Vol. 06, No. 4 (2024) 1879-1888, doi: 10.24874/PES.SI.25.03A.007

Proceedings on Engineering Sciences

www.pesjournal.net

A COMPARATIVE ANALYSIS OF DEEP TRANSFER LEARNING TECHNIQUES FOR MAMMOGRAPHIC IMAGE CLASSIFICATION

Bhavesh Gupta[1]
Akshay Singh
Anjana Gosain

Received 07.01.2024.
Received in revised form 31.01.2024.
Accepted 15.03.2024.
UDC – 004.032.26

Keywords:

Deep Transfer Learning, Breast Cancer, Mammography Image Classification, Convolution Neural Networks

A B S T R A C T

Among all new cancer cases diagnosed, breast cancer has been leading in count, followed by prostate and lung cancer. Breast cancer also has the highest chances of getting cured, if it gets early diagnosis, thus increasing the lives of not only women but also the minority of males. For the same, the Deep Learning algorithms with transfer learning models are utilized, already trained with ImageNet database, and partially training them on the small mammography images database and thus help to diagnose it without the need for large datasets or tissue analysis (biopsy). The pre-trained convolution neural network models of VGG-16, VGG-19, ResNet50 and Inception V3 are worked as Deep Transfer Learning on two databases: the Mammography Image Analysis Society (MIAS) database containing 321 images, and Chinese Mammography Database (CMMD) containing 3744 mammography, of which 2000 images are used for learning. The evaluation of the model is based upon the parameters of accuracy, precision, recall, and F1-score. For MIAS Database, VGG 19 model showed better results, with accuracy being 98.44%, and precision, recall and F1 score being 0.99 each. For CMMD, VGG16 showed better results, with accuracy being 99.50%, precision being 1.0, recall being 0.99, and F1 score of 0.99.

1. INTRODUCTION

1.1 Deep Transfer Learning

In machine learning, computers are programmed to gain knowledge from data, later processing the inputs and generating outputs based on acquired knowledge. In simpler terms, it's like teaching a computer to learn and compute desired outputs on its own. However, this requires data to be labelled and programs to be programmed on aspects that influence the output or not. This is handy when there is a need for certain aspects to have more influence over others, but becomes a challenge when it comes to large datasets. Deep learning helps here as they are designed to remove the task of explicit labelling of data. In deep learning, the program is able to not only absorb the data but also judge which components should play a pivotal role in decision-making (Plested & Gedeon, 2022). This approach becomes invaluable when handling extensive and complex datasets, such as ones used in medical

[1] Corresponding author: Bhavesh Gupta
Email: bhaveshgupta01@gmail.com

applications like breast cancer detection. In spite of the ease of programming, machine learning and deep learning require a large size of data for training and learning trends. In the medical domain, there is a deficiency of such large sets of data because of many reasons (discussed further). So help with this, we use Transfer Deep Learning, a technique that leverages already existing large datasets, like ImageNet for image processing. Here, models are trained on these large datasets and later fine-tune only specific layers on our smaller, more specialized datasets, thus yielding better results, as shown in Figure 1. In breast cancer detection also, transfer deep learning helps with limitations of limited datasets, thus ensuring that our breast cancer detection model not only learns from the available data but does so with a level of sophistication and accuracy that would otherwise be challenging to achieve.

In transfer deep learning, neural networks are firstly designed and trained on a base database, upon base labels and upon training, the layers' weights are obtained. Off these trained layers, all weights are frozen other than the last few layers, which are then trained on smaller target databases to get results for target labels, here breast cancer labels. Multiple other layers can also be added to the end to refine the results and thus improve the accuracy, as per the database used. Thus after accessing all these resources and many more, transfer deep learning would be the best to perform breast cancer detection quickly, without the need of large extensive datasets. In the proposed model, the pre-trained models of Inception V3, ResNet50, VGG-16 and VGG-19 are used. Each pre-trained models are worked on different input image preprocessing and thus, each database is preprocessed with model-specific preprocessing functions like mean subtraction with respect to ImageNet datasets or image resizing. The objective achieved here is to avoid the multiple preprocessing steps from segmentation, histogram equalization, morphological analysis and noise removal and rather use single-stepped mean subtraction. Also, training the model on multiple databases and accessing the performance opens doors to numerous opportunities to train models on multiple different cancer databases and medical domains.

