

Enhancing Transfer Learning Model Performances using Cohen's Kappa Score at Predicting Breast Cancer

Sasanka Sekhar Dalai¹, Bharat jyoti Ranjan Sahu¹, Sashikanta Prusty^{*1}, Jyotirmayee Rautaray²

Submitted: 04/11/2023

Revised: 23/12/2023

Accepted: 05/01/2024

Abstract: Over the past twenty years, cancer has been responsible for a significant number of deaths across the globe. A recent survey was conducted on cancer deaths from 1990 to 2018, where breast cancer (BC) found as the highest death rate, making it a major concern. Thus, minimizing the death ratio can only happen when predicting with a good model at the initial stage. To address this issue, a novel technique using the deep learning-based transfer learning (TL) method has been proposed here to predict this cancer in minimal time as this model has been pre-trained with a lot of images from the ImageNet dataset. Additionally, in this study, we have taken nine different TL models for cancer classification into either benign or malignant classes using 40 epochs at the training phase and compared their performances using predefined performance metrics. Furthermore, to increase the computational time and increase the accuracy of TL models, we have implemented the 'Adam' optimizer during the training phase. The result shows that the EfficientNetB7 model outperforms with a maximum accuracy of 98%, precision of 97.45%, f1-score of 97.42%, and a recall of 97.42% than other models. However, to make better decisions regarding the performance of each model, we further evaluated Cohen's kappa score (CKS) statistical method which specifically indicates how accurately our model can identify the cancer. These experiments have been carried out using Python 3.8.3 programming software on the Jupyter 6.4.3 Notebook application in the Windows 10 operating system.

Keywords: Cancer statistics, Deep Learning, Transfer Learning, Image Classification, EfficientNetB7

1. Introduction

Cancer is one of the leading causes of death, and its prevalence has risen dramatically in the last decade. According to the World Health Organization (WHO) [1], an estimated 9.6 million people have died as a result of this disease, 30-50 percent of cases can be prevented if detected early, and it costs the economy \$1.16 trillion in 2010. With 685,000 deaths 2.3 million women were affected due to BC globally. A study [2] revealed BC cases in the United States, India, and China account for one-third of all cases worldwide. According to a study report [3], with a focus on India, studies show that 60% of breast cancer cases are discovered in advanced stages, making treatment difficult. During the Covid-19 pandemic, research findings reveal there was no significant increase in tumour size or molecular types [4], however, some study research groups [5- 6] showed a slight delay in detection. Although these figures are alarming, early detection of breast cancer results in a survival probability of over 90%, making it the most treatable form of the disease. According to data compiled from cancer treatment facilities, the COVID-19 pandemic's effects included prolonged diagnosis as well as reduced extensive care because much focus was placed on reducing

the number of COVID-19 cases in the nation and around the world [7].

Significant concerns have been made regarding how the National Health Service (NHS) cancer services would be affected by the COVID-19 pandemic preventive actions, including the suspension of national screening programs [8]. The paradigm of cancer care is still being changed by the continuing coronavirus disease-2019 (COVID-19) epidemic. With pre-existing access gaps to cancer services, the impact has been particularly devastating in low- and middle-income nations, causing significant delays and disruptions in cancer care. Excessive measures aimed at preventing the virus's spread defined the lockdown. The majority of patients were unable to travel to the cancer care centers for treatment due to the suspension of public transportation, heavy travel restrictions, and less availability of medical beds. Most likely to be harmed are those individuals with low incomes and those living in smaller towns and villages. In individuals with early-stage invasive breast cancer, it has been demonstrated that each delay in surgery of 60 days is linked to a 26% higher risk of breast cancer-related deaths [9 - 10]. Given the current situation, it is now expected that there will be more delays and disruptions in BC care, which could affect subsequent results. It is crucial to define the delays and their effects to develop systemic mitigation methods for the current and upcoming surges.

Breast cancer causes abnormal growth of the breast and, in

¹ Department of Computer Science & Engineering, Siksha 'O' Anusandhan (deemed to be University), Bhubaneswar-751030, Odisha, India

² Department of Computer Science & Engineering, OTR, Bhubaneswar-751029, Odisha, India

* Corresponding Author Email: sashi.prusty79@gmail.com

some cases, abnormal growth of the tissues (usually referred to as a tumour) surrounding the breast. Breast density refers to the proportionate number of various breast tissue types as seen on a mammogram. However, as shown in Fig. 1, fatty tissues are present around milk-producing glands and arteries and may develop lumps if the breast is harmed or damaged. Secondly, Glandular tissues generate breast milk in addition to seeming white on mammography, it is dense, resembling fibrous tissue. There is scattered fibro glandular breast tissue in about 40% of women. Thirdly, dense tissues have more fibrous and glandular tissue as compared to fatty tissues. The diagnosis of breast cancer is complicated by the thick breast tissue that is present in about 50% of women. The dense breast tissue can make finding malignancies challenging. However, dense tissues are classified into heterogeneously dense or extremely dense, resulting in the proportion of dense and non-dense tissues. Compared to fatty breast tissue, glandular tissue, and fibrous connective tissue comprise a relatively large portion of the tissue in dense breasts. Breast cancer risk is higher for women with dense breasts than for those with fatty breasts, and the risk rises with increasing breast density. It has been clinically found, that the proposed CNN identifies breast density to 98%, which is greater than the current VGG19 and ResNet50 models on 410 images [11]. Hence, in this study, improvised Transfer learning (TL) [12] is another interesting concept of machine learning technique that is being employed in more recent research studies. Thus, in this article, the pertained TL models were taken to classify the abnormal tissues into either fatty, fatty-glandular, or dense-glandular present in the WBCD dataset.

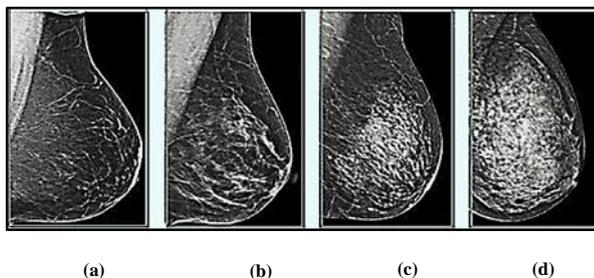


Fig. 1. Four common breast density levels are represented as (a) almost entirely fatty, (b) fibro glandular density; (c) heterogeneously dense, and (d) extremely dense

The main contributions to this study are:

- First, a critical analysis of all cancers worldwide before and during covid-19 pandemic period.
- Secondly, collect the MIAS breast cancer dataset and categorize the classes using EDA techniques.
- Then, designed a method using some common transfer learning models to predict breast cancer.
- Furthermore, an algorithm has been designed to describe the necessary steps to classify abnormal cells.

- Implement image augmentation technique to prevent each classifier from overfitting, to increase training size, and to improve the model accuracy.
- Plot training/validation accuracy/loss ROC curves for every model concerning epochs.
- Finally, design a table to represent the performance report of each model using precision, recall, F1-Score, accuracy, and Cohen kappa score.

The best part of this research follows: Section 2, describes a literature survey on breast cancer before and during the pandemic days; Section 3, reflects the materials and methods to predict malignant cells in the breast; Section 4, shows the experiments and results of all models; Section 5, discussed all the findings achieved in this research; and Section 6, concludes with conclusion and reference.

2. Related Works

This section identifies and describes research works on transfer learning for breast cancer detection. The use of transfer learning algorithms has become more common in recent years, so a survey has been done using various applications of existing transfer learning models even during the widespread of the Covid-19 pandemic. This can be done from the related studies for this paper by using the keywords "transfer learning," "Covid-19," "ML," "BC," and "deep learning." However, researchers have developed many efficient models for detecting BC in the last two decades and found BC as the most deadly disease worldwide, as shown in Fig. 2.

