Early Detection of Breast Cancer Using Deep Learning Algorithms

Dr.T.Ramaswamy , J.Arun Kumar , A.Haripriya ,G.Reethika

Associate professor Department of Electronics

and Communication Engineering Sreenidhi Institute of Science and Technology

Hyderabad, Telangana

Assistant Professor Dept. of ECE SNIST Hyderabad ,Telangana, India

Student Dept. of ECE SNIST Hyderabad, Telangana, India

Student Dept. of ECE SNIST Hyderabad, Telangana, India

India -501301

ramaswamyt@sreenidhi.edu.in, arunk4329@ gmail.com , akulaharipriya8@ gmail.com, gugulothureethika17@gmail.com

***Abstract*:** This investigation into breast cancer diagnosis from mammography pictures uses deep learning algorithms MobileNetV2 and VGG. This research used a publicly accessible dataset of mammograms to train classification algorithms to distinguish between cancerous and noncancerous tissue. Both algorithms' efficacy was measured by a variety of criteria, including accuracy, precision, recall, and F1 score. High levels of accuracy in breast cancer diagnosis were reached by both MobileNetV2 and VGG, with VGG marginally beating MobileNetV2. This work adds to the growing body of evidence that deep learning algorithms may be useful for enhancing the accuracy and efficiency of breast cancer diagnosis using mammography pictures.

***Keywords*:** *F1 Score, Accuracy, Precision, Recall, Malignancy, MobileNetV2, VGG, Breast Cancer, Mammography.*

I.INTRODUCTION

The WHO [1] reports that breast cancer is the most common cancer in women globally. among addition, it accounts for the majority of cancer deaths among females. Breast cancer has the highest fatality rate of all female malignancies and is the most common one. (about 25%) in Malaysia . While about 12.5 percent of women in Europe and the United States face a risk of developing breast cancer, that number is only 5 percent among Malaysian women . Compared to women in other countries, Malaysian women with breast cancer present at a more advanced stage of the illness. If certain signs are present, breast cancer may usually be discovered early. However, many women

who have breast cancer show no signs of illness. Therefore, it is crucial to screen for breast cancer on a regular basis so that it can be detected early. In order to improve the prognosis and increase the likelihood of a successful treatment, early identification of breast cancer is essential. The ability to identify, diagnose, and treat cancer early is crucial since it decreases the patient's probability of dying from the disease. Prolonged intervals between the first discovery of breast cancer and the beginning of therapy are of prognostic concern since any delay in early detection leads to disease progression and treatment complications.

Breast cancer has risen to the top spot as the main cancer killer among women worldwide [1, 2, 3]. Over 40,000 women and roughly 600 men in the United States lost their lives to breast cancer in the most recent year for which data is available from the American Cancer Society. Research published between 2010 and 2021 is taken into account in order to give the most up-to-date information possible for breast cancer diagnosis. We also explore the difficulties and provide suggestions for further study to aid scholars and practitioners in this area. The following are the major findings from this research:

In this essay, we take a look at the state of the art in breast cancer diagnosis using deep learning. Improving breast cancer diagnosis procedures using deep learning. Examining the most widely used data sets for deep learning breast cancer diagnostic algorithms. Discussion of deep learning-based methods for breast cancer detection and a description of the relevant assessment measures. Examining what is known about using deep learning to identify breast cancer and what could be learned in the future.

II.LITERATURE SURVEY

Automatically inferring feature representations from data through a "learning representation" [1, 2] is the goal of deep learning, a kind of machine learning. As opposed to more conventional learning techniques like the support vector machine (SVM), the K-nearest neighbors (KNN), the random forest (RF), etc., deep learning does not require a human-engineered feature for optimal performance [10, 12]. Applications of Recurrent Neural Networks, Convolutional Neural Networks, and Restricted Boltzmann Machines (RNNs), deep auto encoders (AEs), multi-layer perceptrons (MLPs), and adversarial networks (GANs) are only few of the deep learning techniques developed in recent decades [25,26]. These models have been put to use and shown to be effective in a number of contexts.

