Payame Noor University

Control and Optimization in Applied Mathematics (COAM)
Vol. 6, No. 1, Winter-Spring 2021 (11-29), ©2016 Payame Noor University, Iran
DOI. 10.30473/coam.2021.60533.1175 (Cited this article)

Research Article

A Decision Support System Framework Based on Text Mining and Decision Fusion Techniques to Classify Breast Cancer Patients

Mostafa Boroumandzadeh[1], Elham Parvinnia[2*], Reza Boostani[3], Sepideh Sefidbakht[4]

[1,2]Department of Computer Engineering, Shiraz Branch, Islamic Azad University, Shiraz, Iran
[3]Biomedical Group, CSE & IT Department, ECE Faculty, Shiraz University, Shiraz, Iran
[4]Department of Radiology, Medical imaging research center, Shiraz University of Medical Sciences, Shiraz, Iran

Received: August 27, 2021; **Accepted:** November 19, 2021.

Abstract. Medical decision support systems (MDSS) are designed to assist physicians in making accurate decisions. The required data by MDSS are collected from various resources such as physical examinations and electronic health records (EHR). In this paper, an MDSS framework has been proposed to diagnose and classify breast cancer patients (DSS-BC). Medical texts reports (MTR) were embedded, and essential feature vectors combined with EHR were extracted using principal component analysis (PCA). A new method based on a fuzzy $\min-\max$ neural network with hyper box variable expansion coefficient (FMNN-HVEC) was used to determine the molecular subtypes, and the feature vectors were clustered using deep clustering. Also, a new decision fusion algorithm called weighted Yager was proposed based on the F1-Score for each class. This algorithm proposed a mathematical decision fusion technique to determine the Breast Imaging-Reporting and Data System (BI-RADS) and molecular subtypes values with the accuracy of 95.12% and 89.56%, respectively.

Keywords. Decision support system, Text mining, Breast cancer, BI-RADS, Decision fusion.

MSC. 90C34; 90C40.

* Corresponding author
mostafaboroumandzadeh@gmail.com, parvinnia@iaushiraz.ac.ir, boostani@shirazu.ac.ir, sefidbakht@sums.ac.ir
http://mathco.journals.pnu.ac.ir

1 Introduction

Breast cancer is the most common malignancy and the leading cause of cancer death in women [1]. Therefore, early detection, as well as the development of new treatment strategies, is essential to improve the prognosis of breast cancer patients [1, 2]. Breast cancer is a heterogeneous disease that differs in reasons, prognosis, molecular subtypes, and response to treatment [3, 4]. Cancers are commonly described as diseases of the genes. Therefore, an essential method for diagnosis and disease classification is gene expression, often known as molecular subtypes. Four molecular subtypes have been introduced here, including luminal A, luminal B, HER2, and BLBC [5, 6]. Each of the molecular subtypes indicates a distinct recurrence and survival of the disease, which are essential factors in choosing a different treatment. In addition to gene-based diagnosis, another vital point in the follow-up and treatment is the standardization of medical reports for the rapid diagnosis of this disease and its level. In this regard, the BI-RADS [7, 8] was introduced by the American College of Radiology (ACR) to standardize mammography reports. Seven classes of this categorization are from zero to VI and have meanings incomplete, negative, benign findings, probably benign, suspicious abnormality, highly suggestive of malignancy and known biopsy with proven malignancy, respectively. Therefore, using information such as electronic health records, derived from HIS and PACS systems to determine BI-RADS and molecular subtypes can play an important role in prioritizing and advancing treatment planning and reducing physician diagnosis differences. Most of the methods used in this regard are based on text mining [7, 9] and feature classification [10], but each of them has pros and cons [7, 11]. The main contributions of the proposed framework for breast cancer diagnosis (DSS-BC) were summarized with the following process.

a. Mammography reports related to BI-RADS of 5076 patients were collected.

b. A new method was proposed for medical text mining using word2wec [12]. The extracted vectors and HIS features formed a single-features vector and, principal component analysis (PCA)[13] was used to determine the valuable features.

c. The BI-RADS values were predicted using Convolutional Neural Network (CNN) and Naive Bayesian (NB) classifiers.

d. The vectors were clustered in four sections of molecular subtypes using deep clustering. Compared to the clusters whit these classes, the cluster centers' values were compared using the Euclidean distance, and each molecular subtype was assigned to one collection.

e. To improve the FMNN classification algorithm's performance, a new method was proposed, in which the variable expansion coefficient was used to classify molecular subtypes. In this approach, the number of hyper boxes is optimized, which reduces the complexity (increasing the number of hyper boxes) and the model's simplicity (reducing the number of hyper boxes cause to reduce accuracy).

f. A new Yager method was improved based on each evidence (classifier) weighting for the decision fusion scheme to determine the BI-RADS values.

2 Literature Review

Many studies prove that, time reduction of the breast cancer diagnosis can have effective results in effective treatment. Therefore, automation and improving accuracy of optimal drug scheduling medicine and medical processing at each stage of the treatment process can be effective results [14]-[20]. Mammographic findings are not in the form of structured data, and most cases are reported only in the form of free text mammography reports [5, 21]. Hence, various research projects for medical text processing have been presented. Esmaeili et al. [22] proposed a clinical decision support system (CDSS) using data mining techniques to help physicians interpret mammography reports. The highest accuracy was related to k-nearest neighbors (K-NN) and was equal to 84.06%. Boumaraf et al. [23] proposed a BI-RADS classification system by selecting a modified feature using a Genetic Algorithm (GA). The results had a classification accuracy of 84.5%. Borkowski et al. [24] proposed a method for the automatic classification of background parenchyma enhancement (BPE) based on deep CNN and magnetic resonance imaging (MRI) of the breast. The most relared work [11], [25]-[31]focuses solely on detecting BI-RADS to diagnose breast cancer and uses medical images and texts. However, in this paper, a new method for medical text classification was proposed that improved the output vectors. Also, in the classification process, a new method was used to enhance the classification operation for BI-RADS detection. Molecular subtypes were also identified using previous vectors and classified with improved FMNN. This approach enhances the level of accuracy in diagnosing the type of disease and improves the treatment procedure.

3 DSS-BC Framework

This section first describes the proposed method for processing medical reports based on selecting weighted keywords in context and then describes other parts of the framework.

