ISSN: 2278 – 1323
International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)
Volume 1, Issue 10, December 2012

A Review on
Ensemble of Diverse Artificial Neural Networks

Mittal C. Patel[1], Prof. Mahesh Panchal[2]
[1]PG Student, [2]Head of the Computer engineering, Kalol Institute of Technology

KEYWORDS

Artificial neural network
Classifier
Combination of Classifiers
Diversity
Ensemble methods
Generalization error
Neural Network Ensembles

ABSTRACT

Ensemble Data Mining Methods, also known as Committee Methods or Model Combiners, which provides the power of multiple classifiers to achieve better prediction accuracy than any of the individual classifier could on their own. The diversity among the members of ensemble is used to determining its generalization error. The empirical results reveal that the performance of an ensemble is related to the diversity among individual learners in the ensemble and more diversity might be used to obtain better performance. Artificial Neural networks(ANN) are very flexible with respect to incomplete, missing and noisy data and also makes the data to use for dynamic environment. ANN is dependent on how best is the configuration of the net in terms of number of weights, neurons and layers. Diversity in an ensemble of neural networks can be handled by manipulating either input data or output data.

I. INTRODUCTION

Data Mining is the process to analyze the data from all different views and finally forming them into meaningful information. Data Mining is the central step in the KDD process that performs the different tasks like classification, clustering, summarization, regression, task analysis. This paper includes various classification techniques for predicting the class for data instances. Two main techniques, decision trees and neural networks are described here.
Decision tree is faster and easy to classify the data but not works good for the noise data. so, for the solution of noise data, neural network is used which is slower than decision tree but gives better result.SVM, Neural Network, Decision tree, Naïve Bayes and many other techniques are used to solve classification problem. Neural networks was found toprovide better classification accuracy than traditional statisticalmethods in various areas of applications such as business, finance, revenue, health, medicine, engineering, marketing, river monitoring, and Pale oceanography [1], [2], [3], [4], [5].Even though SVM was found to provide better classificationaccuracy than neural network in some applications [6], neuralnetwork was also found to perform better than SVM in varioustasks such as document classification [7], exudate classification[8], bio-activity classification [9], biological microscopicimage classification [10], and learning disability diagnosisproblem [11]. However, it is found that neural network is one of themost widely used methods used to solve the classificationproblem [12].

Combining multiple classifiers performs better than a single classifier but suffers from the difficulty of construction process (diversity of members, parameter tuning, and combination methods). It is well known that combining identical classifiers has no gain and diversity among members is one of the biggest issues in forming successful ensembles.The efforts on the ensemble approach can be categorized into two types: generating individual artificial neural networks and combining individual predictions [4]. Boosting and bagging are popular methods for generating individual networks.
Bagging [13]:Breiman [24] introduced the concept of bootstrapaggregating to construct ensembles. It consists intraining different classifiers with bootstrapped replicas ofthe original training data-set. That is, a new data-set isformed to train each classifier by randomly drawing (withreplacement) instances from the original data-set (usually,maintaining the original data-set size).Hence, diversityis obtained with the resampling procedure by the usageof different data subsets.
Boosting: Boosting (also known as ARCing, adaptive resampling and combining) was introduced by Schapire in1990 [14]. Schapire proved that a weak learner (which isslightly better than random guessing) can be turned into astrong learner in the sense of *probably approximately correct*(PAC) learning framework. AdaBoost [15] is them Boostrep resentative algorithm in this family, it was the first applicableapproach of Boosting, and it has been

ISSN: 2278 – 1323
International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)
Volume 1, Issue 10, December 2012

appointed asone of the top ten data mining algorithms [16].

Paper includes method of generating diverse evolutionary neural networks and then combining these networks.In order to solve the problem of classification and prediction,an ensemble of accurate and diverse neural networkswas found capable of providing better results than a singleneural network [28].

The rest of this paper is organized as follows. **Section2** describes basic of Data Mining.**Section3**explains ensemble techniques. **Section4**consists diversity in Ensemble of classifiers.**Section5** contains the ensemble of Neural Networks.

