The NYU Breast MRI Dataset v1.0

Jan Witowski[a,d,*], Shizhan Gong[b], Nan Wu[b], Linda Moy[a,c,d,e], Laura Heacock[a,c], Beatriu Reig[a,c], S. Gene Kim[a,c,d], Florian Knoll[a,d], and Krzysztof J. Geras[a,b,d,e,*]

[a]Department of Radiology, New York University School of Medicine, 660 First Ave, New York, NY 10016, USA; [b]Center for Data Science, New York University, 60 5th Ave, New York, NY 10011, USA; [c]Perlmutter Cancer Center, NYU Langone Health, 160 E 34th St, New York, NY 10016, USA; [d]Center for Advanced Imaging Innovation and Research, NYU Langone Health, 660 First Ave, New York, NY 10016, USA; [e]Vilcek Institute of Graduate Biomedical Sciences, NYU Grossman School of Medicine

This manuscript was compiled on March 29, 2022

MRI is the highly sensitive modality in breast cancer diagnosis, with a growing range of clinical indications. Large and diverse datasets are necessary for the development of robust and accurate artificial intelligence models, but to this date their existence is uncommon. This technical report introduces *the NYU Breast MRI Dataset*, which consists of 21,537 MRI studies (N=13,463 patients) acquired between 2008 and 2020 at NYU Langone Health. Below we outline its statistics, details of image collection and preprocessing. We specify the process of dataset filtering and label extraction. This dataset includes images, biopsy-proven breast-level cancer labels, as well as BI-RADS risk and background parenchymal enhancement labels. Although this is a private dataset, we are publishing this report to improve reproducibility of our work and to share practices and insights that might be useful to others.

1. Statistics of the dataset

The dataset consists of 21,537 imaging studies from 13,463 patients who underwent breast magnetic resonance imaging between 2008 and 2020. All patient data was obtained with the approval of our institutional review board. Imaging exams are bilateral breast dynamic contrast enhanced MRI (DCE-MRI) studies, acquired using 3.0T or 1.5T magnet MRI scanners. Images were acquired in either sagittal or axial plane by default. Magnevist or Gadavist contrast agents were used. Each study contains T1- and T2-weighted series, subtraction series and and maximum intensity projection images. Some studies include additional series, e.g. ultrafast sequences, which were not included as a part of the dataset.

In addition to the images, the dataset contains associated breast-level cancer labels and auxiliary data about imaging, cancer and demographic characteristics, collected from associated radiology reports, pathology reports and EHR (electronic health record) data. Cancer labels for each study consist of *(a)* a binary label for each breast indicating whether a matching biopsy showed that there is at least one *malignant* finding in the breast, and *(b)* a binary label for each breast indicating whether a biopsy showed that there is at least one *benign* finding in the breast. All cancer labels are biopsy-proven, as described later in this report.

A. Splitting the data into training, validation and test sets. We randomly split studies into three independent subsets: training (60%), validation (15%) and test (25%), based on patients identifiers. After patient randomization, there were 14,198 studies in the training set; 3,516 in the validation set, and 5,958 in the test set. Following additional test set filtering (Sec. 3), ultimately this subset contained 3,936 studies.

B. Breast-level cancer labels. To obtain labels indicating whether each breast of the patient was found to have malignant or benign findings at the end of the diagnostic pipeline, we used pathology reports from breast biopsies and surgical procedures. We have 6,609 exams with at least one pathology report associated with a patient within 120 days before or after the MRI study date. 3,780 (57.2% of biopsied exams; 17.6% of all exams) studies had at least one pathology-proven malignancy. In 5,332 (80.7% of biopsied exams; 24.8% of all exams) studies, at least one pathology report reported benign findings. 2,503 (37.9% of biopsied exams; 11.6% of all exams) MRI studies were associated with both benign and malignant findings. The remaining 14,928 (69.3%) exams were not matched with any pathology report and thus were assigned a negative label, corresponding to the absence of malignant or benign findings in both breasts. The details of the label extraction process from pathology reports are available in the NYU Breast Cancer Screening Dataset report (1).

