Expert Systems With Applications 167 (2021) 114095

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Review

Medical image based breast cancer diagnosis: State of the art and future directions

Mehreen Tariq [a], Sajid Iqbal [a,*], Hareem Ayesha [a], Ishaq Abbas [a], Khawaja Tehseen Ahmad [a], Muhammad Farooq Khan Niazi [b]

[a] Department of Computer Science, Bahauddin Zakariya University, Multan, Pakistan
[b] Bakhtawar Amin Memorial Hospital, Multan, Pakistan

ARTICLE INFO

Keywords:
Breast cancer detection and diagnosis
Transfer learning (TL)
Deep learning (DL)
Machine learning (ML)
Computer aided diagnosis (CAD)

ABSTRACT

The intervention of medical imaging has significantly improved early diagnosis of breast cancer. Different radiological and microscopic imaging modalities are frequently utilized by medical practitioners for identification and categorization of different breast abnormalities by manual scrutiny. The meticulous classification of different breast abnormalities is challenging, because of ambiguous imaging data and due to indistinguishable characteristics of benign and malignant breast lesions. However, with the advent in applications of Artificial Intelligence (AI) in healthcare, researchers have turned their focus towards designing of efficient intelligent computer aided detection and diagnosis systems for prognosis of this catastrophic disease using image processing and computer vision (CV) techniques. An abundance of work could be found in literature on classification of different breast abnormalities, where majority of them has dealt with binary classification (i.e. benign and malignant). In current study, a comprehensive review has been presented to analyze and evaluate state of the art proposed methodologies for breast cancer diagnosis based over commonly used breast screening imaging modalities. The studies under consideration are mainly categorized into statistical machine learning based and deep learning based classifier, where deep classifiers further sub-categorized into models built from scratch and transfer learning based models. A number of factors have been taken to compare the performance of these classification models, on the basis of which some recommendations are provided for researcher to precede this work in future.

1. Introduction

Breast cancer is one of major health issues among women worldwide. It is the second most common type of cancer after lung cancer (Observatory, 2018). Recent studies show that about one in every 9 women becomes the patient of breast cancer (Begum, 2018; Menhas & Umer, 2015). In 2018, almost 2,088,849 (11.6% of total cancer cases) new cases of breast cancer are found around the world from which 626,679 patients died (IARC, 2018). Pakistan has the highest breast cancer rate in Asia (Bhurgri et al., 2000, 2006). (IARC, 2018) report that 34,066 new cases of breast cancer are found in women of Pakistan, which is about 36.8% of total cancer occurrences.

Breast cancer is caused due to the superfluous intensification of cells in breast, which give rise to lumps or tumors. A tumor could be benign or malignant: benign lumps are less likely to invade in its surroundings and are considered to be non-cancerous. Usually, if these lumps do not cause problems like breast pain or dissemination in neighboring tissues then they are not operated and left untreated. There exist different types of breast masses or benign lumps including cysts, fibroadenomas, phyllodes tumor, atypical hyperplasia, fat necrosis and adenosis (D'Orsi et al., 1994). Malignant tumors are cancerous and invasive. If these tumors are not detected and medicated at an earlier stage, they disseminate and destroy the encompassing breast tissues and may also cause metastatic breast cancer. Metastatic breast cancer is evolved when cells from primary breast lump are shattered to other body organs i.e. in liver, brain, bones, or lungs through bloodstream or lymphatic system.

Women breasts mainly incorporate glandular (for milk production) and fatty tissues, lobes (part where milk is produced) and ducts. Breast cancer can be originated in any of these breast constituents. Major types of invasive breast cancer are Ductal and Lobular carcinomas (Li, Anderson, Daling, & Moe, 2003). Women who are tribulated by breast cancer discern different variations in breast such as redness, swelling,

* Corresponding author.
E-mail addresses: sajid.iqbal@bzu.edu.pk (S. Iqbal), tehseen@bzu.edu.pk (K.T. Ahmad).

https://doi.org/10.1016/j.eswa.2020.114095
Received 26 June 2020; Received in revised form 2 October 2020; Accepted 3 October 2020
Available online 20 October 2020
0957-4174/© 2020 Elsevier Ltd. All rights reserved.

scaling, underarm lump, skin irritation, fluid discharge and deformation of breast. Breast cancer is primarily may occur in 5 stages (i.e. Stage 0 to stage IV), where stage 0 delineates non-invasive cancer and stage IV delineates invasive breast cancer. In Pakistan, more than 50% breast cancer patients consult to doctors at advanced stages such as at stage III or IV (Khokher, Qureshi, Mahmood, & Sadiq, 2016). Almost 90,000 new cases of these diseases are found every year in Pakistan, from which 40,000 patients die (Khan, 2009). The cause of increasing mortality rate due to this catastrophic disease is delayed presentation or delayed consultation with doctors, which is due to lack of awareness, lower level of education and prevailing poverty (Khan, Ahmad, Nadeem, & Hussain, 2016). If this disease is detected at an early stage the survival rate can be significantly improved and better treatment options can be availed (Lu et al., 2009). The mortality rate, in women above 50 years of age, can be prevented by $1/3^{rd}$, if routine mammography is done (Vogel, 1999).

Although there is no definite method of breast cancer prevention however the use of several manual and image-based tests can be helpful in detection and diagnosis. First and the most common type of test, which is recommended to women for early detection of this disease is self-scrutiny to savvy about the occurrences of freakish discrepancies in breast. For meticulous assessment and scrupulous diagnosis of breast lesions, radiologists use different imaging modalities such as X-Ray mammography, Ultrasound MRI, Thermography, CT-scan for screening breast cancer. Several breast abnormalities can be studied in breast cancer patients by analyzing these images. For example, in mammograms breast cancer can appear as: micro-calcifications, masses and architectural distortions, where during breast tissue components analysis through whole slide images (WSI) abnormalities that could be detected include: nuclei, epithelium, tubule, stroma and mitotic detection (Hamidinekoo, Denton, Rampun, Honnor, & Zwiggelaar, 2018). The abnormality that is most difficult to find in mammogram is architectural distortion, which is shattering of normal breast architecture without any presence of lump (Balleyguier et al., 2007). Expert radiologist often conflicts during the interpretation of medical breast images i. e. mammogram. To avoid this conflict and subjectivity of radiologists; American College of Radiology (ACR) has developed a standard lexicon namely Breast Imaging Reporting and Dated System (BIRADS) for interpretation and characterization of breast mammogram, ultrasound and MRI images (Magny et al., 2020).

Microscopic analysis of diseased area can accurately identify the breast lesion and its type (Zhang, Zhang, Coenen, & Lu, 2013). However, manual perusal of medical images is a prolonged and tedious task and the availability of highly trained and expert medical practitioners is crucial for it. For example, during visual inspection of breast, there is probability of subjectivity, excess time-consumption and ambiguities. There is plausibility of human misconception during vetting which may cause false positive (FP) or false negative (FN) diagnosis. Manual image analysis can help the practitioners to improve prognosis outcomes. However, technical issues such as inadequate image quality and noise can also raise miss-classification rate which may cause unnecessary biopsies. To overcome these shortcomings and for early prognosis of breast cancer computer aided diagnosis system (CADs) is a preferable alternative. These systems assist radiologists and other practitioners in discerning abnormalities and identifying suspicious features in medical images. Medical images contain a lot of fine-grained information, which can be extracted using CAD systems, by applying image processing and computer vision techniques. CADs assist medical practitioners in detection of cancer, in classifying tumors as benign or malignant and in categorizing tumors into different classes such as Ductal carcinoma in Situ, Invasive carcinoma, lobular carcinoma etc. Moreover, these systems save a lot of time of human experts that is wasted in manual scrutiny of medical images and also help in avoiding unnecessary biopsies. A general architecture of such a (supervised learning based) CAD system for breast cancer detection is depicted in Fig. 1.

Input dataPreprocessingROI extractionFeatures ExtractionClassificationA variety of challenges are faced during the development of such CADs that may include scarcity of enough, well annotated dataset of breast medical images, choice of imaging modality, selection of appropriate segmentation technique for highlighting tumor area from background, choice of features to be extracted from the images for their classification and picking the right classifier for the precise classification of these images. A number of surveys and reviews are done by different authors during past years for providing researchers with a comprehensive guide to tackle with these problems, such as (Cheng, Shan, Ju, Guo, & Zhang, 2010) conducted a survey which mainly focuses on summarizing the approaches for breast cancer detection and classification using breast ultrasound images, (Huang, Liu, Van Der Maaten, & Weinberger, 2017) have done a comprehensive survey to compare the existing segmentation techniques for breast ultrasound images, (Borchartt, Conci, Lima, Resmini, & Sanchez, 2013) present a review based on the recent works on breast disease detection and diagnosis using thermography images, (Cheng, Cai, Chen, Hu, & Lou, 2003) have done a survey on the techniques for the detection and classification of calcifications in mammograms, (Ramani, Valarmathy, & Suthanthira Vanitha, 2013) presents a survey on recent clustering based breast cancer detection techniques using mammograms and in (Nahid & Kong, 2018), the authors have presented a comprehensive review of state of the art breast cancer image classification procedures which recapitulates the available breast datasets, generalized image classification techniques, feature extraction and reduction techniques and the performance measuring criteria. All these studies have presented admirable reviews on the existing automated techniques for breast cancer detection and diagnosis using different types of medical images, but there exist limitations in these works such as, authors are limited to one or two imaging modalities (e.g. Mammography and ultrasound), some of the studies do not belongs to recent years, some studies have discussed only few factors that affect the classification accuracy of breast image classification etc. The motivation of our study is to present a holistic review to analyze and evaluate recently proposed methodologies for breast cancer detection using all the commonly used imaging modalities. The evaluation of these methodologies is done on a number of factors, such as techniques used for preprocessing, segmentation, features extraction, classification, type of imaging modality used, training parameters and learning type etc.

The remainder of the paper is organized as follows: Section 2 explains the applications of machine learning in automated breast cancer detection, Section 3 depicts study methodology, Section 4 presents detailed description of selected research parameters and their selection incentives, Section 5 depicts basic categorization structure of our study,

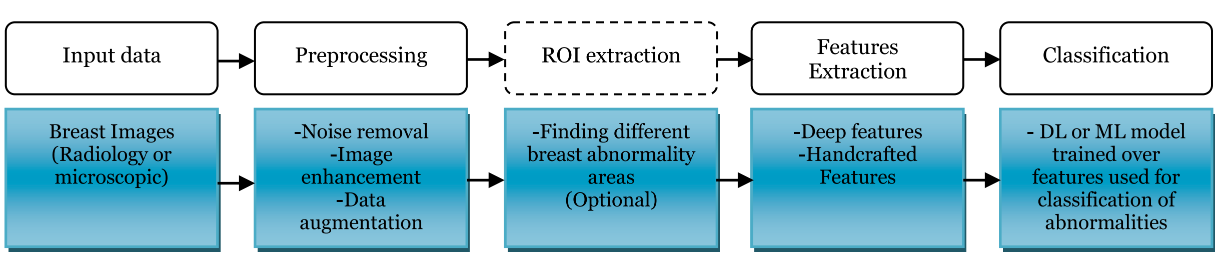

Fig. 1. Generalized CAD architecture for breast cancer detection.

in Section 6 a detailed literature review and evaluation of existing methodologies is given and finally Section 7 describes result and conclusion of our study.

2. Applications of Machine learning (ML) in breast cancer detection

CAD systems aids in automating various tasks related to breast cancer analysis such as it assists in automated risk prediction, computer aided detection and diagnosis, automated mitosis detection for tumor grading, discerning its aggressiveness, prediction of cancer recurrence and in prediction of survival rate. A number of ML based systems have been proposed for accomplishing these tasks as described below:

2.1. Risk prediction of breast cancer

Breast cancer risk assessment is made to predict the threat of developing this disease in women over a specific time interval. By correct stratification according to the risk level and providing more intensive surveillance, preventive measures (i.e. chemoprevention and surgery) to the women at high risk, the risk of breast cancer occurrence (reoccurrence) could be significantly reduced and even prevented (Amir, Freedman, Seruga, & Evans, 2010; Burton et al., 2013). There are several factors that increase the chances of breast cancer occurrence i.e. family history, hormonal factor, reproductive factor, use of alcohol, smoking, mammographic density, body mass index (BMI). Inclusion of these risk factors is necessary for the accurate risk assessment (Evans & Howell, 2007) that leads to the improved treatment and more lives can be saved with significantly reduced treatment cost (Al-Quraishi, Abawajy, Chowdhury, Rajasegarar, & Abdalrada, 2017). For risk assessment in home based patients who have family history, breast cancer scrutinizing is most important (Amir et al., 2010). Several automated risk prediction statistical models have been proposed in literature for estimating and predicting the breast cancer in such patients. In their studies, (Amir et al., 2010; Evans & Howell, 2007), the authors have discussed major traditional statistical risk assessment models (i.e. Gail model, Claus model). The major limitations in these models include: they relied only on some known risk factors, many known risk factors related to family history are not included and none of the model could discriminate that who will develop breast cancer and who has BRCA1 and BRCA2 mutation. However, the task of disease prediction with the help of certain risk factors (i.e. breast density and BRCA1/2 mutation estimate) could be significantly improved by using deep-learning based models.

BRCA1 and BRCA2 are main genes whose mutation is prone to breast cancer. Automated breast cancer risk prediction could be done by analyzing the mutation of these two highly susceptible genes. In his study (Al-Quraishi et al., 2017), the author proposed an ensemble classifier (based on deep neural networks, random forest and Support Vector Machine (SVM)) for the determination of genes susceptible to breast cancer. The characteristics of microarray gene expression dataset are different from traditional classification datasets i.e. it possesses small number of observations with high dimensions, high noise and redundancy, which may arise problems like over-fitting and may affect the accuracy of classifier. These problems vindicate the need of some feature selection criteria for the selection of most suitable genes. Therefore, in (Al-Quraishi et al., 2017), authors have used a correlation based features selection criteria for the selection of most informative set of genes from the highly redundant attributes of dataset.

Apart from age, gender, family history and gene mutation, one of the most significant breast cancer risk factors is mammographic density (MD). High dense breast density also increases the risk of breast cancer. In (Kallenberg, Petersen, Nielsen, Ng, Diao, Igel, Vachon, Holland, Winkel, Karssemeijer, & Lillholm, 2010) an automated unsupervised mammographic risk scoring technique is introduced. The authors have proposed a convolutional sparse auto-encoder (SCAE) network, which learns features from multi-scale unlabeled input images and applied sparsity regularizer on the extracted features. Resulting features are then fed to a classifier for breast density segmentation and its texture scoring.

2.2. Breast cancer detection and diagnosis

Automated detection and diagnosis of breast cancer primarily based on some intermediate steps such as segmentation (breast lesion detection), features extraction and finally classification of detected regions into different categories. Breast lesion detection is done to mark the suspicious area in the whole breast image either by pixel to pixel boundary or by drawing a bounding box around the suspected area. However, it is an optional phase and instead of first extracting suspicious areas and then classifying them, cancer could directly be diagnosed by processing whole breast images with additional cost of processing whole mammograms. Features extracted, using ROI or whole image, are used to categorize the lesions under consideration. These extracted features are fed to a statistical algorithm (ML or DL) for classification.

2.2.1. Features learning

In this task different features are unsheathed from the images that represent their characteristics and attributes. For the precise segmentation and classification of images learning of accurate and most informative features is necessary (Gardezi, Elazab, Lei, & Wang, 2019). Due to the disparity in the attributes of benign and malignant lesions, the extracted feature space is mostly massive and complex. Filtering of optimal set of features is crucial as the excessive use of features affects performance of classifier and increases its complexity (Gardezi et al., 2019). Traditionally, for the segmentation and classification of breast lesions, different types of handcrafted features were extracted from breast images i.e. texture features (Htay & Maung, 2018; Huynh, Li, & Giger, 2016), size, shape, intensity and margin features (Huynh et al., 2016). However, the emergence of DL has significantly improved the whole process of features extraction and consequently improved the performance of subsequent phases (i.e. detection and classification). As concluded in (Donahue, Jia, Vinyals, Hoffman, Zhang, Tzeng, & Darrell, 2014), deep features extracted from convolutional network trained over large dataset have significant generalization ability to perform discrimination tasks, which outperforms traditional techniques based on hand-engineered features and traditional ML methods.

Various DL based features extraction architectures have been proposed for breast cancer detection and classification such as (Wang et al., 2019) have proposed a 7-layer deep CNN and (Guan, Loew, & 2017. Breast Cancer Detection Using Transfer Learning in Convolutional Neural Networks, 2017) have proposed a 6-layer deep CNN for features extraction from mammograms, which further assists in binary classification of these mammograms by feeding the extracted features into a classifier (i.e. SVM and FCN respectively). Several authors have also used pre-trained CNNs for features extraction due to their high performance as compared to traditional systems based on handcrafted features, such as in (Kooi et al., 2017) the authors have compared a transfer learning based pretrained CNN namely VGG16 with a state-of-the-art system based on manually designed features set (both trained over a mammographic image dataset containing 45,000 images) for the detection of mammographic lesions and results shows that CNN surpassed traditional CAD system.

2.2.2. Automated breast cancer detection and classification

For the detection and classification of different breast abnormalities using breast images, either separate or combined automated CAD systems are designed. Traditionally automated breast cancer detection is done by extracting handcrafted features such as in (Punitha, Amuthan, & Joseph, 2018), an automatic ROI extraction and classification methodology is proposed for Breast Mammograms by using a region growing algorithm. A swarm optimization technique namely Dragon Fly Optimization (DFO) is used for initial seed points and thresholds generation.

From the final segmented images, texture features are extracted (using GLCM and GLRM techniques) and fed to feed forward neural network (FFNN) for their classification into benign and malignant. The task of automated breast cancer detection boosted with the advent of DL. Several DL based automated breast lesion detection models have been proposed in literature such as in (Cao, Duan, Yang, Yue, & Chen, 2019). The authors have discussed several architectures for locating lesions in Breast Ultrasound (BUS) images which includes: Fast R-CNN, Faster R-CNN, YOLO, YOLOv3 and SSD.

Apart from DL based models that have been proposed specifically for the detection of breast abnormalities, a number of authors have also used transfer learning and from models built for other problems. A combined DL based CAD system for the detection and classification of breast masses in mammograms have been proposed in (Akselrod-Ballin et al., 2019). In this study, region proposal network (RPN) is used for region of interest extraction followed by Fast-RCNN that is used for the classification of candidate mass windows returned by RPN. For features extraction both RPN and Fast-RCNN have used a modified version of VGG-Net (Simonyan & Zisserman, 2015). For the detection of micro-calcifications in mammograms, in (Mordang et al., 2016) a pixel wise classification approach is proposed based on deep CNN. To overcome class imbalance problem between pixels belonging to calcifications and other breast tissues, a hard mining technique based on two CNNs is applied. First CNN is used for the separation of easy samples, where the second CNN is used to extract hard negative samples. For the training of the proposed CNNs patches from both positive and negative classes are used. For the binary classification of mammograms (Yemini, Zigel, & Lederman, 2019) have used pretrained Google Inception-V3 architecture, (Agrawal, Rangnekar, Gala, Paul, & Kalbande, 2018) have used pretrained VGG-16 followed by a Voting Classifier and (Liang, Bian, Lyu, Zeng, & Ma, 2018) have used AlexNet followed by an ensemble classifier.

2.3. Automated breast image registration

As we have already mentioned, multiple breast imaging modalities are being exploited for automated breast cancer diagnosis, however all of them have their own strengths and limitations. For better analysis and visualization purpose, images of same or different modalities acquired at same or different times are often combined or registered. However automated registration or spatial mapping of breast images is a challenging task due to breast structural dissimilarity in different breast images and due to the distinction in imaging conditions. Basically the determination of transformation function for mapping the points in one image over the corresponding points in another image is called image registration. Different mapping functions and transformations (i.e. affine, elastic and rigid transformations etc) are applied on input images (that could be 2D or 3D breast images) for their alignment based on different similarity measures. These similarity measures could be image's control points, contours, edges, structural, statistical or syntactic descriptors, intensity values etc. An optimal search mechanism is crucial for the extraction of these features. (Guo, Sivaramakrishna, Lu, Suri, & Laxminarayan, 2006)

Image registration could be employed over inter-modality and intra-modality images. A recent study (Haskins, Kruger, & Yan, 2020) has described detailed intervention of deep learning in medical image registration.

2.3.1. Automated detection of breast structural descriptors

Several studies have worked upon the automated detection of different structural descriptors of breast. A few of them are briefly described below:

- Nipple Detection

Breast nipple is an important landmark that assists in automated breast image fusion and image registration. An automated approach for the precise detection of breast nipple using breast thermograms imaging dataset has been proposed in (Abdel-Nasser, Saleh, Moreno, & Puig, 2016), in which adaptive thresholding is employed for nipple's candidate extraction followed by a novel algorithm for the selection of actual nipple.

- Pectoral Muscle (PM) Detection

In addition to other breast markers, accurate detection of breast pectoral muscle is crucial for precise breast image analysis. Moreover, the imaging characteristics of PM are similar as that of breast fibro-glandular tissue, which make its automated detection task more critical. A hybrid approach for the automated PM detection has been proposed in (Pavan et al., 2019), in which author has employed Hough transform for PM edge detection, followed by active contour for its segmentation. Another deep learning based approach has been proposed in (Wang et al., 2019), in which author has utilized deep U-Net architecture for automated detection of pectoral muscle from breast tomosynthesis and x-ray mammograms.

- Breast wall detection

Correct identification of breast wall is a critical task, as it assists in automated breast lesion localization, in quantification of fibro-glandular tissue and it also acts as a major control point in image registration. To perform this task, automated knowledge based as well as deep learning based approaches for automated detection of chest wall has been proposed in (Verburg et al., 2019) by utilizing T1-weighted breast MRI images.

2.4. Mitosis detection

Breast cancer assessment and grading of its aggressiveness is a critical task for radiologists. Normally it is done on the basis of various characteristics like size of tumor, invasiveness of cancer, tumor spreading in the lymph nodes, and tumor spreading in other parts of the body (i.e. metastasis) (Mohebian, Marateb, Mansourian, Mañanas, & Mokarian, 2017). For breast cancer grading, mitotic detection is one of the most important criteria (Albayrak & Bilgin, 2017) that carries considerable diagnostic information necessary for grading of histopathological slides (Saha, Chakraborty, & Racoceanu, 2018). Several challenges are faced during histopathological image analysis for mitotic detection (Saha et al., 2018); in each stage of mitotic division process the shapes of mitotic figures are different. Hematoxylin and eosin (H&E) stained slides contain objects similar to mitotic cells (such as lymphocytes and dense nuclei etc.) which increase the chance of misclassification, moreover manual scrutiny of these slides is a monotonous, time consuming and prone to intra-observer variability. Automating this whole process using DL based models will attenuate these problems to a larger extent. A bulk of work has been done in literature for the development of these systems such as, in (Albayrak & Bilgin, 2017) the authors have proposed a deep CNN for features extraction form cellular structure image patches which are obtained by applying K-mean clustering based segmentation and blob analysis for noise removal. The extracted feature map is then fed to SVM for the final classification of patches into mitotic and non-mitotic. Moreover, in his study (Veta et al., 2015), the authors have presented a comprehensive review on the eleven mitosis detection techniques proposed by the participants of "Assessment of Mitosis Detection Algorithms 2013 (AMIDA13)" challenge, which was basically organized for the performance assessment of different mitosis detection methods on the extracted regions from Whole Slide Images (WSI) on a large dataset.

2.5. Prediction of cancer recurrence

After the initial treatment of breast cancer patients, still there are chances of recurrence of this catastrophic disease. The chance of its recurrence is at peak in 1st two years after initial treatment and these chances diminish slowly in 5 years post-surgery of breast cancer (Saphner, Tormey, & Gray, 1996). This reappearance could be avoided by its timely prediction and subsequently by optimizing the treatment to increase curative efficacy. Prediction of recurrence of breast cancer could be done by using clinicopathologic characteristics of cancer patients i.e. Human Epidermal Growth Factor Receptor 2–Positive (HER2 positive) and Node-negative tumor etc. (Gonzalez-Angulo et al., 2009). Unlike breast cancer detection and diagnosis, there is limited work done for breast cancer prognosis (i.e. recurrence and survival prediction), such as in (Mohebian et al., 2017) authors have used a bagged decision tree (BDT) based statistical classifier for recurrence prediction of breast cancer. In this study, the authors have used clinicopathologic data of 579 women extracted from Cohort database, subsequently applies a statistical classifier and Practical Swarm Optimization (PSO) Algorithm for the selection of more refined and optimal features and fed the final set of extracted features to the classifier (i.e. BDT). The authors have also compared the performance of BDT with SVM, Decision Tree, and Multi-Layer Perceptron (MLP).

2.6. Prediction of survival

Survival analysis is done to predict the chances of patient to stay alive after the initial breast cancer treatment. Several prognostic factors are there which influence survival chances of patient. By accurately analyzing these factors, the probability of a patient to stay alive could be predicted in advance. Several ML approaches have been used to automate this task such as proposed in (Ganggayah, Taib, Har, Lio, & Dhillon, 2019). The authors have used a database collected from a local hospital consisting of 8942 breast cancer patients records, each of which consists of 113 variables from which 89 are considered as less significant prognostic factors and therefore discarded. The remaining 24 variables (i.e. 23 independent and 1 dependent) are used for the training of ML based classifier after preprocessing. The quality of dataset is then assessed by adapting six classifiers i.e. SVM, random forest, decision tree, extreme boost, logistic regression and neural network. The resultant dataset is then clustered into three groups based on receptor status and a features selection algorithm (i.e. Random Forest Explainer) is applied for the selection of significant prognostic factors. The extracted optimal features are then fed to a decision tree classifier for predicting survival status of patient.

3. Methodology

The goal of our study is to analyze, evaluate and review existing Artificial Intelligence (AI) based classification models proposed in the last five years for early detection of breast cancer using medical images. In order to conduct a detailed review of recent methodologies proposed in the domain of interest, we have divided our study into following main steps: 1) identifying research parameters, 2) defining search query and searching for papers, 3) screening retrieved papers and filtering out low quality papers based on exclusion criteria, 4) query updating based on keywords selected from abstracts of relevant papers and searching again, 5) categorizing retrieved papers and 6) performing paper's review. Research parameters that we have selected for the evaluation and comparison of retrieved papers includes: type of imaging modality, dataset type and its size, pre-processing and segmentation techniques, type of extracted features, classifier and regularization technique used for the training of proposed model. The detailed description of these parameters and the motive behind selection of these research parameters for the evaluation of retrieved papers is discussed in Section 4. For online searching of such articles, that are admissible and homologues to our research domain, the searching query must be well defined. We have employed an iterative method of query definition. In which at first searching is done using the query that is defined on the basis of our research problem (i.e. Computer aided systems for breast cancer detection). After the initial retrieval of relevant papers, query is updated by using keywords extracted from abstracts of retrieved papers and the search process proceeded.

