Stacked Classifier Model with Prior Resampling for Lung Nodule Rating Prediction

Vinay Kumar, Ashok Rao and G. Hemanthakumar

Abstract In this work, we are proposing a new machine learning strategy for classification task for imbalanced data. We are using lung image data by Lung Image Database Consortium (LIDC), since LIDC data is a better example for imbalanced dataset. In this work we are using sufficiently large dataset which contains 4,532 nodules extracted from CT images. Later we consider 55 low level nodule image features and radiologists ratings for experiments. This work is being dealt in two stages. (1) data level learning and (2) algorithm level learning. In first stage, we are balancing the dataset prior to classification process. We are using resampling approach for this task. In second stage, we are using ensemble of classifiers to predict lung nodule rating. We are using wide range of classifier models for constructing an ensemble. We use Bagged Decision Tree, naïve Bayes, Boosted Decision Trees, and Support Vector Machine (SVM) in a classifier library. Stacking algorithm is used to combine the different classifier models in library to construct higher level ensemble. We are evaluating the performance of our model on five metrics: Accuracy, precision, recall, F-score and Kappa statistics. Results show that our method yields much improved scores as we are refining at both, data level and algorithm level.

Keywords Stacking · Ensemble of classifier · Resampling · Kappa statistics

V. Kumar (✉) · G. Hemanthakumar
DoS in Computer Science, University of Mysore, Mysore, India
e-mail: gotovinni@gmail.com

G. Hemanthakumar
e-mail: ghk.2007@yahoo.com

A. Rao
Freelance Academician, 165, 11th main, S.Puram, Mysore, India
e-mail: ashokrao.mys@gmail.com

P. P. Swamy and D. S. Guru (eds.), *Multimedia Processing, Communication and Computing Applications*, Lecture Notes in Electrical Engineering 213,
DOI: 10.1007/978-81-322-1143-3_11, © Springer India 2013

1 Introduction

Lung cancer is one of the major medical challenges that the world is facing today. Recent survey shows that mortality rate of people dying because of lung cancer tend to increase year by year. Computer Aided Diagnostics (CAD) is one such system in medical field which are built using computer programs to effectively assist in diagnosis of the diseases. Many such systems are built using image processing as well as pattern recognition techniques. Even though there are a good numbers of CAD systems that are available in the field, still there is lack of intelligent systems which can adopt themselves to change with respect to variation in input environment. These changes refer to imbalance in input data, uncertainty in domain expert prediction and problem in deciding marginal cases. Hence there is still lot of research required to develop such intelligence into systems and schemes. Machine learning technique is one such approach to solve such issues and recently many efforts have been carried out using various machine learning techniques to address above. We have discussed some of state-of-art work in this domain which has been carried out recently in the Sect. 2.

2 Literature Review

In this paper, we brief the literature in two sections. In the first section we discus about some work on CT scan image feature extraction and classification in the perspective of image processing and pattern recognition. In second section we present recent works on machine learning and ensemble of classifier approaches for classification problems. Ekrain et al. [1] investigated several approaches to combine delineated boundaries and ratings from multiple observer and they have used p-map analysis with union, intersection and threshold probability to combine the boundary reading and claimed that threshold probability approach provides good level of agreement. Ebadollahi et al. [2] proposed a framework that uses semantic methods to describe visual abnormalities and exchange knowledge with medical domain.

Nakumura et al. [3] worked on simulating the radiologists perception of diagnostic characteristic rating such as shape, margin, irregularity, Spiculation, Lobulation, texture etc. on a scale of 1–6 and they extracted various statistical and geometric image features including fourier and radiant gradient indices and correlated these features with the radiologists ratings. Significant work towards designing panel of expert machine learning classifier which automates the radiologist work of predicting nodule ratings is done by Dmitriy and Raicu [4]. They have proposed active decorate, a new meta-learning strategy for ensemble of classifier domain. Oza and Tumer [5], presented a survey on applications of ensemble methods covering different fields such as remote sensing, person recognition, one versus all recognition in medicine. In their work they have summarized the most frequently used classifier ensemble methods including averaging, bagging, boosting and order statistics classifiers. They have made a statement that each ensemble method has different properties that make it better suited to particular types of classifiers and applications.

