JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Machine Learning for Breast Cancer Detection and Diagnosis:

S. Meenakshi , Dr. A.S. Aneetha

Research Scholar, Associate Professor,

Vels Institute Of Science, Technology & Advanced Studies.

Abstract

One of the most common diseases that kill women around the world is breast cancer. When a suspicion is expressed, routine exams, typically with digital mammography, should be performed. (DM), magnetic resonance imaging (MRI), and infrared thermography Microwave, microscopic (histological) pictures, and ultrasound it might be advised to use photos, additional instruments, or testing. Now a days, several software and hardware applications use various methods for producing results of a high caliber, particularly the machine learning methods. This publication provides a thorough survey to assess the majority of the precise methods now being used to detect and diagnose breast cancer are both carried out. Besides, varied hardware and software, both for profit and not for profit are discussed in the with their benefits and drawbacks. Detection and diagnosis of breast lesions. This research shows that numerous techniques used for developed.

INDEX

Breast cancer, medical modality, methods for detection and diagnosis, CAD systems, and computer-aided design techniques.

Introduction

Many people around the world are affected by breast cancer, a condition that has an impact on the breast's glandular epithelium (Al-Tam, 2015; Harris & Vogel, 1997).

Other organs are susceptible to the attack of this disease because this lesion typically spreads to encompassing tissues, and even infiltrate the lymphatic or circulatory systems (Harris & Vogel, 1997).

Periodic examinations, which often include digital mammography, ultrasound, microscopic (histological) pictures, magnetic resonance imaging (MRI), infrared thermography, may be proposed by the subject-matter expert if a suspicion is developed (Harris & Vogel, 1997). Therefore, to enhance the survival rate, early detection and diagonsis of breast cancer are crucial.

Numerous factors affect how breast cancer appears, but established researchers are still unsure of exactly what causes healthy cells to develop into cancer. Generally speaking, some elements including especially inherited DNA mutations, density of breasts, and hormones may raise the likelihood that cancer will manifest itself (American Cancer Society, 2020). The women age can raise their risk of breast cancer; for example, invasive breast cancer may occur in one out of every eight women under the age of 45. In contrast, according to the American Cancer Society's 2020 report, there are roughly two to three times as many women over the age of 55 who could get invasive breast cancer. Additionally, a woman's risk of getting cancer rises if her breasts are thick, compared to women who have less dense breasts. In addition to being overweight, drinking alcohol, not exercising, and not having children they all can the risk of breast cancer has increased. (Al-Tam, 2015; Harris & Vogel, 1997). Breastfeeding while a baby is young has been linked to a lower incidence of breast cancer, according to several research (Al-Tam, 2015; Harris & Vogel, 1997). Additionally, by walking for 1.25 to 2.5 hours per week, the chance of developing breast cancer can be lowered to 18%. (Al-Tam, 2015; Harris & Vogel, 1997).

If the cancerous cells are contained, cancer may not spread. Cancer can be when it has penetrated the basal membrane, but it can also occur when cancer cells that have not yet crossed by the way of basal membrane are totally confined within the lobe of the ducts. and migrated into adjacent tissue (National Breast Cancer Registry) 2020 (Foundation). Breast cancer comes in a variety of forms, some of which are uncommon. Lobular carcinoma in situ (LCIS), ductal carcinoma in situ (DCIS), and infiltrating cancer are the three most prevalent kinds. The National Breast Cancer Foundation defines ductal/lobular carcinoma as 2020). However, fewer frequent cancer types exist, like triple-negative, inflammatory breast cancer, Phyllodes tumors, nipple Paget illness, as well as Angiosarcoma.

Modalities are tools that doctors employ to help with breast cancer identification and early detection. Although the most popular form of often utilized treatment for cancer screening, combining other modalities including magnetic resonance imaging, mammography and ultrasound together is an effective strategy to increase the accuracy of early detection (Deo, 2015). For instance, a physician can first mammography should be used to detect breast cancer., plus if any worrisome area have shown up, MRI or Ultrasound may be employed in addition to that.

Since the development of medical modalities, Techniques of Machine learning have utilized to identify and detection of breast cancer (Saxena & Gyan Chandani, 2020). A ML-based CAD model was created in 1993 by Street et al. and implemented for the first time at the University of Wisconsin (Saxena & Gyan Chandani, 2020). In order to dramatically lower the risk of cancers that affect humans, including as skin, breast, brain, prostate, colon, bladder, liver, cervical cancers, various researchers have been working to build a variety of CAD systems.

This essay aims to offer an in-depth analysis of encourage additional research. Additionally, to aid in the investigation, we give a taxonomy of medical modalities in this survey. analysis of breast cancer detection and comprehension employing machine learning approaches to diagnose. the remaining the following organizational structure: The discussion of the breast cancer's morphology and structure. Further, a For detecting, a combination of novel and ancient modalities has been used. and breast cancer classification are presented. Additionally, the Section IV shows the properties of several CAD systems.

Imaging's primary objective is to identify and classify cancer at the time when it is most curable. In order to analyze breast imaging, some techniques have been developed that primarily rely on three concepts: first, determine whether both normal malignant tissues ; second, pinpoint the breast's abnormalities to aid in additional testing or treatment; and third, characterize the irregularities to assist when making decisions upon identification (Cancer et al., 2001).

II. Analysis of breast cancer

However, some women with symptomless cancer making routine imaging breast cancer so crucial (Cancer et al., 2001). Breast cancer is occasionally discovered following sign show up. Certain medical imaging techniques, such as asymmetry, angiogenesis, architectural distortion, tissue masses, and microcalcifications, are capable to detect structural variations in tumors (Cancer et al., 2001). These methods rely on the tissue's Physical, mechanical, electrical, chemical and biological properties. Numerous techniques have been employed during the past 20 years, but only a select few are universally endorsed by doctors.

A. Outdated breast cancer therapies

1) Examination of the breast

Radiologists can examine breast tissue using a mammography (Helvie & Patterson, 2014). This technology flattens or compresses the breast using two plates to stretch the tissue apart, allowing for the clearest images possible (Helvie & Patterson, 2014). Figure 1 depicts the analogue and digital mammography options that are available (LBN Medical, 2020). In contrast to digital mammography, which an x-ray detector with digital beams, low-dose mammography is recorded analog, x-ray teams on used video taps. Since a reader CR , like the XL II Fuji Capsula, is required to turn from an analog to a digital image, it takes longer for the analogue equipment to prepare an image for collection in the PACS system. Full-field digital mammography (FFDM), also characterized as a 3D mammography or breast tomosynthesis, may provide breast images in three dimensions saved on a system, whereas digital mammography can just produce 2D Picture. Additionally, according to Guo et al. (2018), digital mammography has a radiation exposure that is 30–40% lower than analogue mammography. According to numerous studies, there is a Compared to 2D mammography, 3D mammography can identify more breast cancer, particularly in those with denser breast tissue (Society, 2020). The two techniques are used to help with the first diagnosis of breast cancer, Although the atypical breast tissue can be seen, this does not imply that any suspected cancerous tissue is present.

They assist in determining whether or not additional testing is required. According to doctors, the mammography is the procedure most frequently used to find breast cancer early in the world (Helvie & Patterson, 2014). However, only a few types of breast cancer, like lumps and calcifications, can be detected by a mammogram, not all types. The mammography, which can be uncomfortable for younger women and may result in unneeded biopsies, does not effectively detect abnormalities in dense-breast. As a result, additional modalities like MRI or ultrasound can advised (Helvie & Patterson, 2014).

2) Sonography

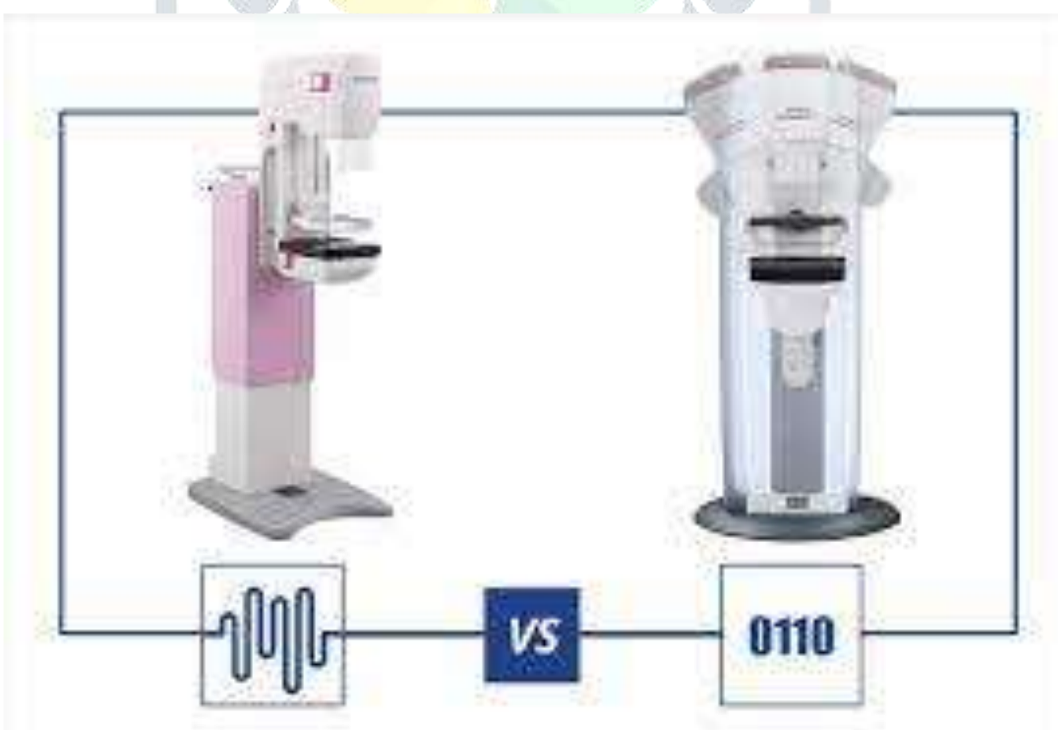


Figure 2 depicts a tool that makes use of sound waves to produce images. This equipment, known as an ultrasound, is primarily used to detect more pronounced breast alterations challenging to detect on mammograms, such as cysts with detefluid, tumors, and other alterations within the thick breast tissues (Guo et al., 2018). Additionally, according to Guo et al. (2018), It is applicable to direct one for a biopsy towards a questionable location in order to remove some cells for cancer testing. This modality is readily available throughout the world, exposes users to no radiation, and is less expensive than many various imaging methods (Guo et al., 2018). The methods for detecting breast cancer using ultrasound imaging include three-dimensional ultrasound, breast cancer computer-aided detection, automatic breast ultrasound, , ultrasound elastography, contrast-enhanced ultrasound, and breast ultrasound imaging (Guo et al., 2018). A form of ultrasound called elastosonography regular technique that can distinguish between malignant and bengin tumors through assessing the firmness or consistency of the tissues, lesions 2018 (Guo et al.). This method isn't 100% effective and can cause lesions. Has a few flaws, such the fact that it can't distinguish tumors from surrounding tissues when the elastic characteristics of lesions are similar. The depth of the elastography image has an impact on its quality. In order to overcome the limitations of elastography and reduce the frequency of needless biopsies, it has been suggested in numerous research to combine B-mode and elastography techniques (Guo et al., 2018). The 2D B-mode image is made up of vivid dots that are ultrasound echoes. Dots' brightness makes it possible to see and count anatomical structures. The capacity of contrast-enhanced ultrasonography to depict breast cancers' vascular structure and perfusion, in addition to numerical variables according to the time or intensity curve, is helpful in separating benign from malignant lesions (Guo et al., 2018). Additionally, a three-dimensional, automated breast scanner US also provide helpful details on breast lesions, but the last choice of which ultrasound technology to use is left up to a subject matter expert like a radiologist. Finally, 3D ultrasound is a technique that may aid since it can show the architecture and geographical locations of a breast tumor, in the early detection and diagnosis of breast cancer. On the other hand, certain studies on the value of utilizing 3D modality are discussed in (Guo et al., 2018). For calculating the receiver operating characteristic curve's area under the curve (0.51, 0.76 correspondingly), 3D is superior to 2D; nevertheless, Combining 3D with mammography will be preferred as this enables attaining to 0.90. (Guo et al., 2018).