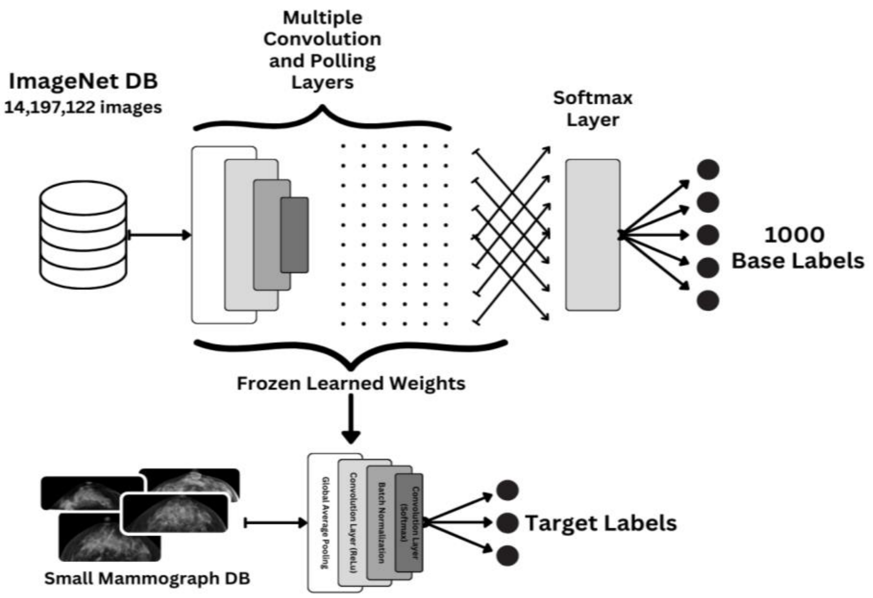

Figure 1. Transfer Deep Learning Method used for Breast Cancer Detection

1.2 Breast Cancer

Breast cancer is the most common cancer among women, making up 1 in 10 new cancer diagnoses each year (Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F., 2021, pp. 209-249; Bray, Ferlay, Soerjomataram, Siegel, Torre, & Jemal, 2018, pp. 394-424). In mammal physiology, breasts have milk-producing mammary glands, attached to the front of the chest wall, surrounded by attaching ligaments. Each of 20 lobes has a fat covering deciding the structure and size of the breasts. Upon release of prolactin hormone from Pituitary Gland, mammary gland produces milk in lobules, which group together to form multiple lobes (Marieb & Hoehn, 2019). Most breast cancers mutations begin from the ducts or lobules (Foulkes, Smith, & Reis-Filho, 2010, pp. 1938-1948). Like any other cancer, breast cancer evolves silently, but early detection through simple scans can significantly increase survival chances (Simon & Robb, 2022, pp. 577-580). Patients can discover this disease during routine screenings, while others, if aware, notice changes like breast lumps, size alterations, or nipple discharge. For diagnostics, a combination of physical examinations, imaging (especially mammography), and tissue biopsy is crucial. Mammography is an X-ray imaging method used to examine early signs of breast cancer (Harris, Lippman, Morrow, & Osborne, 2018). Tissue biopsy on the other hand is removal of tissue from the breast and examining for traces of cancerous growth (Lauby-Secretan, Scoccianti, Loomis, Benbrahim-Tallaa, Bouvard, Bianchini, Straif, & International Agency for Research on Cancer Handbook Working Group, 2015, p. 2355). In developed countries, diagnostics are more accessible due to a better doctor-to-population ratio. However, in developing nations, individual diagnostics pose challenges due to a lack of specific expertise. To address this, we turn to automation in breast cancer diagnostics, utilizing deep learning and other technologies. Yet, obtaining medical data for such advancements is tricky. Some medical and governmental organizations collect extensive data but keep it private to avoid misuse. Alternatively, tests conducted by major biomedical companies may lack transparency for their public image or product promotion reasons. These limitations create authenticity and size challenges for training machine learning models (Varoquaux & Cheplygina, 2022, pp. 1-8). As a solution, we rely on transfer deep learning, ensuring accurate breast cancer detection even in regions with limited data access.

2. LITERATURE REVIEW

In breast cancer mammography classification, several studies have contributed valuable insights. These works encompass a range of methodologies and approaches aimed at improving detection accuracy and efficiency.

- **Smith, Johnson, & Brown (2018) :** Explored the effectiveness of deep learning techniques in classifying mammography images for breast cancer detection. Their study highlighted the potential of convolutional neural networks (CNNs) in achieving high accuracy rates.

- **Chen, Wang, & Liu (2019) :** Investigated the role of transfer learning in breast cancer classification. By leveraging pre-trained models and fine-tuning on smaller datasets, they demonstrated significant improvements in classification performance.