C. BI-RADS labels and terms. We used the associated radiology reports to extract Breast Imaging-Reporting and Data System (BI-RADS) (2) labels for risk scoring, background parenchymal enhancement (BPE), and the amount of fibroglandular tissue (FGT). Patients with marked BPE (the highest level of contrast enhancement of fibroglandural breast tissue) are at higher risk for false negatives in breast imaging (3), which is why extracting information about subgroups is necessary. Both risk scoring and BPE labels follow BI-RADS lexicon (0-6 for risk scoring; 'minimal', 'mild', 'moderate' and 'marked' for BPE). For details on how BI-RADS labels were extracted from radiology reports, please refer to sections 4.B and 4.C.

2. Image collection and preprocessing

In this section, we explain in detail the complete processing pipeline for filtering, selecting and preprocessing of MR images in our data set. This pipeline consists of three major phases: (A) data extraction and verification, (B) image resampling and reorientation, and (C) filtering.

A. Image collection and verification. After an initial query to our institution's PACS, unique studies were identified with the `Study Instance UID` DICOM tag value, meanwhile patients were identified with the `Patient ID` tag. DICOM files were loaded and processed with either the SimpleITK library with a GDCM image reader or a Pydicom library with a Pylibjpeg reader. Series, organized by `Series Instance UID`, have been checked in terms of geometry information. Series that include images without spatial information (`Image Orientation (Patient)` and `Image Position (Patient)`) have been excluded from the dataset, as it is impossible to reliably sort images without geometry data.

*To whom correspondence should be addressed. E-mail: k.j.geras@nyu.edu and jan.witowski@nyulangone.org.

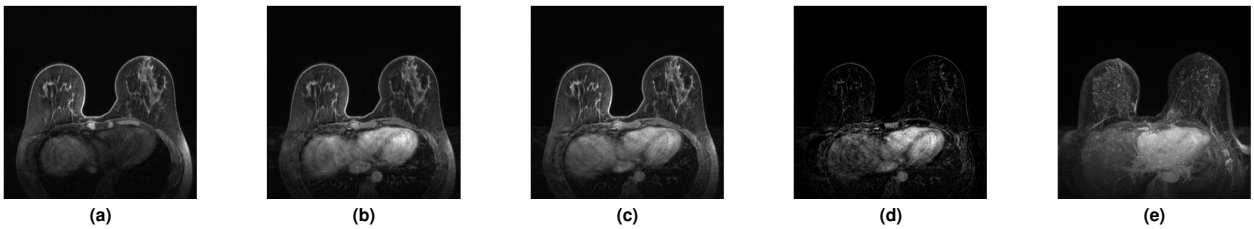

Fig. 1. Example study from the NYU Breast MRI Dataset. Images present selected views from T1 fat-saturated series: **(a)** pre-contrast, **(b)** first post-contrast; **(c)** second post-contrast; **(d)** second post-contrast subtraction, and **(e)** maximum intensity projection. All images show breasts at the same level (slice in the middle of the volume), in axial projection. The presented sample study does not contain any suspicious enhancing masses or areas of ductal enhancement in either breasts and has been rated as BI-RADS 2 (benign). Note that in the process of filtering our dataset, we excluded subtraction and maximum intensity projection images, as those can be easily generated from pre- and post-contrast images only.

B. Reorientation and resampling. Our data set consists of images acquired in different planes, most commonly in the axial or sagittal plane. In recent years, the acquisition protocol at our institution has changed and currently all DCE-MRI studies are by default acquired in the axial plane. To ensure that all pixel matrices in our dataset are consistent in terms of anatomical orientation, we unified the images within the anatomical coordinate system. Direction cosines from the `Image Orientation (Patient)` DICOM tag have been collected to establish the current orientation. We reoriented all studies to the LPS (left-posterior-superior) orientation, which is a common orientation used in medical image processing, and a standard orientation for DICOM axial images.

Because images in our dataset were acquired with a number of scanners and with different acquisition protocols, image spacing in studies is inconsistent. Thus, we resampled all volumes to the same anisotropic spacing of (0.714, 0.714, 1.2), which was the most common spacing in each of the axes (X, Y, Z). Images were resampled with a linear interpolator, implemented in the SimpleITK framework.

C. Overall filtering. To reduce potential noise in data, we established a set of rules for filtering out studies. A flowchart presenting excluded studies is shown in Figure 2.

C.1. Selecting series of interest. In the first step, we defined a set of series names that represent T1-weighed pre- and post-contrast series with fat suppression. This means that at this point we excluded T2-weighted series, subtraction images, reconstructions in other planes than the original acquisition and maximum intensity images. Series name matching was done by comparing the `Series Description` DICOM tag `(0008,103E)` values to a predefined set of acceptable series names. For a whole study to be included to a dataset, it needed to have at least three acceptable series, because our workflow expects to have one pre-contrast and at least two post-contrast T1 images. The lexicon of acceptable series descriptions was verified manually so that it included all relevant pre- and post-contrast series names.

