Journal of Advances in Computer Research
Quarterly pISSN: 2345-606x eISSN: 2345-6078
Sari Branch, Islamic Azad University, Sari, I.R.Iran
(Vol. 7, No. 4, November 2016), Pages: 109-124
www.jacr.iausari.ac.ir

Ovarian Cancer Classification Using Hybrid Synthetic Minority Over-Sampling Technique and Neural Network

Moshood A. Hambali ✉[1], **Morufat D. Gbolagade**[2]
1) Computer Science Dept., Federal University Wukari, Nigeria
2) Computer Science Dept., Al-Hikmah University, Ilorin, Nigeria
hambali@fuwukari.edu.ng; dammyconsult@gmail.com

Received: 2016/04/23; Accepted: 2016/06/25

Abstract

Every woman is at risk of ovarian cancer; about 90 percent of women who develop ovarian cancer are above 40 years of age, with the high number of ovarian cancers occurring at the age of 60 years and above. Early and correct diagnosis of ovarian cancer can allow proper treatment and as a result reduce the mortality rate. In this paper, we proposed a hybrid of Synthetic Minority Over-Sampling Technique (SMOTE) and Artificial Neural Network (ANN) to diagnose ovarian cancer from public available ovarian dataset. The dataset were firstly preprocessed using SMOTE before employing Neural Network for classification. This study shows that performance of Neural networks in the cancer classification is improved by employing SMOTE preprocessing algorithm to reduce the effect of data imbalance in the dataset. To justify the performance of the proposed approach, we compared our results with the standard neural network algorithms. The performance measurement evaluated was based on the accuracy, F-measure, Recall, ROC Area Margin Curve and Precision. The results showed that SMOTE + MLP (with above 96% accuracy) performed better than SMOTE + RBF and standard RBF and MLP.

Keywords: *Artificial Neural Network, RBF, SMOTE, MLP, Data Imbalance, Ovarian Cancer*

1. Introduction

All women are at peril of ovarian cancer, but older women are more prone to this kind of cancer than younger women. About 90 percent of women who get ovarian cancer are above 40 years of age, with the greatest number of ovarian cancers occurring in women of ages 60 years and above. Each year, almost 20,000 women in the United States get ovarian cancer and also, ovarian cancer is the eighth most common cancer and the fifth leading cause of cancer death [1]. Cancer is a disease which develops as a result of over growth of the body cells. Cancer is always named after the part of the body where it starts, even if it spreads to other body parts later.

Ovaries are reproductive glands found only in female species of mankind. The ovaries produce eggs (ova) for reproduction and also serve as the main sources of the female hormones - estrogen and progesterone. The ovaries consist of 3 main types of cells; each of these cells can develop into different types of tumor [2]:

Epithelial tumors start from the outer surface cells that cover the ovary. Most ovarian tumors are epithelial cell tumors. **Germ cell tumors** start from the cells that produce the eggs (ova) and **Stromal tumors** start from structural tissue cells that hold the ovary together and produce the female hormones estrogen and progesterone.

Cancerous epithelial tumors are usually referred to as *carcinomas*. About 85% to 90% of ovarian cancers are epithelial ovarian carcinomas. When someone has ovarian cancer, it usually means that the person has this type of cancer [2]. Timely and correct diagnosis of ovarian cancer can allow proper treatment and as a result reduce the mortality rate. Among machine learning methods, neural networks are about the most common methods used in medical diagnostics [3].

Data mining plays an important role in the medical field by predicting and detecting various diseases [4]. A major application of microarrays has been used for the study and diagnosis of cancer. Recognition of the signals that are symptoms for the disease phenotype and its progression requires the use of robust techniques. Cancer can be identified through the analysis of genetic data. The human genome contains more than 10 million single nucleotide polymorphisms which will be in charge of the difference that lies among human beings.

Artificial neural networks (ANNs) are widely used with applications in science and technology. ANNs are mathematical forms of the human neural design, representing its "learning" and "generalization" capabilities. With this reason, ANNs have its place in the area of artificial intelligence. ANNs are extensively employed in research due to its ability to model highly non-direct systems in which the relationship among the variables is undetermined or very complex.

However, microarray data are usually characterized with class imbalanced data set. This is a problem that is commonly found in the real world application that can cause severe negative effect on classification performance. Microarray datasets with class imbalance pose problems when less observed patterns are of higher relevance; since most of the data mining techniques tend to generalize the patterns observed over the majority data and ignore those observed over small portions of the data.

One of the best approaches to deal with the class imbalance problem is Synthetic Minority Over-Sampling Technique (SMOTE). In this technique, SMOTE generates minority class within the overlapping regions. SMOTE has been widely used to solve the problem of imbalanced dataset in many medical area, such as medical imaging intelligence [5] and prostate cancer staging [6].

In this paper, we proposed new method of hybrid SMOTE and Artificial Neural Network (ANN) to diagnose ovarian cancer from public available ovarian dataset. To justify the performance of the proposed approach, we compared our result with the standard neural network algorithms. The performance measurement evaluated was based on the accuracy, F-measure, Recall, ROC Area and Precision.

The rest of this paper is organized as follows. Section 2.0 presents problem statement. Section 3.0 reviews related works to this study. In section 4.0, we describe the methodology used. Section 5.0 discusses the evaluation metrics used in this work. Section 6.0 contains discussion of the results. Finally, conclusions are presented in section 7.0.

2. Problem Statement

The improvement in data acquisition capacity, low cost of data storage and development of database and data warehousing technology in recent years had led to the advent of high dimensional dataset. Many of these features are irrelevant, redundant and increase the search time and resulting in difficulty to correctly classify medical datasets with class imbalance. The problem is to predict the ovarian cancer classes from large amount of data without any bias among classes. This has been done by doing a comparative study of ANNs classification algorithm using feed-forward multi-layer perceptron (MLP) and radial basis functions neural networks (RBF) with oversampling technique SMOTE.

3. Related Works

Numerous works have been published on applying machine learning (ML) techniques for classification and predictive analysis. In this section, a number of papers are reviewed in relation to data mining contribution in detection and diagnosis of cancer diseases.

