An interpretable automated detection system for FISH-based HER2 oncogene amplification testing in histo-pathological routine images of breast and gastric cancer diagnostics

Abstract

Histo-pathological diagnostics are an inherent part of the everyday work but are particularly laborious and associated with time-consuming manual analysis of image data. In order to cope with the increasing diagnostic case numbers due to the current growth and demographic change of the global population and the progress in personalized medicine, pathologists ask for assistance. Profiting from digital pathology and the use of artificial intelligence, individual solutions can be offered (e.g. detect labeled cancer tissue sections). The testing of the human epidermal growth factor receptor 2 (HER2) oncogene amplification status via fluorescence in situ hybridization (FISH) is recommended for breast and gastric cancer diagnostics and is regularly performed at clinics. Here, we developed an interpretable, deep learning (DL)-based pipeline which automates the evaluation of FISH images with respect to HER2 gene amplification testing. It mimics the pathological assessment and relies on the detection and localization of interphase nuclei based on instance segmentation networks. Furthermore, it localizes and classifies fluorescence signals within each nucleus with the help of image classification and object detection convolutional neural networks (CNNs). Finally, the pipeline classifies the whole image regarding its HER2 amplification status. The visualization of pixels on which the networks’ decision occurs, complements an essential part to enable interpretability by pathologists.

Keywords: FISH imaging, HER2 amplification, gastric/breast cancer, digital pathology, deep learning, image classification, object segmentation and localization, interpretability.

1. Introduction

The HER2 amplification status is an important tumor marker in breast and gastric cancer. It indicates a more aggressive disease with a greater rate of recurrence and mortality. Hence, it influences the decision making for finding an appropriate therapy (Mitri et al., 2012). The amplification of HER2 is detected by assessing FISH images: Fluorescence signals are counted in at least 20 interphase nuclei from tumor regions and are then graded into a HER2 negative or positive status with a high or low amplification (Wolff et al., 2018). To optimize the diagnostics in terms of speed, accuracy, objectivity and interpretability, we developed a comprehensible, multi-step DL-based pipeline (Figure 1A). It mimics the pathologist’s evaluation steps and integrates the decision processes into a report for transparency (Figure 1A.6). The pipeline independently evaluates each nucleus twice by different networks creating a second opinion. The pre-selection of nuclei reduces the risks of isolating overlapping nuclei parts (e.g. signals) or artifacts which could alter the nucleus-specific
classification. The first component is an instance segmentation network designed for cell containing datasets, called StarDist (Schmidt et al., 2018), to detect and extract all individual nuclei (not only 20) from the entire FISH slide (Figure 1A.1). The retrieved nuclei are then classified by a custom image classification CNN and a RetinaNet-based FISH signal detector system (Lin et al., 2017) (Figure 1A.2 and 4). Visualizations, such as class activation maps (CAMs) (Zhou et al., 2016), display the decision making of the pipeline.

Figure 1: (A) Workflow for the automated evaluation of the HER2 gene amplification testing from FISH images. (B) Utilized architectures of the pipeline components. Adapted from Lin et al. (2017); Schmidt et al. (2018). (C) Graphical user interface for interactive access to the pipeline results for pathologists.

2. Results

The pipeline components (Figure 1B) were trained on augmented datasets composed of routinely processed FISH (1,600 × 1,200 px, jpeg format) and single nucleus images. These images often display diverse artifacts, low signal-to-noise ratios and color disparities as a consequence of sample fixation and imaging. Network performances were estimated on validation and test sets using precision, recall and average precision (AP) scores.
The first component, called Nucleus Detector (Figure 1A.1), is based on StarDist (Figure 1B, first panel) and was trained on 62 FISH slides (~7,440 nuclei) to detect individual nuclei. The shape prediction of nuclei as star-convex polygons excludes classification-altering artifacts and overlapping nuclei parts. It achieved a precision score of 0.76 and recall score of 0.65 on the test set (10 images, 810 nuclei) since the differentiation of small and adjacent nuclei (often predicted as single nucleus) remained challenging.

A set of 8,313 single nucleus images was used to train a custom image classification CNN (Figure 1B, second panel; VGG-like, weighted cross-entropy loss), the Nucleus Classifier (Figure 1A.2), for the classification of nuclei into artifact, background, HER2 normal expression (= HER2 negative) and HER2 low or high amplification (= HER2 positive). Thereby, the filter classes (artifact, background) ensure that images with none or more than one nucleus will not be considered for the HER2 amplification testing. On the validation set (1,668 images), the network achieved a precision and recall score of 0.98. CAMs (Figure 1A.2) were used to elucidate the classifications and demonstrated that the classes were recognized based on FISH signal presence and number.

The third component is the Signal Detector (Figure 1A.4) with a RetinaNet architecture (Figure 1B, third panel) was trained on 397 single nucleus images (~5,955 signals). It localizes and classifies FISH signals within a single nucleus into HER2 (single signal), HER2-cluster (non-differentiable amount of signals) and chromosome enumeration probe 17 (CEP17; reference centromeric satellite DNA, single signal). Thus, each nucleus is classified a second time (second opinion). The bounding boxes provide details regarding the number and position of FISH signals per nucleus. The Signal Detector achieved a mean AP of 0.73 on the validation set (39 images, 316 signals). The CEP17 signals (AP: 0.94) were detected very well but the detection and distinction of HER2 (AP: 0.65) and HER2-cluster signals (AP: 0.60) remained challenging. However, mainly crowded HER2 and weak FISH signals were not identified or detected in multiple and distinctly classified boxes.

The nucleus- and FISH image-wide HER2 amplification status is inferred by different ratios (Figure 1A.3 and 5) and thresholds as mentioned in Zakrzewski et al. (2019). The classification thresholds can be modified to the needs of any clinical purposes.

All steps of the pipeline are documented in a report file, which can be used to evaluate the classifications made by the individual pipeline components (Figure 1A.6 and 7).

3. Conclusions

Our pipeline is one step towards the development of a deep learning-based assisting tool for pathologists to cope with the growing number of cancer cases during clinical routine. It can be individually (re-)trained on lab-specific images to reach optimal performance. While increasing the speed of the evaluation, the pipeline additionally enhances objectivity, provides the maximum amount of information for medical reports and can be applied to any FISH-based (e.g. BCR/ABL, BCL/IGH fusions; MYC, BCL6, ALK translocations) analyses. To integrate these novel analysis requirements, it will be necessary to train our pipeline on the individual FISH protocol-specific images. We are optimizing our system to be applicable on whole slide images up to a size of 100k × 100k px or more. The source code for our pipeline can be found on GitLab1.

References