We have used well-known web Academia paper search engines for the searching of pertinent articles. These web resources include: (1) Science Direct, (2) Research Gate, (3) Google Scholar, (4) IEEE Xplore, (5) ACM Digital Library [5] and (6) Microsoft Academia Search [6]. In our study, we have included articles of recent years, specifically published in last 5 years. For this purpose, during search, we have explicitly specified the time limit of articles from 2016 to 2020. In addition to the research articles that are retrieved by searching through the mentioned search engines, we have also included those papers which are given in references of relevant articles (if the paper is of high quality) and the article found browsing personal web pages. In addition to this, we have also explored well-known journals and conferences of relevant domain, such as Journal of Medical Image Analysis, IEEE Access, Pattern Recognition, European journal of Radiology etc. and Conferences, such as IEEE International Conference on Image Processing, IEEE Conference on Computer Vision and Pattern Recognition etc.

We have implied an inclusion–exclusion criterion on the pool of papers retrieved through search query for excluding papers that are irrelevant to the problem for which we are concerned about and those papers which do not fulfill our specified selection criteria. Following are the main conditions which define the selection criteria: 1) Article must be highly relevant to research area under focus. 2) There must be enough information provided in the papers, i.e. the answers of our defined research questions must be present 3) Article must be published within last 5 years (i.e. 2016 to 2020). 4) Article published in a well-known journal or conference (i.e. journal with high impact factor) are preferred 5) papers using benchmark datasets and showing comparable results and 6) Top cited papers are preferred. Those articles which do not fulfill the above mentioned criteria are excluded.

The collected pool of articles is then grouped into different categories. The categorization of articles is depicted in Fig. 10 and described in Section 5. After the grouping of articles, a detailed review of these publications is done in Section 6.

4. Evaluation parameters

Performance is the primary factory in design of CAD system. There are number of factors that affect the system performance and to evaluate and compare proposed work major factors are selected and discussed in this section.

4.1. Imaging modalities

Different imaging modalities are used for screening, diagnosis and adjunctive evaluation of early stage breast cancer to increase survival chances and to reduce the need of treatment (i.e. as depicted in Fig. 2). Currently used breast imaging modalities are: mammography, breast ultrasound, thermography, magnetic resonance imaging (MRI), Scintimammography, optical imaging, positron emission tomography (PET), computed tomography (CT), histopathology and electrical impedance based imaging (Sree, 2011). Each of these modalities possesses different capabilities. Mammography is the most commonly used imaging modality for the early detection of breast cancer. Regular mammography tests prove useful for decreasing the rapidly increasing mortality rate due to breast cancer, as discussed in Kerlikowske, Grady, Rubin, Sandrock, and Ernster (1995) it reduces the mortality rate about 20–30% in women aged 50–74 after 7–9 year of follow-up. Mammography gives efficient results in the detection of micro-calcifications and in fatty breast but its accuracy impaired in case of dense breast and in patients

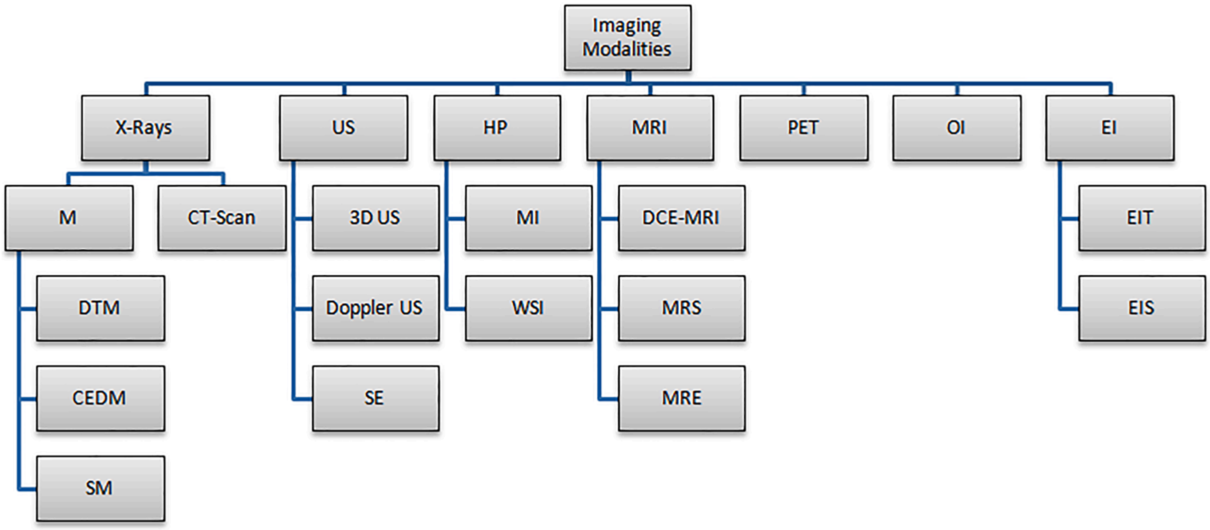

Fig. 2. Imaging modalities. Ultrasound (US), Histopathology (HP), Magnetic Resonance Imaging (MRI), Positron emission tomography (PET), Optical Imaging (OI), Electrical Impedance (EI), Mammography (M), Computed Tomography (CT-Scan), Sonoelastography (SE), Microscopic Images (MI), Whole Slide Images (WSI), Digital Tomosynthesis Mammography(DTM), Contrast-Enhanced Digital Mammography (CEDM), Scintimammography (SM).

with stern scarring and silicone implants (Heywang-Köbrunner, Viehweg, Heinig, & Küchler, 1997). To get better results in detection of invasive breast cancer combination of the different imaging modalities (i.e. mammography, ultrasound and Magnetic Resonance (MR) Mammography) is used (Sree, 2011). There exists different variations of mammography i.e. digital tomosynthesis mammography (Wu et al., 2003) that acquire 3D views of breast, contrast-enhanced digital mammography (CEDM) (Dromain, Balleyguier, Adler, Garbay, & Delaloge, 2009) that uses an intra-venous injection of an iodinated contrast agent in combination with mammography inspection and Scintimammography which helps in detection of small lesions which are occult in dense breast. Mammography alone misses many cancerous cases especially in case of dense breast, therefore employing automated whole breast ultrasound (AWBU) together with mammography significantly improves the accuracy of breast cancer detection (Kelly, Dean, Comulada, & Lee, 2010). Breast ultrasound is also used as an adjunct tool for locating breast lesions. The major drawbacks of traditional ultrasound are: it is not systematic and results are user dependent, correct probe pressure is necessary, however probe pressure may cause tissue shape distortion and lesion's displacement from original position (Riis et al., n.d.). Moreover, Breast ultrasound proves commendable in determination of cysts and palpable cancers but it is difficult to find pre-invasive carcinomas using it (Pamilo, Soiva, Anttinen, Roiha, & Suramo, 1991). There exist several advanced variations of traditional ultrasound which includes: 3D-ultrasound (Riis et al., n.d.), Doppler ultrasound (Kook, Park, Lee, Lee, Pae, & Park, 1999) and Sonoelastography (Ferranti, 2008). Together with mammography, computed tomography (CT) also use X-rays for 2D breast image acquisition. The temperature around pre-cancerous and cancerous tissues is higher as compared to normal tissues, due to their high metabolic rate. By using this factor, breast thermography is used for screening and diagnosis of breast cancer at an early stage (i.e. at least ten years in advance) (Sree, 2011). A variation of traditional MRI is dynamic contrast enhanced-MRI (DCE-MRI) (Heywang-Köbrunner et al., 1997), in which a contrast agent is injected in the patient's body before image acquisition.

Similarly other variations include Magnetic Resonance Spectrography (MRS) (Bolan, Nelson, Yee, & Garwood, 2005), and Magnetic Resonance Elastography (MRE). Positron Emission Tomography (PET) use γ rays for the detection of breast cancer. The major drawback of this methodology is: it is very costly and produces low resolution images. Other less commonly used imaging modalities for breast cancer detection includes: Optical Imaging which imaging use Near Infrared (NIR) wavelength light and electrical Impedance based imaging modalities i.e. Electrical Impedance Tomography (EIT) and Electrical Impedance Scanning (EIS).

4.2. Datasets

Research community needs data to evaluate and compare their proposed image analysis techniques. The quality and quantity of available dataset matters a lot i.e. the dataset must be properly labeled, there must be sufficient contrast between different image segments, explosive noise in images is not endurable and sufficient data must be available for training of proposed model. For detection and diagnosis of breast cancer using Artificial intelligence (AI) based systems, such good quality and large-scale datasets are needed. Different researchers and organizations have constructed assorted breast medical image datasets for training and evaluation of their proposed models. These datasets belong to different imaging modalities, such as shown in Fig. 3. Some of these datasets are open source and publicly available (i.e. kaggle, Amazone, UCI ML repository etc.) whereas other datasets are closed source. Apart from these, some challenge datasets (i.e. Grand Challenge on Breast Cancer Histology images (BACH) and BioImaging-2015) are also publicly available (See Table 2). A brief description of some open source datasets is given below.

4.2.1. Mammographic datasets

4.2.1.1. MIAS database. Mini-MIAS database (Suckling, 1994) is one of the most widely used mammographic dataset. It is collected by Mammographic Image Analysis Society (MIAS), which is an organization of UK research groups. The dataset consists of 322 single slice digital mammograms of size 1024*1024. It is divided into three classes on the basis of characteristics of background tissues i.e. Fatty, Fatty-glandular, Dense-glandular. Each of these classes is further subdivided into normal, benign and malignant. The dataset consists of 206 normal and 116 abnormal (64 benign and 52 malignant) cases (Li, Ge, Zhao, Guan, & Yan, 2018).

MIAS dataset has been used by many authors in their studies for the training and evaluation of their proposed techniques for breast cancer detection. For example, (Htay & Maung, 2018) has used handcrafted features from 120 images of MIAS dataset for the training of KNN classifier, (M., S., A., A., E., H., T., M, 2017) proposed a double thresholding technique for the segmentation of breast mammogram using 4-

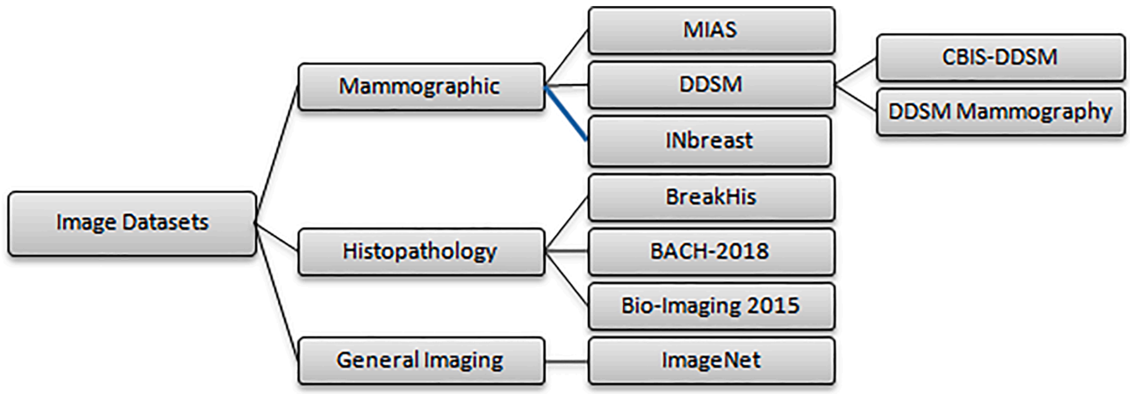

Fig. 3. Medical Imaging Datasets.

Images of MIAS dataset. (Charan, Khan, & Khurshid, 2018) used MIAS dataset for the training and evaluation of proposed 7-layer deep CNN. (Gua, et al, 2017) used deep features extracted from mammograms of MIAS dataset using VGG-16 for their binary classification. As the size of MIAS dataset is small, which may cause over-fitting in some cases therefore (Kral & Lenc, 2016) merged MIAS and DDSM dataset and extracted handcrafted features for the binary classification of mammograms.

The major drawbacks found in MIAS dataset are: 1) The size of dataset is limited, which may cause over-fitting in training DL models. 2) The dataset is imbalanced. From total 320 images 206 cases are normal and 116 cases are abnormal. 3) Mammograms consist of some superfluous data in addition to the mammary area, which needs to be removed before further processing. 4) The dataset didn't contain pre-generated ground-truths; instead it contains the center point and radius of abnormal area. So, extra processing is needed to generate ground-truths. 5) Ground truths only contain an approximate circle around abnormality not the pixel to pixel boundary of abnormality.

4.2.1.2. DDSm. Digital Database for screening mammography (DDSM) (Heath et al., 1998, 2000) is another resource of mammograms, which is extensively used by the mammographic image analysis research community. It consists of 2620 cases of scanned film mammograms along with verified pathology information where each case consists of two mammograms of each breast (CC and MLO view). For abnormal cases, the dataset also includes pixel level ground truths information using chain codes. The images are stored in LJPEG format and the corresponding ground truth information is stored in OVERLAY file. Along with breast mammograms, the dataset also includes patient's information such as age, breast density and image information such as resolution and scanner type which is stored in .ICS file. The whole mammography dataset is divided into three categories: normal, benign without callback, benign and cancerous.

Many authors have used DDSM in their studies for the evaluation of their proposed models. In their study, (Kral & Lenc, 2016) have used LBP features extracted from MIAS and DDSM databases for training SVM classifier, (Guan et al., 2017) has extracted deep features using transfer learning based model VGG-16 for the classification of breast mammograms into benign and malignant.

The major drawbacks found in DDSM dataset are: 1) This image dataset was first released in 1997, at that time computational recourses were limited; therefore dataset was compressed in a non-standard compression format (JPEG) which is not updated for modern computers, 2) In spite of pre-generated ground truth images, dataset consists of information for ground truth generation in the form of chain codes in OVERLAY files, therefore extra computation is required for their generation, 3) The ground-truths of DDSM indicate the general position of lesions not the precise location (Lee et al., 2017), therefore proper segmentation techniques are needed to implement for better features extraction, 4) Some of the images of DDSM database cannot be clearly seen (Lee et al., 2017) and the outlines of lesions are not accurate (Song et al., 2009), therefore assistance of expert radiologists is required for reviewing the ground-truths, 5) Surplus artifacts are present in digitized images due to dust, scratches and some are introduced by scanner (Heath, Bowyer, Kopans, Kegelmeyer, Moore, Chang, & Munishkumaran, 1998).

4.2.1.3. CBIS-DDSm. To overcome above mentioned drawbacks a variation of DDSM database is introduced namely CBIS-DDSM (Curated Breast Imaging Subset of DDSM) (Lee et al., 2017). CBIS-DDSM database consists of a subset of DDSM. There are several questionable images present in DDSM database, such as ground truths indicate the presence of suspicious lesion which are not present in original images, such images are filtered out and the vague images are re-annotated with the help of expert mammographers in CBIS dataset. Moreover, the images are decompressed from LJPEG format to 16-bit grayscale TIFF format and later converted into DICOM format. For making the process of features extraction more accurate, a lesion segmentation technique is implied on images of DDSM before adding them to CBIS-DDSM database. As some studies require only ROI instead of full images, therefore together with full size mammography images, cropped ROIs are also provided in the database which also increases the size of dataset. The database is divided into two categories (i.e. masses and micro-calcifications) and each category further splits into training and testing dataset (i.e. 80% training, 20% testing). As some mammograms contain both masses and micro-calcification therefore they are replicated in both groups. Information about ground truths (i.e. chain codes), patients (i.e. age, breast density etc.) and images (i.e. resolution, scanner type etc.) are stored in . CSV file. The size of database is large enough due to replication of mammograms.

4.2.1.4. DDSM-Mammography. DDSM-mammography is another variation of DDSM database. It consists of 55,890 examples from which 14% are positive cases taken from CBIS-DDSM database and the remaining 86% are negative cases taken from DDSM database. During pre-processing ROIs are extracted from images and resized to 299*299 pixels patches. Different data augmentation techniques are then applied on extracted positive patches. The data is stored in .tfrecords files. The images are annotated with two types of labels as given in Table 1. A major drawback seen in this database is the distribution of negative and positive class (i.e. the dataset is highly imbalanced with 86% of negative class examples). The reason behind this distribution is that, the authors want to develop a realistic dataset, in which the number of negative cases is far more than positive cases. However, to avoid model biasness, some strategy (i.e. weighted cross entropy) for giving more weights to positive class must be employed when developing classification models.

Table 1
Labels of DDSM-Mammography dataset.

	Class	Label
Binary Classification	Normal	0
	Abnormal	1
Multi Classification	Benign calcification	1
	Benign Mass	2
	Malignant Calcification	3
	Malignant Mass	4

4.2.1.5. Inbreast. INbreast dataset (Moreira et al., 2012) is a full field digital mammographic dataset that is collected at Breast Centre in CHSJ Porto. The dataset is collected from a total of 115 cases, from which 90 women are those who's both breasts are affected, where the remaining 25 cases are mastectomy patients. Two images (CC and MLO view) per breast are collected from the women whose both breasts are defected and two images of only one breast are taken from the women who had a mastectomy. The dataset consists of 410 images of resolution 3328 * 4084 and 2560 * 3328 pixels. Information about annotation of images is stored in xml files. These xml files also contain information about number of ROIs.

The images in INbreast dataset contain mainly six types of abnormalities: distortion, calcification, masses, asymmetries, multiple findings and normal. The major advantage of INbreast dataset is: Existing mammography databases such as MIAS provides an approximate circle to mark the abnormal region where in DDSM database a pixel level annotation of abnormalities is provided but that is not accurate as mentioned in (Song et al., 2009), which had a great impact on classification accuracy. The annotations are done by expert field specialist and validated by a second specialist. In case, if the specialists disagree on any mammogram that case is re-analyzed.

One major limitation of this dataset is it only contains 410 images; the size of this dataset is too small for the training of a DL model. Therefore, extra processing must be done to extend data size i.e. by employing data augmentation techniques.

4.2.2. Histopathological

4.2.2.1. BreaKHis. The Breast Cancer Histopathological Image Classification (BreaKHis) (Spanhol, Oliveira, Petitjean, & Heutte, 2016) is a publicly available breast microscopic images dataset, which is built with the participation of P&D Laboratory Pathological Anatomy and Cytopathology, Parana, Brazil. It incorporates 7909 histopathological breast tumor biopsy images of size 700 × 460 pixels, from which 2480 images belong to benign class where the remaining 5429 examples belong to malignant class. Images are assembled from 82 different patients at different magnifying factors (i.e. 40×, 100×, 200×, and 400×). The resulting images are cropped (i.e. to remove black border) and saved in 3-Channel RGB format with 8-bit depth of each channel in PNG format. The database incorporates four types of benign tumors (i.e. Phyllodes Tumor (PT) Tubular Adenoma (TA), fibroadenomas (F), adenosis (A)) and four categories of malignant tumors (i.e. Papillary Carcinoma (PC), Mucinous Carcinoma (MC), Lobular Carcinoma (LC), Ductal Carcinoma (DC). Hence this dataset can be used for multi-classification studies.

4.2.2.2. Bach-2018[3]. As part of 15th International Conference on Image Analysis and Recognition ICIAR-2018, the Grand Challenge on BreAst Cancer Histology Imaging (BACH) (Aresta et al., 2019) was organized. This challenge is mainly comprised of two parts, A and B. In Part A, participants are endowed with Hematoxylin and Eosin (H&E) stained breast histology microscopy image dataset. The task is to automatically classify the provided dataset into four classes (i.e. Normal, Benign, In situ carcinoma, and Invasive carcinoma). In part B, breast histology whole slide images are provided for their pixel wise classification into same four classes. Below is the description of microscopic

Table 2
Publically available mammography and histopathology datasets, Abnormalities: Masses (M), Micro-calcifications (C), Architectural Distortion (D), Asymmetries (AS), Adenosis (A), Fibroadenomas (F), Tubular Adenoma (TA), Phyllodes Tumor (PT), Ductal Carcinoma (DC), Lobular Carcinoma (LC), Mutinous Car-cinoma (MC), Papillary Carcinoma (PC), Imaging Modality: Mammography (M), Histopathology (H), General Imaging (G), Imaging Mode: Digitized Film Mammography (FM), Full Field Digital Mammography (DM), Microscopic (M), Category: Normal (N), Benign (B), Malignant (M), Cancerous (C), Benign without Callbacks (BWC), In-Situ Carcinoma (SC), Invasive Carcinoma (IC).

Database	Modality	Size	Number of Cases	Resolution (bits/pixel)	Magnification Scale	Format	Mode	Type	Abnor-malities	View	Labels	Labels file	Train/Test split	BIRADS	Category	Origin	Year	Size
Mini-MIAS (Suckling, 1994)	M	322	161	8	–	.PGM	FM	Gray	M,C, D, AS	MLO	Center and radius of circle around ROI	.txt	No	No	N,B,M	UK	2003	422 Mb
DDSM (Heath et al., 1998, 2000)	M	10,480	2620	12, 16	–	.LJPEG	FM	Gray	M,C, D, AS	MLO, CC	Pixel wise annotation	OVERLAY	No	Yes	N,B,C, BWC	USA	1999	–
CBIS-DDSM (Lee et al., 2017)	M	10,239	1644	16	–	.DICOM	FM	Gray	M,C, D, AS	MLO, CC	Pixel wise annotation	.CSV	Yes	Yes	N,B,C, BWC	USA	2017	–
DDSM Mammography	M	55,890	–	16	–	.npy, .tfrecords	FM	Gray	M,C, D, AS	MLO, CC	Pixel wise annotation	.npy	Yes	Yes	N,B,C, BWC	USA	2018	6 GB
INBreast (Moreira et al., 2012)	M	410	115	14	–	.DICOM	DM	Gray	M,C, D, AS	MLO, CC	Contour points of the ROI	.XML	Yes	Yes	N,B,M	Portugal	2011	–
BreakHis (Spanhol et al., 2016)	H	9,109	82	8	40×, 100×, 200×, and 400×	.PNG	M	RGB	A,F,TA, PT,DC,LC, MC,PC	N/A	Image wise annotation	.txt	No	N/A	B,M	Brazil	2015	3.98 GB
BACH-2018	H	500	500	8	–	.TIFF	M	RGB	–	N/A	Image wise annotation	.csv	Yes	N/A	N,B,SC, IC	Portugal	2018	10–20 MB
Bio-Imaging 2015 (Pego & Aguiar, 2015)	H	140	140	8	200×		M	RGB	–	N/A	Image wise annotation	–	Yes	N/A		Portuga	2015	–

data provided by BACH-2018. The provided dataset consists of total 500 (i.e. 400 training, 100 testing) breast tissue biopsy microscopic images of size 2048 × 1536 pixels. In training dataset, images are equally distributed into four categories (i.e. Normal, Benign, In situ carcinoma, and Invasive carcinoma), where each category contains 100 images. Images are annotated by expert medical practitioner. Whenever there is a conflict on an image between normal and benign class, the image is discarded. Resultant RGB images are stored in .TIFF format, where image wise labels are saved in .csv file.

4.2.2.3. Bio-imaging-2015. Bio-Imaging 2015 challenge was presented to push forward new methods for automatic classification of breast tumor tissue slides, in the field of digital pathology. Main goal of this challenge is to automatically classify H&E stained breast histology tissue slides into four classes (i.e. normal tissue, benign lesion, in situ carcinoma and invasive carcinoma).

The dataset provided by the challenge to the participants consists of 140H&E stained breast tissue slides microscopic images (i.e. 120 training and 20 testing) of size 2048 × 1536 pixels, which are magnified at a factor of 200×. Training dataset is basically annotated with four labels (i.e. normal tissue, benign lesion, in situ carcinoma and invasive carcinoma) where each category consists of 30 images.

4.2.2.4. Miscellaneous Datasets. As it is already mentioned above that publicly available datasets of breast cancer images are produced using different modalities. Different researchers have used these datasets in their studies for the training of their proposed models, but most of those datasets are of limited size (i.e. consists of a few hundred of examples). The problem arises is that the model overfits when trained over limited size dataset. To overcome this problem, transfer learning approaches are used; in which models are first trained over large datasets of images belonging to any other domain (i.e. any general imaging dataset) and then fine-tuned over the dataset belongs to domain of interest. Such a widely used general imaging dataset is ImageNet (Deng, Dong, Socher, Li, & Li, 2010). Many transfer learning based deep models are pretrained over this dataset i.e. VGG-16, ResNet, Inception-V3 etc.

4.2.3. Ultrasound

4.2.3.1. UDIAT dataset. UDIAT (Yap et al., 2018) is a breast ultrasound (BUS) imaging dataset that consists upon 163 BUS images (i.e. 53 cancerous and 110 benign) of size 760*570 pixels. This dataset are collected UDIAT diagnostic center, Spain in 2012. Images of this dataset are annotated by expert radiologists and incorporate almost all type of breast malignancies (i.e. ductal carcinoma in situ, invasive ductal carcinoma, invasive lobular carcinoma etc). Authors have made this dataset publically available for research purposes, while it is available at: http://www2.docm.mmu.ac.uk/STAFF/m.yap/dataset.php.

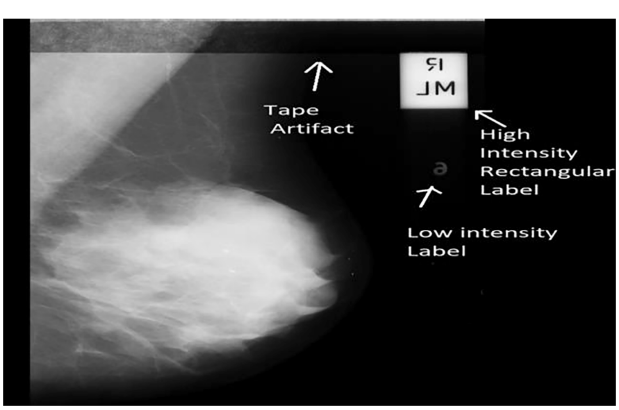

Fig. 4. Types of Noise Artifacts (Sample taken from MIAS Dataset).

