Scientific Journal of Informatics
Vol. 11, No. 3, Aug 2024
p-ISSN 2407-7658 https://journal.unnes.ac.id/journals/sji/index e-ISSN 2460-0040

Optimizing Deep Learning Models with Custom ReLU for Breast Cancer Histopathology Image Classification

Wahyu Adi Nugroho[1], Catur Supriyanto[2]*, Pujiono[3], Guruh Fajar Shidik[4]
[1,2,3,4]Master Program in Informatics Engineering, Universitas Dian Nuswantoro, Indonesia

Abstract.
Purpose: The prompt identification of breast cancer is crucial in preventing the considerable damage inflicted by this dangerous form of cancer, which is widely happened across the globe. This study seeks to refine the efficacy of a deep learning-driven approach for the precise diagnosis of breast cancer by employing diverse bespoke Rectified Linear Units (ReLU) to improve the model's performance and reduce inaccuracies within the system.
Method: This study focuses on analyzing a deep learning approach utilizing the BreakHis dataset with 7,909 images, incorporating changes to the ReLU activation function across different pre-trained CNN models. It then evaluates performance through measurement such as accuracy, precision, recall, and F1-Score.
Result: Based on our experiment results, it can be shown that the DenseNet201 models with a custom LeakyReLU excel beyond the typical ReLU, achieving the highest accuracy, recall, and F1-Score at 99.21%, 99.21%, and 99.11%, respectively. Simultaneously, ResNet152, utilizing LessNegativeReLU (α=0.05), achieved the highest precision at 99.11%. The VGG11 model exhibited the most notable performance enhancement, with improvements ranging from 1.39% to 1.59%.
Novelty: The research is original in optimizing a model for accurate breast cancer diagnosis. The proposed model is superior to the model utilizing the default activation function. This finding indicates that the study significantly enhances performance while effectively minimizing errors, thereby necessitating further exploration into the effectiveness of the customized activation function when applied to other medical imaging modalities.

Keywords: Activation function, Breast cancer, Convolutional neural network, Custom ReLU, Deep learning
Received August 2024 / **Revised** October 2024 / **Accepted** November 2024

INTRODUCTION

Breast cancer is classified as one of the most threatening kinds of cancer worldwide due to its high annual mortality rate [1]. This is evidenced by the fact that the incidence of female deaths caused by this cancer is 25%, or approximately 4.4 million, based on the International Agency for Research on Cancer (IARC) in 2020 [2]. However, not all types of breast cancer have the potential to be harmful. There are at least three common types of breast cancer, namely: benign, in situ, and invasive carcinoma. Breast cancer of the benign kind is classified as a benign type since it just alters the anatomy of the breast [3]. On the other hand, in situ carcinoma breast cancer will affect the lobular duct system and will not spread to other systems [4]. Both types of cancer can be considered relatively benign and can be promptly treated [5]. The most aggressive type of breast cancer that has the most severe impact is invasive carcinoma, or malignant [6]. This kind can rapidly spread to other organs if not detected early. Rapid screening and precise diagnosis of breast cancer have a crucial role in reducing mortality rates caused by cancer [7]. Typically, doctors diagnose cancer by identifying the features and characteristics of the cancer using various medical imaging techniques. One often-used procedure is histopathological imaging. Histopathology is a diagnostic procedure that involves the examination of intact tissue samples obtained through biopsy or surgery using a microscope [8]. However, the process requires a considerable amount of time, as doctors need to do more analysis to accurately detect the presence of cancer. In addition, doctors also have time constraints when it comes to diagnosing and treating patients, due to the increasing number of cases they handle [9]. To address this issue, a smart system is required that can automatically and accurately diagnose breast cancer, thereby assisting doctors in rapidly detecting cancer.

*Corresponding author.
Email addresses: catur.supriyanto@dsn.dinus.ac.id (Supriyanto)
DOI: 10.15294/sji.v11i3.12722

During the past decade, several technologies have rapidly developed, particularly artificial intelligence (AI) technology, which can revolutionize various aspects of life, such as humans, businesses, society, and the environment [10]. Supported by the availability of powerful hardware for computational processes, this technology may be employed as an effective approach for handling the analysis of thousands of medical images in a short amount of time. Computer vision is a field of artificial intelligence that is capable of imitating human visual perception [11]. The prevalent methodology in this domain is the application of deep learning. The deep learning method has an advantage as it can automatically extract important features from given data [12]. There are several deep learning models, including artificial neural networks (ANN), recurrent neural networks (RNN), and convolutional neural networks (CNN) [13]. However, the CNN model is the most commonly used deep learning model for image classification tasks [14]. In the following paragraphs, we present studies using CNN-based methods for breast cancer classification that demonstrate excellent performance with an average accuracy of over 90%. The CNN model is composed of many layers, including a convolutional layer, a pooling layer, and output layers or fully connected layers [15]. The utilization of convolutional and pooling layers makes it possible to extract important features from breast cancer images.

Although deep learning methods, especially CNN, have many benefits and are commonly used for image classification, there is a continuous requirement to improve their performance in the processing of medical images. Multiple studies have been undertaken to improve the ability of the CNN model, particularly for breast cancer classification. Mahati et. al. in 2023 [16] introduced a method to improve the CNN model by using a hybrid CNN-LSTM approach using transfer learning. The objective was to accurately categorize the many kinds and subtypes of breast cancer disease into two distinct classes and eight categories in the BreakHis dataset. The results of the proposed model are contrasted with cutting-edge CNN models including Inception, ResNet50, and VGG16. The models are constructed leveraging three distinct optimizers, such as Adam optimizer, RMSProp optimizer, and SGD optimizer, with different variations of the number of epochs. The suggested model attained an overall accuracy of 99% and 92.5% for binary and multi-class classification, respectively, utilizing the Adam optimizer. Another study conducted by Vandana Kumari and Rajib Ghosh in 2023 [17] also proposed a breast cancer histopathology image classification approach based on transfer learning. Three CNN-based architectures, namely DenseNet201, VGG16, and Xception, are employed as baseline models in the transfer learning approach. The model was evaluated using two public datasets, namely the IDC dataset and BreakHis, achieving the best accuracy of 99.42% and 99.12% in the case of binary classification, respectively. In the same year, 2023, Emmanuel et. al. [18] also conducted a study to classify breast cancer using the CNN method. Instead of starting with a model with randomly assigned weights, this study employs transfer learning to construct four pre-trained models, namely DenseNet201, MobileNetV2, ResNet50, and ResNet101. The dataset is divided into a 70% and 30% ratio for the training and testing datasets, respectively. Based on the experiment, the DenseNet201 model achieved the highest accuracy of 91.37% and a sensitivity of 100% with an image magnification factor of 200x for binary classification.

