05$). Moderate agreement was shown by Cohen's kappa calculation ($K = 0.47$; $p < 0.05$). It is noteworthy that as many as 26 cases were classified to different biological subtypes (according to St. Gallen 2013 recommendations in force at the time) depending on the method of

PI assessment. This research revealed that ImmunoRatio needed further improvement to be a reliable diagnostic tool.

In a further stage of the study, the new reference database consisting of a collection of 95 whole slide images (WSI) of ductal carcinoma in situ (DCIS) was created. For each case one pROI was selected and saved as image file. The reference method of PI calculation was the same as described above (ImageJ). The aim of this stage of research was to build an accurate algorithm for PI quantification with deep learning model and dedicated image processing. The complete data set was randomly divided into 4 chunks: 3 training sets of 25 images each and one testing set containing 20 images. These training sets are used in a k-fold training manner of 3 CNN models in the next stage. Each CNN model (based on the same architecture) is trained on 2 chunks (50 images), which are randomly divided into training and validating subsets in ratio 4:1. The developed solution calculates the Ki-67 proliferation index by utilizing a deep learning model and fuzzy-set interpretations for hot-spots detection. The obtained ROI (region of interest) is used to segment relevant nuclei via classical methods of image processing. The PI value is approximated by relating the surface area occupied by immunopositive nuclei to the total surface area of relevant nuclei. The achieved results were compared to reference values. To increase reliability of the algorithm, several models were trained in a threefold manner and the impact of different hyper-parameters was compared.

It was revealed that applying cell segmentation over a neural-network-based ROI significantly improves the final Ki-67 proliferation index estimation (from MAE of around 0.17 to 0.024–0.11). Moreover, window size has an enormous impact on PI estimation correctness (from 0.11 MAE down to 0.02 for ensemble models). It has been proven that the methods utilizing Otsu's threshold biasing are characterized by a lower number of mismatched cells-background segmentations in most cases and thereby much lower mean proliferation index error (for the single model with 96 px window it lowers MAE from 0.058 to 0.034 and reduces the number of invalid images from 3 to 0). Concluding, the developed method estimates PI with MAE as low as 0.024, which gives a significant advantage over the current state-of-the-art solution.