# Convolutional Layer-Based Feature Extraction in an Ensemble Machine Learning Model for Breast Cancer Classification

Shofwatul 'Uyun[1], Lina Choridah[2], Slamet Riyadi[3], Ade Umar Ramadhan[4]
Department of Information-Faculty of Science and Technology, UIN Sunan Kalijaga, Yogyakarta, Indonesia[1, 4]
Department of Radiology-Faculty of Medicine, Universitas Gadjah Mada, Yogyakarta, Indonesia[2]
Department of Information Technology-Faculty of Engineering, Universitas Muhammadiyah Yogyakarta, Indonesia[3]

***Abstract*—Mammography and ultrasound are the main medical imaging modalities for identifying breast lesions. Computer-assisted diagnosis (CAD) is an important tool for radiologists, helping them differentiate benign and malignant lesions more quickly and objectively. The use of appropriate features in mammography and ultrasound is one of the key factors determining the success of computer-assisted diagnosis (CAD) results for breast cancer systems. The diversity of feature forms and extraction techniques is a challenge. Additionally, the use of a single classification algorithm often causes noise, bias, and is not robust. We propose a convolutional layer-based feature extraction technique in the ensemble learning model for the classification of breast cancer. This study uses 439 mammography images (203 benign, 236 malignant) and 421 ultrasound images (244 benign, 177 malignant). This research consists of several stages, including data pre-processing, feature extraction, classification, and performance evaluation. We used four convolution layer-based feature extraction techniques: simple convolution (SC), feature fusion convolution (FFC), feature fusion depthwise convolution (FFDC), and feature fusion depthwise separable convolution (FFDSC). The model uses five machine learning algorithms (support vector machine, random forest, k nearest neighbours, decision tree, and logistic regression) that are part of ensemble learning. The experimental results show that the use of the FFC convolution layer in ensemble learning has the best performance for both datasets. In the ultrasound data set, the FFC achieved a value of 0.90 in each of the accuracy, precision, recall, specificity, and F1 score metrics. In the mammography data set, the FFC achieved a value of 0.98 on each of the same metrics. These results show the effectiveness of feature fusion in improving classification performance in the soft voting classifier for ensemble learning.**

***Keywords—Ensemble learning; feature extraction; convolutional layer; breast cancer***

## I. Introduction

Breast cancer is the main cause of cancer-related mortality in women. Early detection of cancer, especially breast cancer, will contribute to the treatment process [1]. Currently, computerised tomography (CT), magnetic resonance imaging (MRI), mammography, thermal imaging, and ultrasound are common screening for breast cancer. These methods have their own unique approaches and tools, and the expected results of these methods depend on different factors, so it is recommended to validate the results using multiple methods. Although mammography is considered by many physicians and specialists the gold standard method for the detection of breast cancer, the demand for more reliable methods is increasing. Mammography and ultrasound are critical tools in breast cancer screening, but they have different applications, effectiveness, and limitations. Mammography, a form of X-ray imaging, is considered the gold standard for breast cancer screening in high-income countries because of its ability to detect cancer at an early stage. However, its effectiveness may be limited by the radiographic density of the breasts; In dense breasts, noncalcified cancers are more likely to be missed [2]. This limitation is particularly significant because the diagnostic accuracy is largely dependent on breast density and denser breasts, which pose a challenge for clear imaging [3]. Ultrasound, on the other hand, uses sound waves to create images of breast tissue [4]. It has been shown to have a high detection sensitivity, especially in younger women and those with dense breasts, where mammography may not be as effective.

Computer-assisted diagnosis (CAD) systems have shown significant achievements in improving breast cancer detection and provide complementary tools to traditional diagnostic methods [5]. The effectiveness of the CAD system is emphasized by its ability to detect and classify breast cancer with high precision, sensitivity, and specificity, as demonstrated in various studies. Recent research has shown that the precision of breast cancer detection is greatly influenced by the characteristics of mammography and ultrasound. Although mammography is widely used, its sensitivity is limited, especially in dense breast tissue, which can conceal the presence of tumours [6]. Contrast-enhanced mammography (CEM) is introduced to address some of these limitations, introducing a similar performance in the detection of mammography-occult diseases with higher sensitivity than conventional mammography and higher specificity than ultrasound [7].

The development and application of computer-aided diagnosis (CAD) systems in the detection and classification of breast cancer by mammography and ultrasound have significantly benefitted from various feature extraction methods. In mammography, several feature extraction techniques have been used in different studies. For example, an approach involves the extraction of 16 geometrical characteristics from regions of interest (ROI) in mammograms, which are then analysed using machine learning algorithms to classify mammograms into four classes [8]. Another mammography method was to calculate 271 characteristics in various categories, including shapes, textures, contrasts, and other characteristics, and to calculate additional characteristics derived from the dilated segment [9]. Some methods have been used in ultrasound. A study extracted 855 characteristics, including shapes, contours, and texture, from breast ultrasound images [10]. Another method uses the pyramid of orientated gradient descriptor histograms to obtain a characteristic vector without prior processing of tumour region selection [11]. Automated contouring and morphological analysis were used to calculate 19 practical morphological characteristics [12]. The various forms of features used affect the performance of a classification model.

Some cases that often occur in the learning process using a single classification algorithm are noise, bias, and low accuracy. This is due to the use of non-uniform data samples and the presence of overlapping classes [13]. One method of reducing these issues is to implement the concept of ensemble learning, which is designed to improve the stability and accuracy of machine learning algorithms. The concept of ensemble learning is a paradigm of learning that uses a combination of several models synergistically to improve the quality of predictions collectively [14], [15]. The objective of Ensemble learning is to reduce the bias, variance, or error that often occurs in individual models [16], [17], [18]. The main goal is to achieve predictive results that are more accurate, stable, and generally better than the results of a single model.

