

Oncology Evaluation: failure cases

In this survey you will review cases that were rejected by our automatic verifier model. The model assigns each rejection to one of seven error types. Whenever an essential item is involved, we provide a brief explanation.

For each MCQ you will see:

- the image(s)
- the original caption from the paper
- the paragraph in which the figure was cited ("Context")
- the generated MCQ and its proposed answer
- the essential item that triggered rejection and the verifier's explanation

Please keep in mind that the verifier model is imperfect and may be wrong. We ask you to judge whether you agree with the model's decision for each case and to provide your comments.

Because some context can be very long, we recommend you to skip this part for each question. You can refer to context part when necessary.

1. Caption: A 37-year-old woman with a palpable mass on the upper outer quadrant of the right breast. (a) Cranio-caudal mammogram. (b) Mediolateral Oblique mammogram. Arrows in (a and b) mark the lesion. The triangular skin marker indicating the location of the palpable mass is visible in (b). (c-e) Ultrasound imaging show a circumscribed $1.6 \times 1.8 \times 1.5$ cm hypoechoic mass with flow and post enhancement at 11 o'clock position and 10 cm from nipple corresponding to the mass in mammograms. (f) Ultrasound image from biopsy procedure showing the needle. (g) invasive ductal carcinoma, grade 1, showing infiltrating malignant glands within desmoplastic stroma (H&E stain). (h) high-power magnification showing predominant tubular formation (arrows), mild nuclear atypia with rare/no mitosis (H&E stain). (i) ER immunostain showing strong diffuse positivity with anti-ER antibody.

Context: Tumor differentiation graded on an ordinal scale (3: poorly-differentiated, 2: moderately-differentiated, and 1: well-differentiated) was positively correlated with US measured tumor size ($p = 0.365$, $P < 0.001$), circumscribed ($p = 0.115$, $P = 0.014$) or microlobulated ($p = 0.160$, $P < 0.001$) margins, heterogeneous echo texture ($p = 0.100$, $P = 0.033$), posterior enhancement ($p = 0.224$, $P = 0.009$) and hypoechoic lesions ($p = 0.201$, $P < 0.001$), and negatively correlated with angular or spiculated margins ($p = -0.138$, $P = 0.003$), isoechoic lesions ($p = -0.193$, $P < 0.001$), and posterior acoustic shadowing ($p = -0.235$, $P = 0.006$). The proportional odds assumption could not be satisfied ($P < 0.001$). Hence, a generalized linear model was fit. After adjusting for age, the likelihood for poorly differentiated tumor increased with US measured tumor size (OR: 1.066; CI: 1.042–1.091, $P < 0.001$), and hypoechoic features (OR: 2.044; CI: 1.337–3.126; $P = 0.001$), and decreased with angular or spiculated margins (OR: 0.605; CI: 0.393–0.931; $P = 0.022$), and posterior acoustic shadowing (OR: 0.352; CI: 0.238–0.523; $P < 0.001$). [Figure 1](#) shows an example case of a high-grade, ER/PR positive tumor. [Figure 2](#) shows an example case of a low-grade, ER negative tumor. Follow-up multivariable logistic regression analysis was performed by dichotomizing to high-grade versus low- or moderate-grade tumors. The overall model was significant (likelihood ratio test, $P < 0.0001$) and satisfied Hosmer-Lemeshow goodness-of-fit test ($P = 0.807$). [Figure 3](#) shows the ROC curve for the overall model with the aforementioned four US morphological features and achieved an AUC of 0.799 (CI: 0.752–0.845). Regarding US measured tumor size, the mean (IQR) for well differentiated, moderately differentiated, and poorly differentiated tumors were 10 (7–15) mm, 14 (9–20) mm, and 16.5 (12–25) mm, respectively. [Figure 4](#) shows the ROC curve for the overall model with the aforementioned four US morphological features and achieved an AUC of 0.799 (CI: 0.752–0.845). Regarding US measured tumor size, the mean (IQR) for well differentiated, moderately differentiated, and poorly differentiated tumors were 10 (7–15) mm, 14 (9–20) mm, and 16.5 (12–25) mm, respectively.

Q: A 37-year-old woman presents with a palpable mass in the upper outer quadrant of the right breast. Imaging and biopsy reveal a circumscribed hypoechoic mass with flow, invasive ductal carcinoma, grade 1, with predominant tubular formation, mild nuclear atypia, and strong diffuse ER positivity. Which histopathological feature is most consistent with the tumor grade?

