2024 4th International Conference of Science and Information Technology in Smart Administration (ICSINTESA) | 979-8-3503-7611-1/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICSINTESA62455.2024.10748049

Breast Cancer Image Classification Obtained Through Dynamic Thermography using Deep Learning

Ethan Daniel Timadius
Data Science Program
Computer Science Department
School of Computer Science,
Bina Nusantara University,
Jakarta, Indonesia 11480
ethan.timadius001@binus.ac.id

Reyner Wongso
Data Science Program
Computer Science Department
School of Computer Science,
Bina Nusantara University,
Jakarta, Indonesia 11480
reyner.wongso@binus.ac.id

Thirza Baihaqi
Data Science Program
Computer Science Department
School of Computer Science,
Bina Nusantara University,
Jakarta, Indonesia 11480
thirza.baihaqi@binus.ac.id

Alexander Agung Santoso Gunawan
Computer Science Department,
School of Computer Science,
Bina Nusantara University,
Jakarta, Indonesia 11480,
aagung@binus.edu

Karli Eka Setiawan
Computer Science Department,
School of Computer Science,
Bina Nusantara University,
Jakarta, Indonesia 11480,
karli.setiawan@binus.ac.id

***Abstract*—This paper addresses the vital task of diagnosing breast cancer early and accurately. It acknowledges the difficulty in detecting minor variations in breast cancer patterns through visual examination alone, particularly when analyzing grayscale images. To overcome this challenge, our research concentrates on using image processing models, particularly Convolutional Neural Networks (CNNs), and enhances them with a refined approach to classification. The research methodology encompasses rigorous data collection, preprocessing, and strategic feature selection to enhance model robustness. To address challenges related to data intricacies, the study incorporates transfer learning, refining the model's training process for improved accuracy. The paper not only seeks to elevate diagnostic precision and minimize errors but also aims to provide healthcare professionals with efficient tools for breast cancer detection. Building upon prior research, the study underscores the potential of deep learning methodologies in advancing breast cancer diagnosis. Furthermore, the fine-tuning process involves adapting a pre-trained model to the intricacies of breast cancer patterns, optimizing its performance. The ultimate goal is to contribute to the development of reliable diagnostic tools that empower medical professionals in the fight against breast cancer. The paper concludes by summarizing key findings. Additionally, the model achieved an accuracy of 84%, highlighting its effectiveness in breast cancer detection. The paper also emphasizes the broader significance of leveraging deep learning for enhanced breast cancer detection.**

Keywords: Breast Cancer, Deep Learning, Thermal Image, Image Classification, Convolutional Neural Network

I. INTRODUCTION

Breast cancer is a serious concern throughout the world. It is the second leading cause of death by cancer for women worldwide [1]. In 2018 alone, there have been 2,1 million new cases of breast cancer, and the rate has increased 3,1% per year [2]. Early-stage breast cancer is considered curable, so correct and early diagnosis of breast cancer is the main key to successful treatment [3]. Medical images, obtained using mammograms and dynamic thermograms, are very important in identifying breast cancer [4]. But, in understanding these images, there are often variations in opinion and human error can occur.

Furthermore, mammograms can miss some malignant tumors, particularly in dense breasts, highlighting the need for complementary diagnostic tools. Dynamic thermography offers promise in this regard, as it captures functional information about blood flow and metabolic activity within the breast, potentially revealing abnormalities invisible to standard X-rays. By harnessing the combined power of medical imaging and deep learning, the possibility of accurate and early breast cancer detection, regardless of breast density, presents itself as a beacon of hope in the fight against this prevalent disease [5].

The innovation in dynamic thermography technology and the development of deep learning algorithms present significant potential in enhancing the accuracy of breast cancer diagnosis. This section will delve into the experimental methods applied in this research [6]. The fusion of dynamic thermography data with deep learning techniques aims to identify specific patterns and features indicative of breast cancer presence. A thorough analysis of the interaction between dynamic thermography data and the deep learning model is outlined to provide a comprehensive overview of the proposed approach [7].

