CUOB-ReliefF: Diagnosis of breast cancer by balancing datasets

Zeinab Abbasi, Mohsen Rahmani, Hossein Ghaffarian
Faculty of Engineering
Arak University
Arak, Iran, 38156 – 879
Email:zabasi@gmail.com, m-rahmani@araku.ac.ir, h-ghaffarian@araku.ac.ir

Abstract— **One of the challenges of artificial intelligence and data mining algorithms in the automatic diagnosis of diseases is imbalanced dataset problem. The lack of data balancing will reduce accuracy of the results, which is very dangerous in diseases like breast cancer. This paper presents an algorithm for balancing number of instances in breast cancer datasets. The proposed algorithm uses ReliefF for weighting and ranking of instances. ReliefF is a well-known algorithm for ranking features, but, here, we used it with some modifications to rank the instances. After ranking the instances, based on the weight obtained, a combination of undersampling and oversampling methods is used to balance the dataset. The obtained results from testing the proposed algorithm on two datasets show the effectiveness of this algorithm.**

Keywords-Breast cancer; Imbalanced datasets, ReliefF, Undersampling, Oversampling.

I. INTRODUCTION

The most important challenge for computer-aided diagnosis (CAD) systems in the field of medicine is the correct diagnosis of patients from healthy people. This problem is more important in dangerous and prevalent diseases. One of the most dangerous and epidemic diseases in women is breast cancer. International Agency for Research on Cancer (IARC)[1] predicts 2,888,849 new incident cases of breast cancer in 2018. Breast cancer incidence and mortality are rapidly growing worldwide [2].

A well-timed diagnosis of this disease can prevent its progress and help its treatment faster. One of the applications of data mining is the creation of algorithms for classification and prediction for patient diagnosis. Diagnostic system can be trained by collecting values for a number of healthy and sick people. When new person's information is given to the system, the system declares based on previous values whether it is a patient or not. The correct diagnosis of a patient is not the endpoint, and the patient's information can give a doctor a new viewpoint on the illness and its treatment [3].

Unfortunately, the problem with collecting information is that number of sick people is often less than number of healthy people. Therefore, the dataset does not have enough information to train. In this type of datasets, called "Imbalanced datasets", the prediction error usually increases and algorithms usually classify new items as healthy people[4]. A dataset is imbalanced if the number of instances of one or some of classes is much smaller than the number of instances of other classes [5]. Even, sometimes, the ratio between classes may be 1: 100 or more [6]. A class with fewer instances is a minority class, and majority class is another prevailing class. Unfortunately, in many cases, the data of minority class is more important [7].

Imbalanced datasets are not limited to the medical field and many datasets about natural phenomena and real-world problems are imbalanced. Examples of these areas are: detect oil spill with satellite imagery, learning the pronunciation of words, text classification [8], data retrieval, fraud detection [9], determine the credit of customers for loan payments, telecommunications management, speech recognition[10, 11, 12, 13]and so on.

For the following reasons, conventional classification algorithms are not suitable for classifying imbalanced datasets[5]:

1. Standard classifiers just perform for a balanced dataset, correctly.
2. Standard performance measures, such as accuracy, are not suitable for deciding on the performance of classifiers in imbalanced sets.
3. In the training phase, instances of minority class can be considered as noise, and vice versa.
4. Sometimes instances of minority class overlap with instances of other classes and their separation is not easy.
5. Low number of instances leads to a lack of proper classification of the class and failure in the recognition of minority class instances.

Many researchers develop different techniques to solve the imbalanced sets problem and improve the performance of classifiers. They divide the techniques into three broad categories [14, 15]:

- Algorithm level methods: These methods are based on the correction of previous algorithms to fit them into imbalanced data sets.
- Data level methods: These methods are independent of the classifiers. They are used to reduce the imbalance

rate of the datasets. Data level methods perform resampling methods as a preprocessing operation.