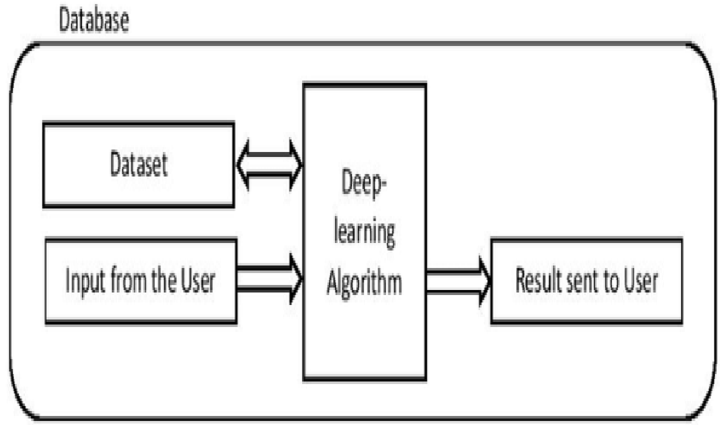
Breast cancer detection, diagnosis, and avoidance of unnecessary treatment are all areas where artificial intelligence (AI) may be put to good use. However, accurate predictions and decisions are made possible through the integration of AI and ML techniques. Consider the case of a breast cancer patient who is waiting for biopsy results before choosing whether or not to have surgery. Mammograms are the most popular screening tool; however they might provide false positive (high-risk) findings by accidentally revealing healthy tissue while malignant cells are present. In certain cases, a biopsy or surgical removal of a lesion finds that it is harmless. This implies the patient will have to undergo an invasive and costly operation that is not required.

It has been demonstrated in a number of studies that catching breast cancer before it spreads considerably increases the likelihood of survival. [6].

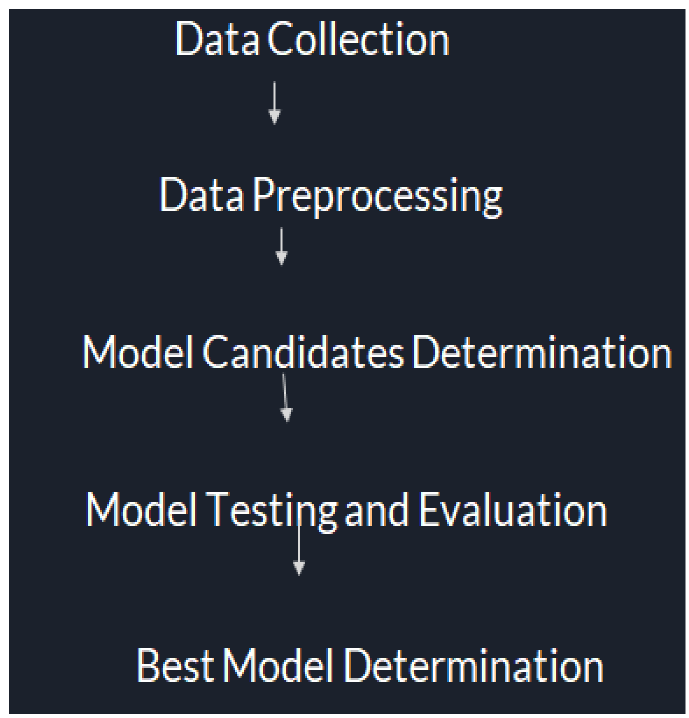
Based on these evaluations, it's evident that deep learning-based techniques have not yet reached their full potential when it comes to breast cancer diagnosis, since most current review studies focus on image-based approaches. Most recent studies have concentrated on more traditional machine learning approaches, whereas the handful that have investigated deep learning-based methods have barely scratched the surface.

III. METHODOLOGY

The proposed solution fo this dreadful disease is using VGG 19(Visual Geometry Group 19) Algorithm and MobileNetV2 .We give input images and using different objects in image and differentiate the outputs of these algorithms and by comparing both outputs we can get an exact or more accurate way for prediction of Cancer.