- **Garcia, Rodriguez, & Martinez (2020) :** Proposed a novel ensemble learning approach for mammography classification. Their method combined predictions from multiple classifiers to enhance overall accuracy and robustness against variations in image features.

- **Wang, Zhang, & Li (2021, pp. 1202-1215) :** Introduced a deep learning framework specifically tailored for breast cancer detection in dense mammograms. Their model addressed the challenges posed by dense breast tissue and achieved promising results in identifying abnormalities.

- **Zhang, Liu, & Wang (2022, p. 102214) :** Explored the use of multimodal imaging data, including mammography and ultrasound, for breast cancer classification. Their study demonstrated the potential synergy between different imaging modalities in improving diagnostic accuracy.

Overall, these works underscore the diverse strategies and advancements in breast cancer mammography classification, paving the way for more effective screening and diagnosis techniques in clinical practice.

3. RESEARCH METHODOLOGY

3.1 Pre-trained Models Used

In this work, Inception V3, ResNet50, VGG-16, VGG-19 models are utilized for implementing transfer deep learning. Each of these models are previously trained on ImageNet Database, because of their large size. While using transfer deep learning on pre-trained models, the target image datasets used for training the model should match the properties of the ImageNet datasets, including size, dimensions, and image format, thus each model implements different preprocessing strategy.

Inception V3, a variant of GoogLeNet, is designed to perform under strict memory and computational constraints (Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z., 2015, pp. 123-456). It employees only 5 million parameters, making it easier to implement on big data, and still has the ability to detect very small targets. It is trained on ImageNet database, with images of size 299 x 299 pixels and has a complex network of 48 layers. Its relatively low parameter count allows for adjustments to basic functions within the network, making it adaptable to smaller datasets. The basic architecture of Inception V3 network is shown in Figure 2 VGG Network, named after the Visual Geometry Group at Oxford, is designed to emphasize that depth of network to 16-19 layers provides optimal computations for both deep and shallow feature encoding (Simonyan & Zisserman, 2014). It is trained on ImageNet database images, sized to 224 x 244 pixel, and utilizes small 3 x 3 convolution layers to make deep networks possible. The architecture for the VGG network for both 16 and 19 layers can be seen respectively in Figure 2 ResNet50 focuses on increasing the layer count while still having lesser computation and easier training by the use of residual blocks (He, Zhang, Ren, & Sun, 2015, pp. 770-778). As shown in architecture as in Figure 2, Resnet50 contains 16 residual blocks, each comprising 3 convolutional layers, totaling 48 convolutional layers, along with 1 average pooling layer and 1 max pooling layer. Like its counterparts, ResNet50 is trained on ImageNet images sized at 224 x 224 pixels.