We further evaluated the correctness of the series description lexicon by manually reviewing sample studies with different series name combinations. There were approximately 50 combinations of series names, coming from various acquisition protocols, that we manually reviewed for abnormalities, e.g. multiple pre-contrast series.

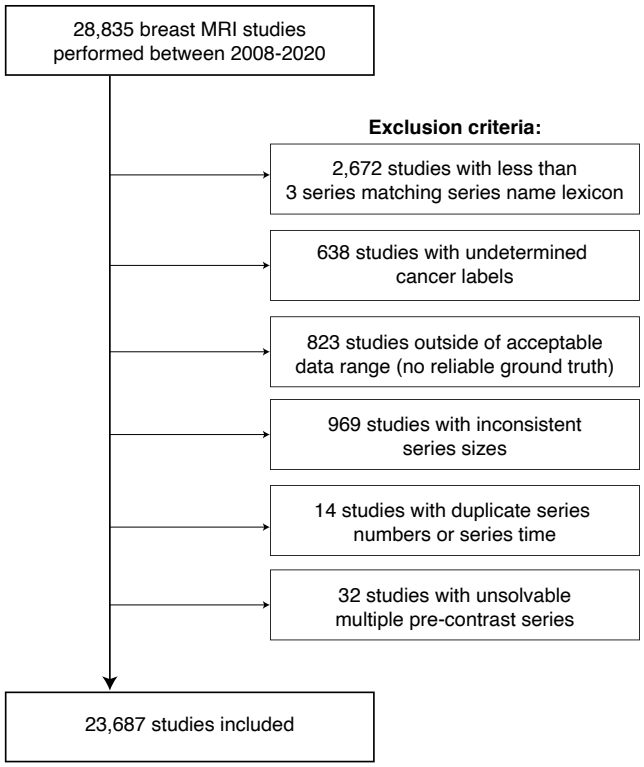

Fig. 2. Full dataset filtering flowchart. This figure presents reasons for exclusion and number of excluded studies in our dataset, as described in detail in section 2C. Please note that studies in the test set underwent additional filtering (not shown here), as described in section 3.

C.2. Series consistency. As a consistency check, we evaluated each study in terms of series acquisition time and series number. If there were at least 2 series in a single study with the same `Series Number (0020,0011)` or `Series Time (0008,0031)` tag values, we excluded that study, as it is impossible to reliably determine the order of the series. There were 14 studies with duplicate Series Numbers or Series Times.

We also use the information from the `Series Time` tag values to establish a correct order of series in a study. Series in all studies were ordered assuming that the first (earliest `Series Time`) series is pre-contrast, and the following series are post-contrast.

We make a note that there might be some variability to

our dataset due to different waiting times between pre- and post-contrast series. In our dataset, post-contrast series were usually acquired 2 minutes after the bolus.

Moreover, we excluded studies that had inconsistent volume sizes. We made an assumption that both pre- and post-contrast series should have the same dimensions. We excluded 969 studies that did not comply with that rule.

C.3. Excluding studies with unreliable cancer labels. Because determining the ground truth for each datum depends on extracting information from pathology reports, we excluded those studies for which pathology reports might be missing. As described in this report, we generate labels by matching studies pathology reports dated 120 days before or after the MRI exam. However, our database for both imaging and pathology reports is time-limited to a range from 2008 to 2020. This means that if the date of the MRI study falls within the first or last 119 days of this time range, it is possible that we are missing a pathology report that was dated outside of the 2008-2020 period. For this reason, these studies are excluded.

Additionally, in some situations, a pathology report was matched with a DCE-MRI study, but our data extraction pipeline was not able to collect information from the report. This can happen, for example, when information about the pathology report was pulled from the hospital system, but its contents were missing.

C.4. Additional heuristics. We noticed rare situations where the accepted series differed from most exams, e.g. by including more than one pre-contrast series. Studies with multiple pre-contrast sequences are possible due to problems with patient compliance, dosage administration or hardware issues. In those situations, pre-contrast series will be repeated and could potentially lead to generation of subtraction images based on the wrong pre-contrast series. To address that, we developed a simple rule-based heuristic algorithm that checks for pre-contrast-specific series names, attempts to fix problems (e.g. by selecting later pre-contrast sequence), and, if an automatic fix is not possible, excludes studies from the dataset. There were 32 studies we excluded because of multiple pre-contrast series.

