European Journal of Radiology Open 5 (2018) 165–170

Contents lists available at ScienceDirect

European Journal of Radiology Open

journal homepage: www.elsevier.com/locate/ejro

Distinction between phyllodes tumor and fibroadenoma in breast ultrasound using deep learning image analysis

Elina Stoffel*, Anton S. Becker, Moritz C. Wurnig, Magda Marcon, Soleen Ghafoor, Nicole Berger, Andreas Boss

Institute for Diagnostic and Interventional Radiology, University Hospital of Zurich, Switzerland

ARTICLE INFO

Keywords:
Breast imaging
Ultrasound
Phyllodes
Fibroadenoma
Deep learning
Computer assisted diagnosis

ABSTRACT

Purpose: To evaluate the accuracy of a deep learning software (DLS) in the discrimination between phyllodes tumors (PT) and fibroadenomas (FA).

Methods: In this IRB-approved, retrospective, single-center study, we collected all ultrasound images of histologically secured PT (n = 11, 36 images) and a random control group with FA (n = 15, 50 images). The images were analyzed with a DLS designed for industrial grade image analysis, with 33 images withheld from training for validation purposes. The lesions were also interpreted by four radiologists. Diagnostic performance was assessed by the area under the receiver operating characteristic curve (AUC). Sensitivity, specificity, negative and positive predictive values were calculated at the optimal cut-off (Youden Index).

Results: The DLS was able to differentiate between PT and FA with good diagnostic accuracy (AUC = 0.73) and high negative predictive value (NPV = 100%). Radiologists showed comparable accuracy (AUC 0.60–0.77) at lower NPV (64–80%). When performing the readout together with the DLS recommendation, the radiologist's accuracy showed a non-significant tendency to improve (AUC 0.75–0.87, p = 0.07).

Conclusion: Deep learning based image analysis may be able to exclude PT with a high negative predictive value. Integration into the clinical workflow may enable radiologists to more confidently exclude PT, thereby reducing the number of unnecessary biopsies.

1. Introduction

Phyllodes tumor (PT) of the breast are rare breast lesions, accounting for less than 1% of all breast tumors. They are typically seen in women aged 35 to 55 years at presentation and are mostly large with a median size of 4 cm [1]. Histologically, they are characterized by "leaf-like" lobulations, from which the name is derived (Greek *phullon* leaf), with more abundant and cellular stroma than that of fibroadenoma (FA). PT are commonly classified into categories of benign, borderline, or malignant on the basis of histological parameters such as mitotic count, cellular atypia, stromal cellularity and overgrowth, and the nature of tumor borders [2]. Histologically, benign PT can be mistaken for FA, whereas at the other end of the spectrum, malignant PT show overlapping features with primary breast sarcomas or spindle cell metaplastic carcinoma. However, regardless of their histology, all PT can recur, where an increased risk of local recurrence is correlated with larger size and malignancy [3–5].

FA is the most common benign tumor of the breast in women under 35 years of age. They present as well-defined, mobile masses that can increase in size and tenderness in response to high levels of estrogen (e.g. during pregnancy or prior to menstruation). Histologically, they are made up of both glandular breast tissue and stromal tissue. In contrast to PT, risk of cancer is usually not increased in FA [6].

In addition to their histopathological similarities, FA are usually indistinguishable from PT on a macroscopic level. Both fibroepithelial tumors are often detected as fast growing breast lumps, and distinguishing PT from FA by means of physical exam is extremely difficult. With increased public awareness and screening, most of the breast tumors are being discovered at earlier stages, when both tumors share a substantial overlap in sonographic features and size [7,8]. Furthermore, sonography cannot distinguish between malignant, borderline and benign PT. Diagnostic evaluation is therefore often extended to the use of invasive diagnostic procedures, such as core-needle biopsies. However, even with the help of histology, diagnosis can be complicated due to sampling errors.

The diagnosis has wider implications that also influence the therapeutic approach to these tumors. Although conservative management is an acceptable strategy in FA, malignant PT should be completely

* Corresponding author at: Institute for Diagnostic and Interventional Radiology, University Hospital Zurich, Raemistrasse 100, 8091 Zurich, Switzerland.
E-mail address: Elina.stoffel@uzh.ch (E. Stoffel).

https://doi.org/10.1016/j.ejro.2018.09.002
Received 13 March 2018; Received in revised form 7 September 2018; Accepted 7 September 2018

enucleated with clear margins due to the high recurrence rate of up to 30% [9] with metastases and death being observed in 22% [2]. Without re-excision, the recurrence rates can be as high as 43%, necessitating an additional operation [10].

