Leveraging Large-Scale Weakly Labeled Data for Semi-Supervised Mass Detection in Mammograms

Supplementary Material

Yuxing Tang[1], Zhenjie Cao[1], Yanbo Zhang[1], Zhicheng Yang[1], Zongcheng Ji[1],
Yiwei Wang[1], Mei Han[1], Jie Ma[2], Jing Xiao[3], Peng Chang[1]

[1]PAII Inc., USA   [2]Shenzhen People's Hospital, China   [3]Ping An Technology, China

Examples of mammography reports

All the mammography reports used in this study are in simplified Chinese. In Figure 1-3, we provide some examples of de-identified text reports and their English translation (automatically translated by Google Translate), together with the probabilistic labels predicted by our NLP model, namely, *Mammo-RoBERTa*.

诊断报告
(Original report)

放射学表现:
致密型乳腺；乳头未见凹陷、变形；乳晕、皮肤未见异常增厚；皮下脂肪清晰，悬韧带未见异常增粗。双侧腺体基本对称，量较致密，呈团片状，密度不均，左乳12点位置见分叶形肿块，约13mm×12mm，呈高密度，边缘似见浸润；腺体后脂肪层未见异常密度灶。双侧血管分布均匀。双侧腋下未见肿大淋巴结。

放射学诊断:
致密型乳腺：1.左乳上份肿块，性质待定，建议手术。BIRADS 4 类。2.右乳未见明确肿块及恶性钙化。BIRADS 1 类。3.双侧乳腺增生，请结合临床及相关检查除外隐匿性病变。

Diagnostic Report
(Automatic translation)

Findings:
Dense breast; no depression or deformation of the nipple; no abnormal thickening of the areola and skin; clear subcutaneous fat and no abnormal thickening of the suspensory ligament. The bilateral glands are basically symmetrical, dense in volume, in a lump, with uneven density. A lobular mass is seen at 12 o'clock in the left breast, about 13mm×12mm, with high density and infiltration at the edges; the fat layer behind the gland No abnormal density foci were found. The blood vessels are evenly distributed on both sides. No swollen lymph nodes were seen in the bilateral armpits.

Impression:
Dense breast: 1. A mass in the upper left breast. The nature is to be determined. Surgery is recommended. BIRADS 4 categories. 2. There is no clear mass and malignant calcification in the right breast. BIRADS Class 1. 3. Bilateral breast hyperplasia, please combine clinical and related examinations to exclude occult lesions.

Probabilistic labels by Mammo-RoBERTa

Breast	Left	Right
Prob. of mass	0.9908	0.0031

Figure 1. Mammography report of Example #1 and its labels generated by *Mammo-RoBERTa*. Reference standard (Ground-truth): a mass is present in the left breast; no clear mass in the right breast. Our NLP labels indicate a confidence score of 0.9908 for a left mass and 0.0031 for a right mass, respectively.

诊断报告
(Original report)

<u>放射学表现:</u>
乳腺所属分型：混合型。
乳头：未见凹陷、固定、变形。
乳晕、皮肤：未见异常增厚。
皮下脂肪：增宽，腺体前缘光整，边界清楚。
悬韧带：未见显影。
腺体结构：密度不均，其间可见圆形及不规则致密影。
肿块及恶变征象：左乳外下象限可疑细小钙化5枚，不伴肿块结节。索带状导管影及团状致密区融合成团。
导管相：未见增生扩张。
血管相：走行自然，未见迂曲、增粗、变形。
腺体后脂肪层：清晰；未见异常。
腋下：未见肿大淋巴结。

<u>放射学诊断:</u>
1.左乳外下象限可疑细小钙化5枚，请结合临床，必要时短期复查。2.双侧乳腺增生。请结合临床，随诊。

Probabilistic labels by Mammo-RoBERTa

Breast	Left	Right
Prob. of mass	0.0100	0.0041

Diagnostic Report
(Automatic translation)