- Hybrid data level and algorithm level methods: These methods first find instances that are more important, give them higher weights, and then by combining several weak classifiers, create a strong classifier.

Sampling techniques, which are part of the data level methods, are divided into two general categories: undersampling and oversampling. The undersampling methods eliminate some of the majority class instances. The oversampling methods add some new minority instances by repeating the previous information or producing artificially [16]. Most researches have only used one of these methods. But this paper combines undersampling and oversampling methods to balance the ratio of two healthy and patient classes.

The proposed algorithm in this paper is based on ReliefF algorithm [17]. Although ReliefF is a ranking and feature selection algorithm, we use ReliefF to rank the instances and sampling. The name of our proposed algorithm is CUOB-ReliefF (Combining Undersampling and Oversampling Based on ReliefF). In CUOB-ReliefF, we first compute a weight for each instance. This weight is calculated using Jaccard index and based on its similarity with other instances in its class and its opposite class. After this, based on the user-defined sampling rate, only the instances with higher weights in the majority class are retained and the rest will be removed. For the minority class, some of the best instances are repeated to reach the desired sampling rate of the user. At the end, the number of instances of the two classes will be approximately equal. The efficiency of this method has been tested on Wisconsin Breast Cancer dataset (WBCD) and Wisconsin Diagnostic Breast Cancer dataset (WDBCD). The results have been improved according to the measures mentioned in the experimental study section.

Rest of the paper is organized as follows. Section II reviews the related works and explains some background contents. Section III presents CUOB-ReliefF algorithm that uses ReliefF for diagnosing the breast cancer. Section IV describes the experimental studies and shows the results of tests. Finally, section V concludes the work and suggests directions for future research.

II. Related Works

In this section, some papers that use data mining techniques to diagnose breast cancer are examined. One of the first applications of machine learning algorithms for detecting breast cancer is in [18]. In this paper, fuzzy logic and genetic algorithms, a kind of evolutionary algorithm, are combined to create an automatic detection system for cancer. The proposed algorithm is tested on the WBCD and the results are presented.

Researchers in [19] develop a breast cancer diagnosis system using Association Rules (AR) and Neural Networks (NN). Association rules investigate the relationships between the data. They use AR to reduce the dimension and find the related features, and then, using the Multi-Layer Perceptron (MLP), a kind of NN, classify the instances of WBCD.

In [20], for all the features, F-score is calculated and the features with larger F-score are selected. F-score is a technique for measuring the discrimination between two sets of real numbers. Authors of the paper use Support Vector Machine (SVM) for classification of the instances. The Grid search method is used to calculate the parameters of the SVM model. The authors repeat his process for all features.

In [21], a step-by-step approach is introduced to create a diagnosis breast cancer system. They first extracted the features with the Principal Component Analysis (PCA), then ranked the features and selected useful features with Signal To Noise Ratio (SNR) and Sequential Forward Selection (SFS) algorithms. In the next step, the selected data is given to SVM, K-Nearest Neighbors (K-NN) and Probabilistic Neural Network (PNN) classifiers to classify the instances into benign/malignant classes.

Authors of [22] propose a hybrid method for managing imbalanced data, especially cancer detection. They used the Synthetic Minority Over-Sampling Technique (SMOTE) algorithm, a well-known Over-Sampling algorithm, to balance the data set. Subsequently, the output data was classified by Artificial Immune Recognition System (AIRS).

In [23], Bagging and Boosting, two methods of combining weak classifiers to create powerful classifiers, have been used. In this method, in a repeat process, a random subset of the majority class, which is equal to the minority class, is selected. Then, two output sets of Bagging and Boosting are given to a classifier to obtain the results. This method is time-consuming for large sets.

Authors of [24], obtain new features of the main to increase the accuracy of the diagnosis of cancer. The K-means algorithm is used to recognize new patterns of benign and malignant tumors. The membership of each tumor to these patterns is calculated and added to the training model as a new feature. Six new features are created from the 32 basic features. Then, SVM classifier is used to detect cancerous instances.