3. Test set filtering

To further minimize noise in the data when evaluating AI models, we performed additional filtering of the test set. A flowchart of this process is shown in Figure 3.

A. Limiting targeted patient population. We made a decision to limit the population in which our AI models are intended to be used. We excluded patient groups in which MRI images can look very different from most patients undergoing DCE-MRI. Those are: patients after bilateral mastectomy, patients after neoadjuvant chemotherapy, and patients with breast implants. We believe that those subgroups require larger representation in the training set for models to perform well. Here, we exclude those groups and do not intend to apply AI models. In total, 640 studies were excluded from the test set.

B. Excluding cases with potentially noisy labels. While neural networks are capable of learning from datasets with noisy data, it is critical to evaluate AI models on datasets with maximally clean data. For this reason, we established stringent

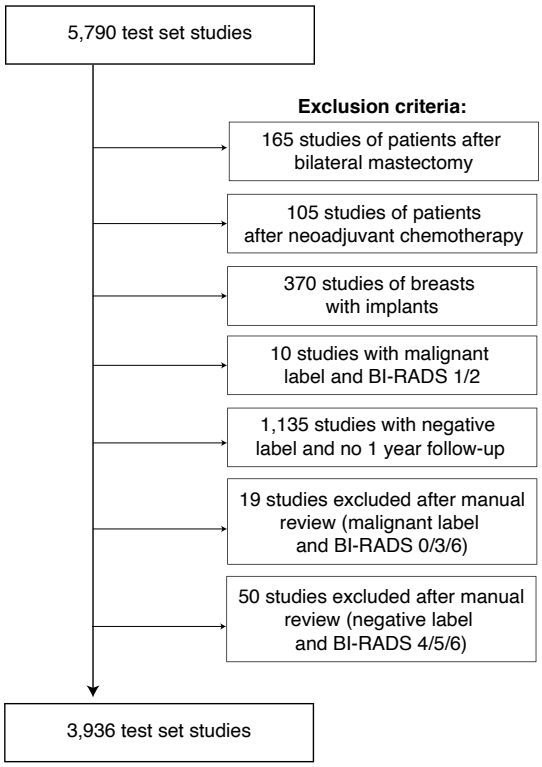

Fig. 3. Additional filtering of the test set. This flowchart presents reasons for exclusion and the number of studies excluded from the test set only. Criteria and reasoning are described in detail in section 3.

rules to exclude test set studies for which labels might be noisy. This process had two steps: automatic filtering and manual revision.

We first excluded studies where the label (from pathology report) was malignant, but BI-RADS assigned to the study was category 1 or category 2, suggesting negative or benign result. Second, we excluded all studies with a negative label (i.e. no pathology report associated with the study), which did not have a 1 year negative follow-up. By "negative follow-up" we mean a situation where in a year after the study date: *(1)* there are no pathology reports associated with the patient; *(2)* there is at least one breast imaging study (mammography, MRI) with BI-RADS category 1, 2 or 3; *(3)* there are no breast imaging studies with BI-RADS category 0, 4, 5 or 6. In summary, we automatically excluded 1,145 studies with potentially noisy labels.

Finally, we manually reviewed studies with potentially conflicting cancer labels and BI-RADS categories. We manually reviewed studies with malignant labels (i.e. associated with at least one pathology report yielding malignant results) and associated with BI-RADS categories 0, 3 or 6. From 378 test set studies matching this rule, 19 were excluded upon a manual review. We also manually reviewed studies with negative labels (i.e. no associated pathology reports and negative follow-up) and associated BI-RADS categories 4, 5 or 6. Out of 204 studies matching this rule, 50 were excluded.

The manual revision was done by board-certified radiologists on our team and involved manually checking DCE-MRI images, radiology reports, pathology reports or patient history to confirm the correctness of the labels.

4. Label extraction

A. Biopsy-proven cancer labels. We extracted breast-level labels for presence of benign or malignant findings from patient pathology reports. Initial processing and data extraction was performed using the same method as described in our screening mammography data report (1). Pathology reports were generated as a result of either breast biopsy or a surgical procedure.