Furthermore, one of the keys to successful learning is the use of the right features. Feature extraction techniques have evolved [19]. In general, features consist of colour, shape, texture, and others. Some previous research related to medical imaging, especially breast cancer, uses features of colour and texture [20]. Deep feature extraction uses pre-trained convolutional neural networks such as VGG-19, SqueezeNet, ResNet-18, and GoogLeNet to distinguish between benign and malignant tumour types in ultrasound images [21],[22] Therefore, it is necessary to experiment with the use of several image feature extraction schemes using convolutional layers.

We propose the use of an ensemble learning concept for breast cancer classification using two medical images, namely mammography and ultrasound. This ensemble learning model uses five machine learning algorithms, including a supervised vector machine, random forest, nearest neighbours k, decision tree, and logistic regression. The ensemble learning model was trained using mammography and ultrasound images that have been feature extracted using four different convolution layers, including simple convolution, feature fusion convolution, feature fusion depthwise convolution, and feature fusion depthwise separable convolution. The four model schemes were evaluated for accuracy, precision, recall, ROC curve, specificity, F1 score, kappa and Matthews correlation coefficient. Some of the main contributions of this research are summarised in the following.

- We propose a novel ensemble learning model for breast cancer classification using mammography and ultrasound.
- The feature extraction technique uses four different convolution layers, namely, simple convolution (SC), feature fusion convolution (FFC), feature fusion depthwise convolution (FFDC), and feature fusion depthwise separable convolution (FFDSC).
- The ensemble learning model uses soft voting with five machine learning algorithms, namely the support vector machine, the random forest, the k closest neighbours, the decision tree, and logistic regression.
- We investigate the use of different feature extraction techniques in the ensemble learning model by evaluating its performance.

The structure of this paper consists of several sections; after the introduction in Section I, Section II describes the proposed method, for Sections III and IV explain the results and discussion on the use of the proposed method. Section V explains the conclusions and future research and ends with acknowledgments.

## II. METHODS

### A. Datasets

This study used primary data from Sardjito Hospital and Kotabaru Yogyakarta Cancer Clinic that have been identified and diagnosed by radiologists. The approval for the private data set was obtained from the Ethics Committee (Ref. No.: KE/FK/1229/EC/2023). The results of the image reading were initially classified into five class categories in BIRADS. The researchers grouped the data from the BIRADS standard into two classes, namely benign and malignant. BIRADS 2 and BIRADS 3 are benign classes, while BIRADS 4 and BIRADS 5 are malignant classes. Details of the amount of data in each benign and malignant class for both types of images can be seen in Table I.

TABLE I. THE INITIAL DATA SET (BEFORE AUGMENTATION)

| **Dataset** | **Class** | | **Total** |
|---|---|---|---|
| | ***Malignant*** | ***Benign*** | |
| Ultrasound | 177 | 244 | 421 |
| Mammography | 236 | 203 | 439 |

### B. Proposed Method

We propose an ensemble learning model using two different modes of mammography and ultrasound data with four convolution layer schemes for feature extraction. In the classification process, five machine learning algorithms (support vector machine, random forest, KNN, decision tree, and logistic regression) are run together in one iteration by voting. In general, we propose a model that consists of four main processes: data preprocessing, feature extraction, classification, and performance evaluation. The use of the model aims to compare the use of four convolutional layer schemes as feature extraction using voting on the determination of classification results. Details can be seen in Fig. 1.