Reid [6] has presented a survey work on several ensemble methods that can accommodate different classifiers for base models types. He has given useful review on heterogeneous ensemble methods with supporting theoretical motivation, empirical results and relationship to other techniques. Kuncheva et al. [7] have mentioned searching for a best classifier is an ill- posed problem because there is no one classifier that is best for all types of data. They have used variety of machine learning techniques to compare the performance of classifier ensembles for fMRI data analysis. Caurana [8] has identified that ensemble method can optimize the performance of the model for classification task. He has experimented with seven test problems and ten performances metric and claimed that ensemble techniques outperformed in all the scenarios.

Datta and Datta el al. [9] have presented their work on adaptive optimal ensemble classifier via bagging and rank aggregation with applications to high dimensional data. In their work they have considered three norm data and simulated microarray dataset. Based on their observation on obtained experimental results they have claimed ensemble classifier performs at the level of best individual classifier or better than individual classifier. They have concluded that for a complex high-dimensional datasets it is wise to consider a number of classification algorithms combined with dimension reduction techniques rather than a fixed standard algorithm. Dzeroski and Zenko et al. [10] have empirically evaluated several state-of-art methods for construction of ensembles of classifiers with stacking and they claimed that stacking method performs at best, comparable to selecting the best classifier from the ensemble by cross validation [5, 11]. Ting and Witten [12] recommended Multi-response Regression (MLR) as suitable for meta-level learning and showed other learning algorithms are not up to the mark as compared to MLR. In this work we are using linear regression as a meta-learner in our stacking model.

3 LIDC Dataset

Lung Image Database Consortium (LIDC) provides lung CT image data which is publically available through National Cancer Institute’s Imaging Archive (web site—http://ncia.nci.nih.gov) [13]. Dataset consists of image data, radiologist’s nodule outline details and radiologist subjective characteristic ratings. The LIDC dataset currently contains complete thoracic CT scans of 399 patients acquired over different periods of time. LIDC data download comes with DICOM image and the nodules information in the XML file. This has information regarding the spatial location information about three types of lesions, they are nodules <3 mm; nodules >3 mm and non-nodules >3 mm in maximum diameter as marked by

Table 1 Overview of the LIDC data subset considered

LIDC data subset	
Number of cases considered	124
Number of instances	14,956
Number of nodules	4,532

panel of 4 expert radiologists. For any lesion greater than 3 mm in diameter XML file contains spatial coordinates of the pixel of nodule outline. Since the number of radiologist in LIDC panel is 4 it is obvious that each nodule >3 mm has 4 nodule outlines. Moreover, any radiologist who identifies the nodule >3 mm also provides subjective ratings for 9 nodule characteristics: Lobulation, internal structure, calcification, subtlety, spiculation, margin, sphericity, texture and malignancy.

LIDC data collection process is in two fold, blinded and unblinded reading session and LIDC did not impose any forced consensus on radiologists, all the lesions indicated by the radiologists at the conclusion of the unblinded reading sessions are recorded and available to the public. With this no consensus on radiologist, lesion >3 mm is marked by a single a radiologist, by two radiologists, by three radiologists or by all four radiologists. The overview of the LIDC data subset we have used in this work has been shown in Table 1.

In our earlier work on lung images [14], we have considered 4,532 nodules that were extracted from LIDC lung image dataset. In this work our objective is to identify the significance of stacking ensemble method on large dataset. Hence we have collected sufficiently large dataset from LIDC CT images. As in literature [4] we are not only concentrating on those nodules which were agreed by all four radiologists, and also it has to be larger in CT scan series to be in dataset. Instead of following same method as in [4] we are considering the entire nodules which may appear in consecutive images in the CT series irrespective of sizes. This is because we have to notice the effect of resampling prior to classification. Therefore, at the end of our dataset preparation we have 4,532 nodules and the details about their distribution in original dataset and resampled dataset are given in Table 2. We have used SMOTE method technique for resampling dataset to make it balanced. The working principle of SMOTE [15, 16] technique will be discussed in Sect. 5.