4.2.4. MRi

4.2.4.1. RIDER breast MRI dataset. Reference Image Database to Evaluate Therapy Response (RIDER) is a publically available breast MRI imaging dataset. It encompasses MRI data of 5 patients and have 153 total images (i.e. 63 normal 28 abnormal) in DICOM file format. (Ibraheem, Rahouma, Hamed, & 2019. Automatic mri breast tumor detection using discrete wavelet transform and support vector machines. NILES, 2019) This dataset is publically available at: https://wiki.cancerimagingarchive.net/display/Public/RIDER+Breast+MRI

4.2.5. Thermography

4.2.5.1. Dmr-ir. DMR-IR is an infrared breast imaging open access dataset. The acquisition of this thermal imaging dataset is done by following a dynamic protocol i.e. 20 sequential infrared images have been taken with the interval of 15 s during 5 min after cooling down of breast using a FLIR camera. The images of dataset are of size 640 × 480 pixels, while the dataset incorporates 37 sequences of cancerous while 19 sequences of healthy patients. The dataset also contains segmented breast images i.e. that depict breast temperature. This dataset is online available at:

4.3. Preprocessing

Preprocessing is a basic component of most automated image analysis systems (Weickert, Weickert, & Schnörr, 2000). Basically, image pre-processing is done for image quality enhancement, noise reduction and for the removal of unwanted, superfluous artifacts from images. It helps in improving the results of subsequent processes (i.e. image segmentation) (Weickert et al., 2000). Breast medical images such as mammograms are difficult to interpret, therefore preprocessing phase is necessary to improve image quality and to make segmentation process more accurate (Bandyopadhyay, 2014; Ponraj & Jenifer, 2011). Different types of noises and imaging artifacts (i.e. tape artifact, high intensity rectangular label) found in MIAS (Suckling, 1994) dataset, that needs to be removed during preprocessing is depicted in Fig. 4. There are different studies which show that use of suitable preprocessing methods can boost the results (Tahir, 2019).

Various denoising filters are applied to remove unwanted information from images i.e. mean filter, median filter (Htay & Maung, 2018), adaptive mean filter (Wang et al., 2019) and adaptive median filter. The major drawback of using low-pass filters (i.e. mean and median filter) is that it not only denoises the input images but also blurs their edges (Ponraj & Jenifer, 2011). To tackle with this problem adaptive mean and adaptive median filters are used for noise removal (Ponraj & Jenifer, 2011). Additionally, breast medical images (i.e. mammogram, ultrasound etc.) possess limited contrast due to which different contrast enhancement techniques such as histogram equalization, adaptive histogram equalization (AHE), contrast limited adaptive histogram equalization (CLAHE) is applied on input images. Different noise removal and enhancement techniques have different impact on lesion segmentation and classification results, such that by using an appropriate image preprocessing technique the results of subsequent operations (i.e. segmentation and classification) can be improved (Rodriguez-Cristerna, Guerrero-Cedillo, Donati-Olvera, Gomez-Flores, & Pereira, 2017; Tahir, 2019), which is a major reason of using preprocessing as an evaluation parameter in our study.

Apart from image enhancement and noise removal techniques, several other operations are also performed during preprocessing phase: which includes image resizing, data augmentation (Taylor & Nitschke, 2017) and normalization. Normally different pretrained deep CNNs have different input layer size and different number of channels, so input images need to be resized and transformed before feeding them to the network. Convolutional neural networks need large size datasets for

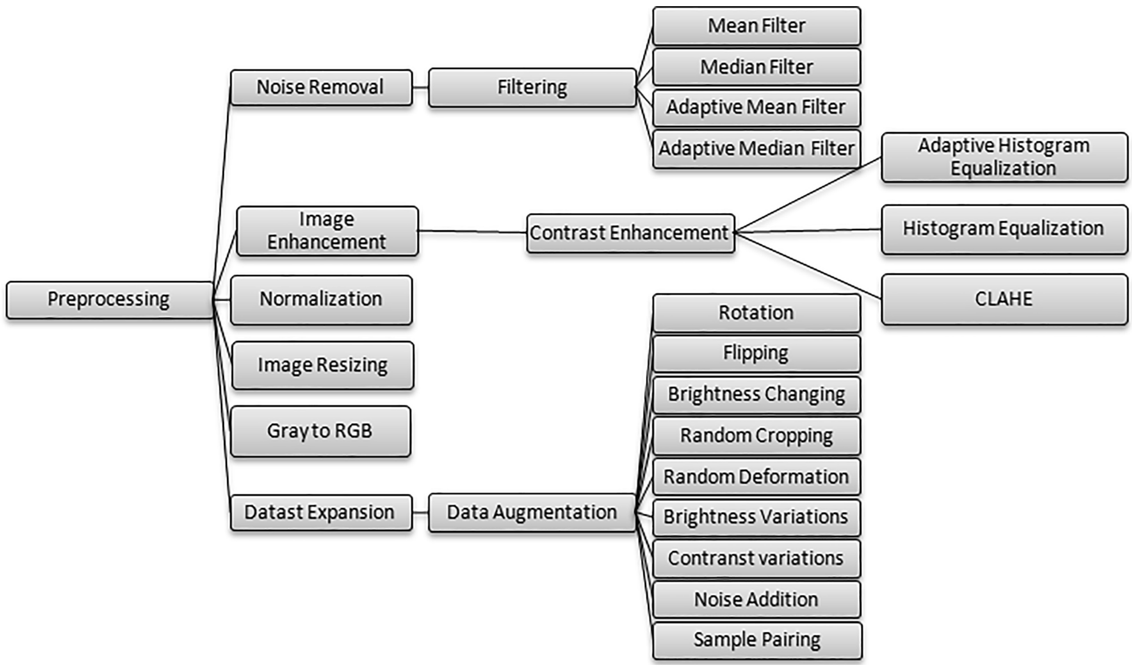

Fig. 5. Preprocessing techniques.

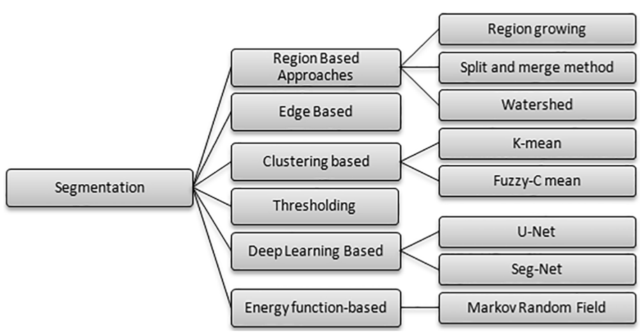

Fig. 6. Segmentation techniques.

training. Usually the size of publicly available datasets is not enough for the training of CNNs, therefore data augmentation technique is applied on it, in which original images are randomly transformed (i.e. random rotation, flipping etc.) and added to the dataset. Fig. 5 represents commonly used preprocessing techniques for problem under discussion.

4.4. Segmentation

Image segmentation is done to partition the images into regions based on similar characteristics such as texture, gray level, brightness, contrast and color. Segmentation is a significant, crucial and most difficult component of image processing and the quality of final results are based on it (Cheng et al., 2010). The precise and exact segmentation of breast images is essential for performing subsequent tasks efficiently

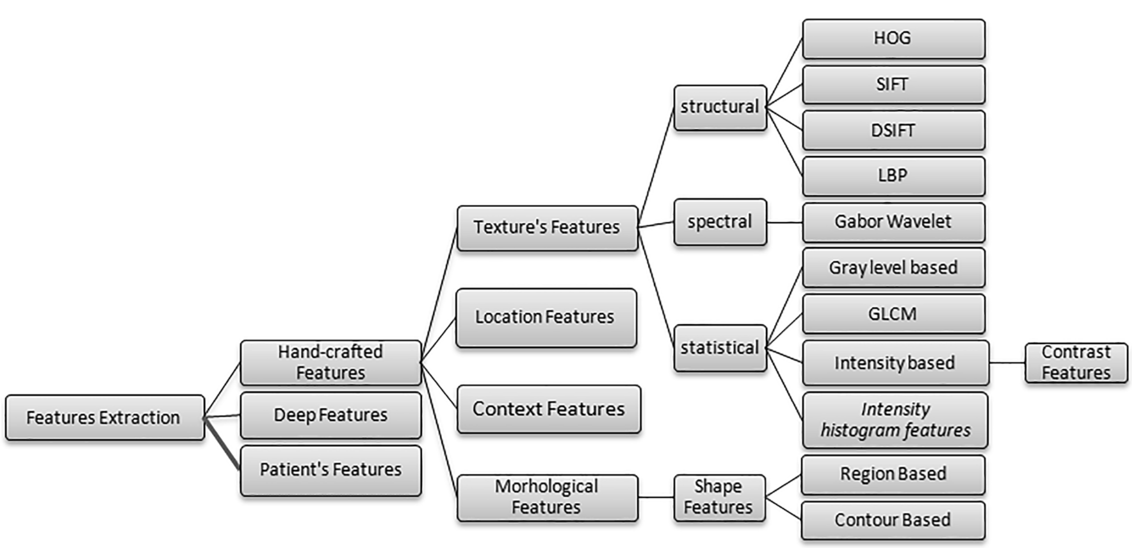

Fig. 7. Types of features.

and accurately. Different authors have used different segmentation techniques in their studies, that can be categorized into thresholding based, region based, edge based, un-supervised clustering based, and DL based segmentation techniques (as depicted in Fig. 6). But there is no comprehensive segmentation method which proves efficient for all type of images and on the other hand not all methods are good for some particular imaging modality (Ramani, Suthanthiravanitha, & Valarmathy, 2012).

The main objectives of breast image segmentation is to separate breast area from background area and to highlight suspicious area (ROI) from rest of the breast region to focus more on detection of abnormalities in the breast region and to minimize the effect of background. The removal of background reduces the search space for finding abnormalities (Ponraj & Jenifer, 2011). Together with all the segmentation techniques discussed above, in some studies bounding box is also used to extract region of interest from images.

4.5. Features extraction and selection

For the discrimination of cancerous and non-cancerous breast cancer lesions a feature set is extracted from the breast images. The extracted feature set could be massive and multifaceted enough for the training of a classification model, therefore extraction and selection of most effectual features is necessary. The accuracy of a classifier mainly depends upon the type features on which it is trained. Even if same classification approach is used in a study and with different feature set the results may be different (Zheng, Yoon, & Lam, 2014). Therefore, selecting the correct feature extraction technique is crucial for the learning of classifier algorithm and for utilizing full capability of a CAD system. Mainly there are two types of features used for the classification of breast images namely handcrafted features and deep features. Additionally a few studies such as (Kooi et al., 2017) have also included patient's features (i.e. age, medical history etc) in their study.

There exists a variety of handcrafted features that could be used for breast cancer detection, different researchers have categorized them differently i.e. in (Cheng et al., 2010) authors have categorized them into texture features, morphological features, model based and descriptor features, in (Ede, n.d.) categorization is done into shape, texture, intensity and hybrid features, (Francis, Sasikala, & Saranya, 2014) have used curvelet based statistical and texture features for the classification of thermograms, (Nahid & Kong, 2018) have classified them into local and global features and further sub-classified local features into textural, structural, descriptor, statistical, BIRADS and Detector features, (Ergin & Kilic, 2014) have introduced a new set of features namely, Histogram of Oriented Gradients (HOG), Scale Invariant Features Transform (SIFT), Dense Scale Invariant Features Transform (DSIFT), Local Configuration Patterns and Gabor Wavelet and (Kooi et al., 2017) have categorized them as Contrast, texture, geometrical, location, context and patient's features.

For the extraction of deep features Convolutional Neural Networks are used, which extracts features by themselves using convolutional layers. The output feature map of convolutional layer consists of abundance of features which are then usually down sampled using pooling layer. However, pooling layer also downscales image, induces translation invariance and reduces location precision (Kooi et al., 2017). To the best of our knowledge we have categorized all types of features that have been used for breast cancer classification in different studies (reviewed in this work) and depicted in Fig. 7. Several researchers have used solely handcrafted features for the classification of breast lesions such as in (Htay & Maung, 2018; Huynh et al., 2016; Kral & Lenc, 2016), some have used only deep features (Chennamsetty, Safwan, & Alex, 2018; Kwok, 2018) and some have used the combination of handcrafted and deep features (Wang et al., 2019). All of these studies have manifested different performance in categorization of breast images.

4.6. Classification

It is the final phase after features extraction and selection, with the help of obtained feature map using an efficient classification model. The frequently used classification models for breast cancer detection can be broadly classified into two main categories i.e. ML based and DL based as depicted in Fig. 8.

In statistical ML approaches, discriminant low level feature vectors (i.e. handcrafted features) are extracted from the images and fed to some computational algorithm which determines a decision boundary in the provided high dimensional features space. Some of the commonly used statistical approaches used by different researchers for the classification of breast cancer images includes: SVM, KNN, Decision Tree, Random Forest and Gradient Boosting etc. A major drawback of this approach is that, the task of features extraction should be manually done by researchers. This problem can be avoided using DL based models, which optimally extract and learn high level features on their own way. Most successful type of DL based model for image analysis is Convolutional Neural Network (CNN) (Litjens et al., 2017). Major benefits of CNN includes: they didn't need to learn separate detectors for detecting similar objects at different places in an image and reduces the number of parameters to a large extent (Litjens et al., 2017). However, a bulk of training data is required for the proper training of the DL model to get accurate results and to avoid overfitting. Usually such large-scale datasets related to problem domain are not publicly available. To overcome this problem either different regularization techniques are applied or pretrained transfer-learning based models are used. As seen by the literature, a number of CAD systems have been proposed based on pretrained transfer learning models which are discussed in **Section 5**. Commonly used CNNs that are pretrained on ImageNet dataset includes AlexNet (Krizhevsky, Sutskever, & Hinton, 2012), VGG-16 (Simonyan & Zisserman, 2015), Inception-Net, Dense Net (G. Huang et al., 2017), Res Net (He, Zhang, Ren, & Sun, 2016) and ResNetXt50 (Xie, Girshick, &

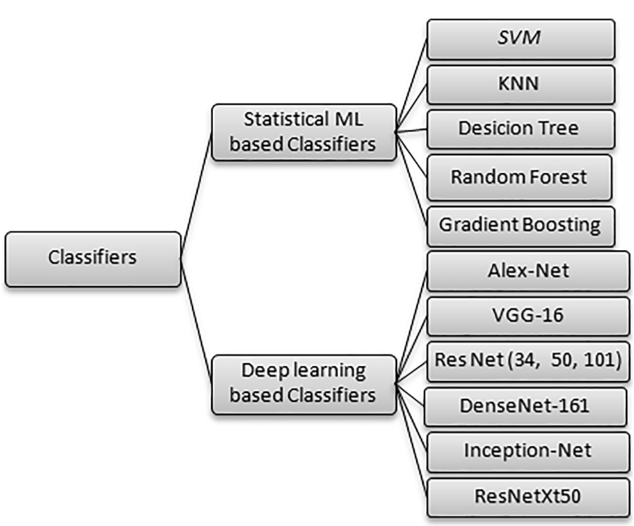

Fig. 8. Types of Classifiers.

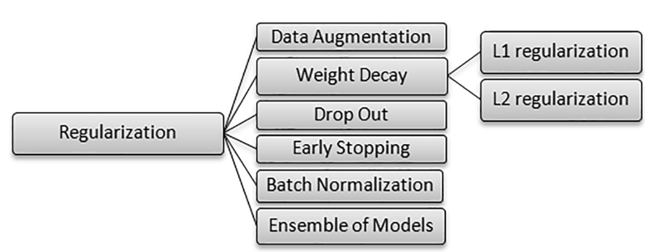

Fig. 9. Types of Regularizations.

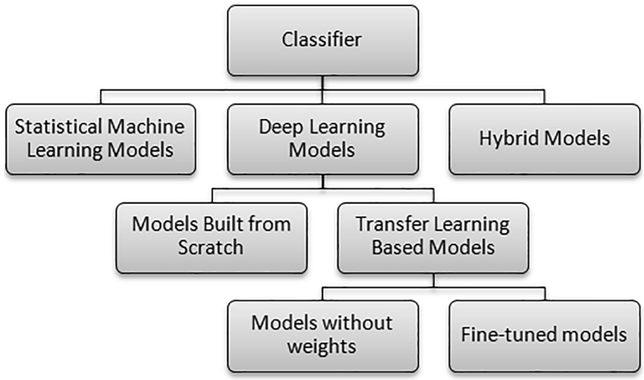

Fig. 10. Categorization Scheme.

Doll, 2017).

4.7. Regularization

A major challenge in ML is to design a classification model, which not only performs well on training data but also generalize well on unseen data (i.e. a model which avoids overfitting). Deep neural networks are prone to overfitting because of their ability to memorize the patterns in training data. A common perception about neural networks is that, the networks with too much capacity have poor generalizability (Caruana, Lawrence, & Giles, 2001). However this is not the veracity, networks with excess capacity could generalize better than smaller networks if some regularization mechanism is applied on it to avoid overfitting (i.e. Early stopping) (Caruana et al., 2001; Smirnov, Timoshenko, & Andrianov, 2014). Massive amount of regularization techniques have been presented such as in this study (Kukačka et al., n.d.), the authors have broadly categorized these techniques into five main types (i.e. data based, network architecture based, error function based, regularization term and optimization based techniques), where in (Moradi, Berangi, & Minaei, 2019) a comprehensive survey on regularization techniques is given in which authors have broadly classified them into thirteen major categories (i.e. Weight decay, Adding noise, Dropout, Drop connect, Ensemble of models, Data augmentation, Early Stopping, Adversarial training, Multi–task learning, Layer–wise pre–training and initialization, Architectural Regularization, Label smoothing and Batch Normalization).

From the bulk of regularization techniques, most commonly used in training of deep CNNs for the detection of breast cancer (according to reviewed work) includes data augmentation, weight decay, dropout, early stopping, batch normalization and ensemble of models (shown in Fig. 9).

5. Categorization scheme

In this study, we have reviewed recently published approaches for computer aided breast cancer detection and diagnosis. These approaches are broadly divided into four major categories, which are: 1) semi-automated approaches 2) statistical ML based classifiers, 2) DL based classification models, 3) hybrid models. The hierarchy of categorization scheme is depicted in Fig. 10 and the brief description about these categories and their sub-categories is discussed below.

In some recent studies, traditional image processing based techniques have been employed for the detection of different breast abnormalities, such techniques are referred here as semi-automated detection approaches. ML based breast cancer detection approaches mainly operate in two stages i.e. radiomics features extraction from input images followed by categorization of these images using statistical classifiers (i.e. KNN, SVM etc). A large number of DL based architectures are also being used for the classification of breast medical images, most common among which is deep convolutional neural network (CNN). The task of convolutional layer is to extract massive number of features from breast images which is subsequently fed to pooling layer for down-scaling and ultimately the final set of features is fed to fully connected layer for the classification of breast images. These deep models are further sub-categorized into 1) Models that are built from scratch (i.e. new deep architecture is designed and trained from scratch over breast images dataset) and 2) Transfer learning based models (i.e. Alex Net (Krizhevsky et al., 2012), VGG-16 (Simonyan & Zisserman, 2015), Inception-Net, Dense Net (G. Huang et al., 2017), Res Net (He et al., 2016) and ResNetXt50 (Xie et al., 2017)), while transfer learning based models are further sub-categorized into following categories 2.1) Weightless Models (i.e. architectures of pretrained models are used alone without their weights by training them on new data from scratch) 2.2) Fine-tuned models (i.e. pre-trained models that are further fine-tuned over breast images dataset).

Apart from the two transfer learning based methodologies discussed above, a third category exists namely features extraction models, in which pretrained models are adopted with their weights, no-fine-tuning of models is done and they are used for extraction of features, which are fed to fully connected layer for classification. However most of the studies have used statistical ML based or other classifiers instead of its fully connected layers, that's why we have added such studies in third major category of our paper discussed below. 3) Last major category of these approaches is hybrid models, in which features are extracted using some deep CNN (i.e. transfer learning based pre-trained) and fed to some classifier (i.e. statistical ML classifier). Another approach is ensemble-based classification where multiple ML models are used with majority-voting based mechanism (Iqbal et al., 2019).

6. Literature review

6.1. Semi-Automated breast tumor detection techniques

In his study (M., S., A., A., E., H., T., M, 2017), a double thresholding technique is proposed for the segmentation of mammograms to detect breast cancer; that is much faster and takes less memory as compared to other manual segmentation technique. Four sample images of size 1024 * 1024 pixels are taken from Mini MIAS dataset (SUCKLING, 1994). On the selected sample of images double thresholding technique is applied for segmentation. In double thresholding technique, two intensity values L and U are selected from the input image. For every pixel in input image, if its value is between L and U then the corresponding pixel in output image is transformed to white otherwise black. To remove unwanted edges from the thresholded image, a mask of size 1024 * 1024 is multiplied by the image. On masked image some morphological operations are applied for the smoothing of borders. The resultant image is added back to the original image for contouring and highlighting the abnormalities areas.

An image processing techniques based automated approach for the segmentation of breast lesions in breast MRI images has been proposed in (Rahman, Hussain, Hasan, Sultana, & Akter, 2020). The main aim of this study is to minimize the processing cost and time required for the automated detection of breast tumors using machine learning approaches. Author has proposed a 3-phases approach i.e. in first stage median filter is applied for noise reduction and contrast enhancement is done, secondly Otsu's thresholding is applied for image segmentation, while at the third stage different morphological operations such as hole filling, shrinkage and dilation operations are applied to remove unwanted regions. A publically available RIDER breast MRI imaging dataset is utilized for the validation of proposed methodology. This study achieves an overall accuracy of 97.33% over 150 test images.

In (Wang, Li, Liu, Liu, Zang, Liu, Dong, & Chang, 2017) a 3-step semi-automated approach has been proposed for the detection of breast lesions in breast CT images. At first stage, a bounding box is drawn manually by the assistance of expert radiologist for local lesion area

Table 3
Semi-automated approaches for breast cancer detection.

Paper	Image Modality	Classification type	Preprocessing	Segmentation	Features	Classifier	Dataset	Dataset Size	Evaluation Metrics	Regularization	Learning Type
(M., S., A., A., E., H., T., M, 2017)	Mammography	Pixel by pixel detection	–	Double Thresholding, Masking, morphological operations	–	Manual	MIAS	4	–	–	Semi-automated
(Rahman et al., 2020)	MRI	Pixel by pixel detection	Median filter, contrast enhancement	'Otsu Thresholding, hole filling, dilation, shrinkage	–	Manual	RIDER Breast MRI database	150	Accuracy = 97.33%	–	Semi-automated
(Wang et al., 2017)	CT	Pixel by pixel detection	Modified histogram Equalization	4 Seed Random walk	–	Manual	Proprietary dataset	50 cases, 630 slices	DICE = 0.886	–	Semi-automated
(Singh, 2016)	CT	Pixel by pixel detection	Contrast enhancement, ROI extraction	Global thresholding, noise removal, Sobel edge detection	–	Manual	Proprietary dataset	1	–	–	Semi-automated

extraction. In the second stage, a modified histogram equalization technique has been applied to enhance the contrast of lesion area and to fill dark region areas inside tumor. At the final stage a four seed random walk algorithm (Grady, 2005) is employed on preprocessed patch for the segmentation of tumor area. A locally collected dataset collected from 50 patients (i.e. 630 CT slices) has been utilized to check the validity of proposed method, while DICE coefficient is utilized for evaluation.

Another simple semi-automated tumor segmentation approach has been proposed in (Singh, 2016), in which author have utilized a breast image collected from an anonymous source for experimentation. In the proposed technique, at first sharpness enhancement is done, than image is manually cropped to extract ROI area. On the extracted ROI, global thresholding segmentation is applied. From the resultant segmented image, noise is removed and Sobel edge detection technique is applied to detect boundary of lesion.

6.2. Discussion

Despite of emerging applications of deep learning in automated diagnosis of breast cancer, still there are some studies that are employing traditional image processing techniques for automated abnormalities detection. The main aim of these methodologies is to reduce the computational cost and memory requirements, however these methods are unable to recognize patterns in dataset (i.e. done in deep learning models), therefore their results are not as accurate as that of ML and DL based approaches. Moreover, these algorithms requires manual parameter tuning (i.e. parameter values are selected on hit-and-trial basis) such as author in (M., S., A., A., E., H., T., M, 2017) have selected threshold values to perform segmentation (See Table 3).

6.3. Traditional ML models

In 2016, Huynh et al., compared the accuracy of SVM for binary classification of breast mammograms by training it over two types of features (ML features and DL features). In first case, region of interest is extracted from mammograms using region growing algorithm. From resultant region of interests, size, shape, margin (i.e. sharpness and spiculation) and intensity (i.e. contrast, texture and average gray level) are extracted and fed to SVM for classification. The classifier achieved an AUC of 0.81 at 5-fold cross validation at a dataset containing 607 images.

In (Kral & Lenc, 2016), the authors have proposed a methodology for the classification of mammographic images into normal and cancerous using uniform Local Binary Patterns (LBP) for texture features extraction. The mammograms are first thresholded using Otsu's thresholding for background removal and then LBP for the segmented image is calculated. The resultant LBP image is represented as a grid of specific cell size. For each cell in grid, a feature vector of size 59 is obtained by calculating the histogram of uniform patterns. The obtained feature vector from each cell is given to SVM as input. The classifier assigns a single value to each cell of the grid. The number of positive regions in the grid is then compared with a specific threshold. If the value is above threshold the classifier classifies the image as cancerous otherwise normal. For the evaluation of proposed model 17,639 images are extracted from DDSM (Heath et al., 1998, 2000) and MIAS (Suckling, 1994) datasets. The best accuracy achieved is 84% at threshold 0.5.

In (Taheri, Hamer, Son, & Shin, 2016), the authors have used SVM for the binary classification of mammograms (i.e. in normal and abnormal). A dataset of 600 mammographic images (i.e. 200 for training and 400 for testing) is used for the training and evaluation of classification model. Images are first preprocessed, during which noise and unwanted artifacts are removed from images and contrast of images is enhanced using histograms equalization. Resultant images are then resized to 256*256 pixels. In the next step, intensity, energy and correlation based features (i.e. using Harris corner detection (HCD)) are extracted from the preprocessed images. An automated threshold

Table 4
Traditional machine learning models.

