IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁘ Impact Factor 7.105 ⁘ Vol. 9, Issue 7, July 2022

DOI: 10.17148/IARJSET.2022.9718

Survey on Deep learning for the detection of breast cancer

Lakshmi Sagar D U[1], Suma N R[2]

[1]Dept. of MCA, Bangalore Institute of Technology (BIT), Bengaluru, India

[2]Dept. of MCA, Assistant Professor, Bangalore Institute of Technology (BIT), Bengaluru, India

Abstract: One in eight women may suffer from breast cancer at some point in their lives, making it a frequent health issue for women. Since radioactive radiation exposure makes breast cancer treatment risky, many women choose to forego getting the disease diagnosed. The non-invasive nature, the danger of radiation, and the specificity of the identification of breast tumors are all problems with breast cancer screening methods. Medical imaging often makes use of deep learning algorithms. This essay seeks to offer a thorough analysis of the benefits and drawbacks of breast cancer screening methods. Deep learning techniques' potential for use in the early identification of breast cancer is being investigated. Investigations are also conducted on the datasets and performance metrics for breast cancer. The directions for breast cancer research in the future are examined. The main goals are to provide an in-depth analysis of this area and to inspire creative researchers.

Keywords: SVM, Machine Learning, Deep Learning, CNN, VGG-16, and Breast Cancer.

INTRODUCTION

Once malignant tumors are formed by aberrant cells interacting with healthy ones, cancer is formed. Breast cancer is the most frequent and deadly illness in the world. Invasive and non-invasive breast cancers are the two main forms. Invasive and cancerous, invasive spreads to other organs. Non-invasive and precancerous, it remains in the original organ. In the end, it leads to a more aggressive form of breast cancer. Breast cancer may be found in the glands and milk ducts that carry milk throughout the body. Breast cancers typically spread to other parts of the body and become more aggressive as a result. It may also spread to other organs thanks to the circulatory system. Different types of breast cancer develop at different rates. According to the World Health Organization, breast cancer claimed the lives of 627,000 women in 2018. Breast cancer is the most important concern, affecting people all over the globe, but it is particularly widespread in the United States. Breast cancer is divided into four distinct subtypes. First, there is Ductal Carcinoma in Situ, a kind of early-stage breast cancer found in the lining of the milk ducts. Breast cancer is most often diagnosed in women with the second type of the disease, which accounts for 70% to 80% of all cases. Inflammatory breast cancer is a third kind of breast cancer in which cancer cells invade the surface and lymphatic vessels of the breast quickly and aggressively. The fourth kind of breast cancer is cancer that has spread to other parts of the body.

Numerous diagnostic procedures, including mammography, ultrasound, MRI, and biopsies, produced the pictures needed for classification. X-rays are used in mammograms to detect breast cancer. If any questionable results are discovered during a mammography screening, the doctor will be informed and the tissues will be tested. After the mammography, an ultrasound is performed. When a suspicious area in your breast is found, a doctor will order an ultrasound. If the tests performed during a symptomatic examination are inconclusive, the doctor will favor a breast MRI. It depicts your condition and viewpoint from that perspective. Biologic testing is the main diagnostic method used to determine if a region is carcinogenic. Because of this, most women who get a breast biopsy do not have a malignant tumor in their breasts.

The classification of breast cancer benefits greatly from machine learning. There are numerous diagnosis procedures, which are illustrated by the photos presented above. Machine learning is used to classify these types of diagnostic images. AI includes the subfield of machine learning. Many developers utilize machine learning to improve the performance of their models by retraining them. The linear data is processed using machine learning. Machine learning produces better outcomes when the data is tiny, but when the data is too vast, it does not. The model is trained using one of three main types of machine learning. Machine learning under supervision uses known data and the supervisor's assistance to complete tasks. Machine learning that is not under supervision is called unsupervised. Less machine learning reinforcement is used. These algorithms pull the most relevant data from prior knowledge to make precise decisions.