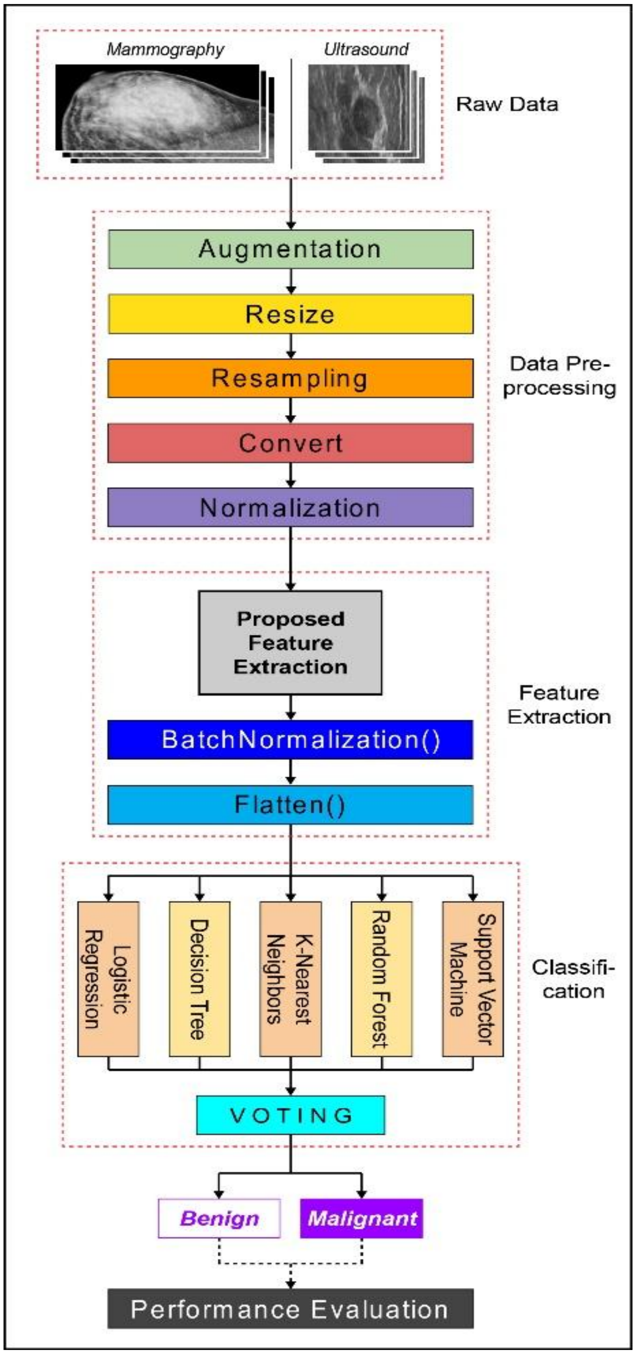


Fig. 1. Proposed ensemble learning model with four different feature extraction schemes.

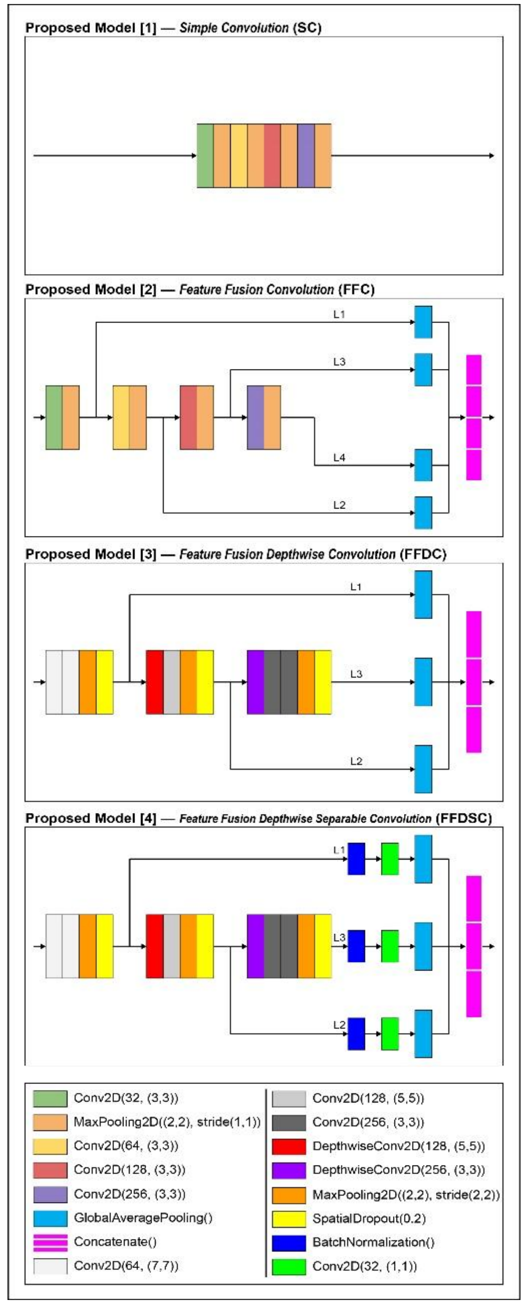


Fig. 2. Convolution layer scheme as feature extraction.

Details of the convolution layers for feature extraction for the four schemes are shown in Fig. 2. The first scheme uses four convolution layers. The second scheme uses feature-fusion convolution in feature extraction with the same layer configuration as in the first scheme. Feature fusion is a technique that combines predictions from multiple machine learning models to create a combination of more than one feature that is used to generate the final prediction [19], [20]. The third scheme uses a depth-wise convolution of the feature fusion, and the last scheme uses a separable depth-wise feature fusion convolution.

*1) Pre-processing:* This preprocessing stage consists of five image improvement techniques before feature extraction, including: augmentation, resize, resampling, convert, and normalisation. Some augmentation techniques in medical images that have been shown to improve classification accuracy are translation, shear, and rotation [5]. Translation is done by shifting the image by a maximum of 15 pixels either towards the positive or negative poles of the x-axis and the y-axis with a probability of 1. Shear is done by shearing the image by a maximum of 15 degrees toward the positive or negative poles of the x-axis and y-axis with a probability of 1. Rotation is done by rotating the image by a maximum of 25 degrees clockwise or counterclockwise. After image enhancement, resize the ultrasound image to 224×224 pixels and the mammography image to 448 x 224 pixels so that the entire data set has a uniform resolution. The total data amounted to 1,000 images for each modal image with various predefined parameters. After that, the data are converted into a numpy array to convert the colour data in the image into numerical data using the numpy library. Labels or classes are extracted using numpy by taking the folder name used in each class. The purpose of converting data into numeric form is to allow the model to recognise the data to be trained. The last step in preprocessing is image normalisation using a rescaling technique with a normalisation factor of 1/255 to change the range from 1-255 to 0-1. The preprocessing data are shown in Table II.