3.1 Weighted Keywords for Text Mining

In the pre-processing step, prepositions and punctuation were removed. To improve the word embedding accuracy, Bigrams with fewer than 40 occurrences were removed, and Bigrams with more than 1000 cases were considered a single word. Then, the dictionary's keywords were selected from the text, and if they were in a negative sentence, the opposite word was selected. On the other hand, if the opposite word did not exist, its inverse vector was calculated. For this reason, the extracted vector from the word2vec step was normalized $[0-1]$, and then its 1's complement was calculated for all its elements. After combining keywords and terms derived from CLEVER, a total of 350 keywords were obtained. To train Word2vec, Skip-gram was used with variable vector length and window width equal to 12, which were obtained based on trial and error. In each report, we used selected keywords and two words before and two words after the keyword to describe the text. Then the weighted arithmetic mean of all the obtained

vectors represents the vector of that report. Each word in the dictionary has a weight value of 2, and other words have 1. Also, to show the corresponding vector of a report, each sentence's vector was calculated, and each sentence was weighted according to the number of keywords. For example, in the sentence: "Both breasts show scattered areas of fibroglandular density" (a sample sentence from a mammography report), for "breast" and "Fibroglandular" keywords, the two words after and before each keyword are shown in Table 1. The vectors corresponding to the other sentences in the report are calculated in the same way. Finally, the weighted mean of the sentences is calculated based on (1) and (2).

Table 1: Calculation of the weighted average of words in Word2vec

Both	$W2Vec_{Both}$	
breast	$W2Vec_{breast}$	W2Vec1=[$W2Vec_{Both}$+2*$W2Vec_{breast}$+$W2Vec_{show}$+$W2Vec_{scattered}$]/5
show	$W2Vec_{show}$	
scattered	$W2Vec_{scattered}$	
areas	$W2Vec_{area}$	
of	Omitted	W2Vec2=[$W2Vec_{area}$+2*$W2Vec_{Fibroglandular}$+$W2Vec_{density}$]/4
Fibroglandular	$W2Vec_{Fibroglan}$	
density	$W2Vec_{density}$	
$W2Vec_{Sentence}$ = [W2Vec1 + W2Vec2]/2		

$$W = [w_1, w_2, \ldots, w_{ns}], \tag{1}$$

$$V_{MT}^{d} = \frac{1}{\sum_{i=1}^{ns} w_i} \sum_{i=1}^{ns} w_i \times W2Vec_{Sentence_i}^{d}. \tag{2}$$

In this equation, the vector V_{MT}^{d} of dimension d, representing the report, ns is the number of sentences in the report, w_i is the number of keywords in the i^{th} sentence, and $W2Vec_{Sentence_i}^{d}$ is the vector of the i^{th} sentence with d dimensions.

3.2 Effective Feature Selection

Principal component analysis (PCA) [13] was used to determine the exact value of d in the Word2vec algorithm and to determine the essential features of HIS. The PCA is the best method for reducing data dimensions. This technique is not limited to dimension reduction. It is also used in other areas such as pattern recognition and facial recognition. All eigenvalues [32] of the covariance matrix are greater than or equal to 0. Here, if we want to reduce the data dimensions, we can remove the nonsignificant components. Of course, this operation is associated with the loss of data [13].

3.3 Deep Clustering to Assign the Molecular Subtypes

Deep clustering has very good results in this field due to its ability to learn high-dimensional data relationships. In this work, only the people who had breast cancer were included for clustering.

Molecular subtypes were divided into 4 clusters based on existing immunohistochemical markers (IHC). The immunophenotypes besides the number of patients in each cluster are shown in Table 2. Luminal A grows slowly and has the best prognosis. Luminal B grows faster than Luminal A and has a lower prognosis. BLBC is more common in women with the BRCA1 gene mutation. This type of cancer is also more common in young and black women. HER2 grows faster than duct cancers and can have a worse prognosis. Each cluster center that was closer to the available data's immunohistochemical results (46 patients) based on the Euclidean distance was assigned to one of the four molecular subtypes. The method presented in [33], a deep learning method with multi-objective clustering, was used here. Deep clustering yields excellent results in this field due to its ability to learn high-dimensional data relationships. The cost function consists of two sections: (1) Locality preserving objective and (2) Clustering-oriented objective. The patient characteristics are shown by $X = \{x_i \in \mathbb{R}^d\}_{i=1}^n$.

Table 2: Molecular subtypes with immunophenotype [34]

Molecular subtypes	Immunophenotype	Number of patients
Luminal A	ER+ and/ or PR+, HER2 -, CK5/ 6±, and Ki67<14%	17
Luminal B	ER+ and/ or PR+, CK5/ 6±, HER2+, or Ki67≥14%; or PR < 20%	12
HER2	ER-, PR-, HER2+, CK5/6±	9
BLBC	ER-, PR-, HER2- (triple-negative), CK5/6+, and/or EGFR+	8

Here, n is the number of patients and d is the number of obtained features. The clustering problem is formulated in (3). Here, $m \ll d$ and y_i is a meaningful representation of x_i. The output is $\{I_i\}_{i=1}^n$ and $I_i \in \{1, 2, \ldots, C\}$, is an index that identifies which data belongs to a cluster. In this paper, deep manifold clustering (DMC) is used. The DMC is a deep network with an input layer and L hidden layers. An autoencoder consisting of two parts, encoder, and decoder, was created. There are $2L+1$ layers in the deep encoder. The local objective is defined based on observations in (4), and (5) calculates the local density. Here, $\alpha \in [0,1]$ is the balancer of x_i and its local neighbors. $\Psi(x_i, x_j) = \frac{1}{2}\left\|x_i - x_j\right\|^2$ represents the error function. k_i is the set of indicators k-NN of x_i. Here, two criteria were calculated to estimate the density of each cluster. (1) The local density ρ_i and (2) its distance (ξ_i) from the cluster with the highest density. $\Omega = \{i\}_{i=1}^n$, and Δ_{ij} denote the distances between $y_i^{(L)}$ and $y_j^{(L)}$ of frequent updates $\{y_i^{(L)}\}_{i=1}^n$[33] and, $\hat{\Delta}$ is the cut off distance. λ is extracted from DMC (output of hidden layer). (6) is used to calculate the ξ_i.

$$F : x \in \mathbb{R}^d \rightarrow y_i \in \mathbb{R}^m, \tag{3}$$

$$\min_{\theta \backsim \theta'} (1-\alpha)\Psi(\bar{x_i}, x_i) + \frac{\alpha}{k} \sum_{j \in k_i} \Psi(\bar{x_i}, x_i), \tag{4}$$

$$\rho_i = \sum_{j \in \Omega \setminus \{i\}} \exp^{-} \frac{\Delta_{ij}}{\hat{\Delta}}, \tag{5}$$

$$\xi_{\lambda_i} = \begin{cases} \min\{\Delta_{\lambda_i\lambda_j}\}, & i \geq 2, j < i, \\ \max\{\xi_{\lambda_i}\}, & i = 1, j \geq 2. \end{cases} \tag{6}$$