It is a subfield of machine learning known as deep learning. Unsupervised data is used in deep learning to learn from the data itself. Neither the information nor its structure may be restricted. A deep learning model is said to be deep and will have more than 2 layers. The top layer is called the input layer, and the bottom layer is called the output layer. The middle

DOI: 10.17148/IARJSET.2022.9718

layer, which has more layers than a neural network, is known as the hidden layer. Neurons refer to the node in which the layer is located.

Deep learning and machine learning vary in that deep learning is more akin to its end objective than machine learning. The breast cancer dataset is classified using a convolution neural network. The images are classified by utilizing a convolutional neural network. The breast cancer image dataset is used.. The photos are provided to CNN along with the corresponding weights as input. To reduce mistakes and improve performance, the weights are changed. The convolution layer, pooling layer, ReLU layer, and fully connected layer are just a few of the many layers that make up CNN. A feature map is used in the convolution layer to extract the given image's features and compress the original image. The image's dimensions are decreased by using a pooling layer. To check whether the valuation of the kernel function falls within a given range, It is utilized as an activation function for the ReLU layer. The model's final layer is the fully connected layer. It aggregates the findings from all layers and uses the softmax algorithm to assign probabilities to each output class.

LITERATURE OF SURVEY

Medical image processing has been the subject of a wide range of recent studies. Researchers in various fields such as computer vision, image processing, and machine learning have found a place in the field of medical imaging. [2] Recently, this researched some of the existing papers to find the most useful and advanced methods used in the existing articles. This chapter details these papers related to our survey.

The model was proposed by Megha Rathi et al. [3] and is based on a hybrid strategy incorporating machine learning. This tactic, which included MRMR feature selection and the use of four different classifiers, was utilized to get the best possible outcomes. The author's research evaluated the effectiveness of these four different classifiers: SVM, Naive Bays, Function Tree, and End Meta. It was found that the SVM was an efficient classifier. to uncover even better results.

Nathan Jacobs, Jinze Liu, and Erik Y. Han's "Whole Mammogram Image Classification using Convolutional Neural Networks" [4], published in 2017. This article presents preliminary research on creating and improving machine learning models for mammography image classification. They looked at 7 different CNN designs and came to the conclusion that the best way to boost classification performance is to use a CNN in conjunction with both data augmentation and transfer learning techniques.

In the publication titled [5] "Research on the Deep Neural Network Computer - Based Aid Detection Approach of Breast Cancer," With the help of features developed with the aid of computers, the CNNs approach is used to categorize the image and detect the method. The accuracy of their method was found to be around 89%, and Consequently, when compared to conventional techniques, the efficiency of the recognition and classification of breast cancer is greatly increased. To automatically extract the distinguishing features and combine the features extracted from the two structures, CNNs with various architectures were pre-trained before being employed.

MODERN BREAST CANCER DETECTION TECHNOLOGIES

Technology in the medical industry has advanced steadily throughout time. [6] Different techniques are employed to detect breast cancer. The most important of these are highlighted in the figure.

- Breast screening is the process of checking for diseases early on before symptoms arise. Physical or imaging-related exams, genetic tests to detect inherited illnesses, and other tests can be carried out as a part of a screening test. Some of the imaging methods used for breast cancer screening include mammograms, magnetic resonance, whole-slide imaging (WSI), and ultrasound.

- Biosensors are analytical tools for measuring the biological properties of body fluids and tissues. Optical, piezoelectric, and electrochemical diagnostic devices are a few types of biosensors.

- In order to identify key features from large, complicated datasets, machine learning (ML) techniques are largely regarded as the preferred strategy for classifying breast cancer types and modeling decision forecasts. The classification of data has produced encouraging results using the KNN, SVM, and DT approaches. Additionally, these techniques significantly aid in clinical diagnosis and decision-making.

