Journal of Engineering Science and Technology
Vol. 17, No. 6 (2022) 3940 - 3955

CORRELATION-BASED FEATURE SELECTION WITH BAG-BASED FUSION SCHEME FOR MULTI-INSTANCE LEARNING APPLICATION

MAZNIHA BERAHIM[1,2,*], NOOR AZAH SAMSUDIN[1], AIDA MUSTAPHA[3]

[1]Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia (Main Campus), 86400 Parit Raja, Batu Pahat, Johor, Malaysia
[2]Center for Diploma Studies, Universiti Tun Hussein Onn Malaysia (Pagoh Campus), KM1, Jalan Panchor, 84600 Muar, Johor, Malaysia
[3]Faculty of Applied Sciences and Technology, Universiti Tun Hussein Onn Malaysia (Pagoh Campus), KM1, Jalan Panchor, 84600 Muar, Johor, Malaysia
*Corresponding Author: mazniha@uthm.edu.my

Abstract

Multi-Instance Learning (MIL) classifies a bag of instances rather than an individual instance. There is a lack of consideration of feature selection in MIL. The large number of features that are irrelevant and redundant in MIL affects the classification performance. Besides that, a genuine label of instance is unknown in MI data and evaluation of relevancy using bag class label cannot be done directly. To address this gap, this paper proposes a Fusion Bag-based Correlation Feature Selection (FBC-FS) technique using multiple bag summarization to accommodate MI data in an effort to increase performance in MI classification. The proposed technique consists of three steps: feature transformation, feature evaluation using the bag correlation and fusion of candidate features. The FBS-FS is evaluated based on the MI dataset (Breast Cancer and Tiger image) with a standard Support Vector Machine, K-Nearest Neighbour (KNN) and Decision Tree. The superior result achieves up to 91.5% AUC when using KNN for the Breast Cancer dataset and the improvement achieves up to 16% with proposed FS compared without performing FS task. The results also proved that correlation measures in evaluating relevance and redundancy criteria with extended parameters to find optimal features contribute highly to the improvement of the classification performance.

Keywords: Bag summary, Correlation measure, Medical image classification, Multi-instance feature selection, Redundant feature, Relevant feature.

1. Introduction

Disease diagnostic accuracy is a major challenge for global healthcare systems [1]. It is estimated that on average the rate of diagnosis error is between 3% to 5% every year worldwide [2]. In medical imaging, a small bias towards the majority class such as a little mark in a retinal image is sufficient to change the image class from normal to abnormal, even though the majority of the images are normal [3]. Therefore, an accurate disease diagnosis is crucial because it is the key for reducing unnecessary surgeries and procedures [4]. Fortunately, Multi-Instance Learning (MIL) is an alternative solution to the problem with medical image diagnostic systems at conventional single image analysis. A bag in MIL is defined as a collection of unlabelled instances.

Region of interest (ROI) in medical images usually distributed over multiple adjacent slices (e.g., in Computed Tomography (CT) or Magnetic Resonance Imaging (MRI)) and having correlations in successive slices [5]. Therefore, features of MI images may provide complementary information, and provide more insights of the disease pathology [6] to achieve class discrimination with higher accuracy [7, 8]. This means MIL is able to offer a more accurate diagnosis [9, 10]. However, the MIL diagnosis is implemented on the full data set (without FS task) in general.

There are limited studies regarding feature selection (FS) that have been designed for MI problems. Conventionally, MI image features are analysed individually and without considering the correlation between different instances [11]. The evaluation and analysis of MI features implemented at instance-based is prone to a few problems. Firstly, it did not consider full information from the structure of bag features. Secondly, it depends on the availability of the target instance class label to identify relevant features which the genuine label of instances is unknown in MI data [12]. Tackling the ambiguity label by assumption may be inappropriate and the instance-based feature evaluation of relevancy using bag class label cannot be done directly.

To address this gap, the objective of this study is to propose a FS technique which evaluates the image features as a whole (at bag-based) instead of analysing as a single instance (instance-based). This means, evaluating the image’s features once from the group of images to obtain optimal features for good performance of classification task. Hence, the FS task is enhanced from instance-based to bag-based process. By implementing the proposed technique, the optimal feature set will be identified, and the accuracy of the medical MI image classification is hoped to be improved. Three contributions of this study include: (1) analysing MI features as multiple bag representation using statistical central tendency measures, (2) the bag feature acquired being new bag-based parameter in measuring correlation of among bag features and correlation between bag feature with a bag class label and (3) a new fusion scheme implement using set theory operations on binary relations in selecting the optimal features with considering two criteria and multiple bag summarization.

The remaining of this paper is organized as follows. Section 2 reviews work on feature selection and feature transformation in MIL. Section 3 presents the research methods including the research framework, the datasets, the proposed technique, and the evaluation metrics. Section 4 presents the results and discussions, and finally Section 5 concludes the paper with some indications for future works.

2. Literature Review

In Multi-Instance Learning (MIL) problems, instances are grouped into bags, and the entire bags are labelled. MIL was first proposed by [13]. MIL is a weak supervised learning framework [14, 15]. Weak supervision in machine learning deals with limited or imprecise class labels such as in the case of MIL, when a class represents a group or bag of instances rather than a single instance. The learning task is to predict the label of bags. Feature Selection (FS) for MIL problems is more challenging than Single-Instance Learning (SIL) due to ambiguous input [16], missing instance labels, and a mix of positive and negative instances in a particular positive or negative bag.

The literature shows that tasks involved in MIL can be performed at two different levels; instance-based and bag-based [17,18]. However, most previous studies implement FS in MIL using instance-based FS technique with different categories: based on interaction with classifier (e.g., MI-AdaBoost [19]), ranking (e.g., MI-FEAR [20], ReliefF-MI [21], Reliability-based FS [22]) and subset selection (e.g., HyDR-MI [23]). However, none consider feature correlation for evaluating redundancy and relevance criteria simultaneously in FS of MIL.