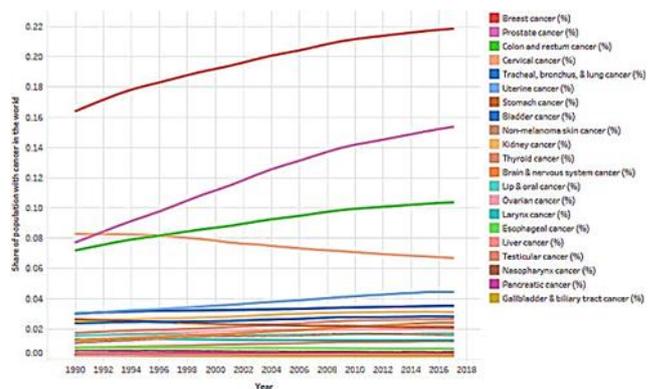


Fig. 2. World's cancer death statistics before the pandemic since the year 1990 to 2018

2.1. Before the covid-19 pandemic i.e. from the year 1990 to 2018

In the past two decades, deaths caused due to cancer have grown continuously worldwide as shown in Fig. 2, where BC sits at the top of the list with having higher death rates than others. However, so much development has already been done in the healthcare field but it still, continues to be prone to a great extent among people. Sana Ullah Khan et al [13] propose a hybrid architecture of GoogleNet, VGGNet,

and ResNet CNN to predict BC significantly. Mehra [14] experimented on the BreakHis dataset using a VGG16+LR, VGG19+LR, and ResNet50+LR model, where VGG16+LR outperforms with 92.60% of accuracy than other models. In the meantime, India recorded 162,468 new cases of breast cancer in 2018 [15], before the pandemic started.

2.2. During covid-19 pandemic i.e. from 2019 onwards

The global delivery of cancer services has been badly affected by the COVID-19 epidemic. These patients are not only thought to be at an increased risk for COVID-19-related morbidity and mortality, but they are also receiving delayed and insufficient cancer care. However, approximately 2.26 million cases are expected to be reported in 2020, BC is the most often diagnosed disease in the world and the main cause of cancer mortality among women. It poses a serious threat to global health [16]. In this comparison study, four TL techniques AlexNet, VGG16, GoogleNet, and RestNet50 are utilized to categorize brain tumours, and it was found as RestNet50 with a greater accuracy rate of 95.8% [17]. However, in the year 2020, almost 42,690 people are anticipated to pass away, including 42,690 women and 520 men [18]. This cross-sectional analysis was carried out in a North Indian tertiary care facility with the Departmental Ethics Committee's blessing. Between May 2020 and March 2021, patients who had just been diagnosed with BC and were receiving treatment at the Department of Radiotherapy and Oncology were assessed [19]. For the analysis, patients who had encountered delays in cancer diagnosis, treatment beginning at diagnosis or relapse, or interruption of ongoing treatment due to COVID-19 were included. As a result of regular screening for all patients anticipated for persistent cancer chemotherapy on July 30, 2020, after a patient scheduled for chemotherapy tested positive. ICMR advises testing the affected people who are admitted for chemotherapy or who are asymptomatic while hospitalized [20]. The global delivery of cancer services has been badly affected by the COVID-19 epidemic. These patients are not only thought to be at increased risk of morbidity and mortality [21], but they

are also receiving delayed and insufficient cancer care [22 - 23]. In contrast, both feature extraction and grey level co-occurrence matrix (GLCM) are fed into k-nearest neighbors (KNN) to obtain accurate figures from breast ultrasonography. The results found an accuracy of 80% [24].

From previous studies, it has been found that transfer learning models significantly identify the malignant cells in different organs of the human body. Additionally, BC is the major cause of death in women globally and needs proper clinical diagnosis to avoid unnecessary deaths these days. That's why, in the further section, a discussion about transfer learning models was carried out, while predicting breast cancer in women.

3. Material & Methods

However, the prevalence of breast screening, particularly in women who consistently undergo annual screening mammography, has been a significant contributor to the decrease in BC mortality through early diagnosis. Despite these initiatives, hurdles nevertheless remain in the way of people getting the care they require. The use of more effective TL techniques has helped make screening mammography more widely available recently. Thus, for this study, first, 329 breast mammogram images from the MIAS dataset have been collected; second, augmenting those images at the training stage to improve the performance; thirdly, models with input-shape = (224, 224, 3) and weight as 'imagenet'; fourth, compile the model with 100 epochs; fifth, plot loss/accuracy metrics between train set and validation set; and finally, design the confusion matrix to represent the model performance. Additionally, each model's performance on the same dataset has been compared to find the best predictive one, as shown in Fig. 3. These experiments are performed using Python 3.8.3 software on Jupyter Notebook 6.4.3 platform as discussed in a further section.

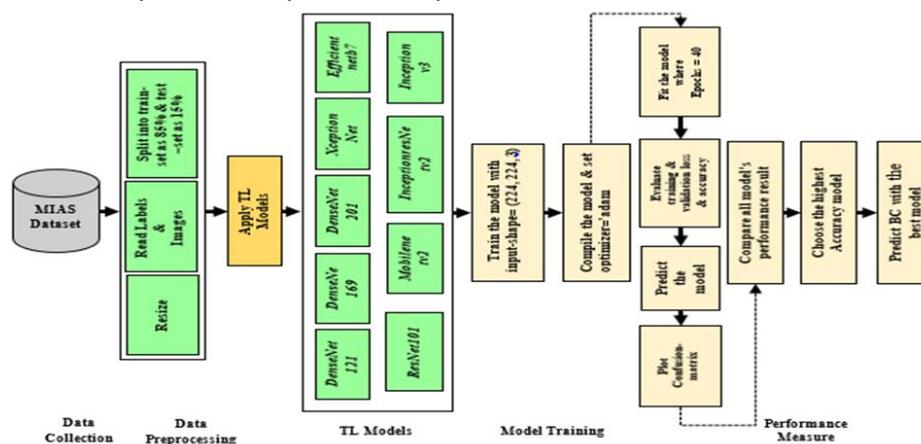


Fig. 3. A proposed methodology design to predict breast cancer on the MIAS image dataset

3.1. Material

In this study, the MIAS image dataset has been collected from the UCI machine repository which is freely available over the internet. As indicated in Table 1, this dataset contains 7 primary classes and their percentage of both benign as well as malignant tissues. Furthermore, Fig. 4 shows the BC patients that are mostly divided into 3 different types of tissues “G”, “D”, and “F”. Although, these 329 unique images are classified into seven different classes according to their abnormality.

Table 1. Classification of seven types of classes

Class	Percentage (%)	B (%)	M (%)
NORM	64.28		
CALC	07.76		
CIRC	07.15		
ARCH	05.90	55.65	44.34
SPIC	05.90		
ASYM	04.66		
MISC	04.35		

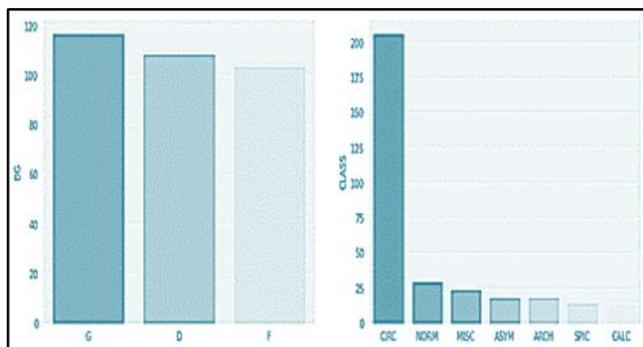


Fig. 4. Representation of three common breast tissues and classes in the MIAS dataset

3.2. Data Pre-processing

This is an initial step to make the raw data usable as they are so complex to handle. Thus, cleaning the dataset is necessary to eliminate redundant, null, or irrelevant data. Although, helps to find useful insights from the huge volume of nonlinear complex data found in medical fields these days. Additionally, the data augmentation technique enlarges the training set by adding the number of copies of data to improve the model's performance. Also, increases the diversity of images by generating the transformed images from the inputs. Moreover, the implementation of this image augmentation technique is necessary to improve the size and the number of images to make a model robust. The `train_test_split` in `sklearn.model_selection` divided the MIAS dataset into 85% as `train_set` and 15% `test_set`, before

fed into TL model.

3.3. Method

The transfer learning process applies specific elements of an existing machine learning model to solve a new but familiar problem, by transferring knowledge that can be applied to a new model in different scenarios or conditions. The idea is to transfer a previously trained task with a generalized view to a related new task, saving the learning algorithm from having to start from scratch. Because the results of the task on BC classification will have an impact on human lives, there is very little room for error. However, twelve deep CNN models were used to address the classification problem, where EfficientNet, InceptionResNetV2, ResNet, DenseNet, MobileNet, and VGG19 produced an accuracy of 99.96%, 99.95%, 99.95%, 99.94%, 99.92%, and 99.89% [25]. Additionally, each model's performance in detail has been described as follows in Fig. 6 to Fig. 14.