TABLE II. DATASET AFTER IMAGE PRE-PROCESSING

| **Dataset** | **Class** | | **Size [Height×Width] (pixel)** | **Total** |
|---|---|---|---|---|
| | ***Malignant*** | ***Benign.Benign.*** | | |
| Ultrasound | 500 | 500 | 224 × 224 | 1.000 |
| Mammography | 500 | 500 | 448 × 224 | 1.000 |

*2) Feature extraction:* This research uses four different CNN architecture schemes in the convolution layer in the feature extraction stage, namely: simple convolution (SC), feature fusion convolution (FFC) [23], feature fusion depthwise convolution (FFDC), and feature fusion depthwise separable convolution (FFDSC). The first scheme with simple convolution consists of four layers of convolution or Conv2D and 32 filters in the first layer, 64 filters in the second layer, 128 filters in the third layer, and 256 filters in the fourth layer. The kernel size used in the convolution layer is 3x3. The pooling process in each layer uses a kernel pool of the maximum value of MaxPooling2D pixels with a size of 2x2. Each convolution layer, Conv2D, uses the Rectified Linear Activation (ReLU) activation function and the ‘same’ value in the padding parameter so that the output image is the same size and does not cut at the edges. The pooling layer, MaxPooling2D, also uses the ‘same’ value for the padding parameter.

The second scheme with FFC, consists of four convolution blocks that are run cumulatively. The first block consists of a Conv2D layer with 32 filters and a kernel size of 3×3. Then a Maxpooling2D layer with a kernel size of 2×2 and standard strides of 1×1 pixels. The second block consists of the layers of the first block plus Conv2D with filter 64 and kernel size 3×3. Then a Maxpooling2D layer with a kernel size of 2×2 and standard strides of 1×1 pixels. The third block consists of the layers of the second block plus Conv2D with filter 128 and kernel size 3×3. Then the Maxpooling2D layer with a kernel size of 2×2 and standard strides of 1×1 pixels. The fourth block consists of the layers of the third block plus Conv2D with filter 256 and kernel size 3×3. Then a Maxpooling2D layer with a kernel size of 2×2 and standard strides of 1×1 pixels. Each Conv2D convolution layer uses the Rectified Linear Activation (ReLU) activation function and the ‘same’ value in the padding parameter so that the output image is the same size and not cut at the edges. The pooling layer, MaxPooling2D, also uses the ‘same’ value for the padding parameter.

The third scheme, FFDC, consists of three convolution blocks that are run cumulatively. The first block consists of two Conv2D layers with 64 filters and a kernel size of 7×7. Then a Maxpooling2D layer with a kernel size of 2×2 and strides of 2×2 pixels. Next, we have the SpatialDropout2D layer with a probability of 0.2. The second block consists of the layers of the first block plus DepthwiseConv2D with a kernel size of 3×3. Then two Conv2D layers with 256 filters and 3×3 kernel size. Next, we have a Maxpooling2D layer with a kernel size of 2×2 and a shift of 2×2 pixels. Then comes the SpatialDropout2D layer with a probability of 0.2. The third block consists of the layers of the second block plus DepthwiseConv2D with a kernel size of 5×5. Then Conv2D with filter 128 and kernel size 5×5. Next, we have the Maxpooling2D layer with a kernel size of 2×2 and a shift of 2×2 pixels. Subsequently, the SpatialDropout2D layer was populated with a probability of 0.2. Each Conv2D and Depthwise2D uses the Rectified Linear Activation (ReLU) activation function and the ‘same’ value in the padding parameter so that the output image is the same size and does not cut the edges. In layer-pooling, MaxPooling2D also uses the ‘same’ value for the padding parameter.

The fourth scheme with FFDSC has the same architecture as FFDC. The difference between the two types of architecture lies at the end of the process of each block; for the FFDSC architecture, a batch normalisation layer is added, then Conv2D with a 32 filter and a kernel size of 1×1 which is often referred to as Pointwise Convolution. In the first scheme, after passing through the data convolution layer, a batch normalisation layer is added to overcome problems caused by changes in input distribution that occur during the training process.

The results of the training process are entered into the flatten layer to change the data dimension from three dimensions to one dimension to facilitate the classification process. However, in the second, third, and fourth schemes before the BatchNormalization and Flatten layers, the GlobalAveragePooling2D (GAP) layer is added with the aim of reducing dimensions, accelerating computation, and reducing overfitting. After passing through the convolution layer in each block, the data are merged in the concatenate layer. The next stage is the process of compiling the model with Adam parameters in the optimiser and binary cross-entropy in loss.

*3) Classification:* In this research, the ensemble learning used is voting on the classifier. Voting is a technique in which various predictive models vote or weigh their predictions, and the final result is taken based on the majority of these votes or weights [24]. The classification process using five algorithms, namely: SVM [25], Random Forest [26], KNN, Decision Tree, and Logistic Regression are used as voting-based classifiers in the designed model. The hyperparameters used for each classifier are the result of a series of experiments that have been carried out to optimise model performance. The SVM used is the SupportVectorClassifier with a linear kernel and a penalty value of 3.0. Random Forest uses an estimator of 100. KNN uses a neighbour count value of 3, uses distance as a weight assessment scheme, and cosine as a distance metric value. Decision Tree uses the entropy value as a criterion to measure split quality. Logistic regression uses L1 as penalty value, liblinear as optimisation algorithm, and a maximum iteration value of 750. In voting classifiers, SVM, regression are trained together in one iteration. The "soft" parameter is used in voting with the aim that the classification results used later are based on the average probability of each classifier, not on the number of dominant classifiers.