2.1. Feature transformation

A multi-instance (MI) dataset means a class is represented by a bag of instances rather than a single instance. Due to this structure, the FS task cannot be implemented directly to the dataset. The instances in a bag have to be transformed into a bag vector before they are classified using the standard SIL classifier in order to classify the bag labels or classes. To handle MI datasets, feature transformation is required especially when the model accuracy is more important than model interpretation [24]. Recent research has shown that the use of bag summarization by transforming functional vectors in one single case into a new bag vector is necessary as a generic learning strategy to tackle the ambiguity of instance class [8]. The single value is meaningful to represent the whole bag. All instances of the same bag are considered equally important in the bag representation [25]. Table 1 lists the variation of MI bag representation by feature transformation which basically uses a single statistical and distance measure. However, the distance measure may ineffective in the case of high-dimensional data [26].

Table 1. Variation of MI feature transformation.

Approach	Comments
Summarization by arithmetic mean, geometric mean, and minimax [27]	The arithmetic mean outperforms the other two methods in a few cases. However, it has been done for classification task whereas not for FS.
Weighted representation transformation [25]	The weighted computed by the minimal Hausdorff distances between the bags and only considered closed instances and used for classification task.
Arithmetic mean [28]	Used for classification task.
Discriminative Bag Mapping [8]	Only consider a small number of selected instances from MI bags and used for classification task.
Histograms utilized as bag representation [29]	Used for FS task in MIL at instance-level. The distance measure used as the criterion to evaluate the feature importance.

All the implementations used the transformation for classification [29]. Some transformations consider all instances such as [27-29] and others consider selected instances [8, 25]. Summarization by mean is the most frequently used and generally considered the best measure of central tendency [30]. However, some cases in analysing numerical features value of real image may exist extreme scores in the distribution, may some scores have undetermined values or open-ended distribution. These possible cases make the other measures of central tendency preferred [30]. To empower the MI image diagnostic, this study proposes a FS technique that involves a few transformations process and considers all instances to look at the image features as a whole. Instead of using a single measure, the proposed technique considers all measures of central tendency to enrich the bag information.

2.2. Feature evaluation

Features thus play an important role during the classification stage [31]. A redundancy criterion determines the degree of dependency among all features, while a relevance criterion determines the degree of dependency between a feature to the class in order to see how well a feature discriminates between the classes. Formulas for calculating a correlation coefficient are available and the choice of measure depends on the type of data to measure the strength of features association. Redundancy among features can be evaluated by using the Pearson Correlation Coefficient (PCC), which is suited to measure a correlation among continuous data. Meanwhile, relevancy features and target class can be evaluated by using the Point Biserial Correlation (PBCC) which suits continuous-binary categorical data [32].

PCC is commonly used in medical research [33, 34] to evaluate features in a dataset such as in Magnetic Resonance Imaging (MRI) dataset [35]. Eq. (1) shows the PCC formula to calculate between the two features *X* and *Y*, with x_i, y_i, and the respective mean for x and y. Refer to Table 3 for the description of Eq. (1).

$$r_{x,y} = \frac{\sum_{i=1}^{n} X.Y}{\sqrt{\sum_{i=1}^{n} X^2}\sqrt{\sum_{i=1}^{n} Y^2}} \text{ where } X = x_i - \bar{x} \text{ and } Y = (y_i - \bar{y}). \quad (1)$$

PBCC is adequate to analyse whether there is weak or strong correlation among features as well as the correlation value in the form of a redundant score to indicate feature goodness in a classification task [36]. However, this study will use the PBCC to assess relevance between the features and the class label of the images. The PBCC is calculated as Eq. (2) [37]. Refer Table 4 for the description.

$$r_{pb}(f_j, c_i) = \frac{\overline{y_0} - \overline{y_1}}{S_j}\sqrt{\frac{n_0 n_1}{n^2}} \quad (2)$$

These existing correlation measures only use instance-based features parameters in calculating the correlation coefficient which need to extend to use for MI data.

3. Research Method

This paper proposes a Fusion Bag-based Correlation Feature Selection (FBC-FS) technique to empower feature selection (FS) tasks to improve performance of Multi-Instance (MI) classification. Firstly, a brief on research the framework of the proposed technique. Secondly, the datasets used are presented. Thirdly, brief on feature transformation followed by feature evaluation. Finally, the classifier used.

3.1. Research framework

Generally, the important four stages of MI image diagnostic consist of image acquisition and pre-processing of MI images, feature extraction, feature selection, and classification [38-40]. However, the first and second stage will not be repeated because it has been completed by [41], where the derivation of numerical MI dataset was obtained and used as benchmark input (refer subsection 3.3 for the details). The main contribution of this study focuses on the third stage, proposes an extended FS technique to select optimal features named Fusion Bag-based Correlation Feature Selection (FBC-FS) which consists of three main steps: transformation (a), evaluation (b) and fusion (c) (refer Fig. 1).

