Received: 6 June 2019 | Revised: 29 December 2019 | Accepted: 3 January 2020

DOI: 10.1002/widm.1357

ADVANCED REVIEW

WIREs DATA MINING AND KNOWLEDGE DISCOVERY WILEY

A review of mammographic region of interest classification

Sena B. Yengec Tasdemir | **Kasim Tasdemir** | **Zafer Aydin**

Department of Electrical and Computer Engineering, Abdullah Gul University, Kayseri, Turkey

Correspondence
Zafer Aydin, Department of Electrical and Computer Engineering, Abdullah Gul University, Kayseri 38080, Turkey.
Email: zafer.aydin@agu.edu.tr

Abstract

Early detection of breast cancer is important and highly valuable in clinical practice. X-ray mammography is broadly used for prescreening the breast and is also attractive due to its noninvasive nature. However, experts can misdiagnose a significant proportion of the cases, which may either cause redundant examinations or cancer. In order to reduce false positive and negative rates of mammography screening, computer-aided breast cancer detection has been studied for more than 30 years and many methods have been proposed by the researchers. In this review, region of interest (ROI) classification methods, which operate on a predefined or segmented ROIs with a focus on mass classification are surveyed. A total of 72 high quality journal and conference papers are selected from the Web of Science (WOS) database that meet several inclusion criteria. A comparative analysis is provided based on ROI extraction methods, data sets and machine learning techniques employed, the prediction accuracies, and usage frequency statistics. Based on the performances obtained on publicly available data sets, the ROI classification problem from mammogram images can be considered as approaching to be solved. Nonetheless, it can still be used as complementary information in breast cancer detection from the whole mammograms, which has room for improvement.

This article is categorized under:
Application Areas > Science and Technology
Technologies > Machine Learning
Technologies > Classification

KEYWORDS

breast cancer, computer-aided diagnosis, deep learning, mammogram, region of interest

1 | INTRODUCTION

Breast cancer is the most common cancer type among women. According to World Health Organization (WHO), 627,000 women died because of this disease, which constitutes 15% of the cancer-related losses in 2018 (WHO, 2019). Moreover, it is estimated that 268,600 new cases of breast cancer will arise in 2019 making breast cancer still the most prevalent cancer type with a rate of 30% (American Cancer Association, 2019).

Early diagnosis is vitally important to develop an effective treatment strategy that will potentially reduce the mortality rate of a cancer. Various tests can be considered to diagnose breast cancer including breast exams, breast ultrasound, mammogram, biopsy, and magnetic resonance imaging. Though biopsy is the only diagnostic procedure that can definitely identify whether a suspicious breast region has cancer or not, different tests may provide complementary information and may

be attractive due to their noninvasive nature. Among those, mammogram is an effective X-ray imaging technique that is used commonly to detect breast cancer at an early stage (American Cancer Association, 2019).

There are two types of mammographic abnormalities: masses and microcalcifications. Masses are usually described as undefined shapes and tartarated tissues (Figure 1a,b) while microcalcifications are formed by small calcium deposits (Figure 1c).

In clinical practice, a mass tissue can be categorized as benign or malignant by a radiologist. However, a significant proportion of the cases can also be misdiagnosed as revealed by subsequent biopsy tests (Bassett, Manjikian III, & Gold, 1990; Eltoukhy, Faye, & Samir, 2012). In order to reduce diagnostic error rates further, computer-aided detection (CAD) systems have been developed, which have the potential to detect, locate, and diagnose the cancer affected areas with high accuracy. It has been estimated that the sensitivity of detection without a CAD system is 80% while this value increases to 90% if a CAD system is employed during the process (Yassin, Omran, El Houby, & Allam, 2018). There are four main stages in a CAD system: preprocessing, region of interest (ROI) extraction, feature extraction, and classification. During the preprocessing stage, noise and artifacts are removed from the mammogram. This is followed by ROI extraction, which selects a subset of image pixels that characterize a potential cancer tissue. In the next step, a representative set of features are computed using the pixel intensities in ROI. Finally, a machine learning algorithm classifies the ROI into one of the disease states (e.g., cancer vs. normal). Automatic detection and classification of the ROI can help to confine the locations of the abnormal tissues to facilitate the decision making process (Yengec Tasdemir, Tasdemir, & Aydin, 2018). Nevertheless, some CAD systems may skip the ROI classification step and use the whole mammogram image to detect cancer directly. This is also true for recently developed deep learning systems, which combine feature extraction and classification steps into a single model that can compute informative features at various levels of abstraction.

The contribution of this article is to provide a comprehensive and systematic review of the state-of-the-art in ROI classification on mammogram images. The organization of the article is as follows: Section 2 summarizes the criteria used to collect the relevant publications; Section 3 includes data sets and methodologies used to assess the performance of ROI classification systems; Section 4 presents the accuracy of the best-performing methods; Section 5 explains the challenges faced during the review phase, makes links with breast cancer detection from full mammograms and evaluates the current status of ROI classification; finally, Section 6 concludes by providing suggestions for future research directions.

2 | RESEARCH METHOD

In this article, the following research questions are considered:

- Which data sets are used in ROI classification?
- Which techniques are used to extract ROIs from mammograms?

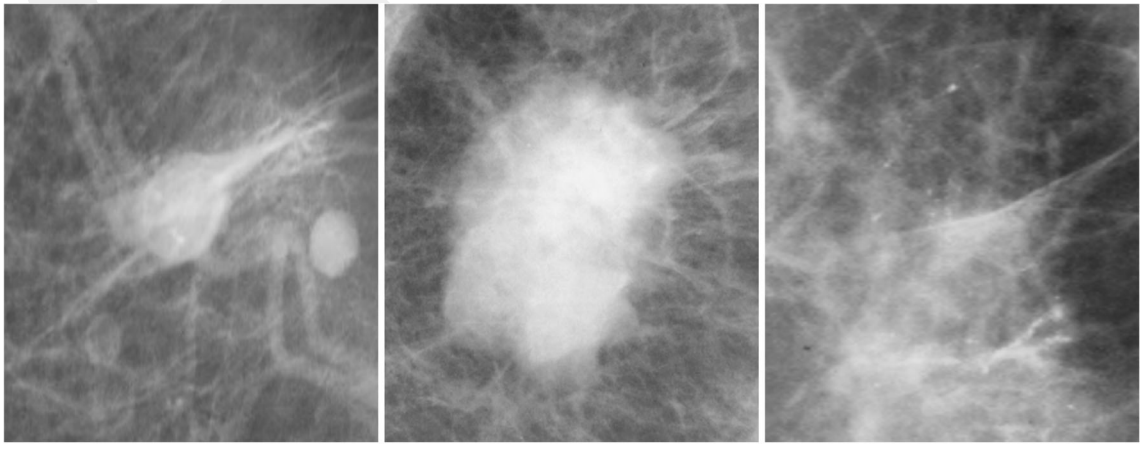

FIGURE 1 Mammographic abnormalities from the MIAS data set (Székely, Tóth, & Pataki, 2006). (a) Circumscribed lesion. (b) Speculated mass. (c) Micro-calcification. MIAS, mammography image analysis society

- Which features are employed as inputs to classification methods?
- Which methods are used to classify ROIs?
- Which metrics are used to assess the performance of CAD systems?

To address these questions, the Web of Science (WOS) database is searched for the following string: "(ALL = ((breast AND cancer) AND (mammogram) AND (ROI OR Region of interest) AND (classification OR machine OR learning OR classifier OR learning) AND (computer-aided diagnosis OR computer-aided detection OR computer assisted diagnosis OR cad))) AND LANGUAGE: (English)". Furthermore, publications in the time interval between 1975 to March 2019 are selected and the following inclusion criteria are applied: (a) journal and conference papers, excluding the posters, reviews, and surveys, (b) the quartile of the publisher is between Q1 and Q2, (c) papers that employ machine learning techniques for classification, (d) papers that use mammograms as input to the CAD system, (e) papers that are cited at least once in the last 2 years. Furthermore, some of the relevant papers which are not written in English are excluded. A total of 249 publications are retrieved by the initial search phrase and after applying the inclusion criteria, a refined set of 72 papers are retained.