3.1 Synthetic Minority Over-Sampling Technique

There are two main approaches for solving data imbalanced; first is to pre-process data by under-sampling the majority instance, and second is to over-sampling the minority instance. SMOTE is recognized as a famous over-sampling method. In SMOTE, the positive class is over-sampled by creating synthetic instances in the decision regions formed by the instance and its *k*-nearest neighbors [7]. SMOTE employs different ways of generating sample in continuous and categorical features. Euclidean distance is evaluated for continuous samples generating and Value Distance Metric is used for nominal features. Applying SMOTE technique in data pre-processing can lead to a better generalization for the classifiers.

Several researchers have applied SMOTE technique to deal with imbalanced microarray data problem. Gao et al., [7] combined SMOTE and particle swarm optimization (PSO) with RBF as classifier to predict the survivability of patients who undergo breast cancer surgery. They proved that SMOTE is effective in increasing the significance of the positive class in the decision region and concluded that their proposed method offers a very competitive solution to other existing methods that deal with imbalanced class problem. Chandana, Leung and Trpkov [6] examined the performance of SMOTE and combination of genetic algorithm (GA) and rough set (RS) to predict the stages of prostate cancer. They concluded that under-sampling and rough sets based features were acknowledged to be the most useful in improving overall performance of their system. Wang, Makond and Chen [8] proposed SMOTE+PSO+C5 to enhance the efficacy of classification of 5-year survivability of breast cancer patients with imbalanced dataset. The result shows that this proposed method has the highest performance on the dataset used. They concluded that standard classifier exclusively cannot improve the classification performance.

3.2 Artificial Neural Networks

There are several works concerning the application of ANNs in medical diagnosis. The concept was first mentioned in 1988 in the pioneering work of Szolovits et al. [9] and since then many papers have been published. The general application of ANNs in medical diagnosis has previously been reviewed by [10]. For instance, ANNs have been used in the diagnosis of following diseases: colorectal cancer [11], multiple sclerosis lesions [12], colon cancer [13], pancreatic disease [14], gynaecological diseases [15], and early diabetes [16]. ANNs have also been applied in the data analysis and diagnostic classification of patients with uninvestigated dyspepsia in gastroenterology [17] and in the biomarkers searching [18]. Sansanee, Siripen, Sutasinee and Nipon [19] trained

Neural Network employing back propagation and achieved an accuracy level on the test data of approximately 94% on breast cancer dataset.

Feng and Lip [20] reported an approach based on a novel radial basis function (RBF) neural network that successfully classified the lymphoma data set with 100% accuracy using only 9 genes. This approach also obtained 100% accuracy in the SRBCT data set and the ovarian data with only 8 genes and 4 genes, respectively. Their method includes two steps. In the first step, they select some genes with the greatest discriminative ability in the training data. In the second step, the selected genes were used to train their RBF neural network and subsequently use the trained network to classify the testing data. Meenakshi, et al., [21], compare the performance analysis of two ANN algorithms (MLP and RBF) to identify the breast cancer prognosis. The results of their research work show that the classification accuracy of MLP classifier to be 79.20% and RBF classifier to be 77.78%, which confirmed that MLP network produce more specific, accurate results compared to RBF. Gursharan and Kulwinder [22] applied supervised multilayer perceptron decision tree to classify lung cancer based datasets. They compared the results of their proposed method with other techniques to check the effectiveness of the proposed method. Their simulation results shows that proposed technique achieved 100% accuracy to classify cancer data sets which is more as compared to other techniques.TP rate, ROC and Precision are highest for proposed method amongst other methods.

The vital issues in Multi-Layer Perceptrons (MLP) design include specifications of the number of hidden layers and the number of units in these layers. The number of input and output units is defined by the problem, determining the number of hidden layer(s) and number of nodes in the hidden layer(s) is a critical decision in the design of neural networks. Too many hidden neurons will lead to many trainable weights, which can result into a neural network becoming erratic and unreliable. On the other hand, few hidden neurons limit the learning ability of a neural network and deteriorate its approximation performance [23], [24]. However, there is no clear rule for determining the number of neurons in hidden layer(s). The usual practice is by trial and error which may not yield an optimal network design and the process is also time consuming [24]. A network with one hidden layer is sufficient to solve most tasks. The universal neural network approximation theorem states that every continuous function that maps intervals of real numbers to some output interval of real numbers can be approximated arbitrarily closely by a multi-layer perception with just one hidden layer. There is no theoretical reason proved to use more than two hidden layers. It has also been seen like that for the vast majority of principal problems. Problems that require two hidden layers are very rarely encountered in real life circumstances. *Rawtani, Rana and Tiwari* [25] suggested that choosing the number of hidden nodes should be one or more than the training points on the curve. The number of neurons in the hidden layers may be equivalent to the control points.

4. Methodology

In the medical research, Cancer research is one of the leading research areas. Exact predictions of various tumor types have great value in granting better treatment and reduce harmfulness on the patients. In earlier days, cancer identification had always been morphologically and clinically based. These methods of cancer classification are boomed to have several demerits in their diagnostic capability. With the advent of data mining techniques and machine learning, thousands of genes can be correctly classified without much stress. In this paper, the proposed method is in two steps. The first step is to use SMOTE to reduce the effect of data imbalance in the dataset. The second step involves classifying using ANN algorithms (RBF and MLP), and then comparing the results of the experiment with standard RBF and MLP algorithms on the dataset without data imbalance algorithm. WEKA Explore (Weka 3.6) is used to implement these algorithms.

4.1 Synthetic Minority Oversampling Technique - SMOTE

SMOTE (Synthetic Minority Oversampling Technique) was proposed to reduce the effect of having few instances of the minority class in the data set. SMOTE adopts an over-sampling approach in which the minority class is over-sampled by creating synthetic examples rather than by over-sampling with replacement [26]. SMOTE generates synthetic instances of the minority class by working on the “feature space” rather than the “data space”. By synthetically producing more instances of the minority class, the inductive learners, such as decision trees or rule-learners algorithm, are able to strengthen their decision regions for the minority class. The process of balancing dataset with nominal (or discrete) and continuous attributes are different in SMOTE. The minority class is over-sampled by considering each minority class sample and introducing synthetic examples along the line of segments joining the k minority class nearest neighbors. Depending upon the amount of over-sampling required, neighbors from the k nearest neighbors are randomly selected [27]. The new synthetic minority samples are created as follows:

For the continuous features

- Determine the difference between a feature vector (minority class sample) and one of its k nearest neighbors (minority class samples).
- Multiply this difference by a random number between 0 and 1.
- Add this difference to the feature value of the original feature vector, thus creating a new feature vector

For the nominal features

- Take majority vote between the feature vector under consideration and its k nearest neighbors for the nominal feature value. In the case of a tie, choose at random.
- Assign that value to the new synthetic minority class sample.

