Araştırma Makalesi
Adıyaman Üniversitesi
Mühendislik Bilimleri Dergisi
21 (2023) 299-306

A NOVEL METHOD FOR BREAST CANCER CLASSIFICATION: A SIGNAL PROCESSING-BASED APPROACH IN ULTRASOUND IMAGES

Şerife GENGEÇ BENLİ[1*], Zeynep AK[2]

[1]Erciyes University, Engineering Faculty, Biomedical Engineering, Kayseri, Turkey

[2]Erciyes University, Graduate School of Natural and Applied Sciences, Biomedical Eng. Dept., Kayseri, Turkey
Geliş Tarihi/Received Date: 23.10.2023 Kabul Tarihi/Accepted Date: 11.12.2023 DOI: 10.54365/adyumbd.1378982

ABSTRACT

Breast cancer, a leading cause of mortality among women worldwide, the importance of accurate and efficient diagnostic methods is emphasized. This study contributes to the literature on breast cancer classification, particularly using breast ultrasound images, with a new method using a signal processing approach. It introduces a novel approach by combining features extracted from signals obtained from breast ultrasound images with signals from Variational Mode Decomposition (VMD) sub-bands. The results demonstrate that utilizing features from both preprocessed raw data and VMD sub-band signals can effectively distinguish benign and malignant breast ultrasound images. Classification performance varied depending on the algorithms and data used. According to the numerical results, the highest classification performance was achieved through the study with balanced data using the artificial neural network method, with an area under the curve value of 0.9971 and an accuracy value of 0.9821.

Keywords: *Breast ultrasound images, Variational mode decomposition, Sub-bands, Classification.*

MEME KANSERİ SINIFLANDIRMASI İÇİN YENİ BİR YÖNTEM: ULTRASON GÖRÜNTÜLERİNDE SİNYAL İŞLEME TEMELLİ BİR YAKLAŞIM

ÖZET

Meme kanseri, dünya genelinde kadınlar arasında ölümün önde gelen nedenlerinden biri olup, doğru ve etkili tanı yöntemlerinin önemi vurgulanmaktadır. Bu çalışma, özellikle meme ultrason görüntülerini kullanarak meme kanseri sınıflandırması alanındaki literatüre yeni bir sinyal işleme yaklaşımı kullanan yöntem ile katkı sağlamaktadır. Çalışma, meme ultrason görüntülerinden elde edilen sinyaller ve Varyasyonel Kip Ayrışımı (VMD) alt bantlarından elde edilen sinyalleri kullanan yeni bir yaklaşım sunmaktadır. Elde edilen sonuçlar ile hem orijinal veriden hem de VMD alt bant sinyallerinden elde edilen özelliklerin iyi huylu ve kötü huylu meme ultrason görüntülerini etkili bir şekilde ayırt edilebileceği gösterilmiştir. Kullanılan algoritma ve uygulanan verilere göre elde edilen sınıflandırma performansları değişmektedir. Sayısal sonuçlara göre, dengelenmiş veriler kullanılarak yapay sinir ağları yöntemi ile yapılan çalışma sonucunda en yüksek sınıflandırma performansı elde edilmiş olup, eğri altında kalan alan değeri 0.9971 ve doğruluk değeri 0.9821 olarak elde edilmiştir.

Anahtar Kelimeler: *Meme ultrason görüntüleri, Varyasyonel kip ayrışımı, Altbantlar, Sınıflandırma.*

1. Introduction

Breast cancer is a prominent cause of mortality among women globally [1,2]. The effectiveness of treatment greatly relies on the accuracy and reliability of early diagnosis, along with the capability to

[*] e-posta[1] : serifegengec@erciyes.edu.tr ORCID ID: https://orcid.org/0000-0002-5527-8574 (Corresponding Author)
e-posta[2] : zeynepak.bmm@gmail.com ORCID ID: https://orcid.org/0000-0001-8621-9465

differentiate between benign and malignant tumors. Commonly used medical imaging methods for breast tissues are known as mammography and ultrasound. Ultrasound is widely regarded as a primary imaging modality for the characterization of breast lesions due to its widespread accessibility, cost-effectiveness, satisfactory diagnostic accuracy, and ability to provide noninvasive imaging capabilities. It allows real-time imaging of breast lesions from different angles and orientations, reducing the risk of false-negative diagnoses [3]. On the other hand, ultrasound waves cannot penetrate the regular muscle system, leading to an inaccurate depiction of the boundaries of breast tumors. Extracting a region of interest (ROI) for subsequent research can also be challenging for imaging. Additionally, it is susceptible to speckle noise, which can make working with ultrasound images difficult.