There are two reasons for using publications in WOS. First, the papers in WOS are mainly published in journals indexed by Science Citation Index which contributes positively to the quality. Second, if all the papers in other databases had also been retrieved, the number of publications that needs to be reviewed would be much higher, which would make it difficult to read and analyze them one by one. For instance, using the same search keywords Google Scholar was able to find 3,300 publications while the IEEE database contained approximately 1,400. On the other hand, concentrating on publications in WOS still gives a clear picture of the current state-of-the-art in mammographic ROI classification. Based on these reasons, other databases are not considered in this work.

3 | SYSTEMATIC ANALYSIS

This section describes the problem definitions, data sets, methods employed in various components of a CAD system, and their preference statistics among the papers reviewed for ROI classification.

3.1 | Problem definitions

Different definitions of the ROI classification problem has been considered in the literature depending on the clinical objective and the scope of the project. While some of the papers aim to identify whether a ROI is mass or not (50 papers out of 72, 69.4%) while others categorize a ROI as normal, malignant or benign (13 papers out of 72, 18.1%). Furthermore, there are studies that concentrate on binary classification of the microcalcification clusters (9 out of 72 paper, 12.5%) in a given ROI.

3.2 | Data sets

Figure 2 illustrates the usage frequency of the data sets developed for assessing the performance of breast cancer and ROI classification methods. Digital database for screening mammography (DDSM) and mammography image analysis society (MIAS) are the most popular data sets that are publicly available. There are also a considerable number of papers that worked on private data sets.

Among the set of 72 publications selected for this review, DDSM and MIAS have been employed 31 (43.06%) and 18 times (25%), respectively. Twenty-nine (40%) publications adopted private data sets such as Nijmegen and Alberta Program (Test, 2004). Furthermore, 14 (19%) studies used hybrid data sets, which are formed by combining multiple data sets and help to evaluate the feasibility and adaptability of the methods. Table 1 presents a comparison among commonly employed databases.

3.2.1 | Digital database for screening mammography

Digital database for screening mammography(DDSM) is a mammographic database that consists of 2,620 film studies. The retrieval process of the mammograms is an effort between Massachusetts General Hospital, Sandia National

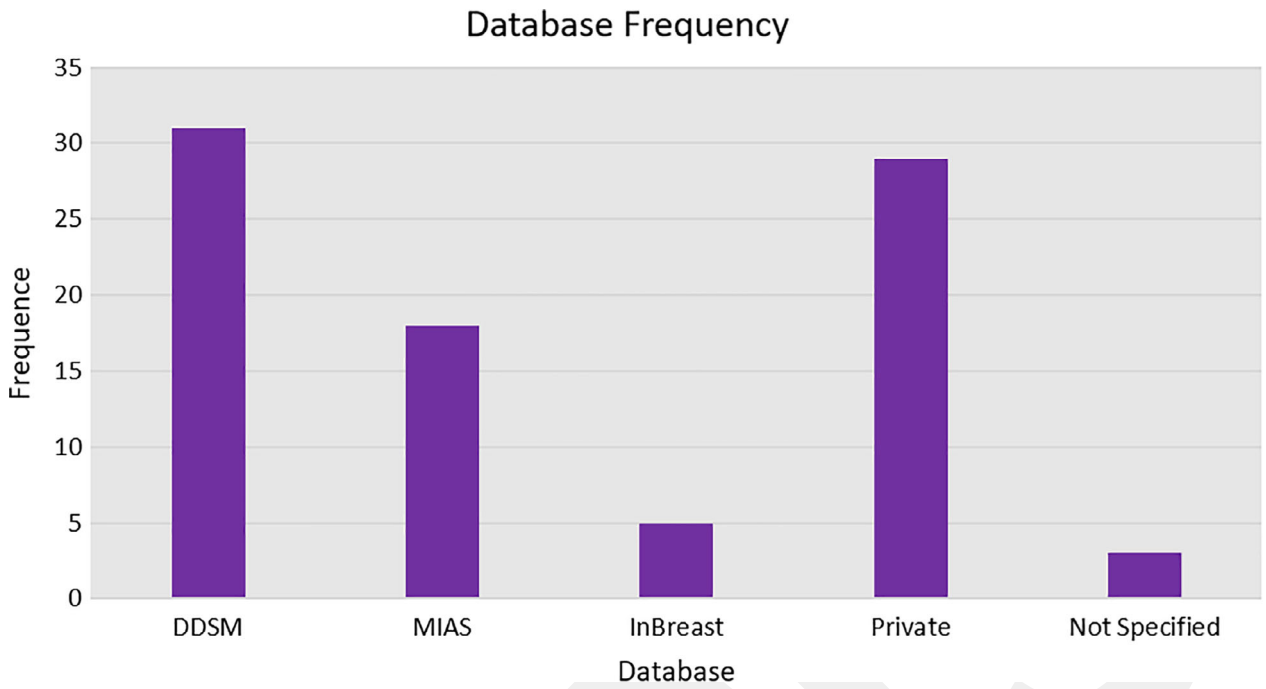

FIGURE 2 Usage frequency of the data sets developed for ROI classification. ROI, region of interest

TABLE 1 Comparison of databases commonly employed in breast cancer research

Quality—database	DDSM	MIAS	INBreast	Nijmegen	Alberta Program
Number of mammograms	10,239	322	410	21	158
Image file format	DICOM	PGM	DICOM	Unknown	Unknown
Size of images (in pixels)	1024 × 1024	1024 × 1024	328 × 4084 2560 × 3328	Unknown	Unknown
Lesion type	All kind	All kind	All kind	Micro calcifications	Architectural distortions
Views	MLO, CC	MLO	MLO, CC	MLO, CC	Unknown
Public access	Available	Available	Available	Not available	Not available

Abbreviations: DDSM, digital database for screening mammography; MIAS, mammography image analysis society.

Laboratories and the University of South Florida Computer Science and Engineering Department. The database contains normal, benign, and malignant cases with verified pathology information. Accessible image format of the database is DICOM. Furthermore, in addition to the full mammograms, whose sizes are 1024 × 1024 pixels, database also contains ROI information (ground truth data and ROI images) (Lee et al., 2017).

3.2.2 | Mammographic image analysis society

Mammographic image analysis society (MIAS) is an UK based organization. The mammographic films are taken from UK National Breast Screening Program. MIAS contains 322 digitized mammograms whose sizes are 1024 × 1024 pixels in PGM format. Film category of the mammograms are as follows: calcifications (benign and malignant), circumscribed masses (B, M), spiculated masses (B, M), architectural distortions (B, M), asymmetries (B, M), miscellaneous (B, M), and normal. The pixel location of the abnormalities are specified by experts and are demonstrated in a file (Suckling J, 1994).

3.2.3 | InBreast

InBreast is a mammographic database which is constructed from mammograms that are retrieved from Centro Hospitalar de S. João [CHSJ], Breast Centre, Porto. It consists of 410 images in DICOM format with sizes 3328 × 4084 or 2560 × 3328 pixels. The mammographic instances have samples of normal mammograms, masses, asymmetries, calcifications, and architectural distortions. The annotation of the data is performed by experts and given to user in an XML format (Moreira et al., 2012).

3.3 | ROI extraction methods

The two main approaches considered to extract the ROIs are cropping and segmentation. 41 out of 72 (56.94%) of the publications cropped the mammographic ROIs using two different methods, either with the help of the ground truth data of the data set(s) or with the help of an expert radiologist who annotates the suspicious areas. On the other hand, 31 out of 72 (43.06%) researchers extracted the regions automatically using segmentation techniques such as model-based, region-based, or counter-based methods.