To determine the cluster-based objective, an equation is defined so that each sample is found based on the distance from the center of the cluster. Here, the kernel function is Gaussian. Also, a linear combination of objectives (7) is used to calculate the final cost function [33]. In this regard, J_1 and J_2 are defined based on the reconstruction layer and hidden layer, respectively. β is also a parameter to balance the contribution of J_1 and J_2.

$$\min_{\theta \backsim \theta'} J = J_1 + J_2 = \sum_{i=1}^{n} \left[(1-\alpha)\Psi(\bar{x_i}, x_i) + \frac{\alpha}{k} \sum_{j \in k_i} \Psi(\bar{x_i}, x_i) \right] + \beta \min_{\theta} \sum_{c=1}^{C} S_{ic} \Psi(y_i^{(L)}, y_{\pi_c}^{(L)}). \tag{7}$$

3.4 BI-RADS Classification

3.4.1 Convolutional Neural Network (CNN)

Machine learning algorithms work at reasonable computational times for specific problems and play an important role in extracting knowledge from data [35]. Convolutional neural networks (CNNs) [36] are a series of deep neural networks commonly used for image, speech, or text analysis in machine learning. In this work, we use CNN as a classifier to detect BI-RADS. The CNN can detect complex relationships between dependent and independent data variables, as well as handle noisy data. The convolution operation is applied to the input. After the convolution layer, in pooling layers, sampling is applied to reduce dimensions and prevent over-fitting [37]. In the back-propagation error, the θ parameter was updated using error minimization. The activator function in the first and second convolution layers is ReLU type. Also, the output layer function is softmax, and the loss function is the square of the mean squares error. It should be noted that the Adam optimization algorithm [38] is used here. This optimizer is an adaptive learning rate optimization algorithm.

3.4.2 Naive Bayesian (NB)

In this work, we also use Naive Bayesian [39] as a classifier to detect BI-RADS. Since this approach is a probabilistic method of classification, the final output is also probabilistic. The NB classifier is scalable with high-dimensional data and is suitable for multi-class classification. The NB classification model is shown in (8).

$$P(C = c|X = x) \propto P(C = c) \prod_{i=1}^{d} P(X_i = x_i|C = c) \rightarrow Model^{Bayesian} \tag{8}$$

Here, X is the features vector, and C is the label of the class. The Naive Bayesian tries to estimate different values of $P(C = c)$ and $P(X_i = x_i|C = c)$ using their repetition rate in educational data [39].

3.5 Decision Fusion to BI-RADS Determination

Here, we propose a framework for predicting class using the combining weight assigned classifier method based on the random mapping. Obtained models from $Model^{CNN}$ and $Model^{Bayesian}$, are used to decision fusion in determining BI-RADS. Yager has introduced and formulated the most efficient method. In this theory, the conflict between the evidence (classifiers) is considered. To solve this problem, Yager, in the first step in contrast with the mass function, defined a new function called the Ground Probability Mass Assignment (q). In this method, according to (9), the value of the ground probability mass assignment can be greater than 0, i.e., the second condition of the mass function has been violated [40].

$$q(\phi) \geq 0. \tag{9}$$

Yager considers only one weight per evidence, which further reduces the accuracy of data sets that do not have a constant ratio of data in each class. Therefore, in this work, a method for assigning weight to the evidence (classifiers) is presented which fits every class offered. So if $\overrightarrow{O}_i(J)$ is the estimation of the i^th classifier for state j (class j) and also $\overrightarrow{A}_i(J)$ is the weight of the i^th classifier of the state j, therefore the values of the mass function are defined as (10). Also, the $\overrightarrow{A}_i(J)$ value is obtained from (11), and based on the trained model.

$$m_i(J) = \overrightarrow{A}_i(J) \circ \overrightarrow{O}_i(J), \tag{10}$$

$$\overrightarrow{A}_i(J) = \frac{2.PPV_i^j.Sensitivity_i^j}{PPV_i^j + Sensitivity_i^j} = F1_i^j\_Score. \tag{11}$$

Based on (12), we classify the objection between the evidence in a set $\overrightarrow{\Omega}_i(J)$ when $\overrightarrow{\Omega}_i(J) = \{\omega_i^1, \ldots, \omega_i^j\}$. The evidence decision fusion formulas is calculated (13), (14), (15) in the proposed method.

$$\overrightarrow{\Omega}_i(J) = 1 - \overrightarrow{A}_i(J), \tag{12}$$

$$q(J^j) = \sum_{\frown J_i^j = J^j} \left[m_1^j(J_1^j) \times m_2^j(J_2^j) \times \ldots \times m_i^j(J_i^j) \right], \tag{13}$$

$$q(\phi)^j = \sum_{\frown m_i^j = \phi} m_i^j, \tag{14}$$

$$m_i(J^j) = \frac{q(J^j)}{1 - q(\phi)^j}. \tag{15}$$

3.6 K-nearest Hyper Box Expansion to Detect Molecular Subtypes

FMNN is a fuzzy neural network with a supervised learning method, one of the best algorithms for classifying patterns using its dynamic structure and one pass-through technique [41]. Here, we used the FMNN to predict molecular subtypes. This network uses a set of hyper boxes to specify a class. Each hyper box has a label representing the data class inside the hyper box

[**?**]. FMNN uses the membership function during the learning process according to (16) to determine the membership value (between 0 and 1) of hyper boxes similar to the input sample [44]. Here, $H_j(X_h^d)$ represents the j^{th} fuzzy hyper box with a minimum of v_{ji}^d, a maximum of w_{ji}^d, and d dimensions. Also η is a sensitivity coefficient parameter between 0 and 1, to adjust the membership speed, which is set by the user. Based on the winning hyper box, the FMNN evaluates the dimensions of the hyper box using the inequality (17) to test the hyper box's scalability. Investigating the winning hyper box's expansion criterion shows that many smaller hyper boxes are constructed adjacent to the winning hyper box, which increases the network's complexity. Therefore, K-nearest neighbor hyper box expansion with variable expansion coefficient (FMNN-HVEC), similar to the base FMNN, uses the winning hyper box to predict the target class (y) for the current input sample. Unlike FMNN, which focuses on just one winner, our method uses a set of hyper boxes closer and has the same class label to determine the final winner. According to Algorithm No.1, the hyper box with the highest membership degree is selected. All its dimensions are compared against the expansion coefficient using the inequality (17).

$$H_j(X_h^d) = \frac{1}{2d}\sum_{i=1} d\Big[\max(0,1-\max(0,\eta\min(1,x_{hi}^d - w_{ji}^d))) + \max(0,1-\max(0,\eta\min(1,v_{ji}^d - x_{hi}^d)))\Big], \tag{16}$$

$$d\theta \geq \sum_{i=1} d(\max(w_{ji}^d, x_{hi}^d) - \min(v_{ji}^d, x_{hi}^d)). \tag{17}$$

Each violation of inequality by the winning hyper box leads to choosing the next nearest hyper box to be checked at the same stage. In this case, the first K-nearest neighbor hyper boxes that are capable of resolving the inequality are selected for the expansion process, and the value of θ is increased by as much as γ using (18).

$$\theta = \theta + \gamma, \qquad \gamma << \theta. \tag{18}$$

If all the K-nearest hyper boxes cannot accept the inequality, a new hyper box is created for the input sample. The logical reason is that before creating a new hyper box, a thorough search of the hyper boxes is done for a winner. In other words, the K-nearest neighbor hyper boxes expansion can prevent the creation of many small hyper boxes.