*4) Perfomance evaluation:* We used 5-fold cross-validation for training and model evaluation to calculate the average over five iterations. Some common indicators to determine classification system performance include true positive (TP), true negative (TN), false positive (FP), and false negative (FN). These four basic indicators can be used to determine eight other metrics such as accuracy, precision (TPR), recall, ROC curve, specivicity, F1 score, kappa and matthews correlation coefficient (MCC), which are defined in Eq. (1)-Eq. (8):

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} * 100\% \quad (1)$$

$$Precision = \frac{TP}{TP+FP} * 100\% \quad (2)$$

$$Recall = \frac{TP}{TP+FN} * 100\% \quad (3)$$

$$ROC\ Curve\ (FPR) = \frac{FP}{(FP+TN)} \quad (4)$$

$$Specificity = \frac{TN}{(TN+FP)} \quad (5)$$

$$\text{F1 Score} = \frac{\text{2x(PrecisionxRecall)}}{\text{(Precision+Recall)}} \quad (6)$$

$$Kappa = \frac{p_o - p_e}{1 - p_e} \quad (7)$$

$$Kappa = \frac{TN\ x\ TN - FP\ x\ FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (8)$$

## III. Results

Ultrasound and mammography data were trained separately using a cross-validation fold selection model, KFold, with five data folds. Training data are evaluated by testing the label of the predicted results against the label of the actual data using a confusion matrix. The confusion matrix results for both modal data show that the model has fairly accurate results characterised by the match between the predicted data label and the actual data label having a very dominant data membership in each class. The validation process of the training data is also analysed using several metrics such as accuracy, precision, recall, RoC curve, specificity, F1 score, Cohen's Kappa (Kappa) and Matthews correlation coefficient (MCC). Information about the data metric: The values of the training validation results can be seen in Table III.

Both graphs in Fig. 3 show the accuracy graph using Eq. (1). Both graphs compare the performance of the four feature extraction schemes. The FFC scheme consistently has the most superior feature extraction capability in ultrasound and mammography images. This is reflected in the higher average validated accuracy compared to the other feature extraction schemes, which is 89.90% on the ultrasound image and 98.20% on the mammography image. Meanwhile, Fig. 4 shows that the FFC scheme has the lowest loss among others for mammography and ultrasound, 3.6404 and 0.6488, respectively.

The Receiver Operating Characteristics (ROC) curve serves to evaluate the performance of binary classification models with a focus on the trade-off between sensitivity (recall) and specificity. Both graphs in Fig. 5 show that the FFC scheme performs better than the other three schemes with AUC values of 0.90 and 0.98 on ultrasound and mammography images, respectively. The precision recall curve serves to evaluate the performance of binary classification models, especially in unbalanced data. Fig. 6. shows the results of using four feature extraction schemes for both images. Similarly to the RoC curve, the AUC value also shows that the FFC scheme has the best feature extraction capability of 0.93 and 0.98 in ultrasound and mammography images. The results of the comparison of feature extraction techniques with four different schemes show that scheme 2 (FFC) has the best performance in ensemble models using five machine learning classification algorithms. SVM, Random Forest, KNN, Decision Tree, and Logistic Regression. The feature extraction process with scheme 4 (FFDSC) is designed to reduce the number of parameters and increase the efficiency of deep learning so that it is not optimally applied to classification algorithms using conventional machine learning (non-deep learning).

TABLE III. THE RESULTS OF THE MEAN METRICS FOR 5-FOLD OF EACH MODAL

| Dataset | Proposed Feature Extraction | Accuracy (%) | Precision (%) | Recall (%) | ROC curve (%) | Specificity (%) | F1 score (%) | Kappa (%) | MCC (%) |
|---|---|---|---|---|---|---|---|---|---|
| Ultrasound | Simple Convolution | 0.8240 | 0.8241 | 0.8240 | 0.1760 | 0.8240 | 0.8240 | 0.6480 | 0.6481 |
| | FF Convolution | 0.8990 | 0.8995 | 0.8990 | 0.1010 | 0.8990 | 0.8990 | 0.7980 | 0.7985 |
| | FF Depthwise Convolution | 0.8930 | 0.8930 | 0.8930 | 0.1070 | 0.8930 | 0.8930 | 0.7860 | 0.7860 |
| | FF depthwise separable convolution | 0.8820 | 0.8822 | 0.8820 | 0.1180 | 0.8820 | 0.8820 | 0.7640 | 0.7642 |
| Mammography | Simple Convolution | 0.9220 | 0.9222 | 0.9220 | 0.0780 | 0.9220 | 0.9220 | 0.8440 | 0.8442 |
| | FF convolutionconvolution | 0.9820 | 0.9821 | 0.9820 | 0.0180 | 0.9820 | 0.9820 | 0.9640 | 0.9641 |
| | FF Depth-wiseDepth-wise Convolution | 0.9630 | 0.9634 | 0.9630 | 0.0370 | 0.9630 | 0.9630 | 0.9260 | 0.9264 |
| | FF Separable Convolutio in Depth | 0.9500 | 0.9507 | 0.9500 | 0.0500 | 0.9500 | 0.9500 | 0.9000 | 0.9007 |

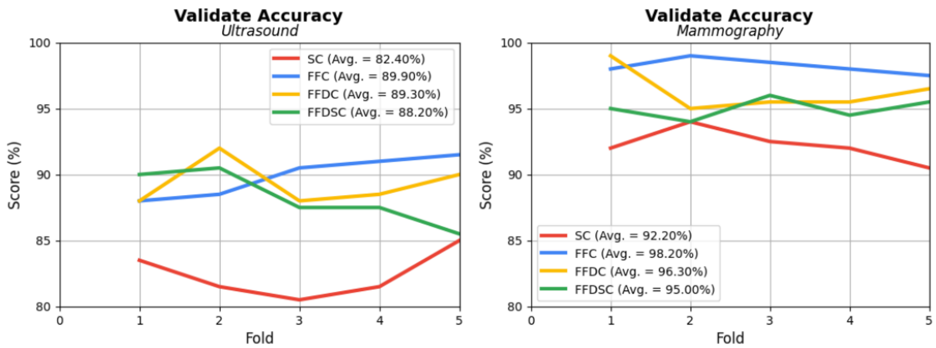


Fig. 3. Accuracy graph in ultrasound and mammography.