3.4 | Feature extraction and dimension reduction methods

The process of feature extraction projects the pixel data into another domain to summarize information contained in the image and to facilitate the decision making process. To date, different feature extraction techniques that compute textural, morphological, and geometrical features have been employed in breast cancer research with the majority of the papers (55.6%) utilizing textural features. The texture analysis has been made at different modalities such as gray level co-occurrence matrix (GLCM; Ancy & Nair, 2018; Ghongade & Wakde, 2017b; Kanadam & Chereddy, 2016; Ribeiro, Costa, Papa, & Romero, 2014) and gray level run-length matrix (GLRM; Mohanty, Senapati, Beberta, & Lenka, 2013). The rest of the papers employed other types of features including features extracted by deep convolutional neural networks (CNNs; Al-masni et al., 2018; Chan, Lo, Sahiner, Lam, & Helvie, 1995; Kooi & Karssemeijer, 2017; Ragab, Sharkas, Marshall, & Ren, 2019; Sun, Tseng, Zhang, & Qian, 2017; Wang et al., 2017). There are also studies (23.6%) that combine features extracted by different methodologies (Choi, Kim, Plataniotis, & Ro, 2014; Dhahbi et al., 2018; Karahaliou et al., 2007; Marrocco, Molinara, D'Elia, & Tortorella, 2010; Mohanty et al., 2013; Mughal, Sharif, & Muhammad, 2017; Sun et al., 2017; Yengec Tasdemir et al., 2018).

Ancy and Nair (2018) propose a method that applies curvelet denoising to the mammograms of MIAS and segments the ROIs by median filter, gray level threshold, and morphological segmentation. From segmented ROIs the GLCM features are constructed and classified by SVM with an accuracy of 81%. Ribeiro et al. (2014) also employ GLCM in which Haralick features are constructed from GLCM to form the feature set. In this work, a private database is used and the Optimum Path Forest (OPF) algorithm is employed with an accuracy of 80%. Kanadam and Chereddy (2016) compare the GLCM's performance to GLAM's. In this article, MIAS database is used and morphological operations are employed for denoising the images. Using an SVM classifier, an accuracy of 97.2% is achieved by GLCM, which is 7% better than the GLAM. Another region-based textural feature matrix GLRM is adapted by Mohanty et al. (2013) and fused with Haralick features to make benign–malignant discrimination of ROIs in DDSM databases. In this study, an accuracy of 97.6% is achieved by a decision tree classifier. Marrocco et al. (2010) make a microcalcification cluster detection on a private database. To achieve this, both statistical and textural features are extracted, fused, and fed into CNN which achieved an AUC score of 91%. Choi et al. (2014) adapt a method that uses texture, shape, intensity, and spiculation features of DDSM ROIs. An AUC of 90% is achieved by an Adaboost classifier. Yengec Tasdemir et al. (2018), make an abnormal and normal discrimination on a private challenge database. In this work, HOG and Haralacik features are extracted from the first level of Wavelet decomposed ROIs. An accuracy of 87.4% is achieved by a random forest classifier. Karahaliou et al. (2007) fuse GLCM, GLRM, and Law's features to make microcalcification cluster detection on DDSM, in which a *k*-nearest neighbor achieved an AUC of 96%. Dhahbi et al. (2018) make a mass–non-mass detection on DDSM, where the feature set is composed of Hilbert's image representation, coaxial rings image representation, Kolmogorov–Smirnov distance approach, maximum regions descriptors, forest fire modeling, GLCM, box-counting fractal dimension, segmentation-based fractal texture analysis, and four-level wavelet packet decomposition. Feature selection is employed and an accuracy of 81.09% is achieved by a random forest classifier.

The feature extraction process can be followed by a dimensionality reduction step, which reduces the number of features that will be employed by the subsequent machine learning classifier(s). Using a high number of features will increase the model training time, include noise into the decision making process and potentially increase the complexity of the classifier causing it to memorize train data while decreasing the predictive performance on new data (a condition known as over-fitting). The most widely preferred dimension reduction algorithms in ROI classification are step-wise feature selection (30%) and projection-based reduction methods (40%) such as principal component analysis (PCA; Mohanty, Rup, Dash, Majhi, & Swamy, 2019; Papadopoulos, Fotiadis, & Likas, 2002; Sun et al., 2017; Wang et al., 2017) or linear discriminant analysis (LDA; Mohanty et al., 2019; Ragab et al., 2019; Shi et al., 2007; Wu et al., 2007).

Wang et al. (2017) employ mini-MIAS database and extract weighted-type fractional Fourier transform (WFRFT) features. Since the number of features is considerably high, PCA is employed to reduce the dimension and an accuracy of 92.3% is obtained with Jaya-FNN. Papadopoulos et al. (2002) employ PCA selected statistical features and a hybrid classification scheme of a rule-based and a neural network sub-system on Nijmegen and MIAS databases. An AUC of 92% is achieved for the MIAS database. Wu et al. (2007) propose a method that employs features derived from the spatial gray level dependence (SLGD) to distinguish masses. LDA is used for attribute selection. Their method achieved a sensitivity of 85% on a private database using an LDA classifier. Shi et al. (2007) fused morphological, textural, and spiculation features. A step-wise LDA is used as the attribute selector and an LDA classifier for making decisions. An AUC of 84% is achieved on DDSM.

3.5 | Machine learning classifiers

Once the feature descriptors are computed they are fed as input to a machine learning classifier, which diagnoses the condition of the ROI. To assess the prediction accuracy of a classifier, typically a cross-validation experiment is performed as a standard and robust technique that includes repeated rounds of model training and testing. Among the publications considered, 20 out of 72 (27.8%) used *k*-fold cross-validation, 14 out of 72 (19.4%) performed a leave-one-out cross-validation, while the rest (52.8%) did not mention explicitly which cross-validation method is used.

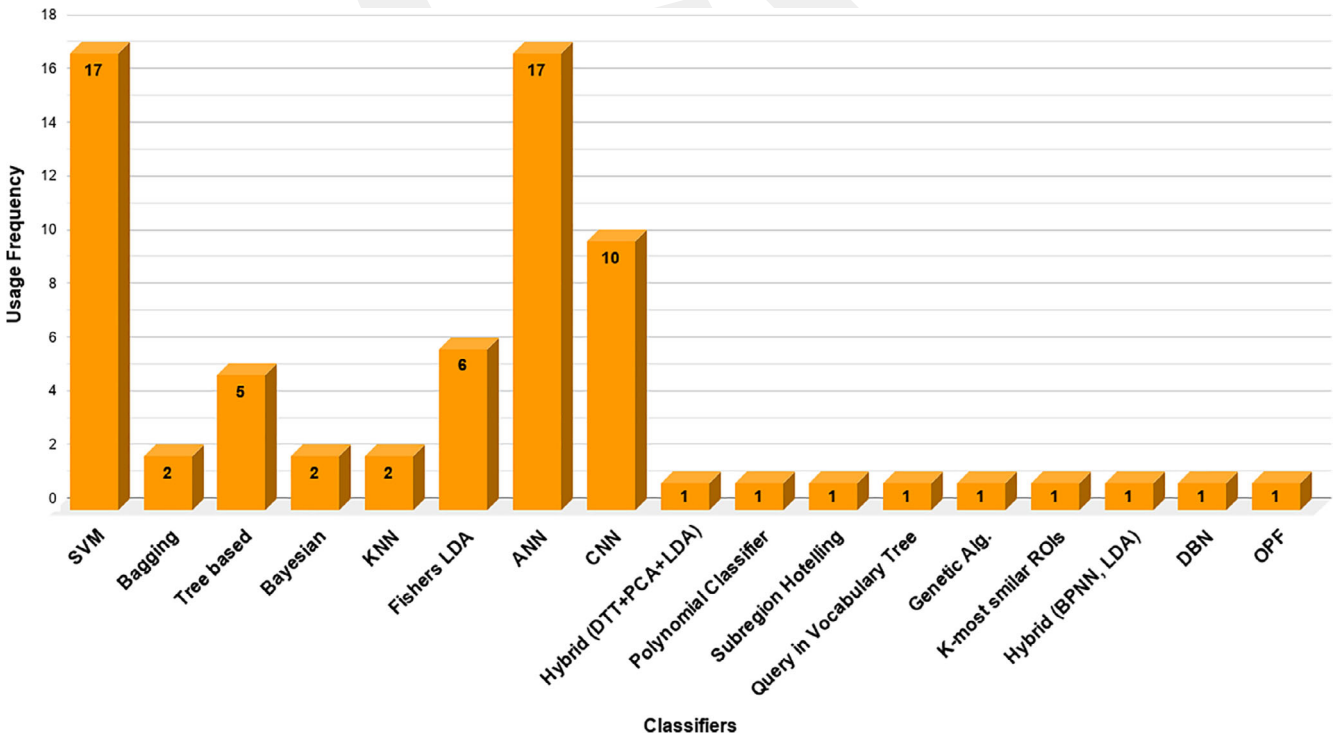

FIGURE 3 Usage frequency of the classifiers for ROI classification. ROI, region of interest

Figure 3 shows the machine learning methods commonly developed for ROI classification from mammogram images. Based on this analysis, the most frequently used models are support vector machines and artificial neural networks. Seventeen out of 72 papers (26.4%) employed a support vector machine (SVM) with different kernels and 17 papers (20.6%) implemented different variations of artificial neural networks. Other methods implemented for ROI classification include deep CNNs, classifiers based on decision trees, and *k*-nearest neighbors.

