BİTLİS EREN ÜNİVERSİTESİ
FEN BİLİMLERİ DERGİSİ

ISSN: 2147-3129 / e-ISSN: 2147-3188

Article Type : Research Article
Received : November 18, 2024
Revised : March 20, 2025
Accepted : April 11, 2025
DOI : 10.17798/bitlisfen.1587411

Year : 2025
Volume : 14
Issue : 2
Pages : 755-776

BREAST CANCER CLASSIFICATION IN ULTRASOUND IMAGING USING COST-SENSITIVE LEARNING AND K-MEANS SMOTE ON THE IMBALANCED BUSI DATASET WITH DEEP FEATURE EXTRACTION

Nedim MUZOĞLU [1]

[1] *University of Health Sciences, Hamidiye Faculty of Health Sciences, İstanbul, Türkiye,*

nedim.muzoglu@sbu.edu.tr

ABSTRACT

Breast cancer is one of the five most common types of cancer that occurs when breast tissue turns into a tumor and mainly affects women. Early diagnosis of the disease is crucial for the patient's lifespan. However, misclassification of malignancy may result in treatment delays and initiate an irreversible process for the patient. This study proposes an approach for classifying ultrasound breast images into malignant, benign, and healthy categories, with a particular emphasis on minimizing false-negative outcomes. The BUSI dataset, characterized by imbalanced class distributions, was used for the breast cancer detection. The dataset was augmented to enhance feature representations using contrast-limited adaptive histogram equalization (CLAHE) to address the class imbalance issue, creating the BUSICL dataset. Features extracted from both datasets with the VGG16 and ResNet50 models were then classified using a support vector machine (SVM). Following the results analysis, the SVM algorithm's cost matrix values were adjusted according to the inverse proportions of class distributions applying a cost-sensitive approach. In addition, the robustness of the proposed methodology is compared with the K-Means SMOTE algorithm. The proposed method achieved an overall accuracy of 99.36%, surpassing the performance of previous comprehensive classification studies using the BUSI dataset.

Keywords: Breast cancer, Clahe, Cost-sensitive learning, K-means smote, Transfer learning.

1 INTRODUCTION

Breast cancer is a type of cancer in which abnormal breast cells in the milk-producing tissue of the breast turn into a tumor and spread throughout the body if precautions are not taken, with approximately 99% of the cases seen in women [1]. According to data from 2022, of the 2.3 million women diagnosed with breast cancer worldwide, 670,000 have died. This disease is one of the five most common cancers in our country, and the disease is responsible for around 5% of deaths in women [2]. Although breast cancer, like many cancers, has no symptoms in the early stages, swelling in the breast, changes in the appearance of the breast or abnormal fluids secreted from the nipple are symptoms that may be observed by the patient in the later stages.

The American College of Radiology has established the protocols for the assessment of images obtained through mammography, magnetic resonance imaging (MRI), and ultrasound (US) devices employed in the radiological diagnosis of breast cancer, ensuring standardized reporting procedures for radiologists [3]. This protocol, which evaluates six scenarios to detect the presence of a benign or malignant tumor obtained from imaging devices, is known as the Breast Imaging Reporting and Data System. In the radiological diagnosis of the disease, mammography, MRI, and ultrasound devices are employed according to the structured protocol. Among these, ultrasound imaging is a real-time diagnostic method that provides a cost-effective and safe approach for diagnosis, as it does not involve radiation and is available in the inventory of almost every healthcare facility. However, unlike mammography and MRI devices, this method is physician-dependent because a radiologist must perform the scan with an ultrasound device to diagnose.

On the other hand, the effectiveness of ultrasound examinations can vary due to several factors, including challenges such as unwanted interferences and speckle noise inherent to the device's operating principles, as well as the physician's proficiency in effectively utilizing the device. These variables collectively influence the accuracy and success of the diagnostic process. To overcome these difficulties, computer-aided diagnostic systems assist physicians in interpreting medical images, thereby reducing the risk of potential errors stemming from radiologists' fatigue, mental state, and experience. For this purpose, these systems have long been supported by mostly artificial intelligence (AI) based algorithms. The inception of research utilizing the detection of breast cancer from ultrasound images using AI is discernible within the literature commencing in the latter part of the 1990s. During this period, scientific

attention has been devoted to breast cancer classification studies that employ machine learning algorithms to classify features obtained through various feature extraction methodologies, some of which are the following. In the study, [4] used the k-means method for the classification of breast lesions. They applied morphology operations and histogram equalization methods to a dataset of proven pathology, consisting of 110 malignant and 140 benign tumors, for breast cancer classification. The study achieved 88.8% binary classification accuracy. [5] utilized Support Vector Machines (SVM) to classify 87 malignant and 303 benign ultrasound images, all confirmed through pathology assessment. Textual features were employed for this purpose. The study yielded a success rate of 93.20%. [6] proposed a study wherein 19 morphological features were extracted from a dataset comprising 34 malignant and 84 benign breast tumors for the classification of breast cancer. Extracted morphological features underwent principal component analysis to reduce dimensionality. Subsequently, SVM was employed to classify the dataset. In the training phase, the study achieved an accuracy of 82% using a 10-fold validation approach.

In the following years, with the development of deep learning algorithms, many studies have been carried out in the field of classification and segmentation with the help of models that automatically extract features in breast cancer detection with high success. These studies have been proposed entirely with deep learning models only or as hybrid models with machine learning and traditional feature extraction methods, and some of them are given. [7] proposed a study on breast cancer diagnosis and classification using the Breast Ultrasound Image (BUSI) dataset. The workflow included preprocessing, segmentation, feature extraction, and classification. They utilized Wiener filtering and contrast enhancement for preprocessing, the Krill swarm algorithm for segmentation, and deep learning models such as Vgg16, Vgg19, and SqueezeNet for feature extraction. In the study, Ragab and colleagues achieved a remarkable success rate of 97.52%. In the [8]'s research, utilizing datasets BUSI and UDIAT in the realm of breast cancer study, is proposed to overcome the challenges posed by class imbalance and the limitations associated with relying solely on global or local features for achieving definitive tumor classification. For this purpose, the study employed the BSMOTE method to generate synthetic data, utilized the ResNet50 model to capture global information, and incorporated the Histogram of Oriented Gradients (HOG) technique to capture local information. This approach yielded a satisfactory performance of 99.14% for the BUSI dataset. [9] introduced a framework comprising a robust deep-learning model with 24 layers designed for the BUSI dataset employed in breast cancer detection. The proposed model includes six convolutional layers,

nine inception modules, and one fully connected layer. Additionally, it consists of the clipped Rectified Linear Unit activation function and cross-channel normalization techniques. This model demonstrated an impressive accuracy rate of 99.35%.

The primary objective of the previously mentioned studies is the classification of breast cancer; however, the vast majority of them include strategies aimed at mitigating the challenges posed by the limited availability of ultrasound images, particularly in cases of malignant class imbalance. Unfortunately, openly shared datasets used for AI-based breast cancer detection from ultrasound images cannot meet the high data demand of deep learning. These datasets are also known to be imbalanced between classes and contain insufficient data. Moreover, models trained on imbalanced data tend to be dominated by samples from the majority class. Machine learning algorithms assign equal weight to misclassification errors, but this can be very problematic, especially in medical diagnosis, because misclassifying a positive case is always more costly than misclassifying a negative case. To solve these problems, research mainly uses data augmentation methods to increase the amount of data. However, the data augmentation technique may result in the loss of the region of interest where the cancerous area is located due to incorrect magnification techniques. This imbalance challenge often requires addressing overfitting or underfitting issues in classification studies. A most commonly employed strategy to address the misclassification challenges associated with minority classes in imbalanced datasets involves the use of synthetic data generation or majority class reduction techniques. The relevant studies conducted in this context are outlined below.