Table 3 gives the instance distribution for malignancy case. The rating for the malignancy is further divided into multiclass such as Highly Unlikely, Moderately Unlikely, Indeterminate, Moderate and Suspicious cases. As we can see in Table 3 that the number of samples for highly likely cases is 572 where as for the cases Moderately Unlikely and Suspicious are 1,285 and 1,146 respectively. It means the number of cases for Moderately Unlikely and Suspicious is almost the double the number of samples in Highly Unlikely cases. In such scenario when we classify such imbalanced dataset, though using good performing classifier, the result will still be biased. This is because the classifier will get more number of samples of some classes and it will get fewer numbers of samples of other classes. Hence the classifier tends to get biased towards the case which has more number of samples.

Table 2 Samples distribution across the class in original dataset and resampled dataset

Class label for malignancy case	Number of samples	
	Original dataset	Resampled dataset
Highly unlikely	572	575
Moderately unlikely	733	742
Indeterminate	1,285	1,219
Moderately	796	854
Suspicious	1,146	1,142
Total number of samples	4,352	4,352

Table 3 Malignancy sample distributions in dataset

Class label	Number of samples
Highly unlikely	572
Moderately unlikely	733
Indeterminate	1,285
Moderate	796
Suspicious	1,146

It is to be noted that majority of real life medical data is indeed imbalanced. This reflects the distribution of such issues across the general population. Thus, working on such data is important since it captures realistic situation much more effectively. Hence we regard the dataset which we are considering in this work as class imbalanced data.

4 Image Feature Extraction

In this work we are using the same set of features which has been used in earlier work [16, 11]. Our image feature set consists of 55 low level image features. In addition to image features we have also used 7 radiologist characteristic ratings making the size of feature set to 62. The details about the image features we have used in this work are given in Table 4.

5 Methodology

In our previous work [11] we have investigated the role of single classifier versus panel of classifiers on LIDC data. In this work we consider smaller subset of data consisting of 212 nodules which are extracted from 50 cases. In [16] we have attempted to notice the significance of homogenous ensemble of classifier and heterogeneous ensemble of classifiers on LIDC data. There we have used

Table 4 Details of low level image features considered

Size features	Shape features	Intensity features
Area	Circularity	Min intensity
Convex area	Roughness	Max intensity
Perimeter	Elongation	Mean intensity
Convex perimeter	Compactness	SD intensity
Equiv diameter	Eccentricity	
Major axis length	Solidity	
Minor axis length	Extent	

Texture features
24 Gabor features are mean and standard deviation of 12 different gabor response images at orientation = 0, 45, 90, 135 and time frequency = 0.3, 0.4, 0.5
13 Haralick features calculated from co-occurrence matrices. Energy, correlation, inertia, entropy, inverse difference moment, sum average, sum variance, sum entropy, difference, average, difference variance, difference entropy, information measure of correlation 1, information measure of correlation 2

DECORATE and stacking ensemble method to construct ensembles. In [14] we had used class imbalanced dataset and single classifier model. Various resampling approaches prior to classification were used and we noticed significant improvement in the results. In this current paper we are using a large dataset, it consists of 4,532 nodules from 124 cases which is a relatively larger dataset compared to the dataset which we have used in our previous works. Here we are considering resampling approach as well as panel of classifiers using stacking method to investigate the performance of classifiers on class imbalanced data. Data resampling and stacking methods are discussed in detail in Sects. 6 and 7.