DenseNet is a type of CNN where each layer connects with every other layer below it, as given in Fig. 5. The main reason behind this is to maximize the information flow between network tiers. The transfer learning methods like DenseNet121, DenseNet169, and DenseNet201 have all been used in this literature. DenseNet-121, DenseNet-169, and DenseNet-201 have [6, 12, 24, 16], [6, 12, 32, 32], and [6, 12, 48, 32] layers in the four dense blocks respectively. This study compares the techniques mentioned above and displays the experimental results and these methods outperform traditional learning methods. In addition, for x_0, x_1, \dots, x_{i-1} inputs, the feature map on the i th layer is represented using non-linear transformation (H_i) as:

$$x_i = H_i([x_0, x_1, \dots, x_{i-1}]) \quad (1)$$

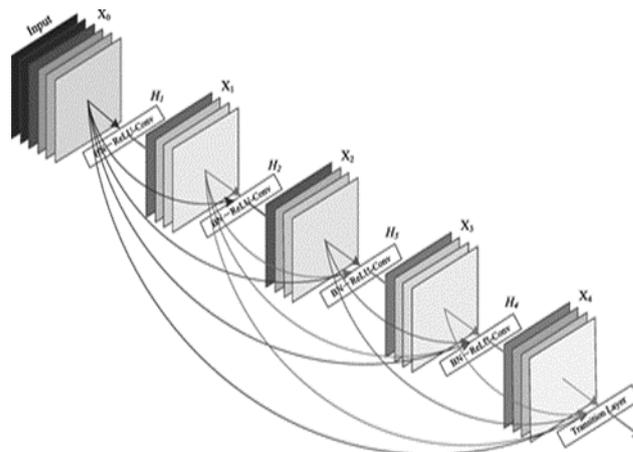


Fig. 5. Representation of dense blocks where layer = 5

3.3.1. DenseNets

This architecture is comprised of 58 blocks and 121 learnable layers.

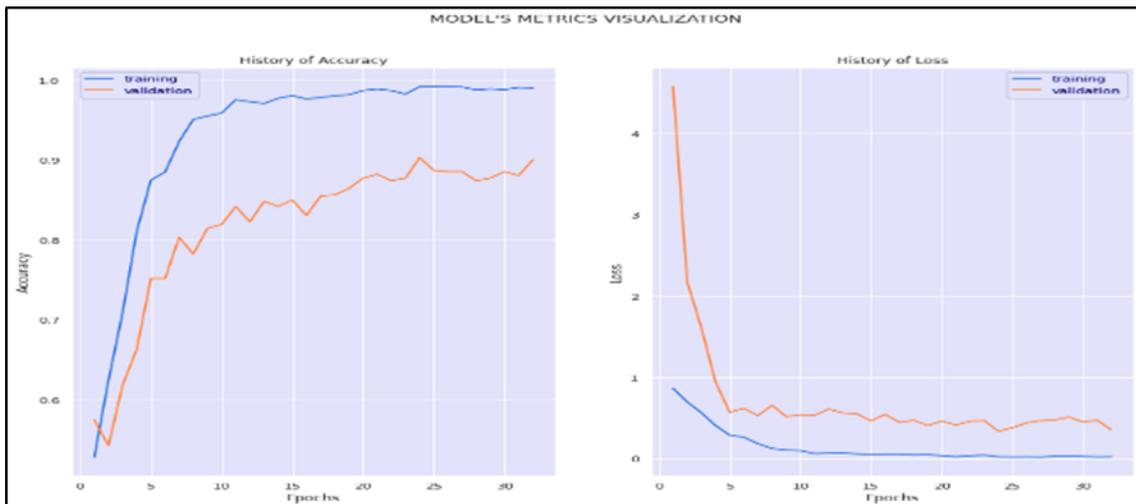
- The first convolution layer is made up of five dense

layers of $(6+12+24+16)*2 = 116$ blocks each, 3 transition layers, and 16 dense layers at the output end. It promises to provide greater accuracy even with fewer constraints and has demonstrated a higher level of accuracy than most learning techniques.

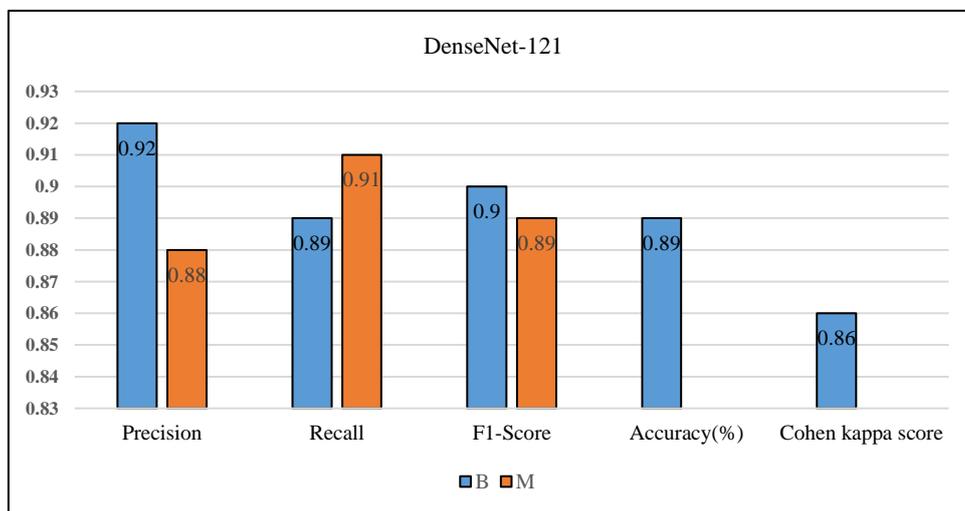
- DenseNet169 is a densely connected network that groups all of its convolution layers into a dense layer block. Its connectivity pattern sets it apart from other networks that have fewer parameters than other learning models. As compared to densenet121, densenet169 is larger (i.e. 55Mb, where 31Mb in the

case of densenet121).

- In addition to having better parameter efficiency, DenseNet also has improved information flow and gradients throughout the network, making them simple to train and reducing overfitting issues.
- The potential for feature reuse by different layers creates an improvement in the performance of the model. DenseNet201 exploits the condensed network that is simple to train and highly efficient as compared to densenet121 and densenet169.

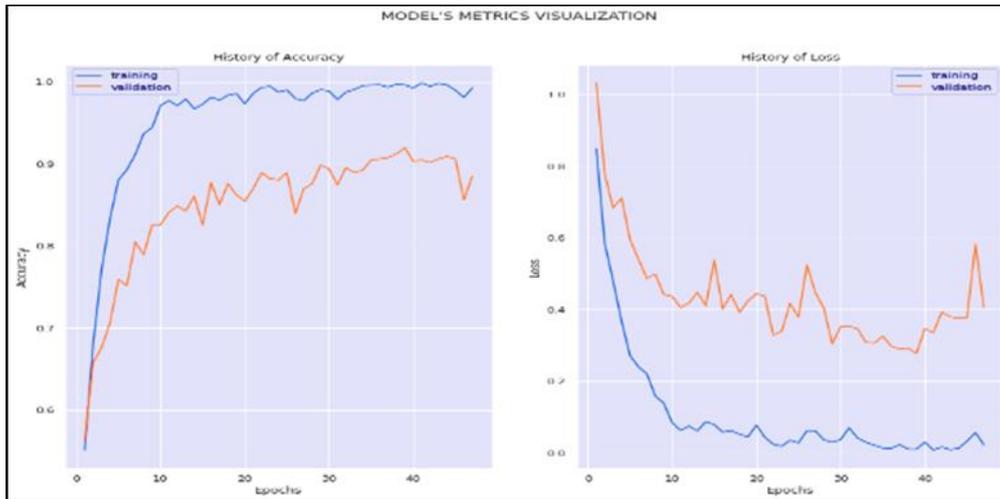


(a)