Paper	Image Modality	Classification type	Preprocessing	Segmentation	Features	Classifier	Dataset	Dataset Size	Evaluation Metrics	Regularization	Learning Type
(Huynh et al., 2016)	Mammography	Binary Classification (malign and benign)	(0–1) normalization	region growing algorithm	size, shape, intensity, margin features	SVM	Local Dataset	607	AUC = 0.81,	–	Supervised
(Kral & Lenc, 2016)	Mammography	Binary (Cancerous and normal)	–	Otsu's Thresholding	LBP (Texture) vector size = 59	SVM	DDSM and MIAS	17,639	Accuracy = 84%	–	Supervised
(Taheri et al., 2016)	Mammography	Binary classification (normal and abnormal)	Adaptive Median Filter, Histogram Equalization, resizing	–	Intensity, Energy, Correlation (using Harris corner detection) features	SVM	–	600 (100 for training, 400 for testing)	Precision = 96.8 Recall = 92.5	–	Supervised
(Arafa et al., 2019)	Mammography	Binary (benign, malignant) Multi (normal, benign, malignant)	Noise and Pectoral Muscle removal	GMM	texture, shape and statistics features	Non-linear SVM (RBF kernel)	MIAS	Training = 30, testing = 60	Accuracy = 92.5% Accuracy = 90%	–	Supervised
(Vijayarajeswari et al., 2019)	Mammography	Binary (normal and abnormal)	Pectoral Maximization method	Masking using Gradient based thresholded mask	Hough Transform (mean, variance, entropy and standard deviation)	SVM	MIAS	95	Accuracy = 94%	–	Supervised
(Htay & Maung, 2018)	Mammography	Binary classification (normal and abnormal)	Median filter, cropping	Otsu's Thresholding	8 Texture features	KNN (k = 1)	MIAS	120	Accuracy = 92%	–	Supervised
(Diaz et al., 2019)	Mammography	Binary classification (benign and malignant)	–	–	6-first order statistical features	KNN (K = 15 w/0 CV) (K = 5 with CV)	MIAS	110	Accuracy = 91.8%	–	Supervised
(Khoulqi & Idrissi, 2019)	Mammography	Binary classification of Micro-calcification (benign and malignant)	Median filter, Michelson's contrast	Top-hat, Watershed, Split and Merge	Wavelet	KNN	MIAS	–	Sensitivity = 100 Specificity = 100	–	Supervised
						SVM			Sensitivity = 100 Specificity = 100		
						DT			Sensitivity = 75 Specificity = 100		
					Tamura	KNN			Sensitivity = 100 Specificity = 100		
						SVM			Sensitivity = 100 Specificity = 75		
						DT			Sensitivity = 83 Specificity = 100		
(Ibraheem et al., 2019)	MRI	Normal and Abnormal	2-D Median filter, resizing	–	Discrete wavelet	SVM	RIDER Breast MRI database	4 cases (153 samples)	Specificity = 96.72%, sensitivity = 96.42%, accuracy = 98.03%	–	Supervised

(continued on next page)

Table 4 (*continued*)

Paper	Image Modality	Classification type	Preprocessing	Segmentation	Features	Classifier	Dataset	Dataset Size	Evaluation Metrics	Regularization	Learning Type
(Lu et al., 2017)	MRI	Normal and Abnormal	–	–	Haralick, morphological, GLRLM, GLDM, Gabor, GGLCM	Ensemble of C45	Proprietary dataset	2336	Sensitivity = 90.19, Specificity = 96.31, AUC = 0.9617	–	Supervised
(Conte et al., 2020)	DCE-MRI	infiltrative and in-situ regions	–	ROI hunter mechanism, GoogleNet followed by ANN	1820 handcrafted features	XGBoost	Proprietary dataset	55 scans	AUC = 0.70	–	Supervised
(Soliman et al., 2019)	Infrared	Normal and Abnormal	Median filter, Gaussian filter	Canny edge detection, dilation, HPP	1st and 2nd order statistical features, GLCM	Back propagation neural network	UFF university dataset	63	Accuracy = 96.51%, Sensitivity = 79.7%, Specificity = 98.25%	–	Supervised
(Abdel-Nasser et al., 2019)	Infrared	Normal and Cancerous	–	–	HOG 4*4, LTR	MLP	DMR-IR	56	AUC = 0.989 Accuracy = 0.958	–	Supervised
(Sakai et al., 2020)	Tomosynthesis	Benign and Malignant	Normalization	ROI extraction	72 shape, specula and texture features	SVM	Aoyama Hospital data + public data	28 + 23	Accuracy = 72.54 AUC = 0.798	–	Supervised
						Random forest			Accuracy = 70.59 AUC = 0.757		
						Naïve Bayes			Accuracy = 60.78 AUC = 0.648		
						MLP			Accuracy = 70.59 AUC = 0.754		

selection technique is used during the selection of corner pixels using HCD technique. The final set of extracted features is then fed to SVM for its training. This model has achieved a precision of 96.8 and recall of 92.5 over 400 mammographic images test dataset.

In (Jothilakshmi & Raaza, 2017), the author has proposed a CAD system for binary classification of breast masses (i.e. in benign and malignant) using breast mammographic images. The proposed system is based on SVM classifier, which is trained over 50 images extracted from MIAS dataset. To remove noise from images median filter is applied and to remove unwanted artifacts thresholded mask is applied on input images, during preprocessing stage. To extract regions of interest segmentation is done using split and merge region based segmentation technique. Thirteen texture based features using gray level co-occurrence matrix (GLCM) are extracted from images and fed to classifier for training. An estimated 94% accuracy is achieved by the proposed methodology.

Another statistical ML classifier based methodology has been proposed by Arafa, El-Sokary, Asad, and Hefny (2019), who have used non-linear SVM classifier for multiclass classification of breast mammographic images (i.e. in normal, benign and malignant). A total of 90 images are used for the training and evaluation of proposed methodology, which are taken from Mini-MIAS dataset. Unwanted noise artifacts and removal of pectoral muscle from training images is done during preprocessing stage for which an automatic pectoral muscle removal technique is proposed. Subsequently, Gaussian Mixture Model (GMM) (Reynolds, 2009) is applied for ROI extraction (i.e. normal, benign and malignant regions). From the resultant segmented ROIs fourteen features are extracted (i.e. shape, statistical and texture based features) and fed to the classifier for training. An accuracy of 90% is achieved using proposed methodology.

For the binary classification of breast mammography images such that in normal and abnormal, in his study (Vijayarajeswari, Parthasarathy, Vivekanandan, & Basha, 2019) the author has proposed a methodology based on SVM classifier. For the training of classifier 95 images are taken from Mini-MIAS (Suckling, 1994) dataset with fatty glandular breast density. During preprocessing stage, masking is done to remove unwanted artifacts and maximization method is applied to remove pectoral muscle from mammograms. On the resultant images canny edge detector followed by Hough transform is applied and subsequently features are extracted. Four types of intensity based features are extracted (i.e. mean, entropy, standard deviation and variance), which are then fed to the classifier for training. An accuracy of 94% is achieved from the proposed approach.

Htay and Maung (2018), in their study, proposed a CAD system for the early detection of breast cancer using mammograms. The system classifies mammographic images into normal and abnormal using KNN classifier. It is evaluated on 120 images taken from MIAS dataset (Suckling, 1994). In preprocessing of input images the unwanted parts of images such as noise and labels are removed using median filter and masking. In the next step, the suspicious mass of the image is segmented by applying Otsu's thresholding. From segmented images, first and second order statistical features are extracted for texture analysis. The extracted features are then fed to KNN for classification. An accuracy of 92% is achieved through this methodology.

(Diaz, Swandewi, & Novianti, 2019) have used KNN classifier for the binary classification of breast mammographic mass images (i.e. in benign and malignant). A total of 110 mass images extracted from Mini-MIAS (Suckling, 1994) dataset are used for the training of classifier in proposed methodology. From each of the confiscated image, six first order texture features (i.e. mean, standard deviation, smoothness, third moment, uniformity and entropy) are extracted and normalized. Extracted features are then fed to KNN for classification. The best accuracy achieved using proposed methodology is 91.8%.

For the detection and binary classification of breast micro-calcifications (i.e. in benign and malignant), in (Khoulqi & Idrissi, 2019) authors have proposed a CAD system using a voting classifier (i.e. based upon KNN, SVM and Decision Tree (DT)). These models are trained over mammographic images taken from MIAS dataset (SUCKLING, 1994). During preprocessing phase, noise is removed using median filtering and contrast of resultant images is enhanced using Michelson's contrast technique. Three types of segmentation techniques are applied over the preprocessed images which are: top hat transformation of mathematical morphology, watershed method under constraint of markers obtained from white hat method and split and merge approach with homogeneity criteria based upon mean, variance and uniformity, where top-hat method has performed best among all others. From the segmented images features such as LBP, Tamura, Wavelet and Discrete cosine transform (DCT) are extracted. Different combination of classifiers and extracted features are tested during experimental phase, where KNN has provided maximum accuracy in majority cases.

An approach for the automated diagnosis of breast cancer using MRI images, has been proposed in (Ibraheem et al., 2019). A publically available breast MRI images dataset is utilized for the validation of proposed methodology, where images of dataset are first preprocessed during which median filter is applied to remove noise and images are resized to 256*256 pixels. Subsequently, discrete wavelet features are extracted from the pre-processed images and PCA is applied for features reduction. Resultant features are then fed to SVM for the classification of images into normal and abnormal.

For the automated classification of breast tumor areas (i.e. in benign and malignant) an automated approach has been proposed in (Lu, Li, & Chu, 2017). Author has utilized a locally collected breast MRI imaging dataset consisting of 1898 negative and 438 positive samples for the validation of proposed methodology. Various types of handcrafted features (i.e. including Haralick, morphological, GLRLM, GLDM, Gabor and GGLCM features) are extracted from the input MRI images. To improve classification accuracy and to reduce computational complexity "Relief" (Kira & Rendell, 1992) features selection method has been employed. To solve imbalanced data problem ensemble method and to reduce generalization error Adaboost has been implemented, while C45 classifier has been used as a base classifier. The proposed methodology has achieved an AUC of 0.9617.

In (Conte, Tafuri, Nunzio, Portaluri, & Galiano, 2020), an automated approach has been proposed for the discrimination of infiltrating and in-situ breast tumors. The proposed methodology is mainly divided into two stages i.e. detection of breast area using intensity based iterative ROI hunter mechanism (i.e. a bounding box is drawn around breast area) followed by false positive reduction mechanism, in which features are extracted using pre-trained Google Net architecture by employing sliding window approach (i.e. window size is set to 30*30 on the basis of length of lesion edge). Extracted features are then fed to a binary classifier specifically artificial neural network for false positive reduction (i. e. patch with less than 10% tumor area is excluded). For the characterization of extracted positive patches (i.e. in infiltrative and in-situ regions) radiomics features extraction is done (i.e. 1820 handcrafted features including shape, first and higher order statistical features, GLCM, GLRLM etc are extracted), which are subsequently fed to Extreme gradient boosting (XGBoost) classifier for classification. The proposed methodology has achieved an AUC and ROC curve of 0.70 by validating the proposed methodology over a DCE-MRI dataset of 55 scans. In (Soliman, Sweilam, & Shawky, 2019), an automated approach has been proposed for breast cancer detection, that is validated over a publically available thermal imaging dataset (Silva et al., 2014). During pre-processing stage color input images are converted into grayscale, median filter is applied for noise removal, Gaussian filter is applied for image smoothness and canny edge detector is applied for the detection of outer boundary of breast, while for further processing detected boundary is dilated. To detect the lower border of breast inframammary line is detected using horizontal projection profile (HPP) approach. Left and right breast are separated by finding bifurcation point which is the intersection of the two inframammary curves. Subsequently, features are extracted from segmented image using first and second order statistical measures. To examine the spatial relationship of pixels GLCM is computed in four directions. Then average value of the features of four GLCMs is computed. The normalized absolute difference of the features of left and right breast is computed to measure the asymmetry between them. The resultant feature vector is then passed to back propagation neural network for classification, which has achieved an accuracy of 96.51%.

For the differentiation of normal and cancerous breast, a novel approach has been proposed by (Abdel-Nasser, Moreno, & Puig, 2019), in which six features extraction methods have been tested (including GLCM, LBP, lacunarity analysis of vascular networks (LVN), gabor filters (GF), local directional number pattern (LDN) and histogram of oriented gradients (HOG)) for the texture analysis of input infrared images. Subsequently, a learning to rank (LTR) method has been employed to learn extracted feature space from the sequence of thermograms taken to analyze temperature changes among them and to produce a compact feature's representation from them. Resultant feature vector is then fed to multi-layered perceptron for the classification of images. The proposed methodology is validated over publically available DMR-IR dataset (Silva et al., 2014). Experimental results shows that HOG features together with LTR method has performed best among other methods i.e. achieved an AUC of 0.989. An automated radiomics features based approach has been proposed in (Sakai et al., 2020), for the classification of benign and malignant breast tumors using breast CT images. Author has utilized two datasets i.e. a locally collected dataset and a publically available breast tomography dataset for the validation of proposed methodology. Images of both datasets are first normalized, to reduce the difference in their intensity values; subsequently region of interest is extracted from input images. From the preprocessed input images 70 radiomics features are extracted (i.e. for shape, specula and texture analysis) and fed to ML classifier for the classification of images (i.e. in benign and malignant). Four types of classifiers are evaluated over extracted features (i.e. random forest, MLP, SVM, Naïve Bayes), however SVM has performed best and achieved an accuracy of 72.54%.

6.4. Discussion

Many authors have used supervised statistical ML based classifiers for the diagnosis of breast cancer using hand-driven features (See Table 4). However majority of them have exploited breast x-ray mammography imaging repositories for the validation of their proposed methodologies, as x-ray mammography is an efficient imaging modality that helps in early stage diagnosis of breast cancer. However, only a few studies have focused upon other modalities (i.e. MRI, tomosynthesis, infrared or thermography). Most of the studies have focused upon binary classification of mammograms (i.e. classifying images as normal and abnormal) or on categorization of masses as benign and malignant, where only (Khoulqi & Idrissi, 2019) have considered upon categorization of micro-calcifications and (Conte et al., 2020) have worked upon the classification of infiltrative and in-situ regions.

However none of the author has contemplated upon the sub-categorization of benign and malignant masses and in detection of other breast abnormalities such as architectural distortions. Due to the convenience of some publically accessible datasets, they are utilized in majority of the studies especially MIAS dataset. As these datasets incorporates noise, superfluous artifacts and low contrast mammographic images, which may effects classification outcomes, therefore many of the studies have employed different preprocessing techniques to tackle these problem. In addition to these, for the withdrawal of efficient features from MLO views of x-ray mammograms, it is essential to eradicate pectoral muscle area from them in preprocessing stage, as there is a significant resemblance in intensity values of tumor and pectoral muscle area (Vijayarajeswari et al., 2019), despite of this fact some studies such as (Htay & Maung, 2018) have not done this task.

To classify mammograms as normal and abnormal, features are extracted from entire mammogram area, which is extracted using some semi-automated segmentation technique i.e. Thresholding, region growing, split and merge, Canny edge detection etc, however efficiency of these techniques based upon the tuned parameters (i.e. threshold), which needs to be manually fine-tuned. Each one of the discussed studies have utilized hand-crafted features (i.e. texture, intensity, morphological etc) for the classification of mammograms, while many of the studies such as (Dhungel, Carneiro, & Bradley, 2016) has manifested that usage of deep features for the classification of abnormalities in mammograms have outperformed classical hand-driven features. Majority of the datasets used for the training of classifiers in proposed methodologies are of limited size. Moreover none of the data extension technique (i.e. data augmentation) has been employed to tackle with limited data problem in any of the mentioned studies, so the model could easily memorize the training data and cause over-fitting.

Overall admissible accuracies have been achieved by all the classifiers used in studies under consideration; however SVM classifier is most widely used for the binary categorization of breast mammograms, while from the comparative study presented in (Khoulqi & Idrissi, 2019) it could be inferred that KNN depicts more accurate results as compared to other classifiers but takes more training time. However, according to the results depicted in (Sakai et al., 2020), SVM has performed best as compared to RF, NB and MLP.

6.5. DL models

6.5.1. Models built from scratch

A new deep CNN model is proposed by (Jiao, Gao, Wang, & Li, 2016) for features extraction from breast mammographic mass images to binary classify them (i.e. in benign and malignant). The proposed model is first trained over ImageNet (Deng et al., 2010) dataset, then subsequently fine-tuned and tested over the 600 mammographic mass images taken from DDSM database. Extracted images are first normalized and then PCA image whitening technique is applied on them. Afterwards, data augmentation technique is applied on enhanced images. Resultant images are fed to the proposed CNN for features extraction, from which two types of features are extracted (i.e. high level and middle level features) on which two linear SVM classifiers are trained individually. During testing, same example is passed through both classifiers, if they show consistent results; the corresponding instance is added to first result list otherwise Euclidian distance of test example is calculated with specific labeled benign and malignant examples of training data. Based on the calculated distance that particular instance gets classified. The proposed model has achieved a classification accuracy of 96.7%.

A 7-layered CNN model has been created by Guan et al., 2017, for the classification of mammograms into benign and malignant. MIAS dataset is used for the training and evaluation of proposed model, where for each mammogram in dataset 95 ROIs are randomly selected and divided into training and testing dataset in a ratio of 15:4. Extracted ROI are then fed to the proposed model from which a maximum accuracy of 0.751 is achieved by training the model for 500 epochs.

In (Platania et al., 2017), author has proposed a deep CNN inspired from YOLO (Redmon, Divvala, Girshick, & Farhadi, 2016) architecture for the detection and binary classification of breast masses (i.e. in benign and malignant) and trained it over the mammographic images taken from DDSM database. Noise removal, background removal and normalization of images are done during preprocessing stage. On the resultant images data augmentation technique (i.e. random rotation) is applied and subsequently mass patches are extracted (i.e. using bounding box) and resized. Training of proposed model is done in two steps; such that at first model is trained over mass patches only, where in second stage it is trained over all patches extracted from whole mammograms. The weights learned by CNN in initial training process are extracted and used to initialize the weights of modified CNN used in main training phase (i.e. weights of model are tuned only if there is an abnormality found in the patch). The proposed framework achieved a classification accuracy of 93.5% and a detection accuracy of 90%.

Charan et al., 2018, proposed a 9-layered CNN for the classification of breast mammograms and trained it over MIAS dataset. Images of training dataset are first preprocessed i.e. morphological closing is done for noise removal and masking is performed for ROI segmentation. The resultant images are then resized and fed to the proposed model, which has obtained 65% accuracy, while performing two types of classifications (i.e. binary and multi-classification).

Another new CNN is proposed by Li et al. (2018) for the binary classification of breast mammographic images. Author has used abnormal images extracted from MIAS dataset (116 images) for the training of proposed model, which are first cropped to remove unnecessary background and unwanted noise, and then global contrast normalization is applied on them. Afterwards, ROIs are extracted from the resultant images using bounding box and data augmentation technique is applied on them to avoid overfitting and for dataset balancing. The resultant data is then fed to the proposed CNN for experimentation i.e. to determine impact of depth of CNN and adaptation of dropout over the generalizability of model. Results show that employing dropout strategy and usage of smaller filter size in deep CNNs improves classification performance.

In another recent study (Arslan, Yasar, & Colak, 2019), the author have proposed a web based CAD system for binary classification (i.e. in benign and malignant) of breast histopathological images using deep CNN. The proposed model is trained over a publically available microscopic breast images dataset namely BreaKHis dataset. Input images are first resized to 192×192 pixels and then fed to proposed 7-layered deep CNN architecture. The model is trained over 6327 and the tested over 1582 instances of dataset and achieved 99% of training and 91.4% of testing accuracy respectively.

For effective early breast cancer detection using mammograms, (Rashed, Seoud, & M, 2019) has proposed a new deep encoder-decoder CNN architecture that is inspired by U-Net architecture (Ronneberger, Fischer, & Brox, 2015). The proposed model is trained individually over two subsets of CBIS-DDSM database (Lee et al., 2017) i.e. one for micro-calcification detection and other for masses detection. A total of 692 masses and 603 micro-calcification subjects of size 1024*1024 are extracted from this database for the training and evaluation of proposed architecture. To improve contrast histogram equalization is done of input images and data augmentation technique is applied on resultant images (i.e. images are randomly rotated). The preprocessed images are then fed to 23 layered proposed CNN for training. The model has achieved classification accuracy of 94.31% and 95.01% during micro-calcifications and masses detection respectively.

For the classification of mammograms Wang et al., 2019, extracted both subjective and objective features. The author has used a proprietary image dataset (i.e. consists of 400 mammograms) for training of proposed model. For noise removal and contrast enhancement between masses and its surroundings, adaptive mean filter is applied on input images. Afterwards, ROIs are extracted using bounding box and sub-divided into 48*48 sub regions. These sub regions are then fed to a 7 layer CNN for feature extraction, which results a feature-vector of length 24. The extracted features are then fed to Unsupervised Extreme Learning Machine (US-ELM) algorithm for clustering the regions into suspicious and non-suspicious. The extracted suspicious regions are then fed to a 10 layer CNN for extracting deep features. Moreover, 5 morphological, 5 texture and 7 density based features are also extracted from these suspicious areas. All of these features are then given to ELM classifier for classification. During experimentation, results of SVM and ELM are compared. Result shows that: classification accuracy of ELM with all four types of features is better than SVM.

A deep learning architecture has been proposed in (Mishra, Prakash, Roy, Sharan, & Mathur, 2020) for the automated classification of normal and cancerous breast thermograms. Author has utilized a publically available breast thermograms DMR-IR dataset (Silva et al., 2014) for the training and validation of proposed model. Input images are first preprocessed (i.e. RGB to gray, normalization and resizing is done) and then further fed to the proposed network (i.e. consisting of 5 2D-convolution, 3 max-pooling, 2 fully connected, batch normalization and dropout layers) for classification. The proposed network has achieved an accuracy of 95.8%.

7. Discussion

Various new models (i.e. CNNs) have been proposed that assists in automated breast cancer diagnosis, where majority of them are intended to recognize abnormalities (i.e. benign and malignant masses) found in breast x-ray mammographic images as compared to images of other modalities. The performance of these models is highly influenced by the quality and quantity of training data i.e. generalizability of models could be improved by training them over large scale datasets (Arslan et al., 2019). Therefore the presence of large amount of breast images data is obligatory for efficient learning of these models. Just like statistical classifiers most of the authors have used publically available breast imaging data resources for the training of their proposed deep architectures. However, majority of the state of the art publically available databases are inadequate (i.e. MIAS, INbreast) for the efficient training of these models and training over such small size datasets may raise problems, like overfitting. Despite of that, many of the proposed deep models have performed well even on small datasets by employing data augmentation (i.e. to extend size of dataset) and regularization (i.e. to avoid overfitting) techniques as depicted in (Li et al., 2018) and (Rashed et al., 2019). Majority of the newly proposed models are intended to binary classify the abnormalities except a few one such as (Charan et al., 2018), who have adopted a two step methodology (i.e. first classify mammograms in normal and abnormal, secondly further classification of abnormalities) for the recognition of multiple abnormalities in mammograms, however this study have not achieved significant performance, due to multiple reasons such as mammograms have been processed without removal of pectoral muscle from them, size of dataset is limited, number of training iterations are small, no data extension and regularization technique have been employed etc. On the other hand (Wang et al., 2019) have also adopted a similar two step methodology but have attained valuable results by utilizing combination of high and low level features. Hence it could be said that training models over low level features in addition to high level features improve their performance. To deal with limited data problem and to improve the authenticity of his proposed methodology, (Jiao et al., 2016) have first trained his proposed model over ImageNet dataset and then fine-tuned it over the breast image dataset and ultimately achieved good results. Another major finding can be observed from the study (Li et al., 2018), in which author has conducted several experiments to visualize the impact of depth of deep models over their performance, where the results demonstrate that the escalation of number of layers and by adaptation of smaller filter size (i.e. extracting low level details), boosts the performance of model. Meanwhile adaptation of suitable regularization technique improves its generalizability. (See Table 5)

7.1. Transfer-learning based models

7.1.1. Weightless models

In (Brancati, Frucci, & Riccio, 2018), authors have compared three different ensemble classifiers of ResNet-34/5-/101 i.e. by training it from scratch, by fine-tuning full trained model and by fine-tuning only last layer of trained model. In first case, the ensemble classifier is taken without pre-trained weights and trained over microscopic breast image dataset provided by BACH-2018. Before training, patches are extracted from images and data augmentation is applied over them (i.e. vertical and horizontal flipping, and rotations of 90 degrees and image resizing). The resultant data is then fed to the proposed ensemble classifier. The proposed model has achieved a validation accuracy of 97.3% and a test accuracy of 86% over the provided dataset by BACH.

Table 5
Models built from scratch.