In the study [10], addressing the class imbalance problem, a model is presented that generates heuristic synthetic data using SMOTE-based k-means clustering, aiming to avoid the influence of unnecessary noise. The experimental results of the model, evaluated on 90 datasets, demonstrate an improvement in classification performance. The study [11], proposes a cluster-based oversampling method for imbalanced data, combining the SMOTE and k-means algorithms. In this work, the Random Forest algorithm is used as the classifier. The model, validated with various UCI datasets, achieves Sensitivity and Specificity values of 99.84% and 99.56%, respectively. In the study presented as Reduced Noise-SMOTE (RN-SMOTE) to address the class imbalance, the synthetic instances generated by SMOTE were subsequently subjected to denoising using Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [12]. RN-SMOTE was applied to nine different datasets and classifiers, and it was reported that the proposed approach enhances classifier performance. In another study [13], a transfer learning approach was used for feature extraction in detecting melanoma, a type of skin

cancer. The artificial samples generated using K-Means SMOTE were then classified with the XGB classifier. Using the ISIC Challenge 2018-19 datasets, the study achieved an accuracy (ACC) of 0.941. Nonetheless, common challenges arise from oversampling methods change the distribution of different classes, potentially leading to the exclusion of valuable data or the generation of unnecessary data.

Another method, cost-sensitive learning, aims to address class-imbalanced issue by assigning distinct cost values to minority and majority classes, rather than attempting to alleviate the misclassification problem through the generation of artificial data. This method has been utilized for various datasets in AI-based disease detection to date and has demonstrated success, and some of them are as follows. [14], utilized the Pima Indian Diabetes, Haberman Breast Cancer, Cervical Cancer Risk Factors, and Chronic Kidney Disease datasets in their study, all of which are highly imbalanced datasets. To deal with this problem, they transformed the objective function of logistic regression, decision trees, extreme gradient boosting, and random forest algorithms into robust cost-sensitive algorithms. They employed the Haberman database for breast cancer classification, the focus of their study, and achieved an accuracy of 80.2%. [15], utilized the Wisconsin Diagnostic Breast Cancer dataset in their classification study for breast cancer diagnosis. They employed the Information Gain (IG) algorithm, based on the entropy value of the system, to select a compact feature subset with maximal discriminative capability from the dataset. They achieved a 98.32% success rate for classification by utilizing an improved cost-sensitive Support Vector Machine (SVM) optimized with particle swarm optimization. In the study [16], given the limited availability of expert dermatologists in developing nations, Ravi has proposed a deep learning-driven model aimed at the precise classification of skin diseases. This model relies on feature classification derived from fine-tuned cost-sensitive EfficientV2 architecture, utilizing support vector machines (SVM) and random forests algorithms. This proposed model attained a remarkable 99% classification accuracy.

This study aims to develop a model for breast cancer detection, focusing on achieving minimal misclassification rates when working with imbalanced ultrasound image datasets. The open-access dataset employed for this purpose exhibits class imbalance, which can adversely impact classification performance. To address this issue, feature enrichment methods and transfer learning models were employed, along with cost-sensitive learning techniques to mitigate the misclassification of malignant lesions. The highlights of the study are as follows.

- To the best of the author's knowledge, cost-sensitive learning approach has been used for the first time in the classification of BUSI dataset.
- The Contrast-Limited Adaptive Histogram Equalization (CLAHE) enhancement method is used to reduce and suppress noise in ultrasound images.
- An essential aspect of this study was the utilization of rich features extracted from the BUSI and BUSICL datasets, which were combined and classified using a machine learning approach employing class imbalance handling methods.
- A key focus of this study was comparing SVM-based breast cancer classification with two distinct methods for addressing class imbalance: K-means SMOTE and Cost-Sensitive Learning.
- The cost-sensitive deep learning feature extraction approach achieved the highest accuracy in breast cancer classification, surpassing state-of-the-art models.

The study comprises five sections: the second section detailing the dataset and methodology, the third section presenting the experimental analysis, and the following sections, which discuss and conclude.

2 MATERIAL AND METHODS

2.1 Dataset

In this study, a publicly available breast ultrasound images dataset (BUSI) [17] was obtained from the open data-sharing platform Kaggle, and its relevant link is provided in the "Availability of Data and Materials" section. Although the BUSI dataset provides the corresponding masks for each image, only the relevant ultrasound images have been used in the scope of this study. The dataset consists of three distinct classes: malignant lesions, benign lesions, and healthy tissue, comprising 780 images. Specifically, within the benign, malignant and normal classes, there are 437 and 210 images, respectively, while the healthy tissue class contains 133 images. In particular, the dataset contains a class imbalance problem. Due to the data-hungry nature of deep learning algorithms, a new dataset named BUSICL was created by applying the CLAHE method to the BUSI dataset, which has limited data, to improve classification accuracy. During the model experimental analysis phase, a partitioning ratio of 80:20 was preferred for the training and test datasets, as higher success rates were achieved compared to alternative ratios. The images, although potentially varying in size, exhibit an

average resolution of 500x500 pixels, are in .png format, and are represented in grayscale. Figure 1 illustrates samples from the distinct classes within the datasets.

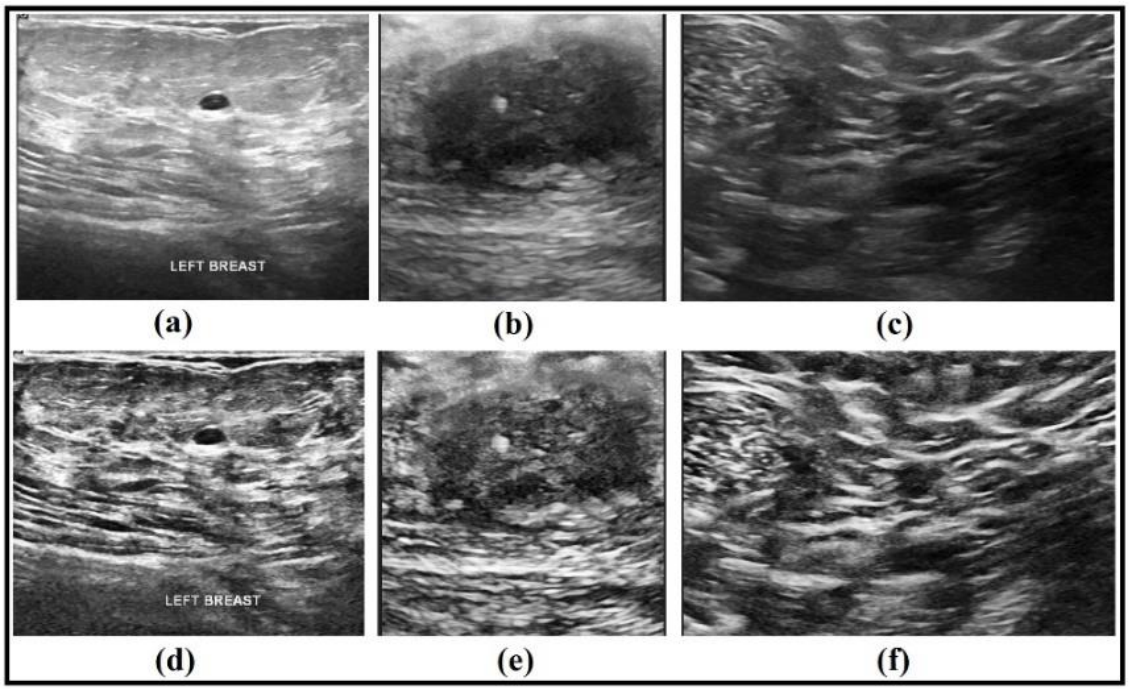

Figure 1. BUSI Dataset , (a) Benign Lesion, (b) Malign Lesion, (c) Healthy Lesion, and BUSICL Dataset (d) Benign Lesion, (e) Malign Lesion, (f) Healthy Lesion.