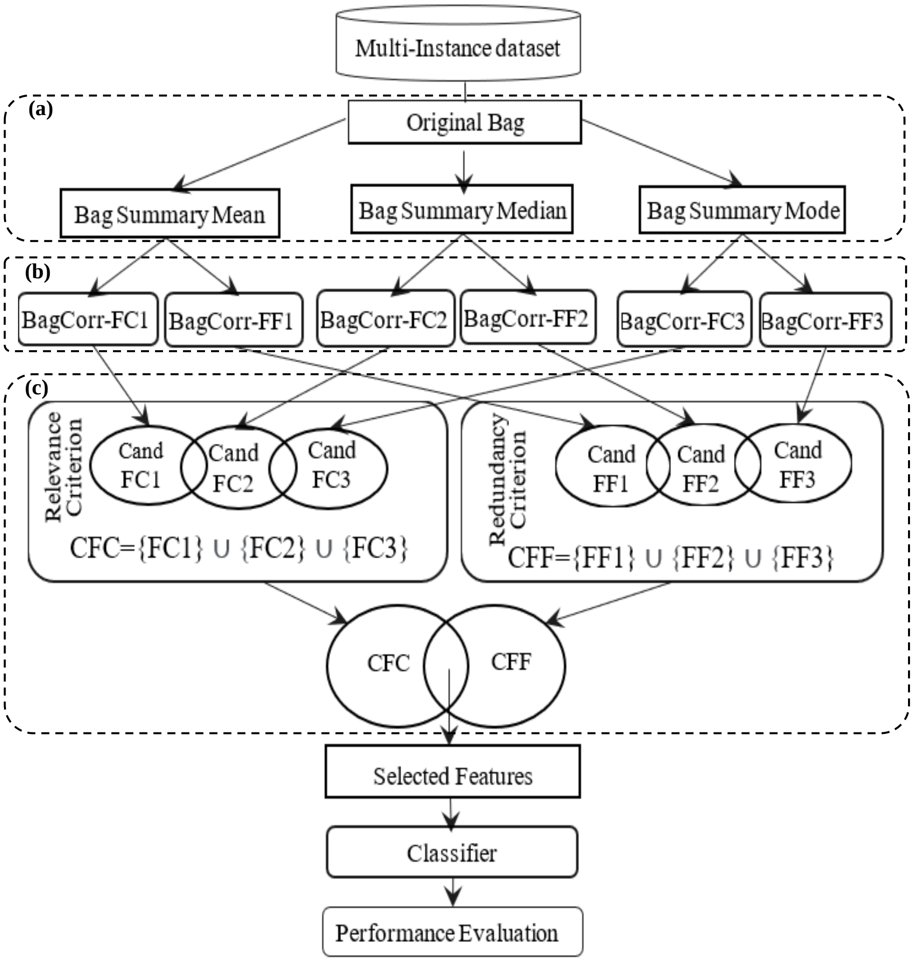

Fig. 1. Framework of the fusion bag-based correlation feature selection.

At the first step, features are transformed from original features to a new bag feature using bag summarization of statistical central measures (refer Fig. 1(a) and Section 3.3). At the second step, the bag features acquired will be evaluated using correlation measures (refer Fig. 1(b) and Section 3.4). The evaluation considered two criteria: relevance (by looking at BagCorr-FC) and redundancy (by looking at BagCorr-FF). The final step is selecting optimal features using set theory operation. The candidate features from multiple bag summarization evaluation were combined to get candidate fused features for each criterion of relevance (Correlation Feature-Class: CFC) and redundancy (Correlation Feature-Feature: CFF) by implementing union operation. Then, the best selected features are fused using intersection operation by looking for features which meet both criteria (refer Fig.1 (c)). Using SIL Classifier, the selected bag features being input for the fourth stage of classification (refer Section 3.5). In the case of disease diagnosis using medical image dataset which refers to predicting the label of image for the subjects. The bag is subsequently classified as positive (disease) or the negative (non-disease). When the classification is completed at the fourth stage, the performance of the diagnosis result will be measured in terms of accuracy and area under the Receiver Operating Characteristic (ROC) curve (AUC) (refer Section 3.6 for details).

3.2. Multi-instance image dataset

The input for this framework is a secondary dataset acquired from a collection of MI dataset defined as vectors on a set of instance features and its bag class labels. The numerical data of MI medical image named UCSB Breast Cancer as the main dataset has been used. A set of extracted features represented as feature vectors of instances. An excerpt of the image is shown in Fig. 2. Each image (called bag) is split into patches (called instances) and the image patches are not labelled [42].

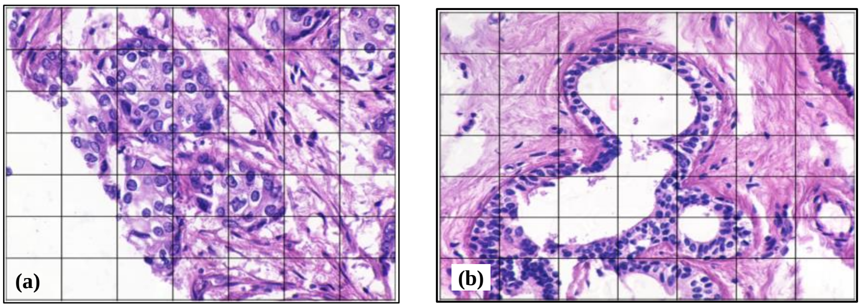

Fig. 2. Sample microscopic images of (a) cancer (or malignant) and (b) non-cancer (or benign) from UCSB Breast Cancer dataset.

The public MI medical image is limited. Therefore, a non-medical image dataset has been used is a Tiger image as comparison. The dataset is a popular benchmark for evaluating new MIL proposals [43] and has the same one characteristic as the UCSB Breast Cancer dataset which is categorized as multimodal positive distribution. The description for all datasets is listed in Table 2. The extracted feature vectors for these dataset are publicly available at [44].

Table 2. Description of benchmark dataset [45].

Dataset	Total Bags	+ Bags	- Bags	Features	Total Instances	Min	Max
UCSB Breast Cancer	58	26	32	708	2002	21	40
Tiger	200	100	100	230	1220	1	13

3.3. Feature transformation into bag summarization of features

To implement the proposed FBC-FS technique, this is a first step as Figure 1(a) represents the bag information as a collection of different numbers of instances into a new feature vector in a single value. The new bag features will contain statistical values computed from the set values of all instances in a bag that corresponds to the original representation. The transformations proposed have been performed on the bags using statistical central tendency measures which are mean, mode and median mapping. The mean measure is a standard measure at the center of the data distribution, the mode measure to the most frequently occurring value in the dataset and the median measure to get the middle value for a dataset. Figure 3 shows the general idea of the bag vector of features.