Paper	Image Modality	Classification type	Preprocessing	Segmentation	Features	Classifier	Dataset	Dataset Size	Training Parameters	Evaluation Metrics	Regularization	Learning Type
(Wang et al., 2019)	Mammography	Binary classification (benign and malignant)	Adaptive mean filter algorithm, Contrast enhancement	Bounding Box + US-ELM clustering of suspicious features	(CNN1 conv = 6, pool = 6, FC = 1) (CNN2 conv = 9, pool = 9, FC = 1) + morphological (5) + texture (5) + density (7)	ELM	Local	400	Kernel = RBF, c = 0.5, and g = 0.0206	Accuracy = 86.50 Sensitivity = 85.10 Specificity = 88.02 AUC = 0.923	×	Extreme
(Charan et al., 2018)	Mammography	Multi-class	Thresholding	Masking	Deep features	CNN (Conv = 6, Pool = 4, FC = 3)	MIAS	322	LR = 0.01, BS = 5, epochs = 50, LR drop factor = 0.2, LR drop period = 0.5	Accuracy = 65%	×	Deep
(Guan et al., 2017)	Mammography	Binary Classification (malign and benign)	Resizing, Grayscale to RGB	bounding box	Deep features	CNN (Conv = 3, Pool = 3, FC = 1)	MIAS	322	Opt = RMSProp, loss = BCE, BS = 15, epochs = 500	Accuracy = 0.751	Dropout	Deep
(Li et al., 2018)	Mammography	Binary Classification (malign and benign)	Cropping, Global contrast normalization, Data augmentation, Local Histogram Equalization, Data Balancing	Bounding box	Deep Features	CNN (Conv = 4, Pool = 2, FC = 2)	MIAS	Original = 116, Augmented = 547	AF = RELU, LR = 0.001, Opt = SGD, RMSProp, Dropout = 0.7	Without Dropout Train.Acc = 98.90% Test.Acc = 86.13% With Dropout Train.Acc = 93.43%Test. Acc = 88.05%	Augmentation, Dropout	Deep
(Jiao et al., 2016)	Mammography	Binary Classification (malign and benign)	PCA whitening, (0–1) normalization, data augmentation	–	CNN (Conv = 5, Pool = 3, FC = 3)	SVM	(Training) ImageNet (Fine tuning) DDSM	Train = 300, Test = 300	Opt = SGD, Loss = Cross Entropy, LR = 0.01 LR = 0.00001	Accuracy = 96.7%	Augmentation	Hybrid
(Rashed et al., 2019)	Mammography	Binary classification (Normal or Mass) Binary classification (Normal or Micro-calcification)	Histogram Equalization, Data augmentation	–	Deep Features	O-Net CNN (Conv = 16 De-conv = 10, Pool = 6, FC = 3)	CBIS-DDSM mass images CBIS-DDSM micro-calcification images	692 603	Opt = ADAM, epoch = 50	Accuracy = 95.01% Accuracy = 94.31%	Augmentation	Deep
(Arslan et al., 2019)	Histopathology	Binary (benign and malignant)	Resizing	–	Deep Features	CNN (Conv = 4, Pool = 4, FC = 1)	BreaKHis	7909	Opt = ADAM, LR = 0.001	Training = 99%, Testing = 91.4%	–	Deep
(Platania et al., 2017)	Mammography	Binary (benign and malignant)	Noise and Background removal, Normalization	Patch extraction + resizing (128*128)	Deep Features	CNN (Conv = 9, Pool = 6, FC = 3)	DDSM	Original = 10,480, Augmented = 25,000	BS = 16, epochs = 100+, LR= (10^{-3}, 10^{-4}, 10^{-5})	Accuracy = 93.5%	Augmentation	Deep

Another transfer learning work is done by (He, Ruan, Long, Wang, & Dataset, 2018) where authors have compared different TL approaches (i.e. used model architecture without pretrained weights and fine-tuned model for features extraction) for the binary classification of patches extracted from breast histopathological WSIs. In first case, the architecture of AlexNet and GoogLeNet is trained from scratch on the 5000 patches extracted from locally collected WSIs dataset. Before feeding the patches to networks, data augmentation is applied to extend dataset size. Final accuracy achieved by AlexNet is 93.40% and by Google's Net is 94.12%.

In (Alom, Yakopcic, Nasrin, Taha, & Asari, 2019), author has used Inception recurrent residual convolutional neural network (IRRCNN) (Alom, Hasan, Yakopcic, Taha, & Asari, 2018) for the binary and multi-class recognition of breast cancer through breast tissue histopathological images. The main motive of IRCNN is to introduce an improved hybrid model that provides better performance by combining the capabilities of residual networks (He et al., 2016), recurrent convolutional neural network (RCNN) (Liang & Hu, 2015) and inception networks (Szegedy, Ioffe, Vanhoucke, & Alemi, 2017), while using less or same number of network parameters. The performance of model is tested over two publically available microscopic breast image datasets i.e. BreakHis (Spanhol et al., 2016) and BioImaging-2015 (Pego & Aguiar, 2015). Data augmentation technique is applied over both datasets to increase their size. Resultant images are then fed to IRRCNN model for training. Two different models are trained individually for the two mentioned datasets. For the performance evaluation of proposed strategy two different performance criteria are tested (i.e. patient level and image level performance).

In (Lu, Loh, & Huang, 2019), author has used transfer learning approach for detection of benign and malignant breast tumors i.e. author has utilized model architecture proposed in (Geras, Wolfson, Shen, Wu, Kim, Kim, Heacock, Parikh, Moy, & Cho, 2017) and trained it over a proprietary dataset collected from a teaching hospital in Taiwan, that consists of 9927 images with resolution 2294*1914 pixels. Images are pre-processed for contrast stretching and noise removal. To solve imbalanced data problem and to improve generalizability; data augmentation technique is applied on the preprocessed dataset. Resultant data is then fed to CNN for validation. The model achieves an accuracy of 0.82, sensitivity of 0.91, specificity of 0.57 and F1 score of 0.88.

In another study (Wang, Qin, et al., 2020), author has proposed a 3D-Inception UNet architecture in combination with a deep fusion supervision mechanism for the automated detection of breast tumors using volumetric breast ultrasound images of 196 patients (i.e. 559 cancerous samples). Author has embedded two types of 3D-inception blocks to deal with large number of network parameters and large computation problem faced in traditional 3D CNNs. Moreover, to avoid parameter tuning of large number of hyper-parameters deep supervision mechanism is employed, in which output of previous deep supervision block and is concatenated with feature map of corresponding up-sampling level of UNet. In addition to these an asymmetric loss has been defined to balance false positive and false negative regions. The proposed methodology has achieved a FP of 3.0 and sensitivity of 95% per BUS volume.

For the automated detection of breast tumors from DEC-MRI images, a deep learning based approach has been proposed in (Benjelloun, El Adoui, Larhmam, & Mahmoudi, 2018). The author has utilized traditional UNet architecture for tumor area segmentation, while collected a dataset consisting of 86 MRI volumes (i.e. 5452 slices) for the training and validation of proposed methodology. The model has achieved an IoU of 76.14%.

A comparative analysis of several state-of-the-art deep segmentation architectures (including VGGNet (Simonyan & Zisserman, 2015), InputCascadeCNN (Havaei et al., 2017), UNet (Ronneberger et al., 2015) and VNet (Milletari et al., n.d.)) has been done in (Kakileti, Dalmia, & Manjunath, 2019). For the training of these models, an anonymised thermal imaging dataset of 180 subjects has been collected. To remove background area from images Otsu's thresholding technique has been applied while to tackle limited data problem and to avoid overfitting, data augmentation technique has been applied. Additionally, batch normalization, L1 and L2 regularization techniques has also been applied. Evaluation metrics including accuracy, dice coefficient and jaccard index are utilized for the evaluation of mentioned models, while VNet has performed best among mentioned models i.e. by achieving an accuracy of 99.6.

In (Caballo, Pangallo, Mann, & Sechopoulos, 2020), author has proposed an automated approach for the detection and discrimination of benign and malignant breast tumors using breast CT images. A proprietary breast CT imaging dataset collected from 69 patients is utilized for the training of proposed architecture. To overcome limited data problem and to maximize the performance of deep learning model three different types of data augmentation techniques have been applied i.e. by extracting patches from other CT views (including sagittal and axial views in addition to coronary view and from planes of symmetry), by applying traditional data augmentation techniques (i.e. rotation, mirroring and shearing etc) and lastly by employing generative adversarial network (GAN), which results in to a total of 34,992 patches. The extracted patches are then fed to state-of-the-art deep network namely UNet for tumors segmentation. From the segmented tumor area radiomics features (i.e. 18 shape features, 327 texture and contour based features) are extracted to measure the stability performance in classification of benign and malignant tumors using multivariate analysis of variance (MANOVA) and inter class correlation (ICC). The proposed methodology has achieved an efficient precision of 0.93.

7.1.2. Fine-tuned models

In another study conducted by (Chang J, Yu J, Han T, Chang HJ, Park E. A method for classifying medical images using transfer learning: a pilot study on histopathology of breast cancer. In2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom) 2017, 2017), authors used a pre-trained CNN; **Google's Inception v3 model** for the binary classification of histopathological images i.e. malignant and benign. The pre-trained model is fine-tuned over microscopic images of BreaKHis dataset (Spanhol et al., 2016). During pre-processing stage, data augmentation technique is applied i.e. images are rotated by 90, 180, 270, mirrored, randomly distorted and added back to the original dataset. The augmented dataset is then classified by the deep convolutional neural network (Google's Inception v3). The final accuracy achieved by the model is 0.89 at 500 training steps.

In (Yemini et al., 2019), the author has proposed a variation of Google's Inception-v3 network for binary classification of mammograms. Google Inception-V3 architecture is a pretrained deep CNN on ImageNet dataset (Deng et al., 2010). The proposed model is fine-tuned over publicly available mammography dataset INbreast (Moreira et al., 2012). Different data augmentation techniques such as image flipping, contrast and brightness changing, and random addition of Gaussian noise are applied, to solve the problem of insufficient data. For data balancing, two tasks are performed. In first approach dataset is artificially expanded by generating artificial mammograms. To do so, malignant regions from positive cases are extracted from random locations and inserted into mammograms of negative cases after rotation at random angles. In the second approach, dataset is balanced by simply the augmentation of malignant cases. Augmented data is then resized to 299*299 in order to feed in Inception-V3 architecture. As ImageNet and INbreast are different datasets, to make the model accurate Inception-V3 is chopped from middle layer and several new layers are added to it. The proposed model achieves AUC of 0.78 by artificially expanding dataset and 0.86 when the malignant cases are increased by augmentation.

Another CNN namely Inception-Resnet-v2 (Längkvist, Karlsson, & Loutfi, 2014) pretrained over ImageNet dataset (Deng et al., 2010) has been utilized by Kwok (2018)for the classification of breast tissue histological images. For fine-tuning of model, 5600 patches are extracted from provided BACH-2018 breast microscopic image dataset. Image resizing and data augmentation technique is then applied over extracted patches and fed to the model for training. Afterwards patches are extracted and scaled from WSIs provided by BACH challenge. The newly generated WSI patches are then passed through previously trained model on microscopic images. From the distance between WSI images predictions and their ground truths, difficulty of patches is computed. The patches with less than 70% of foreground were then discarded and the top patches are sorted according to their difficulty level. The patches with top 40 percentile of difficulty are selected as candidates. The model is retrained on microscopic patches and resultant WSI patches. The results of patches were aggregated from patch-wise predictions back onto image-wise prediction and WSI annotations. It achieved an accuracy of 87% on microscopic image and a score of 0.6929 on WSI images.

A pretrained convolutional neural network VGG16 (Simonyan & Zisserman, 2015) is used by Wang et al. (2018) for classification of H&E-stained breast histological images into four classes. Model is fine-tuned over microscopic breast image dataset provided by ICIAR-2018. To generate sufficient data for the training of CNN, data augmentation technique is implied. Two types of dataset augmentations are done here: 1) Primary dataset augmentation, 2) Sample Pairing. In the proposed methodology, images of the provided dataset are first resized to 256*256 from which random patches of size 224*224 are extracted and then further augmentation techniques are implemented. Model is first fine-tuned over images generated by sample pairing and then on dataset generated by primary data augmentation. The proposed approach achieved an accuracy of 83% on the microscopic test images provided for first task of Grand Challenge.

Koné & Boulmane, 2018, used a hierarchal architecture, for multi-classification of histopathological images. The hierarchy is based on three pretrained ResNetXt50 models (Xie et al., 2017). The top model classifies images into carcinoma and non-carcinoma. The second model further classifies carcinoma images into in situ and invasive carcinoma. The third model classifies non-carcinoma images into normal and benign. Besides the datasets provided by ICIAR, the author has also used Bio-Image Semantic Query User Environment (BISQUE) (Gelasca, Byun, & Boguslaw Obara, 2008) dataset for training. ResNetXt50 is a pre-trained on ImageNet (Deng et al., 2010) dataset, where in this study it is further fine-tuned on dataset provided by BACH and on BISQUE dataset. Before training, the images of dataset are resized to 299*299 and different augmentation techniques are applied on data. During training different learning rates are set for input output and middle layers. The final system gives an accuracy of 81% on test data provided by BACH.

Kohl, Walz, Ludwig, Braunewell, & Baust, 2018, have used transfer learning based pretrained DenseNet (G. Huang et al., 2017) for the multi-classification of breast microscopic images. Model is pretrained over on ImageNet (Deng et al., 2010) dataset and fine-tuned over data provided by BACH-2018. During preprocessing data augmentation technique is applied. The augmented data is then normalized and fed to Dense Net for training. The model is trained in two steps: 1) Fully connected layer is fine-tuned for 25 epochs to avoid over-fitting at a learning rate of $1*10^{-3}$, 2) Whole network is fine-tuned for 250 epochs at a learning rate of $2*10^{-4}$. The proposed architecture achieves an accuracy of 83% on Part A (microscopic images) of BACH challenge.

To explore the performance of MobileNet and NasNet for breast cancer diagnosis (Falconi, Perez, & Aguilar, 2019), have used these networks in his study. The performance of these networks is compared with two other pretrained networks i.e. Inception-V3 and ResNet50. Author has used mammographic images taken from CBIS-DDSM database for the fine-tuning of pretrained networks. However, networks are used as features extractor i.e. weights of only last classification layer are fine tuned and remaining layer are frozen. Before feeding mammograms in the network they are first preprocessed i.e. intensity values normalization and histogram equalization and resizing of images is done. From the resultant preprocessed images two training datasets are built i.e. one by applying Otsu's thresholding and second by simply extracting region of interest using bounding boxes. Segmented images are fed to the network for training, where the two best accuracies are 74.3% and 78.4% that are achieved by MobileNet and ResNet50 without segmentation.

Chennamsetty et al., 2018 used an ensemble approach based on two pretrained CNNs for the classification of microscopic images. The two pretrained CNNs used in this study are ResNet-101 (He et al., 2016) and DenseNet-161 (G. Huang et al., 2017; Q. Huang et al., 2017). Due to the unavailability of sufficient training data weights of the model are imported from previously trained model on ImageNet (Deng et al., 2010) dataset. Before feeding the data into model images are first preprocessed. During preprocessing phase, symmetric padding is done to equalize the height and width of images then images are resized to 224*224 pixels using bilinear interpolation. The resized images are normalized using Z-score normalization. Two datasets are obtained by applying two types of data normalizations: 1) Dataset1 is achieved by using Statistics of ImageNet dataset during normalization, 2) Dataset2 is achieved by using statistics of the provided histological training data during normalization. The preprocessed data is then fed to both pre-trained models for training. Three fine tuned models are achieved after training, by: 1) training ResNet on Dataset1 2) training ResNet on Dataset2 and 3) training DenseNet on Dataset1. The trained ensemble model is tested on the provided challenge test data, by using majority voting scheme for achieving final set of predictions. The accuracy achieved of proposed technique on challenge test data is 87%.

In another study (Brancati et al., 2018), an ensemble of pretrained CNNs is used for the multi-classification of breast histopathological images into 4 classes. The proposed ensemble architecture is based on ResNet with three different configurations (i.e 34, 50 and 101 layers). ResNet (He et al., 2016) is a pretrained CNN on ImageNet (Deng et al., 2010) dataset, where in this study it is fine-tuned over microscopic image dataset provided by BACH-2018. The provided dataset is first preprocessed during which 5 different transformations are applied and central Square patches of size m*m are extracted from images to feed in network. During training, resultant images are fed to all the configurations of ResNet and predictions are obtained on the basis of maximum probability. The proposed methodology achieved an accuracy of 97.3% for multi-class classification on validation dataset, an accuracy of 86% on 100 images provided by BACH and 88.9% on dataset of Bio imaging 2015 challenge.

Marami et al., 2018, in their study presented an ensemble of CNNs for the classification of breast histopathological images. The proposed model is based on four modified inception neural network (Szegedy, Vanhoucke, Ioffe, Shlens, & Wojna, 2016), which are trained on four different subsets of training datasets. The modified Inception network consists of an additional adaptive pooling layer before fully connected layer. The dataset for training and evaluation of proposed model consists of microscopic and WSI images provided by BACH and additionally contain benign tissue images from publicly available BreakHis Dataset. Data is preprocessed before feeding in the network, during which random data augmentation on different subsets of training data is applied to reduce directional biases, to generate richer dataset and to improve generalizability of our classifier. The output probabilities of all four trained models are averaged during the predictions of unseen data. Each of these CNNs was trained on 512×512 images extracted at $20 \times$ magnification from both microscopy and WSI. For microscopic images predictions are obtained by majority voting. For WSI a secondary classifier ResNet34 (He et al., 2016) is applied which is trained over data with manually marked tissue regions to separate tissue from background regions. The proposed scheme gives an accuracy of 84% on microscopic images and 0.5526 on WSI images provided by BACH.

In their study (Guan et al., 2017), three training methodologies based on CNN are compared for the classification of mammograms. Two image datasets are used for training and evaluation of proposed methodology; MIAS and DDSM. In preprocessing stage, ROIs are extracted from mammogram using bounding box. Extracted ROIs are then resized and converted to RGB for further processing. The authors have compared the accuracies of a model built from scratch, features model and fine-tuned model. In features model, a pre-trained VGG-16 network is utilized for features extraction, in which first five convolutional layers are frozen and last fully connected layer, is initialized by random weights. In fine tuning, the weights of model are imported from pretrained VGG-16, first four convolutional layers are kept frozen and the weights of remaining layers are fine-tuned during training. Results show that, features model outperforms model built from scratch while providing a classification accuracy of 0.905. The accuracy difference between second and third methodology is just 0.008 but the second methodology only takes 5% of the total time of fine-tuning.

To compare the results of some state of the art transfer learning based pretrained models (i.e. VGG-16, ResNet50 and InceptionV3) over different training methodologies (i.e. training from scratch, fine-tuning some layers or fine-tuning entire network) (Chougrad, Zouaki, & Alheyane, 2018), have done a detailed experiment using breast mss images extracted from three publically available mammography datasets; DDSM, BCDR-F03 and INBreast. In the preprocessing stage, patches of ROIs are extracted, resized and subsequently global contrast normalization and data augmentation is applied on the extracted patches. Performance of models is compared first by initializing them with random weights and secondly by utilizing pretrained weights. Result shows that models initialized with pretrained weights outperformed the ones with random weights. Next the performances of models are compared by fine-tuning them. In majority cases best performance is achieved by fine-tuning last two convolutional layers. Overall, Inception-V3 has performed best by fine-tuning last two layers which is later tested over MIAS dataset and gives an accuracy of 98.23% and an AUC of 0.99%.

In this study (Jannesari & Elemento, 2018), different versions of fine-tuned ResNet (He et al., 2016) and Inception architectures (Szegedy et al., 2016) are tested (i.e. ResNet V1 50/101/152 and Inception V1/V2/V3/V4) for the classification of histopathological images by fine-tuning last fully-connected layer and by fine-tuning full network. Two datasets are used for validation of these networks; first one is taken from Tissue Micro Array (TMA) database (Marinelli et al., 2008) and second is BreaKHis dataset (Spanhol et al., 2016). In preprocessing phase, data normalization, patches extraction, resizing and subsequently data augmentation is done. The preprocessed images are then given to ResNet and Inception pre-trained models for multi-classification of breast images. Results show that, ResNet outperforms Inception by fine tuning all their layers.

In (Yap et al., 2018), author has compared three methods (including encoder-decoder U-Net architecture (Ronneberger et al., 2015), a patch-based LeNet (Bengio & Haffner, 1998) that takes 28*28 pixels patches of images as input and a transfer learning based pre-trained FCN-AlexNet (Shelhamer, Long, & Darrell, 2017)) for the detection of breast lesions using breast ultrasound (BUS) images. The performance of proposed three approaches has been compared with four state-of-the-art tumor detection algorithms (i.e. Multifractal Filtering (Yap, Edirisinghe, & Bez, 2008), Radial Gradient Index (Drukker et al., 2002), Deformable Part Model (Pons, Martí, Ganau, Sentís, & Martí, 2013) and Rule-based Region Ranking (Shan, Cheng, & Wang, 2012)). Two proprietary datasets has been utilized for the evaluation of proposed approaches that are comprised upon 306 and 163 BUS images, respectively. Second dataset has been made publically available for research purposes. Among all the mentioned methodologies, proposed patch-based LeNet has accomplished best results over 2nd Dataset, while transfer learning based FCN-AlexNet has performed best over 1st Dataset.

For the detection of cancerous regions in BUS images an automated approach has been proposed in (Wang et al., 2018). The author has employed traditional segmentation architecture 3D U-Net as a backbone of proposed methodology, whose parameters are first fine-tuned using C3D (Tran, Bourdev, Fergus, Torresani, & Paluri, 2015) model to avoid over-fitting. For the refinement of probability map and better detection of cancerous regions, a threshold map layer is concatenated at the end of network. Author has also defined a new densely deep supervision (DDS) mechanism for the extraction of distinct features to discriminate cancerous regions from non-cancerous and to avoid gradient vanishing problem, in which concatenated features extracted from each stage of UNet are processed and two DDS loss functions are introduced (i.e. Class balanced Cross Entropy Loss and Overlap Loss) for the development of cancer probability map. The performance of proposed methodology is evaluated over a proprietary dataset of 196 women that are acquired through Invenia ABUS system (GE, USA). Before feeding the data into the network, ABUS volumes of size $96 \times 64 \times 96$ cube are extracted and data augmentation is applied.

In another succeeding study (Wang, Xu et al., 2020), author has proposed an improved methodology for the discrimination of BUS cancerous regions from non-cancerous ones. Similar to previous study, author has utilized 3D U-Net architecture, fine-tuned using weights of C3D model (Tran et al., 2015), while also adopted 3D-dialated convolution for the aggregation of contextual features to improve detection sensitivity for small lesions. Moreover, threshold map for the refinement of probability map and DDS mechanism to avoid gradient vanishing problem together with a hybrid loss (i.e. combination of overlap loss and Dice loss) to solve small positive sample problem has also been employed. The model is trained over a locally collected BUS dataset of 219 females, while the volumes of size 128*64*128 are extracted and data augmentation is done before feeding data to the network. The proposed methodology has outperformed state-of-the-art deep models including FCN (Shelhamer et al., 2017), V-Net (Milletari et al., n.d.), 3D U-Net (Ronneberger et al., 2015) and SegNet (Badrinarayanan et al., n.d.), while achieved a sensitivity of 95% and FP per volume of 0.84.

A comparative analysis of several state of the art deep networks for object detection including Fast-RCNN (Girshick, 2015), Faster-RCNN (Ren, He, Girshick, & Sun, 2017), YOLO (Redmon et al., n.d.), Single Shot MultiBox Detector (SSD), has been done in (Almajalid, Shan, Du, & Zhang, 2019) for the discrimination of breast tumor regions using BUS images. Two pre-trained architectures (i.e. VGG16 (Simonyan & Zisserman, 2015) and ZFNet) have also been embedded with them for features extraction. Due to the unavailability of publically available large BUS images dataset, a new dataset has been collected to accomplish this task, which comprises of 464 malignant and 579 benign tumor cases. Results depict that, SSD300 has performed best as compared to other networks in term of average precision rate, average recall rate and F1 score.

8. Discussion

Many of the authors have utilized transfer learning based pretrained models in their proposed methodologies for automated breast cancer diagnosis. Majority of these studies are based upon fine-tuned models as compared to models without weights. The idea behind using pretrained models without their weights raise from the fact that, a large number of deep models have already been proposed by different authors in literature, which are trained over different imaging datasets (i.e. natural imaging datasets) and have depicted significant performance, so by training such model architectures from scratch on datasets of specific domain could provide significant results, however for training such deep models from scratch large scale datasets are required. In (Alom et al., 2019) author has utilized a pre-proposed deep architecture namely IRRCNN for the binary and multi classification of breast histopathological images, that was previously trained over CIFAR-10 dataset. In this study, model is trained over a large augmented microscopy breast images dataset (BreakHis) and has depicted significant results for both binary and multi-classification of images so does (Lu et al., 2019). Some of the authors have compared the performance of these weightless models with fine-tuned models such, as done in (Brancati et al., 2018) and (He et al., 2018), however results of these studies shows that fine-

Table 6
Models without weights.

Paper	Image Modality	Classification type	Preprocessing	Segmentation	Features	Classifier	Dataset	Dataset Size	Training Parameters	Evaluation Metrics	Regularization	Learning Type
(Brancati et al., 2018)	Histopathology (Microscopy)	Multi-class (Normal, In-situ, Benign or Invasive)	Data augmentation	Patch extraction of size m*m	Deep	Ensemble Classifier of ResNet-34/50/101 based on Maximum Probability (without weights)	Training (400 microscopic images) provided by ICIAR-2018	400	–	Accuracy = 77.3%	Data Augmentation	Transfer Learning (ensemble)
(Lu et al., 2019)	Mammography	Binary (benign and malignant)	CLAHE, median filter, data augmentation	–	Deep	Fully connected deep CNN [14]	Local dataset	9927	LR = 0.0001, AF= ReLu, dropout = 0.5	Accuracy = 0.82, sensitivity = 0.91, specificity = 0.57, F1 = 0.88	Dropout, Data augmentation	Transfer
(S. He, et al., 2018)	Histopathology (WSI)	Binary	Data Augmentation	Patch extraction (256*256)	Deep	AlexNet,	Local (WSI) dataset	5000 patches (121 cases)	LR = 0.001, BS = 8, M = 0.9, max_iter = 25,000	Accuracy = 93.40%	Data Augmentation	Transfer Learning
						GoogLeNet			LR = 0.001, BS = 4, M = 0.9, max_iter = 50,000	Accuracy = 94.12%		
(Alom et al., 2019)	Histopathology (Microscopy)	Binary	data augmentation	–	Deep	IRRCNN (image level)	BrekHis	Original = 7909, Augmented = 166,068	Opt = SGD, M = 0.9, epochs = 150	Accuracy = 97.95 ± 1.07%	Augmentation	Transfer Learning
		Multi								Accuracy = 97.57 ± 0.89%		
		Binary	resizing, cropping, data augmentation	Random patch extraction (winner take all)		IRRCNN (image level)	BioImaging-2015	Original = 249, Augmented = 43,707		Accuracy = 100%		
		Multi								Accuracy = 100%		
(Wang, Qin et al., 2020)	Ultrasound	Normal and Cancerous	Data augmentation	ABUS Volumes Extraction	Deep	3D-Inception UNet	Proprietary dataset	678	LR = 1e-3, Epoch = 60	Sensitivity = 95%, FP = 3.0	Augmentation, batch normalization	Transfer learning
(Benjelloun et al., 2018)	DEC-MRI	Normal and Cancerous	–	–	Deep	U-Net	Proprietary dataset	86 cases (5452 slices)	LR = 0.005, M = 0.99, E = 500, opt = SGD	IoU = 76.14%	batch normalization, Dropout	Transfer learning
(Kakileti et al., 2019)	Thermography	Normal and Cancerous	Data augmentation	Otsu thresholding	Deep	VGGNet	Proprietary dataset	180 subjects	Opt = ADAM, LR = 10-4, 10-5	Accuracy = 81.6	Data Augmentation, batch normalization, L!, L2 regularization , dropout	
						InputCascadeCNN				Accuracy = 87.6		
						UNet				Accuracy = 99.5		
						VNet				Accuracy = 99.6		
(Caballo et al., 2020)	Tomography	Benign and Malignant	Data augmentation	Voxels extraction	Deep	UNet	Proprietary dataset	93 samples, augmented samples = 34,992	BS = 16, opt = ADAM, LR = 0.001, epoch = 50	Dice = 0.93, Sensitivity = 0.92, Precision = 0.93, Conformity = 0.85	–	Transfer learning

Table 7
Fine-tuned models.