6 Dataset Resample

6.1 SMOTE

Synthetic Minority over-sampling Technique (SMOTE) generates synthetic examples by operating in the feature space rather than in the data space [15, 16]. The synthetic examples cause the classifier to create larger and less specific decision regions, rather than smaller and more specific regions. The minority class is over-sampled by taking each minority class sample and introducing synthetic examples along the line segments joining any/all of the k minority class nearest neighbors. The steps involved in SMOTE method is as follows: For each minority observation:- (1) Find its k-nearest minority neighbors (2) Randomly select ‘n’ of these neighbors (3) Randomly generate synthetic samples along the lines joining the minority sample and its ‘n’ selected neighbors (‘n’ depends on the amount of oversampling which is pre defined). The flow diagram of SMOTE method is as shown in Fig. 1.

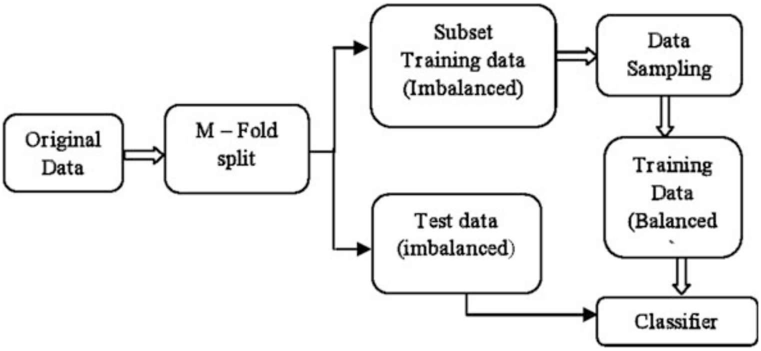
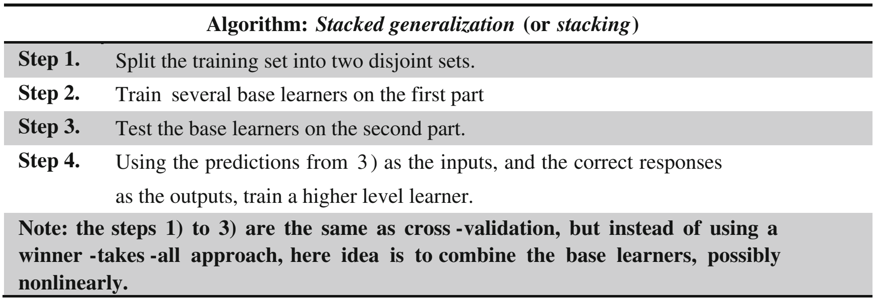


Fig. 1 SMOTE resample work flow diagram

7 Stacking Methodology

In machine learning, ensemble methods use multiple models to obtain better predictive performance than that could be obtained from any of the constituent models [17]. Stacked generalization (or stacking) was first proposed by Wolpert in 1992 [18] is a way of combining multiple models that introduces the concept of a meta-learner. Although it is an attractive idea, it is less used than bagging and boosting in literature. Stacking is a machine learning technique and it is a variant in ensemble literature in that, it actively seeks to improve the performance of the ensemble by correcting the errors. It addresses the issue of classifier bias with respect to training data and aims at learning and using these biases to improve classification and it regarded as stacked generalization. It is concerned with combining multiple classifiers generated using different base classifiers $C1, C2, \ldots Cn$ on a single dataset D, which consist of pattern examples. In the first phase, a set of base level classifiers are generated. In second phase, a meta–level classifier is learned that combines the outputs of the base level classifier [5]. In brief, stacking can be visualized as a method which uses a new classifier to correct the errors of previously learned classifier.

The algorithm of stacking [18] is as given in Fig. 2.



Algorithm: *Stacked generalization* (or *stacking*)	
Step 1.	Split the training set into two disjoint sets.
Step 2.	Train several base learners on the first part
Step 3.	Test the base learners on the second part.
Step 4.	Using the predictions from 3) as the inputs, and the correct responses as the outputs, train a higher level learner.
Note: the steps 1) to 3) are the same as cross -validation, but instead of using a winner -takes -all approach, here idea is to combine the base learners, possibly nonlinearly.	