When reviewing the literature on breast cancer classification using on the breast ultrasound images (BUSI) dataset, it is observed that there is an increasing trend towards machine learning and deep learning approaches in BUSI. Payithra et al., initially attempted to reduce speckle noise in the images using speckle-reducing anisotropic diffusion. Subsequently, active contour-based segmentation was employed in the study to find the ROI. Texture features were extracted to classify the images as normal, benign, or malignant. Three classifiers were utilized, and performance was compared based on classification accuracy. They achieved an accuracy of 83% for the k-Nearest Neighbors (kNN) algorithm, 85% for the decision tree algorithm, and 88% for the random forest classifier [4]. Mishra et al., performed feature selection using a recursive feature elimination-based method by extracting multiple different image features from the tumor region. They classified benign and malignant tumors with Random Forest, Adaboost, and Gradient Boosting classifiers, achieving accuracy rates of 96.7%, 97.4%, and 96.5%, respectively [5]. In their study, Lo et al., extracted breast tissue features from a dataset comprising 48 benign and 21 malignant cases, asserting that these features are more beneficial for clinical diagnosis. They achieved an accuracy of 80.0% in classification using Logistic Regression [6]. Huang et all., extracted 73 features using five breast ultrasound image characteristics, which included grey-level histogram, Grey Level Co-occurrence Matrix (GLCM), histogram of oriented gradients (HOG), shape, and position. To achieve better results in this study conducted with 46 breast ultrasound images diagnosed with tumors, feature selection was performed using the Bicluster score in order to select the best features. Using the selected top 25 features in conjunction with an SVM classifier, they achieved an accuracy of 98.3% [7]. In another study, Huang et al., extracted grey histograms, Grey Level Co-occurrence Matrices (GLCM), and Local Binary Patterns (LBP) from images generated by superpixels. In a classification using the kNN algorithm, which included 160 breast ultrasound (BUS) images in both benign and malignant groups, they achieved an accuracy of 86.5% [8]. Liu et al., utilized edge information for breast ultrasound classification. They generated edge profiles of breast ultrasound images and extracted edge features, including maximum curvature sum, maximum curvature and peak sum, maximum curvature sum, and standard deviation. Subsequently, morphological features were extracted, and classification was performed using the SVM algorithm. To evaluate their method, they used 192 BUS images, achieving an accuracy of 82.69% with edge-based features and 67.31% with morphological features [9]. Kriti et al., examined the impact of different speckle filters on the classification of BUSI, resulting in a total of 149 texture features and 13 morphological features. They conducted classification using SVM and employed Principal Component Analysis (PCA) to select features from different structures. The study was carried out on a total of 100 breast ultrasound images, and the results obtained for different filters were reported as 94.1%, 66.6%, 96.0%, and 68.6% accuracy, respectively [10].

Some researchers have focused on BUSI using a segmentation additionally classification frameworks. Yi et al., designed their work around segmentation with CRA-ENet and classification with SA-Net. CRA-ENet is a network derived from ENet by incorporating a hybrid attention mechanism that can accurately segment breast tumor boundaries and allow precise localization of lesion regions. They achieved a segmentation accuracy of 97.47% and a classification accuracy of 95.84% for BUSI [11]. Sadad et al., applied the Hilbert transform to raw ultrasound images and then applied the watershed transform for segmentation. Methods based solely on texture analysis are quite sensitive to speckle noise and other artifacts. Therefore, a hybrid feature set was developed after the extraction of shape-based and texture features from the breast lesion. They classified using kNN, decision tree, and ensemble algorithms, achieving accuracies of 96.6%, 94.90%, and 97.86%, respectively [12]. Furthermore, as seen in the literature, it is clear that the inclination towards deep learning studies in this subject is increasing day by day [11,13–15].

Upon examining the literature studies, it is observed that different datasets and image processing algorithms are used, and classification studies are conducted in the context of distinguishing BUS images. In the proposed study, the aim is to contribute to the literature in the field of classifying benign and malignant BUS images based on signals obtained from BUS images using signal processing methods. With the use of the employed dataset and methods, the study aims to achieve successful results. In this study, a new method for classifying cancer types based on BUS images has been proposed using the BUSI [16] dataset. For the first time in the literature, the study introduces an innovative approach where various features extracted from signals obtained from BUS images and signals from Variational Mode Decomposition (VMD) sub-bands are extracted and utilized for classification.