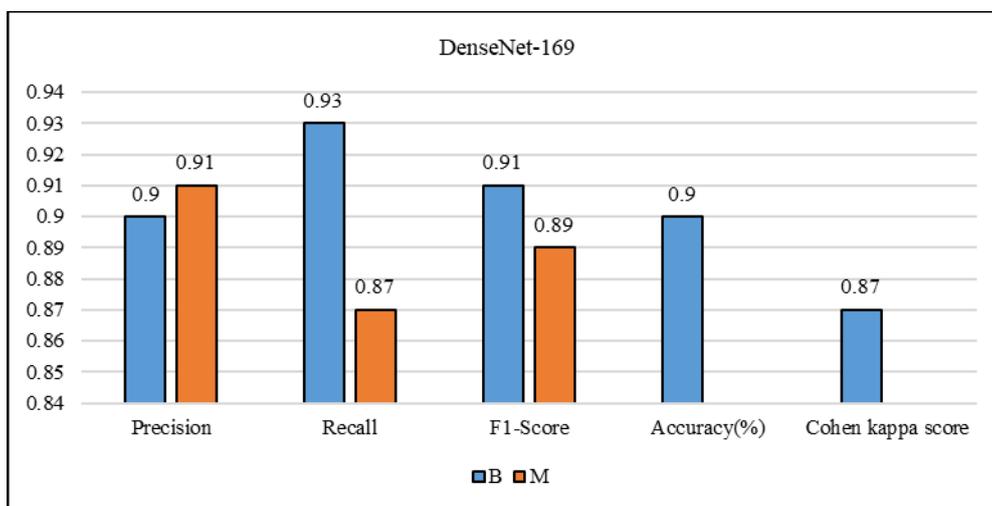


(b)

Fig. 6. Performance evaluation of the DenseNet-121 model using (a) training/validation loss/accuracy, and (b) metrics on the MIAS image dataset

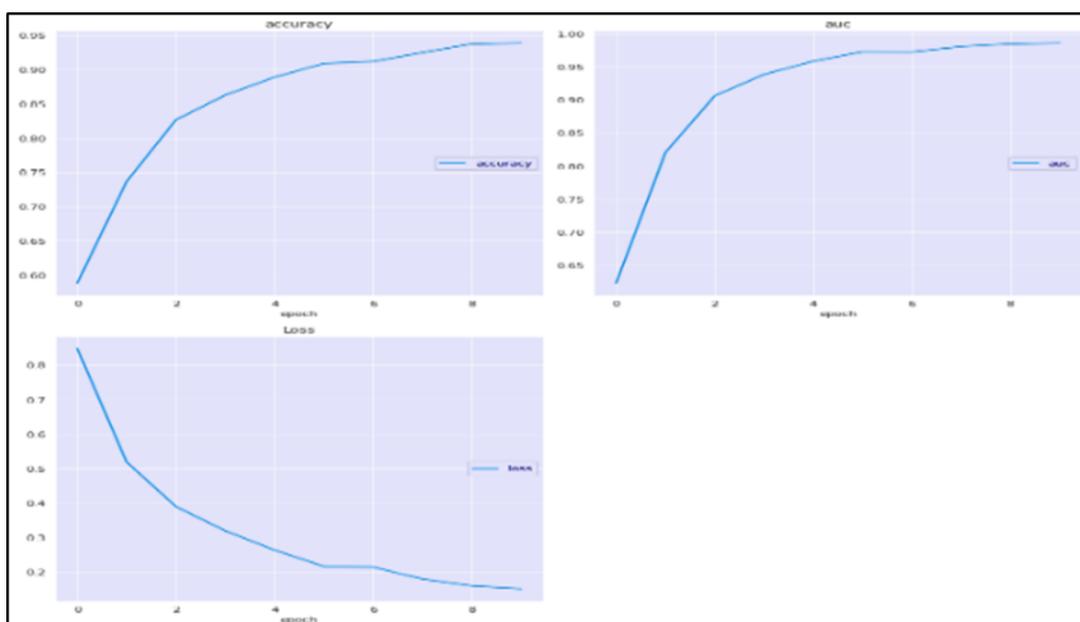


(a)

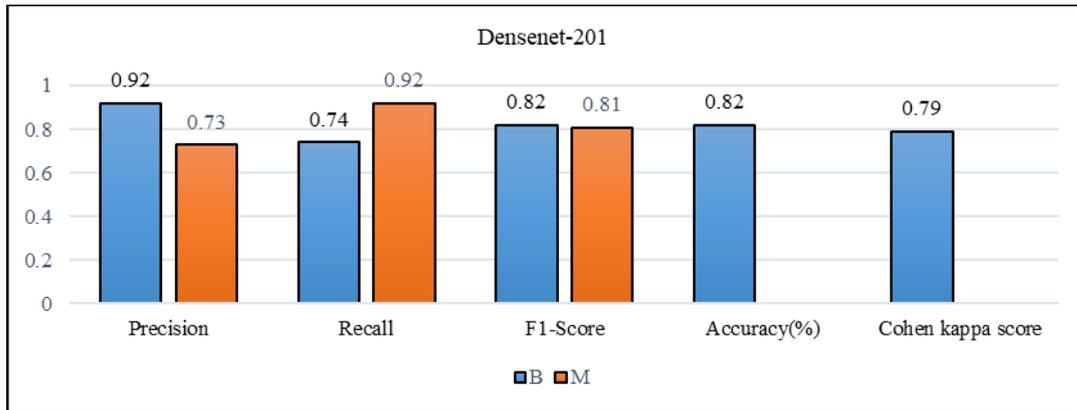


(b)

Fig. 7. Performance evaluation of the DenseNet-169 model using (a) training/validation loss/accuracy, and (b) metrics on the MIAS image dataset



(a)



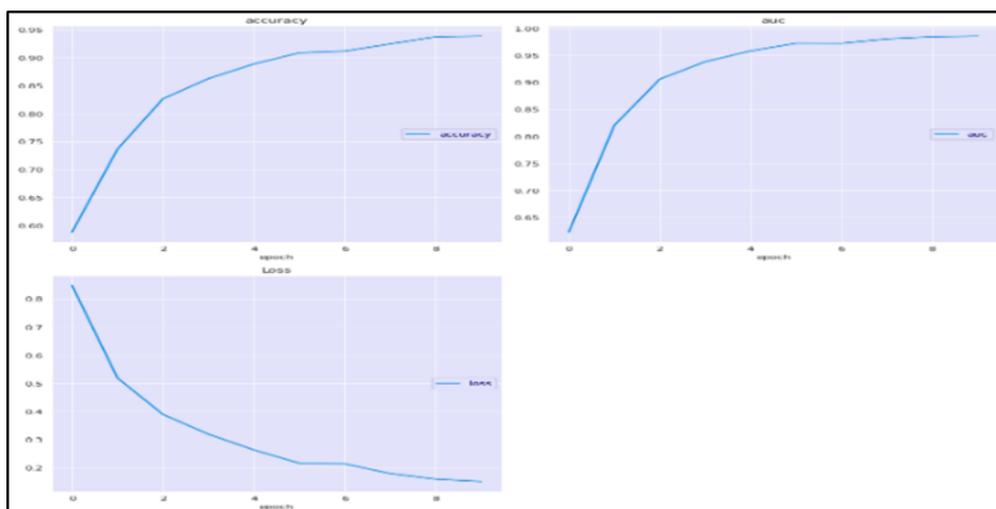
(b)

Fig. 8. Performance evaluation of the DenseNet-201 model using (a) training/validation loss/accuracy, and (b) metrics

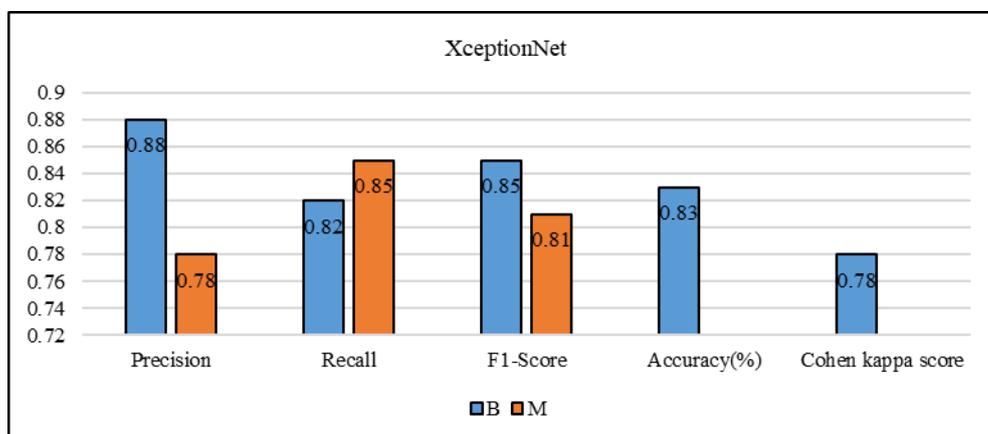
3.3.2. XceptionNet

This is a deep convolutional neural network design based on depth-wise separable convolutions, extracting features from the network using its 36 convolutional layers, organized into

14 modules. Except for the first and last modules, the layers of the networks have linear residual connections surrounding the modules. However, a fine-tuned XceptionNet can extract features from ultrasound tumour images providing improved classification accuracy.



