A Novel Deep Learning System for Breast Lesion Risk Stratification in Ultrasound Images

Ting Liu, Xing An, Yanbo Liu, Yuxi Liu, Bin Lin, Runzhou Jiang, Wenlong Xu, Longfei Cong, and Lei Zhu
Shenzhen Mindray BioMedical Electronics, Co., Ltd., Shenzhen, China

MICCAI2022 Singapore

mindray

Abstract

This paper presents a novel deep learning system to classify breast lesions in ultrasound images into benign and malignant, and into Breast Imaging Reporting and Data System (BI-RADS) six categories simultaneously.

We propose three innovations including:

- **a multitask soft label generating architecture.**
- **a consistency supervision mechanism**, to ensure predictions of two tasks are consistent.
- **a cross-class loss function**, to make a prediction of BI-RADS closer to the annotation.

The proposed method outperformed other methods on two public datasets (BUSI and ADIAT).

Background

- **Breast cancer** has become the most common cancer in women.
- **Ultrasound** has been widely used in breast diagnosis.
- **BI-RADS categorization** (BI-RADS 2, 3, 4a, 4b, 4c, and 5) shows the malignant risk of a lesion [1].
- A higher **BI-RADS category** presents a greater malignant likelihood.
- A suspicious mass that is assessed as BI-RADS 4 or 5 will undergo a **preliminary biopsy** normally.

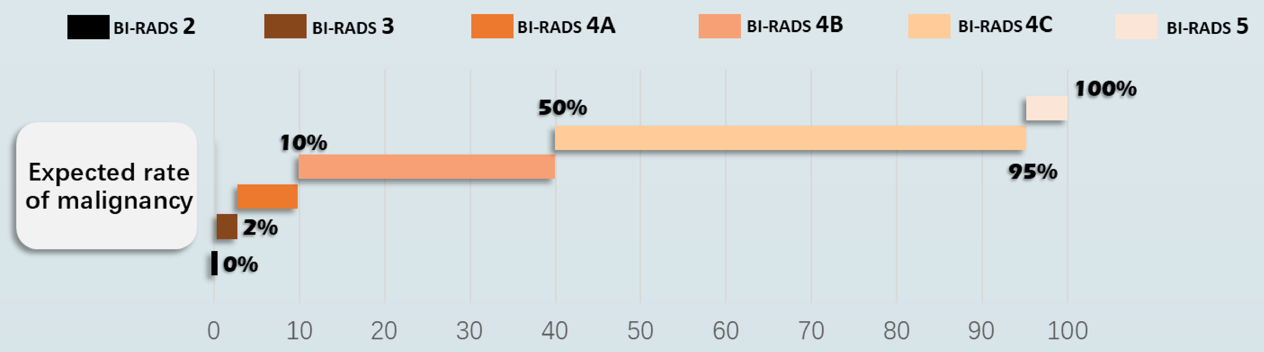

Fig. 1. Expected malignant risk of BI-RADS categories.

Challenges

- Optimize the use of relevance between the pathology and BI-RADS categories to improve classification accuracy.
- Minimize the gap between predicted BI-RADS categories and annotations given by doctors.