The main contributions of this study to the literature can be summarized as follows:
Unlike the other literature studies, this study focuses on breast cancer classification using signal processing methods from breast ultrasound images. This can assist in better analysis of breast ultrasound images and reaching more precise results using the proposed method. This method presents a novel approach by combining different signal features from the BUS image dataset using VMD sub-band signals for classification. This can lead to better classification results.

2. Material and Method

Dataset, preprocessing, feature extraction, data balancing, feature selection and classification methods used for the proposed method are explained in this section.

2.1. Dataset

In this study, data for the analysis of breast lesions were obtained from the publicly available BUSI dataset, which was collected with Dhabyani et al [16]. The dataset includes 780 images, each with a resolution of 500x500 pixels and consists of three categories: normal (133 images), malignant (210 images), and benign (487 images). The aforementioned images were acquired by the use of two LOGIQ E9 ultrasound machines, including a sample size of 600 female patients ranging in age from 25 to 75. Mask images are provided alongside the raw images, making them suitable for use in segmentation or detection purposes. In this recommended study, raw and masked images of 437 benign and 210 malignant images have been used. Classification of raw BUS images and mask BUS images was conducted performing two different methods.

2.2. Preprocessing of Breast Ultrasound Images

Ultrasound images inherently contain speckle noise and are sensitive to segmentation and classification algorithms due to their low contrast. Therefore, various preprocessing steps have been applied to the BUS images. First, noise elimination was attempted using the widely used median filter in ultrasound images. The reason for choosing the median filter is its ability to remove noise while being sensitive to edges. This is crucial because the boundary between tumor tissue and normal tissue plays a significant role in feature extraction [17]. Subsequently, gray level contrast enhancement was applied to the low-contrast ultrasound images to increase the differentiation between tumor tissue and normal tissue [18].

2.3. Sub-band Decomposition Method

In this study, features were extracted from preprocessed raw (p-raw) BUS images and mask images. Two types of images were converted into signals. P-raw and sub-band signals belonging to two types of images were used for feature source classification of tasks. Variational mode decomposition (VMD) was utilized as the sub-band decomposition method described below.

2.3.1. Variational Mode Decomposition

Dragomiretskiy et al. [19] recently developed the VMD approach, which is a data-driven and adaptive signal decomposition method used for separating complex signals into simpler modes, each with a specific frequency and amplitude content. VMD uses an iterative optimization algorithm to decompose the signal into a finite number of modes, each of which is represented by a complex-valued function that oscillates around its local mean. VMD has been applied to a variety of signal processing tasks, including time-frequency analysis, denoising, feature extraction, and classification. The method is particularly useful for separating and analyzing signals with non-stationary and non-linear characteristics. The method is described in depth, together with the relevant mathematical formulae, in the cited work [19].

Signals obtained from p-raw BUS images and mask BUS images were decomposed into different sub-bands. Since the classification performance was optimal at nine for this proposed study, the obtained signals were decomposed into nine sub-bands using the VMD.

2.4. Feature Extraction

For using in the classification with various approaches, during the feature extraction stage, a total of 14 features were acquired from the signals of both the p-raw image and mask image, and 9x14 features were extracted from the VMD sub-bands. These extracted features are mobility, activity, entropy, mean, standard deviation, variation, kurtosis, skewness, maximum, minimum, energy, median, log entropy, and Shannon entropy.

2.5. Data Balancing

Classifying biomedical data is extremely difficult since this data type is typically large and unbalanced. Common methods for resolving the class imbalance include down-sampling and data augmentation. In the proposed study, Adaptive Synthetic sampling approach (ADASYN) was used for data augmentation to generate synthetic minority samples, hence achieving a balanced distribution of samples across the various categories into which the dataset was partitioned [20].

2.6. Feature Selection

The present research used the Least Absolute Shrinkage and Selection Operator (LASSO) technique to conduct feature selection. The LASSO, a statistical technique proposed by Tibshirani [21], is used in regression analysis for the purpose of estimating parameters and selecting variables. This feature selection method has the potential to provide an analytical solution and a low-variance estimate that is easily interpretable within the framework of linear regression.

2.7. Classification

Following the feature extraction and selection processes described previously, a number of machine learning techniques were implemented, and their ability to classify distinct tasks was compared. Support vector machines, multi-perception artificially intelligent networks and Naive Bayes were among these algorithms. The employed algorithms are explained concisely here.

Support Vector Machines (SVMs) are a class of supervised machine learning algorithms that excel in classification tasks by finding an optimal hyperplane in a high-dimensional feature space. They are particularly effective when the dataset is non-linearly separable or when a clear margin of separation is desired [22].