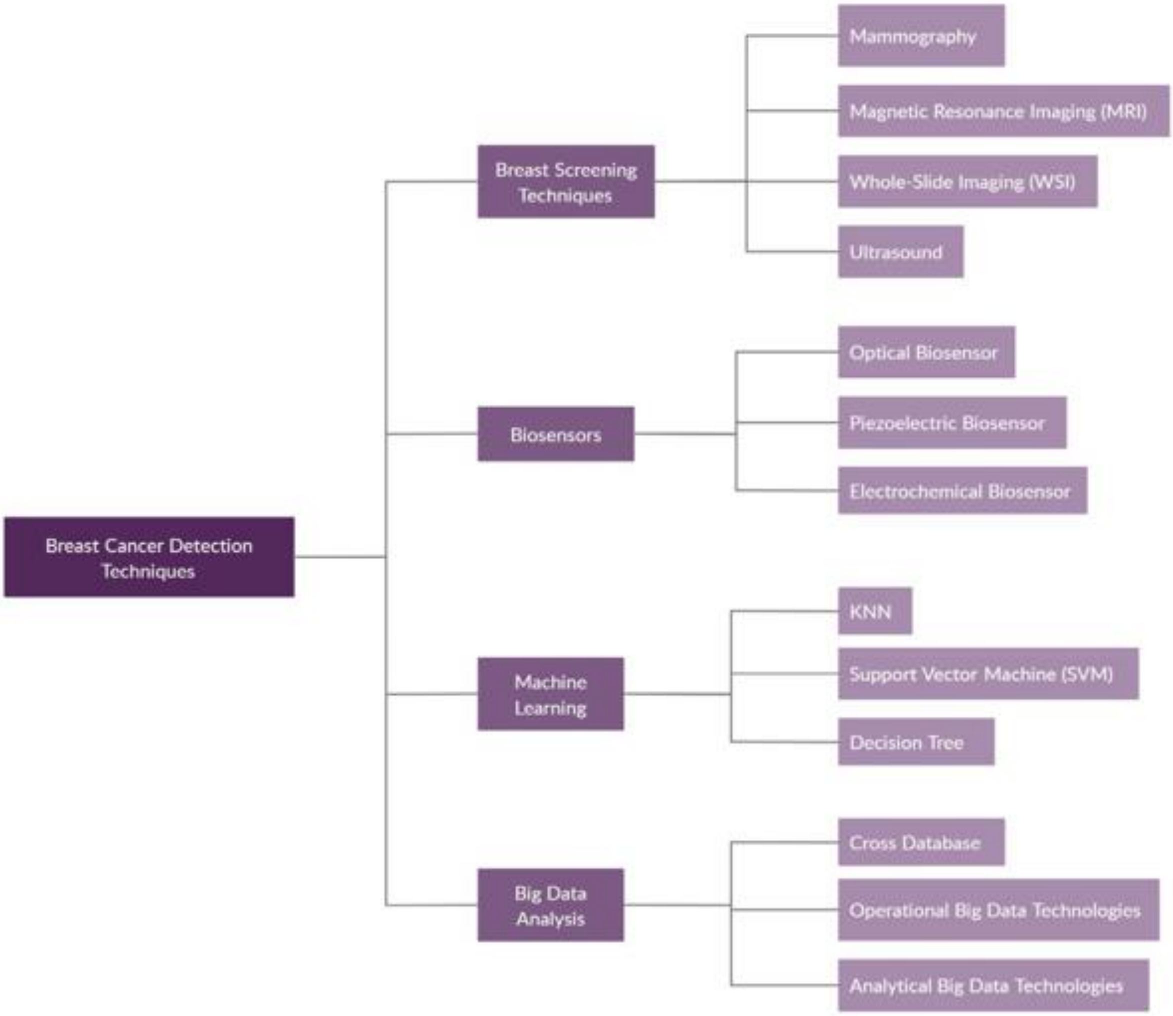
- Specialists can use unstructured data, such as text-based patient documents or photographs, by analyzing data using big data techniques, which has an impact on medical research and ultimately patient care. Some techniques used in big data analysis include cross-database, operational, and analytical big data.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁑ Impact Factor 7.105 ⁑ Vol. 9, Issue 7, July 2022