In 2024, Faisal et. al. [19] carried out a research study to enhance breast cancer image classification performance by employing self-supervised contrastive and transfer learning methods on the BreakHis dataset. The first approach is to expand the training data from a small subset to achieve a solid model with little data. The second method focuses on improving the ResNet50 and Inception architectures to obtain a compact and efficient classification model. The model achieved the highest accuracy performance of 98% on images enlarged by 40x and 200x magnification factors for binary classification. In contrast to previous studies that consistently utilized transfer learning, several other studies focus on modifying the activation function used in the model's architecture [20]–[23]. An activation function is a crucial component in all models based on deep learning [24]. The models facilitate the creation of non-linear abstract representations by performing consecutive linear transformations [25]. As neural networks become more complex, the significance of activation functions becomes increasingly apparent, highlighting their essential role in enabling these networks to understand and represent intricate patterns in data. The Rectified Linear Unit (ReLU) is a default activation function and is commonly employed in the hidden layers of neural networks [26]. This activation function is more time-saving in terms of computational resources than Tanh and Sigmoid functions since it utilizes simple mathematical processes [27]. However, the ReLU has several limitations, referring to the issue of the gradient diminishing to zero and the dying ReLU problem [28]. The ReLU reliably eliminates negative values. Hence, the gradient of these units will be zero, indicating that there will be no weight update during backpropagation.

Based on previous research, several researchers have utilized the advantages of the transfer learning method in training CNN-based models for the classification of breast cancer. The transfer learning method has proven to be superior by achieving sufficiently high performance. However, this method has a drawback concerning the mismatch between the pre-trained model used and the target domain, which is breast cancer [29]. The pre-trained models are often trained using the ImageNet dataset, which consists of 1000 classes with data characteristics that are very different from the breast cancer dataset. Interestingly, several other studies have focused on modifying the activation function within the CNN architecture to enhance model performance and address issues with the default ReLU activation function. Therefore, this study primarily aims to optimize the CNN model's performance in classifying breast cancer histopathology images by fine-tuning and modifying the ReLU activation function with a custom activation function called LessNegativeReLU. Furthermore, this research will compare the performance of models trained with default and custom activation functions to demonstrate the proposed method's improved performance of the models.

METHODS

This research comprises multiple stages to meet its goals, as illustrated in Figure 1. The preliminary phase involves gathering the dataset, which is central to the investigation. This dataset is subsequently divided into three distinct sections with defined ratios. The sets used for training and validation help in developing the model, whereas the testing set is designated for assessing how well the model performs. The research employs a CNN-based model due to its effectiveness in the automatic extraction of features from images, thereby eliminating the necessity for manual feature engineering that is characteristic of conventional techniques. Several pre-trained architectures, namely DenseNet, ResNet, and VGG, are set to be used. To achieve the research objectives, an innovative methodology for adapting the ReLU activation function for neurons responding to negative inputs is introduced. This model's categorization system will classify the data into two groups: benign and malignant, signifying non-cancerous and cancerous breast tumors, respectively. Finally, the performance of the model will be scrutinized through several evaluative metrics, such as accuracy, precision, recall, and F1-Score.

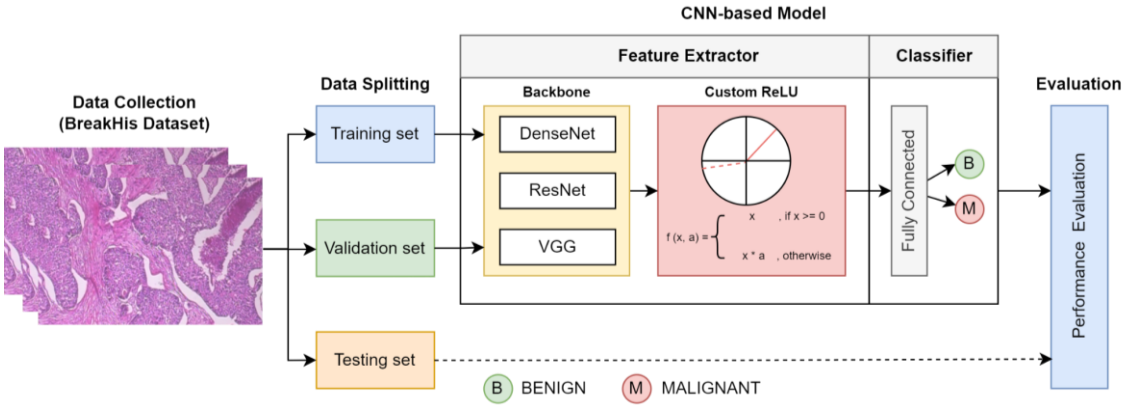


Figure 1. The proposed method

Data collection

This study will employ a publicly available dataset, specifically the BreakHis[2] dataset. This dataset comprises 7,909 microscopic images of breast carcinoma cells collected from 82 patients [30]. The images were taken at several degrees of magnification, including 40X, 100X, 200X, and 400X. The original dataset has 2,480 and 5,429 images for benign and malignant tumor classes, respectively. Benign classes refer to a lesion that does not exhibit any indications of malignancy, such as notable cellular abnormalities, mitosis, or disruption of basement membranes. Benign tumors are usually innocuous, characterized by gradual development, and limited to a single location. Meanwhile, a malignant tumor is a form of abnormal growth that possesses the capacity to invade and annihilate neighboring structures (locally invasive) and disseminate to remote sites, finally resulting in mortality.

[2] https://web.inf.ufpr.br/vri/databases/breast-cancer-histopathological-database-breakhis/

The samples are depicted as pictures of 700 x 460 pixels. The PNG file format contains images with three color RGB channels, each having a depth of 8 bits per channel. This database was established through a collaborative effort with the P&D Laboratory for Pathological Anatomy and Cytopathology, located in Parana, Brazil. Previous studies have extensively utilized this dataset to construct an automated breast cancer diagnostic algorithm. Figure 1 illustrates the distribution of benign and malignant tumor classes after filtering. From the image, it can be deduced that this dataset is categorized as an unbalanced dataset. Meanwhile, Figure 2 displays five samples from each class.

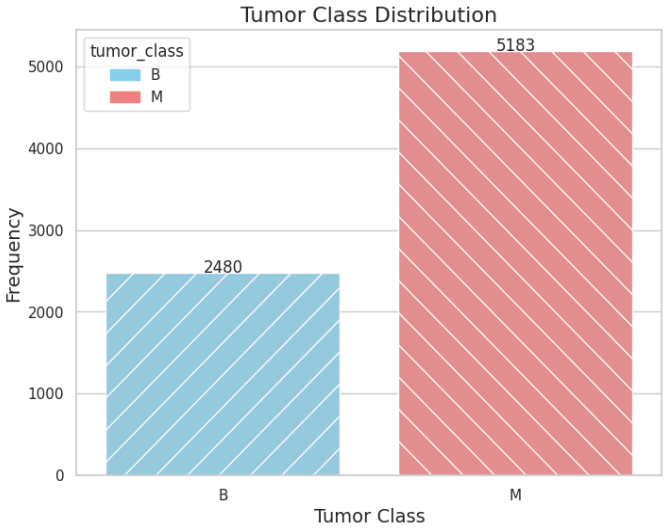


Figure 2. The BreakHIS dataset distribution

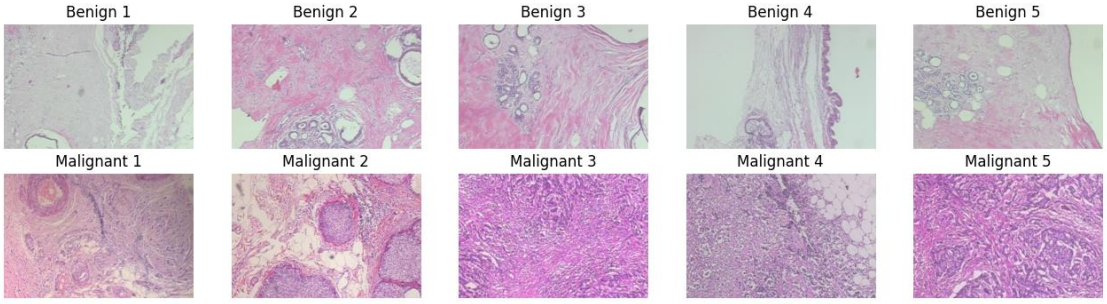


Figure 3. Examples of breast cancer histopathology images (top: benign; bottom: malignant)