4.2 Neural Network

The construction of the neural network involves three different layers with feed forward architecture. This is the most widely used network architecture today. The input layer of this network is a set of neurons, which accepts the input data (feature vectors). The input units (neurons) are fully linked to the hidden layer with the set of neurons. These hidden units (neurons) are also fully linked to the output layer. The output layer produce the response of neural network to the activation pattern applied on the input layer. The data input into a neural network propagate layer-by-layer from input layer to output layer through (none) one or more hidden layers.

Here the average number of control points (3) has been assigned as the number of hidden neurons. In this paper, the following parameters were set for MLP before the experiment commenced:

Hidden Layers = 3

Learning Rate = 0.3
Training Time = 200
Validation Threshold = 20
Momentum = 0.2

4.3 RBF Neural Network

An RBF neural network has three layers as in other neural networks [28]. The first layer is an input layer; the second layer is a hidden layer that includes some radial basis functions, also known as hidden kernels; and the third layer is the output layer.
An RBF neural network can be considered as a mapping of input domain *X* onto the output domain *Y*.

$$y_m(\ddot{x}) = \sum_{i=1}^{n} w_{mi} G\left(\| \ddot{x} - \ddot{t}_i \|\right) + b_m ;$$
$$i = 1, 2, 3, \ldots, N ; \qquad (1)$$
$$m = 1, 2, 3, \ldots, M .$$

Here ‖·‖ stands for the Euclidean norm. M is the number of outputs. N is the number of hidden kernels. $y_m(\ddot{x})$ is output m corresponding to the input $\ddot{x}$. $\ddot{t}_i$ is the center of kernel i. w_{mi} is the weight between kernel i and output m. b_m is the bias on output m. $G(\| \ddot{x} - \ddot{t}_i \|)$ is the kernel function. The most commonly used kernel function for RBF neural networks is Gaussian kernel function as in equation 2:

$$G\left(\| \ddot{x} - \ddot{t}_i \|\right) = exp\ -\left(\frac{\| \ddot{x} - \ddot{t}_i \|^2}{2\sigma_i^2}\right) \qquad (2)$$

Where σ_i is the radius of the kernel i.
The main steps to construct an RBF neural network include:
i. Determine the positions of all the kernels $\ddot{t}_i$,
ii. Determine the radius of each kernel, and
iii. Calculate the weights between the kernels and the output nodes.

4.4 Data Source

To compare the classification performance, we apply these classifiers to the patients with ovarian cancer disease. The microarray datasets are available in the Gene Expression Omnibus (GEO) database. The summary information on microarray datasets is given in Table 1.

Table 1: Summarize information on Dataset

Dataset name	Number of classes	Number of features	Number of samples
Ovarian Cancer	2	15, 155	253

5. Evaluation Metrics

A classifier is evaluated by a confusion matrix as illustrated in table1. The columns indicate the predicted class and the rows show the actual class [29]. In the confusion matrix, True Negative (TN) is the number of negative samples correctly classified, False Positive (FP) is the number of negative samples incorrectly classified as positive, False Negative (FN) is the number of positive samples incorrectly classified as negative and True Positive (TP) is the number of positive samples correctly classified.

Table 2: Confusion Matrix

	Predicted Negative	**Predicted Positive**
Actual Negative	TN	FP
Actual Positive	FN	TP

Overall accuracy is defined in equation 3

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

Overall accuracy is not a suitable parameter for performance evaluation when the data is imbalanced. The nature of some problems requires a fairly high rate of correct detection in the minority class and allows for a small error rate in the majority class while simple overall accuracy is clearly not appropriate in such cases [30]. Actually, overall accuracy is biased over the majority class which contains more samples and this measure does not represent the minority class accuracy.

From the confusion matrix in table 2, the expressions for FP rate, Recall and Precision are derived and are presented in equations 4, 5 and 6.

$$FP\ Rate = \frac{FP}{(TN + FP)} \quad (4)$$

$$TP\ Rate = Recall = \frac{TP}{(TP + FN)} \quad (5)$$

$$Precision = \frac{TP}{(TP + FP)} \quad (6)$$

The main goal of learning from imbalanced datasets is to improve the recall without affecting the precision. However, recall and precision goals can be often conflicting because when increasing the true positive for the minority class, the number of false positives can also be increased; and this will reduce the precision.

6. Results and Discussion

In this study, the results of proposed hybrid of SMOTE and ANN to diagnose ovarian cancer were presented in this section. These results were compared with the standard neural network algorithms.

6.1 Experiment Results

The experiment was performed using Weka Explorer. Firstly, the dataset was uploaded into explorer, SMOTE algorithm was used to reduce the effect of data imbalance (that is, Oversampling the minority class) and second step involves the classification algorithms using MLP / RBF. Since the performance of classifiers will be overestimated when using the Leave-one-out method, we verified our experiment using a random average of 10-fold method (10-Fold cross validation). In table 3 the distribution of dataset before SMOTE and after SMOTE step are presented and visualized plot distribution of Dataset in figure 1 and 2.