System architecture



Flow chart

Data Collection:
Over fitting problems arising from the small number of training images is the model's

starting point. To solve this problem, we first apply data augmentation to enable the application of parameterized transformations including Changing the angle, size, position, and gamma of an image. There are 162 photos of breast cancer in the Breast Cancer Histopathological Image Classification (BreakHis) database, along with 2480 images that are benign and 5429 images that are malignant. Transformational methods are helpful, and they include inversion, rotation, shifting, scaling, and gamma.

Data Preprocessing

Using a simple H&E color normalization method that we found online, we were able to give all of the photographs a consistent look. After that, the images were enlarged to four different sizes, each with a different aspect ratio: 1:1, 1:2, 0.3, and 0.2. These enlarged photos were cut in half such that each of the four sections was the exact same but did not overlap. The original patches were created without any enhancements, and then they were stitched together to form a larger 224x224 pixel image. We managed to get everything we required. Our team joined four adjacent patches that did not overlap to generate a larger, more useful dataset. Since staining and recording can happen under a wide range of circumstances, pathologists are taught to examine histological images from a number of angles. We employed several data augmentation methods, including the image may be flipped horizontally and vertically, rotated, shifted, brightened, magnified, and blurred to represent the actions of a pathologist. and to depict real-world variance. The goal was to simulate the conditions in which a pathologist would use these methods in the course of their practice. Adding new data to the dataset will not compromise its security in any way, regardless of how large it eventually becomes. Data augmentation and patching strategies have been proven to be theoretically possible for histology classification in a published research. New research findings provide credence to this argument. After consensus was reached, each of the new patches was given a name that corresponded to the category of the original image that had been used to create it. The model was trained using image-net weights and then compared its estimates of patch sizes. This was done once model training was finished.

IV.TESTING AND EVALUATION

To determine the overall picture's classification, a patch-wise classifier first analyzes many isolated image patches at different zoom levels. The next phase will include combining the findings of this processing with those of processing the whole picture patches to get an image-wise categorization. The number of completed image patches will be used to label this group. Finding and studying the pictures' essential components is a prerequisite to correctly classifying histology photos into one of the many needed categories. This is necessary before the photographs can be correctly labeled. The provided information elaborates on the form, composition, and structural organization of the nucleus. Cancer cells may be distinguished from normal cells based on their nucleus's size, shape, and color. The lack of certain features allows for the identification of normal cells. This is an essential step in identifying cancerous cells from healthy ones (via characteristics like density and diversity). To distinguish in situ carcinomas from those that have spread a thorough familiarity with tissue architecture is essential. Learned characteristics may be generalized to a broad variety of geographic scales using the categorization approach, from glands to cell nuclei.

Pathologists might benefit from automatic analysis of biological data in making quickly, on-the-spot diagnosis. By isolating and eliminating potentially benign areas, these methods help pathologists save time. They also aid pathologists in making an early diagnosis of breast cancer, which saves lives. Research paradigm, strategy, and framework evaluations are essential. The effectiveness of segmentation and classifiers may be evaluated in a verified system. Therefore, two datasets are required for training and testing purposes. The system has to be tested on a new dataset free from any potential for memorizing.

V.ALGORITHMS

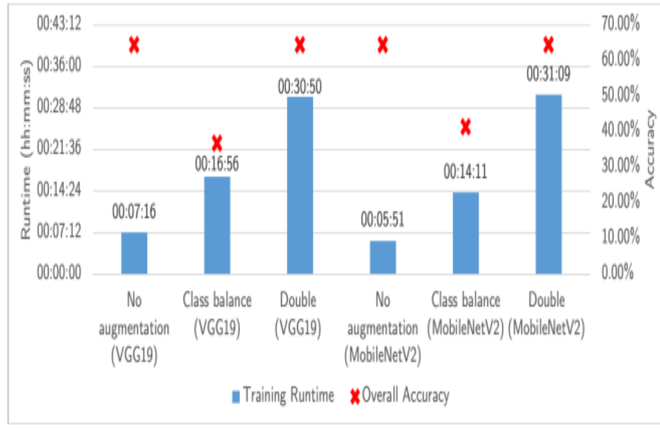
VGG (Visual Geometry Group 19):