Therefore, this paper aims to compare a baseline model, transfer learning, and the results obtained from related studies. In Section 1, the discussion centers on the dataset employed in this research, outlining its characteristics and significance in breast cancer image classification. Section 2 elaborates on the data preprocessing techniques applied to enhance the quality and relevance of the dataset for subsequent analysis. Moving on to Section 3, the model description provides a detailed account of the architecture and configurations of the proposed deep learning model for breast cancer classification using dynamic thermography. Additionally, in Section 3, we delve into the incorporation of transfer learning techniques and their impact on model performance. Subsequently, in Section 4, an exploration into the model evaluation metrics employed reveals the criteria used to assess the performance, accuracy, and robustness of the proposed approach, with comparisons drawn against baseline models and pertinent findings from existing literature. Through this comprehensive examination, insights into the efficacy of the proposed model, including the contributions of transfer learning, in advancing breast cancer diagnosis through dynamic thermography and deep learning are aimed to be provided.

II. LITERATURE REVIEW

Breast cancer requires early and precise diagnosis for effective treatment. Currently, there are a few imaging techniques that are used to detect breast cancer, such as histological images, thermal images, x-ray images (mammography), and breast ultrasound images [8]. In this study, a dataset of x-ray images is utilized as the foundation for developing the model

Histological images, often obtained via x-ray imaging, are crucial for diagnosing conditions like breast cancer. These images offer detailed views of tissue structures, aiding in the detection of abnormal growth patterns and calcifications associated with tumors. Advanced technology and image processing algorithms enhance diagnostic accuracy, enabling timely interventions and personalized treatment strategies [9].

Mammography, a specialized X-ray technique, and breast ultrasound, employing sound waves to produce detailed images of internal breast structures, are distinct yet complementary medical imaging methods used for comprehensive breast tissue examination.

A study conducted by D. L. Birdwell et al. shows that mammography, the gold standard for screening, suffers from limitations like false positives and negatives, especially in dense breasts [10]. Due to this reason, there has been numerous research to improve the classification of breast cancer using machine learning and deep learning as well. As proven in a review conducted by M. Madani et al. the advent of deep learning powered by advanced artificial intelligence (AI) and deep neural network techniques has opened new horizons in breast cancer image classification, enabling more accurate diagnoses. The Convolutional Neural Network (CNN) model they used in identifying the mammogram dataset achieved an AUC of 92.9%, which is better than the baseline method based on a conventional method with an AUC of 91% [11]. Similarly, Mangasarian et al. (1995) devised a linear programming model for breast cancer diagnosis and prognosis, which yielded an accuracy of 94.8%, comparable to that of deep learning models [12]. The current state of deep learning models, if trained on extensive datasets of thermograms and clinical data, utilizing frameworks like TensorFlow and PyTorch, can identify intricate patterns associated with breast cancer. This in turn provides a means to predict the likelihood of breast cancer in new thermograms with much greater accuracy than humans [13].

"Recent studies have shown great promise when it comes to breast cancer detection using deep learning models. For instance, a study by L. A. Habel et al. in 2023, published in Breast Cancer Research, showed that these advanced deep learning models, often created and fine-tuned using tools like Keras, performed better than the traditional risk assessment tools. They did a fantastic job predicting breast cancer risk among over 21,000 women [14].

Another study from 2023, conducted by R. Adam et al. published in Breast Cancer Research, proved that deep learning models using CNN outshines human radiologists in detecting breast cancer on MRI scans [15]. These smart models are usually developed and put into action using Python, along with libraries like scikit-learn to make sure they work well. What's interesting is that deep learning isn't just for diagnosis; it's also reliable in predicting breast cancer recurrence. In a 2021 study published in Breast Cancer Research, these models were shown to be more accurate than the old-school methods, when it came to predicting recurrence by processing the image of ductal carcinoma of the breast [16].

Furthermore, deep learning can identify different types of breast cancer. These different types need different treatment plans. Deep learning models can tell these types apart using various data sources. These sources often involve some bioinformatics tools to handle and analyze the data. And, to make sure these models are spot on, they use popular machine learning libraries to classify those biomarkers. A study conducted by U. Ilhan et al. achieves high processing performances with 95.4% of accuracy in the multi-class breast cancer classification task when compared with state-of-the-art models [17]**.** Another study introduces an innovative deep learning architecture for breast cancer segmentation and classification, using a mammography image dataset. It demonstrates high accuracy in distinguishing between benign and malignant tumors [18].