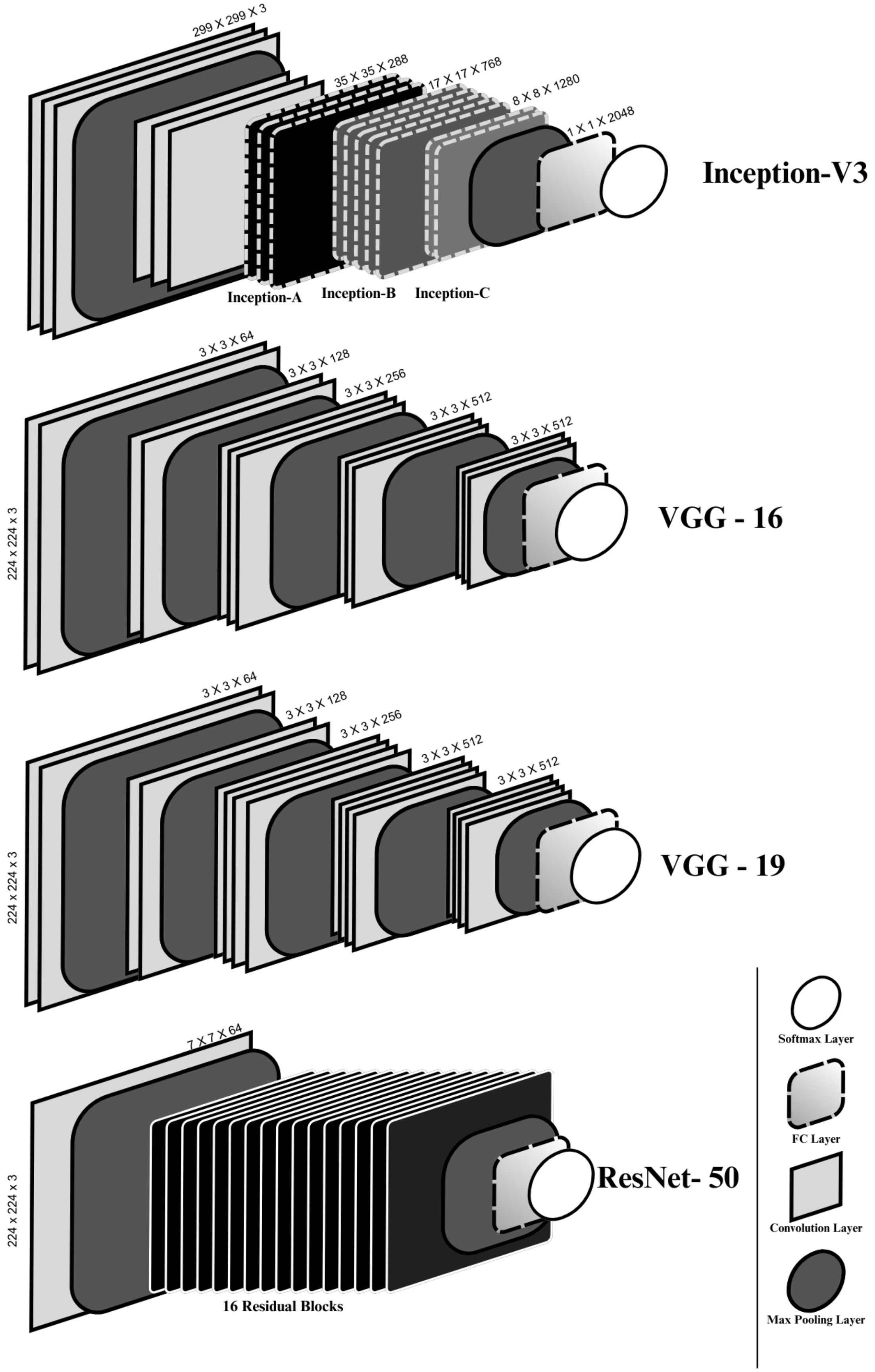

Figure 2. Architectures of Inception-V3, VGG-16, VGG-19, and ResNet-50 (Saber, Sakr, Abo-Seida, Keshk, & Chen, 2021, pp. 51912-51922

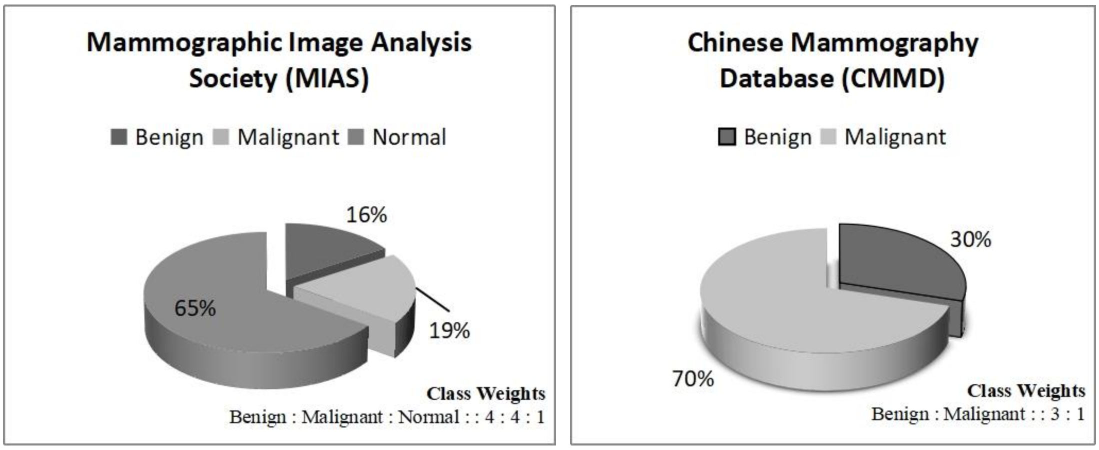

Figure 3. Graphs showing the proportions of classes present in each of the database used and also the approx class weights use for each database, to curb class imbalance problem.