VGG16 Graphics Processing Unit Architecture Design When the network size is quite large; the deep neural network really shines. A, A-LRN, B, C, D, and E are only few of the VGG networks developed by the Visual Geometry Group. In Figure 4, two VGG16 networks (C and D) with a total of 16 layers (13 convolution layers, 3 completely connected layers) are shown. In the C network, the convolution filters are 1, but in the D network, they are bigger. Network D normally employs a 3-sized filter for convolution operations, therefore in addition to training the 134 million parameters, an extra 138 million parameters must be taught. Network E, commonly known as the VGG-19 network, is a deep neural network that comprises of 16 convolution layers and 3 fully linked layers throughout its 19-layer design. All VGG networks include ReLu, however local response normalization is seldom employed due to the training and memory overhead it introduces.

The Convolutional neural network used by VGG has 19 layers. A version of the network that has already been trained using the Image Net database of more than a million photos is available for loading. There are a total of 19 layers in the VGG19 model, including 16 convolution layers, 3 fully linked layers, 5 Max Pool layers, and 1 Soft Max layer.

MOBILE NET V2:

MobileNetV2 is an attempt to design a Convolutional neural network that can function effectively on mobile devices. The foundation of this design is an inverted residual structure in which the bottleneck layers serve as the links between the residual layers. It's a Convolutional neural network with 53 layers that can classify data in real-time despite the limited processing power of mobile devices like smart phones.



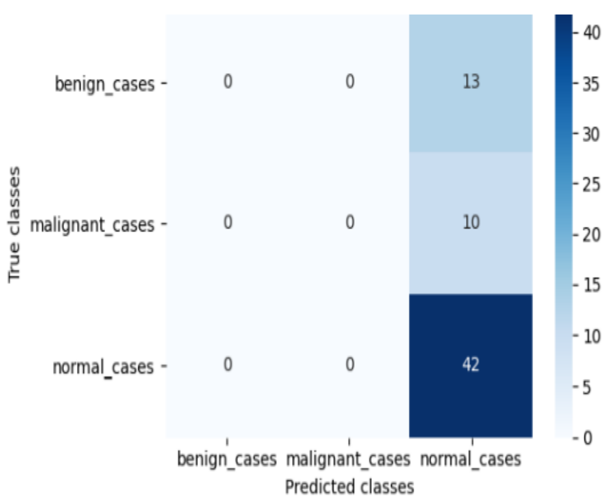
Accuracy Graph

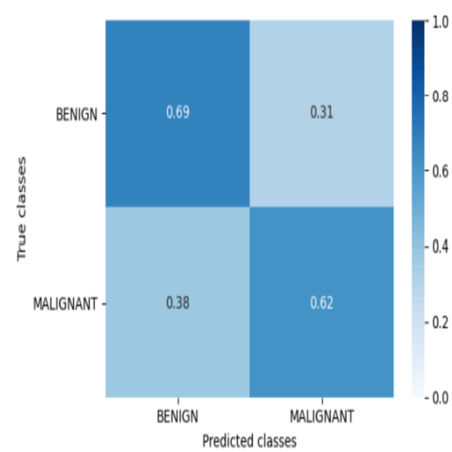
VI.RESULTS

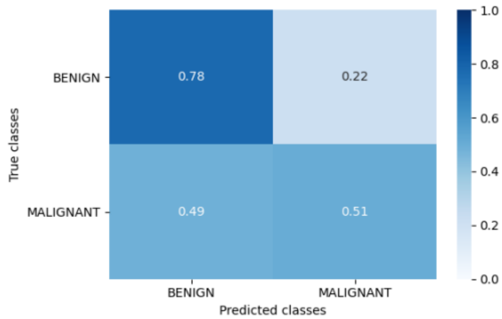
Base Model	Data Augmentation	Overall Accuracy	Precision	Recall	F1
VGG19	*None*	64.62%	41.75%	64.62%	5
	Class Balance	36.92%	48.89%	36.92%	[illegible]
	Double	64.62%	41.75%	64.62%	5
MobileNetV2	*None*	64.62%	41.75%	64.62%	5
	Class Balance	41.54%	40.70%	41.54%	[illegible]
	Double	64.62%	41.75%	64.62%	5