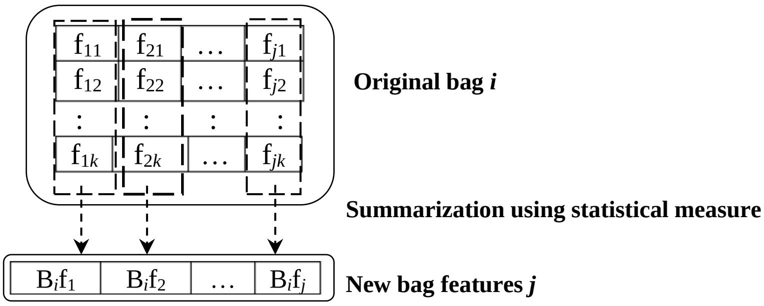

Fig. 3. The new bag feature vector based on statistic measure.

The resulting measures will be used as new bag vectors for the next stage in the proposed technique. Note that feature transformation of bag summarization as a new bag vector is categorized as a bag-based approach for the FS task. This step turns the MIL problem into a mimic SIL problem. However, the single value transformed is not an individual instance but a group of instance information.

3.4. Feature evaluation using bag-based features

The second step in the proposed FBC-FS technique will calculate redundancy and relevance score for each feature using extended parameters of correlation measures.

3.4.1. Redundancy score using modified Pearson correlation

The correlation approach is used to measure the association between the features. For a pair of features (f_j, f_i) or (x, y) the correlation coefficient, r, is given by Eq. (1). Meanwhile, the extended PCC using bag-based parameters with pair-wise bag features is presented in Eq. (3). This correlation is named as BagCorr-FF. Refer Table 3 for the description of Eq. (3).

$$r_{bx,by} = \frac{\sum_{i=1}^{nb}(bX)(bY)}{\sqrt{\sum_{i=1}^{nb}(bX)^2}\sqrt{\sum_{i=1}^{nb}(bY)^2}} \tag{3}$$

where $bX = bX_i - \overline{bX}$ and $bY = \left(by_i - \overline{by}\right)$

The proposed Extended PCC (EPCC) is different from the conventional PCC with its parameter involved in the calculation. EPCC is able to compute redundancy scores which represent the bag feature correlation not instance correlation.

Table 3. Conventional PCC vs. Extended PCC.

Conventional PCC: Eq. (1)	**Extended PCC (Bag-based): Eq. (3)**
$r_{x,y}$= correlation feature *x* and *y*	$r_{bX,bY}$ = correlation bag feature *x* and *y*
x_i= values of *x*-feature	bX_i= values of *x*-bag feature
$\bar{x}$ = mean values of *x*-feature	$\overline{bX}$ = mean values of *x*- bag feature
y_i= values of *y*-feature	bY_i= values of *y*-bag feature
$\bar{y}$ = mean values of *y*-feature	$\overline{bY}$ = average of *y*-bag feature
n = number of features	nb = number of bag features

3.4.2. Relevance score using modified point biserial correlation

PBCC measure is an alternative to PCC when the first variable is continuous, and the second variable is categorical binary. Determining correlation r_{Bpb} with the binary class used to assess the discriminative power of each bag feature *j*. The correlation is named as BagCorr-FC. Refer Table 4 for the description of Eq. (4).

$$r_{Bpb}(f_j, c_i) = \frac{\overline{y_0} - \overline{y_1}}{S_j}\sqrt{\frac{n_0 n_1}{n^2}} \tag{4}$$

Table 4. Conventional PBCC vs. Extended PBCC.

Conventional PBCC: Eq. (2)	**Extended PBCC (Bag-based): Eq. (4)**
$\overline{y_0}$ = average of feature *j* for class 0	$\overline{by_0}$ = average of bag feature *j* for bag class 0
$\overline{y_1}$ = average of feature *j* for class 1	$\overline{by_1}$ = average of bag feature *j* for bag class 1
S_j = standard deviation of feature *j*	bS_j = standard deviation of bag feature *j*
n_0 = respective numbers of 0s class	n_0 = respective numbers of 1s bag class 0
n_1 = respective numbers of 1s class	n_1 = respective numbers of 1s bag class 1
n = the total number features	n = the total number bag features

The parameters used in Eq. (4) differ from conventional Eq. (2) which use instance feature information and ambiguous class information which are not genuine instance's class to evaluate the relevancy score. While the proposed Extended PBCC (EPBCC) involved bag feature information to evaluate the feature relevance by checking its correlation to the real bag class provided in the dataset.

3.5. Classification task

Once the feature was transformed, the MI dataset can now be treated similar to a conventional classification. The acquired selected features by the proposed FBC-FS will then be fed as input to Support Vector Machine (SVM), K-Nearest Neighbour (KNN) and Decision Tree (DT) classifiers in order to predict the image classes. These three are common techniques used in breast cancer diagnosis [46]. SVM is a margin classifier that draws optimal hyperplanes between two classes in the feature vector. KNN predicts the class of observation that is dominant among *k* number of nearest neighbours in the feature vector. DT predicts the class by learning simple decisions. The default parameters used in this study: linear for

SVM and k = 5 for KNN. The dataset was divided into training set and test set using 10-fold cross-validation approach to avoid biased results as follows [47].

3.6. Performance evaluation

The impact of the proposed technique will be evaluated based on the final classification accuracy (Acc.) produced by the standard classifier. The evaluation metrics used include the True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN). These metrics are then represented by a confusion matrix, which allow the performance Acc. of the classifiers to be measured as described in Table 5.

Table 5. Performance measurement based on confusion matrix.