Figure 2. Samsung WS80A Elite, an ultrasound modality (LBN Medical, 2020)

3) Magnetic Resonance Imaging (MRI)

As seen in figure 3, MRI is a technique that may assess the breast tissue density, morphological alterations, and the condition of the armpit, pectoral muscle edge and skin. The use of MRI, which has a large NPV and securely aids in the identification of malignancy, is recommended when cancer is not detected by mammography or clinical means. According to numerous studies, MRI is an effective way to examine young women's breasts, who have a high chance of developing breast cancer. Higher risk applies to women with a strong family history of breast cancer, according to the American Cancer Society. or ovarian cancer should undergo both an annual mammogram and a breast cancer screening MRI (Guo et al., 2018). Those whose chances of having breast cancer over their lifespan that is greater than or equivalent to 20–25% should also use this screening method.

The mammography test should be used in conjunction with the MRI test; it should not be used as a substitute for it. In females with a family or BRCA1/2 mutation propensity, final examination with mammography and MRI increases the metastasis-free existence rate (Guo et al., 2018). In addition, MRI may help when making a ductal carcinoma in situ diagnosis (DCIS) (Guo et al., 2018). On the Sarica and Uluc (Sarica & Uluc, 2014) reported research that indicated that sonographic BI-RADS 4 lesions cannot currently be scored using MRI alone without a biopsy. Furthermore, when using ultrasonographic the specificity (SP) of MRI for BI-RADS 4 lesions was only 56.7%. In MRI breast imaging, true positive lesions need to be identified, a biopsy is still required (Guo et al., 2018). On the other end, few hospitals offer MRI because it is an expensive modality (Cancer et al., 2001).

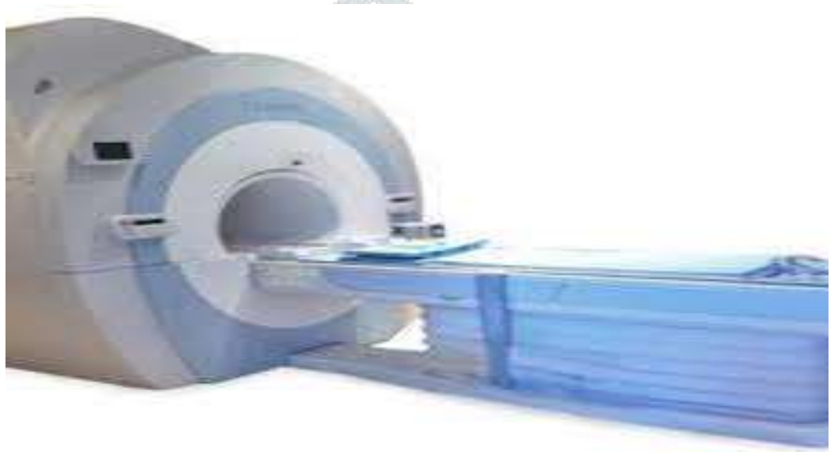


Figure 3 Toshiba Vantage Titan 1.5T, an MRI modality (LBN Medical, 2020)

B. Innovative Breast Cancer Treatments

The most popular screening methods for breast cancer at the moment are mammography, ultrasounds, and breast MRIs. New technologies, for instance, breast tomosynthesis (3D mammography), are being adopted in few facilities, though (Society, 2020). In addition, research has been done on various modalities to see if their effectiveness and quality are comparable to or perhaps even superior to those now in use (Society, 2020):

Positron emission mammography (PEM), optical imaging tests, contrast-enhanced mammography (CEM), elastography and electrical impedance imaging (EIT), are all examples of molecular breast imaging (MBI), additionally referred to as scintimammography or BSGI (breast-specific gamma imaging).

Breast-specific gamma imaging (BSGI), sometimes referred to as scintimammography or molecular breast imaging (MBI), is a kind of mammography. Other types of mammography includes Elastography, Electrical impedance imaging (EIT), Optical imaging tests.

Histopathology, a way of examining infected excised breast tissue under a using a microscope to suggest treatments for the disease, is another device for detection and diagnosing breast cancer (Bagchi et al., 2020). Though, the intricate visual patterns on histopathology slides make it difficult to distinguish between benign and cancerous tissues 2020 (Saxena & Gyanchandani). Consequently, current research has put a strong emphasis on automating this procedure (Zhou et al., 2020).

Gene expression analysis employs DNA microarray technology, particularly to identify mutated genes. According to other studies, Up to 60% of cases may be caused by changes in the BRCA1 and BRCA2 genes of hereditary cases of ovary and breast cancer (National Human Genome Research Institute, 2020). Blood of the patient's

DNA can be sampled in order to identify this mutation. The sample's DNA molecules are then divided into more manageable chunks, with the control DNA being labeled red and the patient's DNA being labeled green using fluorescent dye. In order for the synthetic DNA on a chip to hybridize, the green and red sets are placed in the chip. Genes do not contain a mutation if the chip's standard sequences are joined by the green and red samples; otherwise, this is the case (National Human Genome Research Institute, 2020).

A tool called Fourier-transform infrared spectroscopy (FTIR) a liquid, gas, or solid's absorption or emission in the infrared spectral range has been measured. This tool used employed by several researchers to categorize breast cancer.

Self-examination of breast is the practice of performing routine, independent breast exams (Marcia Boraas, MD, 2019). This procedure is crucial because it aids when it comes to early breast cancer detection cancer, which raises the chances of a successful course of treatment. You must see a doctor if any of these modifications have happened in your breast.:

1) Skin that is puckered, dimpling, or bulging

2) The nipple may undergo any modification, such as shifting its position or making it appear inverted (pushed inward rather than protruding out).

3) pain, rash, redness, or swelling

III The CAD system

Computer-assisted methods of both detection and diagnosis frequently make use of CAD. Generally speaking, CAD is referred to when using computer-aided detection (CADe), it can identify potentially problematic spots in images and as computer-aided diagnosis (CADx) when it can evaluate and differentiate between breast cancer can be benign or aggressive. Additionally, the CAD system now includes a class of systems called "computer-aided simple triage," which employs image analysis and AI to analyze, diagnose, and alert without the radiologist's involvement, inform a radiologist about any suspected findings on the most recent imaging. (Yanase & Triantaphyllou, 2019). In order to detect problematic areas that might be missed on pictures produced by modalities and to distinguish between benign and malignant breast cancers, computer-aided design (CAD) methods were developed (Guo et al., 2018). Segmentation, feature extraction, selection and preprocessing, and classification are the four main stages of a CAD system.

The Lunit INSIGHT MMG technology, developed by the Korean business Lunit using deep learning algorithms, aids radiologists in deciphering mammography pictures (Lunit, 2020). More over 50000 of the 200000 instances in the training set—a total of 200000 cases—are cancer patients. This program can automatically identify lesions on mammograms that may be signs of breast cancer, such as calcification, mass, asymmetry, and distortion. In essence, it uses color or outline to identify worrisome areas in mammography pictures, and it displays an anomaly score that indicates the likelihood of developing breast cancer in the area that was identified (Lunit, 2020). This method is often developed to increase detection rates and reduce memory rates while evaluating mammography images. Additionally, it has a 96% ROC AUC (Area Under the Receiver Operating Characteristic Curve) accuracy rating (Lunit, 2020).

Breastscape is a product made by Olea Medical that includes the BreastApp and BreastLoc programs to assist radiologists. BreastApp is a program that can MRI imaging can be used to find breast cancer automatically and generate score report of a BI-RADS. A diagnostic biopsy tool called BreastLoc also functions with MRI images.Additionally, this program has all of the grids and needles in its catalog loaded (Olea Medical, 2020).

Five programs including T2 Mapping, Textures Analysis, Perfusion-Pharmacokinetics Modeling, Body Diffusion-Weighted Imaging (DWI) - IVIM, and Body Diffusion-Weighted Imaging (DWI) - ADC, have been created by the QUIBIM corporation for the objective of using MRI scans to identify breast cancer (Quibim, 2020).

A lesion's heterogeneity, which may be an aggressive tumor, can be found via textures analysis (Quibim, 2020). T2 Mapping, in contrast, was primarily developed to deliver relaxometry readings and parametric maps. Additionally, The program Body Diffusion-Weighted Imaging (DWI) - IVIM is used to determine molecules of water diffusivity and allows for the distinction between fast and slow water molecules diffusivity (Quibim, 2020). While the application of perfusion-pharmacokinetics modeling can be utilized to categorize the many biological phenotypes of tumor behavior. The ability to compute the tissue water molecule diffusivity is also included into Body Diffusion-Weighted Imaging(DWI) - ADC (Quibim, 2020). Using 101 radiologists, the scientist of (Rodriguez-Ruiz et al., 2019) presented a comparison research to examine the effectiveness of Transpara, including its employed AI algorithms. They found that Transpara performed better than the typical breast radiologist in terms of cancer detection accuracy.

The Volpara Density application, created by Volpara Solutions, uses AI to assess breast density and delivers a reliable score for it (Volpara Health, 2020). Additionally, this application has "Class A" certification and is the FDA assigns a "Class B" classification. (Radboud University Medical, 2020) and is compatible with a facus on 2D and 3D digital modalities. Additionally, Aspen® Breast is Volpara Solutions' most cutting-edge creation. According to reports, A warning system is ASPEN. that can predict the likelihood that a woman will acquire breast cancer, particularly in the decade immediately following her current age and throughout the course of her lifetime (Volpara Health, 2020). The usefulness of adopting MRI screening as an additional tool to work in tandem with mammography screening in women with exceptionally thick breasts was demonstrated in a study that was presented, however (Bakker et al., 2019). Additionally, a collection of mammography pictures with 4 distinct densities were graded using Volpara 1.5, and the women in these photos were then invited to an MRI screening. The number of women diagnosed using this method is lower than simple mammography machine operation (Bakker et al., 2019). Additionally, Volpara was utilized in full-field digital mammograms, which assist precision medicine, to find risk factors in the thick breast. (Brentnall et al., 2019).

Lunchtime Densitas The primary purpose of Denitasai remedies, which also includes Densitas qualityai, Densitas riskai, and Densitas densityai, , which helps radiologists identify and diagnose breast cancer utilizing mammography pictures (Densitas® Inc, 2020). These solutions are typically made for managing health system administrators, putting in place a framework for advanced analytics, and controlling diagnostic pictures by managers, technicians, and quality (Densitas® Inc, 2020). The qualityaiTM tool was developed to manage picture quality, whilst the densityaiTM tool delivers BI-RADS density scale assessments. Additionally, Densitas riskaiTM may evaluate breast cancer risk using just two clinical indicators and image-derived factors (Densitas® Inc, 2020). Densitasai products were classified as "Class B" by the FDA and are approved; nevertheless, the approved unknown level(Radboud University Medical, 2020).

As opposed to that, a device called Vara, created by Merantix Healthcare, is used to identify normal mammograms and flag them for assessment by human experts .

Typically, For human experts, this tool generates pre-written reports., requiring them to review the examinations among the reports produced. However, Vara's screening methodology can be used to evaluate the remaining examinations without the use of warning signs or written reports (Vara, 2020). An application for viewing and reporting mammography images, Vara is created utilizing machine learning approaches (Vara, 2020).

Triage Mammography, an autonomous AI technology created by Zebra Medical Vision, can find worrisome areas in image of a 2D mammogram (Zebra Medical Vision, 2020).

Typically, a platform for analytics imaging receives mammograms from Zebra where a problematic breast area can be found. Triage Mammography the creates another result and sends it to the working station of radiologist (Zebra Medical Vision, 2020).