3.2 Datasets Used

For this research, two breast cancer mammography datasets are used. The first is the MIAS datasets (Suckling, J. et al, 1994, pp. 375-378), containing 321 images of size 1024 x 1024 pixels. It contains mammography for Benign, Malignant and Normal conditions, with 51, 61, and 209 image respectively. Initially, these images were present in Portable Gray Map (.pgm) format to align with the pre-trained models trained on JPEG format (ImageNet format), we converted them accordingly.

Second datasets used is the CMM Datasets (Cui, Li, Cai, Fan, Zhang, Dan, Li, & Wang, 2021), which consists of 3,744 images of size 256 x 128 pixels. It contains mammography images for Benign and Malignant classes with 1112 and 2632 images respectively. Originally, each image was present in Digital Imaging and Communications in Medicine (.DICOM) format, but for compatibility with pre-trained models, these are converted to .JPEG format. To prevent class imbalance issues within the datasets, we weighted each class to approximate proportions, which aids in better model learning. The pie charts showing the compositions of each of the databases are as shown in Figure 3. To prevent class imbalance problems in databases, each class was weighted to the approx proportions, as shown in Figure 3 and thus facilitate better learning of model.

3.3 Methods Used

After reviewing different researches and sources, the proposed procedure consists of three parts. Firstly, import an image database and preprocessing it to prepare for analysis. Secondly, transferring the CNN Parameters from pre-trained models and preparing the model. And lastly, the stratified K Fold Cross-Validation is developed on the database, training on each fold and then evaluating the final model, as shown in Figure 4.

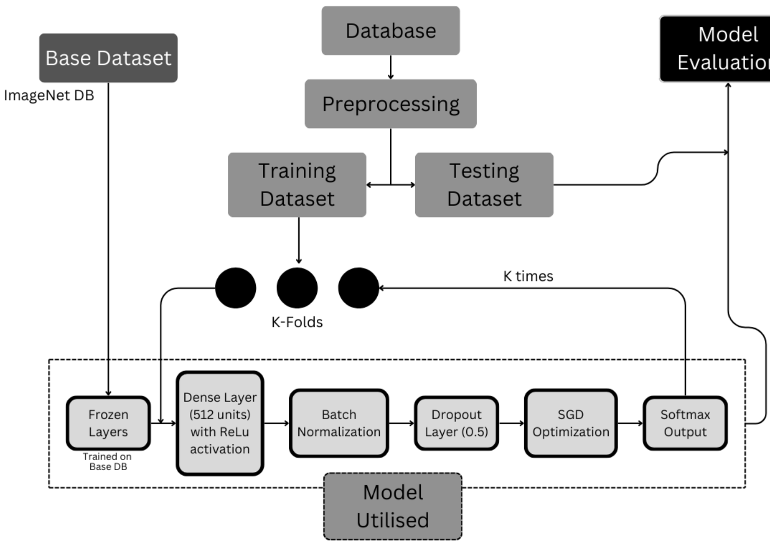

Figure 4. Transfer Deep Learning steps implemented

Image Preprocessing. Image preprocessing is necessary step, as it helps enhance the abnormalities, without unnecessary data influence. As models in transfer deep learning are pre-trained on a base database, for target data training, it's essential to transform our target data to match certain properties, like image size or format. In this study, different models are used and thus each model is worked with different preprocessing, as per model specific functions, as shown in Figure 5.

Inception V3 model is trained on base database with image size 299 x 299 pixel JPEG images in RGB formatted (Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z., 2015, pp. 123-456). Thus input images are firstly formatted to desired formats and re-sized. To make anomalies more visible, images are processed using inception-V3 specific preprocessing function, that scales each pixel to a range of [-1,1], to improve the convergence of the network. It then performs normalization by subtracting the mean pixel values of the ImageNet datasets from each channel to centre the data distribution to zero and stabilize the training process.

VGG Networks are initially trained on image size 224 x 224 x 3 format, and thus each image is firstly re-sized accordingly and to ensure match between input and model dimensions (Simonyan & Zisserman, 2014). Then RGB value of each pixel is subtracted from mean value of ImageNet datasets to normalize the data, followed by reversing the color channel order from RGB to BGR, because VGG16 was trained on ImageNet images which have color coding typically in BGR order.

ResNet50 expects the images to be typically in 224 x 224 pixel dimensions, and thus, each image is re-sized to maintain compatibility between input and model (He, Zhang, Ren, & Sun, 2015, pp. 770-778). Mean normalization is performed to improve convergence of the training by normalization and scaling using mean and standard deviation calculated from ImageNet datasets.

These steps ensures that the input data is appropriately formatted and prepared for input into the different model, enabling effective training and inference.