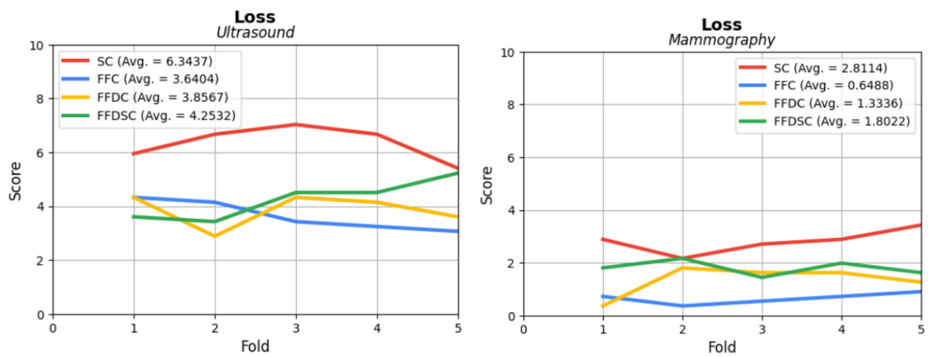


Fig. 4. Graph of loss score on ultrasound and mammography images.

## IV. Discussion

Several studies focus on the classification of breast cancer using several types of medical imaging. The limited number of data sets in medical images, especially hispatological images, is a challenge for deep learning techniques for feature extraction. The use of deep convolutional neural networks (DCNN) and SVM techniques for classification [27]Techniques to overcome the imbalance in the amount of data in a particular class are a challenge in itself, and there are several ways to overcome these limitations, one of which is the use of image enhancement techniques. Image augmentation techniques have been shown to improve classification system performance compared to the original dataset [28]. It has also performed feature extraction using deep learning with several types of CNN architectures (ResNet-18, ResNet-34, ResNet-50, and EfficientNet-B0). The study used mammography images and the classification process based on multiple instances learning [29] In contrast to what was done by [30] with other types of images, namely thermal imaging with classification based on convolutional neural networks. The collaborative use of two image models has also been done by [31] using hispatology and ultrasound images based on transfer learning.

This research has conducted eight different scenarios with four different feature extraction techniques. The test results show that the use of the feature extraction technique of the feature fusion convolution type has been shown to give the best performance compared to the other three feature extraction techniques. For the classification stage, this research uses five machine learning algorithms that are combined into one unit in an ensemble machine learning system with a soft voting classifier. Table IV compares the results of the proposed method with those of other methods from previous studies for the classification of breast cancer.

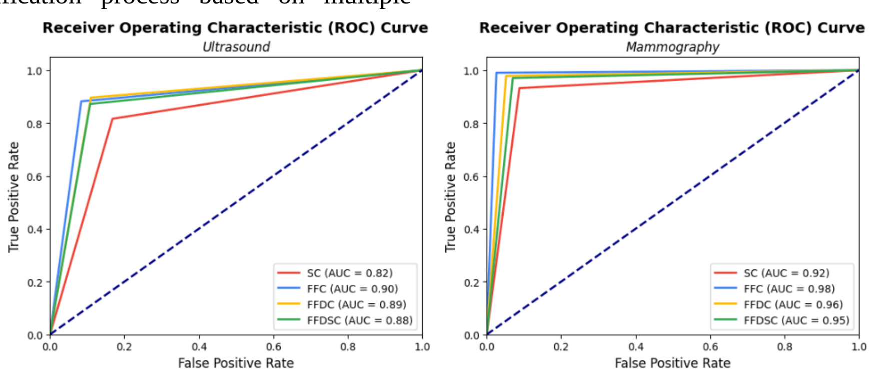


Fig. 5. Receiver Operating Characteristic (ROC) curve in ultrasound and mammography images.

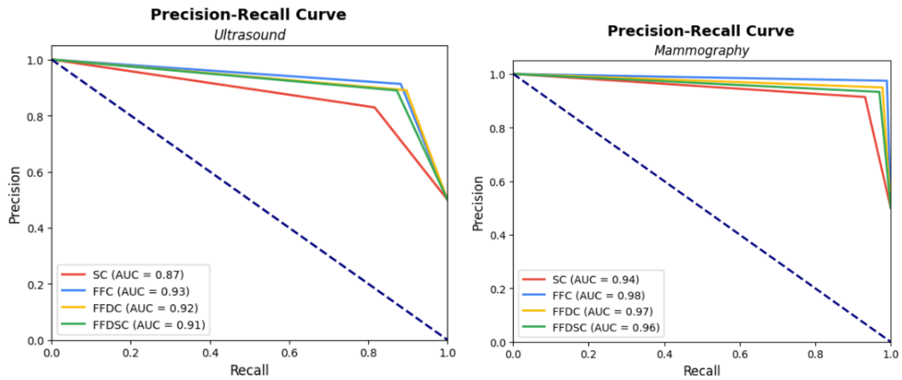


Fig. 6. Precision recall curve on ultrasound and mammography images.