Measurement	Indication of Classification Result
Acc. = (TP+TN)/ N, where N=TN+TP+FN+FP	The score of correct images classified from the total number of samples (N)

Apart from the confusion matrix, the assessment of the goodness of a classifier's prediction from variation of features set in this study can be seen from the ROC curves [48]. Area under the curve (AUC) is acquired from the ROC, which provides a numerical measurement on the classifier's performance. Each point represents a sensitivity (probability of correctly identifying positive image) also known True Positive Rate (TPR) and specificity (probability of correctly identifying negative images) or also known True Negative Rate (TNR) correspond to a particular decision threshold. The performance of classifier by value of AUC can be categorized as excellent (1.0 >AUC > 0.9 or 90%), good (0.9 > AUC > 0.8 or 80%) and not good for disease diagnosis (AUC < 0.7 or 70%).

4. Results and Discussion

The average accuracy (Acc.), average AUC and average ROC curves presented in this section are results per test set of the 10-fold cross validation process using SVM, KNN and DT techniques respectively for the UCSB Breast Cancer and Tiger image. Overall, the performance of multi-instance (MI) classification was improved by the reduced features set by the proposed FBC-FS technique when compared to the full features for all classifiers tested.

The average ROC curves shown in Fig. 4 indicates that the selected features by multiple bag-fused settings can produce classifications with larger AUC on UCSB Breast Cancer using KNN compared to other settings. It also proved the better ROC curve when classified using SVM and DT. Refer Figs. A-1 and A-2 (*Appendix A*).

The average AUC achieved is categorized as excellent performance up to 91.5% for UCSB Breast Cancer and 91.2% for Tiger dataset. Even though the AUC score of DT fed by the selected feature of the proposed technique is slightly lower than 80%, it is still higher than other settings. Refer to the bold score in Table 6. The results indicated that the fusion step with considering all statistical central tendency measures of the multiple bag summary with mean, median, and mode at Fig. 1(c) give comprehensive information to discriminate the class by the reduced features feeder to the classifier. This fusion step as proposed technique automatically caters to open-ended data distribution which may have the possibility of extreme or undetermined correlation score when only focusing on one central tendency measure. This means the selected feature sets have a better classification performance than that produced by the single bag summary of feature and full features. The improvement achieves 2.5-16%.

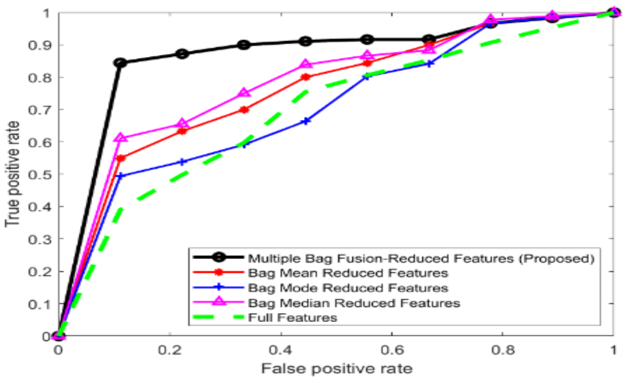

Fig. 4. Comparison ROC Curve on variation of reduced features vs. full features of UCSB breast cancer dataset using K-Nearest Neighbours.

Table 6. Average accuracy and area under ROC curve on MI image with variation of features evaluation setting.

Dataset	Feature Evaluation of Setting	Reduced Feature (%) Bag Mean		Bag Mode		Bag Median		Multiple Bag Fusion (Proposed)		Full (%) -		Proposed FS vs. Without FS (%)
	Classifier	**Acc.**	**AUC**	**Acc.**	**AUC**	**Acc.**	**AUC**	**Acc.**	**AUC**	**Acc.**	**AUC**	**Improved**
UCSB Breast Cancer	SVM	74	74.6	72.2	68.8	82	84.2	**84.9**	**86.3**	78.5	75.4	6.4
	KNN	69.9	79.6	73.4	74.2	80	82.6	**84.3**	**91.5**	68.3	71.7	**16.0**
	DT	62.9	61.7	60.7	54.6	69.5	67.5	**77.5**	**79.2**	68.3	62.5	9.2
Tiger	SVM	83.4	91.1	81.3	**91.5**	84.5	91.1	**84.6**	91.2	80.7	86.8	3.9
	KNN	77.7	86	71.5	78.2	78.3	86.7	**80.4**	**86.8**	77.9	79.1	2.5
	DT	76.1	77.7	74.7	79.8	**76.3**	**81.2**	**76.3**	79.6	73.7	77.3	2.6
Average Score		74	78.5	72.3	74.5	78.4	82.2	**81.3**	**85.8**	74.7	75.5	6.6

In the perspective of feature evaluation in terms of relevance and redundancy criteria, Table 7 shows 93.4% final average reduced number of features for UCSB Breast Cancer image dataset and 97% reduced for dataset Tiger image dataset. Based on the findings from the experiments carried out, the reduced feature set improved the accuracy of both MI image datasets and more effective to UCSB Breast Cancer image.

Table 7. The reduced number of features using FBC-FS.

Feature Set-› / Dataset↓	Full	Reduced Relevance Only	Reduced Selected	% Final Reduced
UCSB Breast Cancer	708	99	47	93.4%
Tiger	230	9	7	97%

The corresponding ROCs are shown in Fig. 5, which show that there is different performance between considering relevance features only and excluding strong redundancy features in the relevance features to feed classifiers. Generally, all MI data have correlated features. Therefore, features with redundancy score more than 0.69 are strongly correlated among features [30, 33] was excluded. Figure 5 results proved that considering both criteria are better of average ROC Curves than considering only single criterion (relevance only) along with the bag-based correlation measures by the proposed FS technique for the KNN classifier applied on UCSB Breast Cancer dataset. It also proved the better curve when the reduced dataset was classified using SVM and DT (refer Figs. A-3 and A-4 (*Appendix A*)). The bag-based summary is an information to extend the parameter used in the redundancy measure (by EPCC) and relevance measure (by EPBCC) to get a bag-based correlation score as Figure 1(b) which gives promising results.