Naive Bayes (NB) is a probabilistic machine learning algorithm based on Bayes' theorem. It is widely used for classification tasks in biomedical applications such as disease diagnosis and medical image analysis. In the context of classification, it helps determine the probability of a data point belonging to a particular class given its features [23].

Artificial neural networks (ANNs) have found numerous applications in the field of biomedicine due to their ability to model complex relationships within data. ANNs may be structured with either a single or several layers. These networks are composed of processing units, also known as nodes or neurons, which are coupled by changeable weights. These weights enable signals to propagate through the network in a parallel and sequential manner. In general, ANNs may be categorized into three layers of neurons: the input layer, which receives information; the hidden layer, which is responsible for extracting patterns and doing the majority of internal processing; and the output layer, which creates and delivers the final network outputs [24].

In this research, we used the Bayesian optimization approach to find values for all of the classifier parameters. The feature vector used as input to the classifier was normalized such that its mean was zero and its standard deviation was one. The study's framework is shown in Figure 1.

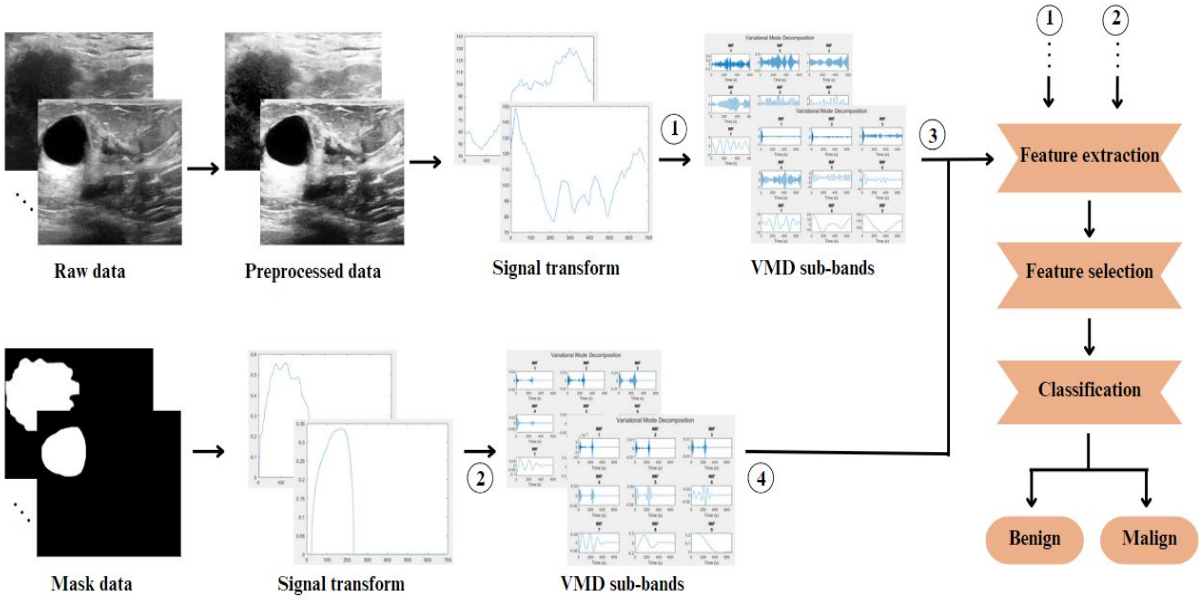

Figure 1. Representation of the classification process of raw data and mask data belonging to BUS images

3. Results and Discussion

This study was performed using the BUS images of 437 benign and 210 malignant. The p-raw images and mask images were transformed to signals and these obtained signals decomposed into nine sub-bands using VMD method. In the feature extraction stage, various features of the p-raw signal and its corresponding 9 VMD sub-bands obtained from each images (mask and p-raw). In addition, among obtained these features, the most important ones have been identified for classification using the LASSO method in both two approaches. The success of VMD sub-bands and p-raw signal in distinguishing benign and malign BUS images has been examined using different approaches, including SVM, Naive bayes, and ANNs. The classification performance measures, such as Area Under the Curve (AUC), accuracy (ACC), precision, and F1 score obtained from binary classification have been evaluated. Tables 1 and 2 show mean classification performance as binary for all images as p-raw signals and VMD sub-band signals, respectively. The examination of these tables reveals the classification performance achieved when utilizing features from both the p-raw signals and the VMD sub-band signals. Throughout the classification study, the training and testing datasets were identified using a 10-fold cross-validation approach. There were ten iterations of each classification procedure, and the provided tables show the average performance in the classification results.