Transfer learning, an increasingly significant concept in deep learning, has been leveraged to enhance the capabilities of Convolutional Neural Networks (CNNs) by employing fine-tuning. In fine-tuning, pretrained CNN models, often sourced from general image datasets like ImageNet, serve as the foundation. Fine-tuning is particularly beneficial when adapting these models to more specific tasks, such as medical image classification. It involves adjusting the model's parameters during training to make it more suitable for the target domain, which, in this case, is the diagnosis of breast cancer using medical images [19]. Fine-tuning essentially allows the model to inherit valuable features and representations from the source task (e.g., classifying natural images), reducing the need for extensive training from scratch. [20]**.**

The research on deep learning's applications in breast cancer diagnosis holds a lot of potential [21]. While challenges like the need for huge datasets and computing power, deep learning has the potential to help change the diagnosis and treatment of breast cancer, offering personalized screening and improved outcomes.

III. RESEARCH METHODOLOGY

A. Dataset

The dataset that will be used in this research is called

"Breast Cancer Diagnosis" which can be accessed from Kaggle, filled with images of breast cancer obtained using dynamic thermography to determine breast cancer. The dataset is divided into two sections, the testing set and training set. Each of those sets have two categories for breast cancer, benign (non-cancerous tumors) and malignant (cancerous tumors). The testing set has 240 total observations, 120 for benign and another 120 for malignant. The training set has a total of 1280 observations, 640 for benign and 640 for malignant.

B. Data Preprocessing

Fig 1. Sample Dataset

The figure above shows samples of breast cancer images obtained using dynamic thermography. When observing the images, there are a few distinct characteristics contained in each classification of data. Data with benign classification has more well-defined tumors and an apparent oval shape (left). Data with malignant classification has less well-defined tumors and no consistency in their shape (right).

Preprocessing using normal methods for medical image data is absolutely essential in computer vision tasks. It transforms raw images into a format that is suitable for neural networks while preserving important features [22]. Several key steps are implemented to ensure those important features are kept and the model has increased effective processing and improved performance [23] [24].

The current dataset has some distinct characteristics. The size of images in this dataset is already uniform. However, the data is still resized, because resizing reduces computation requirements and increases training speed. For the colors of the data, it is already grayscale by default, so there are no other colors other than black, white, and gray. Because of this normalization is done by dividing the intensity of color by 255 to achieve normality. The data has a uniform structure, meaning that it does not have many variations in angle, shapes, and sizes. To combat this, data augmentation techniques are then applied to enhance model generalization, involving flips, zooms, and shifts to generate a more diverse set of images. The data that is augmented is the training set, so before augmentation there needs to be data splitting. The data is split into 80% training set and 20% testing set. These combined steps play a crucial role in optimizing the model's ability to work with the preprocessed data and enhance its overall performance.

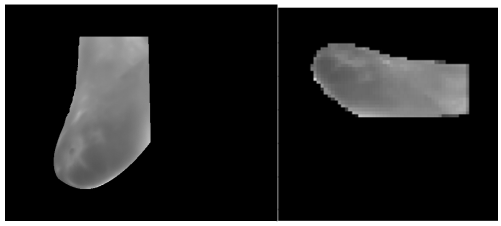

Fig 2. Image Augmentation Result

The figure above shows the result of image augmentation. The left image is the original data, and the right image is the augmented data. The pixel dimension is resized to 64x64, making the image look more blocky and the dimension is uniform. Next, the intensity of the grayscale colors have been normalized. It's not very apparent, however it will help improve the convergence of the model. Data augmentation to the data makes it look more diverse, with the figure above showing changes in the image position.

C. Model Description

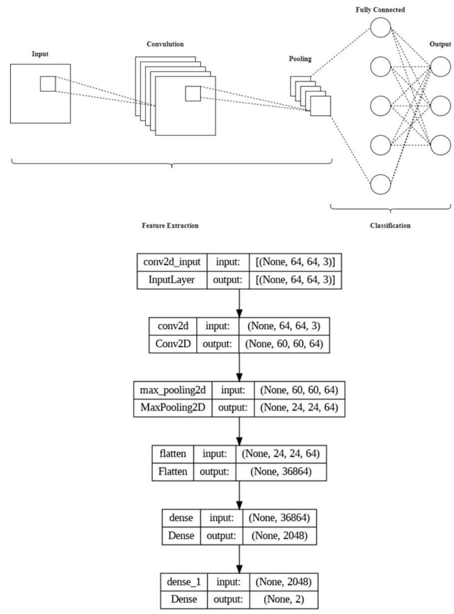

Fig 3. CNN Architecture

In the development of a baseline model for breast cancer image classification, a Convolutional Neural Network (CNN) with the following architecture was employed: an initial Conv2D layer with 64 filters of size 5x5, MaxPooling2D for dimension reduction, and Flatten to flatten the output. Two subsequent Dense layers were utilized, with the first layer containing 2048 ReLU neurons, and the final layer consisting of 2 softmax neurons. The model was compiled using the Adam optimizer, categorical cross entropy loss, and accuracy as the metric.