Data filtering and splitting

The next step entails the process of filtering and dividing the dataset. The study [31] shows that images related to patient ID:13412 have been duplicated in two malignant sub-categories, namely ductal (DC) and lobular (LC) carcinoma. This replication has the potential to mislead the classifier. Therefore, we have applied a filtering process to these images in the split file that was created. Table 1 displays the distribution of the dataset before and after filtering. The total dataset ready for dataset splitting is 7,663. Proper dataset partitioning is essential for developing a breast cancer classification model. The dataset will consist of three components: the training set, the validation set, and the testing set, with proportions of 70%, 20%, and 10%, respectively. The amount of each training, validation, and testing set used in this study is shown in Table 2. To mitigate overfitting and improve the accuracy of the model, it is necessary to employ data augmentation, which is a technique that expands the diversity of the training data [32]. By only applying augmentation to the training set, the model has enhanced generalization skills, allowing it to make more successful predictions on unseen data. Table 3 below provides a concise overview of the augmentation techniques employed in this research. These techniques involve resizing images to dimensions of 256 x 256 pixels, extracting the central portion of the image through a center crop, applying horizontal and vertical flips, and normalizing the dataset using mean and standard deviation (std) values adapted from the BreakHis dataset.

Table 1. Data distribution before and after filtering

Tumor Class	Number of Images	
	Before (Original)	After Filtering
Benign	2,480	2,480
Malignant	5,429	5,183
Total Images	7,909	7,663

Table 2. The amount of each type of data used in this study

Tumor Class	Data			Number of Images
	Training	Validation	Testing	
Benign	1,730	500	250	2,480
Malignant	3,631	1,034	518	5,183
Total Images	5,361	1,534	768	7,663

Table 3. The augmentation method and the value of its parameter

Augmentation	Value
Resize	256 x 256
CenterCrop	224 x 224
RandomHorizontalFlip	True, p=0.5
RandomVerticalFlip	True, p=0.5
Normalize	mean = [0.7862, 0.6261, 0.7654]; std = [0.1065, 0.1396, 0.0910].

CNN-based model

The research utilized three main pre-trained CNN models, including densely convolutional networks (DenseNet), residual networks (ResNet), and visual geometry groups (VGG). DenseNet is a CNN-based model designed by researchers from Facebook AI Research (FAIR) in 2017 [33]. This architecture has several advantages in addressing the issues of weakened gradients and information loss that often occur in conventional neural networks. In DenseNet networks, each layer in the network is directly connected to all previous layers [34]. This means that the result produced by each layer will serve as the input for all subsequent layers, forming dense patterns of connections and integrating information from all previous layers. This results in a highly interconnected architecture and creates a "convergence" effect of information from the whole network. Figure 4 below shows the DenseNet architectures.

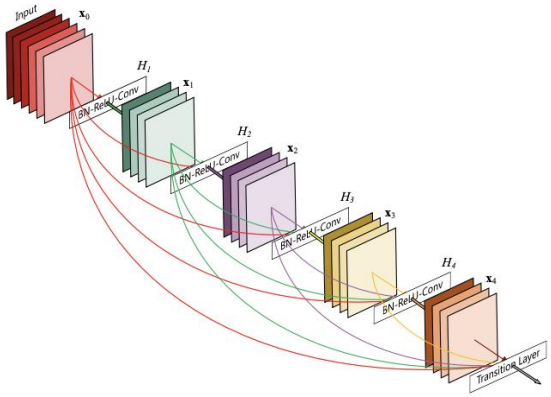


Figure 4. DenseNet architectures [33]

ResNet is a convolutional neural network (CNN) architecture that was released in 2015. It consists of many variants with different variations [35]. ResNet50 is a convolutional neural network (CNN) model that consists of 50 layers. These layers include 48 convolutional layers, 1 pooling layer, and 1 average pooling layer. The purpose of ResNet50 is to effectively handle various computer vision problems. On the other hand, ResNet152, a deeper variant of this architecture, features 152 layers organized into residual blocks, allowing the model to manage more complex tasks by learning deeper and more detailed features, though it requires greater computational resources and training time. The residual block concept across all ResNet variants enables the network to stack layers effectively without compromising performance, addressing the vanishing gradient issue often seen in deep neural networks, thus making ResNet a highly effective architecture for various computer vision applications. Figure 5 illustrates the architectural building blocks for ResNet50, ResNet101, and ResNet152.

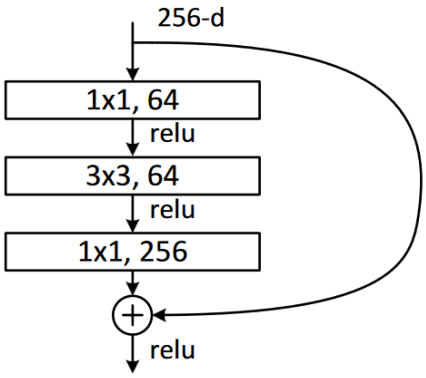


Figure 5. ResNet architectures [35]

VGG is another CNN architecture created by the Visual Geometry Group at Oxford University [36]. It is well-known for its deep structure and straightforward design. The design utilizes a sequence of 3x3 convolutional layers and 2x2 pooling layers, constructed successively to gradually decrease the spatial dimensions of the feature maps. VGG networks are distinguished by their use of compact convolutional filters, which augment their capacity to capture intricate information in pictures. Although VGG models are highly successful, they demand substantial computing resources, rendering them potent yet resource-intensive for image recognition applications. Figure 6 illustrates the VGG architecture.

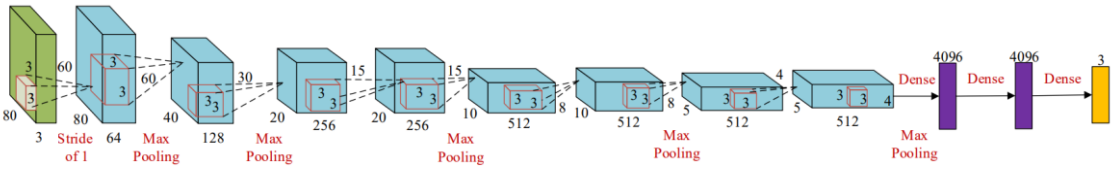


Figure 6. VGG architectures [37]