Table3: Results from running our proposed method on Ovarian Cancer dataset

Steps	Features	Samples
Initial State	15,155	253
After SMOTE	15,155	344

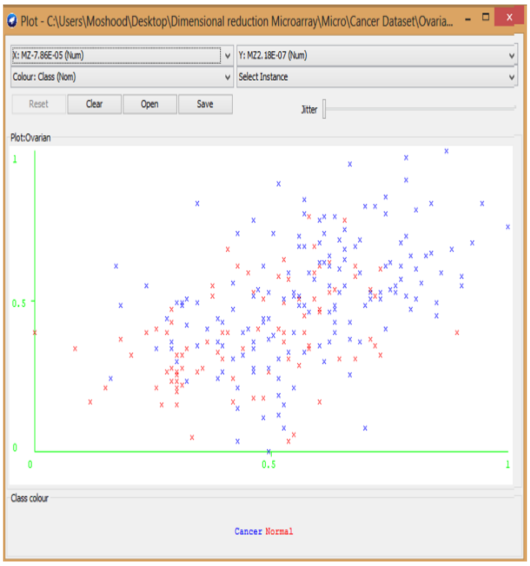

Fig. 1: Initial Ovarian Dataset

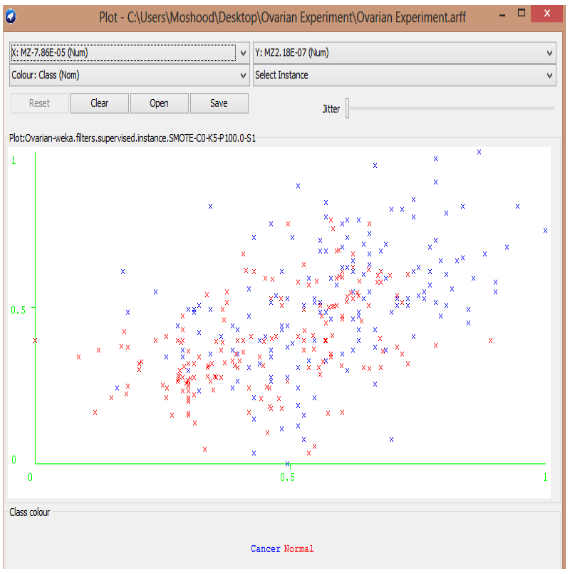

Fig. 2: Dataset After SMOTE

ROC distribution of dataset before and after applying SMOTE on dataset is shown in fig. 3 and 4.

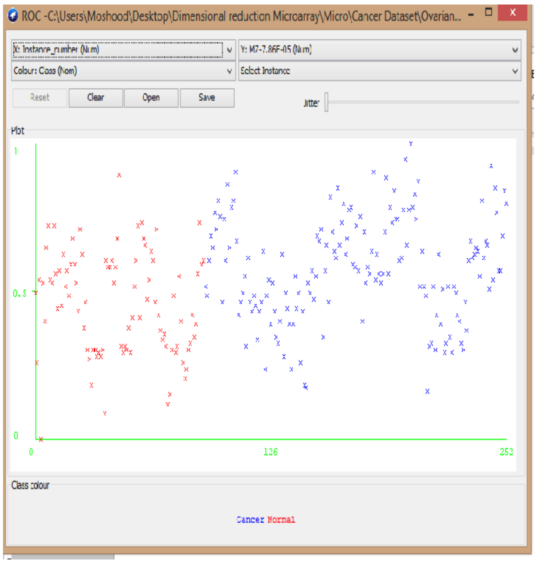

Fig. 3: ROC for Initial Dataset

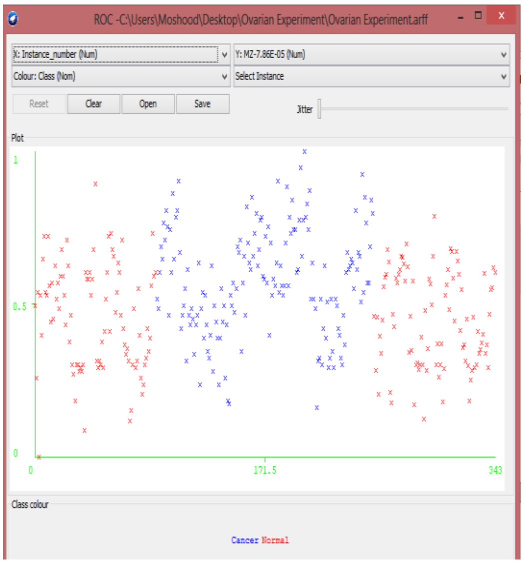

Fig. 4: ROC for Dataset after SMOTE

For ease of visualization, we presented the results in Table 4, 5 & 6 and Fig. 5 to 16. Comparisons of these two proposed classifiers were shown with standard RBF and MLP Classifiers.

Table 4: Performance Metrics for Classifiers

	Correct Classification (%)	Correctly Classified Instances	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
MLP	95.3	241	0.953	0.06	0.952	0.953	0.952	0.991
RBF	83.4	211	0.834	0.151	0.848	0.834	0.837	0.867
SMOTE + MLP	96.8	333	0.968	0.033	0.968	0.968	0.968	0.988
SMOTE + RBF	88.4	304	0.884	0.118	0.884	0.884	0.884	0.921

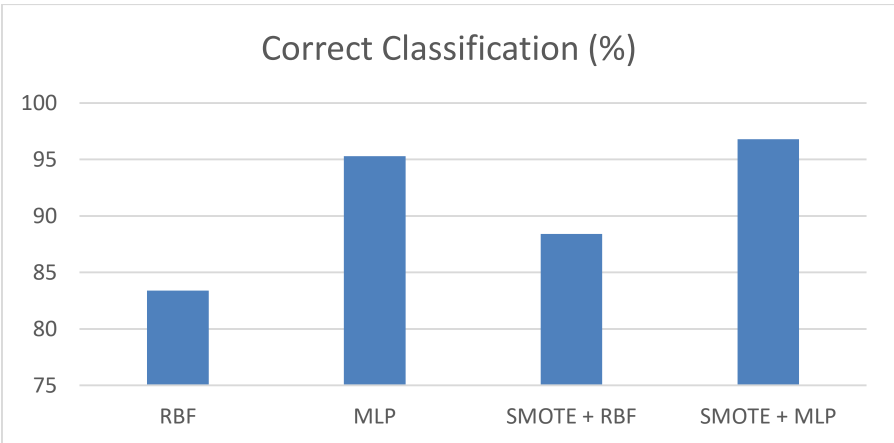

Fig. 5: Classification Accuracy

The proposed methods used both SMOTE and RBF/MLP in order to enjoy the benefits of both methods simultaneously. After running SMOTE resampling method and Classifiers (RBF and MLP) are applied and the accuracy in this condition has increased to above 88 percent (for RBF) and 96 percent (for MLP). This increase is due to the use of SMOTE resampling right before the classifiers to reduce the effect of data imbalance.