The distribution of images within the classes for the proposed approach can be seen in Tables 1.

Table 1. The utilization numbers of BUSI/BUSICL images in the proposed approach

Classes	Train Images	Test Images	Total Images
Benign	350	87	437
Malign	168	42	210
Normal(healthy)	106	27	133
Total Images	624	156	780

2.2 Contrast Limited Adaptive Histogram Equalization

Image enhancement methods in medical imaging allow medical professionals to identify abnormalities, tumors, and other critical details more effectively. Contrast-limited adaptive histogram equalization (CLAHE) was developed to address shortcomings in other histogram methods by improving local contrast and the definition of edges in every region of the image by preventing excessive noise amplification in relatively homogeneous regions [18]. So, many studies employing CLAHE for image enhancement aim to improve contrast while minimizing

the amplification of noise inherent to medical imaging systems [19]. Considering the image details of the resulting dataset, the threshold for the contrast limit was set to ClipLimit 3.0, which involves localized changes in contrast. Additionally, tileGridSize, which adjusts the number of tiles in rows and columns during image enhancement with CLAHE, was set to 8x8.

2.3 Data Augmentation

Data augmentation is an effective approach to reducing overfitting when working with limited datasets. Techniques such as, scaling, cropping, flipping, and rotation can be applied individually or in combination. It operates as follows: During model training, mini-batches undergo random transformations in each iteration, allowing the network to learn from different versions of the same samples. Consequently, while the total number of training samples remains unchanged, the model is exposed to diverse representations, thereby improving generalization. The classes in the BUSI dataset used in this study contain a limited number of samples and are imbalanced. If these issues are not addressed, they may lead to overfitting. In this study, reflection, translation, and scaling data augmentation techniques were used to enhance the generalization capability of the models.

2.4 Deep Learning Models and Parameter Selection

Deep learning models work by extracting features from input images using a sliding kernel, then summarizing these outputs with pooling layers to reduce this feature size. Moreover, in the model development processes initiated by the ILSVRC competitions, numerous models were created with careful consideration of model performance, complexity, the vanishing gradient problem, and parameter size. Therefore, selecting the appropriate model for a given dataset is closely linked to the model's performance.

2.4.1 Vgg16

The Vgg16 [20], a 16-layer deep learning model developed by the Visual Geometry Group at Oxford University, achieved a classification success rate of 92.7% in the 2014 Large Scale Visual Recognition Competition (ILSVRC) on 1000 images in 1000 different categories, and it continues to outperform many models today [21]. VGG16 takes 224x224 tensors as input. It consists of a total of 16 learnable layers. Kernel size, max-pooling filter, and stride values are fixed to 3x3, 2x2, and 2, respectively. Convolution layers and max-pooling layers are then connected to 2 fully connected layers with 4096 channels, and the last output layer evaluates

1000 classes. The output size of the previous layer can be redesigned for datasets with a different number of classes and can be utilized in various studies. Figure 2 illustrates a general representation of the Vgg16 structure.

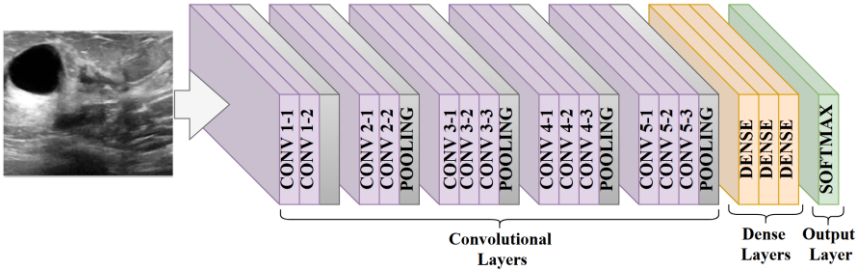

Figure 2. The Structure of Vgg16.

2.4.2 Resnet50

ResNet50 [22] is a deep learning architecture with 50 layers, developed by Microsoft Research in 2015. It is a variant of the Residual Network (ResNet) architecture. The key innovation that distinguishes ResNet architecture from previous architectures is the residual connections. This architectural design incorporates skip connections between the input and output to acquire residual functions. This mechanism effectively addresses the issue of vanishing gradients by facilitating the flow of information to deeper layers. In the residual structure, the convolutional block uses a 1x1 convolutional layer to reduce the number of filters before the 3x3 convolutional layer and then adds the input and output together [23]. The architecture, where the residual input is now represented by X and the output by $F(X) + X$ is presented below. This architecture, commonly referred to as an identity block, is a standard component within ResNet structures. The convolutional block is another type of block used in ResNet structures. The convolutional block contains a convolutional layer within the skip connection to ensure that the input and output size is adjusted as required. Identity block and ResNet50 structures are given in Figure 3 and Figure 4, respectively.

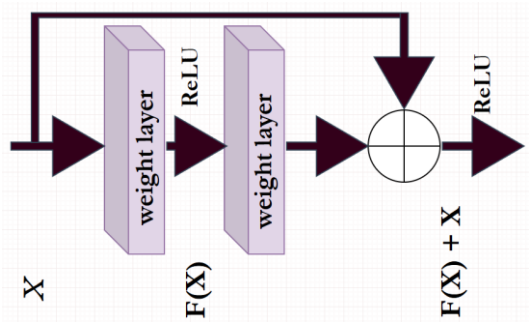

Figure 3. The Structure of Identity Block.

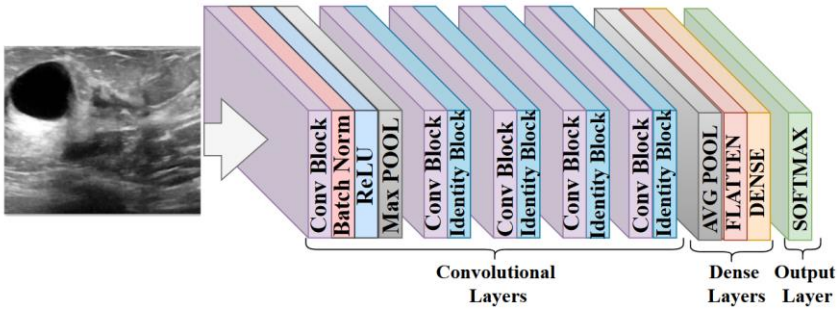

Figure 4. The Structure of Resnet50.

2.4.3 Model Deployment

In this study, after evaluating the performance of various deep learning models and state-of-the-art architectures, the decision was made to proceed with VGG16 and ResNet-50 for further analysis. The reasons for selecting these models are as follows:

VGG16 has demonstrated its effectiveness in numerous studies, owing to its stacked convolutional blocks, and fine-tuning has been shown to enhance its performance. Moreover, ResNet-50 was preferred due to its high accuracy performance, and its ability to overcome the vanishing gradient problem in challenging tasks through its residual structure. Additionally, its use of the global average pooling mechanism allows for the inclusion of global features in the classification process, alongside local features. Upon examining the strengths of both models, the features derived from them generate an expectation of high performance when combined.