Algorithm No. 1: Fuzzy min – max **neural network with K-nearest hyper box variable expansion coefficient algorithm**

Inputs: The number of instances (n), Features (x_{hi}^d), Expansion coefficient (θ), The number of nearest hyper box expansion (K), γ.

Output: Trained model for molecular subtypes prediction ($Model^{MS}$).

Begin

Processing

For i in range (n)

If hyper box does not exist for corresponding class

v_{ji}^d and w_{ji}^d equal i^{th} instance in ($Model^{MS}$)

Else

If (17) is true
Calculate $H_j(X_h^d)$ using (16)
Update v_{ji}^d and w_{ji}^d according to $\max H_j(X_h^d)$ in $(Model^{MS})$
Update $(Model^{MS})$ according to constraint in [43]
Else
For k in $range(K)$
If (17) for $H_j^k(X_h^d)$ is true
Update v_{ji}^d and w_{ji}^d according to $\max H_j(X_h^d)$ in $(Model^{MS})$
$\theta = \theta + \gamma$
Update $(Model^{MS})$ according to constraint in [43]
Elsenew hyper box and **Update** $(Model^{MS})$
Return $(Model^{MS})$
End

3.7 Proposed Framework of the DSS-BC Hybrid Model

In this paper, the framework of a decision support system for breast cancer diagnosis, including the data preprocessing stage, NLP, clustering, classification, and decision fusion, is presented to optimize BI-RADS_classification using improved Yager. The whole process of the DSS-BC hybrid model is shown in Figure 1. Initially, the data set, which includes each individual's medical records, was processed using algorithm one and converted into a vector. In addition to the features extracted from mammography reports, the features obtained from patients' EHR were also used, and essential features were selected using PCA. Some of the samples in our dataset didn't have labels related to molecular subtypes. Therefore, data were clustered using the unsupervised deep clustering method in four clusters, and the values of molecular subtypes were assigned to each cluster according to the Euclidean distance. Deep clustering is suitable for non-uniform data; thus, it can increase molecular clustering accuracy. A trained model was extracted using CNN and Naïve Bayesian to predict BI-RADS values and their fusion results based on the improved Yager. Then, using FMNN-HVEC, the molecular subtypes were classified.

4 Analyzing and Evaluating the Results

4.1 Data Set

Electronic patient health record (EHR) has been collected from HIS and PACS systems in Namazi Hospital of Fars province from 2015 to 2017. The PACS includes electronic records for storing and retrieving medical images and related documents and reports. HIS is an integrated information system designed to cover all aspects of a hospital's operations, such as financial, administrative, medical, EHR, and legal services. According to Table 3, the characteristics

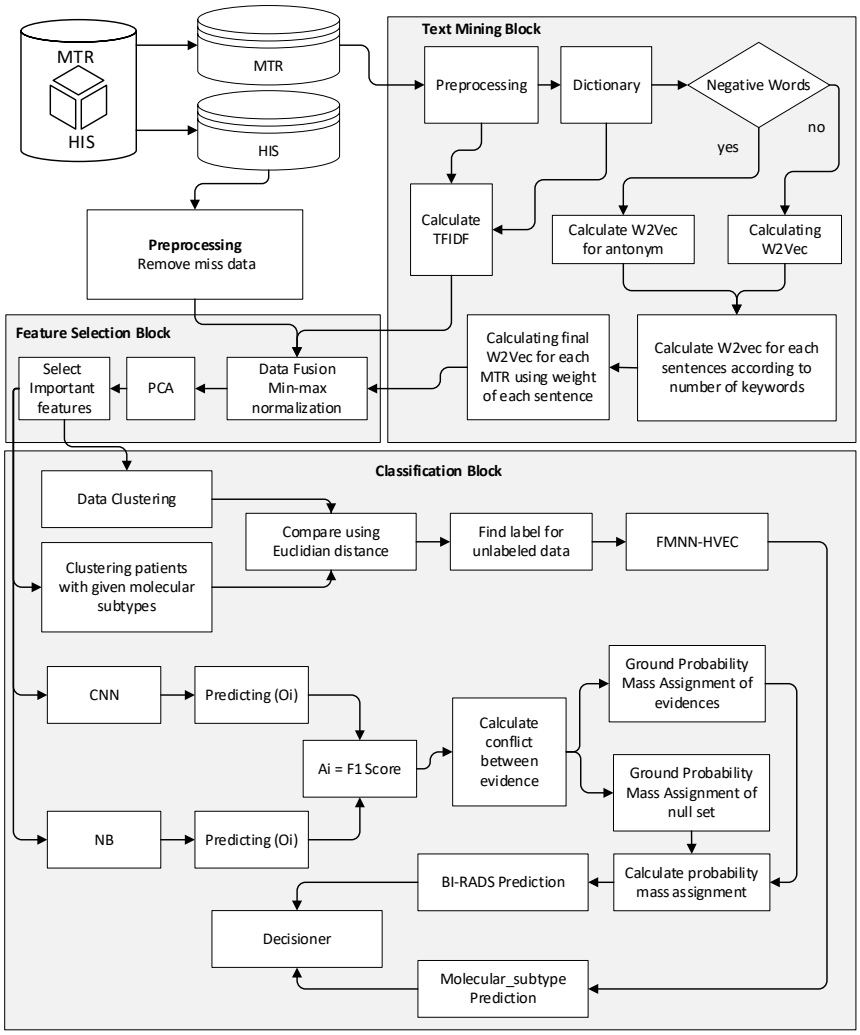

Figure 1: The DSS-BC framework.

extracted from HIS in our data set include age, Inheritance, Regular menstruation, Menopause, Pregnancy history, etc. Also, some of the features and elements extracted from mammography and lexicon text reports are Density (fat, low, equal, or high), Asymmetry (asymmetry, global, focal, or developing), Associated features (skin retraction, nipple retraction, trabecular thickening, parenchyma with no visible mass, etc.), Distribution (diffuse, regional, grouped, etc.), Typically benign (bilateral, right, or bilateral), Suspicious (for instance, coarse heterogeneous, fine pleomorphic, amorphous, fine linear, etc.), size (for instance, 15, 19, and so on), Breast-quad (N or Y), Margin (circumscribed, obscured, micro lobulated, indistinct or, speculated), Shape (oval, round or, irregular), Composition (A: entirely fatty, B: scattered areas of fibro-glandular density, etc.) and so forth. The number of samples in the data set was 5076. In this data set, 25.02% (1270 people), 8.75% (444 people), 30.00% (1523 people), 7.49% (380 people), 12.51% (635 people), 7.49% (380 people) and 8.75% (444 people) were in BI-RADS-0 to BI-RADS-6

levels, respectively. The greatest number was related to BI-RADS-2 and the least number was related to BI-RADS-3. All tests were performed on a desktop computer with an Intel Skylake Core i7-6700 K processor, $4*$ 8 GB DDR RAM, GTX 1080 VGA, and 256GB SSD + 1TB SATA. Python 3.8.7 in the Spyder environment was used for implementation.