D. Model Evaluation Metrics

The performance of the deep learning model is assessed in this image classification research using loss and accuracy measures [25]. By measuring the difference between the predicted classes and the actual classes (ground truth) for a series of photos, Loss measures how effectively the deep learning model is working. A deep learning model that gradually reduces loss is considered best. The number of times the deep learning model's predictions match the actual class labels is measured simply and intuitively by accuracy. A deep learning model that gradually improves in accuracy is ideal. The F1-score is a significant statistical method that is frequently used in addition to loss and accuracy to assess the performance of classification models, particularly deep learning models for image classification. The F1-score provides a fair evaluation of the model's performance, especially in situations where there is a class imbalance, by combining precision and recall into a single metric [26]. Recall is the percentage of true positive forecasts among all real positive instances, whereas precision is the percentage of true positive predictions among all positive predictions. The F1-score measures the model's ability to properly identify positive instances while reducing false positives and false negatives. It is calculated as the harmonic mean of precision and recall. Therefore, keeping an eye on the F1-score in addition to loss and accuracy can offer important insights into how well the deep learning model performs on a variety of classification tasks.

E. Transfer Learning

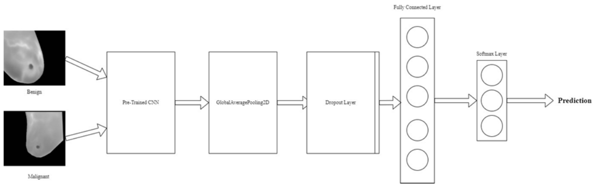

Fig 4. Transfer Learning Architecture

Transfer learning is a powerful technique within the realm of deep learning that plays a crucial role in the rapid and accurate training of Convolutional Neural Networks (CNNs) without initializing their weights from scratch. Instead, it involves importing pre-trained weights from another CNN model, often one that has been extensively trained on a more extensive dataset, such as the ImageNet dataset. These pre-trained weights are derived from several CNN architectures that have achieved high accuracy when classifying a broad range of objects.

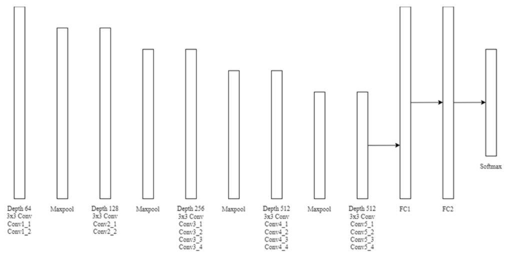

Fig 5. VGG19 Architecture

The transfer learning that is used in this research is VGG19. This is because VGG19 is trained with the diverse ImageNet dataset, which will allow it to learn general features and patterns that can be useful for breast cancer detection. Its straightforward architecture and good performance on various image classification tasks also make it a convenient starting point for experimentation and comparison with more complex models.

In the context of this research, transfer learning is employed as a key strategy to enhance the training of the CNN for breast cancer image classification. There are four primary strategies for implementing transfer learning, however the one that is used is training from scratch. This method uses a state-of-the-art CNN architecture and initiates the training process from scratch, solely relying on the architecture's proven performance on diverse challenging datasets [27] [28].

IV. RESULT AND DISCUSSION

The evaluation of the baseline model, a Convolutional Neural Network (CNN) designed for binary breast cancer classification, revealed a testing accuracy of approximately 58.55% with a testing loss of around 0.84. This initial model, although providing a foundation, exhibited limitations in its performance metrics.