On the other hand, due to the limited number of images in the BUSI dataset utilized in this study, the issue of overfitting during model training often arises. This challenge is particularly evident in datasets with a limited amount of data. To this end, numerous studies have highlighted the effectiveness of fine-tuning pre-trained models using weights acquired from training on ImageNet to enhance model accuracy. As a result, this study employed a transfer learning-based strategy in which the weights of the first ten layers of the VGG16 and ResNet50 models were frozen while the remaining layers were trained. The proposed transfer learning method can be seen in Figure 5.

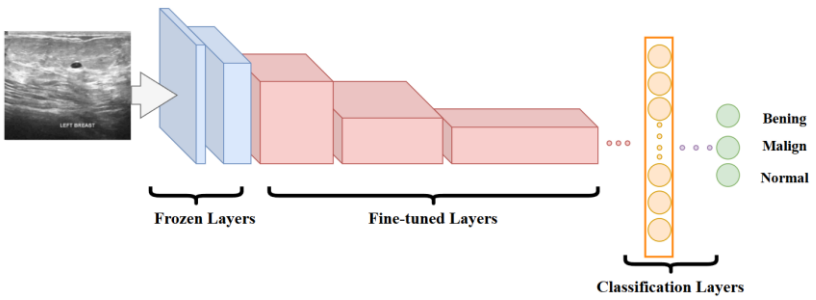

Figure 5. Transfer Learning Approach for the Proposed Model.

Determination of the hyperparameters is also essential for model performance. While a learning rate of $1x10^{-4}$ is commonly used in many deep learning studies, experimental results from this study indicated that starting with a learning rate of $3x10^{-4}$ and halving it after 50 epochs led to a more precise convergence to the local minimum. Moreover, the SGD [24] optimization algorithm, which has demonstrated superior generalization performance in numerous studies, was employed to optimize the model's cost function towards the local minimum. Additionally, based on the results of the experimental studies, it was determined that training the model for 100 epochs would be both sufficient and beneficial. In addition, the batch size is defined based on the computational capacity of the available hardware. Hyperparameters settings used in the training process are shown in Table 2.

Table 2. Hyperparameters settings for the Models.

Hyperparameter	Settings
Optimizer	SGD
Batch Size	32
Epoch	100
Train / Test Split Ratio	80:20
Initial Learning Rate	$3x10^{-4}$
Reduced Learning Rate	$15x10^{-5}$

2.5 Cost-Sensitive Support Vector Machine

So far, successful results have been achieved in binary and multi-class classification applications using SVMs [25]. SVM operates within a cost-insensitive framework, which seeks to minimize the loss function and assigns equal cost values to false positives and false negatives [14]. However, this approach can yield costly outcomes, mainly when dealing with imbalanced medical datasets during classification tasks. In medical diagnosis, particularly, the equality of cost values for false negatives (incorrectly categorizing an individual as healthy despite their illness) and false positives (erroneously classifying an individual as sick despite their health) may result in circumstances challenging to rectify due to treatment delays[15]. In cost-insensitive methods, the cost of all true positive values is zero, while the costs of other misclassified results are set to one. On the other hand, the penalization of incorrect classifications based on the parameters $\boldsymbol{\xi}$ and $\boldsymbol{C}$ to ensure both the maximum margin value, and the minimum classification error of the SVM algorithm, where ‖w‖ is the norm of the weight. This relationship can be defined within the cost matrix, represented as a square matrix in SVM, and is regulated by a parameter $\mathbf{C}$, as shown in Equation 1.

$$\text{Min } \frac{1}{2} ||w||^2 + C\sum_{i=1}^{n} \xi_i \quad (1)$$

In cost-sensitive learning, these cost values are often adjusted by considering the reverse class distribution in the cost matrices. Cost matrix values are determined as follows: The cost matrix values with correctly classified diagonal elements $C(i,i)$ are set to 0, and the FN and FP values, representing malignant and benign classes, are assigned as one and c, respectively, where FN, representing the minority class, must have $c > 1$ [26]. Here, C denotes the regularization parameter and also serves as a regularizer to prevent overfitting [27]. Thus, the objective is to mitigate misclassification induced by imbalanced data classes by minimizing the total misclassification cost of the model.

2.6 K-Means Synthetic Minority Oversampling Technique

Maintaining a balance between classes is critical for optimal model performance in classification tasks. A dataset is considered imbalanced when the number of examples in one or more classes significantly exceeds that of others. In such cases, models trained on imbalanced datasets may lack sufficient information about the minority class, leading to biased predictions favoring the majority class. To address this issue, the Synthetic Minority Over-sampling Technique (SMOTE), which increases the number of minority class examples, has been effectively utilized in numerous studies. Although various variants of the SMOTE method exist, its core principle focuses on identifying the k-nearest neighbors of a minority class sample and generating a synthetic sample as a linear combination of the sample and its neighbors [10]. One such variant is K-SMOTE, introduced by Douzas, Bacao, and Last in 2018, which integrates k-means clustering with the SMOTE technique to generate synthetic samples for the minority class[12]. In the K-Means SMOTE method, the k nearest neighbors of the selected sample is determined from the same cluster. Thus, the aim is to mitigate the drawbacks of SMOTE, specifically the generation of noisy artificial examples, by generating synthetic data from more reliable regions identified through clusters determined by the k-means algorithm.

2.7 Proposed Method

Since it is more costly to misclassify a malignant case than to recognize a benign case as positive, a cost-sensitive hybrid deep learning structure is proposed in this study as a solution to this problem. The BUSI dataset utilized for breast cancer classification exhibits class

imbalance. Initially, a supplementary dataset named BUSICL was generated by applying the CLAHE image enhancement technique to acquire diverse features for classification purposes. This study aims to enhance model performance by utilizing enriched features from both the BUSI and BUSCL datasets. Subsequently, Vgg16 and ResNet50 models were employed to extract features from both datasets. In the subsequent step, a total of 4000 comprehensive features obtained from each dataset through both models were subjected to classification using costs-sensitive SVM via 10-fold cross-validation. To identify malignant values with minimal misclassification error during training, a cost value inversely proportional to the number of classes is assigned in the cost table. For this purpose, the classification performance of the model with the standard SVM was first assessed, followed by an update to the default cost table to account for the misclassified malignant FN values. Finally, the effectiveness of the proposed cost-sensitive approach was compared to the K-Means SMOTE oversampling method, to demonstrate its robustness. The illustration of the proposed structure is depicted in Figure 6.