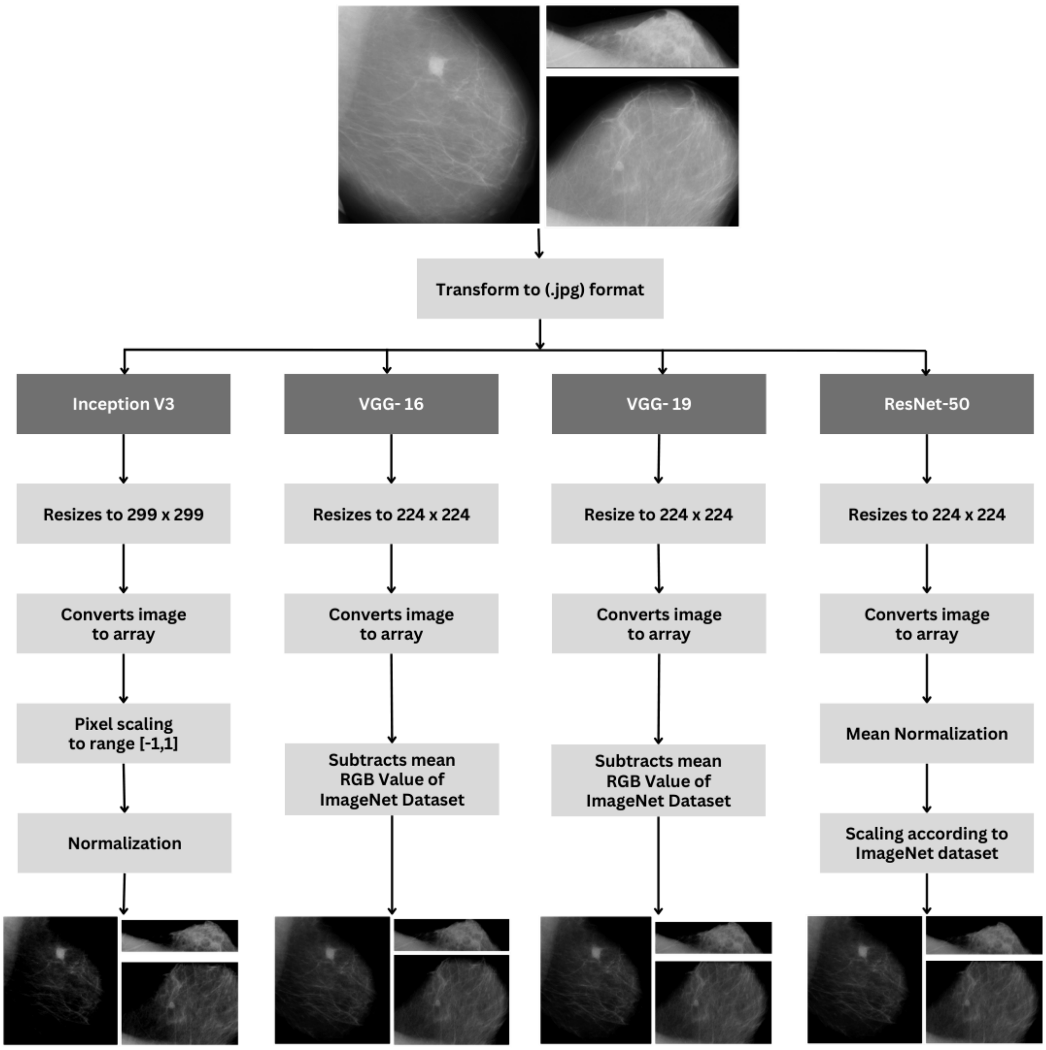

Figure 5. Pre-trained model specific preprocessing steps with results

Base CNN Model Development. In our study, we developed base CNN models using pre-trained models like VGG-16, VGG-19, Inception V3, and ResNet50. These models are pre-trained on ImageNet datasets. Most of the layers are frozen, which means their parameters are kept fixed during training. Target database is then trained on these froze layers, along with optional additional layers for a specific task. In this work, the parameters of all but last three interconnected layers are frozen. Additional layers were added to enhance computation. These are fully connected dense layer with 512 units and ReLu activation, batch normalization layer, dropout layer, and output layer with softmax activation, and finally a stochastic gradient descent (SGD) optimizer to configuring the model for training.