Methodology

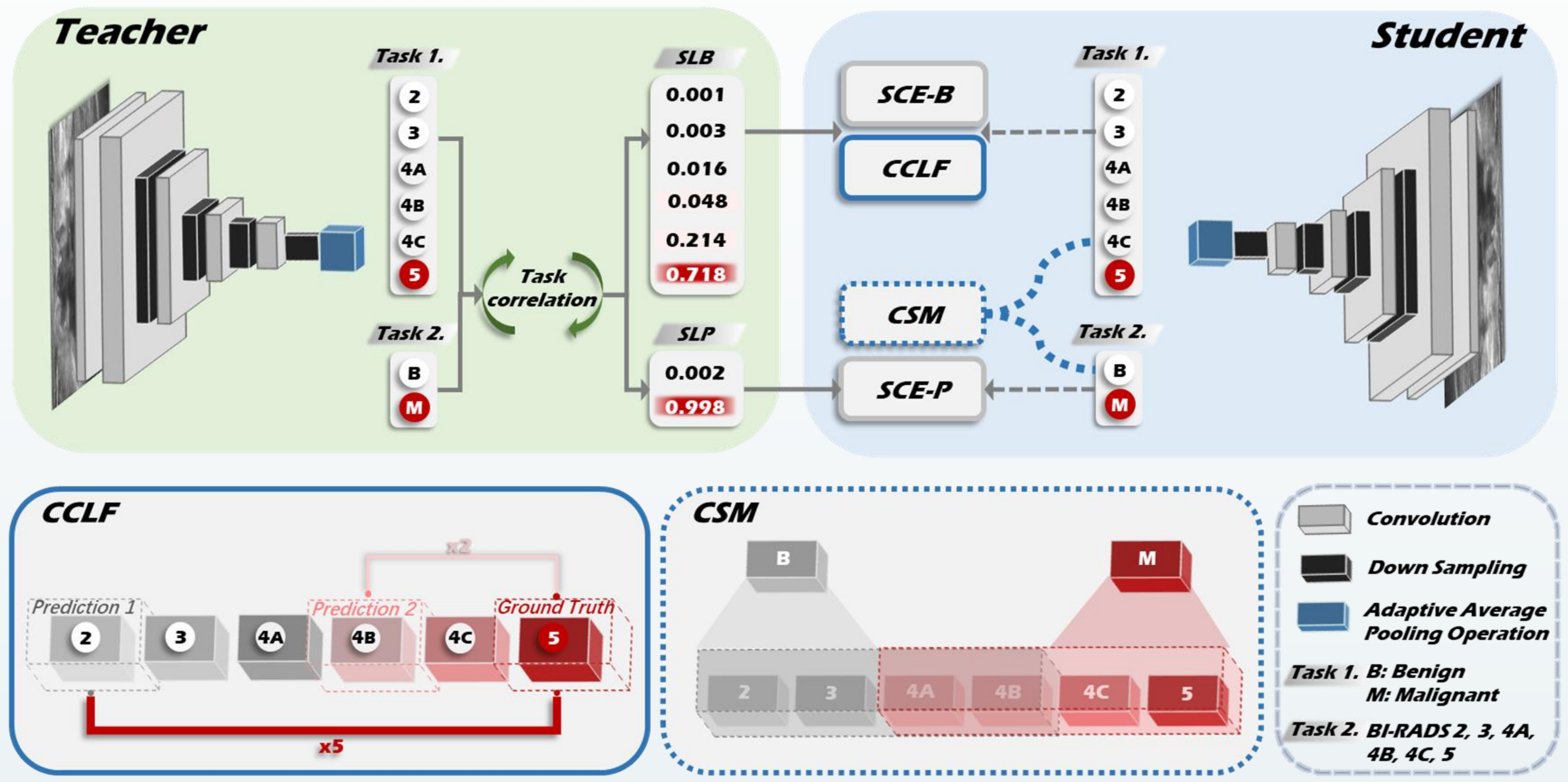

Fig. 2. Overall framework.

Multitask Label Generating Architecture. The proposed framework was based on a teacher-student model. The output in each model consisted of two parts: the BI-RADS and pathology. In order to obtain precise task-related soft labels, only outputs which both BI-RADS and pathology predictions were right were considered. Then, for certain category soft labels of BI-RADS, the average of corresponding predicted BI-RADS probabilities was calculated, meanwhile, the mean of related predicted pathology probabilities was computed to obtain soft labels of pathology.

Consistency Supervision Mechanism (CSM). To utilize the relevance between the BI-RADS and pathology categories to improve the overall classification accuracy and consistency between them, a cross entropy loss between pathology ground-truth and predicted BI-RADS probabilities was proposed, which defined as:

$$L_c = \begin{cases} p_B \cdot log\left(1-\sum_{4c}^{5} b'(x)\right), p_B \geq 0.5 \\ p_M \cdot log\left(1-\sum_{2}^{3} b'(x)\right), & else \end{cases}$$

where $p_B + p_M = 1$. $p_B \geq 0.5$ means a input lesion is benign, otherwise it is malignant.

Cross-Class Loss Function (CCLF). To minimize the gap between predicted BI-RADS categories and annotations, the CCLF was introduced to penalize different degrees of misclassified items with different weights. The CCLF is defined as:

$$L_{cc} = e^{\|n-m\|} \cdot (b'_m(x) - b_m)^2$$

where n and m represent the n^{th} category annotated by radiologists and m^{th} category predicted by the model respectively.

Results

Dataset	Method	AUC^P	ACC	SENS	SPEC	PPV	NPV	AUC^B
BUSI	Shen et al	0.927	-	0.905	0.842	0.672	0.949	-
	Xing et al	0.889	0.843	0.758	0.883	0.751	-	0.832
	Ours	0.900	**0.859**	0.735	**0.916**	**0.803**	0.881	**0.884**
UDIAT	*Zhang et al*	*0.889*	*0.92*	-	-	-	-	-
	Byra et al	*0.893*	*0.840*	*0.851*	*0.834*	-	-	-
	Xing et al	0.870	0.859	0.685	0.945	0.860	-	0.872
	Ours	**0.905**	0.877	0.685	**0.972**	**0.925**	**0.862**	**0.916**

Table1. Performance comparison on BUSI and UDIAT. Our method outperforms others in most metrics on both datasets.