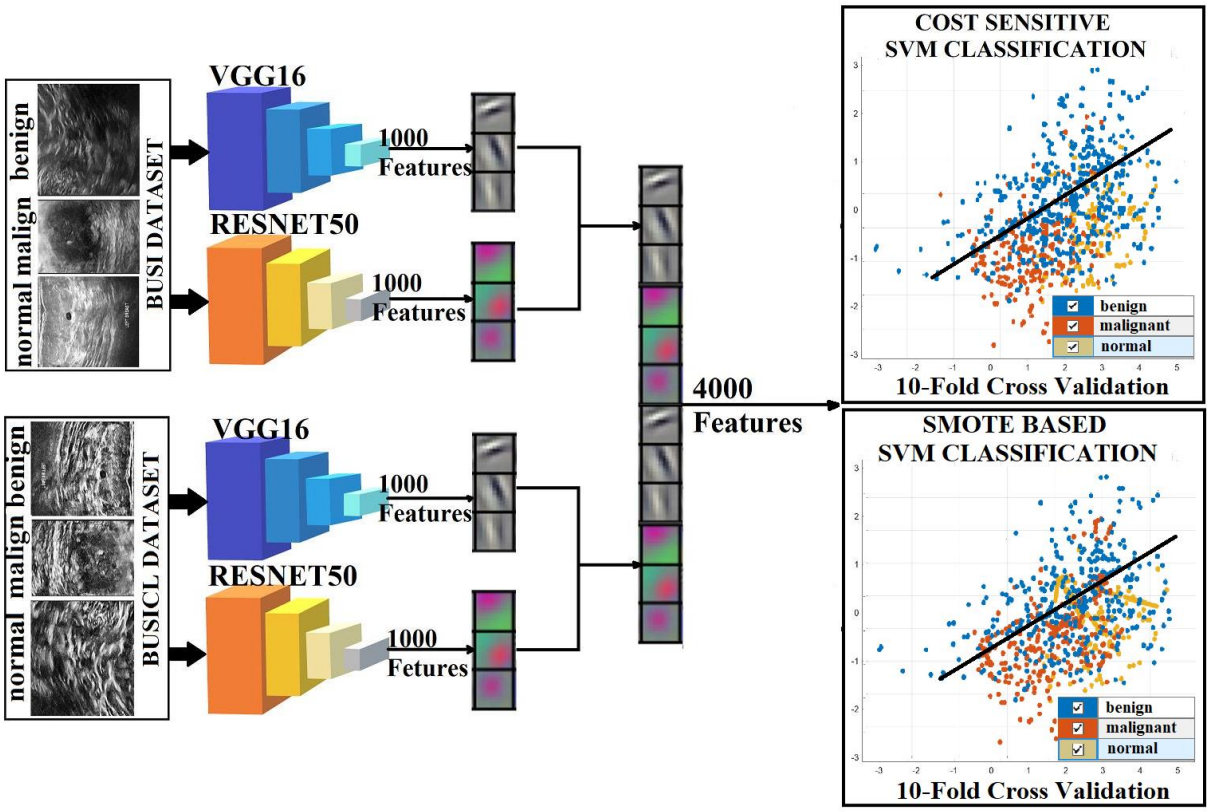

Figure 6. The Functional Block Diagram of the Proposed Approach.

3 EXPERIMENTAL RESULTS

The research was conducted utilizing MATLAB 2022 and KERAS, the high-level API of TensorFlow. The computational system employed is equipped with an RTX 3070 graphics card and 32 GB of RAM on the Windows 10 platform.

The confusion matrix table was utilized to assess the models' prediction outcomes. Subsequently, following the computation of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) values, metrics including accuracy, recall, and f-score were derived [28]. Accuracy measures the overall correctness of a model, while precision measures the accuracy of positive predictions by assessing how many predicted positives are actually positive, and recall, also known as sensitivity or true positive rate, evaluates the model's ability to correctly identify all positive cases, and the F-score is a performance measure defined as the harmonic mean of precision and recall. These metrics can be seen in Equations 2-5.

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$\mathrm{F-Scr} = \frac{2xTP}{2xTP + FP + FN} \tag{4}$$

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

In the initial phase of the study, the fine-tuned Vgg16 and ResNet50 models were trained using the datasets to extract the valuable features with optimal weights. Figure 7 illustrates the confusion matrix table corresponding to the training process in which both datasets are used separately. Specifically, the Vgg16 model achieved an accuracy of 82.05% on the BUSI dataset and 82.69% on the BUSICL dataset. Similarly, the Resnet50 model attained an accuracy of 83.97% on the BUSI dataset and 85.26% on the BUSICL dataset. The results show that the Resnet50 model outperforms Vgg16 in terms of overall accuracy, with further improvement observed when using CLAHE for image enhancement. However, as both the VGG16 and ResNet50 models exhibit significant misclassification rates, it is evident that the classification performance achieved is inadequate for effective breast cancer detection.

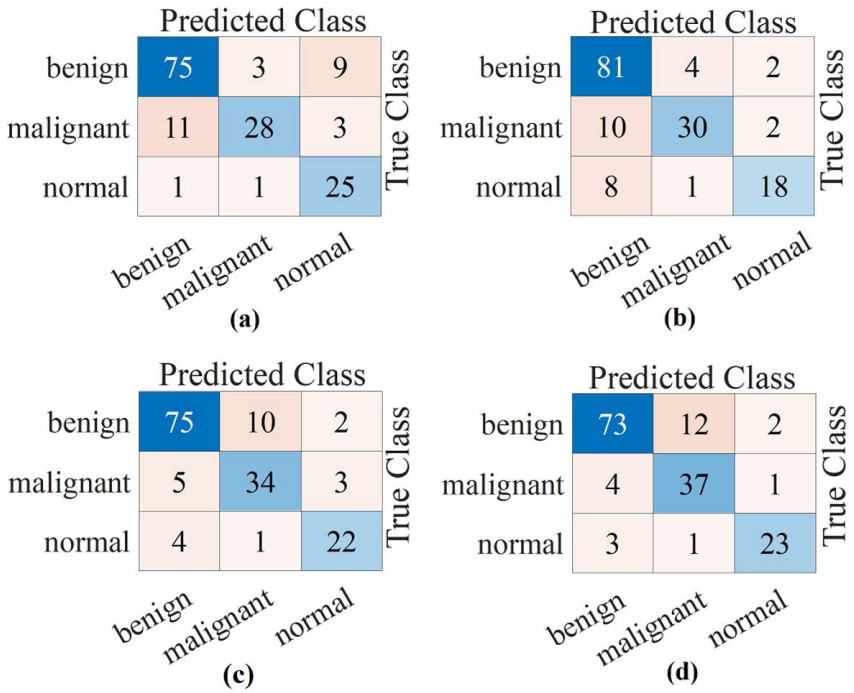

Figure 7. Confusions Matrixes, (a) Vgg16 BUSI Dataset, (b) Vgg16 BUSICL Dataset, (c) ResNet50 BUSI Dataset, (d) ResNet50 BUSICL Dataset.

In addition, Table 3 shows the precision, recall and f-score values derived from the confusion matrix data, which provide additional insight into classification performance. When looking at the class-specific evaluations, it is clear that accurately distinguishing malignancy, a crucial aspect of breast cancer classification, remains a challenge. On the other hand, despite the use of powerful computer hardware, training times are considerably longer for applications in daily life.

Table 3. BUSI, and BUSICL Datasets Classification with Deep Models.