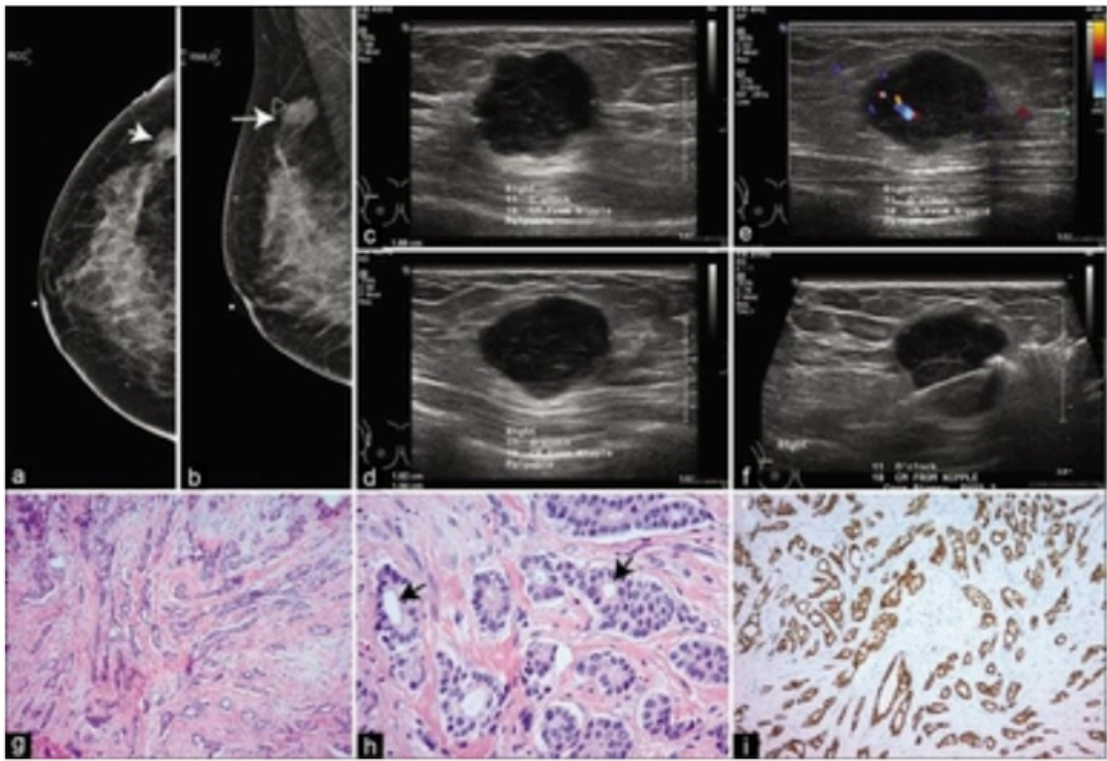
Options:

- A. Poorly differentiated with marked nuclear atypia and frequent mitoses
- B. Moderately differentiated with solid nests and moderate nuclear atypia
- C. Well-differentiated with predominant tubular formation and mild nuclear atypia
- D. In situ carcinoma with comedonecrosis
- E. Metastatic carcinoma with signet ring cells

Answer: C

Verifier response: **Diagnosis Leak** NOT PASS: Stem restates diagnosis from caption, and the question can be answered directly by stem.

To avoid "**Diagnosis Leak**" means "The stem does not restate the diagnosis verbatim, and the answer cannot be directly inferred according to the stem".



Mark only one oval.

- ☐ Yes, I agree with verifier's idea.
- ☐ Other: _____

2. Caption: Screening mammogram of a 59-year-old postmenopausal female. Her calculated BMI is 38.8 (obese). It is matching a predominantly fatty breast parenchyma (ACR B). Her mammogram shows a left LIQ malignant mass lesion (long arrows) and a right UOQ intra-mammary non-specific node (short arrows).

Context: <none>

Q: Which of the following findings is present in the mammogram of the 59-year-old postmenopausal female with a BMI of 38.8?

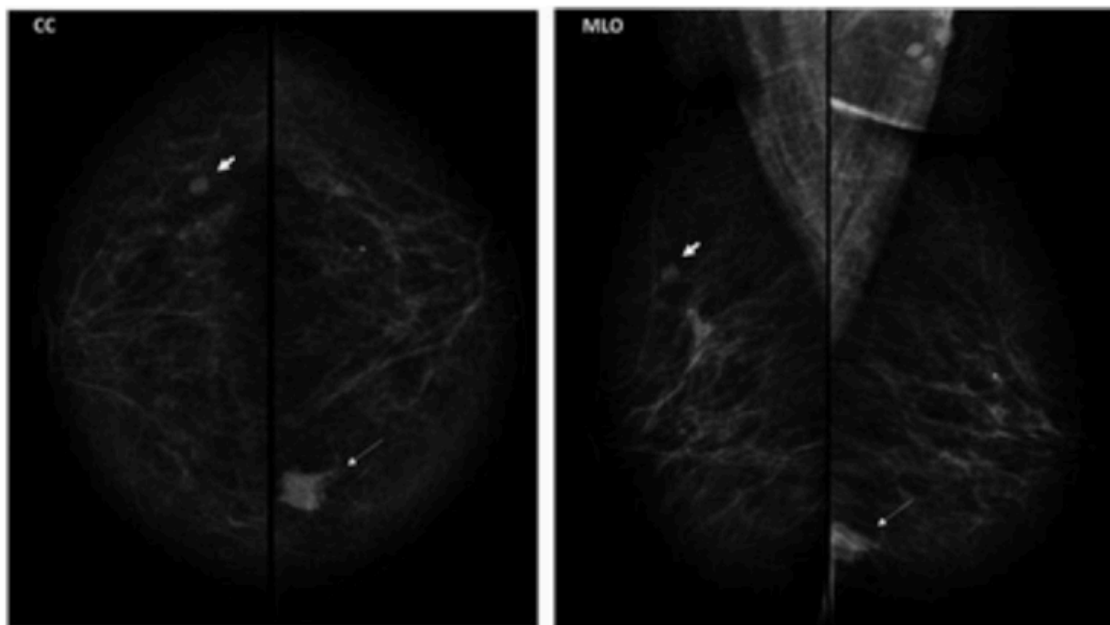
Options:

- A. Left lateral inferior quadrant (LIQ) malignant mass lesion
- B. Right upper outer quadrant (UOQ) intra-mammary non-specific node
- C. Right lateral inferior quadrant (LIQ) malignant mass lesion
- D. Left upper outer quadrant (UOQ) intra-mammary non-specific node
- E. Bilateral fatty breast parenchyma with no focal lesions

Answer: A

Verifier Response: **Single Correct Option** NOT PASS: Multiple keys possible.

"**Single Correct Option**" means "only one option is correct".



Mark only one oval.

- ☐ Yes, I agree with verifier's idea.
- ☐ Other: _____

This content is neither created nor endorsed by Google.

Google Forms