Fig. 2 Stacked generalization (or stacking)

8 Experimental Setup

In this work we have experimented using the following environment. The main objectives of this work is: (1) to notice the performance of stacking ensemble technique on lung nodule prediction data compared to single classifier model (2) to observe the role of how the performance can be boosted if the dataset is made balanced prior to classification algorithm.

For the first set of experiment we have used REPTree (Reduced Error Pruning Tree), Bagging (Bootstrapped Aggregating) and AdaBoost (Adaptive Boosting) algorithms as base classifiers. We have used the same REPTree as a base learning method also for bagging and AdaBoost. The choice of using REPTree in all the cases is to investigate how stacking method performs in homogenous ensemble condition. We have stacked above said four base models with stacking method using linear regression as a meta-level learner.

In the second set of experiments we have used the following model as base learners. REPTree, Naïve Bayes, PART [19] (rule based classifier), Bagging (here the J48decision tree has been used a base classifier), AdaBoost (here we have used Decision Stump as a base learner), Support Vector Machine (here the sequential minimum optimization is used to train the SVM with polynomial kernel). All the above mentioned six base models are learned and stacked using linear regression meta-level learning algorithm. We have run the experiment twice using above said environments, once on original dataset and once on resampled dataset. We have used m- fold cross validation, where the value of m is set to 5.

9 Performance Evaluation

In this work we are evaluating the performance of model using five performances metrics. Accuracy (ACC), Root Mean Squared Error (RMSE), F–Measure, Kappa Statics, Area Under Curve (AUC). Accuracy and F-measure are regarded as thresholded metrics and we have fixed threshold to 0.5 and it means that classifier above the threshold is considered as good performer and classifier which performance below are threshold regarded as under performer. Root Mean squared Error is used as probability metric. Probability metric are minimized when the predicted value for each case is equals to true conditional probability. AUC is used to a rank metric and this metric measures how well the positive cases and negative cases are ordered and viewed. Kappa statics is used as agreement measures, which in turn reflect how well model agrees between the expert prediction and machine prediction. The kappa interpretation scale has been given in Table 5.

Table 5 Kappa statistics interpretation scale

K-value	Strength of agreement
<0	Poor
0–0.2	Slight
0.21–0.4	Fair
0.41–0.6	Moderate
0.61–0.8	Substantial
0.81–1	Almost perfect

10 Results and Discussion

The experiments have been carried out in two different environments as discussed in previous section. In Tables 6 and 7 we have tabulated the results. Each corresponds to different base model and each column corresponds to the obtained performance metric results. For each performance metric there will be two results, which refer to classifier response to original dataset and resampled dataset. We have used the different base classifier for our experiment. This is because it has been claimed in the literature [20] that construction of ensemble is directly proportional to the choice of base learners, the reason behind this is if the base learner is unstable ensemble will get much diversity and works better, if not, ensemble of classifier faces over fitting problem.

10.1 With Homogenous Ensemble Environment

We have tabulated the results and highlighted the best performing model (best in the column). In the entire performance category stacking has outperformed all other models. Only in the case of AUC it is equals with that of bagging method on both original dataset and resampled dataset. It worth noticing that all the classifiers including stacking have gained improvement in accuracy as well as on other metrics when operated on resampled dataset.