II. DATA MINING

Data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both. Knowledge Discovery in Data is the non-trivial process of identifying valid, novel, potentially useful and ultimately understandable patterns from large data sets involving methods such as artificial intelligence, machine learning, statistics and database systems.[1]Ina sense, data mining is the
central step in the KDD process. The other steps in the KDD process are concerned with preparing data for data mining, as well as evaluating the discovered patterns (the results of data mining).

2.1 Steps of KDD process[17]:

The following steps are used to preprocess the large dataset.

- Selection: Obtain data from various sources.
- Preprocessing: Cleanse data.
- Transformation: Convert to common format. Transform to new format.
- Data Mining: Obtain desired results.
- Interpretation/Evaluation: Present results to user in meaningful manner.
- Knowledge presentation: where visualization and knowledge representation techniques are used to present the mined knowledge to the user.

2.2 Data Mining Tasks:

- **Classification** maps data into predefined groups or classes.
- **Regression** is used to map a data item to a real valued prediction variable.
- **Clustering** groups similar data together into clusters.
- **Summarization** maps data into subsets with associated simple descriptions.
- **Link Analysis** uncovers relationships among data.

2.3 Classification techniques

Classification is a data mining (machine learning) technique used to predict group membership for data instances. For example, you may wish to use classification to predict whether the weather on a particular day will be “sunny”, “rainy” or “cloudy”. Popular classification techniques include decision trees and neural networks.[18]

2.3.1 Decision tree[19]

A model consisting of nodes that contain tests on a single attribute and branches representing the different outcomes of the test. A prediction is generated for a new example by performing the test described at the root node and then proceeding along the branch that corresponds to the outcome of the test. If the branch ends in a prediction then that prediction is returned. If the branch ends in a node, then the test at that node is performed and the appropriate branch selected. This continues until a prediction is found and returned.

A Decision Tree Model is a computational model consisting of three parts:

1. Decision Tree
2. Algorithm to create the tree
3. Algorithm that applies the tree to data

Features:

- Processing of Decision tree is basically a search similar to that in a binary search tree.
- Root and each internal node is labeled with a question.
- The arcs represent each possible answer to the associated question.
- Each leaf node represents a prediction of a solution to the problem.
- Leaf node indicates class to which the corresponding tuple belongs.

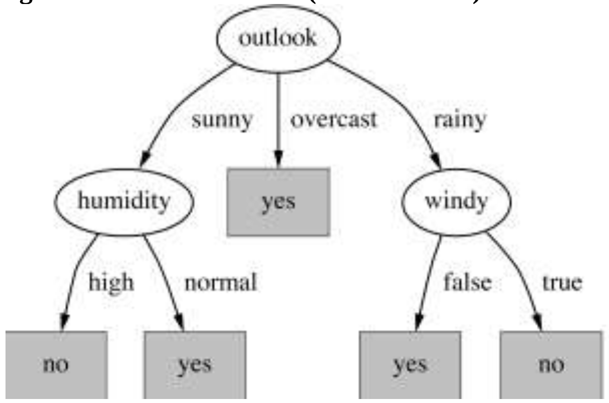
Experimental Study and Analysis:

ISSN: 2278 – 1323
International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)
Volume 1, Issue 10, December 2012

Table 1: Example of Whether Data of Weka 3.6:

ID	Outlook	Temp.	Humidity	Windy	Play?
A	sunny	hot	high	FALSE	No
B	sunny	hot	high	TRUE	No
C	overcast	hot	high	FALSE	Yes
D	rain	mild	high	FALSE	Yes
E	rain	cool	normal	FALSE	Yes
F	rain	cool	normal	TRUE	No
G	overcast	cool	normal	TRUE	Yes
H	sunny	mild	high	FALSE	No
I	sunny	cool	normal	FALSE	Yes
J	Rain	mild	normal	FALSE	Yes
K	Sunny	mild	normal	TRUE	Yes
L	overcast	mild	high	TRUE	Yes
M	overcast	hot	normal	FALSE	Yes
N	rain	mild	high	TRUE	No

Fig.1 Ex. for a decision tree(Weather data)[31]



Attribute Selection:

- Information Gain
  Class P: Outlook = "yes"

Class N: Outlook = "no"

$$Info(D) = I(9,5) = -\frac{9}{14}\log_2(\frac{9}{14}) - \frac{5}{14}\log_2(\frac{5}{14}) = 0.940$$