We matched extracted pathology reports with studies if the pathology report was dated within 120 days before or after the DCE-MRI exam. If there were multiple pathology reports, labels were aggregated. This means that if at least one of associated pathology reports contained malignant findings, the breast in question was assigned a malignant label. Please note that malignant and benign findings are not mutually exclusive.

Additionally, in the test set, studies that were not matched with any pathology reports were assumed to be negative if and only if they had a one-year negative follow-up available, as described in section 3.B.

B. BI-RADS risk assessment labels. Free-text radiology reports for all included studies were collected and matched with exams with accession numbers. For extraction of BI-RADS labels included in reports, we developed a lexicon of phrases used to describe BI-RADS either as a number (e.g. "BIRADS: 3") of a verbose phrase (e.g. "bi-rads: probably benign"). We followed categories as defined by the 5th edition of the ACR BI-RADS Atlas.

In situations where report yielded multiple conflicting BI-RADS categories, reports were reviewed manually to find the correct label.

C. Background parenchymal enhancement labels. Similarly to the BI-RADS assessment, we collected the level of background parenchymal enhancement (BPE) which describes visually estimated enhancement of the fibroglandular tissue in breasts. We followed categories defined by the BI-RADS Breast MRI Lexicon: 'minimal', 'mild', 'moderate' and 'marked'. We did not evaluate whether BPE was symmetrical. We developed a dictionary of descriptors used by radiologists in their reports to describe specific BPE levels.

- *a (minimal)*: 'minimal', 'no', 'minimal to no' [BPE],
- *b (mild)*: 'mild',
- *c (moderate)*: 'moderate',
- *d (marked*): 'marked', 'significant', 'severe', 'extensive', 'extremely dense'.

In situations where the extraction process yielded conflicting BPE labels, those reports were reviewed manually.

5. References

1. Wu N, et al. (2019) The NYU breast cancer screening dataset v1.0, Technical report. Available at https://cs.nyu.edu/~kgeras/reports/datav1.0.pdf.
2. Morris E, Comstock C, Lee C, , et al. (2013) ACR BI-RADS Magnetic Resonance Imaging in *ACR BI-RADS Atlas: Breast Imaging Reporting and Data System*. (American College of Radiology).
3. Giess CS, Yeh ED, Raza S, Birdwell RL (2014) Background parenchymal enhancement at breast mr imaging: normal patterns, diagnostic challenges, and potential for false-positive and false-negative interpretation. *Radiographics* 34(1):234–247.

A. Extended data

A. Mastectomy status extraction. After manual review of a large number of radiology reports associated with the exams, we created several regular expressions to extract post-mastectomy status:

```
[Pp]ost[^\.\n]+(left|right|bilateral|double)(
    nipple[ -]sparing| nipple[ -]sparring)?( breast
    )? mastectom
[Pp]ost[^\.\n]+(left|right|bilateral|double)( total
    )? nipple( and areola)?[ -]sparing mastectom
[Hh]istory[^\.\n]+(left|right|bilateral|double)(
    breast)? mastectom
[Hh]istory( of)?[^\.\n]+(left|right|bilateral|
    double)( breast)? mastectom
(undergone|underwent)( a)? (left|right|bilateral|
    double) mastectom
[Pp]ost[^\.\n]+(left|right|bilateral|double)
    partial( breast)? mastectom
```

Only groups containing words 'left', 'right', 'bilateral' and 'double' were extracted. The word 'double' was converted to 'bilateral' for uniformity. If there were more than two conflicting results, e.g. 'left' and 'bilateral' in the same report, that report was flagged as 'conflict'. Radiology reports that did not yield any results were marked as 'unknown'.

B. Molecular tumor type extraction. To extract Ki-67, HER2+, ER and PR receptor statuses, we manually reviewed pathology reports associated with malignant findings. After the review, we developed a dictionary of regular expressions (regexps) to accurately extract information about the molecular type of lesions.