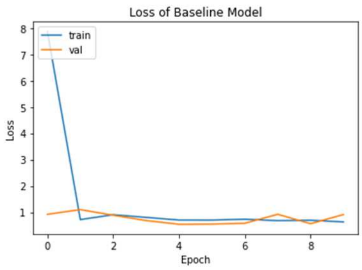

Fig 6. Baseline Model Loss

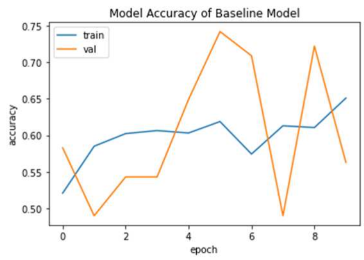

Fig 7. Baseline Model Accuracy

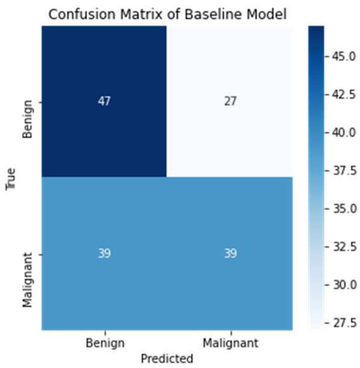

Fig 8. Baseline Model Confusion Matrix

To enhance the model's capabilities, a new architecture leveraging transfer learning with VGG19 pre-trained on ImageNet was introduced. This improved model incorporated a dropout layer to mitigate overfitting and achieved notable advancements. The testing accuracy surged to approximately

84.87%, accompanied by a reduced testing loss to about 0.37. This stark improvement in performance underscores the efficacy of transfer learning in the context of breast cancer image classification.

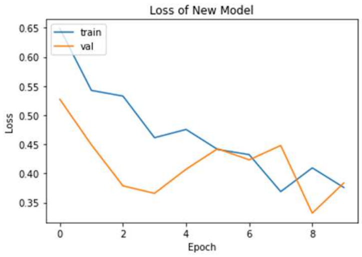

Fig 9.Transfer Learning Model Loss

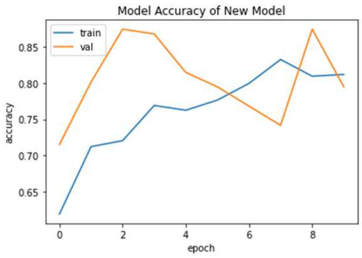

Fig 10.Transfer Learning Model Accuracy

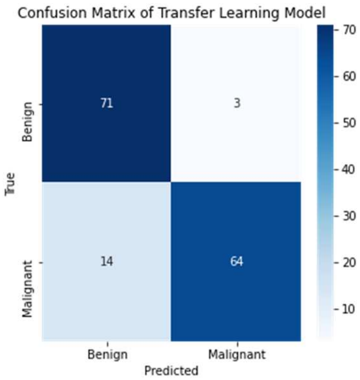

Fig 11. Transfer Learning Model Confusion Matrix

In the conducted experiments, two distinct models were evaluated: a baseline model and a new model that underwent transfer learning. The performance of these models was assessed based on their loss and accuracy metrics over a series of epochs during the training process.

The baseline model demonstrated a rapid decrease in training loss from the initial epochs, stabilizing around a loss of 1. However, the validation loss remained relatively stable, indicating that the model's performance on unseen data was not significantly improving after the initial epochs. This could be indicative of underfitting, where the model is too simple to capture the underlying patterns in the data. The accuracy of the baseline model increased over the epochs for the training set, but fluctuated for the validation set, further supporting the possibility of underfitting.

On the other hand, the new model that underwent transfer learning showed a different pattern. The training loss decreased over the epochs, while the validation loss initially decreased, then increased slightly after epoch 4 before decreasing again. This fluctuation in validation loss could suggest overfitting around epoch 4, where the model might be learning the training data too well and performing poorly on unseen data. The training accuracy of this model increased steadily over the epochs, while the validation accuracy fluctuated, peaking around epoch 3 and then decreasing. This could be another indication of overfitting, as the model performs well on the training data but less so on the validation data.

The convergence of the loss and accuracy lines in both models suggests that the models are learning from the training data. However, the fluctuations in validation loss and accuracy indicate that the models' performance on unseen data is not consistently improving. This highlights the importance of monitoring both training and validation metrics during the training process to ensure the model generalizes well to unseen data. Techniques such as early stopping, dropout, or further tuning of the model parameters can help mitigate overfitting and improve the model's performance.

TABLE I. RESULT COMPARISON

	f1	accuracy	precision	recall
Baseline	0.650	0.585	0.528	0.846
Transfer Learning	0.961	0.848	0.974	0.949

Comparing the two models emphasizes the superiority of the new architecture. The testing accuracy experienced a significant boost from 58.55% to 84.87%, showcasing the potency of leveraging pre-trained models for feature extraction. Concurrently, the testing loss saw a substantial decrease, indicating the model's enhanced predictive capabilities.