(a)



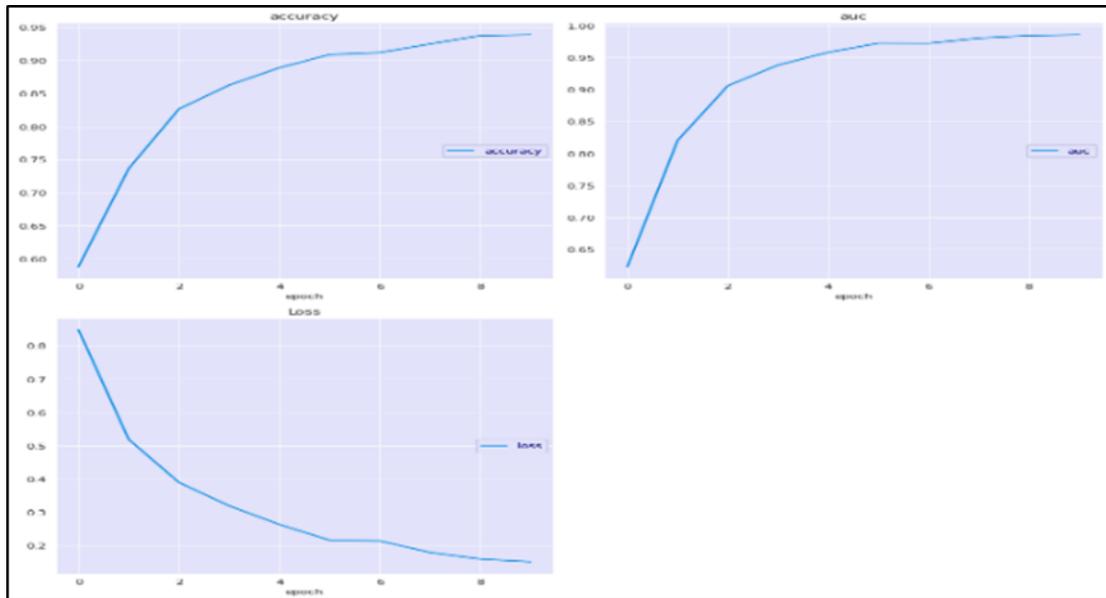
(b)

Fig. 9. Performance evaluation of the XceptionNet model using (a) training/validation loss/accuracy, and (b) metrics on the MIAS image dataset

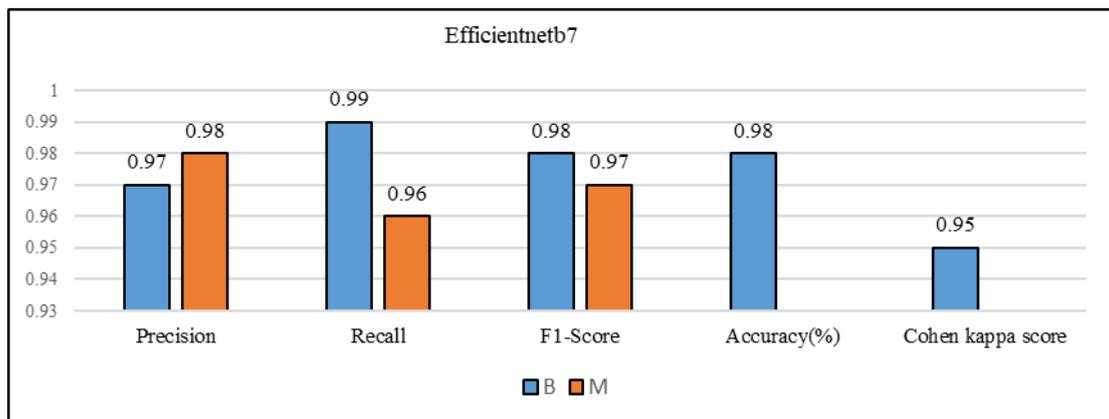
3.3.3. EfficientNetB7

This is an architectural and scaling technique in CNNs network that uses a compound coefficient to scale the dimension of depth, breadth, and resolution effectively [26]. This compound scaling technique increases accuracy and

efficiency even more in comparison to existing ConvNets like MobileNets and ResNet. As the name suggests, this model comprises 7 blocks (B1...B7), and 813 individual layers. After, successfully evaluating this model on the MIAS dataset, it found 98% of accuracy and 95% of Cohen kappa scores, which means nearly 100%.



(a)



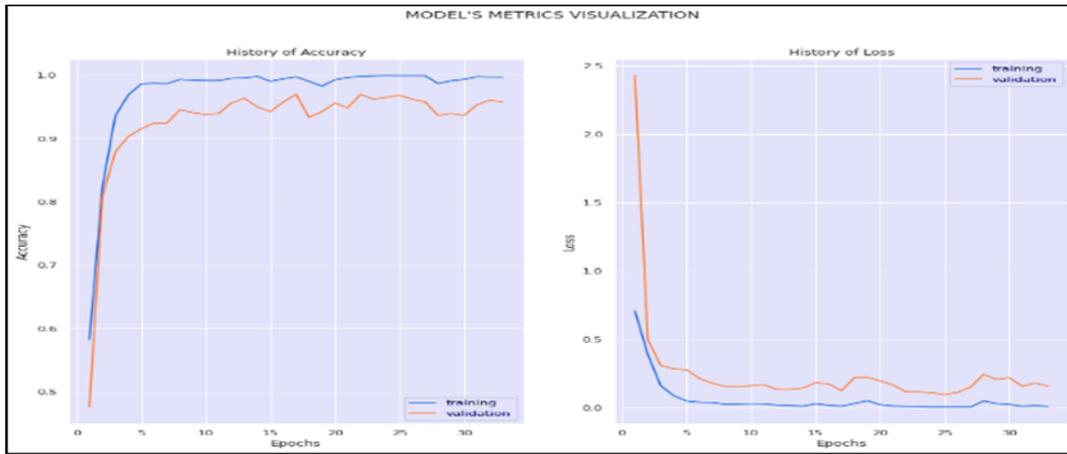
(b)

Fig. 10. Performance evaluation of the EfficientNetB7 model using (a) training/validation loss/accuracy, and (b) metrics on the MIAS image dataset

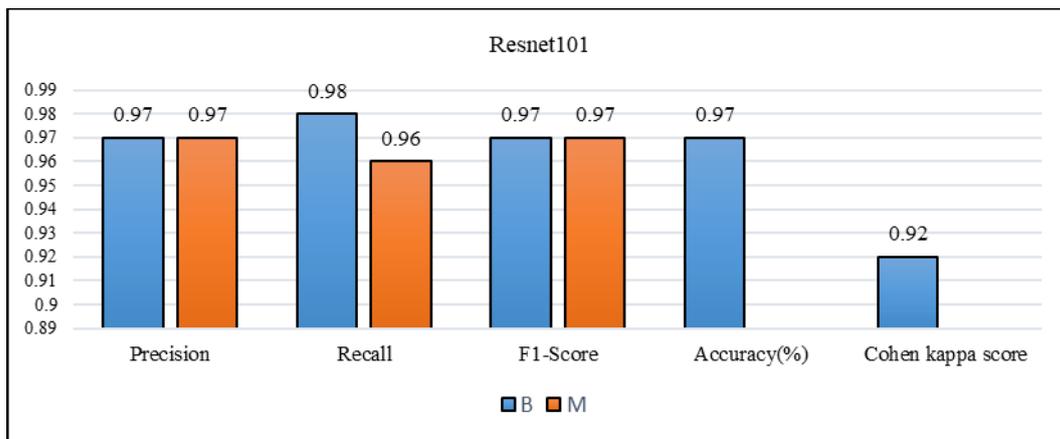
3.3.4. ResNet101

This model includes pathways between layers to stop distortion in complex networks and bottleneck plugs to facilitate speedy training for the model. However, ResNet101 has significantly improved the diagnosis outcomes than those of earlier networks. As its name

implies, this network can classify images into 1000 different object categories. Although, this network has developed rich feature representations for a variety of images, and accepts images with an input size of 224*224*3. Finally, this model resulted in 97% accuracy and 92% Cohen kappa scores, which is the second highest after the previously performed EfficientNetB7 model.