Custom ReLU Activation Functions

Based on the previous discussion, ReLU has several limitations, including the vanishing gradient problem and the dying ReLU problem [28]. While ReLU effectively eliminates negative values, this leads to a gradient of 0 for these units, resulting in no weight updates during backpropagation. To address these issues, this study proposes a custom activation function to change the standard ReLU in the neural network's hidden layers. This modification is intended to assist the model in acquiring a deeper understanding of intricate patterns and improve its overall performance. Figure 7 illustrates the default ReLU activation function, while Figures 8 (a) and 8 (b) show custom ReLU, namely LeakyReLU and LessNegativeReLU, respectively. LeakyReLU assigns a small slope to negative inputs to prevent the dying ReLU problems, while LessNegativeReLU, proposed in this study, scales negative inputs by a smaller factor to retain more information from these values and enhance pattern recognition. The following equations present the equations for each activation function.

$$ReLU\,(x) = \begin{cases} x & if\ x \geq 0 \\ 0 & otherwise \end{cases} \quad (1)$$

$$LeakyReLU\,(x) = \begin{cases} x & if\ x \geq 0 \\ 0.01 * x & otherwise \end{cases} \quad (2)$$

$$LessNegativeReLU\,(x, \alpha) = \begin{cases} x & if\ x \geq 0 \\ x * \alpha & otherwise \end{cases} \quad (3)$$

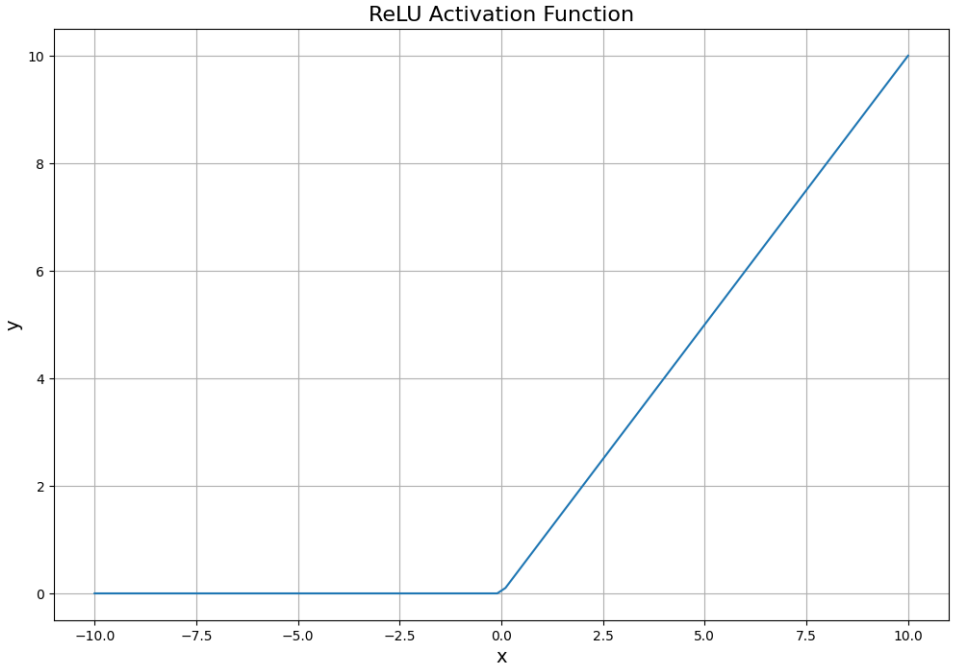
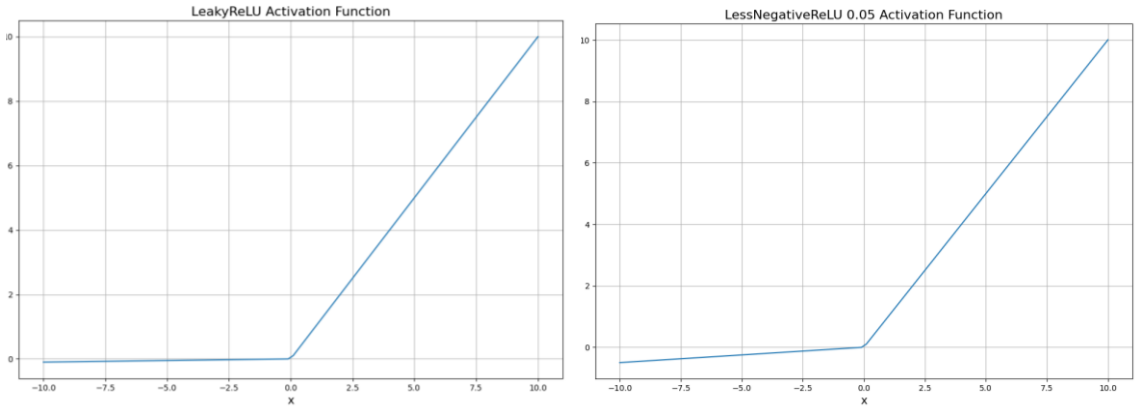


Figure 7. The ReLU activation functions



a) LeakyReLU b) LessNegativeReLU

Figure 8. The custom ReLU activation function (left: LeakyReLU; right: LessNegativeReLU)

Model training

To develop the breast cancer classification model, a variety of hyperparameter configurations are necessitated, which encompass batch size, epochs, learning rate, optimizer, and scheduler, as delineated in Table 4 presented below. The batch size signifies the quantity of samples that will be processed before the modification of the model. The total number of epochs delineates the aggregate number of instances in which the entire dataset is traversed during the training procedure. The learning rate regulates the extent of the increment during the procedure of modifying the parameters of the model. The appropriate learning rate value impacts how quickly the model achieves satisfactory convergence. The most often used optimizer is adaptive moment estimation (Adam), known for its balance between speed and convergence accuracy. A scheduler is employed to dynamically adjust the learning rate throughout the training regimen, thereby facilitating the model's optimal performance. In this study, all experimental tasks, both model training and inference process (model making predictions), were performed on the Ubuntu 22.04.3 operating system, making use of Python 3.10.12 and PyTorch 2.0.1, on a powerful NVIDIA DGX A100 supercomputing setup featuring 20GB of GPU memory. These advanced specifications are crucial for supporting the deep learning research conducted, particularly in training models that process thousands of digital images. Table 5 below shows the experimental scenarios used in this study.

Table 4. Hyperparameters for model training

Hyperparameters	Value
Batch Size	32
Epochs	20
Learning Rate	0.0001 (1e-4)
Optimizer	Adam
Scheduler	StepLR (with step size = 0.1 and gamma value = 0.8)

Table 5. Experiment scenarios

Scenarios	Value
Custom Activation Functions	ReLU (default), LeakyReLU ($\alpha = 0.01$), LessNegativeReLU (with α = [0.03, 0.05, 0.07, 0.09])
CNN-based Model	DenseNet121, Dense201, ResNet50, ResNet152, VGG11