TABLE IV. COMPARISON OF THE PERFORMANCE OF PROPOSED METHOD FOR THE CLASSIFICATION OF BREAST CANCER

| Reference | Feature Extraction | Classification | Image | Parameter | | | |
|---|---|---|---|---|---|---|---|
| | | | | Acc (%) | Loss | RoC | Precision-Recall |
| Kulkarni & Rabidas [28] | U-Net model | | Mammography | 96.81 | - | - | - |
| Hassan et al. [26] | Deep convolutional neural networks (DCNNs) | SVM | Hispatology | 99.24 | - | - | - |
| Bobowicz et al.[28] | ResNet-18, ResNet-34, ResNet-50 and EfficientNet-B0 | Multiple instance learning-based | Mammography | 81.6 | - | 0.90 | - |
| Roslidar et al. [30] | Convolutional neural networks | | Thermal Imaging | - | - | - | - |
| Aaroj et al. [[30] | Transfer Learning | | Hispatology and ultrasound | 99.35 | - | - | - |
| Proposed | Simple convolution (SC), feature fusion convolution (FFC), feature fusion depthwise convolution (FFDC), and feature fusion depthwise separable convolution (FFDSC). | Ensemble machine learning | Mammography | 98.20 | 0.6488 | 0.98 | 0.98 |
| | | | Ultrasound | 89.90 | 3.6404 | 0.90 | 0.93 |

## V. CONCLUSIONS AND FUTURE WORK

This study has compared ensemble learning models with four convolution layer schemes for feature extraction in breast cancer classification using ultrasound and mammography images separately. The four convolution layer schemes used are simple convolution (SC), feature fusion convolution (FFC), feature fusion depthwise convolution (FFDC), and feature fusion depthwise separable convolution (FFDSC). Classification is performed using an ensemble learning soft voting classifier with SVM, Random Forest, KNN, Decision Tree, and Logistic Regression algorithms. The experimental results show that the FFC convolutional layer scheme achieves the best performance for both datasets. In the ultrasound data set, FFC achieved a value of 0.90 in each of the Val-Acc, TPR, Recall, TNR and F1-Score metrics. In the mammography data set, the FFC reached a value of 0.98 on each of the same metrics. These results emphasize the effectiveness of feature fusion in improving classification performance. Future research can focus on exploring more complex fusion techniques, such as using multimodal data or combining classification with deep learning.

## ACKNOWLEDGMENT

The authors gratefully acknowledge the sponsorship of the institute for research and community service UIN Sunan Kalijaga Yogyakarta based on SK Rector No: B-1913/Un.02/L3/TL/05/2024.

## REFERENCES

[1] C. Y. Chou, T. T. Shen, W. C. Wang, and M. P. Wu, “Favorable breast cancer mortality-to-incidence ratios of countries with good human development index rankings and high health expenditures,” Taiwan J Obstet Gynecol, vol. 63, no. 4, pp. 527–531, Jul. 2024, doi: 10.1016/j.tjog.2023.11.012.

[2] A. Glechner et al., “Mammography in combination with breast ultrasonography versus mammography for breast cancer screening in women at average risk,” Cochrane Database of Systematic Reviews, vol. 2023, no. 3, Mar. 2023, doi: 10.1002/14651858.CD009632.pub3.

[3] S. Aslani et al., “Enhancing cancer prediction in challenging screen-detected incident lung nodules using time-series deep learning,” Computerized Medical Imaging and Graphics, vol. 116, Sep. 2024, doi: 10.1016/j.compmedimag.2024.102399.

[4] L. Xiaoming et al., “Application of combined preoperative indocyanine green lymphography and ultrasonography for low-pressure vein localization in secondary lymphedema surgery for breast cancer,” Asian J Surg, vol. 47, no. 1, pp. 289–295, Jan. 2024, doi: 10.1016/j.asjsur.2023.08.121.

[5] M. Hamiane and F. Saeed, “SVM classification of MRI brain images for computer-assisted diagnosis,” International Journal of Electrical and Computer Engineering, vol. 7, no. 5, pp. 2555–2564, 2017, doi: 10.11591/ijece.v7i1.pp2555-2564.

[6] J. Liu et al., “An overview of artificial intelligence in medical physics and radiation oncology,” Journal of the National Cancer Center, vol. 3, no. 3, pp. 211–221, Sep. 2023, doi: 10.1016/j.jncc.2023.08.002.

[7] A. Grażyńska et al., “BIRADS 4 – Is it possible to downgrade lesions that do not enhance on recombinant contrast-enhanced mammography images?,” Eur J Radiol, vol. 167, Oct. 2023, doi: 10.1016/j.ejrad.2023.111062.

[8] N. M. Hassan, S. Hamad, and K. Mahar, “Mammogram breast cancer CAD systems for mass detection and classification: a review,” Multimed

Tools Appl, vol. 81, no. 14, pp. 20043–20075, Jun. 2022, doi: 10.1007/s11042-022-12332-1.

[9] R. Song, T. Li, and Y. Wang, "Mammographic Classification Based on XGBoost and DCNN with Multi Features," IEEE Access, vol. 8, pp. 75011–75021, 2020, doi: 10.1109/ACCESS.2020.2986546.

[10] W. C. Shia, L. S. Lin, and D. R. Chen, "Classification of malignant tumours in breast ultrasound using unsupervised machine learning approaches," Sci Rep, vol. 11, no. 1, Dec. 2021, doi: 10.1038/s41598-021-81008-x.

[11] Y. L. Huang, D. R. Chen, Y. R. Jiang, S. J. Kuo, H. K. Wu, and W. K. Moon, "Computer-aided diagnosis using morphological features for classifying breast lesions on ultrasound," Ultrasound in Obstetrics and Gynecology, vol. 32, no. 4, pp. 565–572, Sep. 2008, doi: 10.1002/uog.5205.