In-depth analysis through classification reports further validates the superiority of the new model, crucially in the context of medical applications. Precision, recall, and f1-score metrics for both benign and malignant classes demonstrate more balanced and improved performance in the new architecture.

The new transfer learning model excels in achieving a high recall, emphasizing its robustness in minimizing the risk of false negatives. The overall accuracy of the new model stands at an impressive 85%, affirming its efficacy in distinguishing between benign and malignant cases, with a particular emphasis on the crucial metric of recall in the medical context.

Visualization of the training and validation metrics over epochs provides insights into the learning dynamics of both models. The baseline model exhibited a modest improvement in training accuracy, but the validation accuracy plateaued. In contrast, the new model showcased consistent improvements in both training and validation accuracy, indicating its ability to generalize well to unseen data.

V. CONCLUSION

Based on the methodology and results presented, it is evident that the Transfer Learning Model (VGG19) significantly outperforms the Baseline Model in binary image classification. The Transfer Learning Model achieved an overall accuracy of 85%, compared to the Baseline Model's 59%. This substantial difference in performance can be attributed to the robustness of the VGG19 architecture, which was fine-tuned for the specific task.

The Baseline Model, despite its simplicity and the application of convolutional and pooling layers for feature extraction, demonstrated a higher rate of false positives and false negatives. In contrast, the Transfer Learning Model, with its additional dense layers and dropout for fine-tuning, resulted in fewer misclassifications, indicating its superior predictive power.

For future research, it would be beneficial to explore other pre-trained models and architectures for transfer learning, such as ResNet, Inception, or EfficientNet, which might offer further improvements in accuracy. Additionally, experimenting with different data augmentation techniques, hyperparameter tuning, and increasing the number of training epochs could potentially enhance the model's performance. Lastly, incorporating a larger and more diverse dataset for training could also contribute to a more robust and generalizable model.

In conclusion, the use of transfer learning, specifically the VGG19 model, has proven to be a more effective approach for binary image classification in this study. However, continuous exploration and experimentation are crucial for further advancements in this field.

VI. REFERENCES

[1] A. Burguin, C. Diorio, and F. Durocher, "Breast cancer treatments: Updates and new challenges," Journal of Personalized Medicine, vol. 11, no. 8, p. 808, 2021.

[2] N. Harbeck et al., "Breast cancer," Nature Reviews Disease Primers, vol. 5, no. 1, 2019.

[3] R. Hong and B. Xu, "Breast cancer: An up-to-date review and Future Perspectives," Cancer Communications, vol. 42, no. 10, pp. 913–936, 2022.

[4] H. Rahman, T. F. Naik Bukht, R. Ahmad, A. Almadhor, and A. R. Javed, "Efficient breast cancer diagnosis from complex mammographic images using deep convolutional neural network," Computational Intelligence and Neuroscience, vol. 2023, pp. 1–11, 2023.

[5] S. W. Fletcher and J. G. Elmore, "Mammographic Screening for Breast Cancer," New England Journal of Medicine, vol. 348, no. 17, pp. 1672–1680, Apr. 2003.

[6] Y. Ohashi and I. Uchida, "Applying dynamic thermography in the diagnosis of breast cancer," IEEE Engineering in Medicine and Biology Magazine, vol. 19, no. 3, pp. 42–51, 2000.

[7] A. Nogales, F. Pérez-Lara, and Á. J. García-Tejedor, "Enhancing breast cancer diagnosis with deep learning and evolutionary algorithms: A comparison of approaches using different thermographic imaging treatments," Multimedia Tools and Applications, Oct. 2023.

[8] Y. Lu, J.-Y. Li, Y.-T. Su, and A.-A. Liu, "A review of breast cancer detection in medical images," 2018 IEEE Visual Communications and Image Processing (VCIP), 2018.

[9] [1]T. L. Le, M. H. Bui, N. C. Nguyen, M. T. Ha, A. Nguyen, and H. P. Nguyen, "Transfer learning for deep neural networks-based classification of breast cancer X-ray images," 2023.

[10] D. L. Birdwell, M. A. Conant, D. W. Lee, and B. F. Geschke, "Mammography sensitivity and cancer detection rates in dense versus non-dense breasts," AJR. American journal of roentgenology, vol. 202, no. 1, pp. 121–126, 2014.