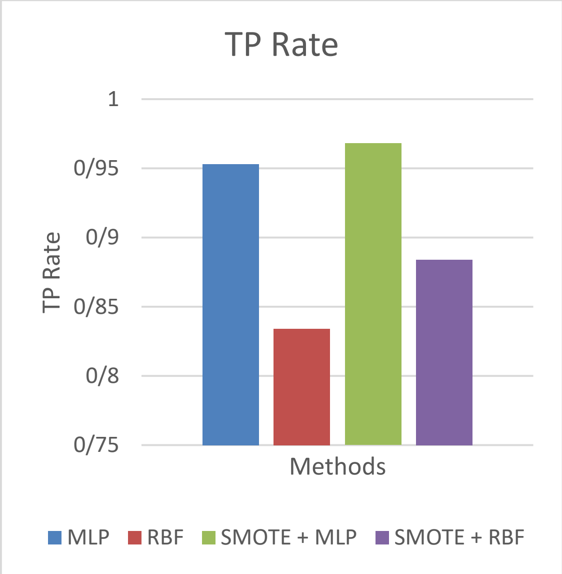

Fig. 6: TP Rate

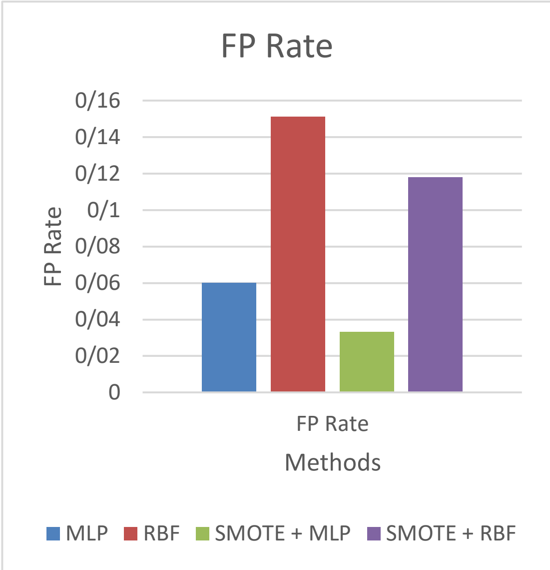

Fig. 7: FP Rate

False Positive (FP) Rate parameter is evaluated for different methods. If FP Rate is high, we can infer that the performance of classification is low because the number of incorrectly classified samples in the minority class is on the upsurge. Reversely, if FP Rate is low, we can construe that the performance of classification is high because the number of incorrectly classified samples for the minority class is on the decline. From fig. 7, it can be observed that SMOTE + MLP has a very low FP rate and RBF only has high FP rate. That is, SMOTE + MLP has high performance classification compared to other methods.

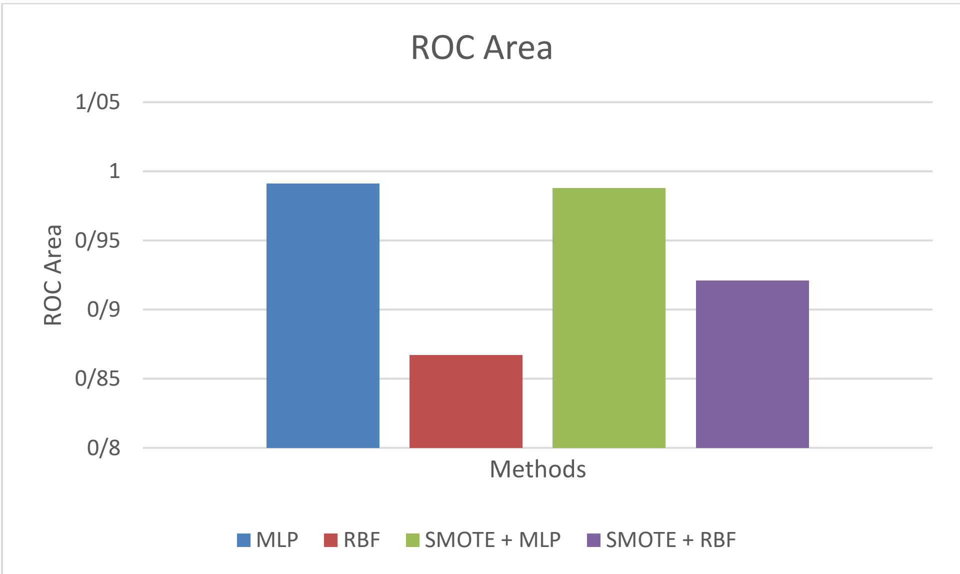

Fig. 8: ROC Area

ROC curves represent the trade-off between values of TP and FP. ROC of methods employed were shown in fig. 8.

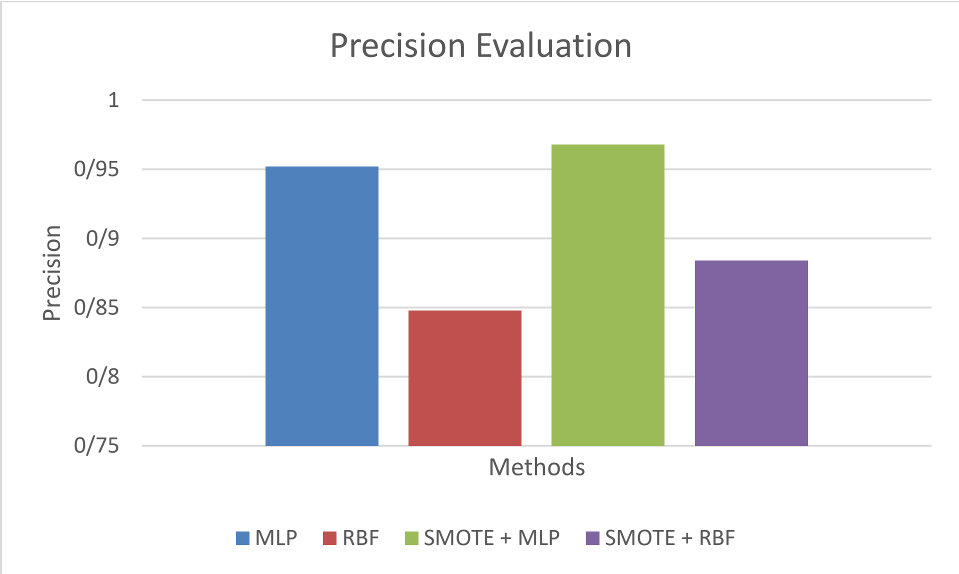

Fig 9: Precision Evaluation

From fig. 9, Precision parameter is calculated for different methods. When only MLP is applied on dataset, the Precision is 0.95. But when SMOTE was applied before MLP, the precision increased with 0.02. With SMOTE on RBF the precision increase from below 0.85 to above 0.88.