Researchers in [25] made a comparison between accuracy and efficiency of some classifiers for the classification of WBCD. These classifiers include SVM, Decision Tree (C4.5), Naive Bayes (NB) and KNN. The results of the experiments show that, in many respects, SVM works better than other classifiers for this problem.

Researchers in [26] use an ensemble of Radial Basis Function Network (RBFN), Generalized Regression Neural Network (GRNN) and Feed Forward Neural Network (FFNN) for the diagnosis of breast cancer. In their method, the dataset is firstly classified with each of the mentioned methods and then the final decision is taken based on the total votes of the classifiers.

In [27], breast cancer is detected by the data mining algorithm in three stages. First, SMOTE algorithm is used to balance the data set with high imbalance rates. In the second step, Bayesian optimization is used to train a set of basic classifiers. In the final step, a stacking method is applied to optimized classifiers and a combinatory classifier is created.

III. PROPOSED ALGORITHM: CUOB-RELIEFF

This section explains our proposed algorithm for breast cancer diagnosis. As previously explained, CUOB-ReliefF is based on a ReliefF algorithm. So, first, ReliefF algorithm is described generally. ReliefF is a well-known feature selection algorithm used as a pre-processing step in many data mining problem. This algorithm is capable of working on multi-class problems, with nominal-numeric features and missing values. The ReliefF calculates the importance of the features by comparing them[28].

The workflow of ReliefF algorithm is as follows:

First, number *L*, the number of replicas of the main loop of the algorithm, is selected by the user, and in each run of this loop, a random instance of the data set is selected. Then *B* number of the nearest instances of the same class, called *Hit set*, and *b* number of nearest instances of the opposite class, called *Miss set*, relative to the selected instance are calculated.The similarity between the random instance with members of the Hit and Miss sets is calculated on the basis of an evaluation function. Then based on the similarity value, a weight for each feature is calculated. The original ReliefF uses Manhattan distance as the evaluation function. After weighing all the features, features with higher weights are selected and the rest are eliminated.

The proposed algorithm, CUOB-ReliefF, uses the same technique as ReliefF, with the difference it is used to select the most valuable instances. In fact, the information and features recorded from a patient or healthy person are stored as an instance. Valuable instances are those that provide more useful information for training the diagnosis system. The goal of this algorithm is to achieve a balance between the number of instance of patient and healthy classes. In CUOB-Relief, a different evaluation function is used, also algorithm is parallelized to increase the speed, especially for large-scale data sets. Pseudo code of CUOB-Relief is shown in Figure 1.

CUOB-ReliefF Algorithm (input:X, B,SR)

1. /**X*= set of training examples */
2. /**B*= number of nearest neighbors to compute*/
3. /**L*= number of instances.*/
4. /**C*= number of classes*/
5. /**SR*= user-defined selection rate (in percent) */
6. For *t*:=1 to *L* do
   a. Find *B* nearest instances to x_t from class y_t (Hit set) by Jaccard index
   b. For each class $c \neq y_t$
      Find *B* nearest instances to x_t from class *c* and add to Miss set by Jaccard index
      End for (line 4.b)
   c. Calculate the weight of each instance x_t:

$$w_t = \sum_{c \neq y_t} \frac{1}{LB} \sum_{(xj,yj) \in Miss} \delta(x_t, x_j) - \frac{1}{LB} \sum_{(xi,yi) \in Hit} \delta(x_t, x_i)$$

   End for (line 6)
7. Sort the weight array w_t.
8. *F= L*(SR/C)*/100 ** Number of final selected instances for each class
9. For each class *c*
   a. If *F*<= count(*c*)
      Run under-sampling ** Select *F* instances with more weight
      else
      Run over-sampling ** Select 50 % of more weighed instances, repeatedly until *F* instances of class *c*is added to the final set.