In order to evaluate the prediction accuracy of the proposed classification methods, researchers employed different evaluation metrics such as the overall accuracy (28 papers out of 72, 38.9%), sensitivity (9 out of 72, 12.5%), and area under ROC curve (AUC) score (35 out of 72, 48.6%).

Deep learning is among the new trends in machine learning with applications in a broad range of topics that brought significant advances in complex and high dimensional problems. Recently, there is an exponentially growing interest for deep learning applications in breast mass classification from mammogram images. Researchers developed various CNN structures both for feature extraction and classification of the mammographic ROIs (Al-masni et al., 2018; Chan et al., 1995; Chougrad, Zouaki, & Alheyane, 2018; Kooi & Karssemeijer, 2017) while others use the CNN structures to extract feature descriptors only (Ragab et al., 2019; Wang et al., 2017) or as a classifier only (Prathibha & Mohan, 2018; Shih-Chung B Lo, Huai Li, Yue Wang, Kinnard, & Freedman, 2002; Sun et al., 2017). Deep learning methods operate particularly well when sufficient number of data samples are available for model training. When this condition is not met, alternative strategies can be considered such as transfer learning (Chougrad et al., 2018; Ragab et al., 2019) and data augmentation, which enlarges the train set by producing new data samples through various transformations (Al-masni et al., 2018; Chougrad et al., 2018; Kooi & Karssemeijer, 2017). A detailed description and comparison of those schemes are available in Section 4.1.

4 | RESULTS

In this section, selected papers are categorized and analyzed in detail with respect to various criteria. Tables 2–5 include the citation information, data set used, the type of the problem studied, ROI extraction method, classification methods developed, and two accuracy metrics (AUC and overall accuracy [Acc]) of a selected subset of the 72 papers considered in this review. The rows are sorted with respect to the level of accuracy achieved. For the papers that employed more

TABLE 2 Summary of ROI classification results on DDSM data set

Reference	Classes	ROI extraction	Classifier	Result
Mohanty et al. (2019)	Mass, non-mass; benign, malignant	Cropped	Hybrid (DTT + PCA + LDA + IGWO-ELM)	AUC = 0.993, Acc = 99.50%
Hussain (2014)	Mass, non-mass	Cropped	SVM	Sn = 0.98, AUC = 0.99
Karahaliou et al. (2008)	Microcalcification	Segmented	Probabilistic NN	Sp = 0.913, AUC = 0.989
Reyad et al. (2014)	Mass, non-mass	Cropped	SVM	Sn = 0.982, Acc = 98.63%
Shen-Chuan Tai et al. (2014)	Mass, non-mass	Segmented	LDA	Sn = 0.9030, AUC = 0.98
do Nascimento et al. (2013)	Mass, non-mass	Cropped	Polynomial classifier	AUC = 0.98
Soares Sérvulo de Oliveira et al. (2015)	Mass, non-mass	Cropped	SVM (RBF kernel)	Sn = 0.986, Acc = 98.88%
Mohanty et al. (2013)	Mass, non-mass	Cropped	Decision tree	AUC = 0.99, Acc = 97.6%
Al-masni et al. (2018)	Benign, malignant	Cropped	CNN (YOLO)	AUC = 0.96, Acc = 97%
Karahaliou et al. (2007)	Microcalcifications	Cropped	KNN	Sp = 0.80, AUC = 0.96

Abbreviations: AUC, area under ROC curve; DDSM, digital database for screening mammography; DTT, discrete Tchebichef transform; IGWO-ELM, improved grey wolf optimization algorithm and extreme learning machine; LDA, linear discriminant analysis; PCA, principal component analysis; ROI, region of interest; SVM, support vector machine.

TABLE 3 Summary of ROI classification results on MIAS data set

Reference	Classes	ROI extraction	Classifier	Result
Ghongade and Wakde (2017b)	Normal, benign, malignant	Segmented	RF, RF-ELM	Sn = 0.97, Acc = 98%
Kim et al. (2012)	Mass, non-mass	Segmented	SVM	AUC = 0.98
Ghongade and Wakde (2017a)	Mass, non-mass	Segmented	Random Forest	Sp = 0.981, AUC = 97.28%
Kanadam and Chereddy (2016)	Mass, non-mass	Segmented	SVM	Sn = 0.952, Acc = 97.2%
Tahmasbi et al. (2011)	Mass, non-mass	Cropped	MLP	FPR = 0.055, AUC = 0.97
Biswas et al. (2017)	Mass, non-mass	Cropped	KNN	Sp = 0.90, Acc = 95%
Wang et al. (2017)	Mass, non-mass	Cropped	Jaya-FNN	Sn = 0.923, Acc = 92.27%

Abbreviations: AUC, area under ROC curve; MIAS, mammography image analysis society; MLP, multilayer perceptron; RF-ELM, random forest and extreme learning machine; ROI, region of interest; SVM, support vector machine.

than one machine learning classifier, only the method that achieves the best prediction accuracy is included. Table S1 shows information and accuracy results of all the 72 publications.

It can be seen from Table S1 that the DDSM data set is frequently employed for ROI classification. In Table 2 the best accuracy score of 99.5% is achieved by Mohanty et al. (2019) who worked on mass versus non-mass and benign versus malignant classification of mammogram images. They extracted features using discrete Tchebichef transform (DTT) from ROIs. The dimensionality of the feature space is reduced by a combination of PCA and LDA algorithms. This is followed by a hybrid classification model that employs an extreme learning machine (ELM) and an improved grey wolf optimization algorithm (IGWO-ELM). The best accuracy is achieved for the mass versus non-mass classification using a five fold cross-validation experiment. In another study on DDSM Hussain (2014) received an AUC score of 99.0% when statistical and LBP features are fused and fed to an SVM classifier. Soares Sérvulo de Oliveira et al. (2015) used taxonomic features to describe the texture of the ROIs and accomplished an overall accuracy score of 98.88% by an SVM. Karahaliou et al. (2008) fused gray level first-order statistics, Wavelet, GLCM, and Law's features. A probabilistic artificial neural network is employed to distinguish the microcalcifications and an AUC of 98% is obtained.

When the other top performing methods that employed DDSM are also considered it can be stated that textural features constitute the most frequently used feature category that achieves high accuracy scores. An AUC score of 98% is obtained when Nascimento et al. (2013) used different types of wavelet algorithms and compared them using a polynomial classifier algorithm. In this work, the best score is reached by Daubechies 8 function of Wavelet. In Shen-Chuan Tai et al. (2014), GLCM attained its best AUC of 0.98 with LDA classifiers on mass classification task on cropped ROIs. It is followed by Mohanty et al. (2013), in which GLCM and GLRM features are employed to distinguish benign–malignant masses from a ROI patch using a decision tree algorithm with an AUC of 99%. Law's features are also commonly employed and achieved the best AUC of 99% in Hussain (2014). In this work, since the number of features is high, feature selection is performed based on the significance measure. The selected features are then fed into an SVM classifier. The highest accuracy score of LBP features is 98.63%, which is accomplished by Reyad et al. (2014) when SVM is employed as the classifier.