Model evaluation

The investigation evaluated efficacy by quantifying significant parameters, encompassing accuracy, precision, recall, and the F1 score. The accuracy metric helps in evaluating a model's performance by analyzing the count of accurately recognized true positives (TP) alongside true negatives (TN). Precision appraises the model's capability to accurately predict positive outcomes by quantifying the ratio of authentic positives to all forecasted positives. Recall constitutes a quantitative assessment of a model's proficiency in accurately recognizing pertinent instances. It is computed by determining the percentage of actual positives that are suitably identified. The F1 score, calculated as the harmonic mean of accuracy and recall, affords a balanced evaluation of classification performance, which is particularly advantageous for addressing class imbalances. The metrics are elucidated in equations 4 to 7.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

$$F1 - Score = \frac{2 * precision * recall}{precision + recall} \quad (7)$$

RESULTS AND DISCUSSIONS

This study follows a previously established methodology, beginning with the preparation and collection of a breast cancer dataset, which is subsequently divided into three sets. Data augmentation is applied exclusively to the training set to mitigate overfitting and improve the model's capacity for generalization. Various custom ReLU activation functions are implemented within a CNN-based model to explore multiple scenarios aimed at optimizing performance and minimizing future diagnostic errors. The model training process is then conducted on a supercomputer server to accelerate both training and evaluation. Fine-tuning is performed on the CNN model, adjusting all layers of the model without freezing the networks, allowing it to better adapt to the specific dataset used in this study. Upon completion of the training, the model's performance in breast cancer classification is assessed using the testing set. The evaluation results are analyzed to determine the influence of the custom activation functions on the model's performance, with the ultimate goal of reducing errors and improving accuracy in early breast cancer diagnosis. Table 6 presents the performance outcomes of the custom ReLU activation functions across various pre-trained models. The results indicate that the suggested method for optimizing the breast cancer model with a custom ReLU activation function has yielded encouraging outcomes, exceeding the performance of the baseline model that used the default ReLU activation function. This demonstrates the efficacy of the proposed method in improving the model's performance.

Table 6. Performance of the custom ReLU on several pre-trained models

CNN-based Model	Activation Function	Evaluation Metrics (%)			
		Accuracy	Precision	Recall	F1-Score
DenseNet121	ReLU (default)	97.62	97.62	97.62	97.62
	LeakyReLU ($\alpha = 0.01$)	97.93	97.36	97.93	97.64
	LessNegativeReLU ($\alpha = 0.03$)	**98.62**	98.42	**98.62**	98.52
	LessNegativeReLU ($\alpha = 0.05$)	98.53	97.94	98.53	98.23
	LessNegativeReLU ($\alpha = 0.07$)	98.51	**98.81**	98.51	**98.66**
	LessNegativeReLU ($\alpha = 0.09$)	98.62	98.42	98.62	98.52
DenseNet201	ReLU (default)	99.11	99.00	99.21	99.11
	LeakyReLU ($\alpha = 0.01$)	**99.21**	**99.01**	**99.21**	**99.11**
	LessNegativeReLU ($\alpha = 0.03$)	98.52	98.22	98.52	98.37
	LessNegativeReLU ($\alpha = 0.05$)	99.01	98.91	99.01	98.96
	LessNegativeReLU ($\alpha = 0.07$)	98.62	98.13	98.62	98.37
	LessNegativeReLU ($\alpha = 0.09$)	98.92	98.43	98.92	98.67
ResNet50	ReLU (default)	98.62	98.42	98.62	98.52
	LeakyReLU ($\alpha = 0.01$)	98.04	97.02	98.04	97.50
	LessNegativeReLU ($\alpha = 0.03$)	98.82	**98.52**	98.82	98.66
	LessNegativeReLU ($\alpha = 0.05$)	**99.02**	98.62	**99.02**	**98.81**
	LessNegativeReLU ($\alpha = 0.07$)	97.73	97.53	97.73	97.63
	LessNegativeReLU ($\alpha = 0.09$)	98.51	98.51	98.51	98.51
ResNet152	ReLU (default)	98.11	98.31	98.11	98.21
	LeakyReLU ($\alpha = 0.01$)	98.62	98.42	98.62	98.52
	LessNegativeReLU ($\alpha = 0.03$)	98.23	97.65	98.23	97.93
	LessNegativeReLU ($\alpha = 0.05$)	**98.80**	**99.11**	**98.80**	**98.89**
	LessNegativeReLU ($\alpha = 0.07$)	98.02	97.83	98.02	97.92
	LessNegativeReLU ($\alpha = 0.09$)	98.23	97.65	98.23	97.93
VGG11	ReLU (default)	96.83	96.93	96.83	96.88
	LeakyReLU ($\alpha = 0.01$)	98.13	97.74	98.13	97.93
	LessNegativeReLU ($\alpha = 0.03$)	**98.42**	**98.32**	**98.42**	**98.37**
	LessNegativeReLU ($\alpha = 0.05$)	98.13	97.74	98.13	97.93
	LessNegativeReLU ($\alpha = 0.07$)	97.82	98.02	97.82	97.92
	LessNegativeReLU ($\alpha = 0.09$)	97.14	96.67	97.14	96.90

Based on Table 6 above, it can be stated that the optimal model is DenseNet201 with a bespoke LeakyReLU activation function. This model outperforms other models with the highest accuracy of 99.21%, recall of 99.21%, and F1-Score of 99.11%. The model with the highest precision is ResNet152, with a custom activation function called LessNegativeReLU (with α= 0.05), achieving 99.11%. Overall, all models trained using unique activation functions, including LeakyReLU and LessNegativeReLU with various alpha values, outperformed the model trained with the regular ReLU activation function. This demonstrates the effective improvement of the deep learning model's performance in classifying breast cancer images. In the DenseNet121 model, the performance improvement between the baseline model that utilizes ReLU as a default activation function and the model with its custom, which achieved the highest performance in all metrics, ranges from 1% to 1.04%. Meanwhile, in DenseNet201, the improvement in model performance is not very significant, ranging from 0% to 0.11%. The performance improvement with the ResNet50 model ranges from 0.1% to 0.4%. The performance improvement with the ResNet152 model ranges from 0.68% to 0.8%. The VGG11 model had the highest performance improvement, ranging from 1.39% to 1.59%.

Figure 9 below displays the plotting of training and validation accuracy for the best model, namely DenseNet201 with LeakyReLU custom activation functions. The graph indicates that the model performs exceptionally well, with no signs of overfitting. It is good benefit of the effectiveness of the applied data augmentation. Furthermore, we provide a visual illustration of the confusion matrix to measure the reliability of the model in accurately categorizing data and detecting classification mistakes, as seen in Figure 10. The confusion matrix aids in analyzing the performance of a model in each class, allowing for clear identification of where the model makes errors and the accuracy of its predictions. Based on the confusion matrix results, the model correctly predicted 248 benign cases and 514 malignant cases. It misclassified 2 benign cases as malignant and 4 malignant cases as benign. These results show that the model is highly accurate in classifying breast cancer images, which is important for helping doctors provide the right treatment for benign or malignant cases.