Using 3D breast ultrasound pictures, the QVCAD System primarily helps radiologists find breast lesions (QView Medica, 2020). Due to their dense breast tissues, which make it difficult to discern between benign and malignant tissue, women who employ the 3D breast ultrasound modality typically get negative mammograms (QView

Medica, 2020). The QVCAD can distinguish between suspicious and normal breast regions thanks to the implementation of pattern recognition techniques and artificial neural networks (QView Medica, 2020). Additionally, the FDA has classified it as "ClassII" (Radboud university medical, 2020).

Breast Screening, a new application described by the authors in (Calisto et al., 2020), is capable of multimodal breast cancer diagnosis and can display various medical pictures (US, MRI, and Mammography) in one interface. By using this program, examining lesions in various medical photos takes less time. This study provides evidence that employing a multimodal view rather than a signal-modal view is preferable when viewing medical images.

A CAD system that may effectively lower the reviewing time as well as the misdetection rate was suggested in (Moon et al., 2020). This system's inputs were Convolutional neural networks (CNNs) were used as the classifier in 3-D automated breast ultrasound (ABUS) images.

YOLO-based deep learning algorithms for a CAD system has been proposed for the identification and finding of breast cancer. (Al-Antari et al., 2020). This system receives a set of mammography pictures from the INbreast and DDSM datasets. The Regular feedforward CNN, ResNet-50 and Inception ResNet-V2 classifiers were developed using the YOLO alarm. The best accuracy is achieved by the InceptionResNet-V2 classifier, which has values of 95.32% for the INbreast dataset and 97.50% for the DDSM dataset.

Furthermore, full-field digital imaging to separate benign tumors from malignancy mammography pictures derived from the INbreast dataset, the authors of (Aly et al., 2021) presented a YOLO-based deep learning algorithms based CAD system. The suggested system's total accuracy was 94.2% using, ResNet and Inception algorithms for feature extraction, the YOLO v3 methodology to detect masses and k-means clustering to produce anchors that matched the original dataset used.

According to a recent study (J. Deng et al., 2020), breast density images were automatically classified using CNN with SE-Attention mechanism. With the use of 18157 images from beast mammography, a new dataset was produced. By identifying the patients who require more attention than others, our approach can give radiologists a trustworthy breast volume diagnosis (J. Deng et al., 2020).

The authors concentrated on using diffuse division and texture analysis approach and fuzzy classifier to aid in the classification of breast density in mammography images (Valencia-Hernandez et al., 2021). The Breast Cancer Digital Repository (BCDR) and the In Breast datasets were used to collect the photos. The suggested method produces trustworthy findings when compared to LIBRA, a completely the computer team at the University of Pennsylvania created an automatic breast density prediction method (Valencia-Hernandez et al., 2021).

4. BI-RADS

The American College of Radiology's (ACR's) Breast Imaging Report and Data System, is used in the majority of countries throughout the world to lessen variation in radiologists' a report for a mammogram, MRIs, or ultrasound imaging (Bell & Weerakkody, 2013). According to table I, BI-RADS has seven evaluation groups, ranging from 0 to 6. The BI-RADS evaluation categories are supported by computer-aided diagnostics as well as digital mammography modalities (Guo et al., 2018) in addition to the conventional DICOM (Digital Imaging and Communications in Medicine). and digital mammography modalities (Guo et al., 2018).

When several findings are found, the BI-RADS categories are allocated in the following order: 1, 2, 3, 6, 0, 4, 5 (Bell & Weerakkody, 2013).

Category	Finding	Description
BI-RADS 0	Incomplete	Additional imaging testing is required at this stage, most likely through ultrasound, MRI, or mammography.
BI-RADS 1	Negative	Symmetrical with no worrisome calcifications, masses, or architectural deformity (Bell & Weerakkody, 2013)
BI-RADS 2	Benign	The likelihood of cancer is 0 percent.
BI-RADS 3	Probably benign	Less than 2% of individuals have cancer.
BI-RADS 4	suspicious for malignancy	Malignancy is more likely than or equivalent to 2-94% likely. You can further categorize these for mammography and ultrasound as follows.
BI-RADS 5	highly suggestive of malignancy	Malignancy is more likely than or equivalent to 95% of the time.
BI-RADS 6	known biopsy-proven malignancy	A prior biopsy revealed a malignancy that had already appeared on a mammogram. In this case, it might be advised to use a mammogram to determine how well a breast cancer treatment is working.

5. SCREENING TEST QUALITY

This section should go over two crucial elements :: sensitivity and specificity of a test as well as its percentage of false positives and false negatives. Two criteria, sensitivity and specificity, are used in conventional testing to determine the quality of imaging modalities used to detect breast cancer. Sensitivity reveals anyone who actually has an illness, whereas specificity reveals anyone who does not (Susan G. Komen, 2020). Both measurements are computed on a scale of 0% to 100a%. False-positive/negative test results should also be accounted for the result. False-negative denotes that a person's test results for breast cancer were negative, despite the fact that they actually have the disease. False-positive test results indicate the presence of breast cancer, while in reality, It is not breast cancer in this person, according to Susan G. Komen (2020). Additional tests will be performed because no modality has 100% sensitivity and specificity.

Breast tissues that are denser result in a significant reduction in mammography's sensitivity; the denser the breast, the less sensitive mammography is (Guo et al., 2018).

In older women, the sensitivity of mammography is greater than that of ultrasonography (US), at 85% and 95%, respectively, according to Guo et al. (2018). However, ultrasound has a 13.3% sensitivity that is higher than mammography in women who are less than or equivalent to 45 years old (Guo et al., 2018).

Recent research suggests that combining mammography and US tests can help in breast cancer early identification and treatment (Guo et al., 2018). The diagnostic precision of tests is increased when mammography and US are combined rather than when mammography alone is used. (Guo et al., 2018)

6. Standard for DICOM

Medical Imaging Technology Association, 2020). DICOM, is a standard for the administration and distribution of medical imaging data in the healthcare sector. To achieve consistency among the many medical imaging modalities, this methodology was created.

Today, all imaging techniques generate images that are saved in the DICOM format and need to be accessed, retrieved, and viewed using specialist software called a medical DICOM viewer. The bady parts and the header are the two components of a DICOM file (Al-Tam, 2015). Some information pertaining to the body part's stored image is contained in the header. Some information pertaining to the body part's stored image is contained in the header.

This information can include patient and physician information, image type (such as JPEG or TIFF), and modality (such as US or MRI). The body houses each stored image pixel. Modalities typically produce DICOM files, which are then exported to a server so that anybody, anywhere, can view them at any time. PACS server is the name of the server, which is a piece of software in charge of gathering DICOM data from various modalities for subsequent use by anyone. To collect the generated medical images for later retrieval, viewing, or annotation by radiologists, many institutions often have their own PACS server (Brühschwein et al., 2020). There are currently a large number of PASC servers and DICOM viewers; some of them are free to use, while others are not and cost a significant amount of money to install and maintain.

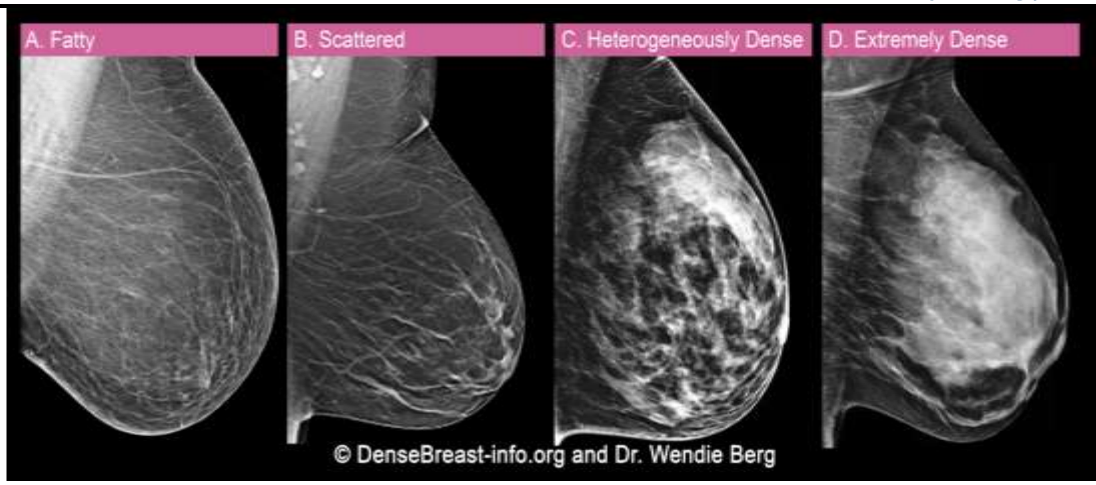
7 DATASETS

A collection of DICOM files has been combined with a few standard datasets, some of which feature radiologists' annotations. Breast cancer comes in two different forms of data sets: extracted includes without medical records and datasets with medical information, a few medical datasets. In addition to the Mammographic Image Analysis Society (Mini-MIAS), there are photographs such as INbreast, Mini and CBIS-DDSM. Wisconsin Breast is a well-known extracted features dataset. EER cancer datasets and the World Bank Cancer Dataset. Curated The updated Version CBIS-DDSM for Screening of the Digital Database, Breast Imaging Subset that is uniform. Breast imaging using mammography (DDSM) and scanned 2620 film (TCIA, 2020). This collection also includes numerous benign, malignant, and normal cases with corroborating pathological information. Only situations with cancerous medical images are collected by CBIS-DDSM, and by converting the photographs to DICOM files, the image quality is increased (TCIA, 2020). The Wisconsin Breast Cancer Dataset, on the other hand, is a different dataset that includes features that were derived from 699 photos that were received from the hospitals at the University of Wisconsin (Frank & Asuncion, 2010). About 2500 mammography images can be found in the Digital Database for Screening Mammography (DDSM), while the MIAS Mini-Mammographic (VCL, 2020) Database contains 322 digitized movies. There are 410 mammography images in the INbreast database, of which 360 were utilized to create 90 cases for women with two breasts (4 photos per case), and the remaining twenty five instances were for patients who had undergone mastectomys (two photos per instance) (Moreira et al., 2012). SEER gathers tens of thousands of features that have been derived from breast cancer and other cancer pictures (National Cancer Institute, 2020). Additionally, the Classification of Histopathological Images for Breast Cancer (BreakHis) (Spanhol et al., 2015) contains 9109 images of the breast in microscopy lesion tissue that 82 patients provided the data and magnified by distinct elements (40X, 100X, 200X, and 400X). Spanhol et al. (2015) assembled this dataset, which mostly consists of 5429 cancerous samples and 2480 benign samples. All image meets the requirements listed below: PNG file format, 700x460 pixels, 3-channel RGB, 8 bits per channel. On the other hand, The Stanford Tissue Microarray Database (Marinelli et al., 2007) contains annotated tissue pictures and related expression information. Following a survey, a number of datasets utilized by researchers were identified. Tens of thousands of images combining MR, MRI, SEG, and CT are present in the datasets Breast-MRI-NACT-Pilot, BREAST-DIAGNOSIS, ISPY1, BREAST-DIAGNOSIS, TCGABRCA, and QIN Breast DCE-MRI (Nahid & Kong, 2017; TCIA, 2020).

The DDSM and MIAS datasets have been used the most effective way to find and diagnose breast cancer, per papers posted on the Springer (http://www.springer.com), Elsevier (https://www.elsevier.com), and IEEE (http://www.ieeexplore.ieee.org) websites (Nahid & Kong, 2017).

8. DENSE BREAST

Having ducts, lobules, fatty, and fibrous tissue as its basic elements, a dense breast is one that contains more glandular or fibrous structures than fats. (society2020). Although thick breasts are typical, most women's breasts get less dense as they age (Society, 2020). It is more challenging to disclose particular breast regions on a mammography to determine what is normal and what is problematic tissue the more dense the breast is. Figure 4.a illustrates a fully fatty breast. Figure 4.b shows scattered areas of fibrous and glandular tissue. Figure 4.d depicts Differentially dense (dense glandular and fibrous tissue). Figure 4.e illustrates highly extremely . Small lesions in the breast are difficult to see due to heterogeneously thick tissues. Although reading tumors on mammograms is poorer with highly thick breast than with heterogeneously dense breast.