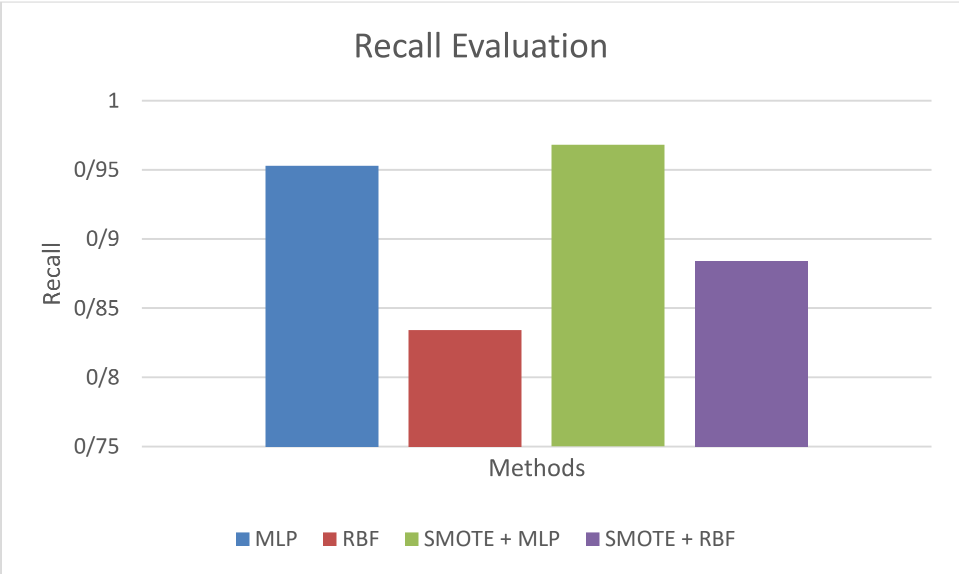

Fig 10: Recall Evaluation

Recall parameter is calculated for different methods. When only MLP is used on the dataset for classification, the recall evaluation is 0.95. After the application of SMOTE, we observed that there was an increase of 0.015 in recall evaluation. This increment shows that the effect of SMOTE on the dataset and helps to increase the performance of classification. The same thing with RBF, which increased from 0.834 to 0.884 after applying SMOTE before employed RBF classification algorithm.

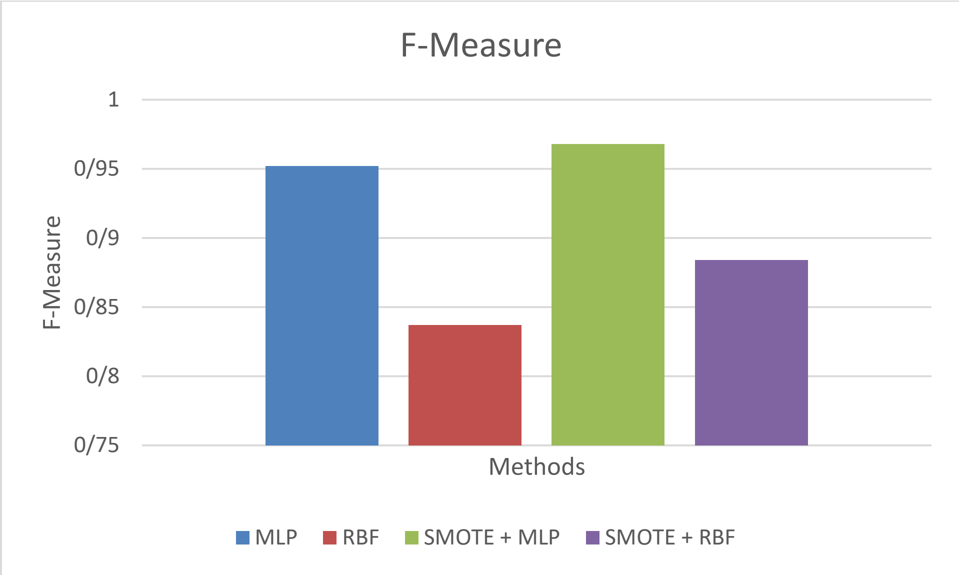

Fig. 11: F-Measure Evaluation

The *F-value* represents the trade-off among different values of TP, FP, and FN**.** SMOTE + MLP has very high F-value while RBF has low value.

Table 5: Classification Error

	RBF	**MLP**	**SMOTE + RBF**	**SMOTE+ MLP**
Mean Absolute Error	0.22	0.0746	0.1736	0.0634
Root Mean Squared Error	0.366	0.1954	0.3156	0.1749
Relative Absolute Error (%)	47.7	16.19	34.84	12.72

From Table 5, we recorded low Classification errors in SMOTE + MLP and very high classification errors in RBF only.

Fig. 12 – 15 show margin curve for the two proposed classifiers and the standard RBF and MLP