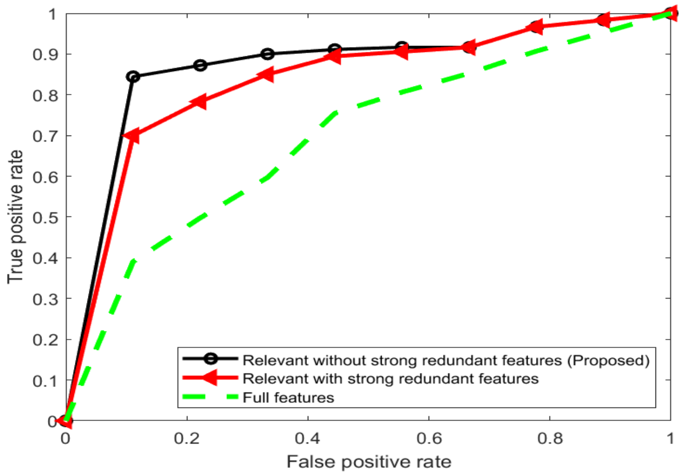

Fig. 5. Comparison ROC Curve on Full Features, with and without redundancy features set for UCSB breast cancer dataset using K-Nearest Neighbor.

Overall, transforming instance-based feature into bag-based feature setting as full information being a good parameter of correlation coefficients measure to assess criteria for features evaluation. This process fills the gap of genuine label of instance is unknown in MI data and evaluation of relevancy using bag class label cannot be done directly. The reduced features by consideration of relevancy score only and the integrated relevance score without strong redundancy score calculated by different bag summary correlation input.

5. Conclusions

Feature selection is a limited issue discussed in image classification problems under MIL. This study proposes Fusion Bag-based Correlation Feature Selection (FBC-FS) technique. By considering the MI data, especially medical image data, have correlated instances as well as correlated features, the proposed approach is featured by 3 steps: transformation-evaluation-fusion. Experiments indicate that the proposed system achieves much improved performance, measured by the accuracy and AUC of ROC, over the approach. These results clearly demonstrate the great potential of the proposed approach in the classification of MI data and cancer MI image data specifically. In the proposed technique, the bag summary transformation contributed to the feature selection task which analyses features as a whole, which was only used for classification tasks in the previous studies. The fused features acquired by the proposed technique are superior in terms of accuracy compared to single bag summary. The three measures complement represents the central tendency of data. As MI images are correlated, it is assumed that the correlated feature problem also exists. It is also proved that consideration of correlation measures to evaluate relevance and redundancy criteria to find optimal features contribute to the improvement. In the future, this study is hoped to explore other statistical summarization measures such as variance, skewness. Furthermore, it is also hoped to explore different datasets.

Acknowledgement

Communication of this research is made possible through monetary assistance by Universiti Tun Hussein Onn Malaysia (UTHM) and the UTHM Publisher's Office via Publication Fund E15216.

References

1. Richens, J.G.; Lee, C.M.; and Johri, S. (2020). Improving the accuracy of medical diagnosis with causal machine learning. *Nature Communications*, 11(1),1-9.
2. Itri, J.N.; Tappouni, R.R.; McEachern, R.O.; Pesch, A.J.; and Patel, S.H. (2018). Fundamentals of diagnostic error in imaging. *Radio Graphics*, 38(5), 1845-1865.
3. Venkatesan, R.; Chandakkar, P.S.; and Li, B. (2016). Simpler non-parametric methods provide as good or better results to multiple-instance learning. *Proceedings of the IEEE International Conference on Computer Vision* (*ICCV*), Santiago, Chile,2605-2613.
4. Singh, B.K.; Verma, K.; and Thoke, A.S. (2015). Adaptive gradient descent backpropagation for classification of breast tumors in ultrasound imaging. *Procedia Computer Science*, 46(1), 1601-1609.
5. Hesamian, M.H.; Jia, W.; He, X.; and Kennedy, P. (2019). Deep learning techniques for medical image segmentation: achievements and challenges. *Journal of Digital Imaging*, 32(4), 582-596.
6. Liu, S.; Cai, W.; Liu, S.; Pujol, S.; Kikinis, R.; and Feng, D. (2015). Subject-centered multi-view feature fusion for neuroimaging retrieval and classification. *Proceedings of the IEEE International Conference on Image Processing* (*ICIP*2015), Quebec City, QC, Canada, 2505-2509.
7. Quellec, G.; Cazuguel, G.; Cochener, B.; and Lamard, M. (2017) Multiple-instance learning for medical image and video analysis. *IEEE Reviews in Biomedical Engineering*, 10(1), 213-234.
8. Wu, J.; Pan, S.; Zhu, X.; Zhang, C.; and Wu, X. (2018). Multi-instance learning with discriminative bag mapping. *IEEE Transactions on Knowledge and Data Engineering*, 30(6), 1065-1080.
9. Shi, B.; Chen, Y.; Zhang, P.; Smith, C.D.; and Liu, J. (2017). Nonlinear feature transformation and deep fusion for Alzheimer's disease staging analysis. *Pattern Recognition*, 63, 487-498.