Table 6 Results from experiment using homogenous ensemble of classifier

Classifier	Accuracy		RMSE		F-measure		Kappa		AUC	
	OD	RD	OD	RD	OD	RD	OD	RD	OD	RD
REPTree	74.09	81.54	0.31	0.26	0.82	0.87	0.68	0.77	0.92	0.95
Bagging	80.26	87.13	0.24	0.20	0.88	0.92	0.75	0.84	0.98	0.99
AdaBoost	74.33	83.18	0.27	0.22	0.83	0.89	0.67	0.79	0.96	0.98
Stacking	81.66	88.18	0.23	0.19	0.89	0.94	0.76	0.85	0.98	0.99

OD original dataset, *RD* resampled dataset

Table 7 Results from experiment using heterogeneous ensemble of classifier

Classifier	Accuracy		RMSE		F-measure		Kappa		AUC	
	OD	RD	OD	RD	OD	RD	OD	RD	OD	RD
REPTree	74.90	81.54	0.31	0.26	0.82	0.87	0.68	0.77	0.92	0.95
Naïve Bayes	36.92	37.53	0.46	0.46	0.46	0.47	0.23	0.24	0.83	0.83
PART	76.98	83.56	0.29	0.25	0.84	0.88	0.71	0.79	0.92	0.95
AdaBoost	36.51	35.48	0.38	0.38	0.70	0.65	0.13	0.13	0.80	0.78
Bagging	75.04	82.24	0.26	0.24	0.84	0.88	0.68	0.77	0.97	0.98
SVM	58.65	58.53	0.36	0.36	0.75	0.76	0.46	0.46	0.90	0.90
Stacking	**81.09**	**86.70**	**0.24**	**0.20**	**0.87**	**0.91**	**0.76**	**0.83**	**0.98**	**0.98**

OD original dataset, *RD* resampled dataset

10.2 With Heterogeneous Ensemble Environment

In the heterogeneous environment we have used different classifier with different base learners. Best example is, we have used REPTree as a base classifier for AdaBoost in previous set up and here we have selected decision stump. When REPTree is used as a base classifier for AdaBoost it has performed very well by obtaining ACC = 74.33 %/83.18 %, RMSE = 0.27/0.22, F- measure = 0.83/0.89, Kappa = 0.67/0.79 and AUC = 0.96/0.98. When we compare the same AdaBoost with decision stump base learner it has given ACC = 36.51 %/35.48 %, RMSE = 0.38/0.38, F- measure = 0.70/0.65, Kappa = 0.13/0.13 and AUC = 0.80/0.78 which shows the choice of base learner is also very important when we deal with ensemble methods. But when we compare the results between the homogenous stacked ensemble and heterogeneous stacked ensemble the results in the all the columns are almost similar. Stacking of classifier can be considered as a information fusion technique. This is because, as we have noticed in our experiments, the meta-learner in the stacking will correct the errors made by the base learners.

11 Conclusion

In this work we have presented experiments to observe the role of machine learning techniques on medical data which happens to be imbalanced. We used machine learning techniques both at data level and algorithmic level. At data level we have used resample technique called SMOTE to make data distribution balanced across the classes in the case prior to classification task. At algorithmic level we have used stacking ensemble method which uses stack of classifiers as a base model, gets their scores and in next phase uses meta-learning algorithm which corrects the error which has occurred in previous stage. Results from experiments shows that machine learning methods outperform at all the levels and we also observed the following.

1. Making the dataset balanced before classification task always improves the result significantly; this can be validated with results from [14].
2. With reference to [11] we can claim that performance of ensemble of classifier is always better when compared to performance of single classifier.
3. Stacking method can be considered as information fusing technique since the results from stacking method outperformed that of any other in the group. Stacking method also performed better compared to other available ensemble method such as bagging or AdaBoost.
4. It has been claimed in literature [6], that generally, heterogeneous ensemble of classifier model performs better than homogenous ensemble of classifiers. However, in our experiments when we compare with respect to some performance metric homogenous ensemble of classifier results show marginally better results. This reveals the fact that the choice of base learning algorithm is very much important in creating ensemble. It can easily happen that a particular data maybe best classified by one particular model of classifier and its ensemble may actually improve the result. On the contrary introducing heterogeneous ensemble of classifiers may actually not improve; perhaps degrade the result even if marginal. So the original data, resampling methods, all play subtle but important role in final performance.
5. Use of ensemble method is similar to the process carried out by human expert, since the output labels from ensemble is produced by a combination rule such as voting. In our experiment we can observe that predictions from stacking method is statistically signification and kappa statics interpret the level of agreement between the expert and that of machine prediction is almost perfect.