Models	Dataset	Classes	Pre (%)	Recall (%)	F-Scr (%)	Acc (%)	Training time(s)
VGG16	BUSI	Benign	86.21	86.21	86.21	82.05	3620
		Malign	66.67	87.50	75.68		
		Normal	92.59	67.57	78.13		
	BUSICL	Benign	93.10	81.82	87.10	82.69	3450
		Malign	71.43	85.71	77.92		
		Normal	66.67	81.82	73.47		
RESNET50	BUSI	Benign	86.21	89.29	87.72	83.97	2340
		Malign	80.95	75.56	78.16		
		Normal	81.48	81.48	81.48		
	BUSICL	Benign	83.91	91.25	87.43	85.26	2210
		Malign	88.10	74.00	80.43		
		Normal	85.19	88.46	86.79		

In the second step of the study, the features derived from the BUSI and BUSICL datasets were combined using a model-based approach. A total set of 2000 features were then employed for training with the SVM algorithm, followed by comprehensive classification procedures. The confusion matrix table, resulting from the classification of 2000 features derived from Vgg16 and Resnet50 models using SVM, along with the precision, recall, F-score, and accuracy values derived from this analysis, are given in Figure 8 and Table 4, respectively. It is seen in Figure 8a and Figure 8b that employing SVM to classify the 2000 features derived from the datasets with rich feature space results in an overall accuracy increase of approximately 10% compared to using SoftMax. The classification accuracies obtained with the VGG16 and ResNet50 models for 2000 features using a standard SVM were 92.95% and 94.87%, respectively.

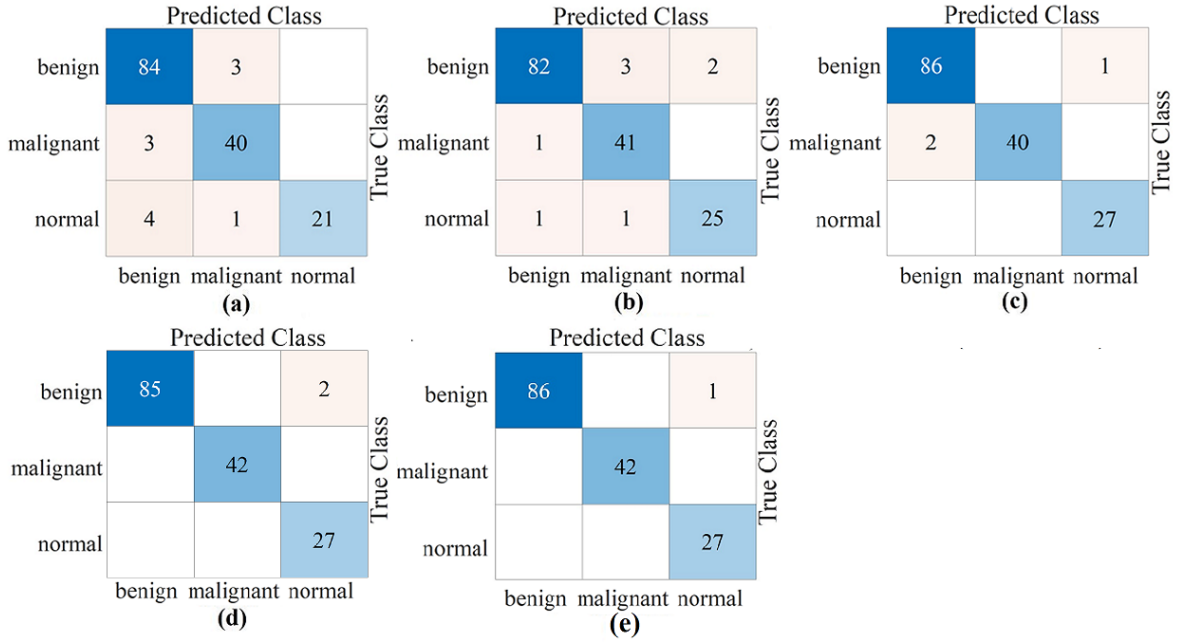

Figure 8. SVM Classification of Concatenated Features of BUSI and BUSICL Datasets, (a) Vgg16 with 2000 Features, (b) ResNet50 with 2000 Features, (c) Vgg16&ResNet50 with 4000 Features, (d)Vgg16&ResNet50 with 4000 Features and SMOTE with SVM, (e) Vgg16&ResNet50 with 4000 Features and Cost-Sensitive SVM.

In the third step of the study, a rich feature space was obtained by combining a total of 2000 features extracted from each model. The primary objective was to minimize misclassification errors, which carry high costs for the patient with the combined 4000 features. Precision, recall, F-score, and accuracy values calculated from the confusion matrix are depicted in Figure 8c and detailed in Table 4 on a class basis. Despite achieving a very high success rate compared to previous steps, costly classification error persists since two out of three errors are in malignant class when using the default setting of SVM.

In the final step of the study, the aim was to minimize misclassification errors, especially for FN evaluations, which incur heavy patient costs, by employing the cost-sensitive SVM and K-means SMOTE approach. Figure 8d shows that the K-means synthetic data oversampling method results in a misclassification error in the non-malignant class. Despite this, the outcome aligns well with the study's objectives. In the cost-sensitive approach, cost matrix values for the SVM can be manually set by considering the inverse of the class sample distribution. To this end, the cell in the cost matrix set by considering the inverse of the class sample distribution. To this end, the cell in the cost matrix corresponding to the FN value of the malignant class was set to 2, while the other values were kept as default. The classification accuracies obtained with the K-means SMOTE and the cost-sensitive approaches for 4000 features were 98.71% and 99.36%, respectively. Figures 8d and 8e present the confusion matrix table obtained with this approach. Moreover, when the training and testing times of the classifier are compared with pretrained models, it is noticeable that it provides faster results for use in mobile-based artificial intelligence applications.

Table 4. Comparative analysis of classification results using standard SVM, cost-sensitive SVM, and K-means with SMOTE-based SVM for the proposed approach.

Model	Dataset	Feature Size	Cost-Sensitive Learning	K-means Smote	Classes	Pre (%)	Recall (%)	F-Scr (%)	Acc (%)	Classifier training time(s)	Classifier testing time(s)
Vgg16	BUSI& BUSICL	2000	No	No	Benign	96.55	92.31	94.38	92.95	6.02	0.90
					Malign	93.02	90.91	91.95			
					Normal	80.77	100	89.36			
ResNet50	BUSI& BUSICL	2000	No	No	Benign	94.25	97.62	95.91	94.87	5.91	0.89
					Malign	97.62	91.11	94.25			
					Normal	92.59	92.59	92.59			
Vgg16& ResNet50	BUSI& BUSICL	4000	No	No	Benign	98.85	97.73	98.29	98.08	10.52	1.20
					Malign	95.24	100	97.56			
					Normal	100	96.43	98.18			
	BUSI& BUSICL	4000	No	Yes	Benign	97.70	100	98.84	98.71	10.70	1.10
					Malign	100	100	100			
					Normal	100	93.10	96.43			
	BUSI& BUSICL	4000	Yes	No	Benign	98.85	100	99.42	99.36	10.44	1.09
					Malign	100	100	100			
					Normal	100	96.43	98.18			

On the other hand, to evaluate the proposed model's robustness, artificial samples generated using the K-means SMOTE are presented in Figure 9 for comparison with the original samples.

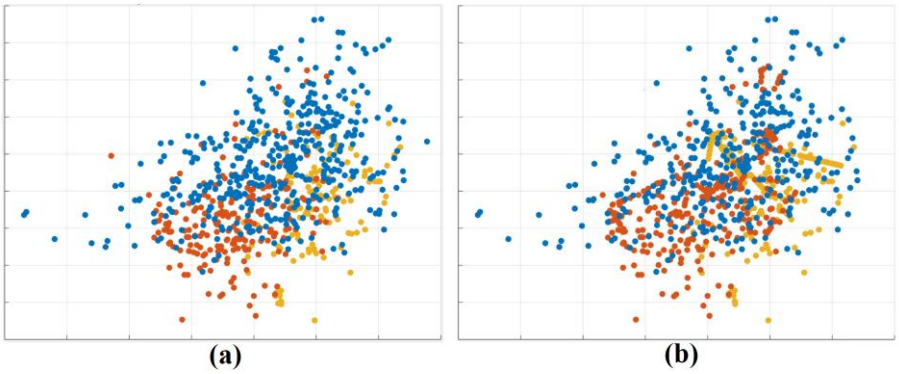

Figure 9. Feature representation of BUSI and BUSICL after concatenation. (a) Features originated from 780 samples from the datasets, (b) The train samples were synthetically increased to 1050 by K-means SMOTE.