Table 1. Binary classification results with features obtained from signals which transformed p-raw and mask BUS images

		Classification				Classification with ADASYN			
Data type	Classification methods	AUC	Accuracy	Precision	F1	AUC	Accuracy	Precision	F1
p-raw image ①	SVM	0.7133	0.6931	0.6131	0.2414	0.7117	0.6522	0.6546	0.6249
	ANN	0.7129	0.7032	0.5781	0.4181	0.8514	0.8045	0.7748	0.8066
	NB	0.5890	0.6818	0.6385	0.1907	0.6610	0.6240	0.6066	0.6211
Mask image ②	SVM	0.8739	0.8352	0.8114	0.7164	0.8489	0.7678	0.7690	0.7436
	ANN	0.8725	0.8267	0.7798	0.7088	0.8842	0.8233	0.7941	0.8164
	NB	0.7981	0.7619	0.7137	0.5532	0.7629	0.70	0.6641	0.6929

Table 2. Binary classification results with features obtained from VMD sub-band signals which transformed p-raw and mask BUS images

		Classification				Classification with ADASYN			
Data type	Classification methods	AUC	Accuracy	Precision	F1	AUC	Accuracy	Precision	F1
p-raw image ③	SVM	0.7112	0.7285	0.6411	0.4747	0.7395	0.6821	0.6788	0.6808
	ANN	0.7146	0.7315	0.6303	0.5080	0.9201	0.8504	0.8088	0.8586
	NB	0.6350	0.6876	0.6118	0.2808	0.7325	0.6864	0.6742	0.6959
Mask image ④	SVM	0.9850	0.9547	0.9624	0.9276	0.9870	0.9592	0.9604	0.9596
	ANN	0.9811	0.9540	0.9533	0.9274	0.9971	0.9821	0.9689	0.9826
	NB	0.9153	0.8579	0.7811	0.7812	0.8897	0.8329	0.8358	0.8341

4. Conclusion

In this study, BUS images, consisting of 437 benign and 210 malignant cases, were subjected to a comprehensive analysis using a signal processing-based approach. The p-raw images and corresponding mask images were transformed into signals and decomposed into nine sub-bands via the VMD method. Feature extraction was performed on both the p-raw signal and the nine VMD sub-bands, for each image (mask and p-raw). Using the LASSO method, the most significant features were determined for classification in both approaches.

In Table 1, classification results using features derived from signals transformed from both p-raw and mask BUS images are provided. The results show varying levels of performance, with the mask image features generally outperforming the p-raw image features, particularly when employed in conjunction with ANN or SVM classifiers. Table 2 presents classification results using features extracted from VMD sub-band signals transformed from both p-raw and mask BUS images.

When examining the given tables, it is observed that the ADASYN method improved the classification performance with ANN on the p-raw images. For example, the ACC value increased from 0.7032 to 0.8045. However, no significant change in classification performance was observed on the masked images. In the classification obtained using VMD signals with ANN on the p-raw images, the ACC value increased from 0.7315 to 0.8504 as a result of the ADASYN method. Through data balancing and classification model with the VMD-ANN method on masked images, the ACC value increased from 0.9540 to 0.9821. Based on the findings, the best classification performance was obtained by employing the ANN method on balanced data, yielding exceptional results, including an AUC value of 0.9971 and an accuracy of 0.9821.

In conclusion, this study showcases the potential of signal processing methods, particularly the VMD approach, in distinguishing between benign and malignant breast ultrasound images. The classification performance improvements, especially when using the mask images and VMD sub-band signals, highlight the promise of this methodology.

These findings provide valuable insights for future research and clinical applications in breast cancer diagnosis and classification, further enhancing the role of signal processing in medical imaging.

Data availability statement

The performed breast ultrasound dataset, generated by Al-Dhabyani et al., can be downloaded from https://doi.org/10.1016/j.dib.2019.104863 (accessed on 10 October 2021).

Conflict of interest

The authors declare no conflict of interest.

References

[1] Fitzmaurice C, Dicker D, et al. The Global Burden of Cancer 2013. JAMA Oncol. 2015;1(4):505–527.

[2] Lima SM, Kehm RD, Terry MB. Global breast cancer incidence and mortality trends by region, age-groups, and fertility patterns. EClinicalMedicine. 2021;7:38:100985.