- Entropy
  Entropy lies between 0 and ln(k) (k is the number of classes). A pure profile has zero entropy (= best). The profile with all class frequencies equal has the largest entropy (= worst), the entropy here is ln(k).

$$Gain(age) = Info(D) - Entropy_{age}(D) = 0.246$$

Root Selection Method:

- Entropy of root node is 0.940
- Gain of "Outlook" = 0.247

Other splits yield:

- Gain(temperature)=0.029
- Gain(humidity)=0.152
- Gain(windy)=0.048

So "outlook" is the best attribute to split on

Result analysis:

Relation: weather
Instances: 14
Attributes: 5
Outlook, temperature, humidity, windy, play
Test mode: 10-fold cross-validation

Confusion Matrix:

a b <-- classified as
8 1 | a = yes
2 3 | b = no

Classifier model: Random Tree
outlook = sunny
| humidity< 77.5 : yes (2/0)
| humidity>= 77.5 : no (3/0)
outlook = overcast : yes (4/0)
outlook = rainy
| windy = TRUE : no (2/0)
| windy = FALSE : yes (3/0)
Size of the tree : 8
Max depth of tree: 3

2.3.2 Artificial Neural Network

An artificial neural network (ANN), often just called a "neural network" (NN), is a mathematical model or computational model based on biological neural networks, in other words, is an emulation of biological neural system. It consists of an interconnected group of artificial neurons and processes information using a connectionist approach to computation. In most cases an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase.

A neuron (or single layer perceptron) is a function of N real variables of the form

$$f(x_1, \ldots, x_N) = \operatorname{sgn}\left(\sum_{i=1}^{N} w_i x_i - \theta\right). \quad (1)$$

Here xi are real variables, (x1, . . . ,xN) takes values in some domain U, wiare real parameters (weights of the neuron), _ is the threshold of activation of the neuron, the function sgn(x) = 1, for x≥0 and is equal to zero for x <0.

Consider the smoothed variant of the above neuron for which instead of the functionsgn,also can use the smooth monotonous increasing function sgm which varies from zero to unity. In particular generally consider the neuron of the form

$$f(x_1, \ldots, x_N) = \mathrm{sgm}\left(\sum_{i=1}^{N} w_i x_i - \theta\right), \qquad \mathrm{sgm}(x) = \frac{1}{1+e^{-x}}. \quad (2)$$

A neural network is a composition of the above neurons.

Result analysis:
Parameters of result from Weka 3.6

Hidden layers = 3
Momentum = 0.2
Validation Threshold = 20
Training time = 500
Learning Rate = 0.3
GUI=True

Relation: weather
Instances: 14
Attributes: 5
Outlook, temperature, humidity, windy,play
Test mode: 10-fold cross-validation

Confusion Matrix:
a b <-- classified as
8 1 | a = yes
2 3 | b = no

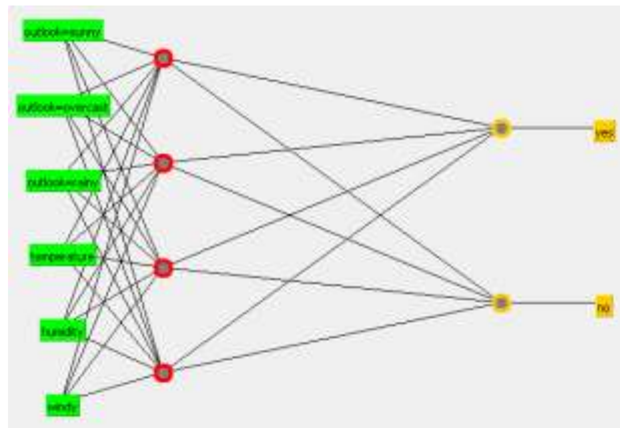


Fig.2 Result of Neural Network for Weather Data when run in the Weka 3.6(GUI=True)

Comparative sty of Decision Tree and Neural Network :