End For (line 9)

Figure 1. Pseudo-code 1.CUOB-ReleifF algorithm.

As shown in Figure 1, CUOB-ReliefF works as follows:

The main *for* loop in line 6, for all instances of the dataset, perform the following procedure: In line 6-a, a set of nearest instances is found from the same class of the selected instance and in line 6-b, the Miss set is created. In line 6-c, using the given formula, a weight is calculated for each instance. Here, δ represents the evaluation function used to calculate the similarity between two instances. Based on this formula, instances that are very similar or repetitive will be taken less weight. In line 7, the samples are sorted according to the calculated weight.

The SR is a user-defined value. This number indicates that what percentage of the total data must be selected. This value is divided equally between two classes. For example, if SR=80, 560 instances from 699 instances of WBCD are selected, half of them from the minority class and the rest from the majority class. As explained below, if a class has fewer instances from this value, new instances will be created from previous examples. Inside the *for* loop in line 9, if the calculated *F* value is less than the number of instances of class *c*, top *F* instances that are more weighted are selected. In fact, the undersampling operation is executed. If the *F* value is greater than the number of instances in class *c*, 50% of the instances with topmost weights would be copied. Repeat operation continues as long as the number of instances of class *c* is equal to *F*. This operation is called oversampling.

Because of finding the nearest neighbors of each instance in an independent way, in multiprocessor systems, the *for* loop in line 6 can run in parallel. Therefore, the runtime of the algorithm decreases, particular for large-scale data sets.

As mentioned before, we use the Jaccard index as an evaluation function δ in COUB-ReliefF. The Jaccard index is a statistic index used for measuring the similarity of two sets. If $x=(x_1,x_2,...,x_n)$ and $y=(y_1,y_2,...,y_n)$ are two sets with all real numbers ≥ 0, their Jaccard index is defined as follows:

$$J(X,Y) = \frac{\sum_i \min(x_i,y_i)}{\sum_i \max(x_i,y_i)} \quad (1)$$

It should be noted that if we want to run this formula on features with negative values, we must normalize the values in the range 0 to 1.If two instances are equal, the result of (1) is 1 and if they are completely different the result is 0.

If *x* and *y* are two nominal (non-numerical) sets, the Jaccard index is defined as follows:

$$J(X,Y) = \frac{|X \cap Y|}{|X \cup Y|} \quad (2)$$

If value of each of the x_i or y_i is unknown, its subscription will be null ($x_i \cap y_i = \emptyset$). Based on type of features (numerical or nominal), we use (1) or (2) to calculate the similarity of two instances. In addition, if a feature has a missing value, the algorithm applies (2). Finally, the result is obtained by calculating the average value of (1) and (2). For each instance, neighbors with a higher Jaccard index, as the nearest instances, are placed in the *Hit* and *Miss* set. Given the line 6-c in the Pseudo-code, if an instance is similar to other instances of its class, its weight will decreased. On the other hand, whatever it looks like to instances in other classes, its weight will increased.

IV. Experimental Results

In this section, the system and data sets specifications are presented and the results of CUOB-ReliefF are compared with the results of some other methods. To implement the CUOB-ReliefF algorithm, we used MATLAB 2013 and Weka. In addition, our hardware had an Intel core 2 Dou 2.26GHz CPU with 3 GB RAM, run a 32-bit Windows 7.

To compare and evaluate the results of the CUOB-ReliefF algorithm with other methods, the results should be trained and tested by classifiers. In this paper, we use three classifier of Weka for evaluation: MLP [29], Random Forest (RF) [30] and C4.5[31]. In addition, we used 5-fold cross-validation strategy to sure from unbiased comparisons of the classification results.

Usually, to compare the performance of classifiers, the standard accuracy is used. Accuracy indicates the number of instances that are correctly classified. However, this measure is not suitable for the imbalanced data sets. Because, if the minority class contains 5% and the majority class contains 95% of the instances, even if the classifier always classifies the instances as the majority class, then its accuracy would be 95%. To calculate the performance in imbalanced sets, the concepts of *FN*, *FP*, *FP*, *TP*, and *TN* are used and measures are defined based on them.