<u>Findings:</u>
Types of mammary glands: mixed type.
Nipple: No depression, fixation, or deformation.
Areola and skin: No abnormal thickening was seen.
Subcutaneous fat: widened, the front edge of the gland is smooth, and the boundary is clear.
Suspension ligament: No development.
Glandular structure: The density of glands is uneven, and round and irregular dense shadows can be seen between them.
Lump and signs of malignant transformation: There are 5 suspicious small calcifications in the lower left quadrant of the breast, without mass nodules. The cord-like duct shadow and the cluster-like dense area fuse into a cluster. Duct phase: no hyperplasia and expansion.
Vascular phase: natural walking, no tortuosity, thickening, or deformation.
Fat layer behind glands: clear; no abnormalities.
Armpit: No swollen lymph nodes were seen.

<u>Impression:</u>
1. There are 5 suspicious small calcifications in the lower outer quadrant of the left breast. Please review it in a short-term basis if necessary. 2. Bilateral breast hyperplasia. Please follow up with the clinic.

Figure 2. Example #2. Reference standard: there were no masses on both breasts. Our *Mammo-RoBERTa* model predicts that the probability of a mass is very low for both breasts.

诊断报告
(Original report)

<u>放射学表现:</u>
双侧乳腺实质构成属不均匀致密型（ACR C型）；乳头未见凹陷、变形；乳晕、皮肤未见异常增厚；皮下脂肪清晰，悬韧带增粗，呈牛角样突起。双侧腺体基本对称，量较多，呈片絮状，密度不均，右乳外上象限见一椭圆形等密度肿块，未见分叶，边缘部分清楚部分模糊，未见明确钙化或毛刺，大小约2.0×1.8cm，右乳另见两枚细小点状及粗大钙化；左乳见两枚细小点状及粗大钙化，未见明确肿块；腺体后脂肪层未见异常密度灶。双侧血管分布均匀。双侧腋下未见肿大淋巴结。

<u>放射学诊断:</u>
(1)双侧乳腺实质构成：不均匀致密型（ACR C型）。(2)右乳外上象限肿块，拟良性可能性大，建议密切结合超声检查。BIRADS 3 (3)双乳良性钙化。BIRADS 2 (4)双侧乳腺增生。请结合临床，随诊。

Probabilistic labels by Mammo-RoBERTa

Breast	Left	Right
Prob. of mass	0.3483	0.9543

Diagnostic Report
(Automatic translation)

<u>Findings:</u>
The parenchymal composition of the bilateral mammary glands was heterogeneous and dense (ACR type C); the nipples were not depressed or deformed; the areola and skin were not abnormally thickened; the subcutaneous fat was clear, and the suspensory ligament was thickened with a bull's-horn-like protrusion. The gland was basically symmetrical bilaterally, with a large amount of flocculent flakes and uneven density. An oval-shaped isointense mass was seen in the upper outer quadrant of the right breast, without lobulation, with partially clear and partially blurred margins and no clear calcification or burr, about 2.0×1.8 cm in size, and two tiny punctate and coarse calcifications in the right breast; two tiny punctate and coarse calcifications in the left breast, without clear mass; no abnormal density foci were seen in the posterior fat layer of the gland There was no abnormal density foci in the posterior fat layer. The blood vessels were evenly distributed bilaterally. No enlarged lymph nodes were seen in the bilateral axillae.

<u>Impression:</u>
(1) Bilateral breast parenchymal structure: uneven and compact type (ACR type C). (2) A mass in the upper outer quadrant of the right breast is likely to be benign. It is recommended to closely combine with ultrasound. BIRADS 3 (3) Benign calcification of both breasts. BIRADS 2 (4) Bilateral breast hyperplasia. Please follow up with the clinic.

Figure 3. Example #3. Reference standard: there is no clear mass in the left breast; a mass is present in the right breast. Our NLP labels indicate a confidence score of 0.3483 for a left mass and 0.9543 for a right mass, respectively.