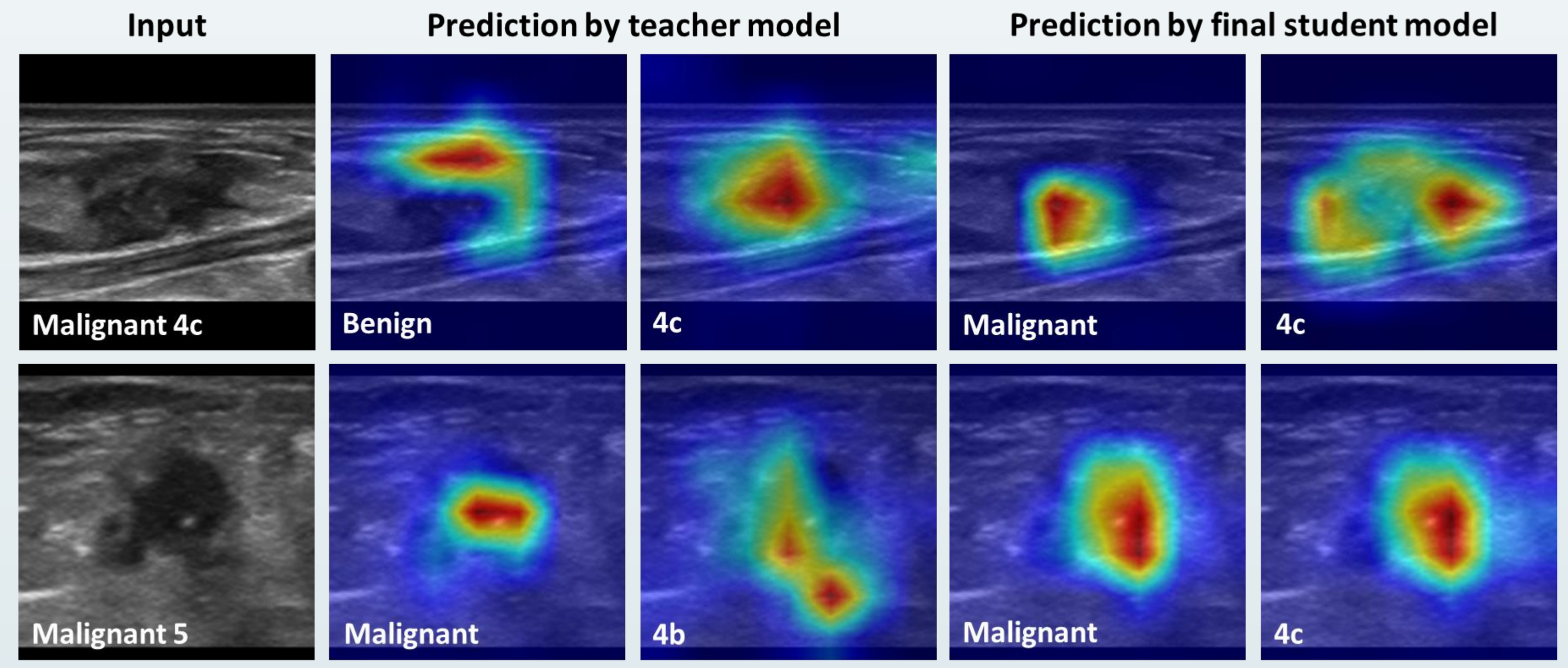

Fig. 3. CAMs [2] between the teacher model and final student model. The final student model was able to correct misclassifications (1st row). The BI-RADS category predicted by student model was closer to annotations (2nd row).

Conclusion

Our method successfully improved the classification accuracy and the consistency between two tasks, and minimized the gap between predicted BI-RADS categories and annotations.

References

1. Medelson EB., Böhm-Vélez M., Berg W.A., et al.: ACR BI-RADS® Ultrasound. In ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System. Reston, VA, American College of Radiology. (2013)
2. Selvaraju., Ramprasaath R., Michael Cogswell., et al.: Grad-cam: Visual explanations from deep networks via gradient-based localization. IEEE international conference on computer vision, pp. 618-626. (2017).