For the MIAS data set, successful methods mainly employ texture-based features. In Table 3, the best accuracy is obtained by Ghongade and Wakde (2017b) with an overall accuracy score of 98%. They extracted textural features from ROIs using GLCM and performed feature selection using the correlation based filter algorithm. This is followed by a hybrid classifier that employs random forest and extreme learning machine (RF-ELM). In the paper by Kim et al. (2012), region-based satellite features are extracted from ROIs and are classified by an SVM. This approach obtained an AUC of 98%. Ghongade and Wakde (2017a) extracted GLCM features, which are selected using a correlation based attribute selection technique. The derived features are classified by an SVM with an AUC of 97.28%. In Kanadam and Chereddy (2016), GLAM features are employed and classified by an SVM with an accuracy of 97.2%. Tahmasbi et al. (2011) extracted Zernike features, which are classified by a multilayer perceptron (MLP) reaching an AUC of 97%.

In addition to publicly available data sets, some studies employed private data sets as summarized in Table 4 (Nugroho et al., 2014; Wei et al., 1997; Youssef et al., 2017) while others used hybrid data sets, which are formed by combining multiple sets so that more generic and robust classification models can be trained (Chougrad et al., 2018; Mughal et al., 2017; Narváez, Díaz, Poveda, & Romero, 2017; Wajid & Hussain, 2015).

TABLE 4 Summary of ROI classification results on private data sets

Reference	Classes	ROI extraction	Classifier	Result
Wei et al. (1997)	Mass, non-mass	Cropped	LDA	AUC = 0.97
Nugroho et al. (2014)	Mass, non-mass	Segmented	MLP	Sn = 0.967, Acc = 96.66%
Gundreddy et al. (2015)	Mass, non-mass	Cropped	K-most smilar ROIs	Sn = 0.712, AUC = 0.96
Youssef et al. (2017)	Benign, malignant	Segmented	SVM	Acc = 96%
Muramatsu et al. (2016)	Benign, malignant	Cropped	ANN	AUC = 0.9
Kooi and Karssemeijer (2017)	Mass, non-mass	Segmented	CNN	AUC = 0.895
Yengec Tasdemir et al. (2018)	Mass, non-mass	Cropped	Random Forest	AUC = 0.84, Acc = 87.5%

Abbreviations: AUC, area under ROC curve; LDA, linear discriminant analysis; MLP, multilayer perceptron; ROI, region of interest; SVM, support vector machine.

TABLE 5 Summary of ROI classification results on hybrid data sets

Reference	Data set	Classes	ROI extraction	Classifier	Result
Wajid and Hussain (2015)	MIAS + InBreast	Benign, malignant	Cropped	SVM (linear kernel)	AUC = 0.99, Acc = 99.73%
Chougrad et al. (2018)	DDSM + InBreast + Private	Benign, malignant	Cropped	CNN (Inception V3)	AUC = 0.99, Acc = 98.94%
Mughal et al. (2017)	MIAS + DDSM	Mass, non-mass	Segmented	SVM	Sp = 0.97, Acc = 97.5%
Narváez et al. (2017)	DDSM + InBreast	Mass, non-mass	Cropped	SVM	AUC = 0.94, Acc = 96.8
Abdel-Nasser et al. (2015)	MIAS + InBreast	Mass, non-mass	Cropped	Linear SVM	AUC = 0.93, Acc = 99.0%
Papadopoulos et al. (2002)	MIAS + Private	Microcalcification	Segmented	ANN (two hidden layer)	Sn = 0.90, AUC = 0.92
Abbas (2016)	MIAS + DDSM	Benign, malignant	Cropped	DBN	Sp = 0.842, AUC = 0.91
Shih-Chung B Lo et al. (2002)	MIAS + Private	Mass, non-mass	Segmented	CNN (multiple circular path)	AUC = 0.89
Ragab et al. (2019)	Cbis-DDSM + DDSM	Benign, malignant	Cropped	CNN (Alexnet) with transfer learning, SVM	AUC = 0.88, Acc = 80.9%

Abbreviations: AUC, area under ROC curve; DBM, deep belief network; DDSM, digital database for screening mammography; MIAS, mammography image analysis society; ROI, region of interest; SVM, support vector machine.

Among studies on private data sets, Wei et al. (1997) obtained an AUC score of 97.0% using global and local descriptors computed from wavelet decomposed ROIs and an LDA classifier. Nugroho et al. (2014) employed a method that uses 40 mammogram images and extracted textural features such as GLCM from segmented ROIs, in which the correlation based attribute selection (CFS) method is employed and features are fed into an MLP classifier that reached an accuracy score of 96.66%. Youssef et al. (2017) developed a model that discriminates the benign and malignant tumors where local descriptors of the segmented ROIs are used as input to an SVM providing an accuracy of 96%. It should be noted that each of these results are achieved on a different and a non-public data set.

Regarding the hybrid data sets, as listed in Table 5 the works of Wajid and Hussain (2015) and Chougrad et al. (2018) stand out, which obtained relatively higher prediction accuracy scores. Wajid and Hussain (2015) combined

MIAS and InBreast data sets and proposed a novel local energy-based shape histogram as the feature set, which is fed into an SVM classifier achieving an accuracy of 99.73%. The best method that uses both hybrid data set and standard machine learning technique (MLT) is proposed by Mughal et al. (2017). In this work, mass vs. non-mass discrimination is performed using an SVM classifier with a feature space that contains texture-based features, intensity and shape or morphological features. Narváez et al. (2017) fused Zernike and Curvelet features for mass vs. non-mass discrimination and obtained an accuracy score of 96.8% by an SVM classifier.

Among the studies that use standard MLTs as classifiers, the best AUC of 99.3% is achieved by Mohanty et al. (2019) using a hybrid classifier on DDSM data set. This is followed by Hussain (2014) and Wajid and Hussain (2015) with an AUC of 99%. Both of the methods employed various types of textural features and an SVM classifier.

Comparing different feature extraction techniques texture-based features achieved the best results based on the experiments performed by standard MLTs on different data sets. Since the mass changes the texture of the ROIs this finding is coherent. Furthermore, SVM is commonly employed by the best-performing methods. Based on the accuracy results, it can be said that it is suitable to use SVM as the standard MLT.

4.1 | Recent results by deep learning

Due to the growing interest on deep learning, this section is devoted to results obtained by deep learning methods. Chougrad et al. (2018) (Table 6) combined DDSM, InBreast, and a Private data set and employed data augmentation and global contrast normalization as preprocessing techniques. Their study is an example that uses a hybrid data set for model development and evaluation. They employed the following deep learning methods with transfer learning: VGG16, ResNet50, and Inception V3. In VGG16, a convolutional block that contains three convolutional operations are followed by an activation and a pooling layer with a total number of 23 layers. In ResNet50, the convolutional block is a residual block while for Inception V3 it is an inception module. A stochastic gradient descent (SGD) algorithm with a small learning rate is used for model training and fine tuning. To avoid over fitting, data augmentation, L2 regularization, and dropout regularization are performed. The output layer contains a softmax layer for final classification. The best accuracy of 98.94% for benign versus malignant classification is achieved when Inception V3 is employed.