Base model	Class weights	Overall Accuracy	Precision	Recall	F1 Score
VGG19	*None*	62.25%	63.66%	62.25%	62.58%
	Balanced	63.96%	63.94%	63.96%	63.95%
	+50% minority	61.15%	62.16%	61.15%	61.45%
MobileNetV2	*None*	65.83%	66.33%	65.83%	66.01%
	Balanced	**67.08%**	**66.50%**	**67.08%**	**66.48%**
	+50% minority	65.05%	65.19%	65.05%	65.12%

Accuracy

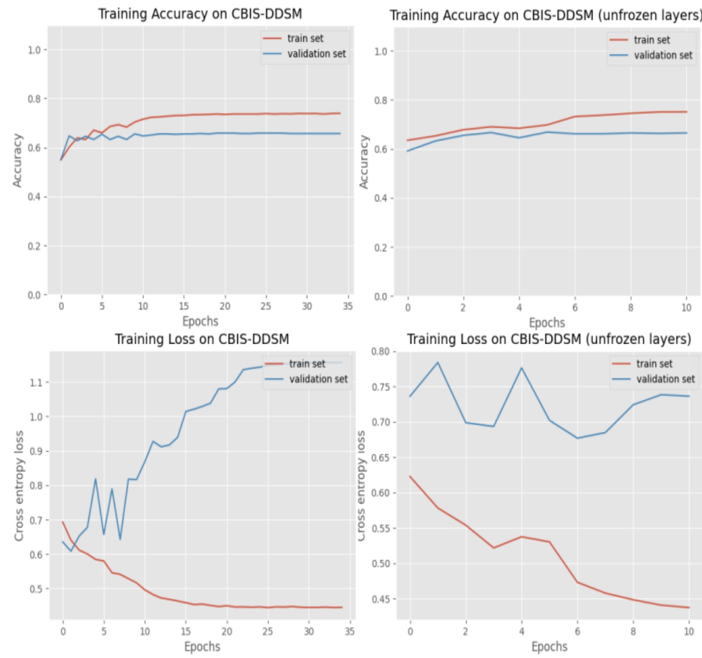






Confusion matrix

Base Model	Whole Image Size (pixels)	Extra conv/ pool layers	Overall Accuracy	Precision	Recall	F1 Sco
VGG19	*224 x 224*	No	63.96%	65.33%	63.96%	64.28
	512 x 512	No	64.59%	64.44%	64.59%	64.51
	1024 x 1024	Yes	59.28%	**66.94%**	59.28%	58.43
MobileNetV2	*224 x 224*	No	62.56%	62.38%	62.56%	62.46
	512 x 512	No	**67.08%**	66.50%	**67.08%**	**66.48**
	1024 x 1024	Yes	OOM	OOM	OOM	OOM



VII.CONCLUSION

Our key contribution is a more accurate model for identifying breast cancers in histopathology images, which was made possible by using deep learning. A pathologist will need to look at the images to establish whether or not the growths are malignant. The CNN model's layered structure allows for both cancer detection and object search through convolution, all while maintaining a manageable level of complexity. In contrast to filter expansion, network expansion occurs during dilation. A convolution procedure added to the equation simplifies the process, which in turn speeds it up. Multi-classification breast cancer detection tests, as opposed to the more frequent binary evaluations, may be used to evaluate the study's results.

VIII.FUTURE SCOPE

To improve the accuracy of our projections, we want to continue developing new models that use a mix of qualitative and quantitative methods. We will also test out other methods, comparing their efficacy as we search for the most advanced prediction model we can.