Table 2:Result of Decision Tree for weather dataset of weka tool

Detailed Accuracy by class:							
	TP Rate	**FP Rate**	**Precision**	**Recall**	**F-Measure**	**ROC Area**	**Class**
	0.889	0.4	0.8	0.889	0.842	0.744	Yes
	0.6	0.111	0.75	0.6	0.667	0.744	No
Weighted Avg.	0.786	0.297	0.782	0.786	0.779	0.744	

Table 3:Result of Artificial Neural Network for weather dataset of weka tool

Detailed Accuracy by class:							
	TP Rate	**FP Rate**	**Precision**	**Recall**	**F-Measure**	**ROC Area**	**Class**
	0.889	0.4	0.8	0.889	0.842	0.733	Yes
	0.6	0.111	0.75	0.6	0.667	0.733	No
Weighted Avg.	0.786	0.297	0.782	0.786	0.779	0.733	

2.3.3 Decision Tree vs Neural Network

There are many differences between these two, but in practical terms, there are three main things to consider: speed, interpretability, and accuracy.

Decision Trees

- Should be faster once trained(both for training and classification.This is because a decision tree inherently "throws away" the input features that it doesn't find useful, whereas a neural net will use them all unless you do some feature selection as a pre-processing step.
- If it is important to understand what the model is doing, the trees are very interpretable.
- Only model functions which are axis-parallel splits of the data, which may not be the case.
- You probably want to be sure to prune the tree to avoid over-fitting.

Neural Networks

- Slower (both for training and classification), and less interpretable.
- If your data arrives in a stream, you can do incremental updates with stochastic gradient descent (unlike decision trees, which use inherently batch-learning algorithms).
- Can model more arbitrary functions (nonlinear interactions, etc.) and therefore might be more accurate, provided there is enough training data. But it can be prone to over-fitting as well.
  - **High Accuracy:** Neural networks are able toapproximate complex non-linear mappings.
  - **Noise Tolerance:** Neural networks are veryflexible with respect to incomplete, missingand noisy data.
  - **Independence from prior assumptions:** Neuralnetworks do not make a priori assumptionsabout the distribution of the data, or the formof interactions between factors.
  - **Ease of maintenance:** Neural networks can be updated with fresh data, making them useful for dynamic environments.
  - Neural networks can be implemented in parallel hardware.
  - When an element of the neural network fails, it can continue without any problem by theirparallelnature.fraud detection,telecommunications, medicine,marketing, bankruptcy prediction, insurance, the list goes on. The following are examples of where neural networks have been used.

III. ENSEMBLE TECHNIQUES[17]

Bagging (or boostrap aggregation) and boosting are two such techniques. Each combines a series of T learned classifiers, C1; C2;…..CT, with the aim of creating an improved composite classifierC*.

3.1 Bagging

Working of Bagging for sample of S set is: For iteration i (i = 1,2,…m), a training set St is sampled with replacement from the original set of samples, S.[20]As Sampling by replacement is used, but Si does not includes all sample of S and some of them occurs more than one time. Each bootstrap sample Si has 63.2% and 36.8% used as of the original training data and as test set respectively. A classifier Ci is learned for each training set, Si .To classify an unknown sample, Y, each classifier Ct will give its class prediction and one vote is counted for it.The bagged classifier, C*, counts the votes and assigns the class with the most votes to.[20] Bagging can be applied to the prediction of continuous values by taking the average value of each vote, rather than the majority. [20]

Advantages [20]

- Bagging works well if the base classifiers are unstable.
- It Increased accuracy because it reduces the variance of the individual classifier.
- Bagging seeks to reduce the error due to variance of the base classifier.
- Noise-tolerant, but not so accurate

3.2 Boosting [30]

The version of boosting investigated in this paper is AdaBoost.M1.(Freund and Schapire) Instead of drawing a succession of independent bootstrap samples from the original instances boosting maintains aweight for each instance the higher the weight themore the instance influences the classifier learned. Ateach trialthe vector of weights is adjusted to reflectthe performance of the corresponding classifier, withthe result that the weight of misclassified instancesis increased. The final classifier also aggregates thelearned classifiers by voting, but each classifiers voteis a function of its accuracy. The boosting algorithm can be extended for the prediction of continuous values.