DOI: 10.17148/IARJSET.2022.9718



Uses of Deep Learning

- Automatic voice recognition [7].
- Digit recognition for calligraphy.
- Cancer screening
- Recognition of images

METHODOLOGY

A Convolutional Network for Detection and Classification (VGG16)

The deep neural network model known as VGG16 was introduced by K. Simonyan and A. Zisserman from Oxford University in their article "Very CNN for High-Resolution Recognition." With more than 14 image net datasets separated into 1000 classes, ImageNet's top-5 accuracy rate for the model is 92.7 percent. [8] One of the most well-known models submitted at ILSVRC-2014 was this one. It improves AlexNet by replacing AlexNet's enormous kernel-sized filtering with a series of 3x3 kernel-sized filters. The week-long learning of VGG16 took place on the NVIDIA Titan Black GPUs.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⋇ Impact Factor 7.105 ⋇ Vol. 9, Issue 7, July 2022

DOI: 10.17148/IARJSET.2022.9718

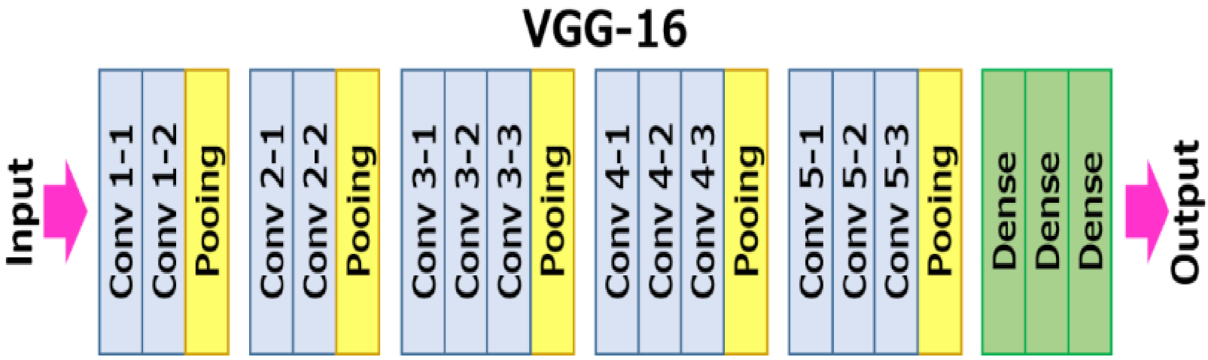


Figure: VGG – 16.

Data Set

The ImageNet collection consists of around 22,000 categories and 15 million high-resolution photos with annotations. Human labelers gathered the media files and applied labels to them using the crowdsourcing platform offered by Amazon Mechanical Turk. Since 2010, the Pascal Visual Object Competition has included the ImageNet Massive Visual Recognition Challenges, an annual competition. According to ILSVRC, each of the subset's 1000 categories has around 1000 images. [9] The collection consists of about 1.2 million train pictures, 50,000 validating photos, and 150,000 check shots. Different ImageNet images have different resolutions. The photographs were consequently scaled down to 256 x 256, a set resolution. The mid 256 x 256 patch is removed from the image after resizing a rectangular image.

The Architecture

VGG16 is depicted in the structure below.