The synthetic K-means SMOTE oversampling technique was employed to balance the class distribution within the training dataset. In this study, the number of samples in the minority classes was increased to match the majority class, which included 350 samples. As a result, the total number of samples in the training dataset reached 1050.

Table 5. Comparison with state-of-the-art studies for BUSI Dataset.

Authors	Methods	Recall (%)	F-Scr (%)	Classes Names	Acc (%)
[29]	Multi-Headed CNN, image processing, contour detection.	92	92	Benign, malign, healthy.	92.31
[30]	CNN, optimizer settings	97.37	97.25	Benign, malign.	96.24
[31]	Adaptive Feature fusion, local binary pattern features, SVM.	94.7	97.1	Benign, malign.	96.9
[32]	Traditional Feature extraction, Grid search optimization, Feature selection.	95.80	95.88	Benign, malign	97.81
[33]	Transfer Larning, Fine tuning.	98.22	98.54	Benign, malign, healthy	98.73
[8]	Over-sampling technique, feature fusion with Resnet50 hog features.	99.14	99.14	Benign, malign	99.14
[34]	Data Augmentation, Transfer Learning, Feature Selection.	99.3	99.3	Benign, malign, healthy	99.3
[9]	CNN, Cross-channel normalization techniques	99.60	99.66	Benign, malign, healthy	99.35
Proposed approach	Transfer Learning, K-Means SMOTE, Cost-Sensitive SVM, CLAHE	99.38	99.36	Benign, malign, healthy	99.36

4 DISCUSSION

Breast cancer classification using ultrasound scans often faces the challenge of class imbalance, primarily due to the limited number of malignant class samples in the datasets. The

BUSI dataset, which exhibits class imbalance, is one of the most well-known datasets in this field. Another notable limitation of the dataset is its insufficiency of samples for data-hungry deep learning models across all classes. The proposed model leverages two key approaches to address these two critical issues. Firstly, the transfer learning approach addresses underfitting issues arising from a limited number of samples during training, thereby enhancing classification accuracy. The results obtained using transfer learning methods on the BUSI dataset are presented in Table 3. The number of misclassified examples exceeds the acceptable threshold for medical studies, making the outcomes costly. In the subsequent step, the enriched feature pool addresses the class imbalance through cost-sensitive and oversampling approaches. To compare the performance of cost-sensitive and oversampling methods, SVM classifier is used for both approaches. Table 4 shows that the cost-sensitive approach is more successful than the synthetic oversampling method. Therefore, the success of the proposed model, as presented in Table 5, demonstrates that it provides a robust approach for handling imbalanced datasets.

This study successfully classified breast cancer with minimal misclassification errors through the proposed approach. This study's overall accuracy value of 99.36% achieved is particularly noteworthy, as it does not contain FN misclassification results. Numerous studies have examined the BUSI breast cancer dataset using various methods. However, this study emphasizes the necessity of considering the critical distinction of highlighting those approaches assuming equal cost values for non-malignant and malignant tumors may be insufficient, as it distinguishes it from others. This study emphasizes the classification of breast cancer with high accuracy and prioritizes the classification of malignant tumors with minimum misclassification errors on a class-specific basis. Upon examination of the study by [9], which yielded results closest to those of this study, it is noted that it includes two misclassification errors. Consequently, the results of this study suppress previous studies. Table 5 compares studies using the BUSI dataset and the recommended approach.

On the other hand, despite the proposed model's successful results, it is not an end-to-end system, as it involves multiple steps and requires manual processes. Moreover, the dataset comprises malignant, benign, and healthy classes. A more comprehensive breast cancer dataset that includes malignancy grade values would be more beneficial for assisting radiologists in their evaluations.

5 CONCLUSION AND SUGGESTIONS

The approach proposed for breast cancer classification in this study aims to reduce misclassified FN values. FN values pose a more significant challenge than other erroneous results because they delay the patient's treatment process. With this method, classification success was first increased by obtaining a richer feature pool, and then the focus shifted to reducing FN values using a cost-sensitive SVM. To our knowledge, the cost-sensitive SVM approach proposed in this study has been used for breast cancer for the first time. The results demonstrate that cost-sensitive approaches can yield successful outcomes even in the challenges arising from imbalanced datasets, which stands out as the most disadvantageous aspect of this study. The successful outcomes of this study have led to plans for further research on the cost-sensitive approach in different imbalanced datasets.

Statement of Research and Publication Ethics

The study is complied with research and publication ethics.

Artificial Intelligence (AI) Contribution Statement

This manuscript was entirely written, edited, analyzed, and prepared without the assistance of any artificial intelligence (AI) tools. All content, including text, data analysis, and figures, was solely generated by the author.

Availability of data and material

The dataset can be accessed through this link:

https://www.kaggle.com/datasets/anaselmasry/datasetbusiwithgt/data

REFERENCES

[1] M. Arnold *et al.*, "Current and future burden of breast cancer: Global statistics for 2020 and 2040," *Breast*, vol. 66, no. June, pp. 15–23, 2022, doi: 10.1016/j.breast.2022.08.010.

[2] J. A. Malik *et al.*, "Drugs repurposed: An advanced step towards the treatment of breast cancer and associated challenges," *Biomed. Pharmacother.*, vol. 145, p. 112375, 2022, doi: 10.1016/j.biopha.2021.112375.

[3] M. Eghtedari, A. Chong, R. Rakow-Penner, and H. Ojeda-Fournier, "Current status and future of BI-RADS in multimodality imaging, from the AJR special series on radiology reporting and data systems," *Am. J. Roentgenol.*, vol. 216, no. 4, pp. 860–873, 2021, doi: 10.2214/AJR.20.24894.

[4] D. R. Chen *et al.*, "Classification of breast ultrasound images using fractal feature," *Clin. Imaging*, vol. 29, no. 4, pp. 235–245, 2005, doi: 10.1016/j.clinimag.2004.11.024.

[5] Y. L. Huang and D. R. Chen, "Support vector machines in sonography: Application to decision making in the diagnosis of breast cancer," *Clin. Imaging*, vol. 29, no. 3, pp. 179–184, 2005, doi: 10.1016/j.clinimag.2004.08.002.

[6] 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 Obstet. Gynecol.*, vol. 32, no. 4, pp. 565–572, 2008, doi: 10.1002/uog.5205.

[7] M. Ragab, A. Albukhari, J. Alyami, and R. F. Mansour, "Ensemble Deep-Learning-Enabled Clinical Decision Support Ultrasound Images," *Biology (Basel).*, vol. 11, p. 439, 2022.

[8] B. Abhisheka, S. K. Biswas, and B. Purkayastha, "HBMD-Net: Feature Fusion Based Breast Cancer Classification with Class Imbalance Resolution," *J. Imaging Informatics Med.*, no. 0123456789, 2024, doi: 10.1007/s10278-024-01046-5.