Advantage [17]

- Boosting tends to achieve more accuracy than bagging
- Boosting focuses on misclassified tuples so it risks overfitting

Limitation

- Boosting can fail to perform well given insufficient data. This observation is consistent with the Boosting theory. Boosting also does not perform well
- When there is a large amount of classification noise (i.e. training and test examples with incorrect class labels).Boosting is also very susceptible to noise in the data.

Comparison between Bagging and Boosting[17]

Bagging is noise-tolerant, produce better class probability estimates. It is not so accurate. It is related to random sub-sampling.

While **Boosting** is very susceptible to noisy data, produces rather bad class probability estimates. It is related to windowing.

IV. ENSEMBLE AND DIVERSITY[29]

The key idea in ensemble research is; if a classifier or predictor is unstable then anensemble of such classifiers voting on the outcome will produce better results – betterin terms of stability and accuracy. While the use of ensembles in Machine Learning (ML) research is fairly new, the idea that aggregating the opinions of a committee ofexperts will increase accuracy is not new.

The Codorcet Jury Theorem states that:

If each voter has a probability p of being correct and the probability of a majorityof voters being correct is M, then p > 0.5 implies M > p. In the limit, approaches 1, for all p > 0.5, as the number of voters' approaches infinity.

This theorem was proposed by the Marquis of Condorcet in 1784 (Condorcet, 1784) –a more accessible reference is (Nitzan&Paroush, 1985). It is known that M willbe greater that p only if there is diversity in the pool of voters. And also known that theprobability of the ensemble being correct will only increase as the ensemble grows ifthe diversity in the ensemble continues to grow as well. Typically the diversity of theensemble will plateau as will the accuracy of the ensemble at some size between 10and 50 members.

In ML research it is well known that ensembling will improve the performance ofunstable learners. Unstable learners are learners where small changes in the trainingdata can produce quite different models and thus different predictions.

5.1 Different measures of diversity

Many different ways are available to measure the diversity but in a continuous output(regression) problem Squared error measure is used.(Krogh &Vedelsby, 1995).It is described by following equation:[29]

$$a_i(x_k) = [V_i(x_k) - \overline{V}(x_k)]^2 \tag{3}$$

Where ai= the ambiguity of the ith classifier on example xk, randomly drawn from anunknown distribution,
Vi=ith classifier
V= theensemble predictions.
Error from the ensemble is: E =E Bar-ABar
E= Overall Error of the ensemble
E bar=generalization error (ensemble components)
A=is the ambiguity of theensembleHowever, for classification the most commonly used error measure is a simple 0/1loss function, so a measure of ambiguity in this case is:

$$a_i(x_k) = \begin{cases} 0 \text{ if } class V_i(x_k) = class\overline{V}(x_k) \\ 1 \text{ otherwise.} \end{cases} \tag{4}$$

where this time the classifier and ensemble outputs for the case labeled as $x_k$are classes instead of real numbers.

V. ENSEMBLE NEURAL NETWORKS

The bagging technique for Ensemble ANN's improves the classification performanceas compared to the technique that applies onlya single pair of opposite networks or only a single network.Boosting provides diversity by manipulating each training set according to the performance of the previous classifier [21]. Furthermore, manipulating the input features can also provide diversity in the ensemble [22].AnEnsemble of ANN's is a set different Artificial Neural Networks means they are diverse in nature.Ensemble of ANN's is achieved by two methods,

1: Method to generate set of artificial neural networks.

-Schapire's *Boosting* [23]and Breiman's *Bagging* [24] are prevailing methods for generating individual networks.

2: Method to combine individual predictions of artificial neural networks.

-To combine individual predictions of artificial neural networks two methods Simple averaging and weighted averaging are preferred.

To combine individual predictions of classifiers two methods Majority voting and *plurality voting* are preferred.

Majority voting judges a prediction to be the final output if more than half of the individual networks vote to theprediction. [25]

Plurality voting judges a prediction to be the final output if the prediction ranks first according to thenumber of votes. [25].

Voting: The result supported by the majority of ANN's is the output of the ensemble.

- Bayesian method: This method takes each ANN'ssignificance into account by allowing the errorpossibility of each ANN to affect the ensemble'sresults [27].

-Weighted average: The weighted average multiplies theweight w to the output of each ANN when averagingthe outputs.[26]

-Gating: Gating is a method for choosing the fittestANN by utilizing the information from the learningdata. Four steps are required to do this.