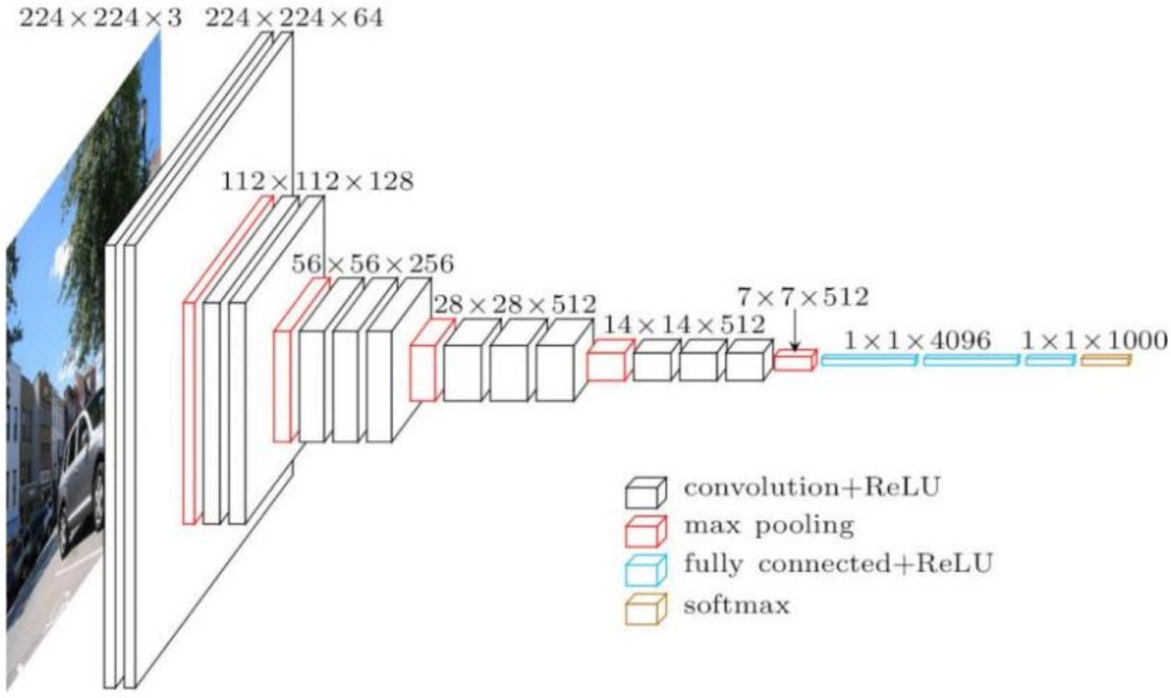


Figure 1. VGG16 Architecture

Diagram: VGG16 Structure.

The RGB values are input to the cov1 layer with a fixed input size of 224 x 224. The picture is subjected to convolutional (Conv.) filtering using a 3 x 3 visual field before being combined. [10] One of the configurations also uses 1x1 convolution filters, which may be thought of as an input channel sequence function. To ensure that the resolution is maintained after convolution, the spatial padding for the convolutional input is 1 bit for 3x3 convolution. One of the configurations also uses 1x1 convolution filters, which may be thought of as an input channel sequence function. To ensure that the resolution is maintained after convolution, the spatial padding for the convolutional input is 1 bit for 3x3

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⋇ Impact Factor 7.105 ⋇ Vol. 9, Issue 7, July 2022

DOI: 10.17148/IARJSET.2022.9718

convolution. A stack of convolutional layers is followed by the three fully connected (FC) layers, the third of which conducts 1000-way ILSVRC classification and contains 1000 channels. The first two FC levels each include 4096 channels. The soft-max layer is the last one. In every network, the FC layers are configured in the same manner.
Rectification is a non-linear process that affects all buried strata equally (ReLU). Notably, none of the networks make use of the Local Reaction Normalization (LRN) technique, which increases the amount of time needed for calculation and the amount of memory needed without improving the quality of the ILSVRC dataset.

CONFIGURATIONS

The figure displays the ConvNet installations. [11] They are referred to by their names, the nets (A-E). Among network A's 11 weight levels & network E's 19 weight layers, there are only minor details that differ between all arrangements from the standard architectural design. The first layer of a convolutional layer has a width of 64, and the subsequent layers increase in breadth by a factor of two while maintaining maximum pooling until the layer width reaches 512.

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

Figure: Conv Net Configuration.