9. THICK BREAST

Radiologists evaluate breast density using the BIRADS A or 1, BIRADS B or 2, BIRADS C or 3, and BIRADS D or 4 scales, as per the American College of Radiology (2019). BI-RADS A indicates a breast that is fatty, BI-RADS B indicates a breast that is largely made up having fatty tissues sporadic patches of dense tissue. Furthermore, according to BI-RADS C, the breast is made up of both dense and fatty tissues. Last but not least, BIRADS D shows that the breast is primarily made up of sturdy tissues.

10. AREA OF RESEARCH

This article aims to evaluate a number of studies regarding breast cancer screening and diagnosis using computer-aided design (CAD) systems that analyze medical images obtained from diverse modalities using machine learning techniques. It is important to consider the following issues to make this survey as helpful as possible:

- What imaging techniques have been applied to the early identification and treatment of breast cancer?
- What CAD applications are being used the most in hospitals?
- What modern CAD systems use machine learning techniques?
- How can one tell whether a breast image is malignant or benign, What standards are used to identify breast cancer?
- What data sources do CAD systems or researchers use the most frequently to find and diagnose breast cancer?
- What criteria are to evaluate the CAD system's effectiveness?

Several digital databases were also investigated. To start, a search was done from January 2017 to August 2020 to see whether there were any surveys or reviews that were similar to this one. With the help of the table II search words criterion, 3,505 studies were located. The words "survey" or "review" were not typically used in the titles of these research. As a result, similar publications with titles including "survey" or "review" were found using the https://academic.microsoft.com/ search engine. In addition, a second search criteria were conducted in the other databases listed in table II with the same restrictions as the first search criteria in order to compile an exhaustive list of related publications free of duplication. Only 12 of the 45 studies are thought to be relevant to this survey. In the end, papers by Bharati et al. (2020), Bagchi et al. (2020), Das et al. (2020), Kajala & Jain (2020), Nahid & Kong (2017, 2020), Gardezi et al. (2019), Saxena & Gyanchandani (2020), Raghavendra et al. (2019), Yassin et al. (2018, 2020), Zhou et al. The primary objective is to provide coverage for Without duplicating anything, apply each machine learning method that has been applied in research articles. As a result, any repeated approaches that were cited numerous times in the publications that were collected were only stated once in this survey. The last step is to search exactly the identical databases list and lookup phrases listed below in omitted table II the phrases "survey" or "review" to locate any new articles published between 2017 and 2020 that made use of the obtained machine learning techniques. The search parameters are based on a standard that consists of a list of terms used in the research engine of each database to collect as many machine learning approaches as is practical for this survey.

TABLE II

Database	Search in	Search terms
IEEE	**http://www.ieeexplore.ieee.org**	(("Document Title":"breast cancer") AND "All Metadata":"machine learning"), 312 results were founded.
Science Direct	**http://www.sciencedirect.com**	1. "breast cancer" "machine learning" "review", 2072 results were founded. 2. "breast cancer" "machine learning" "survey", 690 results were founded. 1. Year: 2017-2020 Title: "breast cancer" "machine learning", in this criterion, the final number of founded papers were reduced to only 31 papers compared to the previous criteria.
ACM	**https://dl.acm.org**	[Publication Title: "breast cancer"] AND [Publication Title: "machine learning"] AND [Publication Date: (01/01/2017 TO 12/31/2020)], 10 results were founded
Microsoft Academic	**https://academic.microsoft.com/**	1. "breast cancer" "machine learning" "review", 26 results were founded. 2. "breast cancer" "machine learning" "survey", 19 results were founded.
Springer	**http://www.springerlink.com**	1. '"machine learning" AND ("survey")' within 2017-2020, 66 results were founded. Where the title contains= "breast cancer", with the exact phrase = "machine learning", start year="2017" and end year="2020", with at least one of the words= "survey". 2. "machine learning" AND ("review") within 2017-2020, 195 results were founded.
Pubmed	**http://www.ncbi.nlm.nih.gov/pubmed/**	(("breast cancer"[Title]) AND ("machine learning"[Title])), 84 results were founded, in period between 2017-

Nevertheless, it's possible that some pertinent studies weren't reviewed inadvertently. The majority of pertinent studies were reviewed, but only those that met the following requirements are included:

- The works must have been published between January 2017 and August 2020.
- A minimum of one medical technique was applied.
- Nearly all forms include cases of breast cancer that are found and diagnosed without a biopsy. As oppsed, several articles were utilizing different examinations, such as medical photographs from histopathology, which are discussed without going into great depth (Saxena & Gyanchandani, 2020). Images are made utilizing the close study of a biopsy or surgical specimen under a microscope.
- It is necessary to use one or more methods for machine learning.
- We remove papers that do not include performance indicators like sensitivity and specificity.
- The table II journal list excludes any papers that are not on it.

11. LEARNING FROM MEDICINE MACHINES

Machine learning (ML) is a subcategory of artificial intelligence (AI) that uses algorithms to identify patterns and make predictions within a set of data. This data can consist of numbers, text, or even photos. Under ideal conditions, machine learning allows humans to interpret data more quickly and more accurately than we would ever be able to on our own.

Artificial intelligence happens when humans synthetically create a sense of human-like intelligence within a machine. For machine learning, this means programming machines to mimic specific cognitive functions that humans naturally possess, such as perception, learning, and problem-solving.

Machine learning and artificial intelligence can be used to enhance user experience, anticipate customer behavior, monitor systems to detect fraud, and can even help healthcare providers detect life-threatening conditions. Many of us benefit from and interact with machine learning on a daily basis. Some common examples include:

- Recommendation algorithms on your favorite streaming services.
- Automatic helplines and chatbots.
- Targeted ads.
- Automated quotes from financial institutions.

Machine learning has been extensively used to identify and treat human diseases, particularly liver, brain, lung, and breast cancers (C. Deng et al., 2020). It has also been used to predict the protein structure, recognize speech, and detect objects. In the last few decades, it has become a potent tool to help decision making due to its capacity to learn from a vast amount of data. For the time being, a doctor can employ a variety of machine learning approaches to identify and diagnose breast cancer. These methods can predict the progression of cancer and the likelihood of a cancer recurrence following therapy in addition to aiding regarding breast cancer early detection. Additionally, the likelihood of death or the survival rate might be discussed at a particular period. Additionally, applying machine learning can greatly reduce human error and improve the robustness and reliability of results, both of which contribute to the development of robust systems (Kajala & Jain, 2020). Machine learning techniques are still far from acting intelligently like people do, even with all of these skills.

12. ALGORITHMS CREATED BASED ON FEATURES

For machine techniques approaches to determine whether an image is malignant or not, prior actions on the input images are required. The pre-processing stage of recent machine learning-based system advancements includes the use of several approaches like image noise reduction, cropping, and quality improvement (Bagchi et al., 2020; Kajala & Jain, 2020). Because of this, feature extraction, segmentation, pre-processing and and classification and selection are some of the processes that must be finished for CAD systems, as shown in figure 5.

There are various techniques that can be used for each level. Image enhancement, Cropping, and de-noising, are preprocessing techniques, whereas thresholding, region-based, boundary-based, and template matching are segmentation techniques. Additionally, the book Feature extraction and selection includes methods for removing color, form, and texture from pictures. Last but not least, the classification divides benign, images into cancer, or normal using machine techniques approaches for instance, supervised and unsupervised techniques.

A. Stage of preparation

In order to remove the unwanted labels and increase the contrast level of medical pictures with noise, such as those created by mammography, ultrasound, and MRI scans, specific treatments must be applied (Kajala & Jain, 2020). At this stage, a variety of techniques were used, including multiplicative, additive, impulse, multiplicative, uniform, shot and periodic noise. Additionally, many filters, including Median, Mean, Gaussian and Wiener filters, can be used (Bagchi et al., 2017).

Since the preliminary stage is crucial it can be difficult to detect masses in medical pictures because their characteristics can be difficult to see and occasionally resemble a breast's natural cells (Kajala & Jain, 2020). Additionally, the breast's microscopic calcium deposits are more contrasted using methods like dyadic wavelet processing, this region can be identified because it is different from the surrounding areas (Kajala & Jain, 2020). While masses feature spiculated architecture, varied densities, and poor contrast in contrast to microcalcifications. It is possible to boost the resolution using median filtering and Contrast Limited Adaptive Histogram Equalization (CLAHE). 94.4% for specificity and 96.2% for sensitivity of mass detection (Kajala & Jain, 2020).

B. Stage of segmentation

In order to decide on a medical imaging region of interest (ROI), this stage is in charge of segmenting the images. To achieve the necessary ROI, common techniques including boundary-based segmentation, thresholding, template matching segmentation and region-based segmentation may be used.

1) Segmentation based on boundaries

In general, edges and discontinuities in an image's intensity are vital as they provide information about an object's border (Kajala & Jain, 2020).. As a result, based on these discontinuities, detection methods can be used in order to illustrate image segmentation and object recognition. In order to segment breast ROI pictures, numerous boundary-based segmentation techniques have been employed. These techniques can spot gaps or sharp modify in a grayscale picture; sadly, the method for locating the edge is application-dependent; there is no standard method. One of the fundamental methods for detecting edges is the use of gradient filters and high pass filters (Bagchi et al., 2020). Additionally, edges can be found using the derivatives of the first and second orders, however the first order often misses them, especially in noisy images. The second-order derivative, on the other hand, is more noise-resistant than the first-order counterpart, making it more resilient. Along with these techniques, some researchers used the Butterworth high-pass filter, the Speculation Filter, in conjunction with Non-linear Polynomial Filtering, Sobel edge detection, Color Gradient-based Geodesic Active Contour, Difference of Gaussian (DoG) filter and snake-based algorithm, to find edges in medical images (Bagchi et al., 2020; Kajala & Jain, 2020).

2) Threshold

Gray Level Thresholding, some thresholding strategies include the Minimum Error Method, the Maximum Entropy Method and the Otsu's Method that can be used (Kajala & Jain, 2020). On the basis of the grey level histogram, thresholding processes are frequently used to take away the portions of photos those don't have any crucial information in them. Then, segments were created using the disparity in pixel intensities between the backdrop and the necessary picture. The fundamental disadvantage is that the threshold disregards the spatial information contained in the images; as a result, the contiguousness of the segmented portions is disregarded. Additionally, other techniques can be used in conjunction with thresholding to improve the result, showing in (Bagchi et al., 2020) This modifies the threshold approach in order to avoid over-segmentation. Furthermore, research utilizing thresholding mammography image segmentation produced a sensitivity of 80% and a false-positive rate of 0.32% for each image., according to the scientist of (Bagchi et al., 2020). Additionally, some scientists used the edge detection component of the three classes threshold segmentation strategy for the selected photos (Bagchi et al., 2020).

The authors of (Torres et al., 2019) proposed a fully automated thresholding method employing the mammogram's morphological analysis, which makes use of The suggested algorithm can calculate breast density. According to some researchers, radiologists are capable of manually segmenting dense tissues, which improves classifier performance arts more quickly compared to when segmentation isn't used (Torres et al., 2019); regrettably, this process takes a lot of time and requires instruction from the expert.

3) Template matching

The main disadvantage of this approach is that it can only be applied when a lesion has been previously identified. Some researchers have utilized the Sech template, which they used in conjunction utilizing thresholding to deal with the unclear areas in medical images (Kajala & Jain, 2020). While this was going on, Other researchers combined dynamic programming, local cost function, and template matching. to detect tumor locations more efficiently (Kajala & Jain, 2020).