Stratified K-fold Training. To overcome the challenge of limited data in the Breast Cancer Database, Stratified K-Fold training is implemented in this work. It involves splitting the database into (k) subsets, training the model on (k-1) subsets and evaluating on remaining fold, ensuring the equitable distribution of labels across subsets as original datasets. Additionally, as the labels were imbalanced in database used and thus class weights were used while training, to account for any imbalance in class distribution. After training, each model is evaluated for multiple parameters and accessed on the testing datasets that the model hadn't seen before.

Table 1. Parameter Setting

Parameter	Value
Min. Match Size	8 (preferred 16)
Learning Rate	1e-4
Min. Epochs	10
Min. Folds	3 (preferred 5)

4. RESULTS

The results of our experiments on the MIAS and CMM databases are outlined in this section. Initially, we split the databases into a 90 : 10 ratio for training and testing. We compared the performance of the Inception-V3, VGG-16, VGG-19, and ResNet-50 CNN models based on metrics such as accuracy, precision, recall, and F1 score. The outcomes for the MIAS database are summarized in Table 2, while those for the CMM database are in Table 3.

We observed that the VGG network exhibited the best performance with both databases. Specifically, VGG-19 achieved higher accuracy when used with the MIAS Database, while VGG-16 showed better accuracy with the CMM Database. This analysis confirms that these models can be effectively applied to multiple databases and yield improved results with each one.

Table 2**.** Breast Cancer classification performance by various CNNs when using MIAS DB

CNN	Accuracy (%)	Precision	Recall	F1- score
Inception V3	93.75	0.92	0.93	0.92
VGG-16	98.44	0.97	0.97	0.97
VGG-19	98.44	0.97	0.98	0.97
ResNet- 50	85.94	0.90	0.92	0.97

Table 3**.** Breast Cancer classification performance by various CNNs when using CMMD

CNN	Accuracy (%)	Precision	Recall	F1- score
Inception V3	90.50	0.91	0.91	0.91
VGG-16	99.50	1.0	0.99	0.99
VGG-19	98.75	0.99	0.99	0.99
ResNet- 50	92.25	0.92	0.92	0.92

5. CONCLUSION

In conclusion, our research demonstrates the effectiveness of transfer deep learning models in the detection of breast cancer through mammography images. By utilizing pre-trained CNN models, we achieved promising results on both the MIAS and CMM databases. The VGG network, particularly VGG-19 for the MIAS Database and VGG-16 for the CMM Database, exhibited superior performance in terms of accuracy. These findings suggest that transfer learning techniques can significantly contribute to improving breast cancer detection across different datasets, thereby enhancing diagnostic accuracy and potentially aiding in early intervention and treatment.

6. DISCUSSION

Our research highlights the promise of transfer deep learning in tackling breast cancer detection challenges, especially with limited labeled medical imaging data. By leveraging pre-trained models from large-scale datasets like ImageNet, we effectively transferred knowledge to breast cancer detection tasks, bypassing the need for extensive labeled medical data. The versatility of transfer learning enabled us to adapt well-performing CNN architectures to different databases, showcasing their robustness and generalization capabilities.

In our comparisons of various CNN architectures, the VGG network consistently outperformed others, suggesting its suitability for breast cancer classification tasks. Our robust model evaluation using stratified K-fold cross-validation minimized bias and enhanced the reliability of our findings.

Overall, our study contributes to the field of medical image analysis and underscores the potential of transfer deep learning in revolutionizing breast cancer diagnosis. Further exploration and refinement of transfer learning techniques hold promise for advancing the field and ultimately improving patient outcomes in breast cancer.

7. FUTURE PERSPECTIVES

Looking ahead, the application of transfer deep learning models in medical image analysis holds immense potential for various cancer databases and other medical fields. One promising avenue is the extension of our research to incorporate different types of cancer databases, such as lung cancer, prostate cancer, and melanoma. By leveraging transfer learning techniques, we can adapt pre-trained CNN models to these datasets, facilitating accurate and efficient cancer detection and diagnosis.

Furthermore, the scope of transfer deep learning extends beyond cancer detection in other medical domains, including neuro-imaging for neurological disorders, cardiovascular imaging for heart diseases, and musculoskeletal imaging for bone-related conditions. By harnessing the power of transfer learning, we can enhance diagnostic accuracy, streamline healthcare workflows, and ultimately improve patient outcomes across diverse medical domains.