Utilization and application
Unfortunately, VGGNet has two significant flaws:

1. Training moves very slowly.
2. The overheads of the network architecture are rather substantial.

IARJSET

ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⋇ Impact Factor 7.105 ⋇ Vol. 9, Issue 7, July 2022

DOI: 10.17148/IARJSET.2022.9718

VGG16 is larger than 533 megabytes owing to the intricacy and number of nodes that are fully coupled to one another. As a direct consequence of this, implementing VGG becomes a challenging task. The VGG16 technique is used in a wide variety of deep learning problems requiring the classification of pictures. However, smaller network architectures are typically favored. But because it is simple to use, it makes a fantastic building block for instructional purposes to implement.

a. Pytorch
b. Tensorflow
c. Keras

RESULT

ILSVRC-2012 and ILSVRC-2013 competitions are where VGG16 performs considerably better than the older versions. [12] The winner of the contest, the ILSVRC-2013, Clarifai, achieved 11.2 percent with an intensive training dataset and 11.7 percent without it. The VGG16 result greatly outperforms both scores. The VGG16 output is also vying with Google Net for the title of classification task champion with a 6.7 percent error rate. When it comes to single-net performance, the VGG16 architecture outperforms a single GoogLeNet by 0.9%, achieving the best result of 7.0 % test error.

Method	top-1 val. error (%)	top-5 val. error (%)	top-5 test error (%)
VGG (2 nets, multi-crop & dense eval.)	**23.7**	**6.8**	**6.8**
VGG (1 net, multi-crop & dense eval.)	24.4	7.1	7.0
VGG (ILSVRC submission, 7 nets, dense eval.)	24.7	7.5	7.3
GoogLeNet (Szegedy et al., 2014) (1 net)	-	7.9	
GoogLeNet (Szegedy et al., 2014) (7 nets)	-	**6.7**	
MSRA (He et al., 2014) (11 nets)	-	-	8.1
MSRA (He et al., 2014) (1 net)	27.9	9.1	9.1
Clarifai (Russakovsky et al., 2014) (multiple nets)	-	-	11.7
Clarifai (Russakovsky et al., 2014) (1 net)	-	-	12.5
Zeiler & Fergus (Zeiler & Fergus, 2013) (6 nets)	36.0	14.7	14.8
Zeiler & Fergus (Zeiler & Fergus, 2013) (1 net)	37.5	16.0	16.1
OverFeat (Sermanet et al., 2014) (7 nets)	34.0	13.2	13.6
OverFeat (Sermanet et al., 2014) (1 net)	35.7	14.2	-
Krizhevsky et al. (Krizhevsky et al., 2012) (5 nets)	38.1	16.4	16.4
Krizhevsky et al. (Krizhevsky et al., 2012) (1 net)	40.7	18.2	-

VGG16 comparison

Figure: VGG16 comparison.

It was proven that representation depth improves classification accuracy and that a standard ConvNet architecture with significantly more depth can deliver cutting-edge results on the ImageNet validation set.

CONCLUSION

Breast cancer, which is both the most common and the most lethal illness, is the one that is the most difficult to diagnose. The number of women who are diagnosed with breast cancer every year also rises, but the likelihood of a full recovery falls. ML and deep learning are two methods that are used in the breast cancer detection process. Research conducted in the past reveals that the use of machine learning strategies yields superior outcomes in their industry. In the prior study, a variety of machine learning algorithms, together with a certain set of data augmentation and enrichment procedures in order to achieve higher levels of performance. On the other hand, it has been shown that machine learning functions more effectively with linear data. In addition, it has been concluded from earlier research that the data was in the form of pictures while the system was malfunctioning. This was deduced from the findings of the prior study. A novel approach has been implemented in order to solve the problem using methods based on machine learning. The newly developed technique of deep learning is used rather often in the field of data science. CNN, a technology that is based on deep learning, is used in order to classify the photo data pertaining to breast cancer. The image dataset is used extensively by

DOI: 10.17148/IARJSET.2022.9718

CNN. As was found in the prior research, which came to the same result, CNN performed far better than other machine learning algorithms.