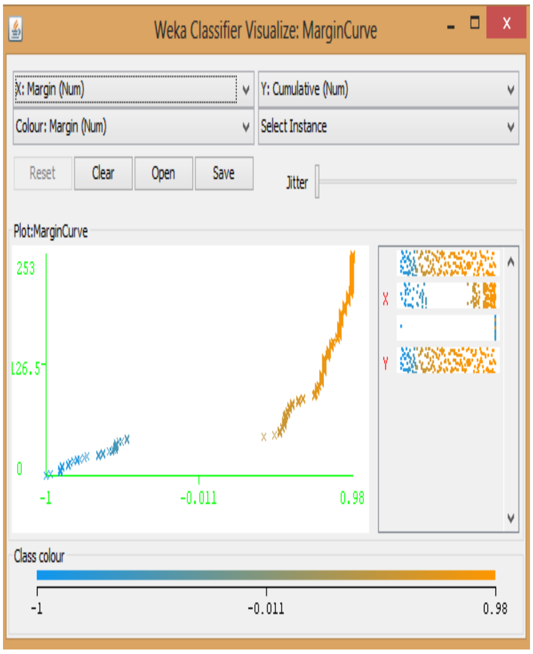

Fig. 12: Visualize Margin Curve RBF

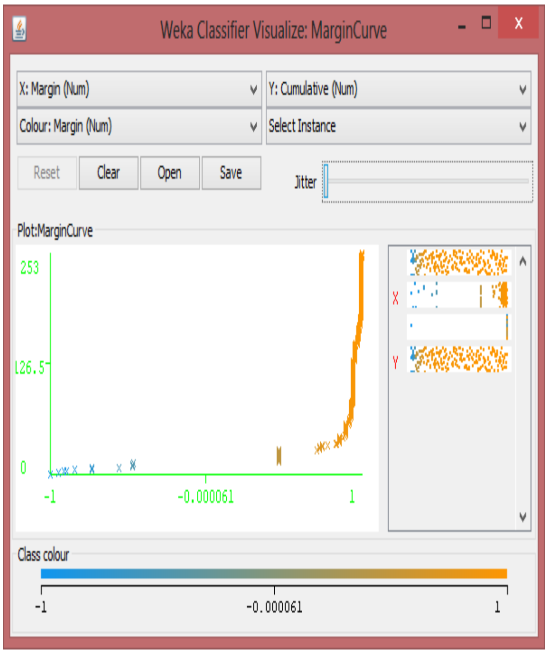

Fig. 13: Visualize Margin Curve MLP

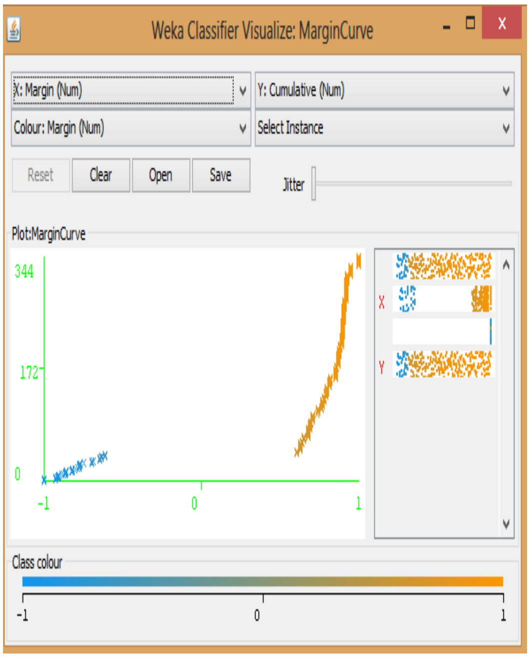

Fig. 14: Visualize Margin Curve SMOTE+RBF

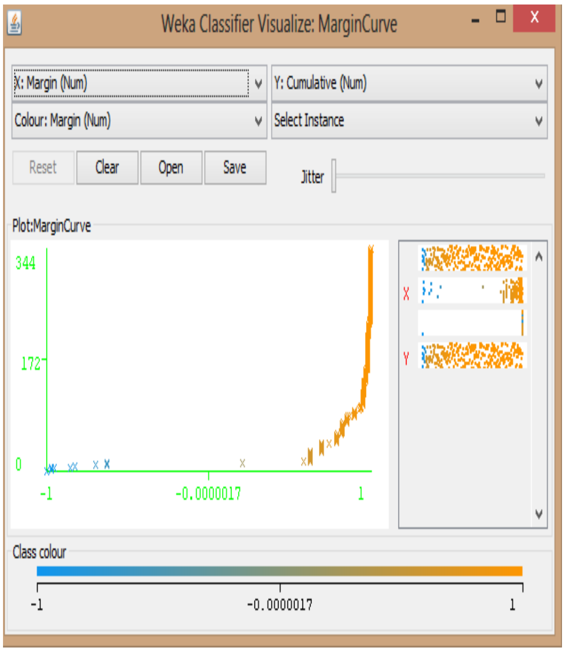

Fig. 15: Visualize Margin Curve SMOTE+MLP

The margin curves of the proposed algorithms and standard RBF/MLP algorithms are shown in fig. 12 - 15 respectively. The margin curve prints the cumulative frequency of the two actual class probabilities, if it is predicted to be positive with probability *p*, the margin is *p-1*. The negative values denote classification errors, meaning that the dominant class is not the correct one.

The fig. 15 depicts that the majority of instances are correctly classified by SMOTE + MLP model, since they are centralized in the area of probability one (the right part of the graph). On the other hand, the classified instances of RBF and MLP algorithms shown in fig. 12 and 13 are not concentrated in that area, revealing a significant deviation.

7. Conclusion

In this research work, we were able to minimize the bias inherent in the learning procedure due to the class data imbalance in dataset, and upsurge the sampling weights for the minority class. Many research works on imbalanced datasets have generally established that unequal class distribution in the dataset, the results always tend to be biased towards the majority class. The reasons for poor performance of the existing classification algorithms on imbalanced data sets are due to the following: 1.They are accuracy driven i.e., their goal is to minimize the overall error to which the minority class contributed insignificantly. 2. They assume that datasets are equally distributed for all the classes. 3. They also assume that the different classes generated the same cost errors. With unbalanced data sets, data mining learning algorithms produce degenerated models that do not take into account the minority class as most data mining algorithms assume balanced data set.

In reference to the results of the proposed methods, we obtained increase in classification accuracy of above 88 percent (for SMOTE +RBF) and 96 percent (for SMOTE+MLP) and lesser error rate in the proposed methods compared with standard MLP and RBF algorithms. This is due to effect of reducing data imbalance in the dataset using SMOTE algorithm.

Also, we showed that SMOTE + MLP performed better than SMOTE + RBF and standard RBF and MLP in all performance metrics we used. Finally, our results are also in line with [21] which showed that MLP network produced more specific, accurate results compared to RBF.

References

[1] AMS, "Ovarian Cancer," American Cancer Society Publication, 2015. www.cancer.org. Retrieved: 02-02-2016.