4) Segmentation depending on region

Another technique for segmenting images is region-based segmentation, commonly known as region-growing approaches. If the requirements for class similarity are met, satisfied, the surrounding pixels are checked based on criteria for resemblance or smoothing, and added to the region class. All neighboring pixels in the area undergo a repeat of this procedure. As a result, this algorithm can recognize regional characteristics like texture, color, and grayscale that are comparable. Along with this technique, other operations like merge, spilt and uniform blocking can be used to get decent results. This technique was utilized in research publish in (Bagchi et al., 2020) Mean Based Region Growing Segmentation (MRGS), which was also utilized in another study stated in Bagchi et al. (2020), was utilized to divide up mammography pictures. Additionally, several researchers used an Artificial Neural Network with training (ANN) threshold with an automated region expanding segmentation.

5) Template matching

The main disadvantage of this approach is that it can only be applied when a lesion has been previously identified. Some researchers have utilized the Sech template, which they used in conjunction with thresholding to address the ambiguous regions in medical pictures (Kajala & Jain, 2020). While this was going on, Other researchers combined dynamic programming, local cost function, and template matching. to detect tumor locations more efficiently (Kajala & Jain, 2020).

C. Stage of segmentation Feature Selection and extraction

Contour-based and region-based approaches may be used to extract and choose an image feature, shape, such as color or texture. According to the entire region, the regional approach provides shape features, whereas the contour-based method bases them on boundary information (Kajala & Jain, 2020). As a type of texture feature, several features are reported, including structural or geometric, model-based, statistical and transform-based features. According to Bagchi et al. (2020), Typically, sets of features can be collected with the goal of finding the perfect set for a subset that will allow them to grade cancer. Additionally, The extracted feature may frequently be unnecessary or redundant, hence a feature selection stage is carried out to determine which features are most crucial. The scientists derived ROI segmentation using the mean and standard deviation Evaluation and outlook for cancer (Veta et al., 2012). though other scientists calculated area with the lowest intensity value, and standard deviation, small and large axes, in addition to mean and to be provided the minimum intensity values for each separated region applying a binary decision tree with training, into clustering (Petushi et al., 2004).

A point and edge patterns, as well as their location within the hierarchy, are what determine the structural characteristics. Although geographic distribution of the picture pixels' brightness levels is the subject of statistical properties including variance, mean, skewness,standard deviation, mean, kurtosis, and entropy. They can also be of first or second order. In contrast the Gray Level Co-occurrence Matrix (GLCM), which is a second-order statistical tool, a given pixel's linked intensity is disclosed in the first statistical order, show the relationship between a specific pair of texture-related measures and contrast, energy, correlation, and homogeneity (X. Liu & Tang, 2013). Others have employed Local Binary Pattern (LBP), a technique that combines structural, statistical, and textural analysis techniques and can reveal an explanation of a pixel's connection to its neighbor by using a binary design (Kajala & Jain, 2020). Often, the initial order is a sample and requires little computational effort. but the second place produces superior results despite the fact that the computational cost increases exponentially with increasing statistical order (Bagchi et al., 2020).

The authors of (Belsare et al., 2015) suggested a classification method based on linear discriminant analysis. The suggested technique was 100% accurate in classifying 70 histopathology pictures of the breast. For the breast nuclear grading tumors in histopathology pictures, the region covariance descriptors geodesic mean served as the foundation for a textural-based feature approach that was created (A. M. Khan et al., 2015). An algorithm for automatically detecting breast cancer using textural features was proposed in (Ojansivu et al., 2013). Additionally, a fusion approach was used used in (Gardezi & Faye, 2015), merging finished curvelet sub-band features and LBP (CLBP), with an accuracy of 96.68%. was attained as well as a decrease in the frequency of false-positive results. (Gandomkar et al., 2019) implemented an approach that combines segmentation and texturing that can identify traits from histopathology images for grading cancer. In addition, two novel extracted feature methods— Run Difference Method (RDM) and Square Centroid Lines Gray Level Distribution Method (SCLGM) —were introduced in (Ganesan et al., 2014).

High dimensional features and evolutionary algorithms have been combined in recent years to reduce redundancy and improve accuracy (Yeh & Chan, 2017). In addition, a study was given (Gastounioti et al., 2018) for obtaining characteristics employing a lattice-based scaffold of breast parenchyma method to assess heterogeneity of breast tissues. Additionally, The Convolutional Neural Network (CNN) was utilized to cut down on the features that were used.

D. Classifications

In respect to breast cancer images, the term "classification" generally means refers to grouping them into three categories: malignant, benign and normal. The categorization model is executed in two stages: testing and training. The classifier is first given a dataset having labels as an input, which the input dataset gives the classifier the opportunity to train and learn. The classifier can then be utilized for a testing set of samples from unidentified classes after it has trained on the current input data set. A testing set of samples can then be used with the classifier from unidentified classes after it using the current input data set to train, Decision trees, Naive Bayes, artificial neural networks (ANN), support vector machines (SVMs), knearest neighbor (KNN), Gaussian Mixture models, SVM combined with Bayes classifier, convolutional neural networks (CNN) and random forests are currently the most well-liked ML methods that have been used in CAD systems extensively (Bagchi et al., 2020; Das et al., 2020; Kajala & Jain, 2020).

1) Decision tree

The label of the private information is displayed at a tree's leaf node in which the classified data are presented and an inside node shows a feature. Additionally, the categorization outcome is contained in the leaf node of the tree, which is traveled from leaf to root. Using Multi Verse Optimizer (MVO) and Gradient Boosting Decision Tree (GBDT), a new ensemble learning technique is implemented in (Tabrizchi et al., 2020), one of the numerous recent works to employ this method to categorize breast cancer. Wisconsin Diagnostic Breast Cancer and Wisconsin Breast Cancer were the datasets that are used. According to Tabrizi et al. (2020), The suggested strategy succeeds in 0.9876% accuracy, 0.9764% specificity, 0.9943% sensitivity 0.9764% specificity. Contrarily, the Gaussian Light Gradient Boost Decision Tree Classification (GLGBDTC), an amalgam method, which is primarily employed to improve the identification of breast cancer, is reported in (Ezhilraman et al., 2020). In this study, C4.5 the popular decision tree algorithms is decision trees. are used. The writers of (Bagchi et al., 2020) reported the release an update to C4.5 dubbed EC.4.5, whose performance was five times better than that of C4.5. Additionally, utilizing the Wisconsin Breast Cancer dataset, In order to predict breast cancer, k-means and decision trees were used (Marne et al., 2020).

Another study (Hamim et al., 2020) based on genes for breast cancer examined the combined with Fisher selection in C5.0 of features depending on score. The C5.0 decision tree had the highest accuracy (93.28%) of all the classification methods used, including fisher-score based feature selection, logistic regression, artificial neural networks and support vector machines. Among 97 patients' 24481 gene expressions a dataset with microarray breast carcinoma was used in all tests.

Researchers gave a comparison of decision making

To establish which is the most precise approach for classifying breast cancer, this study looked into tree, Nave Bayes, KNN, and SVM algorithms. demonstrates that decision trees perform better than the other options (G. Kumar & others, 2019).

2) Naive Bayes

Due to its adaptability, Nave Bayes is an easy-to-use and accurate classification technique. It has been put to use in a variety of applications. (Arar & Ayan, 2017). Additionally, several researchers breast cancer prediction algorithm was put into use (G.Shaikh & Ali, 2020; Kumar & others, 2019). the Lemons, In a research published in 2020, the authors compared the accuracy of comparing Nave Bayes and random methods for breast cancer diagnosis Forest. Unfortunately, random forest's effectiveness was more trustworthy in terms of accuracy than the Naive Bayes approach accuracy with a score of 97.82%

3) Artificial Neural Network (ANN)

An approach to artificial intelligence known as ANN aims to emulate the physiological neural networks seen in the human brain. Since ANN a group of pattern recognition algorithms that generally comprised artificial neurons, which are a collection of connected nodes, it may learn using certain input data (Bagchi et al., 2020). Additionally, there are other ANN architectures available, with the multilayer perceptron (MLP) architecture is always used popular. Numerous studies have recently examined this method's capacity to identify breast cancer.

In remote and underserved areas, a unique technique for the early identification of breast cancer was proposed. The database of a collection of mammographic images is available from the Mammographic Image Analysis Society (MIAS), was utilized. Preprocessing detection was successfully accomplished using a wavelet-based image processing technique. Different classification techniques are used, including K-Nearest Neighbor classifiers, J48 decision trees, random forests and multilayer perceptron neural networks. Random forest, K-NN, and decision tree classifiers all perform worse than the neural network classifier in terms of final results.

In order to assist the radiologist in the detection phase, a fuzzy c-mean algorithm (FCM) method was presented (Faisal & El Abbadi, 2020). Additionally, to extract relevant characteristics, discrete wavelet transformation (DWT) and principal component analysis (PCA) has been used. for the ANN classifier's data input. The MIAS database's mammography scans were the main source of the original pictures.

Five supervised machine learning techniques—SVM, ANNs, random forests, KNN and logistic regression—were compared to see which produced the best results (Marcia Boraas, M.D., 2019). The UCI machine learning repository provided the Wisconsin Breast Cancer dataset was the one used in this study (Frank & Asuncion, 2010). This study shows that utilizing ANNs resulted in the highest accuracy, precision, and F1 score, with values of 98.57%, 97.82%, and 0.9890%, respectively.

4) Support Vector Machines (SVMs)

Data can be divided into categories using the learning classifier method Support Vector Machine (SVM) (Bagchi et al., 2020). This technique has been widely used by researchers to categorize breast cancer.
The following supervised learning classification methods were compared: SVM, Navie Bayes, logistic regression, decision tree and KNN classifier (A. Kumar & Poonkodi, 2019). The comparison was based on the accuracy for detecting breast cancer and the confusion matrix. By reaching 98.24% accuracy, the suggested approach, the other classifiers offered performed better than Kernel SVM with PCA.

According to A. A. Khan and Arora (2018), traditional mammography is a painful treatment that exposes the body to dangerous Xrays. The thermograph technology was looked into as a potential improvement. A. A. Khan and Arora (2018) used a mastology research database called the DMR-Database, to extract some of the thermogram pictures. In order to separate the left and right breasts' textural characteristics, Gabor filters are used, and Support vector machine (SVM) is based on the asymmetry of the breasts' textures to identify breast tumors. The suggested approach achieved an accuracy of 84.5%; as a result, the starting stages of breast cancer using thermography.

A histopathology-based feature technique has been taken into consideration for the diagnosis and classification of breast cancer. (Singh & Kumar, 2020). Random Forest and the K-Nearest Neighbor (KNN)and around six varieties of SVM have been examined using as input data, the BreakHis dataset. The results demonstrate that the proposed cubic SVM classifier had a maximum accuracy of 92.3%.

When employing bi-modal ultrasonography to diagnose breast cancer, the scientist of (Gong et al., 2020) suggested a fresh multiple view machine with a deep neural network supporting it. (MDNNSVM).

The dataset GSE76275 contains 265 samples, was used to gather gene expression microarray data (Chen et al., 2020). Feature elimination using recursion (RFE) was removal (RFE) was employed to identify the ideal feature subset after testing the SVM and KNN algorithms. In comparison to other commonly used classifiers like SVM, KNN, SVM-PCA, KNN-PCA, SVM-RFE-SVM, KNN-PCA and SVM-RFE-KNN, the recommended SVM-RFE-SVM technique performed well.

Using the Wisconsin breast cancer dataset, a study was published in (Vrigazova, 2020) to suggest a changes of the SVMs it can achieve a great precision in malignancy detection tumors of 99.6%. Furthermore, in comparison to some of the methods indicated in (Vrigazova, 2020), a low error rate was reported.

The researcher of (Hamouda et al., 2020) presented a study that used blood analysis data from the Coimbra Dataset to predict the development of breast cancer. Support vector machines (SVM) were employed as the classifiers, and the SVM classifier was optimized using the Grid Search Algorithm.

Another study used superior accuracy in the Gabor algorithm, which uses the standard deviationand mean to extract characteristics to detect cancerous followed by the data are pre-processed using a adaptive histogram equalization and a median filter. Finally, a selection of mammography pictures from the Mini-MIAS dataset were classified using the SVM classifier. (Thakare et al., n.d.).