(a)



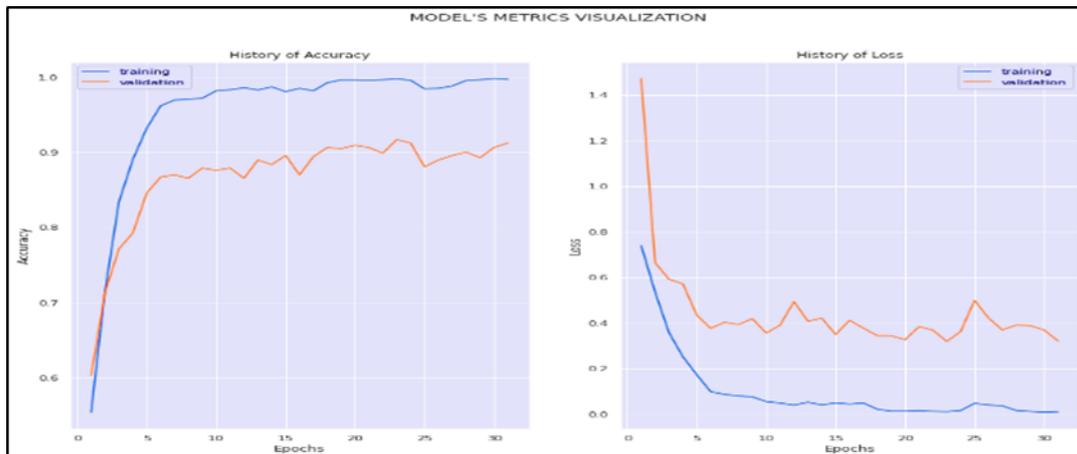
(b)

Fig. 11. Performance evaluation of the Resnet101 model using (a) training/validation loss/accuracy, and (b) metrics on the MIAS image dataset

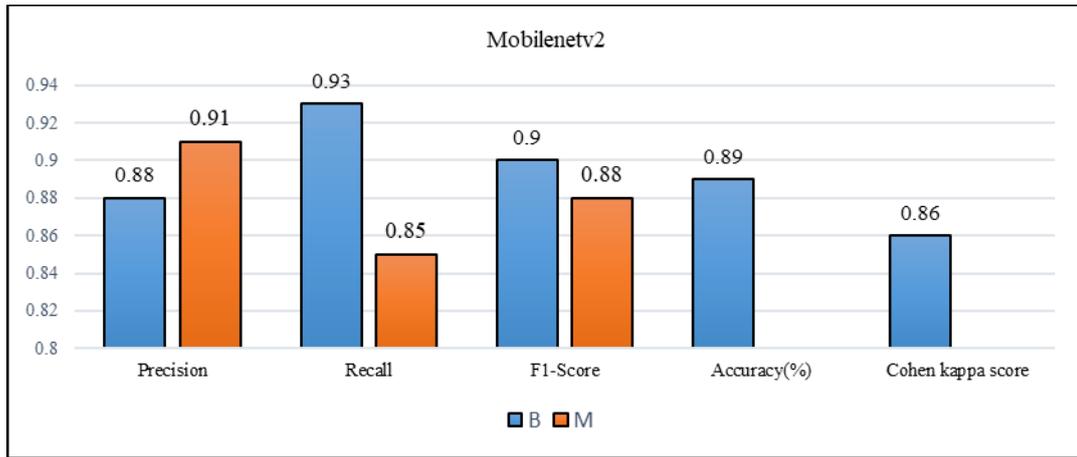
3.3.5. MobileNetv2

These are based on a streamlined architecture containing 53 layers deep and use depth-wise separable convolutions to build modular deep neural networks. Additionally, the

hyper-parameters allow this model to be built based on the limitations of the task that effectively trade off both latency and accuracy. After successfully evaluating this model on the MIAS dataset, it produced an accuracy score of 89% and a Cohen kappa score of 86% as shown in Fig. 12.



(a)



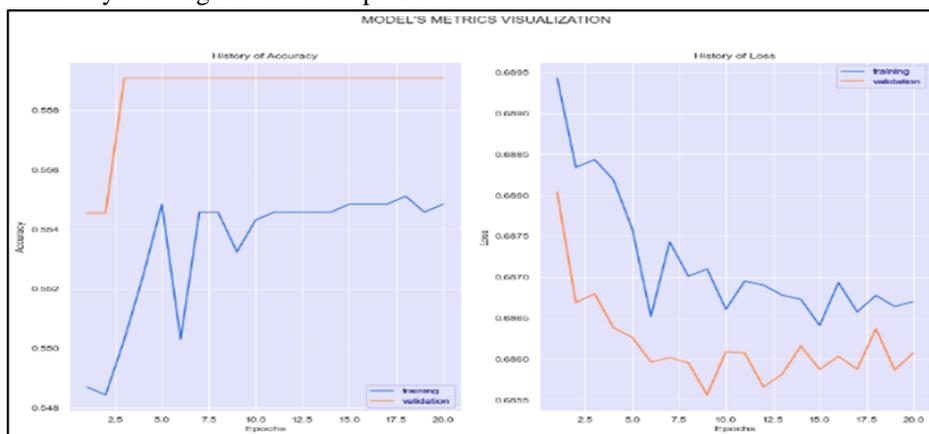
(b)

Fig. 12. Performance evaluation of Mobilenetv2 model using (a) training/validation loss/accuracy, and (b) metrics on MIAS image dataset

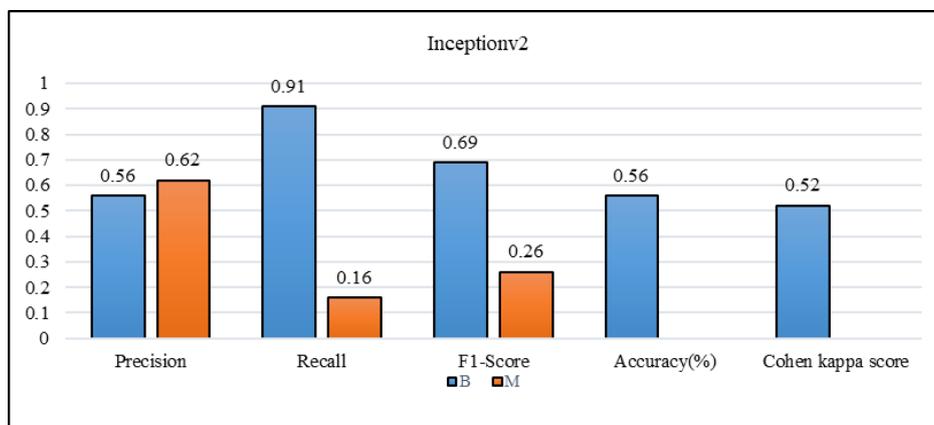
3.3.6. InceptionresNetv2

This contains 164 layers and is trained with millions of images from ImageNet. This network carries images of input size 299*299 and has learned rich feature representations for a variety of images. A filter expansion

layer is placed after each Inception block to increase the dimensionality of the filter bank. This method has been implemented successfully on the MIAS image dataset and resulted in an average score of 56% as illustrated below in Fig. 13.



(a)



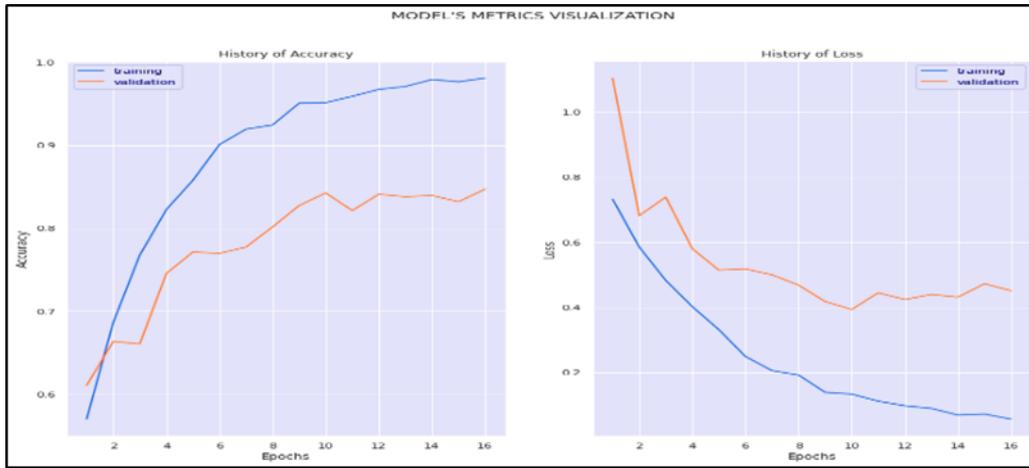
(b)