[11] M. Madani, M. M. Behzadi, and S. Nabavi, "The role of deep learning in advancing breast cancer detection using different imaging modalities: A systematic review," Cancers, vol. 14, no. 21, p. 5334, 2022.

[12] O. L. Mangasarian, W. N. Street, and W. H. Wolberg, "Breast Cancer Diagnosis and Prognosis Via Linear Programming," Operations Research, vol. 43, no. 4, pp. 570–577, Aug. 1995, doi: 10.1287/opre.43.4.570.

[13] M. Ensafi, M. R. Keyvanpour, and S. V. Shojaedini, "A New method for promote the performance of deep learning paradigm in diagnosing breast cancer: improving role of fusing multiple views of thermography images," Health and Technology, vol. 12, no. 6, pp. 1097–1107, Oct. 2022.

[14] L. A. Habel, S. E. Alexeeff, N. Achacoso, V. A. Arasu, A. Gastounioti, L. Gerstley, et al., "Examination of fully automated mammographic density measures using LIBRA and breast cancer risk in a cohort of 21,000 non-Hispanic white women," Breast Cancer Research, vol. 25, no. 1, p. 5, 2023.

[15] R. Adam, K. Dell'Aquila, L. Hodges, T. Maldjian, and T. Q. Duong, "Deep learning applications to breast cancer detection by Magnetic Resonance Imaging: A literature review," Breast Cancer Research, vol. 25, no. 1, 2023. doi:10.1186/s13058-023-01687-4 vol. 22, no. Supp01, 2022.

[16] G. Fernandez et al., "Development and validation of an AI-enabled digital breast cancer assay to predict early-stage breast cancer recurrence within 6 years," Breast Cancer Research, vol. 24, no. 1, 2022.

[17] U. Ilhan, K. Uyar, and E. I. Iseri, "Breast Cancer Classification Using Deep Learning," Advances in Intelligent Systems and Computing, pp. 709–714, 2021.

[18] S. Ramesh, S. Sasikala, S. Gomathi, V. Geetha, and V. Anbumani, "Segmentation and classification of breast cancer using novel deep learning architecture," Neural Computing and Applications, vol. 34, no. 19, pp. 16533–16545, May 2022.

[19] D. Benavente, G. Gatica, and J. González-Feliu, "Balanced medical image classification with transfer learning and Convolutional Neural Networks," Axioms, vol. 11, no. 3, p. 115, 2022.

[20] H. E. Kim et al., "Transfer learning for medical image classification: A literature review," BMC Medical Imaging, vol. 22, no. 1, 2022.

[21] A. Gastounioti, S. Desai, V. S. Ahluwalia, E. F. Conant, and D. Kontos, "Artificial Intelligence in mammographic phenotyping of breast cancer risk: A narrative review," Breast Cancer Research, vol. 24, no. 1, 2022.

[22] L. M. Mina and N. A. M. Isa, "A Review of Computer-Aided Detection and Diagnosis of Breast Cancer in Digital Mammography," Journal of Medical Sciences, vol. 15, no. 3, pp. 110–121, Apr. 2015.

[23] A.Rama, A.Kumaravel, C. Nalini, "Preprocessing Medical Images for Classification using Deep Learning Techniques," International Journal of Innovative Technology and Exploring Engineering, vol. 8, no. 9S3, pp. 711–716, Aug. 2019. 10.35940/ijitee.i3147.0789s319

[24] Ş. Öztürk and B. Akdemir, "Effects of histopathological image pre-processing on convolutional neural networks," Procedia Computer Science, vol. 132, pp. 396–403, 2018.

[25] Z. Qian, "Do We Need Fully Connected Output Layers in Convolutional Networks?," arXiv.org, Apr. 28, 2020.

[26] D. J. Hand, P. Christen, and N. Kirielle, "F*: an interpretable transformation of the F-measure," Machine Learning, vol. 110, no. 3, pp. 451–456, Mar. 2021, doi: 10.1007/s10994-021-05964-1.

[27] Y. Zhu et al., "Heterogeneous transfer learning for image classification," Proceedings of the AAAI Conference on Artificial Intelligence, https://ojs.aaai.org/index.php/AAAI/article/view/8090 (accessed Nov. 1, 2023).

[28] I. Kandel and M. Castelli, "Transfer learning with convolutional neural networks for diabetic retinopathy image classification. A Review," Applied Sciences, vol. 10, no. 6, p. 2021, 2020