[3] Gong X, Zhou H, Gu Y, Guo Y. Breast ultrasound image classification with hard sample generation and semi-supervised learning. Biomedical Signal Processing and Control. 2023;86:105196.

[4] Pavithra S, Vanithamani R, Justin J. Computer aided breast cancer detection using ultrasound images. Materials Today. 2020;33(7):4802–4807.

[5] Mishra A, Roy P, Bandyopadhyay S, Das S. Breast ultrasound tumour classification: A Machine Learning—Radiomics based approach. Expert Systems. 2021;38:e12713.

[6] Lo CM, Chang RF, Huang CS, Moon WK. Computer-Aided Diagnosis of Breast Tumors Using Textures from Intensity Transformed Sonographic Images. In: 1st Glob. Conf. Biomed. Eng. 9th Asian-Pacific Conf. Med. Biol. Eng. Springer International Publishing, Cham. 2015;124–127.

[7] Huang Q, Yang F, Liu L, Li X. Automatic segmentation of breast lesions for interaction in ultrasonic computer-aided diagnosis. Information Sciences. 2015;314:293–310.

[8] Huang Q, Huang Y, Luo Y, Yuan F, Li X. Segmentation of breast ultrasound image with semantic classification of superpixels. Med Image Anal. 2020;61:101657.

[9] Liu Y, Ren L, Cao X, Tong Y. Breast tumors recognition based on edge feature extraction using support vector machine. Biomedical Signal Processing and Control. 2020;58:101825.

[10] Kriti, Virmani J, Agarwal R. Effect of despeckle filtering on classification of breast tumors using ultrasound images. Biocybernetics and Biomedical Engineering. 2019;39(2):536–560.

[11] Yi S, Chen Z, Yi L, She F. CAS: Breast Cancer Diagnosis Framework Based on Lesion Region Recognition in Ultrasound Images. Journal of King Saud University - Computer and Information Sciences. 2023;35(8):101707.

[12] Sadad T, Hussain A, Munir A, Habib M, Ali Khan S, Hussain S, Yang S, Alawairdhi M. Identification of Breast Malignancy by Marker-Controlled Watershed Transformation and Hybrid Feature Set for Healthcare. Applied Sciences. 2020;10(6):1900.

[13] Pacal I. Deep Learning Approaches for Classification of Breast Cancer in Ultrasound (US) Images. Iğdır Üniversitesi Fen Bilimleri Enstitüsü Dergisi. 2022;12(4):1917–1927.

[14] Jiménez-Gaona Y, Rodríguez-Álvarez MJ, Lakshminarayanan V. Deep-Learning-Based Computer-Aided Systems for Breast Cancer Imaging: A Critical Review. Applied Sciences. 2020;10(22):8298.

[15] Zhang G, Zhao K, Hong Y, Qiu X, Zhang K, Wei B. SHA-MTL: soft and hard attention multi-task learning for automated breast cancer ultrasound image segmentation and classification. International Journal of Computer Assisted Radiology Surgery. 2021;16(10):1719–1725.

[16] Al-Dhabyani W, Gomaa M, Khaled H, Fahmy A. Dataset of breast ultrasound images, Data in Brief. 2020;28:104863.
[17] Khusna DA, Nugroho HA, Soesanti I. Performance analysis of edge and detailed preserved speckle noise reduction filters for breast ultrasound images. 2015 2nd International Conference on Information Technology Computer, and Electrical Engineering 2015:76–80.
[18] Gupta S, Kaur Y. Review of Different Local and Global Contrast Enhancement Techniques for a Digital Image. International Journal of Computer Applications. 2014;100(18):18–23.
[19] Dragomiretskiy K, Zosso D. Variational Mode Decomposition. IEEE Transactions on Signal Processing. 2014;62(3):531–544.
[20] He H, Bai Y, Garcia EA, Li S. ADASYN: Adaptive synthetic sampling approach for imbalanced learning. in: 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence). 2008:1322–1328.
[21] Tibshirani R. Regression Shrinkage and Selection Via the Lasso. Journal of the Royal Statistical Society: Series B. 1996;58(1):267–288.
[22] Cortes C, Vapnik V. Support-vector networks. Machine Learning. 1995;20:273–297.
[23] Rish I. An Empirical Study of the Naïve Bayes Classifier. IJCAI 2001 Workshop Empiral Methods in Artificial Intelligence. 2001;3(22):41-46.
[24] Zhang G, Hu MY, Eddy Patuwo B, Indro DC. Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis. European Journal of Opererational Research. 1999:116(1):16–32.