The following are regexps for Ki-67 status extraction:

```
([Kk]i-?67\s*\(30-9; Ventana\):)\s*(<?[0-9]+\s?%)
[Kk][Ii]-?67:?\s(<?\d?\d?\d\s?%)
[Kk][Ii]-?67 is estimated at (<?\d?\d?\d\s?%)
[Kk][Ii]-?67 shows a (<?\d?\d?\d\s?%)
[Kk][Ii]-?67 proliferation index\sis approximately
    (<?\d?\d-?\d?\d%)
[Kk][Ii]-?67 is (<?\d?\d?\d\s?%)
[Kk][Ii]-?67.*is (very high)$
[Kk][Ii]-?67.*is.+(high)$
[Kk][Ii]-?67[^\.]+\s(\d?\d?-?\d?\d%)
[Kk][Ii]-?67\s+Percentage of Positive Nuclei:\s
    +(<?\d?\d\s?%)
[Kk][Ii]-?67[^\.]+(<\d?\d?-?\d?\d%)
```

If Ki-67 status was reported numerically, we further established whether it was equal or greater to 14%, as this is often a criterion for luminal A breast cancer subtypes.

For HER-2 status extraction, the following regexps were used:

```
Her2/Neu \(4B5, Ventana\):\s+(\d\+?)
Her2/Neu \(4B5, Ventana\):\s+(N\/A)
[Hh][Ee][Rr]-?2 Score = ([\d.,]{1,6})
[Hh][Ee][Rr]-?2\s(negative)
[Hh][Ee][Rr]-?2\s(positive)
[Hh][Ee][Rr]-?-?2 \(4B5; Ventana\)\sInterpretation
    :\s(\d\+?)
[Hh][Ee][Rr]-?2\s(\d\+?)
[Hh][Ee][Rr]-?2\s(NEGATIVE|negative)
[Hh][Ee][Rr]-?2\s(POSITIVE|positive)
[Hh][Ee][Rr]-?2\/neu:\s(\d\+?)
(negative)\sfor\s[Hh][Ee][Rr]-?2
[Hh][Ee][Rr]-?2[^\.\n]+\((\d\+?)\)\.
Her-?2[^\.\n]+(negative)
H[Ee][Rr]2 (\(?-\)?)[.\n]
H[Ee][Rr]-?2:\s?(\d\+)
H[Ee][Rr]2\/neu is (equivocal)
ER[\-\+]/PR[\-\+]/H[Ee][Rr]-?2([\-\+])
H[Ee][Rr]-?2\s?(\(\+\))
(negative) for ER, PR and HER-?2
H[Ee][Rr]-?2 is (\d\+?)
H[Ee][Rr]-?2 \((\d\+)
H[Ee][Rr]-?2: (\d\+?)
[Hh][Ee][Rr]-?2\s?(\-|\+)
[Hh][Ee][Rr]-?2[^\.\n]+is ([Pp]ositive|[Nn]egative)
[Hh][Ee][Rr]-?2 \(4B5; Ventana\)\nInterpretation:
    ([Pp]ositive|[Nn]egative|[Ii]ndeterminate)
[Hh][Ee][Rr]-?2Neu Oncoprotein[^\.\n]+(POSITIVE|
    NEGATIVE|[Pp]ositive|[Nn]egative)
[Hh][Ee][Rr]-?2Neu Oncoprotein[^\.\n]+(\d\+)
[Hh][Ee][Rr]-?[^\.\n]+(\d\+)
[Hh][Ee][Rr]-?2?:?\n\n?(POSITIVE|NEGATIVE|[Pp]
    ositive|[Nn]egative|INDETERMINATE|[Ii]
    ndeterminate)
[Hh][Ee][Rr]-?2? \(IHC\):?[^\.\n]+(POSITIVE|
    NEGATIVE|[Pp]ositive|[Nn]egative|INDETERMINATE
    |[Ii]ndeterminate)
```

Results of this search were grouped into HER2-negative (score 0 or 1+), equivocal/borderline (score 2+) and positive (score 3+). Sometimes results were expressed verbatim as 'negative' or 'positive', and not as a numerical value.

The following are regular expressions for estrogen receptor (ER) status:

```
Estrogen\sReceptor\s\(SP1;\sVentana\):\s+(\d?\d
    ?\%?)
Estrogen\sReceptor\s\(SP1;\sLabvision\):\s+(\d?\d
    \%?)
(ER[\+\-])
ER\sInterpretation:\s(Positive|Negative)
ER\s(\d?\d\%?)
(positive)\sfor\sER
ER and PR (negative)
ER\s\((>?\d?\d\%?)
(positive)\sfor\s[Ee]strogen
ER\/PR(\+)
ER[^.\n]+(\d\d%)
ER, PR ([Nn]egative|[Pp]ositive)
ER\s?&\s?PR\s([Pp]ositive|[Nn]egative)
ER\sAND\sPR\s(NEGATIVE|POSITIVE)
(unspecified) [Ee]strogen
(negative) for [Ee]strogen
[Ee]strogen receptor (positive)
```