Table 3: Extracted features from HIS

Feature	Range	Feature	Range
Age	Discrete numbers	Breast side	Left, Right, or Bilateral
Inheritance	0-1	Abnormal nipple discharge	0-1
Regular menstruation	0-1	Breast pain	0-1
Menopause	0-1	Disease history	0-1
Pregnancy history	0-1	Associated features	0-1
Breast size	Discrete numbers	Alcohol consumption	0-1
Mass shape	Oval, round or irregular	Birth control pills	0-1
Marital status	0-1	Breastfeeding History	Discrete numbers
Sport activities	0-1	Smoking	0-1

4.2 Normalization of Standard Deviation on All Data

Here, normalization by the $\mathbf{min-max}$ method [45] was used. If x is a property, then $X^{'}$the expected value of the property is calculated as (19). $x_{\min}$ represents the minimum value of the features and, $x_{\max}$ represents the maximum value of the features. $x_{\max}^{new}$ and $x_{\min}^{new}$ values are the maximum and minimum of the preset boundaries. When $\mathbf{min-max}$ normalization is applied to the raw data, each attribute is placed in a new range of values. This normalization method has the advantage that all relationships in the data are fully preserved.

$$X^{'} = (x_{\max}^{new} - x_{\min}^{new}) \frac{x - x_{\min}}{x_{\max} - x_{\min}} x_{\min}^{new}. \tag{19}$$

4.3 Evaluation Metrics

First, the data was divided into ten subsets, and then for each subset of data, the system was trained according to the proposed framework. The average of the evaluation metrics is reported $(10 - foldcross - validation)$. The main diameter shows the number of correct diagnoses and the other arrays represent incorrect diagnoses. Four metrics are essential in confusion matrix and describe the following [46]: **True Positive (TP)**: Positive class predicted as a Positive label. **False Positive (FP)**: Negative class predicted as a Positive label. **False Negative (FN)**: Positive class predicted as a Negative label. **True Negative (TN)**: Negative class predicted as a Negative label. Since we have seven classes for BI-RADS and four classes for molecular subtypes, each class' evaluation metrics were obtained separately. Using this matrix, parameters such as Specificity, Sensitivity, PPV, NPV, F1, and Accuracy are calculated [34, 47].

4.4 Experimental Results and Discussion

Figure 2, depicts the accuracy of CNN for detecting BI-RADS using text mining and HIS. Since the number of obtained features from text mining is unknown and some extracted features from HIS have no significant effect on classification, we used PCA to determine the number of appropriate and significant features. The number of features with dimensions $[100, 150, 200, 250, 300]$ and HIS were selected and then classified using CNN. With the increasing dimension, the accuracy of classification increases, and this value decreases in dimensions higher than 200. In many studies [48], with increasing dimensions, word2vec quality, and consequently, accuracy decreased, which was also investigated with decreasing and increasing of dimensions. Finally, since the maximum accuracy was obtained in the dimension of 200 and equal to 90.16%, this value was used as the base dimension for other processes. This value is 79.62% for the NB classifier in the dimension of 200. Table 4, shows the results of the confusion matrix for DSS-BC and the BI-RADS classification and molecular subtypes. This table shows BI-RADS results for seven classes in Table 4a, and the results of molecular subtypes for four classes in Table 4b. Table 5 shows the specificity, sensitivity, PPV, NPV, F1, and accuracy for BI-RADS classification, respectively. The corresponding patients' levels with BI-RAD-2 and BI-RADS-6 have the highest specificity with 95.27% and 94.59%, respectively, and the lowest diagnosis is related to BI-RADS-0 (90.16%). The specificity rate for diagnosing healthy people (BI-RAD-1) is 94.58%. This indicates that DSS-BC also detects healthy individuals with high specificity.

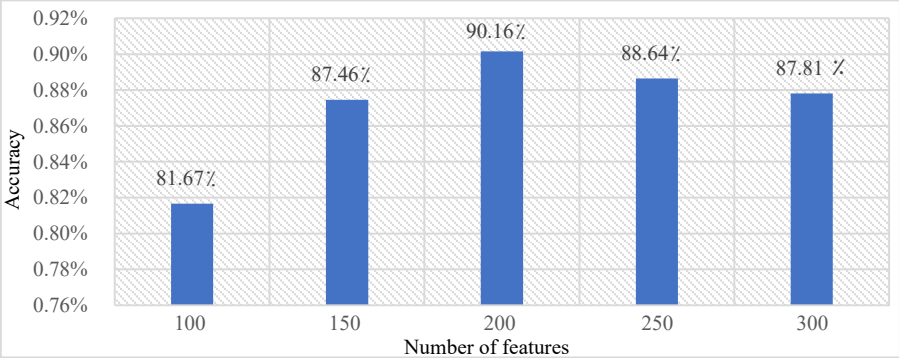

Figure 2: The variation of accuracy with the number of features.

Table 4: The confusion matrix values

(a)

Class	BI-RADS						
BI-RADS-0	1145	2	3	3	11	3	4
BI-RADS-1	16	420	8	8	8	5	3
BI-RADS-2	23	5	1451	6	6	3	5
BI-RADS-3	20	8	14	355	8	4	6
BI-RADS-4	19	2	19	3	579	5	3
BI-RADS-5	22	5	13	1	11	357	3
BI-RADS-6	25	2	15	4	12	3	420

(b)

Class	molecular subtypes			
Luminal A	309	10	17	6
Luminal B	7	321	4	5
HER2	5	4	636	14
BLBC	12	7	10	472

Table 5: The evaluation metrics (BI-RADS)