When a classifier runs on a set of data, four outputs may be achieved:

- *TP*: The system correctly detects a positive case.
- *TN*: The system correctly detects a negative case.
- *FN*: That is, a positive case is detected negative.
- *FP*: That is, a negative case is detected positive.

Based on these outputs, different measures are defined. We use four common evaluation measures for imbalanced sets, defined in the following equations [32]:

$$Accuracy(\%) = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

$$Sensitivity(\%) = \frac{TP}{TP+FN} \quad (4)$$

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

$$G-Mean = \sqrt{Sensitivity\ \times Specificity} \quad (6)$$

As noted above, the accuracy is not an appropriate measure for evaluating algorithms on imbalanced data sets. However, since this metric is still used in many articles and is the most common way for assessing the performance of a classifier[5], we also used it here.

In this article, we used WBCD and WDBCD data sets to test cancer diagnosis using the proposed method.

These two data sets are downloaded from UCI machine learning repository[33]. WBCD was collected by Dr. William H. Wolberg in the University of Wisconsin – Madison Hospitals from 1989 to 1991 [34]. WDBCD was collected by Dr. William H. Wolberg and two of his colleagues in 1995 [35]. Features are computed from a digitized image of a Fine Needle Aspirate (FNA) of a breast mass. They describe the features of the cell nuclei present in the image. Specifications of these data sets are shown in Table I. Table II shows the results for WBCD. In this table, the results of the original data and the SMOTE algorithm, along with CUOB-ReliefF with three *SR* values are presented. The values used in SMOTE algorithm are:

- Seed used for random sampling = 1
- Percentage of SMOTE instances to create = 100
- Number of nearest neighbors = 5.

TABLE I. Data sets specifications.

	Features	Instance	Benign	Malignant	Missing value	Imbalance rate
WBCD	11	699	458	241	yes	1.9
WDBCD	32	569	357	212	no	1.68

TABLE II. Classification measures for WBCD.

Classifier	Algorithm	Accuracy	Sensitivity	Specificity	G-Mean
C4.5	***Original dataset***	94.277	95.633	91.701	93.646
	SMOTE	96.383	95.196	97.51	96.346
	CUOB-ReliefF 131%	96.506	95.633	97.38	96.502
	CUOB-ReliefF 80%	**97.857**	96.428	**99.286**	**97.846**
	CUOB-ReliefF 50 %	96.571	**96.571**	96.571	96.571
MLP	***Original dataset***	95.279	95.633	94.606	95.118
	SMOTE	96.063	96.07	96.058	96.063
	CUOB-ReliefF 131%	96.506	94.978	98.034	96.494
	CUOB-ReliefF 80%	**97.678**	96.071	**99.286**	**97.665**
	CUOB-ReliefF 50 %	94.857	**93.143**	96.571	94.841
RF	***Original dataset***	96.709	97.161	95.85	96.503
	SMOTE	97.021	96.506	97.51	97.007
	CUOB-ReliefF 131%	98.035	98.035	98.035	98.035
	CUOB-ReliefF 80%	**99.285**	**98.571**	**100**	**99.283**
	CUOB-ReliefF 50 %	96.857	97.143	96.571	96.857

The 131% selection rate in the CUOB-ReliefF algorithm means the majority class is fully selected and, in fact, undersampling has not been performed. As can be seen, the best SR is 80%. Although this amount is empirically selected, there is a reasonable reason for this. Due to the low number of instances, the SR 50% causes a large number of instances to be deleted. Also, selecting the rate 131%causes the undersampling does not run and some instances of the majority class with low quality are brought in the final set. These reasons reduce the quality of output and decrease the value of evaluation measures. It is also observed that the best results are obtained by RF classifiers.