[9] A. Raza, N. Ullah, J. A. Khan, M. Assam, A. Guzzo, and H. Aljuaid, "DeepBreastCancerNet: A Novel Deep Learning Model for Breast Cancer Detection Using Ultrasound Images," *Appl. Sci.*, vol. 13, no. 4, 2023, doi: 10.3390/app13042082.

[10] G. Douzas, F. Bacao, and F. Last, "Improving imbalanced learning through a heuristic oversampling method based on k-means and SMOTE," *Inf. Sci. (Ny).*, vol. 465, pp. 1–20, 2018, doi: 10.1016/j.ins.2018.06.056.

[11] Z. Xu, D. Shen, T. Nie, Y. Kou, N. Yin, and X. Han, "A cluster-based oversampling algorithm combining SMOTE and k-means for imbalanced medical data," *Inf. Sci. (Ny).*, vol. 572, pp. 574–589, 2021, doi: 10.1016/j.ins.2021.02.056.

[12] A. Arafa, N. El-Fishawy, M. Badawy, and M. Radad, "RN-SMOTE: Reduced Noise SMOTE based on DBSCAN for enhancing imbalanced data classification," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 8, pp. 5059–5074, 2022, doi: 10.1016/j.jksuci.2022.06.005.

[13] C. C. Chang, Y. Z. Li, H. C. Wu, and M. H. Tseng, "Melanoma Detection Using XGB Classifier Combined with Feature Extraction and K-Means SMOTE Techniques," *Diagnostics*, vol. 12, no. 7, 2022, doi: 10.3390/diagnostics12071747.

[14] I. D. Mienye and Y. Sun, "Performance analysis of cost-sensitive learning methods with application to imbalanced medical data," *Informatics Med. Unlocked*, vol. 25, 2021, doi: 10.1016/j.imu.2021.100690.

[15] N. Liu, J. Shen, M. Xu, D. Gan, E. S. Qi, and B. Gao, "Improved Cost-Sensitive Support Vector Machine Classifier for Breast Cancer Diagnosis," *Math. Probl. Eng.*, vol. 2018, 2018, doi: 10.1155/2018/3875082.

[16] V. Ravi, "Attention Cost-Sensitive Deep Learning-Based Approach for Skin Cancer Detection and Classification," *Cancers (Basel).*, vol. 14, no. 2, p. 5872, 2022, doi: doi.org/10.3390/cancers14235872.

[17] W. Al-Dhabyani, M. Gomaa, H. Khaled, and A. Fahmy, "Dataset of breast ultrasound images," *Data Br.*, vol. 28, p. 104863, 2020, doi: 10.1016/j.dib.2019.104863.

[18] Y. R. Haddadi, B. Mansouri, and F. Z. I. Khodja, "A novel medical image enhancement algorithm based on CLAHE and pelican optimization," *Multimed. Tools Appl.*, no. 0123456789, 2024, doi: 10.1007/s11042-024-19070-6.

[19] D. Garg, N. K. Garg, and M. Kumar, "Underwater image enhancement using blending of CLAHE and percentile methodologies," *Multimed. Tools Appl.*, vol. 77, no. 20, pp. 26545–26561, 2018, doi: 10.1007/s11042-018-5878-8.

[20] D. Theckedath and R. R. Sedamkar, "Detecting Affect States Using VGG16, ResNet50 and SE-ResNet50 Networks," *SN Comput. Sci.*, vol. 1, no. 2, pp. 1–7, 2020, doi: 10.1007/s42979-020-0114-9.

[21] N. Muzoğlu, M. K. Karaslan, A. M. Halefoğlu, And S. Yarman, "Prediction of the Prognosis of Covid-19 Disease Using Deep Learning Methods and Boruta Feature Selection Algorithm," *Afyon Kocatepe Univ. J. Sci. Eng.*, vol. 22, no. 3, pp. 577–587, 2022, doi: 10.35414/akufemubid.1114346.

[22] W. Gómez-Flores, M. J. Gregorio-Calas, and W. Coelho de Albuquerque Pereira, "BUS-BRA: A breast ultrasound dataset for assessing computer-aided diagnosis systems," *Med. Phys.*, vol. 51, no. 4, pp. 3110–3123, 2024, doi: 10.1002/mp.16812.

[23] Y. Peng, L. Zhang, S. Liu, X. Wu, Y. Zhang, and X. Wang, "Dilated Residual Networks with Symmetric Skip Connection for image denoising," *Neurocomputing*, vol. 345, pp. 67–76, 2019, doi: 10.1016/j.neucom.2018.12.075.

[24] S. Ruder, "An overview of gradient descent optimization algorithms," Sep. 2016, [Online]. Available: http://arxiv.org/abs/1609.04747.

[25] A. Iranmehr, H. Masnadi-Shirazi, and N. Vasconcelos, "Cost-sensitive support vector machines," *Neurocomputing*, vol. 343, pp. 50–64, 2019, doi: 10.1016/j.neucom.2018.11.099.

[26] Zhou, Zhi-Hua, and Xu-Ying Liu. "On multi-class cost-sensitive learning." *Computational Intelligence,* 26.3, pp. 232-257, 2010, doi: 10.1111/j.1467-8640.2010.00358.

[27] Mienye, Ibomoiye Domor, and Yanxia Sun. "Performance analysis of cost-sensitive learning methods with application to imbalanced medical data." *Informatics in Medicine Unlocked,*vol. 25, 2021, doi: /doi.org/10.1016/j.imu.2021.100690.

[28] M. N. Hossin, M., Sulaiman, "A Review of Evaluation Metrics in Machine Learning Algorithms," *Med. Image Anal.*, vol. 80, no. 2, p. 102478, 2022, [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1361841522001256.

[29] R. K. Pathan *et al.*, "Breast Cancer Classification by Using Multi-Headed Convolutional Neural Network Modeling," *Healthc.*, vol. 10, no. 12, pp. 1–17, 2022, doi: 10.3390/healthcare10122367.

[30] A. K. Mishra, P. Roy, S. Bandyopadhyay, and S. K. Das, "Achieving highly efficient breast ultrasound tumor classification with deep convolutional neural networks," *Int. J. Inf. Technol.*, vol. 14, no. 7, pp. 3311–3320, 2022, doi: 10.1007/s41870-022-00901-4.

[31] H. Chen, M. Ma, G. Liu, Y. Wang, Z. Jin, and C. Liu, "Breast Tumor Classification in Ultrasound Images by Fusion of Deep Convolutional Neural Network and Shallow LBP Feature," *J. Digit. Imaging*, vol. 36, no. 3, pp. 932–946, 2023, doi: 10.1007/s10278-022-00711-x.

[32] H. Özcan, "BUS-CAD: A computer-aided diagnosis system for breast tumor classification in ultrasound images using grid-search-optimized machine learning algorithms with extended and Boruta-selected features," *Int. J. Imaging Syst. Technol.*, vol. 33, no. 5, pp. 1480–1493, 2023, doi: 10.1002/ima.22873.

[33] F. Z. Reguieg and N. Benblidia, *Ultrasound breast tumoral classification by a new adaptive pre-trained convolutive neural networks for computer-aided diagnosis*, vol. 83, no. 15. Springer US, 2024.

[34] K. Jabeen *et al.*, "Breast Cancer Classification from Ultrasound Images Using Probability-Based Optimal Deep Learning Feature Fusion," *Sensors*, vol. 22, no. 3, 2022, doi: 10.3390/s22030807.