We assessed very deep CNNs for massive picture categorization in this investigation. It was shown that representation depth improves classification accuracy and that cutting-edge effectiveness on the ImageNet validation set may be attained by using a traditional ConvNet architecture with significantly more depth. Additionally, we demonstrate in the appendix how effectively the models generalize to a variety of functions and datasets, comparable to or surpassing more intricate recognition pipelines constructed using shallower picture representations. The outcomes provide more evidence of the significance of depth in pictorial representations.

REFERENCES

[1] A Literature Review on the Application of Deep Learning to the Detection of Breast Cancer To properly reference this article, please mention the full name of the author, Kumar Sanjeev Priyanka, as well as the conference series number, 1022-012071, which was presented at the 2021 IOP Conf.

[2] (Rashed, Gaber, and Karam) Rashed, Edward A. (1922). A completely automated technique has been developed to diagnose breast cancer using thermal images and deep learning. DOI: 10.1371/journal.pone.0262349. doi:10.1371/journal.pone.0262349. You may read it online at that address.

[3] Deep Learning for the Early Detection of Breast Cancer An Indian city called Mysuru, home to the departments of Ece Shwetha K, Sindhu S, Spoorthi M, and Chaithra D, is where it is situated. ISSN: 2278-0181 The International Journal of Engineering Research and Technology is published on the website www.ijert.org, a non-profit media organization.

[4] A journal called (IJRASET), ISSN 2321-9653, IC Value: 45.98, SJ Impact Factor: 7.538, Volume 10, Issue 3, March 2022, is accessible at www.ijraset,com

[5] A comparison of breast cancer detection equipment and algorithms Prince Pipariya, Shlok Patel, Yash Amethiya, and Manan Shah are students at the Pandit Deendayal Energy University's Department of Chemical Engineering School of Technology in Gandhinagar, Gujarat, India.

[6] Predicting breast cancer using an optimized deep learning technique The suspect has been identified as Hager Saleh. Sara F. Abd-el-Ghani is the first author. Hurghada Egypt: Alyami, Hashem3, and Alosaimi, Wael 4 1 Department of Computer Science, College of Computers and Information Technology, Taif University, P.O. Box 11099, Taif, Saudi Arabia 21944, Saudi Arabia, Department of Computer and Information Technology, College of Computer and Information Technology, Taif University 2 Luxor University, Faculty of Computing and Information, Luxor, Egypt 3 Alosaimi Wael Alosaimi's email, may be found here: w.osaimi@tu-edu-esa.

[7] Mehedi Masud1, Amr E. Eldin Rashed1, M. Shamim Hossain2,3, CNN-based models for detection of breast cancer, Springer-Verlag London Ltd., a component of Springer Nature 2020

[8] 7 May 2020; NIT, Silchar, India; researcher, Hyderabad, India; Department of Computer Science and Engineering. Detection of breast cancer through machine learning In addition to Anji Reddy Vaka and Sonia Badal, Sudheer Reddy K. also participated.

[9] Proceedings of the International Society for Modeling and Simulation 2021, 97(8) 511–527. Doi: 10.1177/0037549721996031 - The Authors 2021 The evolutionary algorithm was used to evolve convolutional neural network parameters for the breast cancer classification challenge of Khatereh Davoudi and Parimala Thulasiraman.

[10] In the case of Mohamed AA, Rashed A, Gaber A, and Karam O (2022) To identify breast cancer using thermograms and a deep learning network, You may find it at https://doi.org/10.1371/journal-pone.0026349.

[11] Using Deep Process to Improve Mammography Screening for Breast Cancer Lauren Margolies1, Joseph H Rothstein 2, Eugene Fluder 3, Russell McBride 4, and Weiva Sieh 3 are all members of the group Li Shen 1. Updated on August 29th, 2019.