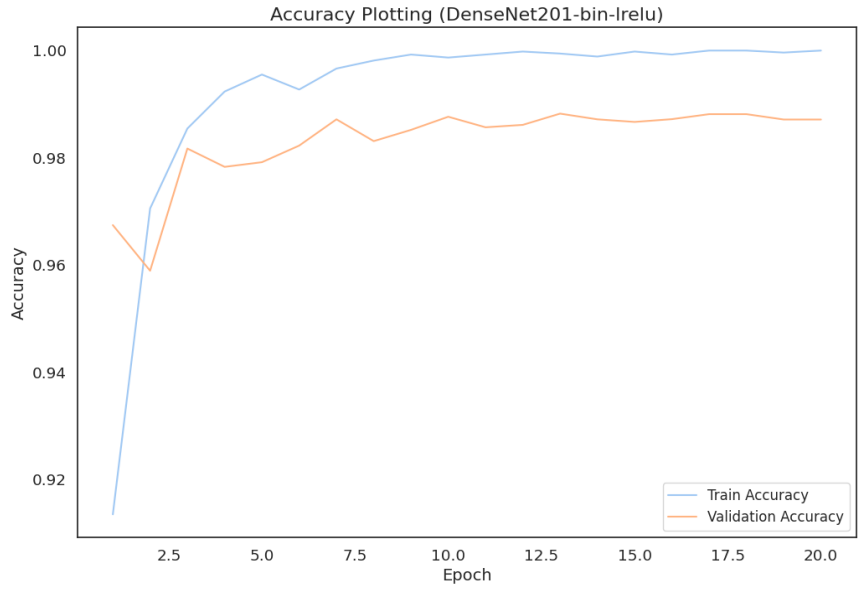


Figure 9. The plotting of training and validation accuracy on the best models

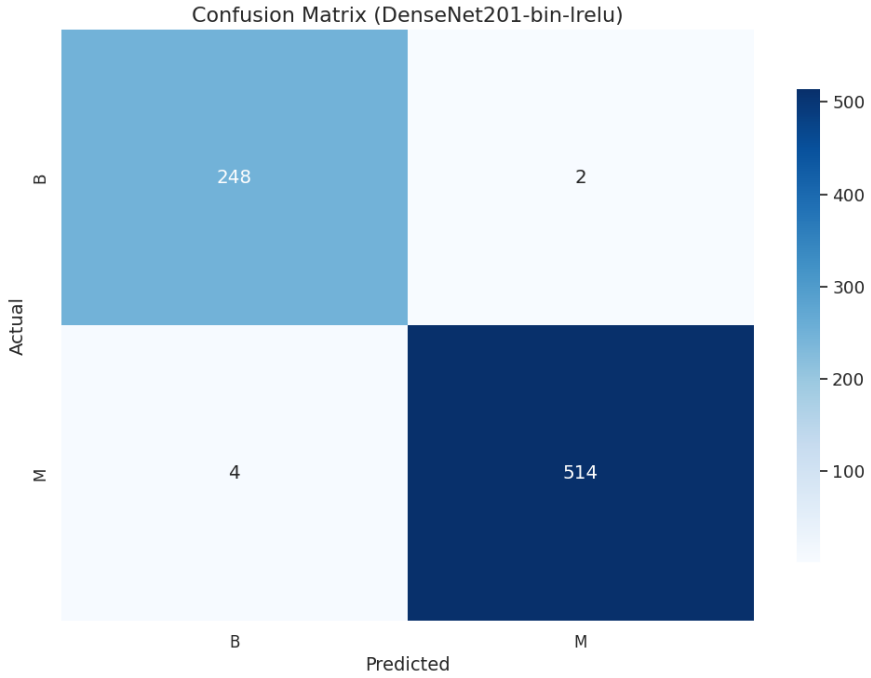


Figure 10. The plotting of the confusion matrix on the best models

To create a breast cancer diagnostic system that is both precise and efficient, this study not only measures the performance of several models using various metrics but also examines the time it takes for training and inference, as shown in Table 7 below. Our objective is to analyze how model size, complexity, and the use of custom activation functions influence the time required for training. By measuring the duration of the training process, our objective is to comprehend the influence of these factors on training efficiency and identify the optimal configurations that balance performance with training time. This technique enables us to select models that not only deliver optimal performance but also demonstrate efficiency in both the training and inference steps. Knowing the training duration for each model enables us to assess their effectiveness, guaranteeing that we choose designs that are not only precise but also quick and resource-efficient. This eventually improves the usability and scalability of the diagnostic system.

Table 7. Runtime performance on several pre-trained models

CNN-based Model	Activation Function	Runtime in Seconds (s)	
		Training Time	Inference Time
DenseNet121	ReLU (default)	546.23	2.34
	LeakyReLU ($\alpha = 0.01$)	762.63	2.43
	LessNegativeReLU ($\alpha = 0.03$)	754.79	2.30
	LessNegativeReLU ($\alpha = 0.05$)	752.70	2.28
	LessNegativeReLU ($\alpha = 0.07$)	754.45	2.33
	LessNegativeReLU ($\alpha = 0.09$)	754.92	2.33
DenseNet201	ReLU (default)	811.13	2.99
	LeakyReLU ($\alpha = 0.01$)	1,159.97	3.07
	LessNegativeReLU ($\alpha = 0.03$)	1,142.03	2.74
	LessNegativeReLU ($\alpha = 0.05$)	1,134.26	2.93
	LessNegativeReLU ($\alpha = 0.07$)	1,137.47	2.91
	LessNegativeReLU ($\alpha = 0.09$)	1,138.10	2.96
ResNet50	ReLU (default)	436.17	2.09
	LeakyReLU ($\alpha = 0.01$)	561.93	2.10
	LessNegativeReLU ($\alpha = 0.03$)	566.25	2.11
	LessNegativeReLU ($\alpha = 0.05$)	576.38	2.39
	LessNegativeReLU ($\alpha = 0.07$)	564.81	2.07
	LessNegativeReLU ($\alpha = 0.09$)	572.21	2.24
ResNet152	ReLU (default)	847.54	2.82
	LeakyReLU ($\alpha = 0.01$)	1,140.92	3.26
	LessNegativeReLU ($\alpha = 0.03$)	1,138.58	3.20
	LessNegativeReLU ($\alpha = 0.05$)	1,133.73	3.06
	LessNegativeReLU ($\alpha = 0.07$)	1,131.15	3.00
	LessNegativeReLU ($\alpha = 0.09$)	1,131.43	2.98
VGG11	ReLU (default)	532.73	2.61
	LeakyReLU ($\alpha = 0.01$)	636.90	2.85
	LessNegativeReLU ($\alpha = 0.03$)	648.10	3.03
	LessNegativeReLU ($\alpha = 0.05$)	642.44	2.78
	LessNegativeReLU ($\alpha = 0.07$)	671.92	3.14
	LessNegativeReLU ($\alpha = 0.09$)	706.06	3.61

Based on Table 7 above, it can be generally argued that larger models require more training time compared to smaller ones. Additionally, models utilizing custom ReLU activation functions tend to have longer training durations than their baseline counterparts. This suggests that while the proposed methods deliver improved performance, they also incur increased training times, highlighting the need to balance performance gains with training efficiency. The model with the shortest training time is the baseline ResNet50, which requires 436.17 seconds, whereas DenseNet201 with Custom LeakyReLU has the longest training time at 1,159.97 seconds. Notably, ResNet50 not only excels in training speed but also achieves the fastest inference time, processing approximately 768 test images in just 2.07 seconds, or about 2.695 milliseconds (ms) per image. This demonstrates that ResNet50 is highly suitable for real-time breast cancer diagnosis systems, providing both efficient performance and high accuracy. Even the model with the highest performance still maintains a rapid inference time, processing images in 3.07 seconds, or approximately 3.997 ms per image. These results make the model suitable for developing an accurate and efficient breast cancer diagnosis assistant system to help doctors make the initial diagnoses.