Class	Specificity	Sensitivity	PPV	NPV	F1	Accuracy
BI-RADS-0	90.16%	98.28%	97.78%	96.63%	98.03%	
BI-RADS-1	94.58%	97.90%	89.74%	98.10%	93.64%	
BI-RADS-2	95.27%	97.56%	96.80%	95.85%	97.18%	
BI-RADS-3	93.42%	97.65%	86.95%	97.10%	91.99%	95.12%
BI-RADS-4	91.18%	97.79%	91.90%	97.67%	94.75%	
BI-RADS-5	93.95%	97.76%	86.65%	97.75%	91.87%	
BI-RADS-6	94.59%	97.60%	91.62%	98.45%	97.52%	
Average	93.30%	97.79%	91.63%	97.37%	94.57%	

The sensitivity value for healthy people in Table 5 is 97.90%, which indicates the high performance of healthy individuals' diagnosis. In general, for each BI-RADS class, the sensitivity value is more than 97%, and its maximum is related to the BI-RADS-0 class (98.28%). The PPV value in Table 5 for healthy people is 89.74%, which indicates the decent performance of the proposed DSS-BC method in diagnosing healthy people and other BI-RADS classes. The lowest value corresponds to BI-RADS-5 (86.65%), and the highest value corresponds to class BI-RADS-0 (97.78%). The NPV value in Table 5 for healthy people is 98.10%, which indicates the excellent performance of the proposed method. It is suitable for other classes, with a minimum of 95.85% for BI-RADS-2 and a maximum of 98.45% for BI-RADS-6. The harmonic mean (F1) for the BI-RADS results is given in Table 5. The maximum F1 for BI-RADS diagnosis is 98.03% (BI-RADS-0), and the minimum is 91.87% (BI-RADS-5), which indicates the appropriate diagnosis of the proposed method. Also, in the correct differentiation of patient and healthy cases, the accuracy or testability in the BI-RADS classification is equal to 95.12%. Table 6, shows the specificity, sensitivity, PPV, NPV, and F1 for various molecular subtypes, respectively.

Table 6: The evaluation metrics (Molecular subtypes)

Algorithm	Class	Specificity	Sensitivity	PPV	NPV	F1	Acc	F1(46_Pa)	Acc(46_Pa)
DSS-BC	Luminal A	89.91%	91.46%	88.51%	90.17%	89.96%	89.56%	68.68%	65.82%
	Luminal B	82.55%	94.14%	84.73%	90.83%	89.19%		64.35%	
	HER2	75.16%	93.64%	74.91%	91.35%	83.24%		59.02%	
	BLBC	75.27%	94.26%	75.26%	90.17%	83.70%		68.50%	
White box[49]	Luminal A	82.16%	85.61%	86.53%	77.81%	86.07%	84.26%	50.36%	55.94%
	Luminal B	79.64%	83.46%	77.12%	81.31%	80.17%		51.98%	
	HER2	79.64%	83.46%	77.12%	81.31%	80.17%		47.69%	
	BLBC	70.81%	85.36%	77.48%	83.12%	77.48%		56.36%	

According to Table 6, the highest specificity for molecular subtypes in DSS-BC is related to Luminal A (89.91%), and the lowest is related to HER2 (75.16%). Here, White Box [49] has been used to compare molecular subtypes in DSS-BC. The specificity in DSS-BC is higher than the White Box method, so that the maximum specificity in DSS-BC is 89.91%, while the maximum specificity in White Box is 82.16%. The highest sensitivity value in molecular subtypes is related to BLBC (94.26%), and the lowest is related to Luminal A (91.46%). The values show that the performance of DSS-BC was more appropriate than the White box (maximum value: 85.61%) in terms of sensitivity values. The PPV value for detecting molecular subtypes for all classes is between 74.92% (HER2) and 88.51% (Luminal A), which indicates the proper diagnosis of each class. DSS-BC also performed better than the White box method

in terms of the PPV parameter. The minimum NPV value for detecting molecular subtypes in Table 6 is 90.17% (Luminal A), and the maximum is 91.90% (BLBC). The NPV for Luminal A in the White box is 77.81% and for the BLBC is 83.12%, indicating that the DSS-BC performed better than the White box. The maximum F1 for detecting molecular subtypes is 89.96% (Luminal A), and the minimum is 83.24% (HER2), which is better than the White box method. The accuracy value in DSS-BC is 89.56%, while the White box method's accuracy is 84.26%. So DSS-BC performed better than the White box. In the last two columns of Table 6, the evaluation metrics (F1 and Accuracy) of the proposed method (FMNN-HVEC) are presented for patients whose molecular subtypes were known in the database (46 patients). Due to the high importance of the molecular subtype factors to determine the type and level of the disease, as well as its four classes, the results will be significant. Figure 3, compares the BI-RADS results with the SVM [26], PART [26], Blue Based BN [50], and BBAS [50] methods; since the values of each class are reported as averages in this paper, so here to the comparison of DSS-BC, the mean values of seven classes are calculated per evaluation metric.

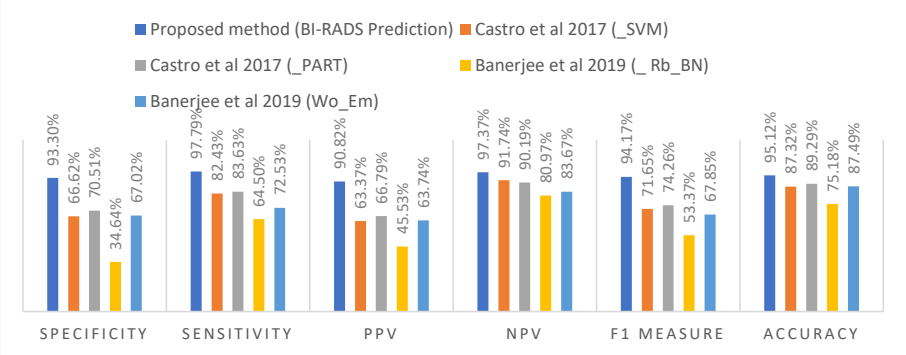

Figure 3: The comparison of DSS-BC with other methods.

In Figure 3, in addition to the graph, numerical results are also reported. In this figure, the parameters of specificity, sensitivity, PPV, NPV, F1, and accuracy are reported in different methods, which shows that DSS-BC performed better than other methods. As a result, by combining the evaluation metrics, it is clear that the proposed method has performed well in identifying BI-RADS classes, molecular subtypes and ultimately helped to patient follow-up. Therefore, since a new way for text processing has been proposed, useful HIS features have been used alongside text processing results. Decision fusion has also been used to increase accuracy. Therefore, the performance of the proposed method has been improved compared to similar tasks in terms of BI-RADS detection and the determination of molecular subtypes.