10. Carbonneau M.-A.; Cheplygina, V.; Granger, E.; and Gagnon, G. (2017). Multiple instance learning: A survey of problem characteristics and applications. *Pattern Recognition*, 77, 329-353.
11. Cheplygina, V.; Sørensen, L.; Tax, D.M.J.; Pedersen, J.H.; Loog, M.; and de Bruijne, M. (2014). Classification of COPD with multiple instance learning. *Proceedings of the* 22*nd International Conference on Pattern Recognition* (*ICPR* 2014), Stockholm, Sweden, 1508-1513.
12. Carbonneau, M.A.; Granger, E.; and Gagnon G. (2016). Witness identification in multiple instance learning using random subspaces. *Proceedings of the* 23*rd International Conference on Pattern Recognition* (*ICPR* 2016), Cancun, Mexico, 3639-3644.
13. Dietterich, T.G.; Lathrop, R.H.; and Lozano-Pérez, T. (1997). Solving the multiple instance problem with axis-parallel rectangles. *Artificial Intelligence*, 89(1-2), 31-71.
14. Xu, Y.; Mo, T.; Feng, Q.; Zhong, P.; Lai, M.; and Chang, E.I-C. (2014). Deep learning of feature representation with multiple instance learning for medical image analysis. *Proceedings of the* 2014 *IEEE International Conference on Acoustics, Speech and Signal Processing* (*ICASSP*), Florence, Italy,1626-1630.
15. Sudharshan, P.J.; Petitjean, C.; Spanhol, F.; Oliveira, L.E.; Heutte, L.; and Honeine, P. (2019). Multiple instance learning for histopathological breast cancer image classification. *Expert Systems with Applications*, 117, 103-111.
16. Sun, Y.Y.; Ng, M.K.; and Zhou, Z.-H.H. (2010). Multi-instance dimensionality reduction. *Proceedings of the Twenty-Fourth Conference on Association for the Advancement of Artificial Intelligence* (*AAAI* 2010), Atlanta, Georgia, USA, 587-592.
17. Cheplygina, V.; Tax, D.M.J.; and Loog, M. (2015). On classification with bags, groups and sets. *Pattern Recognition Letters*, 59(1), 11-17.
18. Mercan, C.; Askoy, S.; Mercan, E.; Shapiro, L.G.; Weaver, D.L.; and Elmore, J.G. (2018). Multi-instance multi-label learning for multi-class classification of whole slide breast histopathology images. *IEEE Transactions on Medical Imaging*, 37(1), 316-325.
19. Yuan, X.; Hua, X.-S.; Wang, M.; Qi, G.-J.; and Wu, X.-Q. (2007). A novel multiple instance learning approach for image retrieval based on adaboost feature selection. *Proceeding of the IEEE International Conference on Multimedia and Expo*, Beijing, China, 1491-1494.
20. Latham, A.C. (2016). *Multiple-instance feature ranking*. PhD Thesis of Case Western Reserve University.
21. Zafra, A.; Pechenizkiy, M.; and Ventura, S. (2013). ReliefF-MI: An extension of relieff to multiple instance learning. *Neurocomputing*, 75(1), 210-218.
22. Gan, R. and Yin, J. (2013). Feature selection in multi-instance learning. *Neural Computing Application*, 23(1), 907-912.
23. Zafra, A.; Pechenizkiy, M.; and Ventura, S. (2013). HyDR-MI: A hybrid algorithm to reduce dimensionality in multiple instance learning. *Information Sciences*, 222(1), 282-301.
24. Bugata, P.; and Drotar P. (2020). On some aspects of minimum redundancy maximum relevance feature selection. *Science China Information Sciences*, 63(1),1-15.

25. Yuan, L.; Zhao, L.; and Xu, H. (2015). Multi-instance learning via instance-based and bag-based representation transformations. *Proceedings - International Conference on Image Processing*(*ICIP*2015), Québec City, Canada,, 2771-2775.
26. Aggarwal, C.C.; Hinneburg, A.; and Keim, D.A. (2001). On the surprising behavior of distance metrics in high dimensional space. *Lecture Notes in Computer Science*, 1973(1), 420-434.
27. Dong, L. (2006). *A comparison of multi-instance learning algorithms*. Master Thesis, University of Waikato, New Zealand.
28. Taser, P.Y.; Birant, K.U.; and Birant, D. (2019). Comparison of ensemble-based multiple instance learning approaches. *Proceedings of the IEEE International Symposium on Innovations in Intelligent Systems and Applications* (*INISTA* 2019), Sofia, Bulgaria,1-5.
29. Zhao, Y.; Yang, F.; Fang, Y.; Liu, H.; Zhou, N.; Zhang, J.; Sun, J.; Yang, S.; Menze, B.; Fan, X.; and Yao, J. (2020). Predicting lymph node metastasis using histopathological images based on multiple instance learning with deep graph convolution. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*(*CVPR*), Seattle, WA, USA, 4836-4845.
30. Mukaka, M.M. (2012). A guide to appropriate use of correlation coefficient in medical research. *Malawi Medical Journal*, 24(3), 69-71.
31. Guha, R.; Ghosh, K.K.; Bhowmik; S.; and Sarkar, R. (2020). Mutually informed correlation coefficient (MICC)-a new filter based feature selection method. *Proceedings of the IEEE Calcutta Conference* (*CALCON* 2020), Kolkata, India, 54-58.
32. Pooja, S.B.; and Siva Balan, R.V. (2019). Point biserial correlated feature selection of weather data. *International Journal of Engineering and Advanced Technology*, 8(6), 1854-1857.
33. Schober, P.; and Schwarte, L.A. (2018). Correlation coefficients: Appropriate Use and Interpretation. *Anesthesia and Analgesia*, 126(5), 1763-1768.
34. Rovetta, A. (2020). Raiders of the lost correlation: A guide on using Pearson and Spearman coefficients to detect hidden correlations in medical sciences. *Cureus*, 12(11): e11794.
35. Sugianela Y.; and Ahmad, T. (2020). Pearson correlation attribute evaluation-based feature selection for intrusion detection system. *International Conference on Smart Technology and Applications* (*ICoSTA*), 1-5.
36. Auffarth, B.; López, M.; and Cerquides, J. (2010). Comparison of redundancy and relevance measures for feature selection in tissue classification of CT images. Lecture. Notes Computer Science, 6171, 248-262.
37. Santos, M.S.; Abreu1, P.H.; Germán, R.-B.; and García-Laencina, P.J. (2018). Improving the classifier performance in motor imagery task classification: What are the steps in the classification process that we should worry about? *International Journal of Computational Intelligence Systems*, 11(1), 1279-1293.
38. Adel, M.; Garali, I.; Pan, X.; Fossati, C.; Gaidon, T.; Wojak, J.; Bourennane, S.; and Guedj, E. (2020). *Alzheimer's disease computer-aided diagnosis on positron emission tomography brain images using image processing techniques*. In Computer methods and programs in biomedical signal and image processing. London, United Kingdom: IntechOpen.