Paper	Image Modality	Classification type	Preprocessing	ROI	Features	Classifier	Dataset	Dataset Size	Training Parameters	Evaluation Metrics	Regularization	Learning Type
(Chang et al., 2017)	Histopathology	Binary	Data Augmentation		Deep Features	Google's Inception-V3	BreaKHis	7909	Epochs=500	Accuracy=0.89%	Data augmentation	Transfer Learning
(Yemini et al., 2019)	Mammography	Binary (malign and benign)	Resizing, Data Augmentation, Artificially tumor insertion Resizing, Augmentation, Augmentation of malignant images	×	Deep feature	Modified Google Inception-V3 architecture	**INbreast**	410	5-fold cross validation	AUC=0.78 AUC =0.86	Data Augmentation	Transfer Learning
(Scotty Kwok, 2018)	Histopathology	Multi-class	Patch extraction, Resizing , Data Augmentation	–	Deep features	Inception-Resnet-v2	ICIAR-2018 grand challenge	400	Dropout = 50%, LR =0.001, M = 0.9, BS = 64, epochs = 25	Accuracy = 87%	Data Augmentation, Dropout	Transfer Learning
(Wan, Sun, Ma & Fang., 2018)	Histopathology	Multi-class	Resize, Data Augmentation		Deep Features	VGG-16	BACH-2018	400	LR = 0.0001, decay_rate = 0.97 times / 200 steps	Validation Accuracy = 92.5% Test Accuracy = 83%	L2 weight decay, Dropout, Data augmentation	Transfer Learning
(Koné & Boulmane, 2018)	Histopathology	Multi-class	Image Resizing (229*229), Data Augmentation		Deep Features	ResNetXt50	BACH-2018	400	BS = 10, epochs = 60	Test Accuracy = 81%	Batch Normalization, Dropout	Transfer Learning
(Kohl et al., 2018)	Histopathology	Multi-class	Data Augmentation		Deep Features	Dense Net VGG-19 Inception-V3	BACH-2018	400	BS=32, epoch1=25 LR1=$1*10^{-3}$ epoch2=200, LR2=$2*10^{-4}$ 5-fold CV	V. Acc=94%, T. Acc=83% V. Acc 0.925 V. Acc =91.25%	pre-trained model's default	Transfer Learning
(Falconi et al., 2019)	Mammography	Binary(Benign and Malignant)	Normalization, Histogram Equalization, Resizing	ROI patches Otsu's Threshold	Deep Features	MobileNet Resnet50 InceptionV3 NAsNet MobileNet Resnet50 InceptionV3 NAsNet	CBIS-DDSM	Train=1318, Test=378	Eochs=6000, LR=0.01. BS=100	Accuracy=74.3% Accuracy=78.4% Accuracy=68.9% Accuracy=73.1% Accuracy=66.0% Accuracy=62.4% Accuracy=67.5% Accuracy=68.0%	–	Transfer Learning
(Sai Saketh Chennamsetty et al., 2018)	Histopathology	Multi-class	symmetric padding, Resizing, Zscore normalization	–	Deep features	Ensemble Classifier ResNet-101 (DS1) DenseNet-161 (DS1) DenseNet-161 (DS2)	ICIAR-2018 grand challenge	400	LR = 0.0001, opt= ADAM, norm Loss = cross entropy, epochs = 30, LR = 0.0001, opt = ADAM	Accuracy = 87%		Transfer Learning
(Brancati et al., 2018)	Histopathology	Multi-class	Patch extraction, Data Augmentation	–	Deep Features	Ensemble Classifier of ResNet-34/50/101 (fine-tuning full model) (fine-tuning last layer)	ICIAR-2018 grand challenge	400	m= 308, k = 80%, L (number of ResNet layers) = 34, 50, 101	Validation Accuracy = 97.3%, Test Accuracy = 86% Validation Accuracy = 92%	Data Augmentation	Transfer Learning
(Marami et al., 2018)	Histopathology	Multi-class	Data augmentation	×	Deep Features	Ensemble of Inception-v3 CNN	Bio-Imaging 2015 BACH-2018			Accuracy = 94.4% Accuracy = 84.0%	Data Augmentation, (Batch Norm, Label Smoothing)	Transfer Learning

(continued on next page)

Table 7 (*continued*)

Paper	Image Modality	Classification type	Preprocessing	ROI	Features	Classifier	Dataset	Dataset Size	Training Parameters	Evaluation Metrics	Regularization	Learning Type
(Chougrad et al., 2018)	Mammography	Binary(Benign and Malignant)	Global histogram Normalization, Data Augmentation	Patch Extraction, Resizing	Deep Features (fine-tuned models)	VGG-16 ResNet65 Inception-V3	DDSM+BCDR-F03+INBreast	6116	Opt=SGD, LR=1e-4, epochs=90 BS=128, CV=5-fold	Accuracy=98.64% Accuracy=98.77% Accuracy=98.94%	Data Augmentation, DropOut, L2 weight decay	Transfer Learning
(Jannesari & Elemento, 2018)	Histopathology	Multi-classification	Data augmentation		Deep Features	Inception-V1 (tune FC layer) Inception-V1 (tune full net) ResNet-V1-50 (tune FC layer) ResNet-V1-50 (tune full net) ResNet-V1-101 (tune FC layer) ResNet-V1-101 (tune full net)	BreaKHis + Tissue Micro Array (TMA)	7909+6402	Epochs=3000	ACC=0.917, AUC=0.917 ACC=0.971, AUC=0.98 ACC=0.993, AUC=0.996 ACC=0.998, AUC=0.996 ACC=0.917, AUC=0.917 ACC=0.996, AUC=0.999	Batch Normalization, Drop Out, Data Augmentation,	Transfer Learning
(Shuyue Guan et al., 2017)	Mammography	Binary (malign and benign)	Resizing, Grayscale to RGB	Bounding Box	**Deep** (from VGG-16 by freezing first 5 layers) **Deep** (from VGG-16 by freezing first 4 layers) **Deep** (from VGG-16 by freezing first 5 layers)	VGG-16 (Feature Model) VGG-16 (Fine Tuned Model) VGG-16 (Feature Model)	MIAS DDSM	322 10,480	Opt=RMsprop, loss=Binary Cross Entropy, BS = 15, epochs = 500 DO = 0.5, opt = Nadam, loss = Binary Cross Entropy, BS = 20, epochs = 500, 10-fold cross validation, LR = $1{*}e^{-4}$	Accuracy = 0.906 Accuracy = 0.914 Accuracy = 0.950	Dropout	Transfer Learning
(Yap et al., 2018)	Ultrasound	Binary (malign and benign)	– – –	Patch extraction (28*28) – –	Deep Features	Patch based LeNet U-Net FCN-AlexNet	Dataset 1 Dataset 2 Dataset 1 Dataset2 Dataset 1 Dataset2	306 163 306 163 306 163	LR = 0.01, epochs = 60, DO=0.33, 10-fold CV Opt=Adam, LR=0.0001 epochs=300, 10-fold CV LR=0.001, epochs=60, dropout =33%, 10-fold CV	FP=0.10 F-score=0.88 FP=0.14 F-score=0.86 FP=0.21 F-score=0.86 FP=0.28 F-score=0.75 FP=0.16 F-score=0.91 FP=0.17 F-score=0.89	Dropout	Transfer Leaning
(Wang, Xu et al., 2020)	Ultrasound	Normal and Abnormal	Data augmentation	Volumes Extraction	Deep Features	U-Net with Threshold map and DDS mechanism	Proprietary dataset	559	LR=1e-4, epochs=30000	Sensitivity=93%, FP=2.2	Data augmentation	Transfer Learning
(Wang, Xu et al., 2020)	Ultrasound	Normal and Abnormal	Data augmentation	Volumes Extraction	Deep Features	U-Net (dilated convolution) with Threshold map and DDS	Proprietary dataset	614	LR=1e-4, epochs=40000	Sensitivity=95%, FP=0.84	Data augmentation,	Transfer Learning

(*continued on next page*)

Table 7 (*continued*)

Paper	Image Modality	Classification type	Preprocessing	ROI	Features	Classifier	Dataset	Dataset Size	Training Parameters	Evaluation Metrics	Regularization	Learning Type
						mechanism with hybrid loss					Batch Normalization	
(Almajalid et al., 2019)	Ultrasound	Benign and Malignant	Resizing	–	ZFNet	SSD300	Proprietary dataset	1043	–	APR=96.89 ARR=67.23 F1=79.38	Pretrained model's default	Transfer learning

tuned models outperforms models trained from scratch in tumor classification problem. Additionally, several authors have utilized state-of-the-art encoder decoder segmentation architectures (i.e. UNet, VNet, 3D-Inception UNet etc) in their studies and trained them from scratch for the detection of different breast abnormalities. As these architectures are very deep and may prone to over-fitting, several regularization techniques (i.e. data augmentation, dropout, batch normalization etc) have been employed in all of these studies. According to (Kakileti et al., 2019), among state-of-the-art segmentation architectures (including UNet, VNet, VGG-Net and inputCascadeCNN), VNEt has depicted best detection results over breast thermography dataset.

In addition to data augmentation another alternative solution to tackle with limited data problem is to use transfer learning based pre-trained models to achieve better performance. A lot of work has been done for the diagnosis of breast cancer using such pre-trained models that are fine-tuned over breast imaging datasets (i.e. mammography, histopathology, and ultrasound datasets). Most of such studies have used histopathological breast tissue images for multi classification and x-ray mammography images for binary classification, while breast ultrasounds have been utilized for the detection or segmentation of breast abnormalities. Almost every study has involved data augmentation technique to extend training data and regularization techniques to avoid over-fitting. Preprocessing is applied on images in accordance to the deep CNN used i.e. images are resized according to the input layer size of model. In case of histopathological images no segmentation technique has been applied, however simply patches are extracted from them for further processing. In addition to limited data problem, another significant problem is imbalanced dataset, such as in (Yemini et al., 2019) author has embedded artificially generated malignant image patches to balance the dataset, where the results show that balancing of data has significantly improves classification performance. Several studies discussed above, belongs to the participants of BACH challenge 2018 in which they have employed different pretrained models such as Inception-Resnet-v2 (Kwok, 2018), VGG-16 (Wang et al., 2018), DenseNet (Kohl et al., 2018) and ResNetXt50 (Koné & Boulmane, 2018) networks and trained them over the provided challenge data, while Inception ResNet-v2 has performed best among all of the mentioned CNNs. Meanwhile another methodology has achieved same accuracy as that of Inception-ResNet V2 in BACH-2018, which is an ensemble of ResNet-101 and two DenseNet-161 modules (i.e. trained over differently normalized datasets) proposed by (Chennamsetty et al., 2018). This study highlighted the importance of training data normalization, as the author has utilized two same configurations of DenseNet-161 in his proposed ensemble just by training it over differently normalized datasets. Apart from (Chennamsetty et al., 2018) two other participants of this challenge i.e. (Marami et al., 2018) and (Brancati et al., 2018) have also presented deep CNNs based ensemble classifiers. They have used different versions of same CNN models in their proposed ensembles, however their achieved performance is not as good as of (Chennamsetty et al., 2018).

It could be seen from (Guan et al., 2017), (Brancati et al., 2018) and (Jannesari & Elemento, 2018), that fine-tuning whole model has great impact over accuracy as compared to fine-tuning only specific layers. In (Wang et al., 2018), author has explored that the detection performance of a deep network could be significantly improved by importing weights from a different pretrained network i.e. the proposed methodology in this study has outperformed state-of-the-art deep segmentation architectures. (See Tables 6 and 7)

8.1. Hybrid models

Huynh et al., 2016, have trained and evaluated three SVM classifiers for the binary categorization of breast lesions as benign and malignant. The authors have compared the classification accuracies of SVM by training it on subjective and objective features extracted from mammograms taken from a proprietary dataset. At first patches of size 512 *

Table 8
Hybrid Models.

Paper	Image Modality	Classification type	Preprocessing	Segmentation	Features	Classifier	Dataset		Training Parameters	Evaluation Metrics	Regularization	Learning Type
(Huynh et al., 2016)	Mammography	Binary Classification (malign and benign)	(0-1) normalization, resizing	ROI extraction using seed	AlexNet CNN (FC6 layer)	SVM	Local Dataset	607	5-fold cross validation	AUC = 0.81	–	Hybrid
(Cao, et al., 2018)	Histopathology	Multi-class	–	–	Handcrafted (PTFAS+GLCM) ResNet-18 ResNeXt NASNet-A ResNet-152 V GG-16 All	RFSVM	BACH-2018	400	No of tree =500, (SVM) c={0.01, 0.1, 1, 10, 100, 1000}	V.ACC=76% V.ACC=78% V.ACC=81% V.ACC=78% V.ACC=83% V.ACC=78% V.Acc=93% T. Acc=79%	–	Hybrid
(Liang et al., 2018)	Mammography	Binary Classification (malign and benign)	Gaussian low-pass filter	–	Shape(11), GLGM (12), AlexNet CNN	Ensemble LR,RF and SVM	Local	212	LOOCV	Accuracy =0.964 Sensitivity=0.94 Specificity=0.99 Youden Index = 0.925	–	Ensemble
(He et al., 2018)	Histopathology (WSIs)	Binary (normal, cancerous)	Patch extraction, Data Augmentation	–	AlexNet GoogLeNet AlexNet, GLCM, LBP GoogLeNet, GLCM, LBP	SVM	Local Dataset	5000 (121 cases)	(AlexNet) LR= 0.001, BS=8, M=0.9, max_iter=25,000 (GoogLeNet) LR=0.001, BS=4, M=0.9, max_iter=50,000	Accuracy=94.79% Accuracy=95.15% Accuracy=98.83% Accuracy=98.90%	Data Augmentation	Hybrid
(Wang et al., 2018)	Histopathology	Multi-class	Resize, Data Augmentation	–	Deep Features using VGG-16	SVM	BACH-2018	400	Lr = 0.0001, decay_rate = 0.97 times / 200 steps	Accuracy = 81.25%	Data augmentation	Hybrid
(Agrawal et al., 2018)	Mammography	Binary Classification (malign and benign)	CLAHE, horizontal flipping, downsizing	–	VGG-16 (from fc-2 layer)	KNN Decision Tree Gradient Boosting Voting Classifier	MIAS	322	K=1, Max_depth = 9, min_sample_split = 2 Lr = 0.1, no of estimators = 100, max_depth = 10, min_sample _Solit = 2 Soft_voting, **weights:** KNN = 2, DT = 1.02, GB = 1	Accuracy = 80% Accuracy = 80% Accuracy = 75% Accuracy = 85%	–	Hybrid Hybrid Hybrid Ensemble
(Wang et al., 2018)	Mammography	Binary (Benign, Malignant)	Patch extraction, resizing, data augmentation	Patch extraction	Deep features using Inception-V3 (feature refinement using attention) + Clinical features	RNN	(BCDR-F03)	736	Weight decay=0.001, dropout=0.2, Opt= RMSprop ,loss=cross entropy, BS=64, epochs=2000, LR=10e^{-4}, decay=0.5	Accuracy=85% AUC=0.89	Data Augmentation, Dropout, Weight decay	Hybrid
(Zhang et al., 2019)	Mammography	Binary	Crop, Resize	RPN	Mobile-Net (Conv 2D 11) Inception-V2 (Mix-4e) ResNet101 (Conv4) Inception-v3 (Mixed-6e) Inceptin-ResNet-v2 (Mixed-6a)	RCNN	DDSM	5045	Opt=SGD, M=0.9, BS (for RPN)= 256, BS(for classifier)= 64, LR=3e-3 LR=2e-4 LR=3e-4 LR=3e-4 LR=1e-3	Precision=0.60 Precision=0.769 Precision=0.769 Precision=0.80 Precision=0.85	Batch Normalization	Hybrid
	Mammography							322+2620			–	Hybrid

(continued on next page)

Table 8 (*continued*)

Paper	Image Modality	Classification type	Preprocessing	Segmentation	Features	Classifier	Dataset		Training Parameters	Evaluation Metrics	Regularization	Learning Type
Ahmed et al., 2020)		Binary (Benign and Malignant)	Normalization, noise removal, muscle removal	Patch extraction (512*512)	ResNet-101 Xception65	RCNN DeepLab V3	MIAS+CBI-DDSM		LR=0.001, epochs=1000	MAP=80Accuracy=98 MAP=72 Accuracy=95		
(Yap et al., 2020)	Ultrasound	Binary (malign and benign)	Gray to RGB	–	Inception-ResNet-v2	Faster-RCNN (RGB)	Dataset 1 Dataset2	306 163	LR=0.001, epochs=100	IoU=0.8535 Recall=0.9358 Precision=0.8818 F1=0.9080 IoU=0.8254 Recall=0.8221 Precision=0.8481 F1=0.8349	Pretrained-model's default	Transfer Learning
(Hu et al., 2020)	DCE-MRI T2w Combined	Binary (malign and benign)	– MIP –	Fuzzy C-mean	VGG-19	SVM	Proprietary dataset	927	Kernel=RBF, CV=5-fold	AUC=0.85 AUC=0.78 AUC (image fusion)= 0.85 AUC (feature fusion)= 0.87 AUC (classifier fusion)=0.86	Pretrained-model's default	Transfer Learning
(Mambou et al., 2018)	Thermograms	Normal and cancerous	RGB to gray	–	Inception V3	SVM	DMR-IR	64	LR=0.0001, Epochs=15, steps=3900	Accuracy=100%	Pretrained-model's default	Transfer learning
(Zhang et al., 2018)	Mammograms 3D-Tomosynthesis	Normal and Cancerous	Data augmentation	–	AlexNet	Shallow CNN	Proprietary dataset		BS=256,LR=0.001, Dropout=0.5, L2-reg=0.0001 BS=16,LR=0.0001, Dropout=0.5, L2-reg=0.0001	ROC=0.7274 ROC=0.6632	Data augmentation, Dropout, L2 regularization	Transfer learing

512 are extracted from mammograms which are then down-sampled to 256*256 and normalized. The three classifier trained in this study are described as follows: In first case, SVM is trained on features extracted from normalized patches by using a pre-trained CNN namely AlexNet (Krizhevsky et al., 2012), In second case, handcrafted size, shape, intensity and margin features are extracted and then classified using SVM, In third case, both deep and handcrafted features are combined and fed to SVM for classification. To get the best set of features a separate SVM classifier is trained over features of each layer and evaluated by ROC metric. During layers comparison of AlexNet, results show that, at Conv4 model's performance is at peak and model performs best when both subjective and objective features are combined. In the proposed model, Fc6 is used for features extraction due to its relatively low dimensionality and high performance.

Extraction of six different types of features for the classification of BACH-2018 microscopic images is done by Cao, Bernard, Heutte, and Sabourin (2018). The authors have used five pretrained deep CNNs on ImageNet Dataset for feature extraction, which includes: ResNet-18 (He et al., 2016), ResNet-152 (He et al., 2016) ResNeXt (Xie et al., 2017), NASNet-A (Zoph, Vasudevan, Shlens, & Le, 2018) and VGG16 (Simonyan & Zisserman, 2015). In addition to deep features, handcrafted features are also extracted from the images. Parameter-Free Threshold Adjacency Statistics (PFTAS) and GLCM features are combined to form the feature group. The provided sample size for training is much smaller, while the resultant features extracted from six groups have very high dimensionality. Multi-View Random forest kernel SVM (RFSVM) (Cao et al., n.d.) is used to efficiently combine all these features based on their performance. Results show that classification accuracy significantly increases by combining deep features with handcrafted features. An accuracy of 79% is achieved on Part A of BACH challenge.

In a study by (Liang et al., 2018), a radiomics model is proposed, for the classification of malignant and benign breast lesions using both deep and handcrafted features (HCF) (i.e. shape and GLGM), extracted from CC and MLO views of mammograms. Before feature extraction, Gaussian filter is applied for noise removal and images are resized to 512*512. In particular, ALexNet (Krizhevsky et al., 2012) is used for deep features extraction. To select optimal set of features, T-test is applied on them. A radiomics based classification model is constructed via weighting multi-trained classifiers (including logistic regression (LR), random forest (RF) and support vector machine (SVM)) trained over the extracted HCF and deep features. Author has used a proprietary dataset consisting of 212 full field digital mammographic images. The proposed model achieves an accuracy of 96.4% in breast lesion classification.

In their study (He et al., 2018), a hybrid model is proposed for the classification of Histopathological whole slide images. For the validation of proposed methodology a proprietary H&E stained histological whole slide images dataset of 121 patients are used. In preprocessing phase, samples of size 256*256 are extracted from images, at 20 × magnification. From these extracted patches noisy patches are manually discarded and data is fed to CNN after applying data augmentation. Two different TL approaches are tried for classification of normal and cancerous images. In first case, the basic architecture of AlexNet (Krizhevsky et al. 2012) and GoogLeNet (Szegedy et al., 2015) without pretrained weights is trained over the original dataset. In the second case, pretrained AlexNet and GoogleNet over ImageNet (Deng et al., 2010) dataset are used for features extraction. Redundant and irrelevant features are then excluded based on differences between positive and negative labels. Due to the complexity of Histopathological images, GLCM and LBP features are also extracted for image texture characterization. Finally SVM is trained over the extracted features. Results shows that SVM trained over combined features of GoogleNet, LBP and GLCM gives an accuracy of 98.90%.

(Wang et al., 2018), compared the classification accuracy of a pretrained convolutional neural network VGG16 with a hybrid model that consists of VGG-16 + SVM, for the multi-classification of breast histopathological images. First case has already been discussed in Section 7.2. In hybrid approach, pre-trained VGG-16 is used as features extractor, which is first fine-tuned using dataset provided by BACH-2018. Data augmentation technique is applied on dataset before feeding the images to the network. Features extracted from the fine-tuned model are then fed to SVM for the classification of microscopic images into four classes. The model has achieved an accuracy of 81.25% on validation set and an accuracy of 80.6% on 36 images of Bioimaging 2015 dataset.

A hybrid approach has been employed for the classification of mammographic images to detect breast cancer by Agrawal et al. (2018). The proposed architecture is based on two classifiers: first one is transfer learning based pretrained CNN (VGG-16) (Simonyan & Zisserman, 2015) on ImageNet (Deng et al., 2010) dataset and the second is an ensemble of linear classifier. For the evaluation of the proposed model, MIAS dataset (SUCKLING, 1994) is used. In preprocessing stage, CLAHE is applied for contrast enhancement of images, then the images are downsized from 1024*1024 to 224*224. Since, ImageNet and mammography images datasets are very different from each other, therefore instead of just relying on the pretrained model an ensemble of linear classifiers is used on the output of pretrained CNN which consists of KNN, Decision Tree and Gradient Boosting. The output of CNN is passed to these linear classifiers which gave score to each label which is further passed to Voting Classifier for results.

A novel architecture for automated breast cancer diagnosis using mammographic images have been proposed in ()(Wang et al., 2018). In his study, the author has proposed a Deep Neural Network based on Multi-View data (MV-DNN) for the classification of mass patches extracted from multiple views of mammograms. The dataset used for the validation of proposed architecture is Breast Cancer Digital Repository (BCDR-F03) that consists of 736 breast film images together with associated clinical data. At first stage of proposed methodology, mass patches are cropped from multiple views of mammograms (i.e. CC and MLO) and resized to 299*299 pixels. Subsequently, data augmentation is applied on the extracted patches. Meanwhile, the provided clinical data is normalized (i.e. min–max normalization is done). Extracted mass patches are then fed to modified Inception-V4 model. Inception-V4 is a pretrained CNN on ImageNet dataset, which is further fine-tuned here on breast images dataset. To give more attention to mass pixels instead of background pixels in extracted features map (i.e. to get more informative feature maps) an attention based method is built similar to model proposed in (Xie & Tu, 2017).

The refined feature maps extracted from multiple views of mammograms together with clinical information are then fed to RNN (particularly Long Short Term Memory (LSTM) (Sutskever, Vinyals, & Le, 2014)) for final classification of masses (i.e. in benign and malignant). The proposed architecture (MV-DNN) achieves an accuracy of 85% and AUC of 0.89 on BCDR dataset.

In the article (Zhang, Wang, Zhang, & Mu, 2019), author has implemented Faster Recurrent Convolutional Neural Network (RCNN) for localization and binary classification of breast masses (i.e. in benign and malignant). Meanwhile, five feature extraction models i.e. inception-V2, inception-V3, ResNet101, Inception ResNet V2 and Mobilenet are also evaluated to analyze their impact on the performance of classifier. The author has used images extracted from DDSM database for fine-tuning and evaluation of proposed model. At first preprocessed images are fed to the five mentioned pretrained features extraction models which results corresponding feature maps. Region proposal network then used to extract regions of interest (candidate regions) from the extracted features maps. Subsequently, from the taken out candidate regions feature vectors are extracted, resized using ROI pooling layer and fed to fully connected layer for detection and classification of breast masses. Results shows that the best performance is achieved by using Inception ResNet-V2 (i.e. avg. precision = 0.85) and worst by using MobileNet (i.e. avg. precision = 0.60) as a feature extractor in proposed architecture.