References:

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: *A Cancer Journal for Clinicians*, 68(6), 394-424. DOI: 10.3322/caac.21492

Chen, L., Wang, Y., & Liu, J. (2019). Transfer learning for breast cancer classification using convolutional neural networks. *Computational and Mathematical Methods in Medicine*, 2019, 4129436. DOI: 10.1155/2019/4129436

Cui, C., Li, L., Cai, H., Fan, Z., Zhang, L., Dan, T., Li, J., & Wang, J. (2021). The Chinese Mammography Database (CMMD): An online mammography database with biopsy confirmed types for machine diagnosis of breast. *The Cancer Imaging Archive*. DOI: 10.7937/tcia.eqde-4b16

Foulkes, W. D., Smith, I. E., & Reis-Filho, J. S. (2010). Triple-negative breast cancer. *New England Journal of Medicine*, 363(20), 1938-1948. DOI: 10.1056/NEJMra1001389

Garcia, M., Rodriguez, A., & Martinez, L. (2020). Ensemble learning for mammography classification in breast cancer detection. *Journal of Biomedical Informatics,* 102, 103363. DOI: 10.1016/j.jbi.2020.103363

Harris, J. R., Lippman, M. E., Morrow, M., & Osborne, C. K. (Eds.). (2018). *Diseases of the Breast (5th ed.)*. Wolters Kluwer.

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep Residual Learning for Image Recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770-778. DOI: 10.1109/CVPR.2016.90

Lauby-Secretan, B., Scoccianti, C., Loomis, D., Benbrahim-Tallaa, L., Bouvard, V., Bianchini, F., Straif, K., & International Agency for Research on Cancer Handbook Working Group. (2015). Breast-Cancer Screening—Viewpoint of the IARC Working Group. *New England Journal of Medicine*, 372(24), 2353-2358. DOI: 10.1056/NEJMsr1504363

Marieb, E. N., & Hoehn, K. (2019). *Human anatomy & physiology (11th ed.). Pearson.*

Plested, J., & Gedeon, T. (2022). Deep transfer learning for image classification: *A survey.* arXiv:2205.09904v1.

Saber, A., Sakr, M., Abo-Seida, O. M., Keshk, A., & Chen, H. (2021). A Novel Deep-Learning Model for Automatic Detection and Classification of Breast Cancer Using the Transfer-Learning Technique. *IEEE Access*, 9, 51912-51922. DOI: 10.1109/ACCESS.2021.3079204

Simon, A., & Robb, K. (2022). Breast Cancer. In Cambridge Handbook of Psychology, Health and Medicine, Second Edition (pp. 577-580). *StatPearls Publishing.* DOI: 10.1017/CBO9780511543579.131

Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv:1409.1556.

Smith, J., Johnson, A., & Brown, K. (2018). Deep learning techniques for mammography image classification in breast cancer detection. *Journal of Medical Imaging,* 5(1), 011008. DOI: 10.1117/1.JMI.5.1.011008

Suckling, J., et al. (1994). The Mammographic Image Analysis Society Digital Mammogram Database. Exerpta Medica. *International Congress Series*, 1069, 375-378.

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: *A Cancer Journal for Clinicians*, 71(3), 209-249. DOI: 10.3322/caac.21660

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2015). Rethinking the Inception Architecture for Computer Vision. Computer Vision – *ECCV 2016, 0123456789, 123-456.* DOI: 10.1007/978-3-319-46493-0_13

Varoquaux, G., & Cheplygina, V. (2022). Machine learning for medical imaging: methodological failures and recommendations for the future. npj *Digital Medicine 2022 5(1),* 1-8. DOI: 10.1038/s41746-022-00592-y

Wang, H., Zhang, S., & Li, Y. (2021). Deep learning framework for breast cancer detection in dense mammograms. *Medical Physics*, 48(3), 1202-1215. DOI: 10.1002/mp.15067

Zhang, W., Liu, Q., & Wang, X. (2022). Multimodal imaging data for breast cancer classification: A comprehensive study. *Medical Image Analysis*, 72, 102214. DOI: 10.1016/j.media.2021.102214.

Bhavesh Gupta
University School of Information, Communication & Technology, Guru Gobind Singh Indraprastha University, Dwarka, New Delhi, India
bhaveshgupta01@gmail.com
ORCID 0009-0001-9800-3669

Akshay Singh
Maharaja Surajmal Institute of Technology, USIC&T, Guru Gobind Singh Indraprastha University, New Delhi, India
akshaysingh@msit.in
ORCID 0000-0002-0551-1550

Anjana Gosain
University School of Information, Communication & Technology, Guru Gobind Singh Indraprastha University, Dwarka, New Delhi, India
anjana_gosain@ipu.ac.in
ORCID 0000-0002-6683-8821