Table III provides performance metrics for WDBCD. The selection rate 125% indicates that only oversampling has been done. As can be seen here, CUOB-ReliefF 80% achieves the best values for the evaluation measures(except sensitivity) using the RF classifier. Perhaps the reason why, as in the previous case, the SR 80% does not achieve the optimal amount in all evaluations, is that the majority class has more important information in WBCD, and eliminating its instances reduces performance.

In Table IV, the number of instances sampled for each class is compared by different algorithms. As can be seen, SMOTE algorithm does not completely balance the dataset and sometimes the number of instances of minority class is more than the majority class. In addition, although this algorithm has created the most number of instances in total, it has actually had lower performance. Increasing the number of instances in the large-scale dataset may cause side effects, such as the need for more storage space and stronger processors for saving and classifying data.

In Table V, the obtained results from CUOB-RELIEFF 80% are compared with the best results from some articles. Other papers mentioned in the related works, tested their methods on other data sets or did not mention the values for all measures, therefore, they are not listed in this table. Here, it is also clear that the results obtained with the CUOB-RELIEFF are better than all other methods in all of the evaluation metrics (except one case, with a slight difference).

TABLE III. CLASSIFICATION MEASURES FOR WDBCD

Classifier	Algorithm	Accuracy	Sensitivity	Specificity	G-Mean
C4.5	*Original data set*	93.849	94.398	92.924	93.926
	SMOTE	95.134	**95.518**	94.811	95.164
	CUOB-ReliefF 125%	**95.225**	94.382	96.067	**95.221**
	CUOB-ReliefF 80%	95.164	93.421	**96.916**	95.061
	CUOB-ReliefF 50 %	89.436	90.845	88.028	89.425
MLP	*Original data set*	96.309	97.759	93.868	95.794
	SMOTE	96.927	**98.599**	95.519	97.047
	CUOB-ReliefF 125%	**98.174**	98.034	**98.315**	**98.174**
	CUOB-ReliefF 80%	95.824	96.930	94.714	95.815
	CUOB-ReliefF 50 %	94.366	96.479	92.253	94.343
RF	*Original data set*	95.431	97.199	92.453	94.796
	SMOTE	96.799	96.919	96.698	96.808
	CUOB-ReliefF 125%	98.315	**97.753**	98.876	98.313
	CUOB-ReliefF 80%	**98.461**	97.368	**99.559**	**98.45**
	CUOB-ReliefF 50 %	92.957	95.774	90.141	92.915

TABLE IV. NUMBER OF SAMPLED INSTANCES FOR WBCD AND WDBCD.

Dataset	Algorithm	Instances	Benign	Malignant
WBCD	*SMOTE*	940	458	482
	CUOB-ReliefF 131 %	916	458	458
	CUOB-ReliefF 80%	560	280	280
	CUOB-ReliefF 50 %	350	175	175
WDBCD	*SMOTE*	781	357	424
	CUOB-ReliefF 125 %	712	356	356
	CUOB-ReliefF 80%	455	228	227
	CUOB-ReliefF 50 %	284	142	142

TABLE V. COMPARISON OF CUOB-RELIEFF WITH RELATED WORKS

Dataset	Algorithm	Accuracy	Sensitivity	Specificity	G-Mean
WBCD	*Asri(SVM) [25]*	97.13	97.38	96.265	96.821
	Osareh(SVM-RBF) [21]	98.80	95.45	99.63	97.517
	CUOB-ReliefF 80%	**99.285**	**98.571**	**100**	**99.283**
WDBCD	*Osareh(SVM-RBF) [21]*	96.33	96.85	93.11	94.961
	Yavuz(Ensemble Of NN) [26]	96.43	**97.62**	95.71	96.66
	CUOB-ReliefF 80%	**98.461**	97.368	**99.559**	**98.458**