In a latest study by (Ahmed et al., 2020), a hybrid methodology is proposed for the semantic segmentation and classification of breast masses, which is based upon two state of the art DL based segmentation frameworks i.e. Mask-RCNN (He, Gkioxari, Dollár, & Girshick, 2017) and DeepLab-V3 (Chen, Papandreou, Schrof, & Adam, 2017). Two publically available mammographic breast image datasets i.e. MIAS and CBIS-DDSM, are used for the performance analysis of proposed methodology. For noise and pectoral muscle removal, the author has proposed a preprocessing mechanism. Patches (512*512 pixels) are extracted from preprocessed images which are further fed to above mentioned two deep models. For features extraction, ResNet-101 and Xception65 pretrained CNNs are used as a backbone in Mask-RCNN and DeepLab respectively. The proposed methodology achieves an accuracy of 98% and 95% over Mask-RCNN and DeepLab-V3 respectively.

An automated BUS ROI extraction technique has been presented in (Yap et al., 2020), in which author has utilized Faster RCNN with Inception-ResNet-v2 deep architectures for object detection. To fulfill the deficiency of large-scale dataset and to improve the overall performance, transfer learning based approach together with a newly proposed RGB method has been utilized. The proposed approach is evaluated over two BUS image datasets (Yap et al., 2018) consisting of 306 and 163 images respectively. In the suggested 3-channel method, original grayscale image is converted into 3-channed RGB image by concatenating it with sharpened and contrast enhanced images. From the 3-channel images feature vectors are extracted using pretrained Inception-ResNet-v2 architecture (Längkvist et al., 2014), while extracted features are then fed to pretrained Faster-RCNN architecture for region proposal candidate extraction and subsequently for their refinement and classification. Results show that, suggested architecture has achieved notable enhancement on intersection over union (IOU).

In another study (Hu, Whitney, & Giger, 2020), author has utilized multi-parametric MRI images (including DCE and T2W MRI) for the binary classification of breast tumors (i.e. benign and malignant). Author has compared five different techniques, in which a pretrained CNN namely VGG-19 is utilized for features extraction while Fuzzy C-mean algorithm has been employed for automated region of interest extraction. In single sequence method, features are extracted from the images of both modalities and fed to SVM separately for the classification of lesions. Secondly three other methods have been investigated (i. e. feature fusion, image fusion and classifier fusion). In feature fusion scheme, features extracted from both modalities are combined as ensemble of features and then fed to the SVM classifier for classification, in image fusion technique a 3-channel RGB image has been constructed by fusing images of both modalities and keeping the third channel empty for features extraction, while in the third classifier fusion methodology a soft voting classifier is embedded in front of DCE and T2W based single sequence SVM classifier. For the validation of proposed methodologies a dataset collected from 616 patients (i.e. 927 unique breast lesions) of both modalities have been utilized. Author has also tested results of mentioned techniques after performing image registration over images of both modalities. Results of the proposed methodology have been mentioned in the Table above.

Based on an extensive review, an automated approach for the classification of breast thermograms has been proposed in (Mambou, Maresova, Krejcar, Selamat, & Kuca, 2018), which is validated over DMR-IR dataset (Silva et al., 2014). Author has employed pretrained Inception-V3 architecture for features extraction and classification of input thermograms (i.e. in normal and cancerous), but if the output probability of cancerous region is between 0.5 and 0.6, then the extracted feature vector is fed to SVM for classification. The proposed methodology has achieved accuracy of 100% in 15 epochs and 3900 steps by tuning the model's parameter over 64 training images.

In (Zhang et al., 2018), a comparative analysis of ten different deep learning architectures have been done for the classification of 2-D mammograms and 3-D tomosynthesis breast images. The proposed deep models are trained over a locally collected dataset consisting upon 2-D and 3-D breast images acquired from 793 negative and 125 positive patients. Due to limited number of samples in dataset, data augmentation technique (i.e. flipping and random rotation of images) has been applied; however for 3-D images augmentation is applied at training to minimize storage utilization. Two pretrained deep architectures including AlexNet and ResNet50 have been utilized for features extraction from input images, which are followed by shallow CNNs for the classification of images. Moreover, the proposed shallow networks have also been tested without employing transfer learning approach; however transfer learning approach (i.e. pre-trained Alex-Net followed by proposed shallow CNN) has outperformed other networks in classification of both 2D and 3D images (i.e. achieved ROC of 0.7274 and 0.6632 respectivily). In practice, 3D breast imaging data proves more powerful for abnormalities detection as compared to 2D, but as it is challenging to handle 3D data for automated detection, classification results of 3D breast tomosynthesis are not as efficient as 2D mammograms.

8.2. Discussion

One of the main limitations in statistical ML based classifiers is that, they are trained over limited number of features which is overcome by DL based models where thousands of features were extracted from images for their classification but they require large amount of training time. In hybrid models, both of these problems are solved, i.e. as the features are extracted using pretrained CNNs and subsequently tumors categorized using statistical classifiers. The usage of imaging modalities and preprocessing techniques in hybrid models is same as observed in previous cases, however in underlying studies; many authors have combined deep features with handcrafted features for the classification of breast images. No doubt efficient classification accuracy could be achieved using deep features but by combining hand-driven features together with deep features this accuracy could be improved (Huynh et al., 2016), (Cao, et al., 2018), (Liang et al., 2018) and (He, et al., 2018). While using deep models as a features extractor, a commonly faced problem is the selection of layer to extract features, to tackle with this problem author has trained different SVM classifiers on features extracted from each layer of CNN and selected the layer that achieves efficient accuracy. Usually features are extracted from a single features model; however in (Cao, et al., 2018) author has used deep features extracted from five different deep feature models together with handcrafted features for the classification breast histopathological images. This model has depicted significant training accuracy, however not performed as well on test data. One reason behind it could be that a sparse range of features has been extracted by using different feature models; training classifier by using only most informative features could improve the performance (i.e. features selection criteria is needed). In addition to deep models that are designed solely to perform breast images classification, significant work has been done for both localization and categorization of breast abnormalities in mammograms such as done in (Wang et al., 2018), (Zhang et al., 2019) and in (Ahmed et al., 2020). The two key findings about mammograms, which need to be focused upon, are: 1) different views of mammograms (i.e. CC and MLO) encompasses distinct information about abnormalities, i.e. detection of abnormalities could be much more improved by combining features of multiple views of mammograms, that in our case have not been done by majority of the studies except (Wang et al., 2018) and 2) features extracted from ROIs (i.e. abnormality areas) are much more informative (i.e. they contribute more in categorization of benign and malignant lesions) so giving them more importance as compared to other features by owing some weightage criteria could improve the results but this thing has been also ignored by most studies except (Wang et al., 2018). As these studies have used deep models as a features extractors instead of classifier, therefore no deep models related regularization technique (i.e. weights decay, dropout etc) have been used except data augmentation. Instead of using a single classifier, usage of ensemble of classifiers could improve accuracy as shown in (Agrawal et al., 2018). In medical image analysis, 3D data proves more informative as compared to 2D images i.e. breast CT images proves more helpful in abnormalities detection as compared to x-ray mammography, but to deal with 3D data for automated abnormalities detection is more challenging, time consuming and computationally hard, therefore, to get high performance over such data is more difficult (Zhang et al., 2018) (See Table 8).

9. Findings and recommendations

The major findings of this study include: for the automated detection and diagnosis of breast cancer mammography and histopathological images are most widely used, where mammograms are not used for multi classification of breast lesions (i.e. normal, benign, malignant and their further sub-categories). Majority of these systems are trained over publically available datasets. The quality and size of datasets have a great impact on the accuracy of models. Certain limitations exist in these publicly available datasets as discussed in Section 4.2. Using a good quality and large scale dataset could further improve the accuracy of these systems. Most of the authors who have used limited amount of training data have applied data augmentation and regularization techniques to avoid over-fitting and to improve models performance. As patches extracted from histopathological images are invariant to translation and rotation, therefore multiple kinds of augmentations are applied almost in every study, however in case of other modalities i.e. mammography only limited augmentations could be applied within some studies. Due to limited contrast in breast mammograms, different contrast enhancement and noise removal techniques are applied during their preprocessing. For histopathological images no segmentation technique is applied, however semi-automated regions of interest extraction techniques are used in case of other modalities, however this task could be improved using automated DL based segmentation techniques.

To achieve an efficient accuracy and generalizability from deep models a large amount of training dataset is required. However, if such a large scale dataset is not present transfer learning is used. According to the best of our knowledge most of the work for automated breast cancer detection is done based on transfer learning. These models are used differently in different studies i.e. some studies have used architectures of pretrained models without their weights, some have fine-tuned specific layers of these pretrained models, some have fine-tuned whole networks and some have used them just for features extraction by which statistical classifier is trained.

Moreover, in most studies ensemble of either DL based or statistical ML based classifier outperforms single classifier. Combination of deep and handcrafted features gives better results. Training models over more productive features instead of large quantity of sparse features (i.e. employing some features selection criteria) could improve the classification results. Majority of the mentioned studies have used single view mammograms for the diagnosis of breast cancer, while neglecting the plenty of productive information found in multiple-views of mammograms. All of the studies have focused upon finding a particular abnormality (i.e. either masses or calcifications) using a single CAD, however no such system has been proposed that could efficiently found multiple abnormalities at a time. Major concentration of most authors is to find breast masses as compares to other breast abnormalities.

In addition to mammography and histopathology we have also studied some papers based on ultrasound and MRI but their main focus is to detect breast lesions instead of diagnosing them, such as in (Almajalid et al., 2019) author have used UNet (Ronneberger et al., 2015) architecture for the segmentation of breast tumor in breast ultrasound (BUS) images. In (Benjelloun et al., 2018) author have used UNet architecture for the segmentation of breast lesions in breast MRI images. In majority of such studies, state-of-the-art deep segmentation architectures has been utilized for detection of breast abnormalities, however none of the study has put effort for the development of some new segmentation architecture. Moreover, from (Zhang et al., 2018) it could be concluded that getting efficient performance in automated analysis of 3D imaging data is more challenging, computationally hard and more time and space consuming as compared to 2D data, however in practice such datasets proves more informative as compared to 2D data. Therefore, efforts are required for designing efficient techniques that could achieve better results by utilizing the tendency of 3D-datasets.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Abdel-Nasser, M., Moreno, A., & Puig, D. (2019). Breast cancer detection in thermal infrared images using representation learning and texture analysis methods. *Electron, 8*. https://doi.org/10.3390/electronics8010100

Abdel-Nasser, M., Saleh, A., Moreno, A., & Puig, D. (2016). Automatic nipple detection in breast thermograms. *Expert Systems with Applications, 64*, 365–374. https://doi.org/10.1016/j.eswa.2016.08.026

Agrawal, S., Rangnekar, R., Gala, D., Paul, S., Kalbande, D., 2018. Detection of Breast Cancer from Mammograms using a Hybrid Approach of Deep Learning and Linear Classification. 2018 Int. Conf. Smart City Emerg. Technol. ICSCET 2018 1–6. https://doi.org/10.1109/ICSCET.2018.8537250.

Ahmed, L., Iqbal, M. M., Aldabbas, H., Khalid, S., Saleem, Y., & Saeed, S. (2020). Images data practices for Semantic Segmentation of Breast Cancer using Deep Neural Network. *Journal of Ambient Intelligence and Humanized Computing*. https://doi.org/10.1007/s12652-020-01680-1

Akselrod-Ballin, A., Karlinsky, L., Alpert, S., Hashoul, S., Ben-Ari, R., & Barkan, E. (2019). A CNN based method for automatic mass detection and classification in mammograms. *Computer Methods in Biomechanics and Biomedical Engineering, 7*, 242–249. https://doi.org/10.1080/21681163.2017.1350206

Al-Quraishi, T., Abawajy, J., Chowdhury, M.U., Rajasegarar, S., Abdalrada, A.S., 2017. Breast cancer risk assessment prediction using an ensemble classifier. 30th Int. Conf. Comput. Appl. Ind. Eng. CAINE 2017 177–183.

Albayrak, A., Bilgin, G., 2017. Mitosis detection using convolutional neural network based features. CINTI 2016 - 17th IEEE Int. Symp. Comput. Intell. Informatics Proc. 335–340. https://doi.org/10.1109/CINTI.2016.7846429.

Almajalid, R., Shan, J., Du, Y., Zhang, M., 2019. Development of a Deep-Learning-Based Method for Breast Ultrasound Image Segmentation. Proc. - 17th IEEE Int. Conf. Mach. Learn. Appl. ICMLA 2018 1103–1108. https://doi.org/10.1109/ICMLA.2018.00179.

Alom, M. Z., Hasan, M., Yakopcic, C., Taha, T. M., & Asari, V. K. (2018). Improved inception-residual convolutional neural network for object recognition. *Neural Computing and Applications*, 1–15.

Alom, M. Z., Yakopcic, C., Nasrin, M. S., Taha, T. M., & Asari, V. K. (2019). Breast Cancer Classification from Histopathological Images with Inception Recurrent Residual Convolutional Neural Network. *Journal of Digital Imaging, 32*, 605–617. https://doi.org/10.1007/s10278-019-00182-7

Amir, E., Freedman, O. C., Seruga, B., & Evans, D. G. (2010). Assessing women at high risk of breast cancer: A review of risk assessment models. *Journal of the National Cancer Institute, 102*, 680–691. https://doi.org/10.1093/jnci/djq088

Arafa, A., El-Sokary, N., Asad, A., & Hefny, H. (2019). Computer-aided detection system for breast cancer based on GMM and SVM. *Arab Journal of Nuclear Sciences and Applications, 52*, 142–150. https://doi.org/10.21608/ajnsa.2019.7274.1170.

Aresta, G., Araújo, T., Kwok, S., Chennamsetty, S. S., Safwan, M., Alex, V., et al. (2019). BACH: Grand challenge on breast cancer histology images. *Medical Image Analysis, 56*, 122–139. https://doi.org/10.1016/j.media.2019.05.010

Arslan, A.K., Yasar, S., Colak, C., 2019. Breast cancer classification using a constructed convolutional neural network on the basis of the histopathological images by an interactive web-based interface. 3rd Int. Symp. Multidiscip. Stud. Innov. Technol. ISMSIT 2019 - Proc. 1–5. https://doi.org/10.1109/ISMSIT.2019.8932942.

Badrinarayanan, V., Kendall, A., Cipolla, R., Member, S., n.d. SegNet : A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation 1–14.

Balleyguier, C., Ayadi, S., Van Nguyen, K., Vanel, D., Dromain, C., & Sigal, R. (2007). BIRADS™ classification in mammography. *European Journal of Radiology, 61*, 192–194. https://doi.org/10.1016/j.ejrad.2006.08.033

Bandyopadhyay, S., 2014. Pre-processing of Mammogram Images.

Begum, N. (2018). Breast cancer in Pakistan: A looming epidemic. *Journal of College of Physicians and Surgeons Pakistan, 28*, 87–88. https://doi.org/10.29271/jcpsp.2018.02.87.

Bengio, Y., Haffner, P., 1998. Gradient-Based Learning Applied to Document Recognition 86.

Benjelloun, M., El Adoui, M., Larhmam, M.A., Mahmoudi, S.A., 2018. Automated Breast Tumor Segmentation in DCE-MRI Using Deep Learning. 2018 4th Int. Conf. Cloud

Comput. Technol. Appl. Cloudtech 2018 1–6. https://doi.org/10.1109/CloudTech.2018.8713352.

Bhurgri, Y., Bhurgri, A., Hassan, S. H., Zaidi, S. H. M., Rahim, A., Sankaranarayanan, R., et al. (2000). Cancer incidence in Karachi, Pakistan: First results from Karachi Cancer Registry. *International Journal of Cancer, 85*, 325–329. https://doi.org/10.1002/(SICI)1097-0215(20000201)85:3<325::AID-IJC5>3.0.CO;2-J

Bhurgri, Y., Bhurgri, A., Nishter, S., Ahmed, A., Usman, A., Pervez, S., et al. (2006). Pakistan - Country profile of cancer and cancer control 1995–2004. *The Journal of the Pakistan Medical Association, 56*, 124–130.

Bolan, P. J., Nelson, M. T., Yee, D., & Garwood, M. (2005). Imaging in breast cancer: Magnetic resonance spectroscopy. *Breast Cancer Research, 7*, 149–152. https://doi.org/10.1186/bcr1202

Borchartt, T. B., Conci, A., Lima, R. C. F., Resmini, R., & Sanchez, A. (2013). Breast thermography from an image processing viewpoint: A survey. *Signal Processing, 93*, 2785–2803. https://doi.org/10.1016/j.sigpro.2012.08.012

Brancati, N., Frucci, M., & Riccio, D. (2018, June). Multi-classification of breast cancer histology images by using a fine-tuning strategy. In International Conference Image Analysis and Recognition (pp. 771-778). Springer, Cham.

Burton, H., Chowdhury, S., Dent, T., Hall, A., Pashayan, N., & Pharoah, P. (2013). Public health implications from COGS and potential for risk stratification and screening. *Nature Genetics, 45*, 349–351. https://doi.org/10.1038/ng.2582

Caballo, M., Pangallo, D. R., Mann, R. M., & Sechopoulos, I. (2020). Deep learning-based segmentation of breast masses in dedicated breast CT imaging: Radiomic feature stability between radiologists and artificial intelligence. *Computers in Biology and Medicine, 118*, Article 103629. https://doi.org/10.1016/j.compbiomed.2020.103629

Cao, H., Bernard, S., Heutte, L., Sabourin, R., n.d. Dissimilarity-based representation for radiomics applications.

Cao, Z., Duan, L., Yang, G., Yue, T., & Chen, Q. (2019). An experimental study on breast lesion detection and classification from ultrasound images using deep learning architectures. *BMC Medical Imaging, 19*, 1–9. https://doi.org/10.1186/s12880-019-0349-x

Caruana, R., Lawrence, S., & Giles, L. (2001). *Overfitting in neural nets: Backpropagation, conjugate gradient, and early stopping.* Syst: Adv. Neural Inf. Process.

Chang J, Yu J, Han T, Chang HJ, Park E. A method for classifying medical images using transfer learning: a pilot study on histopathology of breast cancer. In2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom) 2017 Oct 12 (pp. 1-4). IEEE.

Charan, S., Khan, M. J., & Khurshid, K. (2018). Breast cancer detection in mammograms using convolutional neural network. 2018 International Conference on Computing, Mathematics and Engineering Technologies: Invent, Innovate and Integrate for Socioeconomic Development, ICoMET 2018 - Proceedings, 2018-Janua, 1–5. https://doi.org/10.1109/ICOMET.2018.8346384.

Chen L-C, Papandreou G, Schrof F, Adam H (2017) Rethinking atrous convolution for semantic image segmentation. arXiv:1706.05587.

Cheng, H. D., Cai, X., Chen, X., Hu, L., & Lou, X. (2003). Computer-aided detection and classification of microcalcifications in mammograms: A survey. *Pattern Recognition, 36*, 2967–2991. https://doi.org/10.1016/S0031-3203(03)00192-4

Cheng, H. D., Shan, J., Ju, W., Guo, Y., & Zhang, L. (2010). Automated breast cancer detection and classification using ultrasound images: A survey. *Pattern Recognition, 43*, 299–317. https://doi.org/10.1016/j.patcog.2009.05.012

Chennamsetty, S. S., Safwan, M., & Alex, V. (2018, June). Classification of breast cancer histology image using ensemble of pre-trained neural networks. In International Conference Image Analysis and Recognition (pp. 804-811). Springer, Cham.

Chougrad, H., Zouaki, H., & Alheyane, O. (2018). Deep convolutional neural networks for breast cancer screening. *Computer Methods and Programs in Biomedicine, 157*, 19–30. https://doi.org/10.1016/j.cmpb.2018.01.011

Conte, L., Tafuri, B., Nunzio, G. De, Portaluri, M., Galiano, A., 2020. Breast Cancer Mass detection in DCE-MRI using Deep-Learning Features Followed by Discrimination of Infiltrative vs In Situ Carcinoma through a Machine-learning approach. https://doi.org/10.20944/preprints202008.0179.v1.

D'Orsi, C. J., Adler, D. D., Ikeda, D. M., Mendelson, E. B., Monsees, B. S., Newstead, G. M., et al. (1994). Breast imaging. *Radiology, 190*, 936–938. https://doi.org/10.1148/radiology.190.3.8115659

Deng, J., Dong, W., Socher, R., Li, L.-J., & Li, K. (2010). ImageNet: A large-scale hierarchical image database 248–255. *Li Fei-Fei*. https://doi.org/10.1109/cvpr.2009.5206848

Dhungel, N., Carneiro, G., & Bradley, A. P. (2016, October). The automated learning of deep features for breast mass classification from mammograms. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 106-114). Springer, Cham.

Diaz, R. A. N., Swandewi, N. N. T., & Novianti, K. D. P. (2019). *Malignancy Determination Breast Cancer Based on Mammogram Image With K-Nearest Neighbor* (Vol. 1,, 233–237.

Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E., Darrell, T., 2014. DeCAF: A deep convolutional activation feature for generic visual recognition. 31st Int. Conf. Mach. Learn. ICML 2014 2, 988–996.

Dromain, C., Balleyguier, C., Adler, G., Garbay, J. R., & Delaloge, S. (2009). Contrast-enhanced digital mammography. *European Journal of Radiology, 69*, 34–42. https://doi.org/10.1016/j.ejrad.2008.07.035

Drukker, K., Giger, M. L., Horsch, K., Kupinski, M. A., Vyborny, C. J., & Mendelson, E. B. (2002). Computerized lesion detection on breast ultrasound. *Medical Physics, 29*, 1438–1446. https://doi.org/10.1118/1.1485995

Ede, F.P., n.d. A SURVEY FEATURE EXTRACTION TECNIQUES FOR MASS DETECTION IN DIGITAL MAMMOGRAM.

Elisa Drelie Gelasca, Jiyun Byun, Boguslaw Obara, B.S.M., 2008. EVALUATION AND BENCHMARK FOR BIOLOGICAL IMAGE SEGMENTATION Elisa Drelie Gelasca , Jiyun Byun , Boguslaw Obara , B . S . Manjunath University of California , Santa Barbara 93106-9560 ,. IEEE Int. Conf. Image Process. 1816–1819.

Ergin, S., & Kilic, O. (2014). A new feature extraction framework based on wavelets for breast cancer diagnosis. *Computers in Biology and Medicine, 51*, 171–182. https://doi.org/10.1016/j.compbiomed.2014.05.008

Evans, D. G. R., & Howell, A. (2007). Breast cancer risk-assessment models. *Breast Cancer Research, 9*, 1–8. https://doi.org/10.1186/bcr1750

Falconi, L.G., Perez, M., Aguilar, W.G., 2019. Transfer Learning in Breast Mammogram Abnormalities Classification with Mobilenet and Nasnet. Int. Conf. Syst. Signals, Image Process. 2019-June, 109–114. https://doi.org/10.1109/IWSSIP.2019.8787295.

Ferranti, C., 2008. Role of sonoelastography in non-palpable breast lesions 2381–2389. https://doi.org/10.1007/s00330-008-1032-8.

Francis, S. V., Sasikala, M., & Saranya, S. (2014). Detection of breast abnormality from thermograms using curvelet transform based feature extraction. *Journal of Medical Systems, 38*. https://doi.org/10.1007/s10916-014-0023-3

Ganggayah, M. D., Taib, N. A., Har, Y. C., Lio, P., & Dhillon, S. K. (2019). Predicting factors for survival of breast cancer patients using machine learning techniques. *BMC Medical Informatics and Decision Making, 19*, 1–17. https://doi.org/10.1186/s12911-019-0801-4

Gardezi, S. J. S., Elazab, A., Lei, B., & Wang, T. (2019). Breast cancer detection and diagnosis using mammographic data: Systematic review. *J. Med. Internet Res., 21*, 1–22. https://doi.org/10.2196/14464

Geras, K.J., Wolfson, S., Shen, Y., Wu, N., Kim, S.G., Kim, E., Heacock, L., Parikh, U., Moy, L., Cho, K., 2017. High-Resolution Breast Cancer Screening with Multi-View Deep Convolutional Neural Networks 1–9.

Girshick, R., 2015. Fast R-CNN. Proc. IEEE Int. Conf. Comput. Vis. 2015 Inter, 1440–1448. https://doi.org/10.1109/ICCV.2015.169.

Gonzalez-Angulo, A. M., Litton, J. K., Broglio, K. R., Meric-Bernstam, F., Rakkhit, R., Cardoso, F., et al. (2009). High risk of recurrence for patients with breast cancer who have human epidermal growth factor receptor 2-positive, node-negative tumors 1 cm or smaller. *Journal of Clinical Oncology, 27*, 5700–5706. https://doi.org/10.1200/JCO.2009.23.2025

Grady, L., 2005. Multilabel random walker image segmentation using prior models. Proc. - 2005 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognition, CVPR 2005 I, 763–771. https://doi.org/10.1109/CVPR.2005.239.

Guan, S., Loew, M., 2017. Breast Cancer Detection Using Transfer Learning in Convolutional Neural Networks. 2017 IEEE Appl. Imag. Pattern Recognit. Work. 1–8. https://doi.org/10.1109/AIPR.2017.8457948.

Guo, Y., Sivaramakrishna, R., Lu, C. C., Suri, J. S., & Laxminarayan, S. (2006). Breast image registration techniques: A survey. *Medical and Biological Engineering and Computing, 44*, 15–26. https://doi.org/10.1007/s11517-005-0016-y

Hamidinekoo, A., Denton, E., Rampun, A., Honnor, K., & Zwiggelaar, R. (2018). Deep learning in mammography and breast histology, an overview and future trends. *Medical Image Analysis, 47*, 45–67. https://doi.org/10.1016/j.media.2018.03.006

Haskins, G., Kruger, U., & Yan, P. (2020). Deep learning in medical image registration : A survey. *Machine Vision and Applications, 31*, 1–18. https://doi.org/10.1007/s00138-020-01060-x

Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., et al. (2017). Brain tumor segmentation with Deep Neural Networks. *Medical Image Analysis, 35*, 18–31. https://doi.org/10.1016/j.media.2016.05.004

He K, Gkioxari G, Dollár P, Girshick R (2017) Mask r-cnn. In: Proceedings of the IEEE international conference on computer vision. pp 2961–2969.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Identity mappings in deep residual networks. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 9908 LNCS, 630–645. https://doi.org/10.1007/978-3-319-46493-0_38.

He, S., Ruan, J., Long, Y., Wang, J., & Dataset, A. (2018). Combining Deep Learning with Traditional Features for Classification and Segmentation of Pathological Images of Breast Cancer. 2018 11th International Symposium on Computational Intelligence and Design (ISCID), 01, 3–6. https://doi.org/10.1109/ISCID.2018.00007.