Additionally, we also compared our approach with previous research that aligns with the goals of this study, which uses the same datasets. As shown in Table 8, our method outperforms earlier studies in all metrics like accuracy, precision, recall, and F1-Score. These improvements indicate that our model is more effective in diagnosing breast cancer. Higher accuracy means fewer misclassifications, while better precision and recall show the model can correctly identify both benign and malignant cases. The improved F1-Score reflects a satisfactory balance between precision and recall, making the model reliable. Overall, these results highlight the strength of our approach in addressing the limitations of previous methods and offer a more effective system for breast cancer classification. This success not only surpasses existing solutions but also has the potential to enhance intelligent diagnostic systems, leading to more accurate early diagnoses and better treatment decisions for patients.

Table 8. A comparative analysis of performance with previous research

References	Year	Method	Evaluation Metrics (%)			
			Accuracy	Precision	Recall	F1 Score
[38]	2021	ResNet50 + Augmentation	99.01	-	-	-
[39]	2021	Intrinsic Feature Learning and GCN	98.42	-	-	-
[40]	2022	ResNet50 + Image Sharpening	95.00	-	-	-
[41]	2022	Hybrid CNN	86.55	-	-	79.19
[42]	2023	ResNet101 + Augmentation	96.09	96.26	96.09	96.08
[17]	2023	Transfer learning using DCNN	99.12	-	-	-
[43]	2024	Convolutionally-Enhanced ViT	89.43	-	-	-
Ours	**2024**	**DenseNet201 + Custom ReLU**	**99.21**	**99.01**	**99.21**	**99.11**

CONCLUSION

This paper presents a comprehensive analysis of optimizing deep learning models through customized ReLU activation functions to improve breast cancer histopathology image classification. It involves a comparative evaluation of various CNN-based models, including DenseNet121, DenseNet201, ResNet50, ResNet152, and VGG11, utilizing the public BreakHis dataset, which consists of two classes: benign and malignant. The research addresses challenges associated with vanishing gradients and dying ReLU problems by implementing a custom ReLU approach across all models. The results show that DenseNet201 with custom LeakyReLU achieved the highest accuracy, recall, and F1 score of 99.21%, 99.21%, and 99.11%, respectively. Meanwhile, ResNet152 with custom LessNegativeReLU ($\alpha = 0.05$) achieved the best precision of 99.11%. VGG11 exhibited the most significant performance improvement over the ReLU baseline, with the increase that ranging from 1.39% to 1.59%. These findings substantiate the effectiveness of the proposed method in enhancing the performance of deep learning models for breast cancer classification. However, despite these promising results, the proposed method leads to longer training times compared to the baseline models. Future research should focus on optimizing training efficiency through techniques such as adjustments to model architecture or more efficient optimization algorithms. Additionally, further investigations are necessary to evaluate the benefit of this custom activation function on various medical image analysis tasks.

ACKNOWLEDGEMENT

The author extends deep appreciation to the Ministry of Education, Culture, Research, and Technology for funding this research through the Master's Thesis Research Scheme in 2024.

REFERENCES

[1] F. S. Lobato, J. E. Alamy Filho, G. B. Libotte, and G. M. Platt, "Optimizing breast cancer treatment using hyperthermia: A single and multi-objective optimal control approach," *Appl. Math. Model.*, vol. 127, pp. 96–118, Mar. 2024, doi: 10.1016/j.apm.2023.11.022.

[2] F. Guida *et al.*, "Global and regional estimates of orphans attributed to maternal cancer mortality in 2020," *Nat. Med.*, vol. 28, no. 12, pp. 2563–2572, Dec. 2022, doi: 10.1038/s41591-022-02109-2.

[3] A. B. Nassif, M. A. Talib, Q. Nasir, Y. Afadar, and O. Elgendy, "Breast cancer detection using artificial intelligence techniques: A systematic literature review," *Artif. Intell. Med.*, vol. 127, p. 102276, May 2022, doi: 10.1016/j.artmed.2022.102276.

[4] H. Yao, X. Zhang, X. Zhou, and S. Liu, "Parallel Structure Deep Neural Network Using CNN and RNN with an Attention Mechanism for Breast Cancer Histology Image Classification," *Cancers (Basel).*, vol. 11, no. 12, p. 1901, Nov. 2019, doi: 10.3390/cancers11121901.

[5] M. Nasser and U. K. Yusof, "Deep Learning Based Methods for Breast Cancer Diagnosis: A Systematic Review and Future Direction," *Diagnostics*, vol. 13, no. 1, p. 161, Jan. 2023, doi: 10.3390/diagnostics13010161.

[6] R. Ha *et al.*, "Convolutional Neural Network Using a Breast MRI Tumor Dataset Can Predict Oncotype Dx Recurrence Score," *J. Magn. Reson. Imaging*, vol. 49, no. 2, pp. 518–524, Feb. 2019, doi: 10.1002/jmri.26244.

[7] Babita and D. R. Nayak, "RDTNet: A residual deformable attention based transformer network for breast cancer classification," *Expert Syst. Appl.*, vol. 249, p. 123569, Sep. 2024, doi: 10.1016/j.eswa.2024.123569.

[8] R. Rashmi, K. Prasad, and C. B. K. Udupa, "Breast histopathological image analysis using image processing techniques for diagnostic purposes: A methodological review," *J. Med. Syst.*, vol. 46,