References

1. Varutbangkul E, Mitrovic V, Raichu D, Furst J (2008) Combining boundaries abd rating from multiple observers for predicting lung nodule characteristics. In: IEEE international conference on biocomputing, bioinformatics and biomedical technologies, pp 82–87
2. Ebadollahi S, Johnson DE, Diao M (2008) Retrieving clinical cases through a concept space representation of text and images. SPIE Medical Imaging 2008: PACS and Imaging Informatics. 6919(7). ISBN: 9780819471031
3. Nakumura K, Yoshida H, Engelmann R, MacMahon H, Kasturagawa S, Ishida T et al (2000) Computerized analysis of the likelihood of malignancy in solitary pulmonary nodules with use of artificial neural networks. Radiology 214(3):823–830
4. Zinovev D, Raicu D, Furst J, Armato SG (2009) Predicting radiological panel opinions using a panel of machine learning classifiers. Algorithms 2:1473–1502. doi:10.3390/a2041473
5. Oza NC, Tumer K (2008) Classifier ensembles: select real-world applications. Inf Fusion 9(1):4–20
6. Reid S (2007) A review of heterogeneous ensemble methods. Department of Computer Science, University of Colorado at Boulder
7. Kuncheva LI, Rodriguez JJ (2010) Classifier ensemble for fMRI data analysis: an experiment, magnetic resonance imaging, vol 28. Elsevier Publications, pp 583–593

8. Caruana R, Niculescu-Mizil A, Crew G, Ksikes A (2004) Ensemble selection from libraries of models. In: 21st international conference on machine learning, Banff, Canada
9. Datta S, Pihur V, Datta S (2010) An adaptive optimal ensemble classifier via bagging and rank aggregation with application to high dimension data. BioMed Central 1471-2105/11/427, BMC Bioinformatics
10. Dzeroski S, Zenko B (2004) Is combining classifiers with stacking better than selecting the best one? Mach Learn 54:255–273, Kluwer Academic Publishers
11. Vinay K, Rao A, Hemantha Kumar G (2011) Comparative study on performance of single classifier with ensemble of classifiers in predicting radiological experts ratings on lung nodules. In: Indian international conference on artificial intelligence (IICAI). ISBN: 978-0-9727412-8-6, pp 393–403
12. Ting KM, Witten IH (1999) Issues in stacked generalization. J Artificial Intell Res 10:271–289
13. National Center for Biotechnology Information http://www.ncbi.nlm.nih.gov
14. Vinay K, Rao A, Hemantha Kumar G (2012) Sampling driven approaches for lung nodule characteristic rating predication. In: The 3rd international conference on intelligent information systems and management (IISM), ISBN No.: 978-93-90716-96-1
15. Chawla NV, Bowye KW, Hal LO, Kegelmeye WP (2002) SMOTE: synthetic minority over-sampling technique. J Artificial Intell Res 16:321–357
16. Vinay K, Rao A, Hemantha Kumar G (2011) Computerized analysis of classification of lung nodules and comparison between homogeneous and heterogeneous ensemble of classifier model. In: 3rd national conference on computer vision, pattern recognition, image processing and graphics, 978-0-7695-4599-8/11, IEEE doi:10.1109/NCVPRIPG.2011.56, pp 231–234
17. Polikar R (2006) Ensemble based systems in decision making. IEEE Circuits Syst Mag 6(3):21–45. doi:10.1109/MCAS.2006.1688199
18. Wolpert DH (1992) Stacked generalization. Neural Networks 5(2):241–259
19. Frank E, Witten IH (1998) Generating accurate rule sets without global optimization. In: Shavlik J (ed) Machine learning: proceedings of the fifteenth international conference. Morgan Kaufmann Publishers, San Francisco
20. Polikar R (2009) Ensemble learning. Scholarpedia 4(1):2776