IX.REFERENCES:

[1]. Pacilè S., Lopez J., Chone P., Bertinotti T., Grouin J.M., Fillard P. Improving breast cancer detection accuracy of mammography with the concurrent use of an artificial intelligence tool. Radiol. Artif. Intell. 2020;2:e190208. doi: 10.1148/ryai.2020190208

[2]. Yari Y., Nguyen T.V., Nguyen H.T. Deep learning applied for histological diagnosis of breast cancer. IEEE Access. 2020;8:162432–162448. doi: 10.1109/ACCESS.2020.302155 7

[3]. IHME . Cancer's Deadly Toll Grows in Less Developed Countries as New Cases Increase Globally. Institute for Health Metrics and Evaluation; Seattle, WA, USA: 2016. [(accessed on 25 September 2022)]. Available

[4]. Mammary Duct Ectasia-Symptoms and Causes. Mayo Clinic; Rochester, MN, USA: 2022. [(accessed on 18 July 2022)]. Available

[5]. Suzuki T. Research on analysis of final diagnosis and prognostic factors, and development of new therapeutic drugs for malignant tumors (especially malignant pediatric tumors) Yakugaku Zasshi. 2020;140:229–271. doi: 10.1248/yakushi.19-00178.

[6]. . Spanhol F.A., Oliveira L.S., Petitjean C., Heutte L. A Dataset for Breast Cancer Histopathological Image Classification. IEEE Trans. Biomed. Eng. 2015;63:1455–1462. doi: 10.1109/TBME.2015.2496264.
7. Krizhevsky A., Sutskever I., Hinton G.E. ImageNet classification with deep convolutional neural networks. Commun. ACM. 2017;60:84–90. doi: 10.1145/3065386.

[7]. 8. Szymak P., Gasiorowski M. Using pretrained alexnet deep learning neural network for recognition of underwater objects. NAŠE MORE Znanstveni Časopis za More i Pomorstvo. 2020;67:9–13. doi: 10.17818/NM/2020/1.2. [

[8]. 9. Starke S., Leger S., Zwanenburg A., Leger K., Lohaus F., Linge A., Schreiber A., Kalinauskaite G., Tinhofer I., Guberina N., et al. 2D and 3D convolutional neural networks for outcome modelling of locally advanced head and neck squamous cell carcinoma. Sci. Rep. 2020;10:15625. doi: 10.1038/s41598-020-70542-9.

[9]. 10. Kashyap R. Breast Cancer Histopathological Image Classification Using Stochastic Dilated Residual Ghost Model. Int. J. Inf. Retr. Res. 2022;12:1–24. doi: 10.4018/IJIRR.289655.

[10]. 11. Aldhyani T.H.H., Verma A., Al-Adhaileh M.H., Koundal D. Multi-Class Skin Lesion Classification Using a Lightweight Dynamic Kernel Deep-Learning-Based Convolutional Neural Network. Diagnostics. 2022;12:2048 . doi: 10.3390/diagnostics12092048.
12. Simonyan K., Zisserman A. Very deep convolutional networks for large-scale visual recognition;

Proceedings of the International Conference on Learning Representations; Vancouver, BC, Canada. 30 April–3 May 2020. [Google Scholar]

[11]. 13. Shen L., Margolies L.R., Rothstein J.H., Fluder E., McBride R., Sieh W. Deep learning to improve breast cancer detection on screening mammography. Sci. Rep. 2019;9:12495. doi: 10.1038/s41598-019-48995-4.

[12]. 14. Koné I., Boulmane L. Lecture Notes in Computer Science. INESC TEC; Porto, Portugal: 2018. Hierarchical ResNeXt models for breast cancer histology image classification; pp. 796–803.

[13]. 15. Gardezi S.J.S., Elazab A., Lei B., Wang T. Breast cancer detection and diagnosis using mammographic data: Systematic review. J. Med. Internet Res. 2019;21:e14464. doi: 10.2196/14464.