Traditionally, patients with breast masses that cannot clearly be identified as FA or PT will usually undergo complete surgical excision or mastectomy, for the fear of overlooking a potentially malignant tumor. Therefore, accurate identification and differentiation of PT preoperatively is critical to appropriate surgical planning, avoiding operative complications resulting from inadequate excision or surgical overtreatment. Most FA do not need surgical treatment at all. In these cases, biopsies are essentially an unnecessary physical, psychological and financial burden for the patient [11].

Deep learning is a type of machine learning that was inspired by the structure and function of the brain. It imitates the mammalian visual cortex in processing data using artificial neural networks (ANNs) that contain hidden layers. The deep learning software (DLS) learns to extract meaningful features from images to then make inferences and decisions on its own. “Meaningful” in this context stands for “helping to solve the problem at hand”, in our case discriminating FA from PT. This data-driven method has shown promising results in recent years, as opposed to older more algorithmic approaches with hand-crafted features, which may often yield many arbitrary features not useful for the problem at hand. Hence, the use of deep learning in radiology as a method of differentiating and diagnosing tumors is a rapidly growing field [12]. Although, as with any diagnostic test, false-positive results can occur, the sensitivity of deep learning e.g. in mammography has reached numbers of up to 84%, equaling or surpassing the diagnostic accuracy of seasoned specialists [13]. Deep learning can be integrated into the assessment of sonographically detectable lesions and could be performed in the initial evaluation of indeterminate breast tumors (illustrated in Fig. 1).

In this retrospective, single-center study, we aimed to evaluate the precision of a DLS in the discrimination between PT and FA.

2. Materials and methods

2.1. Ultrasound examination and reference standard

This retrospective study was approved by the IRB, who waived the need for informed consent. All patients from a two-year period (July 2013 – July 2015) were reviewed for the presence of PT with histology as a reference standard (n = 11). From the remaining patients, a random subset with histologically secured diagnosis of a FA was taken (n = 15). Due to the low number (n = 4), FA with histopathological phyllodes features were counted towards one of the other groups. Since the management at our institution for those lesions is surgical excision, they were counted as PT. Median lesion diameter (long axis) was 21.5 mm (interquartile range 18–26 mm) for FA and 26.0 mm for PT (19–37 mm, p = 0.25). Lesion volume as calculated with all three diameters and the ellipsoid formula was also not significantly different (13.6 vs. 24.3 cm^3, p = 0.55). Mean age ± 95% confidence interval was 33.6 ± 15.2 years. All examinations were performed on the same type of ultrasound device (Logiq E9, GE Healthcare, Chicago, IL, USA) with the same reconstruction setting (“Breast”). For large lesions, multiple focus points were used. Functional ultrasound images were not consistently acquired and hence not included for analysis (i.e. with doppler or elastography overlay). For lesions depicted in multiple images, all available data was used, resulting in a total of 50 images of FA and 36 images of PT. The raw DICOM images were converted into lossless,

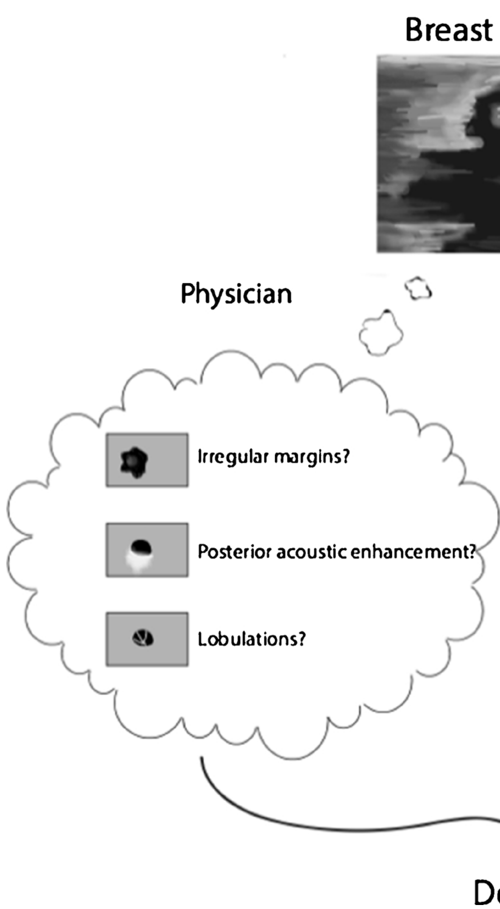

Fig. 1. Proposed integration of a deep learning based software into the clinical workflow. Deep learning image analysis has the ability to evaluate features, which are not perceptible to the human reader and may thus augment the evaluation of the radiologist.