The following are regular expressions for progesterone receptor (PR) status:

```
ER and PR (negative)
ER\/PR(\+)
ER, PR ([Nn]egative|[Pp]ositive)
ER\s?&\s?PR\s([Pp]ositive|[Nn]egative)
ER\sAND\sPR\s(NEGATIVE|POSITIVE)
Progesterone\sReceptor\s\(1E2;\sVentana\):\s+(\d?\d
    ?\%?)
(PR[\+\-])
PR\sInterpretation:\s(Positive|Negative)
PR\s(\d?\d\%?)
(positive)\sfor\sPR
PR\s\((>?\d?\d\%?)
(positive)\sfor\s[Pp]rogesterone
PR[^.\n]+(\d\d%)
(unspecified) [Pp]rogesterone
(negative) for [Pp]rogesterone
[Pp]rogesterone receptor (positive|negative)
PR Score = [^\n]+([Pp]ositive|[Nn]egative)
```

C. Exam indication extraction. To extract the indication for MRI exam, we reviewed radiology reports associated with examinations. After the review, we developed a dictionary of regular expressions (regexps) to accurately extract information on exam indication in several categories.

Regexps in a radiology report associated with high-risk screening:

```
presenting for (a )?screening
presenting for (a )?high[- ]risk screening
presents for (a )?screening
high[- ]risk screening
```

```
routine screening
clinical history: screening
indication: screening
continued annual mr(i)? screening
```

Regexps in a radiology report associated with follow-up or surveillance examination:

```
six months follow-up
(presenting|presents) for (a )?six-month(s)? follow
    -up
(presenting|presents) for (a )?6[- ]month(s)?
    follow-up
(indication|clinical history): short interval (six-
    month )?follow-up
(indication|clinical history): 6[- ]month(s)?
    follow-up
(presenting|presents) for [^\.\n]+follow[- ]up
(presenting|presents) for follow[- ]up
(presenting|presents) for [^\.\n]+followup
clinical indication:[^\.\n]+short interval follow[-
     ]up
clinical indication:[^\.\n]+short interval followup
this is (a )?six-month(s)? follow[- ]up
this is (a )?six-month(s)? followup
mr(i)? performed for surveillance
(presenting|presents) for surveillance
(presenting|presents) for surveillance
(presenting|presents) for mr(i)? surveillance
clinical indication: (left |right )?breast ca
high-risk surveillance
annual (screening )?surveillance
high-risk evaluation
for surveillance mri
routine surveillance mri.
```

Regexps in a radiology report associated with workup:

```
evaluate (a )?questioned area
follow-up (mri ?)(status )?post (benign )?biopsy
presenting for further evaluation
presenting for evaluation of reported
indication:[^\.;\n]+for mri evaluation[^\.;\n]+
    findings
presenting for [^\.\n]+(nipple discharge|nipple
    inversion|pain|palpable|palpated|swelling|
    asymmetry)
mri[^\.;\n]+to evaluate[^\.;\n]+(nipple discharge|
    nipple version|pain|palpable|palpated|swelling|
    asymmetry)
mri[^\.;\n]+to further evaluate[^\.;\n]+(nipple
    discharge|nipple version|pain|palpable|palpated
    |swelling|asymmetry)
presenting [^\.\n]+problem solving
clinical information:[^\.;\n]+(nipple discharge|
    nipple version|pain|palpable|palpated|swelling|
    asymmetry)
underwent mri for (nipple discharge|nipple
    inversion|pain|palpable|palpated|swelling|
    asymmetry)
mri was advised.
mri for (further )?evaluation.
clinical history:[^\.;\n]+(nipple discharge|nipple
    version|pain|palpable|palpated|swelling|
    asymmetry)
for which mri was recommended.
```

Regexps in a radiology report associated with implant evaluation:

```
evaluate for implant
evaluation of implant
```

Regexps in a radiology report associated with evaluating the extent of disease/preoperative planning:

```
extent of disease
evaluation of disease
(indication|clinical history): (presurgical|
    preoperative)
presenting for (presurgical|preoperative)
here for (presurgical|preoperative)
```

Regexps in a radiology report associated with assessing response to treatment:

```
response to treatment
assess (for )?treatment
evaluation of treatment
```