Heath M, Bowyer K, Kopans D, Moore R, Kegelmeyer WP. The digital database for screening mammography. InProceedings of the 5th international workshop on digital mammography 2000 Jun 11 (pp. 212-218). Medical Physics Publishing.

Heath, M., Bowyer, K., Kopans, D., Kegelmeyer, P., Moore, R., Chang, K., Munishkumaran, S., 1998. Current Status of the Digital Database for Screening Mammography 457–460. https://doi.org/10.1007/978-94-011-5318-8_75.

Heywang-Köbrunner, S. H., Viehweg, P., Heinig, A., & Küchler, C. (1997). Contrast-enhanced MRI of the breast: Accuracy, value, controversies, solutions. *European Journal of Radiology, 24*, 94–108. https://doi.org/10.1016/S0720-048X(96)01142-4

Htay, T.T., Maung, S.S., 2018. Early Stage Breast Cancer Detection System using GLCM feature extraction and K-Nearest Neighbor (k-NN) on Mammography image. Isc. 2018 - 18th Int. Symp. Commun. Inf. Technol. 345–348. https://doi.org/10.1109/ISCIT.2018.8587920.

Hu, Q., Whitney, H. M., & Giger, M. L. (2020). A deep learning methodology for improved breast cancer diagnosis using multiparametric MRI. *Scientific Reports, 10*, 1–12. https://doi.org/10.1038/s41598-020-67441-4

Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q., 2017. Densely connected convolutional networks. Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017 2017-Janua, 2261–2269. https://doi.org/10.1109/CVPR.2017.243.

Huang, Q., Luo, Y., & Zhang, Q. (2017). Breast ultrasound image segmentation: A survey. *International Journal of Computer Assisted Radiology and Surgery, 12*, 493–507. https://doi.org/10.1007/s11548-016-1513-1

Huynh, B. Q., Li, H., & Giger, M. L. (2016). Digital mammographic tumor classification using transfer learning from deep convolutional neural networks. *Journal of Medical Imaging, 3*, Article 034501. https://doi.org/10.1117/1.jmi.3.3.034501

IARC. (2018). *The Global Cancer Observatory – Pakistan, 066*, 2018–2019.
Ibraheem, A.M., Rahouma, K.H., Hamed, H.F.A., 2019. Automatic mri breast tumor detection using discrete wavelet transform and support vector machines. NILES 2019 - Nov. Intell. Lead. Emerg. Sci. Conf. 88–91. https://doi.org/10.1109/NILES.2019.8909345.
Iqbal, Sajid, et al. (2019). Deep learning model integrating features and novel classifiers fusion for brain tumor segmentation. *Microscopy Research and Technique, 82*(8), 1302–1315.
Jannesari, M., & Elemento, O. (2018). Breast Cancer Histopathological Image Classification : A Deep Learning Approach. 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), (Idc), 2405–2412.
Jiao, Z., Gao, X., Wang, Y., & Li, J. (2016). A deep feature based framework for breast masses classification. *Neurocomputing, 197*, 221–231. https://doi.org/10.1016/j.neucom.2016.02.060
Jothilakshmi, G. R., & Raaza, A. (2017, January). Effective detection of mass abnormalities and its classification using multi-SVM classifier with digital mammogram images. In 2017 International Conference on Computer, Communication and Signal Processing (ICCCSP) (pp. 1-6). IEEE.
Kakileti, S. T., Dalmia, A., & Manjunath, G. (2019). Exploring deep learning networks for tumour segmentation in infrared images. *Quantitative InfraRed Thermography Journal, 00*, 1–16. https://doi.org/10.1080/17686733.2019.1619355
Kallenberg, M., Petersen, K., Nielsen, M., Ng, A.Y., Diao, P., Igel, C., Vachon, C.M., Holland, K., Winkel, R., Karssemeijer, N., Lillholm, M., 2010. IEEE TMI SPECIAL ISSUE ON DEEP LEARNING 1 Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring 1–10.
Kelly, K. M., Dean, J., Comulada, W. S., & Lee, S. J. (2010). Breast cancer detection using automated whole breast ultrasound and mammography in radiographically dense breasts. *European Radiology, 20*, 734–742. https://doi.org/10.1007/s00330-009-1588-y
Kerlikowske, K., Grady, D., Rubin, S. M., Sandrock, C., & Ernster, V. L. (1995). Efficacy of screening mammography: a meta-analysis. *Jama, 273*(2), 149–154..
Khan, M. A. (2009). Students ' Corner Letter to the Editor The necessity of awareness of Breast Cancer amongst women in Pakistan. *JPMA, 59*, 2009.
Khan, N., Ahmad, R., Nadeem, M., Hussain, I., 2016. Influence of Education and Socio-Economic Factors on Stage of Cancer Diagnosis : A Study in Pakistani Population Introductio n 225–228.
Khokher, S., Qureshi, M. U., Mahmood, S., & Sadiq, S. (2016). Determinants of advanced stage at initial diagnosis of breast cancer in Pakistan: Adverse tumor biology vs delay in diagnosis. *Asian Pacific Journal Cancer Prevention, 17*, 759–765. https://doi.org/10.7314/APJCP.2016.17.2.759
Khoulqi, I., Idrissi, N., 2019. Breast cancer image segmentation and classification. ACM Int. Conf. Proceeding Ser. https://doi.org/10.1145/3368756.3369039.
Kira, K., Rendell, L., 1992. kira - A practical approach to feature selection - 1992.pdf. Ninth Int. Conf. Mach. Learn.
Kohl, M., Walz, C., Ludwig, F., Braunewell, S., & Baust, M. (2018). In *Assessment of breast cancer histology using densely connected convolutional networks* (pp. 903–913). Cham: Springer.
Koné, I., & Boulmane, L. (2018, June). Hierarchical ResNeXt Models for Breast Cancer Histology Image Classification. In International Conference Image Analysis and Recognition (pp. 796-803). Springer, Cham.
Kooi, T., Litjens, G., van Ginneken, B., Gubern-Mérida, A., Sánchez, C. I., Mann, R., et al. (2017). Large scale deep learning for computer aided detection of mammographic lesions. *Medical Image Analysis, 35*, 303–312. https://doi.org/10.1016/j.media.2016.07.007
Kook, S., Park, H., Lee, Young-rae, Lee, Young-uk, Pae, W., Park, Y., 1999. Evaluation of Solid Breast Lesions with Power Doppler Sonography 231–237.
Kral, P., Lenc, L., 2016. LBP features for breast cancer detection. Proc. - Int. Conf. Image Process. ICIP 2016-Augus, 2643–2647. https://doi.org/10.1109/ICIP.2016.7532838.
Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).
Kukačka, J., Golkov, V., Cremers, D., n.d. Regularization for Deep Learning : A Taxonomy 1–23.
Kwok, S. (2018). Multiclass classification of breast cancer in whole-slide images. In International Conference Image Analysis and Recognition (pp. 931-940). Springer, Cham.
Längkvist, M., Karlsson, L., & Loutfi, A. (2014). Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. *Pattern Recognition Letters, 42*, 11–24. https://doi.org/10.1016/j.patrec.2014.01.008
Lee, R. S., Gimenez, F., Hoogi, A., Miyake, K. K., Gorovoy, M., & Rubin, D. L. (2017). Data Descriptor: A curated mammography data set for use in computer-aided detection and diagnosis research. *Scientific Data, 4*, 1–9. https://doi.org/10.1038/sdata.2017.177
Li, B., Ge, Y., Zhao, Y., Guan, E., Yan, W., 2018. Benign and malignant mammographic image classification based on Convolutional Neural Networks, ACM International Conference Proceeding Series. https://doi.org/10.1145/3195106.3195163.
Li, C. I., Anderson, B. O., Daling, J. R., & Moe, R. E. (2003). Trends in incidence rates of invasive lobular and ductal breast carcinoma. *Journal of the American Medical Association, 289*, 1421–1424. https://doi.org/10.1001/jama.289.11.1421
Liang, C., Bian, Z., Lyu, W., Zeng, D., Ma, J., 2018. A Deep Features-based Radiomics Model for Breast Lesion Classification on FFDM. 2018 IEEE Nucl. Sci. Symp. Med. Imaging Conf. NSS/MIC 2018 - Proc. 1–4. https://doi.org/10.1109/NSSMIC.2018.8824722.
Liang, M., & Hu, X. (2015). Recurrent convolutional neural network for object recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3367-3375).
Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., et al. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis, 42*, 60–88. https://doi.org/10.1016/j.media.2017.07.005
Lu, H.C., Loh, E.W., Huang, S.C., 2019. The Classification of Mammogram Using Convolutional Neural Network with Specific Image Preprocessing for Breast Cancer Detection. 2019 2nd Int. Conf. Artif. Intell. Big Data, ICAIBD 2019 9–12. https://doi.org/10.1109/ICAIBD.2019.8837000.
Lu, W., Li, Z., & Chu, J. (2017). A novel computer-aided diagnosis system for breast MRI based on feature selection and ensemble learning. *Computers in Biology and Medicine, 83*, 157–165. https://doi.org/10.1016/j.compbiomed.2017.03.002
Lu, W. L., Jansen, L., Post, W. J., Bonnema, J., Van de Velde, J. C., & De Bock, G. H. (2009). Impact on survival of early detection of isolated breast recurrences after the primary treatment for breast cancer: A meta-analysis. *Breast Cancer Research and Treatment, 114*(3), 403.
M., S., A., A., E., H., T., M., 2017. Breast Cancer Detection with Mammogram Segmentation: A Qualitative Study. Int. J. Adv. Comput. Sci. Appl. 8, 117–120. https://doi.org/10.14569/ijacsa.2017.081016.
Magny SJ, Shikhman R, Keppke AL. Breast, Imaging, Reporting and Data System (BI RADS) [Updated 2020 Jan 8]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2020 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK459169/.
Mambou, S. J., Maresova, P., Krejcar, O., Selamat, A., & Kuca, K. (2018). Breast cancer detection using infrared thermal imaging and a deep learning model. *Sensors (Switzerland), 18*. https://doi.org/10.3390/s18092799
Marami, B., Prastawa, M., Chan, M., Donovan, M., Fernandez, G., & Zeineh, J. (2018). In *Ensemble Network for Region Identification in Breast Histopathology Slides* (pp. 861–868). Cham: Springer.
Marinelli, R. J., Montgomery, K., Liu, C. L., Shah, N. H., Prapong, W., Nitzberg, M., et al. (2008). The stanford tissue microarray database. *Nucleic Acids Research, 36*(Suppl. 1), 1–7. https://doi.org/10.1093/nar/gkm861
Menhas, R., & Umer, S. (2015). Breast cancer among Pakistani women. *Iranian Journal of Public Health, 44*, 586–587.
Milletari, F., Navab, N., Ahmadi, S., n.d. V-Net : Fully Convolutional Neural Networks for 1–11.
Mishra, S., Prakash, A., Roy, S. K., Sharan, P., & Mathur, N. (2020). In *Breast Cancer Detection using Thermal Images and Deep Learning* (pp. 211–216). IEEE.
Mohebian, M. R., Marateb, H. R., Mansourian, M., Mañanas, M. A., & Mokarian, F. (2017). A Hybrid computer-aided-diagnosis system for prediction of breast cancer recurrence (HPBCR) using optimized ensemble learning. *Computational and Structural Biotechnology Journal, 15*, 75–85. https://doi.org/10.1016/j.csbj.2016.11.004
Moradi, R., Berangi, R., Minaei, B., 2019. A survey of regularization strategies for deep models, Artificial Intelligence Review. Springer Netherlands. https://doi.org/10.1007/s10462-019-09784-7.
Mordang, J. J., Janssen, T., Bria, A., Kooi, T., Gubern-Mérida, A., & Karssemeijer, N. (2016). In *Automatic microcalcification detection in multi-vendor mammography using convolutional neural networks* (pp. 35–42). Cham: Springer.
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. https://doi.org/10.1016/j.acra.2011.09.014
Nahid, A., Kong, Y., 2018. Involvement of Machine Learning for Breast Cancer Image Classification : A Survey 2017.
Pamilo, M., Soiva, M., Anttinen, I., Roiha, M., & Suramo, I. (1991). Ultrasonography of breast lesions detected in mammography screeningm. *Acta Radiologica, 32*, 220–225. https://doi.org/10.3109/02841859109177552
Pavan, A. L., Vacavant, A., Alves, A. F., Trindade, A. P., & de Pina, D. R. (2019). Automatic identification and extraction of pectoral muscle in digital mammography. In World Congress on Medical Physics and Biomedical Engineering 2018 (pp. 151-154). Springer, Singapore.
Pego A, Aguiar P, Bioimaging 2015: 2015, Available from: http:// www.bioimaging2015.ineb.up.pt/dataset.html.
Platania, R., Shams, S., Yang, S., Zhang, J., Lee, K., & Park, S. J. (2017, August). Automated breast cancer diagnosis using deep learning and region of interest detection (bc-droid). In Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics (pp. 536–543).
Ponraj, D., & Jenifer, M. (2011). A survey on the preprocessing techniques of mammogram for the detection of breast cancer. *The Journal of Emergency, 2*, 656–664.
Pons, G., Martí, R., Ganau, S., Sentís, M., Martí, J., 2013. Feasibility study of lesion detection using deformable part models in breast ultrasound images. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 7887 LNCS, 269–276. https://doi.org/10.1007/978-3-642-38628-2_32.
Punitha, S., Amuthan, A., & Joseph, K. S. (2018). Benign and malignant breast cancer segmentation using optimized region growing technique. *Future Computing and Informatics Journal, 3*, 348–358. https://doi.org/10.1016/j.fcij.2018.10.005
Rahman, M., Hussain, M.G., Hasan, M.R., Sultana, B., Akter, S., 2020. Detection and Segmentation of Breast Tumor from MRI Images Using Image Processing Techniques. Proc. 4th Int. Conf. Comput. Methodol. Commun. ICCMC 2020 720–724. https://doi.org/10.1109/ICCMC48092.2020.ICCMC-000134.
Ramani, R., Suthanthiravanitha, S., & Valarmathy, S. (2012). A survey of current image segmentation techniques for detection of breast cancer. *International Journal of Engineering Research and Applications, 2*, 1124–1129.

Ramani, R., Valarmathy, S., & Suthanthira Vanitha, N. (2013). Breast cancer detection in mammograms based on clustering techniques- A survey. *International Journal of Computers and Applications, 62*, 17–21. https://doi.org/10.5120/10123-4885

Rashed, E., Samir Abou El Seoud, M., 2019. Deep learning approach for breast cancer diagnosis. ACM Int. Conf. Proceeding Ser. 243–247. https://doi.org/10.1145/3328833.3328867.

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788).

Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 39*, 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031

Reynolds, D. A. (2009). Gaussian mixture models. *Encyclopedia of Biometrics, 741*.

Riis, C., Lernevall, A., Sorensen, F.B., Nygaard, H., n.d. 3D Ultrasound-Based Evaluation of Lesions in the Uncompressed Breast 151–155.

Rodriguez-Cristerna, A., Guerrero-Cedillo, C.P., Donati-Olvera, G.A., Gomez-Flores, W., Pereira, W.C.A., 2017. Study of the impact of image preprocessing approaches on the segmentation and classification of breast lesions on ultrasound. 2017 14th Int. Conf. Electr. Eng. Comput. Sci. Autom. Control. CCE 2017 7. https://doi.org/10.1109/ICEEE.2017.8108826.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 9351, 234–241. https://doi.org/10.1007/978-3-319-24574-4_28|.

Saha, M., Chakraborty, C., & Racoceanu, D. (2018). Efficient deep learning model for mitosis detection using breast histopathology images. *Computerized Medical Imaging and Graphics, 64*, 29–40. https://doi.org/10.1016/j.compmedimag.2017.12.001

Sakai, A., Onishi, Y., Matsui, M., Adachi, H., Teramoto, A., Saito, K., et al. (2020). A method for the automated classification of benign and malignant masses on digital breast tomosynthesis images using machine learning and radiomic features. *Radiological Physics and Technology, 13*, 27–36. https://doi.org/10.1007/s12194-019-00543-5

Saphner, T., Tormey, D. C., & Gray, R. (1996). Annual hazard rates of recurrence for breast cancer after primary therapy. *Journal of Clinical Oncology, 14*, 2738–2746. https://doi.org/10.1200/JCO.1996.14.10.2738

Shan, J., Cheng, H. D., & Wang, Y. (2012). Completely automated segmentation approach for breast ultrasound images using multiple-domain features. *Ultrasound in Medicine & Biology, 38*, 262–275. https://doi.org/10.1016/j.ultrasmedbio.2011.10.022

Shelhamer, E., Long, J., & Darrell, T. (2017). Fully convolutional networks for semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 39*, 640–651. https://doi.org/10.1109/TPAMI.2016.2572683

Silva, L. F., Saade, D. C. M., Sequeiros, G. O., Silva, A. C., Paiva, A. C., Bravo, R. S., et al. (2014). A new database for breast research with infrared image. *Journal of Medical Imaging and Health Informatics, 4*, 92–100. https://doi.org/10.1166/jmihi.2014.1226

Simonyan, K., Zisserman, A., 2015. Very deep convolutional networks for large-scale image recognition. 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc. 1–14.

Singh, A., 2016. Breast tumour detection using segmentation technique from CT scan.

Smirnov, E. A., Timoshenko, D. M., & Andrianov, S. N. (2014). Comparison of regularization methods for imagenet classification with deep convolutional neural networks. *AASRI Procedia, 6*, 89–94. https://doi.org/10.1016/j.aasri.2014.05.013

Soliman, O.O., Sweilam, N.H., Shawky, D.M., 2019. Automatic Breast Cancer Detection Using Digital Thermal Images. 2018 9th Cairo Int. Biomed. Eng. Conf. CIBEC 2018 - Proc. 110–113. https://doi.org/10.1109/CIBEC.2018.8641807.

Song, E., Jiang, L., Jin, R., Zhang, L., Yuan, Y., & Li, Q. (2009). Breast mass segmentation in mammography using plane fitting and dynamic programming. *Academic Radiology, 16*, 826–835. https://doi.org/10.1016/j.acra.2008.11.014

Spanhol, F. A., Oliveira, L. S., Petitjean, C., & Heutte, L. (2016). A dataset for breast cancer histopathological image classification. *IEEE Transactions on Biomedical Engineering, 63*, 1455–1462. https://doi.org/10.1109/TBME.2015.2496264

Sree, S. V. (2011). Breast imaging: A survey. *World Journal of Clinical Oncology, 2*, 171. https://doi.org/10.5306/wjco.v2.i4.171

SUCKLING J, P. (1994). The mammographic image analysis society digital mammogram database. Digital Mammo, 375-386.

Sutskever, I., Vinyals, O., Le, Q. V., 2014. Sequence to sequence learning with neural networks. Adv. Neural Inf. Process. Syst. 4, 3104–3112.

Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. A. (2017, February). Inception-v4, inception-resnet and the impact of residual connections on learning. In Thirty-first AAAI conference on artificial intelligence.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2016. Rethinking the Inception Architecture for Computer Vision. Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2016-Decem, 2818–2826. https://doi.org/10.1109/CVPR.2016.308.

Taheri, M., Hamer, G., Son, S.H., Shin, S.Y., 2016. Enhanced breast cancer classification with automatic thresholding using SVM and Harris corner detection. Proc. 2016 Res. Adapt. Converg. Syst. RACS 2016 56–60. https://doi.org/10.1145/2987386.2987420.

Tahir, Bilal, et al. (2019). Feature enhancement framework for brain tumor segmentation and classification. *Microscopy Research and Technique, 82*(6), 803–811.

Taylor, L., Nitschke, G., 2017. Improving Deep Learning using Generic Data Augmentation.

The Global Cancer Observatory, G., 2018. Breast Cancer. Source: Globocan 2018. World Heal. Organ. 876, 2018–2019.

Tran, D., Bourdev, L., Fergus, R., Torresani, L., Paluri, M., 2015. Learning spatiotemporal features with 3D convolutional networks. Proc. IEEE Int. Conf. Comput. Vis. 2015 Inter, 4489–4497. https://doi.org/10.1109/ICCV.2015.510.

Verburg, E., Wolterink, J. M., de Waard, S. N., Išgum, I., van Gils, C. H., Veldhuis, W. B., et al. (2019). Knowledge-based and deep learning-based automated chest wall segmentation in magnetic resonance images of extremely dense breasts. *Medical Physics, 46*, 4405–4416. https://doi.org/10.1002/mp.13699

Veta, M., van Diest, P. J., Willems, S. M., Wang, H., Madabhushi, A., Cruz-Roa, A., et al. (2015). Assessment of algorithms for mitosis detection in breast cancer histopathology images. *Medical Image Analysis, 20*, 237–248. https://doi.org/10.1016/j.media.2014.11.010

Vijayarajeswari, R., Parthasarathy, P., Vivekanandan, S., & Basha, A. A. (2019). Classification of mammogram for early detection of breast cancer using SVM classifier and Hough transform. *Measurement, 146*, 800–805. https://doi.org/10.1016/j.measurement.2019.05.083

Vogel, V. G. (1999). Assessing risk of breast cancer: Tools for evaluating a patient's 5-year and lifetime probabilities. *Postgraduate Medicine, 105*, 49–58. https://doi.org/10.1080/00325481.1999.11444922

Wang, C., Li, M., Liu, X., Liu, Zaiyi, Zang, Y., Liu, Zhenyu, Dong, D., Chang, C., 2017. Semi - automated Enhanced Breast Tumor Segmentation for CT Image 648–651.

Wang, H., Feng, J., Zhang, Z., Su, H., Cui, L., He, H., et al. (2018). Breast mass classification via deeply integrating the contextual information from multi-view data. *Pattern Recognition, 80*, 42–52. https://doi.org/10.1016/j.patcog.2018.02.026

Wang, K., Khan, N., Chan, A., Dunne, J., & Highnam, R. (2019). In *Deep Learning for Breast Region and Pectoral Muscle Segmentation in Digital Mammography* (pp. 78–91). Cham: Springer.

Wang, N., Bian, C., Wang, Y., Xu, M., Qin, C., Yang, X., et al. (2018). In *Densely Deep Supervised Networks with Threshold Loss for Cancer Detection in Automated Breast Ultrasound* (pp. 641–648). Cham: Springer.

Wang, Y., Sun, L., Ma, K., & Fang, J. (2018). In *Breast Cancer Microscope Image Classification Based on CNN with Image Deformation* (pp. 845–852). Cham: Springer.

Wang, Y., Qin, C., Lin, C., Di, L., Xu, M., Luo, X., Wang, T., Li, A., Ni, D., 2020a. 3D Inception U-net with Asymmetric Loss for Cancer Detection in Automated Breast Ultrasound. Med. Phys. https://doi.org/10.1002/mp.14389.

Wang, Y., Wang, N., Xu, M., Yu, J., Qin, C., Luo, X., et al. (2020). Deeply-supervised networks with threshold loss for cancer detection in automated breast ultrasound. *IEEE Transactions on Medical Imaging, 39*, 866–876. https://doi.org/10.1109/TMI.2019.2936500

Wang, Z., Li, M., Wang, H., Jiang, H., Yao, Y., Zhang, H., et al. (2019). Breast cancer detection using extreme learning machine based on feature fusion With CNN deep features. *IEEE Access, 7*, 105146–105158. https://doi.org/10.1109/access.2019.2892795

Weickert, J., Weickert, J., Schnörr, C., 2000. PDE-Based Preprocessing of Medical Images Introd uction.

Wu, T., Stewart, A., Stanton, M., McCauley, T., Phillips, W., Kopans, D. B., et al. (2003). Tomographic mammography using a limited number of low-dose cone-beam projection images. *Medical Physics, 30*, 365–380. https://doi.org/10.1118/1.1543934

Xie, S., Girshick, R., Doll, P., 2017. Aggregated Residual Transformations for Deep Neural Networks http://arxiv.org/abs/1611.05431v2. https://doi.org/10.1109/CVPR.2017.634.

Xie, S., & Tu, Z. (2017). Holistically-nested edge detection. *International Journal of Computer Vision, 125*, 3–18. https://doi.org/10.1007/s11263-017-1004-z

Yap, M. H., Edirisinghe, E. A., & Bez, H. E. (2008). A novel algorithm for initial lesion detection in ultrasound breast images. *The Journal of Applied Clinical Medical Physics, 9*, 181–199. https://doi.org/10.1120/jacmp.v9i4.2741

Yap, M. H., Goyal, M., Osman, F., Martí, R., Denton, E., Juette, A., et al. (2020). Breast ultrasound region of interest detection and lesion localisation. *Artificial Intelligence in Medicine, 107*, Article 101880. https://doi.org/10.1016/j.artmed.2020.101880

Yap, M. H., Pons, G., Martí, J., Ganau, S., Sentís, M., Zwiggelaar, R., et al. (2018). Automated breast ultrasound lesions detection using convolutional neural networks. *IEEE J. Biomed. Heal. Informatics, 22*, 1218–1226. https://doi.org/10.1109/JBHI.2017.2731873

Yemini, M., Zigel, Y., Lederman, D., 2019. Detecting Masses in Mammograms using Convolutional Neural Networks and Transfer Learning. 2018 IEEE Int. Conf. Sci. Electr. Eng. Isr. ICSEE 2018 1–4. https://doi.org/10.1109/ICSEE.2018.8646252.

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 Trans. Nanobioscience 17, 237–242. https://doi.org/10.1109/TNB.2018.2845103.

Zhang, Y., Zhang, B., Coenen, F., & Lu, W. (2013). Breast cancer diagnosis from biopsy images with highly reliable random subspace classifier ensembles. *Machine Vision and Applications, 24*, 1405–1420. https://doi.org/10.1007/s00138-012-0459-8

Zhang, Z., Wang, Y., Zhang, J., Mu, X., 2019. Comparison of multiple feature extractors on Faster RCNN for breast tumor detection. 2019 8th Int. Symp. Next Gener. Electron. ISNE 2019 1–4. https://doi.org/10.1109/ISNE.2019.8896490.

Zheng, B., Yoon, S. W., & Lam, S. S. (2014). Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms. *Expert Systems with Applications, 41*, 1476–1482. https://doi.org/10.1016/j.eswa.2013.08.044

Zoph, B., Vasudevan, V., Shlens, J., & Le, Q. V. (2018). Learning Transferable Architectures for Scalable Image Recognition. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 8697–8710. https://doi.org/10.1109/CVPR.2018.00907.