Step 1: In the learning stage, all ANN's generate a datalist for which they have produced correct outputs.

Step 2: In the test stage, the learning data most similar tothe input data are searched.

Step 3: An ANN which has recognized the searchedlearning data correctly is selected.

Step 4: The chosen ANN is applied to the test data.[26]

VII CONCLUSION

An ensemble of accurate and diverse neural networks was found capable of providing better results than a single neural network. In the normal process of utilizing neural network ensemble, each network component has to be trained and then the outputs obtained from all networks in the ensemble are combined. However, there are situations that outputs from the networks are differed from one another. if two classifiers produce different errors on new input data then both classifiers are considered to be "diverse". Diversity in an ensemble of neural networks can be handled by manipulating input data or output data.A network that is too complex may fit the noise, not just the signal, leading to over fitting.

VIII ACKNOWLEDGEMENT

I am very thankful to Prof. Mahesh Panchal and my family to support me for accomplishment of this research paper successfully.

REFERENCES

[1] B. A. Malmgren, M. Kucera, J. Nyberg, and C. Waelbroeck, "Comparison of Statistical and Artificial Neural Network Techniques for Estimating Past Sea Surface Temperatures from Planktonic Foraminifer Census Data," *Pale oceanography*, vol. 16, no. 5, pp. 520–530, 2001.

[2]Y.-M. Wang, S. Traore, and T. Kerh, "Monitoring Event-Based Suspended Sediment Concentration by Artificial Neural Network Models," *WSEAS Transactions on Computers*, vol. 7, no. 5, pp. 559–568, May 2008.

[3] M. Paliwal and U. A. Kumar, "Neural networks and statistical techniques:A review of applications," *Expert Systems with Applications*,vol. 36, pp. 2–17, 2009.

[4] C. Botoca, R. Bardan, M. Botoca, and F. alexa, "Prostate Cancer Prognosis Evaluaion Assisted by Neural Networks," *WSEAS Transactions on Computers*, vol. 9, no. 2, pp. 164–173, February 2010.

[5] P. Hajek and V. Olej, "Municipal Revenue Prediction by Ensembles of Neural Networks and Support Vector Machines," *WSEAS Transactions on Computers*, vol. 9, no. 11, pp. 1255–1264, November 2010.

[6] D. Meyer, F. Leisch, and K. Hornik, "The support vector machine under test," *Neurocomputing*, vol. 55, pp. 169–186, 2003.

[7] L. Manevitz and M. Yousef, "One-class document classification via Neural Networks," *Neurocomputing*, vol. 70, no. 7-9, pp. 1466–1481, 2007.

[8] A. Osareh, M. Mirmehdi, B. T. Thomas, and R. Markham, "Comparative Exudate Classification Using Support Vector Machines and Neural Networks," in *MICCAI '02: Proceedings of the 5th International Conference On Medical Image Computing and Computer-Assisted Intervention-Part. II*. London, UK: Springer-Verlag, 2002, pp. 413–420.

[9] J. Z. Shah and N. Salim, "Neural networks and support vector machines based bio-activity classification," in *Proceedings of the 1st International Conference on Natural Resources Engineering & Technology 2006*, Putrajaya, Malaysia, July 2006, pp. 484–491.

[10] I. Maglogiannis, H. Sarimveis, C. T. Kiranoudis, A. A. Chatziioannou, N. Oikonomou, and V. Aidinis, "Radial Basis Function Neural Networks Classification for the Recognition of Idiopathic Pulmonary Fibrosis in Microscopic Images," *IEEE Transactions on Information Technology in Biomedicine*, vol. 12, no. 1, pp. 42–54, 2008.

[11] T.-K. Wu, S.-C. Huang and Y.-R. Meng, "Evaluation of ANN and SVM classifiers as predictors to the diagnosis of students with learning disabilities," *Expert Syst. Appl.*, vol. 34, no. 3, pp. 1846–1856, 2008.

[12] H. Sug, "Investigating Better Multi-layer Perceptrons for the Task ofClassification," *WSEAS Transaction on Computers*, vol. 9, no. 5, pp.475–484, May 2010.