TABLE 6 Results of ROI classification using deep learning approaches

Reference	Data set	Classes	ROI extraction	Classifier	Result
Chougrad et al. (2018)	DDSM + InBreast + Private	Benign, malignant	Cropped	CNN (Inception v3)	AUC = 0.99, Acc = 98.94%
Al-masni et al. (2018)	DDSM	Benign, malignant	Cropped	CNN (YOLO)	AUC = 0.96, Acc = 97%
Ragab et al. (2019)	Cbis-DDSM + DDSM	Benign, malignant	Cropped	CNN (Alexnet) with transfer learning, SVM	Acc = 87.2, AUC = 0.94
Marrocco et al. (2010)	Private	Microcalcification	Segmented	CNN	TP = 0.98, AUC = 0.91
Abbas (2016)	MIAS + DDSM	Benign, malignant	Cropped	DBN	Sp = 0.842, AUC = 0.91
Kooi and Karssemeijer (2017)	Private	Mass, non-mass	Segmented	CNN	AUC = 0.895
Shih-Chung B Lo et al. (2002)	MIAS + Private	Mass, non-mass	Segmented	CNN (multiple circular path)	AUC = 0.89
Prathibha and Mohan (2018)	DDSM	Benign, malignant	Cropped	CNN—orthogonal Ripplet Type II	Acc = 85.4%
Sun et al. (2017)	Private	Mass, non-mass	Cropped	SSl-CNN	AUC = 0.881, Acc = 82.43%

Abbreviations: AUC, area under ROC curve; DDSM, digital database for screening mammography; MIAS, mammography image analysis society; ROI, region of interest; SVM, support vector machine.

In a second study, Al-masni et al. (2018) (Table 6) adapted a YOLO based CNN, which has 24 convolutional layers with different kernel sizes, max-pooling layers with a window size of 2×2 followed by two fully connected layers. As the activation function, leaky rectified linear unit is employed. The hyper-parameters are optimized using the training data set, which constitutes 80% of the data set. The testing evaluations are performed by five fold cross-validation. The system detects the mass location with 99.7% accuracy and can discriminate benign cases from malignants with an accuracy of 97%.

Abbas (2016) proposed a novel CAD system which distinguishes malignant masses from benign masses (Table 6). The system combines hand crafted features with deep learning and makes the final decision by a softmax layer. The speed-up robust features (SURF) and local binary pattern variance features (LBPV) are extracted from ROI database which is constructed from a set of 300 benign and 300 malignant masses by employing both DDSM and mini-MIAS databases. In the second step, the features are transformed into invariant features using *k*-means clustering and probability density function. Afterwards, in order to construct deep invariant features, invariant features are fed into multilayer deep learning neural network which contains deep belief network (DBN) and a restricted Boltzmann machine (RBM). Fine tuning of hyper-parameters is performed on the training set, which constitutes 60% of the database. The system is compared to two different CNN algorithms in literature. The proposed DeepCAD algorithm is moderately higher in the AUC score (93%) than the other CNN structures (80%, 70%) which use CNN both as a feature extractor and a classifier.

Kooi and Karssemeijer (2017) employed a VGG like CNN as the classifier and feature extractor. In this work, both CC and MLO images are used from 18,366 women, and an AUC score of 89.5% is achieved. Shih-Chung B Lo et al. (2002) employed hand crafted features from 144 mammograms and used multiple circular path CNN structure as the classifier. They obtained an AUC score of 89%. Ragab et al. (2019) employed a pretrained AlexNet structure as feature extractor and imparted an SVM classifier to the end of the last fully connected layer. In this work, SVM outperformed and gave the accuracy of 87.2% while AlexNet gave an accuracy of 71% when it is used as the classifier. Prathibha and Mohan (2018) employed different kinds of CNN architectures such as CNN-Bandelet, CNN-ORT II, CNN-Tetrolet on DDSM data base and achieved accuracies of 84.6, 85.4, and 80.96%, respectively. Sun et al. (2017) presented a graph based semi-supervised learning using deep CNN on a private database which consists of 3,158 regions of interests (1,434 malignant, 1,724 benign). Twenty-one hand crafted features are extracted for each ROI. Dimensionality reduction algorithm (MDS) is implemented. Furthermore, CNN results are compared with SVM and artificial neural network (ANN). The CNN gave the AUC of 88.18%, while ANN and SVM gave AUC of 83%, 85.35%, respectively.

Among the studies listed above, deep learning approaches are used as feature extractor, classifier, or both. For example, Chougrad et al. (2018) showed that the accuracy improves when pretrained Inception V3 is used both as a feature extractor and a classifier on a hybrid data base. In that work, benign and malignant classification is performed on mammographic ROIs. Al-masni et al. (2018) used YOLO based CNN for both mass localization (accuracy of 99.7%) and benign-malignant classification (accuracy of 97%). In Chougrad et al. (2018) CAD system worked better than the other studies when accuracy is the baseline in the benign-malignant discrimination. Al-masni et al. (2018) is the only work which makes mass localization using deep learning in this review with an accuracy of 99.7%. Abbas (2016), Shih-Chung B Lo et al. (2002), and Sun et al. (2017) used deep learning approaches as classifier. In those works, mostly texture-based features are extracted and are fed into different types of CNNs. The maximum accuracy is achieved by the CAD system of Abbas (2016). In that study, well-known textural features in the literature are used by a DBN. Moreover, a hybrid database is employed, which made the model more generic. Kooi and Karssemeijer (2017) and Ragab et al. (2019) employed CNNs as feature extractor and used gradient boosted tree (GBT) and SVM classifiers, respectively. The CAD system of Ragab et al. (2019), which uses the most preferred MLT in the literature, performed better than Kooi and Karssemeijer (2017) with an AUC of 94%.

5 | DISCUSSION

Some challenges are encountered during the review phase. Most of them are related to the difficulty of standardization, which makes it difficult to compare the methods and/or the results achieved. This section will summarize these, make links to breast cancer detection from the whole mammogram, and summarize the current status of mammographic mass classification in ROIs.

5.1 | Difficulties in standardization

Most of the studies included in this review used only one data set some of which may be publicly unavailable. This makes it difficult to compare the proposed model to methods developed in other publications because a prediction model that performs well on one data set may not have the same success rates on another data set. Furthermore, each mammogram case (including its level of difficulty) can be different (Cheng et al., 2006). Therefore, the best approach would be to perform experiments on public and multiple data sets.

There are also other sources of incomparability. For instance, researchers used different metrics to assess the performance of CAD systems including AUC, overall accuracy, sensitivity, specificity, and FP rate. Another common problem is in some papers, the testing methodology is not clearly stated. Thirty-eight out of 72 (52.8%) papers did not even mention the experimental testing methodology that is used to assess the prediction accuracy of the CAD system, while others used methods such as k-fold cross-validation (CV), or leave-one-out cross-validation (LOOCV). It is well known in statistical machine learning that choosing a different k value as the number of cross-validation folds may inject variance into accuracy estimates of the classifiers.

5.2 | Relation to breast cancer detection from whole mammograms

The ultimate goal of mammography screening should be to identify whether a person has breast cancer or not. There is an ongoing research on automatic detection of breast cancer from the whole mammograms. Dhungel et al. (2017) employed a deep learning based method, achieving an AUC score of 0.74 in ROI classification. In the same study, they achieved an AUC score of 0.91 for breast cancer detection when they employed manually segmented regions, simultaneously. In two other studies, Zhou et al. (2017) achieved an accuracy of 60.9% while Zhang et al. (2018) used CNN with transfer learning and data augmentation achieving an AUC score of 0.73. As a result, it can be seen that because of the descriptors and the size of the mammographic image, it is more difficult to classify whether a full mammogram contains cancer or not as compared to whether a small patch of the mammogram contains cancer or not or it is a mass tissue or not (Zhang et al., 2018). Recently, an international competition has been organized titled Digital Mammography DREAM Challenge. Participants initially had access to a pilot data set that contains 500 images to prepare their source codes as Docker containers, which were later submitted to a computer cloud system that included a training set of more than 600,000 mammogram images each labeled as normal (label: 0) or cancer (label: 1). As a result of this challenge, Therapixel's best-performing model achieved an AUC of 0.85 (Gilbert, Le, Hickman, Wang, & Huang, 2019), which shows that breast cancer detection from mammograms still has significant room for improvement.

5.3 | Status of mass classification in ROIs

We can state that the mass classification problem in ROIs is approaching to be solved according to the performances achieved on public data sets though there may still be challenging cases. For instance, the 500 images of the pilot data set of the DREAM challenge contained 34 cancer positive and 466 cancer negative images. Recently, Yengec Tasdemir et al. (2018) developed and compared ROI classification methods on this pilot data set and achieved an accuracy rate of 87.5%. In this work, the ROI regions were labeled by an expert radiologist who is an Associate Professor in one of the established university hospitals in Turkey. However the radiologist was not able to locate the ROI in 6 of the 34 images due to the difficulty of those cases.