In (Venkata & Lingamgunta, 2020), the classification of breast cancer utilizing three picture changes—ultrasound, magnetic resonance imaging and mammography —was explored. The suggested work used SVM to classify the shared phenotypic characteristics of the medical images generated by the three different modes. In comparison to mammography and ultrasound images, the final results demonstrate that MRI image lesions can be more precisely identified by the SVM classifier.

To obtain the most crucial features, the Boruta feature selection method was used (Aroef et al., 2020). Additionally, with accuracy rates of 90% and 95%, the random forest and the SVM machine learning models were applied.

5) K-nearest neighbor (KNN)

According to Bagchi et al. (2020), supervised categorization with KNN technique that generates Depending on the k value or the close-by data points, new data points may be added. This technique has been widely employed by researchers to identify and diagnose breast cancer in recent years. For instance, utilizing mammography pictures from the CBIS-DDSM dataset, k-nearest neighbor (KNN) and support vector machine (SVM) classifiers were taken into consideration in (Mohan & others, 2020). After removing noise and undesired artifacts with a using a local binary pattern (LBP) and a gray level co-occurrence matrix (GLCM), features use a 2D median filter to be retrieved. The end findings demonstrate that KNN surpasses SVM in terms of final precision, with KNN achieving 100% compared to SVM's 96%.

On a collection of mammography picture taken from the Wisconsin Breast Cancer Database, a study was conducted to compare the performance of closest neighbor (KNN) and artificial neural networks (ANNs) (O'Shea, 2020). ANN and KNN both had 95.24% and 100% accuracy, respectively.

On the breast cancer microarray dataset, the subspace KNN method and for the initial disease diagnosis, stacked autoencoders (SAE) were used. (Adem, 2020). Then, these hybrid techniques were used on a collection of pictures acquired from the Kent Ridge-2 database.

These combined techniques can handle datasets with large dimensions and uncertainty, and accuracy of 91.24% was attained. (Pharswan & Singh, 2020) presented a survey to look into the top classifiers between SVM and KNN. The crucial features were extracted using GLCM. This study demonstrates that accuracy-wise, SVM outperformed KNN achieving a value of 94%; in addition, SVM outperformed KNN in terms of recall and F1 score.

6) Gaussian Mixture models

It is a classifier that uses training data to group several related points and predict an unknown point. This classifier was employed by the authors of (Ezhilraman et al., 2020) to detect breast cancer. In a different study, a classifier based to forecast the molecular properties, the creation of the Gaussian mixture model (GMM) of tumors using data on mRNA expression (Prabakaran et al., 2019).

7) Convolutional Neural Networks (CNN)

For identifying intricate aspects in data, CNN is a neural network with many layers. (Bagchi et al., 2020). This classifier has recently been widely used by researchers to identify and rounded breast. For instance, the researchers of (H.-C. Lu et al., 2019) employed CNN and using 9000 scans and deep learning algorithms to find and categorize breast cancer obtained from a Taiwanese teaching hospital. Data augmentation, tcontrast-limited adaptive histogram equalization, the median filter, and preprocessing techniques are used. The last results demonstrate that the pre-processing picture technique outperforms the unprocessed image model in terms of accuracy. The remaining 30% of the dataset serves as testing data, while the remaining as validation data, 10% of the dataset is chosen. With preprocessed images, the suggested model's overall specificity, sensitivity, and F1 score were 0.57, 0.91, and 0.88 respectively, compared to 0, 0.79, 0.88 when using the suggested model without reprocessing the images.

In a different work, (Z. Wang et al., 2019), the researchers proposed using US-ELM, and CNN and for component separation and grouping, unsupervised extreme learning machine. Second, an 8-layer CNN architecture was used to create a fusion deep feature set that collected twenty in-depth features and integrated them with the tumour's additional five form, seven density and five texture features. In order to determine if a benign or malignant breast tumor is present, an extreme learning machine (ELM) classifier was employed as an input to the newly formed each mammogram's fusion deep feature set. In this study, 400 mammograms were used as mammography pictures.

Wahab et al. (2019) segmented data using a pre-trained CNN, and a hybrid-CNN for classifying histopathology pictures as cancerous or not. Additionally, the dataset for this study was produced by the an organization that analyzes medical images Edinburgh (2020).

8) Random Forest

A method of group learning for categorization is used, and it is called random forest. (Y. Lu et al., 2018) evaluate some recently proposed methods using computer vision and machine learning for diagnosing breast cancer. On mammography and histology images, the detection performance of various approaches is compared and examined. Three types of imaging are x-ray (mammography), histological imaging, and ultrasound imaging, modalities that are utilized to improve the diagnosis of breast cancer. Each type of modality already has its own set of datasets. This study demonstrates that mammography or histology imaging can effectively deep learning-based techniques are used to identify breast cancer.. On histology pictures, the ScanNet approach provides the highest accuracy.

obtained using a random forest classifier and Adaboost on mammography pictures. Additionally, employing Adaboost and a mammography picture classifier using random forests produced an outstanding result.

For the purpose of predicting breast cancer, Extreme Gradient Boosting (XGBoost) and Random Forest were used (Kabiraj et al., 2020). This paper's primary objective was to categorize recurrence. accurately predict no-recurrence events, and it obtained 74.73% and accuracy of 73.63%. Machine Learning at UCI Repository served as the source of the data (Frank and Asuncion, 2010), While trimming the mode and mean was used to prepare the data.

9) Logistic Regression (LR)

In machine learning, the most famous classification methods is LR. This technique has been utilized by some academics to extract the key features, while others use it to classify data.

The scientist of (Khandezamin et al., 2020) suggested an approach that involved two steps: first, logistic regression was used to extract key features, and then, a neural network using the Group Approach Data Handling (GMDH) technique was used to distinguish between cases that were malignant and those that were benign. In this work, we explore three datasets: Wisconsin Diagnostic Breast Cancer (WDBC), Wisconsin Breast Cancer Database (WBCD) and Wisconsin Prognostic Breast Cancer (WPBC). The accuracy attained utilizing the , WDBC, and WBCD WPBC datasets was 99.4%, 99.6%,and 96.9%, respectively, according to the proposed method.

(Zhou et al., 2020) proposed a fresh model based on logistic regression that was utilized to base on microarray expression data, classify breast cancer tumor samples without shrinking the microarray data matrix. The National Center for Biotechnology Information (NCBI) provided three datasets for download: GSE20711 90, GSE65194 178 and GSE25055 310 samples. The proposed method's lowest performance level was 80%.

By using Logistic Regression (LR), K-Nearest Neighbors (KNN), and other techniques, a comparative analysis in order to detect breast cancer was carried out in (MurtiRawat et al., 2020). Using principal component analysis in a group setting (PCA). Breast cancer in Wisconsin had been recognized. In The Ensemble Learning classifier's accuracy surpassed the competition, scoring 99.30%, compared to 98.60% K-Nearest Neighbors was used to produce this result, and 97.90% of utilizing Logistic Regression, accuracy was obtained.

13. ALGORITHMS LEARNED BASED ON FEATURES

As previously said, an individually engineered handcrafted feature is one that a data scientist creates. To put it another way, a group of features, including histograms, corner detection, and edge detection, are defined before they are extracted. Instead, these options may operate automatically collected by teaching a machine technique method to recognize and glean the necessary details. CNN is famous to used Machine learning techniques for obtaining important qualities (Arajo et al., 2017).

14. DIMENSIONS REDUCTION

In essence, to extract features procedure creates a more-dimensional feature vector, where more of the recovered Features might serve little purpose or be superfluous. The classification of these features will take many times and have a higher computational expense the more extracted features there are (Bagchi et al., 2020). Additionally, large feature vectors present still another problem., particularly when there is a lack of training data. This overfitting of the training dataset can prevent the ML model from correctly identifying invisible images. Dimension reduction is thus a crucial procedure to improve a system's performance when categorizing a collection of photos having a significant dimensional component.

The scientist of (Bagchi et al., 2020) also indicated two methods for reducing the extent: choosing the crucial aspects or adding new dimensions. First to pick important qualities, several researchers have used heuristic strategies like sequential selection both in the forward and backward directions. Additionally, various process techniques such the genetic algorithm, boosting, simulated annealing grafting, and particle swarm optimization have been employed. Second, various methods, including linear discriminant analysis, independent component analysis, principal component analysis and manifold learning, have been utilized to build new dimensions (Bagchi et al., 2020).

In general, the performance is improved by dimensions of reduction, but excessive reduction might result in the loss of some important features; as a result, some employ dropout (Hinton et al., 2012) or regularization (Ng, 2004) approaches to address the deep learning excessive fitting problem. PCA is also utilized in order to reduce dimensions, according to Siregar et al. (2020). In addition, another investigation (Obaid et al., 2019) reduced dimensions using PCA and Linear Discriminant Analysis (LDA).

CONCLUSION

One of the most common diseases in the world that kills women is breast cancer. The complexity of this disease's anatomy makes it more difficult for researchers to discover and diagnose it. To aid in the analysis and classification of breast cancer, numerous pieces of technology and software have been developed. The goal of this study is to undertake a thorough analysis of the most recent machine learning-based methods for diagnosing and detecting breast cancer. There is a list of medical modalities with their advantages and disadvantages, such as mammography, US and MRI. Additionally, a number of both commercial and non-commercial CAD systems, along with advantages and disadvantages, have been compiled.

The steps of the CAD system used to examine and categorize breast cancer are also specifically discussed segmentation, selection and selection, feature extraction classification, and preprocessing are the five-machine learning-based steps of the CAD system. Each stage's most popular techniques are shown. Additionally, a number of datasets for breast cancer are offered, collecting either extracted features or medical images.

In general, this survey comprehensively contrasts current machine learning methods for medical pictures. It demonstrates how improvements in machine learning techniques produce encouraging outcomes that can help radiologists or doctors with detecting breast cancer early and diagnosing it.

This study shows that while more methods have been developed to help in breast cancer detection and diagnosis, no single, ideal method is capable of doing so. Additionally, a comprehensive system that can handle several modalities and provide 100% accuracy is still difficult to develop due to the unique nature of breast cancer treatments that have been used in various ways. Due to these restrictions, additional research must be done in order to stay up with the growing risk and save people from this fatal illness.

REFERENCES

(ACS), A. C. S. (2020). Breast Density and Mammogram Reports | Dense Breast Tissue. Retrieved July 06, 2020, from https://www.cancer.org/cancer/breastcancer/screening-tests-and-earlydetection/mammograms/breast-density-and-yourmammogram-report.html

Adem, K. (2020). Diagnosis of breast cancer with Stacked autoencoder and Subspace kNN. Physica A: Statistical Mechanics and Its Applications, 551, 124591.

Al-Antari, M. A., Han, S.-M., & Kim, T.-S. (2020). Evaluation of deep learning detection and classification towards computer-aided diagnosis of breast lesions in digital X-ray mammograms. Computer Methods and Programs in Biomedicine, 196, 105584.

Al-Tam, R. M. (2015). Diversifying medical imaging of breast lesions (Doctoral dissertation) [University of Algarve,sapientia.ualg.pt]. http://hdl.handle.net/10400.1/8120

Aly, G. H., Marey, M., El-Sayed, S. A., & Tolba, M. F. (2021). YOLO Based Breast Masses Detection and Classification in Full-Field Digital Mammograms. Computer Methods and Programs in Biomedicine, 200, 105823.

American Cancer Society. (2020). Breast Cancer Risk Factors and Prevention Methods. Retrieved July 09, 2020, from https://www.cancer.org/cancer/breast-cancer/risk-andprevention.html

American College of Radiology. (2019). ACR Atlas, BI-RADS. Retrieved July 06, 2020, from https://www.acr.org/- /media/ACR/Files/%0ARADS/BI-RADS/MammographyReporting.pdf

Arar, Ö. F., & Ayan, K. (2017). A feature dependent Naive Bayes approach and its application to the software defect prediction problem. Applied Soft Computing, 59, 197–209.