5 Conclusion

The Breast Imaging-Reporting and Data System (BI-RADS) was introduced by American College of Radiology (ACR) to standardize mammography reports. This standard can be used to

accurately prioritize treatment progress. However, this approach still suffers some disadvantages such as a discrepancy in physicians' opinions on the outcome of BI-RADS value which may affect the determination of more accurate treatments. On this account, the EHR information was suggested to be used for the identification of molecular subtypes besides BI-RADS. In this paper, a hybrid model including unstructured data (medical reports) and structured data (HIS) was developed. After text processing, the medical reports were converted into vectors so that each vector eventually represented a mammography report using Word2vec. To select the essential features, PCA was used and in the following, molecular subtype values were determined based on the Euclidean distance and the results were generalized to cluster members. In this work, deep clustering led to the detection of more accurately labeled molecular subtypes. CNN and NB were used to classify BI-RADS and they were combined through an improved Yager. A new FMNN scheme was used to classify molecular subtypes called FMNN-HVEC. This method reduced network complexity and increased classification accuracy. K-fold cross-validation was used with $K = 10$, and specificity, sensitivity, PPV, NPV, F1 value, and accuracy were evaluated. The maximum results of metrics for estimating BI-RADS were 95.27%, 98.28%, 97.78%, 98.45%, and 98.03%, respectively. These results were 89.91%, 94.26%, 88.51%, 91.90%, and 89.96% for estimating the molecular subtypes in the maximum state. The accuracy of BI-RADS and molecular subtypes diagnostic were 95.12% and, 89.56% respectively. In conclusion, the proposed method for converting medical text records to vectors and using HIS to diagnose BI-RADS and identify molecular subtypes based on medical literature and HIS features will help physicians improve their decisions. Therefore, physicians can evaluate the patient's therapeutic routine more accurately.

Limitations and Future Works

Using streaming data is recommended for medical diagnoses in online environments such as the Internet of Things (IoT) and incremental learning techniques. This work has considered a stationary dataset. In future work, it is suggested to use learning techniques to diagnose in dynamic environments such as IoT, where data increase over time.

Acknowledgment

We are grateful to the Department of Computer Engineering, Islamic Azad University, Branch of *Shiraz*, Iran.

Compliance with Ethical Standards

This paper does not contain any studies with human participants performed by any of the authors.

References

[1] Sefidbakht S., Haseli S., Khalili N., Bazojoo V., Keshavarz P., Zeinali-Rafsanjani B. (2021). "Can shear wave elastography be utilized as an additional tool for the assessment of non-mass breast lesions?", Ultrasound, 1742271X21998721.

[2] Bray F., Ferlay J., Soerjomataram I., Siegel R. L., Torre L. A., Jemal A. (2018). "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries", CA: A Cancer Journal for Clinicians, 68(6), 394-424.

[3] Bi W. L., Hosny A., Schabath M. B., Giger M. L., Birkbak N. J., Mehrtash A., Aerts H. J. (2019). "Artificial intelligence in cancer imaging: clinical challenges and applications", CA: a cancer Journal for Clinicians, 69(2), 127-157.

[4] Sim L. L. W., Ban K. H. K., Tan T. W., Sethi S. K., Loh T. P. (2017). "Development of a clinical decision support system for diabetes care: A pilot study", PloS one, 12(2), e0173021.

[5] Gandomkar Z., Suleiman M. E., Demchig D., Brennan P. C., McEntee M. F. (2019). "BI-RADS density categorization using deep neural networks", In Medical Imaging 2019: Image Perception, Observer Performance, and Technology Assessment (Vol. 10952, p. 109520N), International Society for Optics and Photonics.

[6] O'Connor J. P., Rose C. J., Waterton J. C., Carano R. A., Parker G. J., Jackson A. (2015). "Imaging intratumor heterogeneity: role in therapy response, resistance, and clinical outcome", Clinical Cancer Research, 21(2), 249-257.

[7] Boyer B., Canale S., Arfi-Rouche J., Monzani Q., Khaled W., Balleyguier C. (2013). "Variability and errors when applying the BIRADS mammography classification", European journal of radiology, 82(3), 388-397.

[8] Sefidbakht S., Jalli R., Izadpanah E. (2015). "Adherence of Academic Radiologists in a Non-English Speaking Imaging Center to the BI-RADS Standards of Reporting Breast MRI", Journal of Clinical Imaging Science, 5.

[9] Guo D., Wang Q., Liang M., Liu W., Nie J. (2019). "Molecular cavity topological representation for pattern analysis: a NLP analogy-based Word2Vec method", International Journal of Molecular Sciences, 20(23), 6019.

[10] Dehghan P., Mogharabi M., Zabbah I., Layeghi K., Maroosi A. (2018). "Modeling Breast cancer using data mining methods", (In Persian).

[11] Gao H., Bowles E. J. A., Carrell D., Buist D. S. (2015). "Using natural language processing to extract mammographic findings", Journal of Biomedical Informatics, 54, 77-84.

[12] KENNETH W. C. (2017). "Word2Vec", Natural Language Engineering, 23(1), 155-162.

[13] Kumar I., Virmani J., Bhadauria H. S., Thakur S. (2019). "A breast tissue characterization framework using PCA and weighted score fusion of neural network classifiers", In Classification Techniques for Medical Image Analysis and Computer Aided Diagnosis (pp. 129-151). Academic Press.

[14] Akbarin M. M., Shirdel A., Bari A., Mohaddes S. T., Rafatpanah H., Karimani E. G., Torshizi R. (2017). "Evaluation of the role of TAX, HBZ, and HTLV-1 proviral load on the survival of ATLL patients", Blood Research, 52(2), 106-111.

[15] Boostani R., Karimzadeh F., Nami M. (2017). "A comparative review on sleep stage classification methods in patients and healthy individuals", Computer Methods and Programs in Biomedicine, 140, 77-91.

[16] Boroumandzadeh M., Parvinnia E. (2021). "Automated classification of BI-RADS in textual mammography reports", Turkish Journal of Electrical Engineering & Computer Sciences, 29(2), 632-647.

[17] Kamali T., Boostani R., Parsaei H. (2013). "A multi-classifier approach to MUAP classification for diagnosis of neuromuscular disorders", IEEE Transactions on Neural Systems and Rehabilitation Engineering, 22(1), 191-200.

[18] Nezam T., Boostani R., Abootalebi V., Rastegar K. (2018). "A novel classification strategy to distinguish five levels of pain using the EEG signal features", IEEE Transactions on Affective Computing, 12(1), 131-140.

[19] Torshizi R., Karimani E. G., Etminani K., Akbarin M. M., Jamialahmadi K., Shirdel A., Rafatpanah H. (2017). "Altered expression of cell cycle regulators in adult T-cell leukemia/lymphoma patients", Reports of Biochemistry & Molecular Biology, 6(1), 88.

[20] Zarei H. (2016). "The control parametrization enhancing technique for multi-objective optimal control of HIV dynamic", Control and Optimization in Applied Mathematics, 1(2), 1-21.