39. Fenwaa, O.D.; Ajalab, F.A.; and Akuc, A.M. (2015). Medical image classification with convolutional neural network. 13*th International Conference on Control Automation Robotics & Vision* (*ICARCV*), 844-848.
40. Lashari, S.A.; and Ibrahim., R. (2013). A Framework for medical images classification using soft set. *Procedia Technology*, 8C (2013), 537-545.
41. Kandemir, M.; Zhang, C.; and Hamprecht, F.A. (2014). Empowering multiple instance histopathology cancer diagnosis by cell graphs. *The Series Lecture Notes in Computer Science*, 8674, 228-235.
42. Møllersen, K.; Hardeberg, J.Y.; and Godtliebsen, F. (2020). A probabilistic bag-to-class approach to multiple-instance learning. *Data*, 5(2), 56.
43. Herrera, F.; Ventura, S.; Bello, R.; Cornelis, C.; Zafra, A.; Sánchez-Tarragó, D.; and Vluymans, S. (2016). *Multiple instance learning: Foundations and algorithms*. (1st ed.). Cham: Springer International Publishing.
44. Cheplygina, V. (2018). A repository of MIL datasets. https://figshare.com/articles/MIProblems_A_repository_of_multiple_instance_learning_datasets/6633983.
45. Cheplygina, V.; and Tax, D.M.J. (2015). Characterizing multiple instance datasets. *Lecture Notes in Computer Science*, 9370, 15-27.
46. Yassin, N.I.R.; Omran, S.; El Houby, E.M.F.; and Allam, H. (2018). Machine learning techniques for breast cancer computer aided diagnosis using different image modalities: A systematic review. *Computer Methods and Programs in Biomedicine*, 156, 25-45.
47. Komárek, T.; and Somol, P. (2019). Multiple instance learning with bag-level randomized trees. *Lecture Notes in Computer Science* (*including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics*) 11051, 259-272.
48. Chen, R.-C.; Dewi, C.; Huang, S.-W.; and Caraka, R.E (2020). Selecting critical features for data classification based on machine learning methods. *Journal of Big Data*, 7(1), 52.

Appendix A

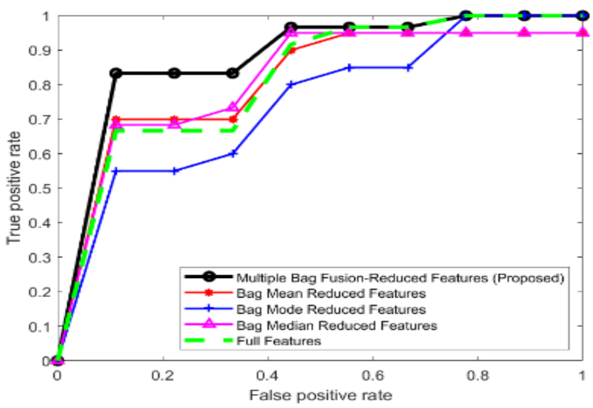

Fig. A-1. Comparison ROC curve on variation of reduced features vs. full features of UCSB breast cancer dataset using support vector machine.

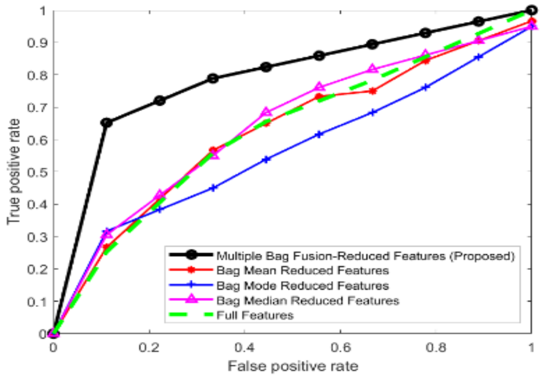

Fig. A-2. Comparison ROC Curve on variation of reduced features vs. full features of UCSB breast cancer dataset using decision tree.

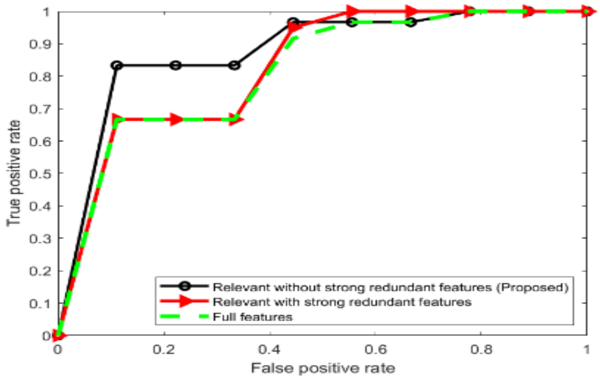

Fig. A-3. Comparison ROC curve on full features, with and without redundancy features set for UCSB breast cancer dataset using support vector machine.

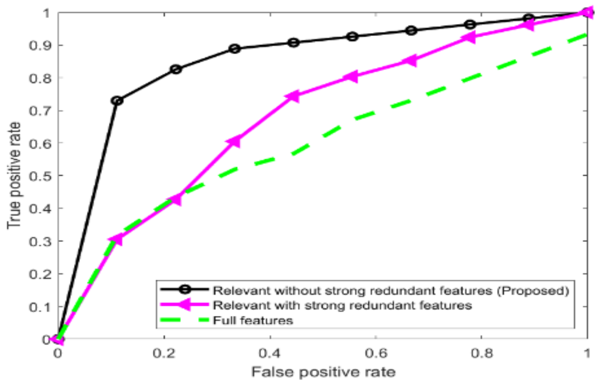

Fig. A-4. Comparison ROC curve on full features, with and without redundancy features set for UCSB breast cancer dataset using decision tree.