Araújo, T., Aresta, G., Castro, E., Rouco, J., Aguiar, P., Eloy, C., Polónia, A., & Campilho, A. (2017). Classification of breast cancer histology images using convolutional neural networks. PloS One, 12(6), e0177544.

Aroef, C., Rivan, Y., & Rustam, Z. (2020). Comparing random forest and support vector machines for breast cancer classification. Telkomnika, 18(2), 815–821.

Bagchi, S., Huong, A., & Tay, K. G. (2017). Investigation of different spatial filters performance toward mammogram de-noising. Int. J. Integr. Eng, 9(3), 49–53.

Bagchi, S., Tay, K. G., Huong, A., & Debnath, S. K. (2020). Image processing and machine learning techniques used in computer-aided detection system for mammogram screening-A review. International Journal of Electrical and Computer Engineering, 10(3), 2336.

Bakker, M. F., de Lange, S. V, Pijnappel, R. M., Mann, R. M., Peeters, P. H. M., Monninkhof, E. M., Emaus, M. J., Loo, C. E., Bisschops, R. H. C., Lobbes, M. B. I., & others. (2019). Supplemental MRI screening for women with extremely dense breast tissue. New England Journal of Medicine, 381(22), 2091–2102.

Bell, D. J., & Weerakkody, Y. (2013). Breast imagingreporting and data system (BI-RADS) | Radiology Reference Article | Radiopaedia.org. RADIOPAEDIA. Retrieved July 07, 2020, from https://radiopaedia.org/articles/breast- Journal of Scientific Research, Volume 65, Issue 5, 2021 282 Institute of Science, BHU Varanasi, India imaging-reporting-and-data-system-bi-rads

Belsare, A. D., Mushrif, M. M., Pangarkar, M. A., & Meshram, N. (2015). Classification of breast cancer histopathology images using texture feature analysis. Tencon 2015-2015 IEEE Region 10 Conference, 1–5.

Bharati, S., Podder, P., & Mondal, M. (2020). Artificial neural network based breast cancer screening: a comprehensive review. ArXiv Preprint ArXiv:2006.01767.

Brentnall, A. R., Cohn, W. F., Knaus, W. A., Yaffe, M. J., Cuzick, J., & Harvey, J. A. (2019). A case-control study to add volumetric or clinical mammographic density into the Tyrer-Cuzick breast cancer risk model. Journal of Breast Imaging, 1(2), 99–106.

Brühschwein, A., Klever, J., Hoffmann, A. S., Huber, D., Kaufmann, E., Reese, S., & Meyer-Lindenberg, A. (2020). Free DICOM-Viewers for Veterinary Medicine. Journal of Digital Imaging, 33(1), 54–63.

Calisto, F. M., Nunes, N., & Nascimento, J. C. (2020). Breastscreening: On the use of multi-modality in medical imaging diagnosis. Proceedings of the International Conference on Advanced Visual Interfaces, 1–5.

Cancer, I. of M. (US) and N. R. C. (US) C. on T. for the E. D. of B., Nass, S. J., Henderson, I. C., & Lashof, J. C. (2001). 2 Breast Imaging and Related Technologies. In Mammography and Beyond: Developing Technologies for the Early Detection of Breast Cancer. Developing Technologies for the Early Detection of Breast Cancer. National Academies Press (US. http://www.ncbi.nlm.nih.gov/books/NBK222350/

Chen, G., Xie, X., & Li, S. (2020). Research on Complex Classification Algorithm of Breast Cancer Chip Based on SVM-RFE Gene Feature Screening. Complexity, 2020.

Das, A., Nair, M. S., & Peter, S. D. (2020). Computer-aided histopathological image analysis techniques for automated nuclear atypia scoring of breast cancer: a review. Journal of Digital Imaging, 33(5), 1091–1121.

Deng, C., Ji, X., Rainey, C., Zhang, J., & Lu, W. (2020). Integrating Machine Learning with Human Knowledge. Iscience, 101656. Deng, J., Ma, Y., Li, D., Zhao, J., Liu, Y., & Zhang, H. (2020). Classification of breast density categories based on SEAttention neural networks. Computer Methods and Programs in Biomedicine, 193, 105489.

Densitas® Inc. (2020). Densitas: Automated Breast Density, Quality, and Risk Software. Retrieved July 06, 2020, from https://densitas.health/

Deo, R. C. (2015). Machine learning in medicine. Circulation, 132(20), 1920–1930.

Ezhilraman, S. V., Srinivasan, S., & Suseendran, G. (2020). Gaussian Light Gradient Boost Ensemble Decision Tree Classifier for Breast Cancer Detection. In Intelligent Computing and Innovation on Data Science (pp. 31–38). Springer.

Faisal, Z., & El Abbadi, N. K. (2020). Breast Cancer Recognition by Computer Aided Based on Improved Fuzzy c-Mean and ANN. International Conference on New Trends in Information and Communications Technology Applications, 246–257.

Frank, A., & Asuncion, A. (2010). UCI Machine Learning Repository [http://archive. ics. uci. edu/ml]. Irvine, CA: University of California. School of Information and Computer Science, 213, 2.

Gandomkar, Z., Brennan, P. C., & Mello-Thoms, C. (2019). Computer-Assisted Nuclear Atypia Scoring of Breast Cancer: a Preliminary Study. Journal of Digital Imaging, 32(5), 702–712.

Ganesan, K., Acharya, U. R., Chua, C. K., Min, L. C., & Abraham, T. K. (2014). Automated diagnosis of mammogram images of breast cancer using discrete wavelet transform and spherical wavelet transform features: a comparative study. Technology in Cancer Research \& Treatment, 13(6), 605–615.

Gardezi, S. J. S., Elazab, A., Lei, B., & Wang, T. (2019). Breast cancer detection and diagnosis using mammographic data: Systematic review. Journal of Medical Internet Research, 21(7), e14464.

Gardezi, S. J. S., & Faye, I. (2015). Fusion of completed local binary pattern features with curvelet features for mammogram classification. Applied Mathematics \& Information Sciences, 9(6), 3037.

Gastounioti, A., Oustimov, A., Hsieh, M. K., Pantalone, L., Conant, E. F., & Kontos, D. (2018). Using convolutional neural networks for enhanced capture of breast parenchymal complexity patterns associated with breast cancer risk. Academic Radiology, 25(8), 977–984.

Gong, B., Shen, L., Chang, C., Zhou, S., Zhou, W., Li, S., & Shi, J. (2020). BI-Modal Ultrasound Breast Cancer Diagnosis Via Multi-View Deep Neural Network SVM. 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), 1106–1110.

Guo, R., Lu, G., Qin, B., & Fei, B. (2018). Ultrasound Imaging Technologies for Breast Cancer Detection and Management: A Review. In Ultrasound in Medicine and Biology (Vol. 44, Issue 1, pp. 37–70). https://doi.org/10.1016/j.ultrasmedbio.2017.09.012

Hamim, M., El Moudden, I., Moutachaouik, H., & Hain, M. (2020). Decision Tree Model Based Gene Selection and Classification for Breast Cancer Risk Prediction. International Conference on Smart Applications and Data Analysis, 165–177.

Hamouda, S., Hassan, A., Wahed, M. E., Ail, M., & Farouk, O. (2020). Tuning to Optimize SVM Approach for Breast Cancer Diagnosis with Blood Analysis Data. Available at SSRN 3537067.

Harris, K. M., & Vogel, V. G. (1997). Breast cancer screening. In Cancer and Metastasis Reviews (Vol. 16, Issues 3–4, pp. 231–262). https://doi.org/10.1023/a:1005893126451

Helvie, M. A., & Patterson, S. K. (2014). Chapter 11: Imaging analysis: Mammography. Diseases of the Breast.

Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. ArXiv Preprint ArXiv:1207.0580. ICAD. (2020).

ICAD Inc. | Home. Retrieved July 06, 2020, from https://www.icadmed.com/home.html

Kabiraj, S., Raihan, M., Alvi, N., Afrin, M., Akter, L., Sohagi, S. A., & Podder, E. (2020). Breast cancer risk prediction using XGBoost and random forest algorithm. 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 1–4. Journal of Scientific Research, Volume 65, Issue 5, 2021 283 Institute of Science, BHU Varanasi, India

Kajala, A., & Jain, V. K. (2020). Diagnosis of breast cancer using machine learning algorithms-a review. 2020 International Conference on Emerging Trends in Communication, Control and Computing (ICONC3), 1–5.

Khan, A. A., & Arora, A. S. (2018). Breast cancer detection through Gabor filter based texture features using thermograms images. 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC), 412–417.

Khan, A. M., Sirinukunwattana, K., & Rajpoot, N. (2015). A global covariance descriptor for nuclear atypia scoring in breast histopathology images. IEEE Journal of Biomedical and Health Informatics, 19(5), 1637–1647.

Khandezamin, Z., Naderan, M., & Rashti, M. J. (2020). Detection and classification of breast cancer using logistic regression feature selection and GMDH classifier. Journal of Biomedical Informatics, 111, 103591.

Kheiron medical. (2020). Kheiron Medical. Retrieved July 06, 2020, from https://www.kheironmed.com/

Kumar, A., & Poonkodi, M. (2019). Comparative study of different machine learning models for breast cancer diagnosis. In Innovations in soft computing and information technology (pp. 17–25). Springer.

Kumar, G., & others. (2019). Breast Cancer Detection Using Decision Tree, Naïve Bayes, KNN and SVM Classifiers: A Comparative Study. 2019 International Conference on Smart Systems and Inventive Technology (ICSSIT), 683–686.

LBN Medical. (2020). High quality fefurbished and used Mammography Equipment »LBN Medical. Retrieved July 11, 2020, from https://lbnmedical.com/solutions/

Lemons, K. (2020). A Comparison Between Naïve Bayes and Random Forest to Predict Breast Cancer. International Journal of Undergraduate Research and Creative Activities, 12(1).

Liu, J., Cheng, H., Lv, X., Zhang, Z., Zheng, X., Wu, G., Tang, J., Ma, X., & Yue, X. (2020). Use of FT-IR spectroscopy combined with SVM as a screening tool to identify invasive ductal carcinoma in breast cancer. Optik, 204, 164225.

Liu, X., & Tang, J. (2013). Mass classification in mammograms using selected geometry and texture features, and a new SVM-based feature selection method. IEEE Systems Journal, 8(3), 910–920.

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 International Conference on Artificial Intelligence and Big Data (ICAIBD), 9–12.

Lu, Y., Li, J. Y., Su, Y. T., & Liu, A. A. (2018). A review of breast cancer detection in medical images. 2018 IEEE Visual Communications and Image Processing (VCIP), 1–4.

Lunit. (2020). LunitINSIGHT MMGAI solution for Mammography. Retrieved July 09, 2020, from https://www.lunit.io/en

Marcia Boraas, M.D., F. (2019). Breast Self-Exam: How to Check for Lumps and Other Breast Changes. Breastcancer.Org. Retrieved July 06, 2020, from https://www.breastcancer.org/symptoms/testing/types/self_e xam

Marinelli, R. J., Montgomery, K., Liu, C. L., Shah, N. H., Prapong, W., Nitzberg, M., Zachariah, Z. K., Sherlock, G. J., Natkunam, Y., West, R. B., & others. (2007). The Stanford tissue microarray database. Nucleic Acids Research, 36(suppl_1), D871--D877.

Marne, S., Churi, S., & Marne, M. (2020). Predicting Breast Cancer using effective Classification with Decision Tree and K Means Clustering technique. 2020 International Conference on Emerging Smart Computing and Informatics (ESCI), 39–42.

Medical Image Analysis Group Eindhoven. (2020). Dataset | Tumor Proliferation Assessment Challenge 2016. https://tupac.tue-image.nl/node/3

Medical Imaging Technology Association. (2020). Current Edition. Retrieved July 06, 2020, from https://www.dicomstandard.org/current

Mohan, P., & others. (2020). A Comparison Between Knn And Svm For Breast Cancer Diagnosis Using Glcm Shape And Lbp Features. 2020 Third International Conference On Smart Systems And Inventive Technology (Icssit), 1058–1062.