[13] MikelGalar, Alberto Fern´andez, EdurneBarrenechea, HumbertoBustince, *Member, IEEE*,and Francisco Herrera, *Member, IEEE*.
"A Review on Ensembles for the Class Imbalance Problem: Bagging-, Boosting-, and Hybrid-Based Approaches."*Presented at IEEE Transactions on systems, An, And cybernetics-part c: Applications and reviews 1.*

[14] R. E. Schapire, "The strength of weak learnability," *Mach. Learn.*, vol. 5, pp. 197–227, 1990.

[15] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *J. Comput. Syst. Sci.*, vol. 55, no. 1, pp. 119–139, 1997.

[16] X. Wu, V. Kumar, J. Ross Quinlan, J. Ghosh, Q. Yang, H. Motoda, G.J. McLachlan, A. Ng, B. Liu, P. S. Yu, Z.-H. Zhou, M. Steinbach, D.J. Hand, and D. Steinberg, "Top 10 algorithms in data mining," *Knowl.Inf. Syst.*, vol. 14, pp. 1–37, 2007.

[17]Prachi S.Adhvaryu1, Prof. Mahesh Panchal "A Review on Diverse Ensemble Methods for Classification", Kalol Institute of Technology.

IOSR Journal of Computer Engineering (IOSRJCE), Volume 1, Issue 4 (May-June 2012), PP 27-32 www.iosrjournals.org

[18]Mrs. Kiruthika M. Asst. Professor, ComputerDept, Mrs. Dipa Dixit Lecturer, Computer Dept., Fr.C.R.I.T."Mining access patterns using classification"*International Journal of Engineering Science and Technology*.Vol. 2(7), 2010, 2895-2902

[19]Nikunj C. Oza, Ph.D."Ensemble Data Mining Methods".*NASA Ames Research Center*, USA.

[20]Amit Kumar, R.M.Sharma. "A New Congestion Control Approach on TFRC OverWired and Wireless Networks".*IOSR Journal of Computer Engineering (IOSRJCE)* Vol. 1, Issue 4 (May-June 2012), PP 01-08

[21] H. Schwenk and Y. Bengio." Boosting Neural Networks". *IOSR Journal of Computer Engineering (IOSRJCE.)*Vol.1, Issue 4 (May-June 2012), PP 28-31

[22] T. G. Dietterich." Ensemble Methods in Machine Learning". *Multiple Classifier Systems,*vol. 1857 ofLecture Notes in Computer Science, pp1–15. Springer, 2000.

[23] Schapire RE. "The strength of weak learnability".*Machine Learning* 1990; 5(2): 197-227.

[24] L.Breiman L. "Bagging predictors". *Machine Learning* 1996; 24(2): 123-140.

[25] Zhi-Hua Zhou*, Yuan Jiang, Yu-Bin Yang, Shi-Fu Chen "Lung Cancer Cell Identification Based on Artificial Neural Network Ensembles".National Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093, P.R.China.*Artificial Ingelligence in Medicine*, vol.24, no.1, pp.25-36, @Elsevier,2002.

[26] Kyung-Joong Kim_, Sung-Bae Cho "Evolutionary ensemble of diverse artificial neural networks using speciation"*Neurocomputing* 71 (2008) 1604–1618

[27] L. Xu, A. Krzyzak, C.Y. Suen," Methods of combining multiple classifiers and their applications to handwriting recognition*", IEEE Trans. Systems Man Cybernet.* SMC-22 (3) (1992) 418–435.

[28] L. K. Hansen and P. Salamon. "Pattern Analysis and Machine Intelligence". In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 12, PP 993–1001, October 1990.

[29]Gabriele Zenobi, Pádraig Cunningham, Using "Diversity in Preparing Ensembles of ClassifiersBased on Different Feature Subsets to MinimizeGeneralization Error" Gabriele.Zenobi@cs.tcd.ie Padraig.Cunningham@cs.tcd.ie

[30]"Bagging, Boosting, and C4.5J. R". QuinlanUniversity of SydneySydney, Australia 2006.

[31] Frank Keller."Decision Trees-onnectionist and Statistical Language Processing".*ComputerlinguistikUniversitat des Saarlandes.*keller@coli.uni-sb.de. PP 1-29.