6 | CONCLUSION

This article reviews the current status of ROI classification in breast cancer research. It can be conducted from the review that most of the studies (49 out of 72 papers) employed public data sets due to their availability. Moreover, image pixel information is used with machine learning algorithms to classify whether a ROI is abnormal or normal.

In this review, the prediction methods are categorized as standard MLTs and deep learning techniques. The published studies are appropriate in terms of the data sets and methods employed. For instance, deep learning methods require large data sets and most of the publications that use deep learning employ combined data sets. Regarding

feature extraction techniques, texture-based features stand out as successful feature descriptors. This is also meaningful as mammogram images contain significant amount of textural information. Among the classification methods, SVM and ANN are extensively employed. This is followed by CNN, especially in recent years when the deep learning techniques have begun to be a trend. All three are powerful classification methods that are suitable for complex problems.

The standardization and the comparability of the papers were challenging. It was not easy to compare the performance of the proposed methods because different metrics are used in publications. Moreover, performance metrics are limited (i.e., some papers only present accuracy). As another source of incomparability most of the studies use only one data set which makes it unfavorable to compare with other methods in the literature.

As a future work, to further sharpen the prediction accuracy of the methods, the demographic information about the patient can also be included into the decision process. This may include the smoking status of the patient, age, whether the patient has children and whether his/her parents have breast cancer. As a second future direction, the ROI based predictions can be used as complementary information in breast cancer detection from the whole mammograms, which is still an open problem.

ACKNOWLEDGMENTS

Sena B. Yengec Tasdemir, is supported by the Turkish Higher Education Council's 100/2000 PhD fellowship program.

CONFLICT OF INTEREST

The authors have declared no conflicts of interest for this article.

AUTHOR CONTRIBUTIONS

Sena B. Yengec Tasdemir: Data curation; formal analysis; methodology; resources; and writing-original draft. **Kasim Tasdemir:** Formal analysis; project administration; resources; and supervision. **Zafer Aydin:** Project administration; supervision; and writing-original draft.

ORCID

Zafer Aydin https://orcid.org/0000-0001-7686-6298

RELATED WIREs ARTICLES

A novel comparative study on breast cancer detection using different types of classification techniques
Wavelet-based scaling indices for breast cancer diagnostics

REFERENCES

Abbas, Q. (2016). DeepCAD: A computer-aided diagnosis system for mammographic masses using deep invariant features. *Computers*, *5*, 28.

Abdel-Nasser, M., Rashwan, H. A., Puig, D., & Moreno, A. (2015). Analysis of tissue abnormality and breast density in mammographic images using a uniform local directional pattern. *Expert Systems with Applications*, *42*, 9499–9511.

Al-masni, M. A., Al-antari, M. A., Park, J.-M., Gi, G., Kim, T.-Y., Rivera, P., ... Kim, T.-S. (2018). Simultaneous detection and classification of breast masses in digital mammograms via a deep learning YOLO-based CAD system. *Computer Methods and Programs in Biomedicine*, *157*, 85–94.

American Cancer Association. (2019). Cancer Facts and Figures 2019. Technical Report. Retrieved from https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2019/cancer-facts-and-figures-2019.pdf

Ancy, C. A., & Nair, L. S. (2018). *An Efficient CAD for Detection of Tumour in Mammograms Using SVM*. Proceedings of the 2017 IEEE International Conference on Communication and Signal Processing, ICCSP 2017, vol. 2018-Janua, 1431–1435. IEEE.

Bassett, L. W., Manjikian, V., III, & Gold, R. H. (1990). Mammography and breast cancer screening. *Surgical Clinics of North America*, *70*, 775–798.

Biswas, R., Nath, A., & Roy, S. (2017). *Mammogram Classification Using Gray-Level Co-occurrence Matrix for Diagnosis of Breast Cancer*. In Proceedings - 2016 International Conference on Micro-Electronics and Telecommunication Engineering, ICMETE 2016, 161–166. IEEE.

Chan, H.-P., Lo, S.-C. B., Sahiner, B., Lam, K. L., & Helvie, M. A. (1995). Computer-aided detection of mammographic microcalcifications: Pattern recognition with an artificial neural network. *Medical Physics*, *22*, 1555–1567.

Cheng, H.-D., Shi, X., Min, R., Hu, L., Cai, X., & Du, H. (2006). Approaches for automated detection and classification of masses in mammograms. *Pattern Recognition*, *39*, 646–668.

Choi, J. Y., Kim, D. H., Plataniotis, K. N., & Ro, Y. M. (2014). Computer-aided detection (CAD) of breast masses in mammography: Combined detection and ensemble classification. *Physics in Medicine and Biology*, *59*, 3697–3719.

Chougrad, H., Zouaki, H., & Alheyane, O. (2018). Deep convolutional neural networks for breast cancer screening. *Computer Methods and Programs in Biomedicine*, *157*, 19–30.

Dhahbi, S., Barhoumi, W., Kurek, J., Swiderski, B., Kruk, M., & Zagrouba, E. (2018). False-positive reduction in computer-aided mass detection using mammographic texture analysis and classification. *Computer Methods and Programs in Biomedicine*, *160*, 75–83.

Dhungel, N., Carneiro, G., & Bradley, A. P. (2017). *Fully Automated Classification of Mammograms Using Deep Residual Neural Networks*. In 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), 310–314. IEEE.

Eltoukhy, M. M., Faye, I., & Samir, B. B. (2012). A statistical based feature extraction method for breast cancer diagnosis in digital mammogram using multiresolution representation. *Computers in Biology and Medicine*, *42*, 123–128.

Ghongade, R. D., & Wakde, D. G. (2017a). *Computer-Aided Diagnosis System for Breast Cancer Using RF Classifier*. In 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), 1068–1072. IEEE.

Ghongade, R. D., & Wakde, D. G. (2017b). *Detection and Classification of Breast Cancer from Digital Mammograms Using RF and RF-ELM algorithm*. In 2017 1st International Conference on Electronics, Materials Engineering and Nano-Technology, IEMENTech 2017, 1–6. IEEE.

Gilbert, F., Le, E., Hickman, S., Wang, Y., & Huang, Y. (2019). Artificial intelligence in breast imaging. *Clinical Radiology*, *74*(5), 357–366.

Gundreddy, R. R., Tan, M., Qiu, Y., Cheng, S., Liu, H., & Zheng, B. (2015). Assessment of performance and reproducibility of applying a content-based image retrieval scheme for classification of breast lesions. *Medical Physics*, *42*, 4241–4249.

Hussain, M. (2014). False-positive reduction in mammography using multiscale spatial Weber law descriptor and support vector machines. *Neural Computing and Applications*, *25*, 83–93.

Kanadam, K. P., & Chereddy, S. R. (2016). Mammogram classification using sparse-ROI: A novel representation to arbitrary shaped masses. *Expert Systems with Applications*, *57*, 204–213.

Karahaliou, A., Boniatis, I., Skiadopoulos, S., Sakellaropoulos, F., Arikidis, N., Likaki, E., ... Costaridou, L. (2008). Breast cancer diagnosis: Analyzing texture of tissue surrounding microcalcifications. *IEEE Transactions on Information Technology in Biomedicine*, *12*, 731–738.

Karahaliou, A., Skiadopoulos, S., Boniatis, I., Sakellaropoulos, P., Likaki, E., Panayiotakis, G., & Costaridou, L. (2007). Texture analysis of tissue surrounding microcalcifications on mammograms for breast cancer diagnosis. *The British Journal of Radiology*, *80*, 648–656.

Kim, D. H., Choi, J. Y., & Ro, Y. M. (2012). *Region Based Stellate Features for Classification of Mammographic Spiculated Lesions in Computer-Aided Detection*. In Proceedings - International Conference on Image Processing, ICIP, 2821–2824. IEEE.

Kooi, T., & Karssemeijer, N. (2017). Classifying symmetrical differences and temporal change in mammography using deep neural networks. *Journal of Medical Imaging*, *4*, 1.