Fig. 13. Performance evaluation of the Inceptionv2 model using (a) training/validation loss/accuracy, and (b) metrics on the MIAS image dataset

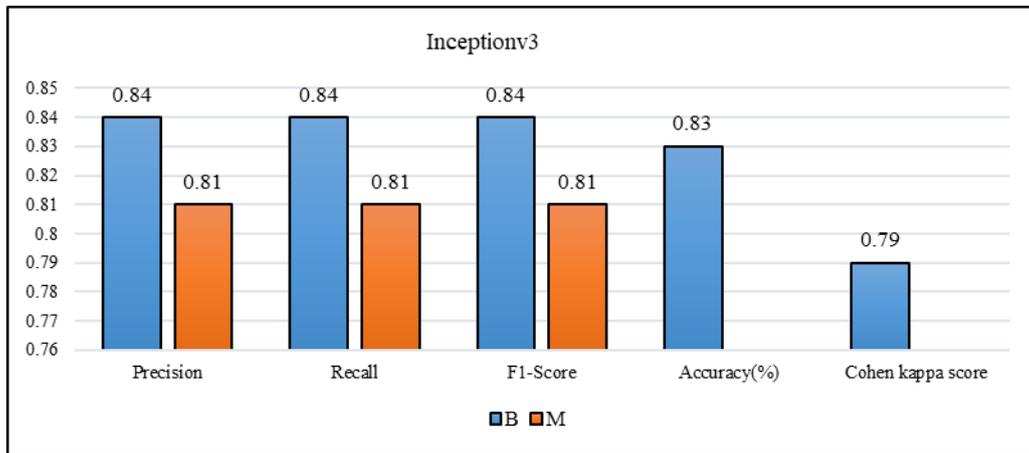
3.3.7. Inceptionv3

This is a type of CNN model that contains 48 layers and is trained with millions of data from ImageNet. Building blocks that are both symmetric and asymmetric make up the model itself. The model makes considerable use of batch

normalization and is used across a variety of medical areas. According to Zhang [27], Inceptionv3 has improved the performance of DL models significantly in past years. Thus, in this study, the Inceptionv3 model has been taken to predict the disease and achieve 80% accuracy.



(a)



(b)

Fig. 14. Performance evaluation of the Inceptionv3 model using (a) training/validation loss/accuracy, and (b) metrics

4. Results

Making a better decision for an ML model on a given dataset can be measured using performance metrics means how well a model classifies the disease using certain parameters. Additionally, a performance report has been generated, which contains accuracy as ‘*auc*’, precision as ‘*p*’, recall as ‘*r*’, and f1-score as ‘*f1s*’ on the test set. These values are being calculated based on t_p as a true positive, t_n as a true negative, f_p as a false positive, and f_n as a false negative.

$$p = \frac{t_p}{t_p + f_p} \quad (2)$$

$$r = \frac{t_p}{t_p + f_n} \quad (3)$$

$$auc = \frac{t_p + t_n}{t_p + f_p + t_n + f_n} \quad (4)$$

$$f1s = 2 * \frac{p * r}{p + r} \quad (5)$$

4.1. Statistical Analysis using CKS

However, for this multi-class classification issue as explained in the preceding subsection, metrics like auc, p/r, or f1s fail to provide each classifier success. Nonetheless, might encounter an issue with unbalanced classes. In the meantime, CKS is a good indicator that can measure the performance of ML models based on actual and predicted (i.e. rater 2) values.

$$p_{actual} = \frac{t_p + t_n}{N} \quad (6)$$

Where, N = total images in the MIAS dataset (i.e. 329).

$$p_{predicted} = \left\{ \frac{(t_p + f_n)}{N} * \frac{(t_p + f_p)}{N} + \frac{(f_n + t_n)}{N} * \frac{(f_p + t_n)}{N} \right\} \quad (7)$$

$$CKS = \frac{p_{actual} - p_{predicted}}{1 - p_{predicted}} \quad (8)$$

Table 2, comprises of classification report for all nine models, where it found Efficientnetb7 outperforms than other eight models with 98% accuracy and a Cohen kappa

score of 95%, specifying perfect agreement between two raters. Besides that, the proposed method successfully classified both benign and malignant tumours belonging to their respective classes, as shown in Fig. 15.

Table 2. Performance analysis on MIAS image dataset using transfer learning models

Model	Class	Precision	Recall	F1s	Auc (%)	CKS
DenseNet-121	B	0.90	0.93	0.91	90.34	87.35
	M	0.91	0.87	0.89		
DenseNet-169	B	0.92	0.89	0.90	89.83	85.95
	M	0.88	0.91	0.89		
Densenet-201	B	0.92	0.74	0.82	81.97	79.23
	M	0.73	0.92	0.81		
XceptionNet	B	0.88	0.82	0.85	83.05	78.05
	M	0.78	0.85	0.81		
Resnet101	B	0.97	0.98	0.97	96.91	92.25
	M	0.97	0.96	0.97		
Mobilenetv2	B	0.88	0.93	0.90	89.19	86.45
	M	0.91	0.85	0.88		
Inceptionv2	B	0.56	0.91	0.69	56.76	52.39
	M	0.62	0.16	0.26		
Inceptionv3	B	0.84	0.84	0.84	82.63	79.10
	M	0.81	0.81	0.81		
Efficientnetb7	B	0.97	0.99	0.98	98.00	95.00
	M	0.98	0.96	0.97		

5. Discussion

The most prevalent and dangerous cancer, breast cancer, drastically reduces lifetime. The automatic classification of malignant tumours is a difficult task. However, COVID-19 pandemics are continuing to be prone globally. In the meantime, patients were unable to obtain cancer services during the first wave, and surprisingly massive delays and interruptions in breast cancer diagnosis and treatment. The pandemic's logistical challenges accounted for the majority of the delays. Fig. 2, shows the cancer statistics from the year 1990 to 2018 worldwide, where BC seems to be more number of people were affected by this disease. Also, in the literature studies, it has been seen how crucial it is to streamline cancer services during pandemics and guarantee that patients who are anticipated to suffer the most due to delays continue to receive treatment in past years. This specifies BC had significantly affected a huge number of people in the world. Thus, in this study, the authors have mainly focused on breast cancer and tried to develop a

methodology to predict BC diseases effectively.

The key terms of this research are discussed as follows:

- Fig. 3, describes every step from scratch to end at classifying BC using a proposed diagram with the help of nine different TL models. Furthermore, an algorithm has been designed in the material and method section that describes all the steps followed to find abnormal cells in breast mammograms.
- Fig. 4, represents the structural view of the MIAS dataset (i.e. 329 patient data) and seven classes based on the abnormality issues.
- Figs. 6 to 14, shows each model's training/validation loss/accuracy result using a roc plot and the classification report based on performance metrics.
- Besides that, Table 2 shows the classification results of all nine models using different metrics such as precision, recall, f1-score, accuracy, and Cohen kappa

statistics.

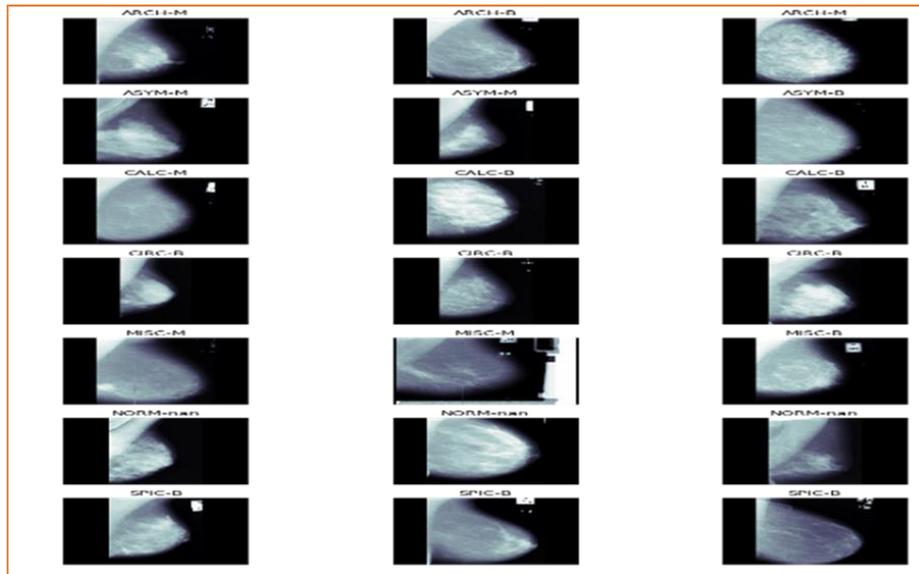


Fig. 15. The outcome of breast images and their respective classes after the prediction

From the above experiments, it can be said that all models had performed well with significant results but Efficientnetb7 outperformed with approximately 98% accuracy and 95% Cohen kappa score specifying that perfect agreement between both raters.