[12] S. Beneddine, "Nonlinear input feature reduction for data-based physical modeling," J Comput Phys, 2023, [Online]. Available: https://www.elsevier.com/open-access/userlicense/1.0/

[13] L. Liu, W. Gao, H. Yu, and D. E. Keyes, "Overlapping multiplicative Schwarz preconditioning for linear and nonlinear systems," J Comput Phys, vol. 496, Jan. 2024, doi: 10.1016/j.jcp.2023.112548.

[14] A. Sharma, D. Goyal, and R. Mohana, "An ensemble learning-based framework for breast cancer prediction," Decision Analytics Journal, vol. 10, Mar. 2024, doi: 10.1016/j.dajour.2023.100372.

[15] L. Alzubaidi et al., "MEFF – A model ensemble feature fusion approach for tackling adversarial attacks in medical imaging," Intelligent Systems with Applications, vol. 22, Jun. 2024, doi: 10.1016/j.iswa.2024.200355.

[16] K. Tembhare, T. Sharma, S. M. Kasibhatla, A. Achalere, and R. Joshi, "Multi-ensemble machine learning framework for omics data integration: A case study using breast cancer samples," Inform Med Unlocked, vol. 47, Jan. 2024, doi: 10.1016/j.imu.2024.101507.

[17] I. D. Mienye and Y. Sun, "A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects," 2022, Institute of Electrical and Electronics Engineers Inc. doi: 10.1109/ACCESS.2022.3207287.

[18] M. H. A. M. H. Himel, P. Chowdhury, and M. A. M. Hasan, "A robust encoder decoder based weighted segmentation and dual staged feature fusion based meta classification for breast cancer utilizing ultrasound imaging," Intelligent Systems with Applications, vol. 22, Jun. 2024, doi: 10.1016/j.iswa.2024.200367.

[19] R. Massafra et al., "Radiomic feature reduction approach to predict breast cancer by contrast-enhanced spectral mammography images," Diagnostics, vol. 11, no. 4, Apr. 2021, doi: 10.3390/diagnostics11040684.

[20] S. Uyun and L. Choridah, "Feature selection mammogram based on breast cancer mining," International Journal of Electrical and Computer Engineering, vol. 8, no. 1, pp. 60–69, Feb. 2018, doi: 10.11591/ijece.v8i1.pp60-69.

[21] H. E. Kim, A. Cosa-Linan, N. Santhanam, M. Jannesari, M. E. Maros, and T. Ganslandt, "Transfer learning for medical image classification: a literature review," Dec. 01, 2022, BioMed Central Ltd. doi: 10.1186/s12880-022-00793-7.

[22] M. Ilyas et al., "Deep Learning based Classification of Thyroid Cancer using Different Medical Imaging Modalities: A Systematic Review," VFAST Transactions on Software Engineering, vol. 9, no. 4, pp. 1–17, 2021.

[23] I. U. Haq, H. Ali, H. Y. Wang, C. Lei, and H. Ali, "Feature fusion and Ensemble learning-based CNN model for mammographic image classification," Journal of King Saud University - Computer and Information Sciences, vol. 34, no. 6, pp. 3310–3318, Jun. 2022, doi: 10.1016/j.jksuci.2022.03.023.

[24] G. Żabiński, J. Gramacki, A. Gramacki, E. Miśta-Jakubowska, T. Birch, and A. Disser, "Multi-classifier majority voting analyses in provenance studies on iron artefacts," J Archaeol Sci, vol. 113, Jan. 2020, doi: 10.1016/j.jas.2019.105055.

[25] V. Jain, A. Jain, A. Chauhan, S. S. Kotla, and A. Gautam, "American Sign Language recognition using Support Vector Machine and Convolutional Neural Network," International Journal of Information Technology (Singapore), vol. 13, no. 3, pp. 1193–1200, Jun. 2021, doi: 10.1007/s41870-021-00617-x.

[26] Dhiyaussalam and S. 'Uyun, "Optimization of Random Forest Hyperparameters with Genetic Algorithm in Classification of Lung Cancer," in 6th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), 2023, pp. 1–7.

[27] A. M. Hassan, A. Yahya, and A. Aboshosha, "A framework for classifying breast cancer based on deep features integration and selection," Neural Comput Appl, vol. 35, no. 16, pp. 12089–12097, Jun. 2023, doi: 10.1007/s00521-023-08341-2.

[28] S. Kulkarni and R. Rabidas, "Fully convolutional network for automated detection and diagnosis of mammographic masses," Multimed Tools Appl, vol. 82, no. 29, pp. 44819–44840, Dec. 2023, doi: 10.1007/s11042-023-14757-8.

[29] M. Bobowicz et al., "Attention-Based Deep Learning System for Classification of Breast Lesions—Multimodal, Weakly Supervised Approach," Cancers (Basel), vol. 15, no. 10, May 2023, doi: 10.3390/cancers15102704.

[30] R. Roslidar et al., "A Review on Recent Progress in Thermal Imaging and Deep Learning Approaches for Breast Cancer Detection," 2020, Institute of Electrical and Electronics Engineers Inc. doi: 10.1109/ACCESS.2020.3004056.

[31] S. Arooj et al., "Breast Cancer Detection and Classification Empowered With Transfer Learning," Front Public Health, vol. 10, Jul. 2022, doi: 10.3389/fpubh.2022.924432.