Lee, R. S., Gimenez, F., Hoogi, A., Miyake, K. K., Gorovoy, M., & Rubin, D. L. (2017). A curated mammography data set for use in computer-aided detection and diagnosis research. *Scientific Data*, *4*, 170177.

Marrocco, C., Molinara, M., D'Elia, C., & Tortorella, F. (2010). A computer-aided detection system for clustered microcalcifications. *Artificial Intelligence in Medicine*, *50*, 23–32.

Mohanty, A. K., Senapati, M. R., Beberta, S., & Lenka, S. K. (2013). Texture-based features for classification of mammograms using decision tree. *Neural Computing and Applications*, *23*, 1011–1017.

Mohanty, F., Rup, S., Dash, B., Majhi, B., & Swamy, M. N. S. (2019). A computer-aided diagnosis system using Tchebichef features and improved grey wolf optimized extreme learning machine. *Applied Intelligence*, *49*, 983–1001.

Moreira, I. C., Amaral, I., Domingues, I., Cardoso, A., Cardoso, M. J., & Cardoso, J. S. (2012). Inbreast: Toward a full-field digital mammographic database. *Academic Radiology*, *19*, 236–248.

Mughal, B., Sharif, M., & Muhammad, N. (2017). Bi-model processing for early detection of breast tumor in CAD system. *European Physical Journal Plus*, *132*, 266.

Muramatsu, C., Hara, T., Endo, T., & Fujita, H. (2016). Breast mass classification on mammograms using radial local ternary patterns. *Computers in Biology and Medicine*, *72*, 43–53.

Narváez, F., Díaz, G., Poveda, C., & Romero, E. (2017). An automatic BI-RADS description of mammographic masses by fusing multiresolution features. *Expert Systems with Applications*, *74*, 82–95.

do Nascimento, M. Z., Martins, A. S., Neves, L. A., Ramos, R. P., Flores, E. L., & Carrijo, G. A. (2013). Classification of masses in mammographic image using wavelet domain features and polynomial classifier. *Expert Systems with Applications*, *40*, 6213–6221.

Nugroho, H. A., Faisal, N., Soesanti, I., & Choridah, L. (2014). Analysis of Computer Aided Diagnosis on Digital Mammogram Images. Retrieved from https://www.semanticscholar.org/paper/Analysis-of-Computer-Aided-Diagnosis-on-Digital-Nugroho-Faisal/ccbce161e75fb9309ca6c12a2652d35459bd466f

Papadopoulos, A., Fotiadis, D., & Likas, A. (2002). An automatic microcalcification detection system based on a hybrid neural network classifier. *Artificial Intelligence in Medicine*, *25*, 149–167.

Prathibha, G., & Mohan, B. C. (2018). *Mammograms Classification Using Multiresolution Transforms and Convolution Neural Networks*. In 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 1–7. IEEE.

Ragab, D. A., Sharkas, M., Marshall, S., & Ren, J. (2019). Breast cancer detection using deep convolutional neural networks and support vector machines. *PeerJ*, *7*, e6201.

Reyad, Y. A., Berbar, M. A., & Hussain, M. (2014). Comparison of statistical, LBP, and multi-resolution analysis features for breast mass classification. *Journal of Medical Systems*, *38*, 100.

Ribeiro, P. B., Da Costa, K. A., Papa, J. P. and Romero, R. A. (2014). *Optimum-Path Forest Applied for Breast Masses Classification*. In Proceedings - IEEE Symposium on Computer-Based Medical Systems, 52–55. IEEE.

Tai, S.-C., Chen, Z.-S., & Tsai, W.-T. (2014). An automatic mass detection system in mammograms based on complex texture features. *IEEE Journal of Biomedical and Health Informatics*, *18*, 618–627.

Shi, J., Sahiner, B., Chan, H.-P., Ge, J., Hadjiiski, L., Helvie, M. A., ... Cui, J. (2007). Characterization of mammographic masses based on level set segmentation with new image features and patient information. *Medical Physics*, *35*, 280–290.

Shih-Chung B Lo, S.-C. B., Huai Li, H., Yue Wang, Y., Kinnard, L., & Freedman, M. (2002). A multiple circular path convolution neural network system for detection of mammographic masses. *IEEE Transactions on Medical Imaging*, *21*, 150–158.

Soares Sérvulo de Oliveira, F., Oseas de Carvalho Filho, A., Corrêa Silva, A., Cardoso de Paiva, A., & Gattass, M. (2015). Classification of breast regions as mass and non-mass based on digital mammograms using taxonomic indexes and SVM. *Computers in Biology and Medicine*, *57*, 42–53.

Suckling J, P. (1994). The mammographic image analysis society digital mammogram database. *Digital Mammo*, 375–386.

Sun, W., Tseng, T. L. B., Zhang, J., & Qian, W. (2017). Enhancing deep convolutional neural network scheme for breast cancer diagnosis with unlabeled data. *Computerized Medical Imaging and Graphics*, *57*, 4–9.

Székely, N., Tóth, N., & Pataki, B. (2006). A hybrid system for detecting masses in mammographic images. *IEEE Transactions on Instrumentation and Measurement*, *55*, 944–952.

Tahmasbi, A., Saki, F., & Shokouhi, S. B. (2011). Classification of benign and malignant masses based on Zernike moments. *Computers in Biology and Medicine*, *41*, 726–735.

Test, S. (2004). Alberta program for the early detection of breast cancer—2001/03 biennial report.

Wajid, S. K., & Hussain, A. (2015). Local energy-based shape histogram feature extraction technique for breast cancer diagnosis. *Expert Systems with Applications*, *42*, 6990–6999.

Wang, S., Rao, R. V., Chen, P., Zhang, Y., Liu, A., & Wei, L. (2017). Abnormal breast detection in mammogram images by feed-forward neural network trained by Jaya algorithm. *Fundamenta Informaticae*, *151*, 191–211.

Wei, D., Chan, H.-P., Petrick, N., Sahiner, B., Helvie, M. A., Adler, D. D., & Goodsitt, M. M. (1997). False-positive reduction technique for detection of masses on digital mammograms: Global and local multiresolution texture analysis. *Medical Physics*, *24*, 903–914.

WHO. (2019). Breast cancer. Retrieved from https://www.who.int/cancer/prevention/diagnosis-screening/breast-cancer/en/

Wu, Y.-T., Wei, J., Hadjiiski, L. M., Sahiner, B., Zhou, C., Ge, J., ... Chan, H.-P. (2007). Bilateral analysis based false positive reduction for computer-aided mass detection. *Medical Physics*, *34*, 3334–3344.

Yassin, N. I., Omran, S., El Houby, E. M., & Allam, H. (2018). Machine learning techniques for breast cancer computer aided diagnosis using different image modalities: A systematic review. *Computer Methods and Programs in Biomedicine*, *156*, 25–45.

Yengec Tasdemir, S. B., Tasdemir, K., & Aydin, Z. (2018). *ROI Detection in Mammogram Images Using Wavelet-Based Haralick and Hog Features*. In 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), 105–109.

Youssef, Y. B., Abdelmounim, E., Zbitou, J., Elharoussi, M., & Najib Boujida, M. (2017). Comparison machine learning algorithms in abnormal mammograms classification. *International Journal of Computer Science and Network Security*, *17*, 19–25.

Zhang, X., Zhang, Y., Han, E. Y., Jacobs, N., Han, Q., Wang, X., & Liu, J. (2018). Classification of whole mammogram and tomosynthesis images using deep convolutional neural networks. *IEEE Transactions on Nanobioscience*, *17*, 237–242.

Zhou, H., Zaninovich, Y., & Gregory, C. (2017). Mammogram classification using convolutional neural networks. Retrieved from http://ehnree.github.io/documents/papers/mammogram_conv_net.pdf

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Yengec Tasdemir SB, Tasdemir K, Aydin Z. A review of mammographic region of interest classification. *WIREs Data Mining Knowl Discov*. 2020;10:e1357. https://doi.org/10.1002/widm.1357