[21] Boroumandzadeh M., Parvinnia E. (2021). "Proposing a clinical decision support system for breast cancer diagnosis", Journal of Knowledge & Health in Basic Medical Sciences, 15 (3), 54-66.

[22] Esmaeili M., Ayyoubzadeh S. M., Ahmadinejad N., Ghazisaeedi M., Nahvijou A., Maghooli K. (2020). "A decision support system for mammography reports interpretation", Health Information Science and Systems, 8(1), 1-8.

[23] Boumaraf S., Liu X., Ferkous C., Ma, X. (2020). "A new computer-aided diagnosis system with modified genetic feature selection for bi-RADS classification of breast masses in mammograms", BioMed Research International, 2020.

[24] Borkowski K., Rossi C., Ciritsis A., Marcon M., Hejduk P., Stieb S., Berger N. (2020). "Fully automatic classification of breast MRI background parenchymal enhancement using a transfer learning approach", Medicine, 99(29).

[25] Bozkurt S., Gimenez F., Burnside E. S., Gulkesen K. H., Rubin D. L. (2016). "Using automatically extracted information from mammography reports for decision-support", Journal of Biomedical Informatics, 62, 224-231.

[26] Castro S. M., Tseytlin E., Medvedeva O., Mitchell K., Visweswaran S., Bekhuis T., Jacobson R. S. (2017). "Automated annotation and classification of BI-RADS assessment from radiology reports", Journal of Biomedical Informatics, 69, 177-187.

[27] Destrempes F., Trop I., Allard L., Chayer B., Garcia-Duitama J., El Khoury M., Cloutier G. (2020). "Added value of quantitative ultrasound and machine learning in BI-RADS 4–5 assessment of solid breast lesions", Ultrasound in Medicine & Biology, 46(2), 436-444.

[28] Gupta A., Banerjee I., Rubin D. L. (2018). "Automatic information extraction from unstructured mammography reports using distributed semantics", Journal of Biomedical Informatics, 78, 78-86.

[29] Percha B., Nassif H., Lipson J., Burnside E., Rubin, D. (2012). "Automatic classification of mammography reports by BI-RADS breast tissue composition class", Journal of the American Medical Informatics Association, 19(5), 913-916.

[30] Sippo D. A., Warden G. I., Andriole K. P., Lacson R., Ikuta I., Birdwell R. L., Khorasani R. (2013). "Automated extraction of BI-RADS final assessment categories from radiology reports with natural language processing", Journal of Digital Imaging, 26(5), 989-994.

[31] Zhang X., Zhang Y., Zhang Q., Ren Y., Qiu T., Ma J., Sun Q. (2019). "Extracting comprehensive clinical information for breast cancer using deep learning methods", International Journal of Medical Informatics, 132, 103985.

[32] Chakraborty D., Chiracharit W., Chamnongthai K. (2021). "Video shot boundary detection using principal component analysis (PCA) and deep learning", In 2021 18th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON) (pp. 272-275). IEEE.

[33] Chen D., Lv J., Zhang Y. (2017). "Unsupervised multi-manifold clustering by learning deep representation", In Workshops at the thirty-first AAAI Conference on Artificial Intelligence.

[34] Huber K. E., Carey L. A., Wazer D. E. (2009). "Breast cancer molecular subtypes in patients with locally advanced disease: impact on prognosis, patterns of recurrence, and response to therapy", In Seminars in Radiation Oncology (Vol. 19, No. 4, pp. 204-210). WB Saunders.

[35] Alesheykh R. (2016). "Comparative analysis of machine learning algorithms with optimization purposes", Control and Optimization in Applied Mathematics, 1(2), 63-75.

[36] Kalchbrenner N., Grefenstette E., Blunsom P. (2014). "A convolutional neural network for modelling sentences", ArXiv Preprint arXiv:1404.2188.

[37] Xu Q., Zhang M., Gu Z., Pan G. (2019). "Overfitting remedy by sparsifying regularization on fully-connected layers of CNNs", Neuro Computing, 328, 69-74.

[38] Jais I. K. M., Ismail A. R., Nisa S. Q. (2019). "Adam optimization algorithm for wide and deep neural network", Knowledge Engineering and Data Science, 2(1), 41-46.

[39] Leung K. M. (2007). "Naive bayesian classifier", Polytechnic University Department of Computer Science/Finance and Risk Engineering, 2007, 123-156.

[40] Akram M., Shahzadi G. (2020). "A hybrid decision-making model under q-rung orthopair fuzzy Yager aggregation operators", Granular Computing, 1-15.

[41] Rizzi A., Panella M., Mascioli F. F. (2002). "Adaptive resolution min-max classifiers", IEEE Transactions on Neural Networks, 13(2), 402-414.

[42] Shinde S. V., Kulkarni U. V. (2014). "Mining classification rules from fuzzy min-max neural network", In Fifth International Conference on Computing, Communications and Networking Technologies (ICCCNT) (pp. 1-7). IEEE.

[43] Patrick K. S. (1992). "Fuzzy Min-Max Neural Networks-Part1: Classification", IEEE Transactions on Neural Networks, 2, 776-786.

[44] Mohammed M. F., Lim C. P. (2017). "Improving the Fuzzy Min-Max neural network with a K-nearest hyperbox expansion rule for pattern classification", Applied Soft Computing, 52, 135-145.

[45] Patro S., Sahu K. K. (2015). "Normalization: A preprocessing stage", arXiv preprint arXiv:1503.06462.

[46] Tharwat A. (2018). "Classification assessment methods", Applied Computing and Informatics. Advance online publication.

[47] Duan K. B., Keerthi S. S. (2005). "Which is the best multiclass SVM method? An empirical study", In International workshop on Multiple Classifier Systems (pp. 278-285). Springer, Berlin, Heidelberg.

[48] Li B., Drozd A., Guo Y., Liu T., Matsuoka S., Du X. (2019). "Scaling word2vec on big corpus", Data Science and Engineering, 4(2), 157-175.

[49] Wu M., Zhong X., Peng Q., Xu M., Huang S., Yuan J., Tan T. (2019). "Prediction of molecular subtypes of breast cancer using BI-RADS features based on a "white box" machine learning approach in a multi-modal imaging setting", European Journal of Radiology, 114, 175-184.

[50] Banerjee I., Bozkurt S., Alkim E., Sagreiya H., Kurian A. W., Rubin D. L. (2019). "Automatic inference of BI-RADS final assessment categories from narrative mammography report findings", Journal of Biomedical Informatics, 92, 103137.

How to Cite this Article:
Boroumandzadeh, M., Parvinnia, E., Boostani, R., Sefidbakht, S. A Decision Support System Framework Based on Text Mining and Decision Fusion Techniques to Classify Breast Cancer Patients. Control and Optimization in Applied Mathematics, 2022; 6(1): 11-29. doi: 10.30473/coam.2021.60533.1175