V. CONCLUSION

One of the most dangerous and epidemic diseases in women is breast cancer. Unfortunately, due to the imbalanced data sets collected in this case, data mining algorithms are not able to diagnose it quite accurately. This paper presents a hybrid algorithm to solve the imbalanced dataset problem. The effectiveness of CUOB-ReliefF was tested on two breast cancer datasets (WBCD and WDBCD). CUOB-ReliefF first uses the ReliefF algorithm to rank the instances of data set and then balances the majority and minority classes by combining undersampling and oversampling methods. In addition, this algorithm parallelizes ReliefF on a multi-core CPU to reduce its runtime. The results presented in the experiments section show the performance and effectiveness of CUOB-ReliefF for diagnosis of breast cancer. In the future, we plan to use this algorithm to improve the detection of some other illnesses.

REFERENCES

[1] I. A. f. R. o. C. (IARC). [Online]. Available: http://gco.iarc.fr/.

[2] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre and A. .. Jemal, ""Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries."," CA: a cancer journal for clinicians, 2018.

[3] L. J. Mena and J. A. Gonzalez, "Machine Learning for Imbalanced Datasets: Application in Medical Diagnostic," in Flairs Conference, 2006.

[4] H. Parvin, B. Minaei-Bidgoli and H. Alinejad-Rokny, "A new imbalanced learning and dictions tree method for breast cancer diagnosis," Journal of Bionanoscience, vol. 7, no. 6, pp. 673-678, 2013.

[5] G. Haixiang, L. Yijing, J. Shang, G. Mingyun, H. Yuanyue and G. Bing, "Learning from class-imbalanced data: Review of methods and applications," Expert Systems with Applications, vol. 73, pp. 220-239, 2017.

[6] A. Anand, G. Pugalenthi, G. B. Fogel and P. N. Suganthan, "An approach for classification of highly imbalanced data using weighting and undersampling," Amino acids, vol. 39, no. 5, pp. 1385-1391, 2010.

[7] L. Yijing, G. Haixiang, L. Xiao, L. Yanan and L. Jinling, "Adapted ensemble classification algorithm based on multiple classifier system and feature selection for classifying multi-class imbalanced data," Knowledge-Based Systems, vol. 94, pp. 88-104, 2016.

[8] Y. Liu, H. T. Loh and A. .. Sun, "Imbalanced text classification: A term weighting approach," Expert systems with Applications, vol. 36, no. 1, pp. 690-701, 2009.

[9] C. Phua, D. Alahakoon and V. Lee, "Minority report in fraud detection: classification of skewed data," Acm sigkdd explorations newsletter, vol. 6, no. 1, pp. 50-59, 2004.

[10] B. W. Yap, K. Abd Rani, H. A. Abd Rahman, S. Fong, Z. Khairudin and N. Nik Abdullah, "An application of oversampling, undersampling, bagging and boosting in handling imbalanced datasets," in In Proceedings of the First International Conference on Advanced Data and Information Engineering , Singapore, 2014.

[11] S. Kotsiantis, D. Kanellopoulos and P. Pintelas, "Handling imbalanced datasets: A review," GESTS International Transactions on Computer Science and Engineering, vol. 30, no. 1, pp. 25-36, 2006.

[12] G. Wu and E. Y. Chang, "Class-boundary alignment for imbalanced dataset learning," in In ICML 2003 workshop on learning from imbalanced data sets II, Washington, DC, 2003.

[13] N. V. Chawla, "Data mining for imbalanced datasets: An overview," in Data mining and knowledge discovery handbook, Boston, MA: Springer, 2009, pp. 875-886.

[14] M. Zięba, J. M. Tomczak, M. Lubicz and J. .. Świątek, "Boosted SVM for extracting rules from imbalanced data in application to prediction of the post-operative life expectancy in the lung cancer patients," Applied soft computing, vol. 14 , pp. 99-108, 2014.

[15] S. García and F. Herrera, "Evolutionary undersampling for classification with imbalanced datasets: Proposals and taxonomy," Evolutionary computation, vol. 17, no. 3, pp. 275-306, 2009.