Moon, W. K., Huang, Y.-S., Hsu, C.-H., Chien, T.-Y. C., Chang, J. M., Lee, S. H., Huang, C.-S., & Chang, R.-F. (2020). Computer-aided tumor detection in automated breast ultrasound using a 3-D convolutional neural network. Computer Methods and Programs in Biomedicine, 190, 105360.

Morais-Rodrigues, F., Silv?erio-Machado, R., Kato, R. B., Rodrigues, D. L. N., Valdez-Baez, J., Fonseca, V., San, E. J., Gomes, L. G. R., Dos Santos, R. G., Viana, M. V. C., & others. (2020). Analysis of the microarray gene expression for breast cancer progression after the application modified logistic regression. Gene, 726, 144168.

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(2), 236–248.

MurtiRawat, R., Panchal, S., Singh, V. K., & Panchal, Y. (2020). Breast Cancer Detection Using K-Nearest Neighbors, Logistic Regression and Ensemble Learning. 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), 534–540.

Nahid, A.-A., & Kong, Y. (2017). Involvement of machine learning for breast cancer image classification: a survey. Computational and Mathematical Methods in Medicine, 2017.

National Breast Cancer Foundation, I. (2020). Types of Breast Cancer - National Breast Cancer Foundation. Retrieved July 07, 2020, from https://www.nationalbreastcancer.org/types-of-breastcancer/

National Cancer Institute. (2020). How to Request Access to SEER Data - SEER Datasets. Retrieved July 05, 2020, from https://seer.cancer.gov/data/access.html

National Human Genome Research Institute. (2020). DNA Microarray Technology Fact Sheet. Retrieved July 08, 2020, from https://www.genome.gov/about-genomics/factsheets/DNA-Microarray-Technology

Ng, A. Y. (2004). Feature selection, L 1 vs. L 2 regularization, and rotational invariance. Proceedings of the Twenty-First International Conference on Machine Learning, 78. Journal of Scientific Research, Volume 65, Issue 5, 2021 284 Institute of Science, BHU Varanasi, India

O'Shea, B. (2020). K-nearest neighbors algorithm (KNN) and artificial neural networks (ANN) accurately predicting malignancy of breast cancer (BC).

Obaid, H. S., Dheyab, S. A., & Sabry, S. S. (2019). The impact of data pre-processing techniques and dimensionality reduction on the accuracy of machine learning. 2019 9th Annual Information Technology, Electromechanical Engineering and Microelectronics Conference (IEMECON), 279–283.

Ojansivu, V., Linder, N., Rahtu, E., Pietikäinen, M., Lundin, M., Joensuu, H., & Lundin, J. (2013). Automated classification of breast cancer morphology in histopathological images. Diagnostic Pathology, 8(1), 1–4.

Olea Medical. (2020). breastscape Breast MRI | Olea Medical. Retrieved July 06, 2020, from https://www.oleamedical.com/en/solutions/women-health/breastscape

Petushi, S., Katsinis, C., Coward, C., Garcia, F., & Tozeren, A. (2004). Automated identification of microstructures on histology slides. 2004 2nd IEEE International Symposium on Biomedical Imaging: Nano to Macro (IEEE Cat No. 04EX821), 424–427.

Pharswan, R., & Singh, J. (2020). Performance Analysis of SVM and KNN in Breast Cancer Classification: A Survey. In Internet of Things and Big Data Applications (pp. 133– 140). Springer.

Prabakaran, I., Wu, Z., Lee, C., Tong, B., Steeman, S., Koo, G., Zhang, P. J., & Guvakova, M. A. (2019). Gaussian mixture models for probabilistic classification of breast cancer. Cancer Research, 79(13), 3492–3502.

Quibim. (2020). Suites | Quibim. Retrieved July 06, 2020, from https://quibim.com/biomarkers

QView Medica. (2020). Home | QView Medical, Inc. Retrieved July 13, 2020, from https://www.qviewmedical.com/

R2 Technologies Corporation. (2020). Home - R2 Technologies Corporation - A specialty recruiting firm - Alpharetta GA. https://www.r2techcorp.com/

Radboud university medical. (2020). AI for radiology. Retrieved July 06, 2020, from https://grandchallenge.org/aiforradiology/

Raghavendra, U., Gudigar, A., Rao, T. N., Ciaccio, E. J., Ng, E. Y. K., & Acharya, U. R. (2019). Computer-aided diagnosis for the identification of breast cancer using thermogram images: a comprehensive review. Infrared Physics \& Technology, 102, 103041.

Rodriguez-Ruiz, A., Lång, K., Gubern-Merida, A., Broeders, M., Gennaro, G., Clauser, P., Helbich, T. H., Chevalier, M., Tan, T., Mertelmeier, T., & others. (2019). Stand-alone artificial intelligence for breast cancer detection in mammography: comparison with 101 radiologists. JNCI: Journal of the National Cancer Institute, 111(9), 916–922.

Sarica, O., & Uluc, F. (2014). Additional diagnostic value of MRI in patients with suspicious breast lesions based on ultrasound. The British Journal of Radiology, 87(1041), 20140009.

Saxena, S., & Gyanchandani, M. (2020). Machine learning methods for computer-aided breast cancer diagnosis using histopathology: a narrative review. Journal of Medical Imaging and Radiation Sciences, 51(1), 182–193.

ScreenPoint Medical. (2020). Welcome to ScreenPoint Medical - Home. https://screenpoint-medical.com/

Shaikh, T. A., & Ali, R. (2020). A CAD Tool for Breast Cancer Prediction using Naive Bayes Classifier. 2020 International Conference on Emerging Smart Computing and Informatics (ESCI), 351–356.

Singh, S., & Kumar, R. (2020). Histopathological image analysis for breast cancer detection using cubic SVM. 2020 7th International Conference on Signal Processing and Integrated Networks (SPIN), 498–503.

Siregar, A. T. M., Siswantining, T., Bustamam, A., & Sarwinda, D. (2020). Comparison of supervised models in hepatocellular carcinoma tumor classification based on expression data using principal component analysis (PCA). AIP Conference Proceedings, 2264(1), 30002.

Society, A. C. (2020). Breast Cancer Early Detection and Diagnosis. Retrieved July 11, 2020, from https://www.cancer.org/content/dam/CRC/PDF/Public/8579 .00.pdf

Spanhol, F. A., Oliveira, L. S., Petitjean, C., & Heutte, L. (2015). A dataset for breast cancer histopathological image classification. Ieee Transactions on Biomedical Engineering, 63(7), 1455–1462.

Su, K. Y., & Lee, W. L. (2020). Fourier transform infrared spectroscopy as a cancer screening and diagnostic tool: a review and prospects. Cancers, 12(1), 115.

Susan G. Komen. (2020). Quality of Screening Tests. Retrieved July 12, 2020, from https://www.komen.org/breastcancer/screening/quality-of-screening-tests/

Syed, L., Jabeen, S., & Manimala, S. (2018). Telemammography: a novel approach for early detection of breast cancer through wavelets based image processing and machine learning techniques. In Advances in Soft Computing and Machine Learning in Image Processing (pp. 149–183). Springer.

Tabrizchi, H., Tabrizchi, M., & Tabrizchi, H. (2020). Breast cancer diagnosis using a multi-verse optimizer-based gradient boosting decision tree. SN Applied Sciences, 2(4), 1–19.

TCIA. (2020). Collections - The Cancer Imaging Archive (TCIA) Public Access - Cancer Imaging Archive Wiki. Retrieved July 10, 2020, from https://wiki.cancerimagingarchive.net/display/Public/Collect ions

Thakare, R. S., Deshmukh, S. M., & Raut, V. R. (n.d.). AUTOMATIC BREAST SEGMENTATION AND CANCER DETECTION USING SVM.

Torres, G. F., Sassi, A., Arponen, O., Holli-Helenius, K., Lääperi, A.-L., Rinta-Kiikka, I., Kämäräinen, J., & Pertuz, S. (2019). Morphological area gradient: System-independent dense tissue segmentation in mammography images. 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 4855–4858.

Valencia-Hernandez, I., Peregrina-Barreto, H., Reyes-Garcia, C. A., & Lopez-Armas, G. C. (2021). Density map and fuzzy classification for breast density by using BI-RADS. Computer Methods and Programs in Biomedicine, 200, 105825.

van Leeuwen, K. G., Schalekamp, S., Rutten, M. J. C. M., van Ginneken, B., & de Rooij, M. (2021). Artificial intelligence in radiology: 100 commercially available products and theirscientific evidence. European Radiology, 31(6), 3797–3804.

Vara. (2020). Vara | Home. Retrieved July 09, 2020, from https://www.vara.ai/

VCL. (2020). Mammographic Image Analysis Homepage - Databases. Retrieved July 14, 2020, from https://www.mammoimage.org/databases/

Venkata, M. D., & Lingamgunta, S. (2020). Breast Cancer Multi Modality Image Analysis Using Pheneotype features by SVM. Journal of Science and Technology, 5(1), 52–60.

Veta, M., Kornegoor, R., Huisman, A., Verschuur-Maes, A. H. J., Viergever, M. A., Pluim, J. P. W., & Van Diest, P. J. (2012). Prognostic value of automatically extracted nuclear morphometric features in whole slide images of male breast cancer. Modern Pathology, 25(12), 1559–1565. Volpara Health. (2020). Breast Screening | Breast Health Platform |

Volpara Health. Retrieved July 06, 2020, from https://www.volparahealth.com/breast-health-platform/

Vrigazova, B. P. (2020). Detection of Malignant and Benign Breast Cancer Using the ANOVA-BOOTSTRAP-SVM. Journal of Data and Information Science, 5(2), 62.

Wadali, J. S., Sood, S. P., Kaushish, R., Syed-Abdul, S., Khosla, P. K., & Bhatia, M. (2020). Evaluation of free, open-source, web-based DICOM viewers for the Indian national telemedicine service (eSanjeevani). Journal of Digital Imaging, 33(6), 1499–1513.

Wahab, N., Khan, A., & Lee, Y. S. (2019). Transfer learning based deep CNN for segmentation and detection of mitoses in breast cancer histopathological images. Microscopy, 68(3), 216–233.

Wang, S., Wang, Y., Wang, D., Yin, Y., Wang, Y., & Jin, Y. (2020). An improved random forest-based rule extraction method for breast cancer diagnosis. Applied Soft Computing, 86, 105941.

Wang, Z., Li, M., Wang, H., Jiang, H., Yao, Y., Zhang, H., & Xin, J. (2019). Breast cancer detection using extreme learning machine based on feature fusion with CNN deep features. IEEE Access, 7, 105146–105158.

Yanase, J., & Triantaphyllou, E. (2019). A systematic survey of computer-aided diagnosis in medicine: Past and present developments. Expert Systems with Applications, 138, 112821.

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

Yeh, J. Y., & Chan, S. (2017). Population-based metaheuristic approaches for feature selection on mammograms. 2017 IEEE International Conference on Agents (ICA), 140–144.

Zebra Medical Vision. (2020). Mammography – Zebra Medical Vision | Medical Imaging & AI. Retrieved July 06, 2020, from https://www.zebra-med.com/solutions/mammography

Zhou, X., Li, C., Rahaman, M. M., Yao, Y., Ai, S., Sun, C., Wang, Q., Zhang, Y., Li, M., Li, X., & others. (2020). A comprehensive review for breast histopathology image analysis using classical and deep neural networks. IEEE Access, 8, 90931–90956.

Zou, L., Yu, S., Meng, T., Zhang, Z., Liang, X., & Xie, Y. (2019). A technical review of convolutional neural networkbased mammographic breast cancer diagnosis.

Computational and Mathematical Methods in Medicine, 2019.

Zuluaga-Gomez, J., Zerhouni, N., Al Masry, Z., Devalland, C., & Varnier, C. (2019). A survey of breast cancer screening techniques: thermography and electrical impedance tomography. Journal of Medical Engineering \& Technology, 43(5), 305–322.

***