no. 1, p. 7, Jan. 2022, doi: 10.1007/s10916-021-01786-9.
[9] F. Atban, E. Ekinci, and Z. Garip, "Traditional machine learning algorithms for breast cancer image classification with optimized deep features," *Biomed. Signal Process. Control*, vol. 81, p. 104534, Mar. 2023, doi: 10.1016/j.bspc.2022.104534.
[10] Y. K. Dwivedi, A. Sharma, N. P. Rana, M. Giannakis, P. Goel, and V. Dutot, "Evolution of artificial intelligence research in Technological Forecasting and Social Change: Research topics, trends, and future directions," *Technol. Forecast. Soc. Change*, vol. 192, p. 122579, Jul. 2023, doi: 10.1016/j.techfore.2023.122579.
[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.
[12] J. Liu and Y. Jin, "A comprehensive survey of robust deep learning in computer vision," *J. Autom. Intell.*, vol. 2, no. 4, pp. 175–195, Nov. 2023, doi: 10.1016/j.jai.2023.10.002.
[13] M. M. Taye, "Understanding of Machine Learning with Deep Learning: Architectures, Workflow, Applications and Future Directions," *Computers*, vol. 12, no. 5, p. 91, Apr. 2023, doi: 10.3390/computers12050091.
[14] M. Radak, H. Y. Lafta, and H. Fallahi, "Machine learning and deep learning techniques for breast cancer diagnosis and classification: a comprehensive review of medical imaging studies," *J. Cancer Res. Clin. Oncol.*, vol. 149, no. 12, pp. 10473–10491, Sep. 2023, doi: 10.1007/s00432-023-04956-z.
[15] A. Anton, N. F. Nissa, A. Janiati, N. Cahya, and P. Astuti, "Application of Deep Learning Using Convolutional Neural Network (CNN) Method For Women's Skin Classification," *Sci. J. Informatics*, vol. 8, no. 1, pp. 144–153, May 2021, doi: 10.15294/sji.v8i1.26888.
[16] M. M. Srikantamurthy, V. P. S. Rallabandi, D. B. Dudekula, S. Natarajan, and J. Park, "Classification of benign and malignant subtypes of breast cancer histopathology imaging using hybrid CNN-LSTM based transfer learning," *BMC Med. Imaging*, vol. 23, no. 1, p. 19, Jan. 2023, doi: 10.1186/s12880-023-00964-0.
[17] V. Kumari and R. Ghosh, "A magnification-independent method for breast cancer classification using transfer learning," *Healthc. Anal.*, vol. 3, p. 100207, Nov. 2023, doi: 10.1016/j.health.2023.100207.
[18] E. O. Simonyan, J. A. Badejo, and J. S. Weijin, "Histopathological breast cancer classification using CNN," *Mater. Today Proc.*, Nov. 2023, doi: 10.1016/j.matpr.2023.10.154.
[19] F. Bin Ashraf, S. M. M. Alam, and S. M. Sakib, "Enhancing breast cancer classification via histopathological image analysis: Leveraging self-supervised contrastive learning and transfer learning," *Heliyon*, vol. 10, no. 2, p. e24094, Jan. 2024, doi: 10.1016/j.heliyon.2024.e24094.
[20] L. Alzubaidi *et al.*, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," *J. Big Data*, vol. 8, no. 1, p. 53, Mar. 2021, doi: 10.1186/s40537-021-00444-8.
[21] M. Agarwal, S. Gupta, and K. K. Biswas, "A new Conv2D model with modified ReLU activation function for identification of disease type and severity in cucumber plant," *Sustain. Comput. Informatics Syst.*, vol. 30, p. 100473, Jun. 2021, doi: 10.1016/j.suscom.2020.100473.
[22] R. K. Vasanthakumari, R. V. Nair, and V. G. Krishnappa, "Improved learning by using a modified activation function of a Convolutional Neural Network in multi-spectral image classification," *Mach. Learn. with Appl.*, vol. 14, p. 100502, Dec. 2023, doi: 10.1016/j.mlwa.2023.100502.
[23] C. Közkurt, A. Diker, A. Elen, S. Kılıçarslan, E. Dönmez, and F. B. Demir, "Trish: an efficient activation function for CNN models and analysis of its effectiveness with optimizers in diagnosing glaucoma," *J. Supercomput.*, vol. 80, no. 11, pp. 15485–15516, Jul. 2024, doi: 10.1007/s11227-024-06057-1.
[24] I. Vallés-Pérez, E. Soria-Olivas, M. Martínez-Sober, A. J. Serrano-López, J. Vila-Francés, and J. Gómez-Sanchís, "Empirical study of the modulus as activation function in computer vision applications," *Eng. Appl. Artif. Intell.*, vol. 120, p. 105863, Apr. 2023, doi: 10.1016/j.engappai.2023.105863.
[25] J. Heaton, "Ian Goodfellow, Yoshua Bengio, and Aaron Courville: Deep learning," *Genet. Program. Evolvable Mach.*, vol. 19, no. 1–2, pp. 305–307, Jun. 2018, doi: 10.1007/s10710-017-9314-z.
[26] A. F. Agarap, "Deep Learning using Rectified Linear Units (ReLU)," Mar. 2018.
[27] C. Nwankpa, W. Ijomah, A. Gachagan, and S. Marshall, "Activation Functions: Comparison of trends in Practice and Research for Deep Learning," *Math. Comput. Sci.*, pp. 1–20, 2018.
[28] T. Szandała, "Review and Comparison of Commonly Used Activation Functions for Deep Neural

Networks," 2021, pp. 203–224. doi: 10.1007/978-981-15-5495-7_11.
[29] Z. Zhao, L. Alzubaidi, J. Zhang, Y. Duan, and Y. Gu, "A comparison review of transfer learning and self-supervised learning: Definitions, applications, advantages and limitations," *Expert Syst. Appl.*, vol. 242, p. 122807, May 2024, doi: 10.1016/j.eswa.2023.122807.
[30] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "A Dataset for Breast Cancer Histopathological Image Classification," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 7, pp. 1455–1462, Jul. 2016, doi: 10.1109/TBME.2015.2496264.
[31] Y. Benhammou, B. Achchab, F. Herrera, and S. Tabik, "BreakHis based breast cancer automatic diagnosis using deep learning: Taxonomy, survey and insights," *Neurocomputing*, vol. 375, pp. 9–24, Jan. 2020, doi: 10.1016/j.neucom.2019.09.044.
[32] A. Mumuni and F. Mumuni, "Data augmentation: A comprehensive survey of modern approaches," *Array*, vol. 16, p. 100258, Dec. 2022, doi: 10.1016/j.array.2022.100258.
[33] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," *Cornell Univ.*, pp. 1–9, Aug. 2016, doi: https://doi.org/10.48550/arXiv.1608.06993.
[34] A. Susanto, C. A. Sari, E. H. Rachmawanto, I. U. W. Mulyono, and N. Mohd Yaacob, "A Comparative Study of Javanese Script Classification with GoogleNet, DenseNet, ResNet, VGG16 and VGG19," *Sci. J. Informatics*, vol. 11, no. 1, pp. 31–40, Jan. 2024, doi: 10.15294/sji.v11i1.47305.
[35] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," Dec. 2015.
[36] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," 2015, doi: /10.48550/arXiv.1409.1556.
[37] D. Wang, F. Qiu, and X. Liu, "Control Method Based on Deep Reinforcement Learning for Robotic Follower with Monocular Vision," *J. Phys. Conf. Ser.*, vol. 1229, no. 1, p. 012043, May 2019, doi: 10.1088/1742-6596/1229/1/012043.
[38] Y. Yari, H. Nguyen, and T. V. Nguyen, "Accuracy Improvement in Binary and Multi-Class Classification of Breast Histopathology Images," in *2020 IEEE Eighth International Conference on Communications and Electronics (ICCE)*, Jan. 2021, pp. 376–381. doi: 10.1109/ICCE48956.2021.9352142.
[39] S. Boumaraf, X. Liu, Z. Zheng, X. Ma, and C. Ferkous, "A new transfer learning based approach to magnification dependent and independent classification of breast cancer in histopathological images," *Biomed. Signal Process. Control*, vol. 63, p. 102192, Jan. 2021, doi: 10.1016/j.bspc.2020.102192.
[40] S. V Rao and K. Raghavendra, "Breast Cancer Detection on Histopathological Data," in *2022 Fourth International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT)*, Dec. 2022, pp. 01–05. doi: 10.1109/ICERECT56837.2022.10059681.
[41] S. Singh and R. Kumar, "Breast cancer detection from histopathology images with deep inception and residual blocks," *Multimed. Tools Appl.*, vol. 81, no. 4, pp. 5849–5865, Feb. 2022, doi: 10.1007/s11042-021-11775-2.
[42] M. Isthigosah, A. Sunyoto, and T. Hidayat, "Image Augmentation for BreaKHis Medical Data using Convolutional Neural Networks," *sinkron*, vol. 8, no. 4, pp. 2381–2392, Oct. 2023, doi: 10.33395/sinkron.v8i4.12878.
[43] S. Tariq, R. Raza, A. B. Sargano, and Z. Habib, "Magnification Independent Breast Cancer Analysis Using Vision Transformer," *Multimed. Tools Appl.*, Jul. 2024, doi: 10.1007/s11042-024-19685-9.