[2] U.S. Department of Health and Human Services, "Ovarian Cancer: Inside Knowledge, Get the Facts about Gynecologic Cancer," Centers for Disease Control and Prevention (CDC) Publication, 2012.

[3] J. Piecha, "The Neural Network Selection for a Medical Diagnostic System using an Artificial Data Set," Journal of Computing and Information Technology, 9(2), pp.123-132, 2001.

[4] F. M. Ba-Alwi and M. Hintaya, "Comparative Study for Analysis the Prognostic in Hepatitis Data: Data Mining Approach," International Journal of Scientific & Engineering Research, 4(8), 2013.

[5] W. Juanjuan, X. Mantao, W. Hui and W. Z. Jiwu, "Classification of Imbalanced Data by using the SMOTE Algorithm and Locally Linear Embedding," In Int. Conf. on Signal Processing Proceedings, 4129201, 2007.

[6] S. Chandana, H. Leung and K. Trpkov, "Staging of Prostate Cancer using Automatic Feature Selection, Sampling and Dempster-Shafer Fusion," Cancer Informatics,7, pp. 57-73, 2009.

[7] M. GAO, X. Hong, S. Chen and C. J. Harris, "On Combination of SMOTE and Particle Swarm Optimization Based Radial Basis Function Classifier for Imbalanced Problems," In Proceedings of the Int. Joint Conf. on Neural Networks, 6033353, pp. 1146-1153, 2011.

[8] K. J. Wang, B. Makond and K. H. Chen, “A Combination Algorithm for 5-year Survivability of Breast Cancer Patient,” In Proceeding of Translational Bioinformatics Conf., 2012.

[9] P. Szolovits, R. S. Patil and W. Schwartz, Artificial Intelligence in Medical Diagnosis. Ann Intern Med. 108, pp. 80–87, 1988.

[10] E. Alkim, E. Gürbüz, and E. Kiliç, “A Fast and Adaptive Automated Disease Diagnosis Method with an Innovative Neural Network Model,” Neural Networks, 33, pp. 88–96, 2012.

[11] L. Spelt, B. Andersson, J. Nilsson and R. Andersson, “Prognostic Models for Outcome Following Liver Resection for Colorectal Cancer Metastases: A Systematic Review,” Eur Journal Surg Oncol. 38, pp. 16–24, 2012.

[12] D. Mortazavi, A. Kouzani and H. Soltanian-Zadeh, “Segmentation of Multiple Sclerosis Lesions in MR Images: A Review,” Neuroradiology. 54, pp. 299–320, 2012.

[13] F. Ahmed, “Artificial Neural Networks for Diagnosis and Survival Prediction in Colon Cancer,” Mol Cancer, 4 (29), 2005.

[14] A. Bartosch-Härlid, B. Andersson, U. Aho, J. Nilsson and R. Andersson, “Artificial Neural Networks in Pancreatic Disease,” Br J Surg. 95, pp. 817–826, 2008.

[15] C. Siristatidis, C. Chrelias, A. Pouliakis, E. Katsimanis and D. Kassanos, “Artificial Neural Networks in Gyneacological Diseases: Current and Potential Future Applications,” Med. Sci. Monit. 16, pp. 231–236, 2010.

[16] Shankaracharya, D. Odedra, S. Samanta and A. Vidyarthi, “Computational Intelligence in Early Diabetes Diagnosis: A Review,” Rev Diabet Stud. 7, pp. 252–262, 2010.

[17] F. Pace and V. Savarino, “The Use of Artificial Neural Network in Gastroenterology: The Experience of the First 10 Years,” Eur. J. Gastroenterol Hepatol. 19, pp. 1043–1045, 2007.

[18] B. Bradley, “Finding Biomarkers is Getting Easier,” Ecotoxicology. 21, pp 631–636, 2012.

[19] S. Auephanwiriyakul, S. Attrapadung, S. Thovutikul and N. Theera-Umpon “Breast Abnormality Detection in Mammograms using Fuzzy Inference System,” IEEE International Conference on Fuzzy Systems, Reno, Nevada, U.S.A., May 2005, pp.155 -160, 2005.

[20] F. Chu and L. Wang, “Applying RBF Neural Networks to Cancer Classification Based on Gene Expressions,” International Joint Conference on Neural Networks Sheraton Vancouver Wall Centre Hotel, Vancouver, BC, Canada July 16-21, 2006, PP. 1930 -1934, 2006.

[21] C. Meenakshi, M. Govindarajan and A. M. Sameeullah, “Detection of Breast Cancer Using MLP and RBF Classifiers,” IMS Manthan, 5 (1), pp. 35 – 38, 2010.

[22] G. Singh and K. S. Mann, “Supervised Multilayer Perceptron Network Based Decision Tree Algorithm for Cancer Classification,” International Journal of Enhanced Research in Science Technology & Engineering, 4 (6), pp. 106-110, 2015.

[23] F. Amato, et al., “Artificial Neural Networks in Medical Diagnosis,” Journal of Applied Biomedicine. 11, pp. 47–58, 2013.

[24] M. D. Olagoke, A. A. Ayeni and M. A. Hambali, “Short Term Load Forecasting Using Neural Network and Genetic Algorithm,” International Journal of Applied Information Systems, 10(4): pp. 22 – 28, 2016.

[25] L. Rawtani, J. L. Rana and A. K. Tiwari, “Number Hidden Nodes for Shape Preserving ANN Representation of a Curve,” IJCNN '01. International Joint Conference on Neural Networks, Washington, DC, pp. 138-143, 2001.

[26] M. Naseriparsa and M. M. R. Kashani, “Combination of PCA with SMOTE Resampling to Boost the Prediction Rate in Lung Cancer Dataset," International Journal of Computer Applications, 77(3), 2013.

[27] V. Ganganwar, "An Overview of Classification Algorithms for Imbalanced," International Journal of Emerging Technology and Advanced Engineering, 2 (4), pp. 42 - 47, 2012.

[29] H. He and E. Garcia, "Learning from Imbalanced Data," IEEE Transactions on Knowledge and Data Engineering, 21 (9), pp. 1263-1284, 2009.

[30] N.V. Chawla, "Data Mining for Imbalanced Datasets: An Overview," Data Mining and Knowledge Discovery Handbook, Springer, pp. 853-867, 2005.