6. Conclusion

However, from the above analysis, it can be concluded that cancer will be cured if detected and prevented at early stages. However, in the past two years, covid-19 pandemic has recently violated the cancer screening program and resulted in a maximum number of deaths globally. Thus, timely awareness and proper diagnosis techniques might help to minimize uncertain cancer deaths. In contrast to this, an accurate and efficient technique is required in medical firms that not only cures the patient completely but also identifies the disease at the initial stages. In this regard, the MIAS dataset has been collected and implemented with nine different TL models to identify malignant cells from the breast mammograms and also compare their performances to find the best predictive model. These experimental findings and results are shown in Table 2. This demonstrates that the Efficientnetb7 model outperforms with 98% accuracy and 95% Cohen kappa score. This implies a baseline classifier for breast cancer research and might help other researchers as well as doctors in diagnosing other cancers shortly. A complete clinical evaluation is essential since BC can disclose itself in a wide range of ways in the future.

Author Contributions

Sasanka Sekhar Dalai: Conceptualization, Methodology, Software Validation, **Bharat jyoti Ranjan Sahu:** Formal Analysis, Writing—Review and Editing, Supervision

Sashikanta Prusty: Writing—Original Draft Preparation, Investigation, Resources, Editing, **Jyotirmayee Rautaray:** Formal Analysis, Data Curation.

Conflicts of Interest

The authors declare no conflicts of interest.

References

- [1] <https://www.who.int/news-room/fact-sheets/detail/breast-cancer>
- [2] S. Suhani et al., “An audit of over 1000 breast cancer patients from a tertiary care center of Northern India,” *PubMed Breast Disease*, Vol. 39(2), pp 91-99, 2020. DOI: 10.3233/BD-190435, PMID: 32310154.
- [3] S. Prusty et al., “SEMeL-LR: An improvised modeling approach using a meta-learning algorithm to classify breast cancer,” *Engineering Applications of Artificial Intelligence*, 129, 107630, 2024.
- [4] A. S. İlgün and V. Özmen, “The Impact of the COVID-19 Pandemic on Breast Cancer Patients,” *European Journal Breast Health*, Vol. 18(1), pp 85-90, 2021. DOI: 10.4274/ejbh.galenos.2021.2021-11-5, PMID: 35059596, PMCID: PMC8734528.
- [5] G. Vanni et al., “Delay in Breast Cancer Treatments During the First COVID-19 Lockdown. A Multicentric Analysis of 432 Patients,” *Anticancer Research*, Vol. 40(12), pp 7119-7125, 2020. DOI: 10.21873/anticanres.14741. Epub 2020 Dec 7. PMID: 33288611.
- [6] V. Ozmen et al., “Factors affecting breast cancer treatment delay in Turkey: a study from Turkish Federation of Breast Diseases Societies,” *European*

- Journal Public Health, Vol. 25(1), pp 9-14, 2015. DOI: 10.1093/eurpub/cku086, PMID: 25096257, PMCID: PMC4304375.
- [7] M. D A. Bashir and N. Begam, "Impact of COVID-19 Pandemic on Cancer Screening in India: Current Situation, Challenges and Way Forwards," *Asian Pacific Journal of Cancer Care*, Vol. 6(1), pp 145-150, 2021. DOI: 10.31557/apjcc.2021.6.S1.
- [8] BJS Commission Team. BJS commission on surgery and perioperative care post-COVID-19. *Br J Surg* 2021;108:1162–1180.
- [9] R. J Bleicher et al., "Time to surgery and breast cancer survival in the United States," *JAMA Oncol* 2016; 2:330-9. Back to cited text no. 4.
- [10] M Mateo et al., "Time to surgery and the impact of delay on triple negative breast cancers and other phenotypes," *J Clin Oncol* 2018; 36 15 Suppl: e12606.
- [11] S. Prusty, S. Patnaik, S. K. Dash, "SKCV: Stratified K-fold cross-validation on ML classifiers for predicting cervical cancer," *Frontiers in Nanotechnology*, 2022, 4, 972421.
- [12] W. Karl, M. Taghi and W. DingDing, "A survey of transfer learning," *Journal of Big Data*, Vol. 3(9), 2016, DOI: 10.1186/s40537-016-0043-6.
- [13] S. Khan, N. Islam, Z. Jan, I. U. Din, and J. P. C Rodrigues, "A novel deep learning based framework for the detection and classification of breast cancer using transfer learning," *Pattern Recognition Letters*, Vol. 125, pp 1-6, ISSN 0167-8655, 2019. DOI: 10.1016/j.patrec.2019.03.022.
- [14] R. Mehra, "Breast cancer histology images classification: Training from scratch or transfer learning?," *ICT Express*, 4(4), 247-254, 2018.
- [15] W. Nawaz, S. Ahmed, M. Tahir and H. Khan, "Classification of Breast Cancer Histology Images Using ALEXNET," *Image Analysis and Recognition*, pp 869-876, 2020. DOI: 10.1007/978-3-319-93000-8_99.
- [16] S. Prusty, P. Das, S. K. Dash, S. Patnaik, and S. G. P. Prusty, "Prediction of Breast cancer using integrated machine learning-fuzzy and dimension reduction techniques," *Journal of Intelligent & Fuzzy Systems*, vol. 45, no. 1, pp. 1633-1652, 2023. DOI: 10.3233/JIFS-223265
- [17] G. S Bhattacharyya, D. C Doval, C. J Desai, H. Chaturvedi, S. Sharma, and S. P. Somashekhar, "Overview of breast cancer and implications of overtreatment of early-stage breast cancer: An Indian perspective," *JCO Glob Oncol*; 6:789-98, 2020.
- [18] American Cancer Society, *Cancer Facts & Figures 2020*, no. 4, American Cancer Society, Atlanta, 2020.
- [19] H. Sung et al., "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA Cancer J Clin* 2021; 71: 209–49.
- [20] A. Gupta, B. S. Yadav, V. K. V. Gade, D. Rana, D. Bharti, and D. Dahiya, "Delays and disruptions in breast cancer care due to coronavirus disease-2019 pandemic: Real-World data from India," *Annals of Oncology Research and Therapy*, 2022, 2(1), 23.
- [21] Indian Council of Medical Research. *Advisory on Strategy for COVID-19 Testing in India (Version VI)*, dated 4th September 2020. Accessed February 04, 2021, at:
- [22] https://www.icmr.gov.in/pdf/covid/strategy/Testing_Strategy_v6_04092020.pdf
- [23] S. Prusty, S. K. Dash, S. Patnaik, et al., "EPD: an integrated modeling technique to classify BC," In *2023 International Conference in Advances in Power, Signal, and Information Technology (APSIT)* (pp. 651-655). IEEE.
- [24] S. Prusty, S. K. Dash, and S. Patnaik, "A novel transfer learning technique for detecting breast cancer mammograms using VGG16 bottleneck feature," *ECS Transactions*, 107(1), 733, 2022.
- [25] K. Djunaidi, H. B. Agtriadi, D. Kuswardani and Y. S. Purwanto, "Gray level co-occurrence matrix feature extraction and histogram in breast cancer classification with ultrasonographic imagery," *Indones. J. Electr. Eng. Comput. Sci*, 22(2), 187-192, 2020.
- [26] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," In *International conference on machine learning* (pp. 6105-6114), 2020, PMLR. <https://arxiv.org/pdf/1905.11946v5.pdf>.
- [27] S. Prusty, S. Patnaik and S. K. Dash, "EfficientNetB7: To predict lung and colon cancer at early stages," In *AIP Conference Proceedings (Vol. 2819, No. 1)*. AIP Publishing.
- [28] P. Zhang, Z. Ma, Y. Zhang, X. Chen, and G. Wang, "Improved Inception V3 method and its effect on radiologists' performance of tumor classification with automated breast ultrasound system," *Gland Surgery*, 10(7), 2232, 2021, doi: 10.21037/gS-21-328.