[16] M. M. Rahman and D. N. Davis, "Addressing the class imbalance problem in medical datasets," International Journal of Machine Learning and Computing, vol. 3, no. 2, p. 224, 2013.

[17] E. Š. M. R.-Š. Igor Kononenko, "Overcoming the myopia of inductive learning algorithms with RELIEFF," Applied Intelligence, 7(1), pp. 39-55, 1997.

[18] C. A. Pena-Reyes and M. Sipper, "A fuzzy-genetic approach to breast cancer diagnosis," Artificial intelligence in medicine, vol. 17, no. 2, pp. 131-155, 1999.

[19] M. Karabatak and M. C. Ince, "An expert system for detection of breast cancer based on association rules and neural network," Expert systems with Applications, vol. 36, no. 2, pp. 3465-3469, 2009.

[20] M. F. Akay, "Support vector machines combined with feature selection for breast cancer diagnosis," Expert systems with applications , vol. 36, no. 2, pp. 3240-3247, 2009.

[21] A. Osareh and B. Shadgar, "Machine learning techniques to diagnose breast cancer," in In Health Informatics and Bioinformatics (HIBIT), 2010 5th International Symposium on, 2010.

[22] K. J. Wang and A. M. Adrian, "Breast cancer classification using hybrid synthetic minority over-sampling technique and artificial immune recognition system algorithm," Int J Comput Sci Electron Eng (IJCSEE) , vol. 1, no. 3, pp. 408-412, 2013.

[23] H. Parvin, B. Minaei-Bidgoli and H. Alinejad-Rokny, "A new imbalanced learning and dictions tree method for breast cancer diagnosis," Journal of Bionanoscience, vol. 7, no. 6, pp. 673-678, 2013.

[24] B. Zheng, S. Won Yoon and S. S. Lam, "Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms," Expert Systems with Applications , vol. 41, no. 4, pp. 1476-1482, 2014.

[25] H. Asri, H. Mousannif, H. Al Moatassime and T. Noel, "Using machine learning algorithms for breast cancer risk prediction and diagnosis," in Procedia Computer Science , 2016.

[26] E. Yavuz, C. Eyupoglu, U. Sanver and R. Yazici, "An ensemble of neural networks for breast cancer diagnosis," in In Computer Science and Engineering (UBMK), 2017 International Conference on, 2017.

[27] T. Cai, H. He and W. Zhang, "Breast Cancer Diagnosis Using Imbalanced Learning and Ensemble Method," Applied and Computational Mathematics, vol. 7, no. 3, pp. 146-154, 2018.

[28] M. Robnik-Šikonja and I. Kononenko, "Theoretical and empirical analysis of ReliefF and RReliefF," Machine learning, vol. 53, no. 1-2, pp. 23-69, 2003.

[29] F. Rosenblatt, Principles of neurodynamics. perceptrons and the theory of brain mechanisms(No. VG-1196-G-8), CORNELL AERONAUTICAL LAB INC BUFFALO NY, 1961.

[30] L. Breiman, "Random forests," Machine learning, vol. 45, no. 1, pp. 5-32, 2001.

[31] J. R. Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers, 1993.

[32] Q. Gu, L. Zhu and Z. Cai, "Evaluation measures of the classification performance of imbalanced data sets," in International Symposium on Intelligence Computation and Applications, Berlin, 2009.

[33] "UCI machine learning repository," [Online]. Available: https://archive.ics.uci.edu/ml/index.php.

[34] W. H. Wolberg and O. Mangasarian, "Multisurface method of pattern separation for medical diagnosis applied to breast cytology," in Proceedings of the National Academy of Sciences, U.S.A., December 1990.

[35] W. Wolberg, W. Street and O. Mangasarian, "Machine learning techniques to diagnose breast cancer from fine-needle aspirates," Cancer Letters , vol. 77, pp. 163-171, 1994.