

Multi-Instance Heterogeneous Classifiers with Extended Space forest

Mr. Vikas Singh

Abstract— Multi-Instance Heterogeneous Classifiers with Extended Space forest (MIHC_ES) is a new method for Feature Set Generation along with efficient heterogeneous ensembling of classifier employed for linear classification problem. Extended Space feature generation is new and efficient system of generating new feature from original feature set. Ensemble classification system consists of multiple classifiers in which each classifier set consist of classifier instance of same type. In heterogeneous ensembling each classifier in the classifier set have multiple instance of same type of classifier together with different heterogeneous classifiers used for active learning. This set of Heterogeneous classifier within ensemble is capable of changing number of instances of each classifier type within the ensemble based maximum and minimum accuracy achieved .

The three major algorithm adopted for this experiment is collaborating extended space forest and stably sized heterogeneous ensembling of classifier and rotation forest .For Heterogeneous Ensembles (HE), experimental evaluations show that HE constructs heterogeneous ensembles that outperform homogeneous ensembles composed of any one of the classifier types, as well as it outperforms AHE on many analysis data set. We in this system leveraged the advantage of AHE over other methods by adapting instances of classifier type in overall in the ensemble during learning and the target data set is composed of target class labels.

Index Terms— Extended space forest , Heterogeneous ensemble , multi-instance classifier ,rotation forest.

INTRODUCTION

A classifier ensemble is a group of intricated classification model, referred to as base classifier, whose individual decision are combined in an effort to improve overall final prediction performance. The problem of combining classifier has been widely studied in literature[1]. It is proved by many studies that ensembling of classifier is more effective than a single classifier based approach[13]. The initial step involved in construction of classification system starts with creating different training dataset from the original dataset such as bagging[2], boosting[3], random subspaces[4]. The existing ensemble methods create different training dataset by deleting or weighting samples in ESAHE the Extended Space Forest (ESF)[5] is used which adds new feature(extended spaces) to the original dataset thereby increasing number of features in the feature set of data set , this is column wise concatenation at the end Rotation Forest[14] is used to create subset from extended data set for the purpose of achieving diversity. ESF is obtained from the original training dataset by applying various operations. Initially the features (attributes) of the original set are randomly permuted then several optimal space extending operator (sum, difference, comp, divide, tanh, two linear) are applied to two paired original features

Manuscript received Oct, 2015.

Mr. Vikas Singh,,Computer science , SBP Pune University/ AISSMSCOE

and are measured for individual accuracy through adaptive heterogeneous ensemble(HE). For each new training set generation, all the features are sorted randomly. This extended dataset is then fed into the classifier ensemble which will learn from this new training set and will accurately classify the query.

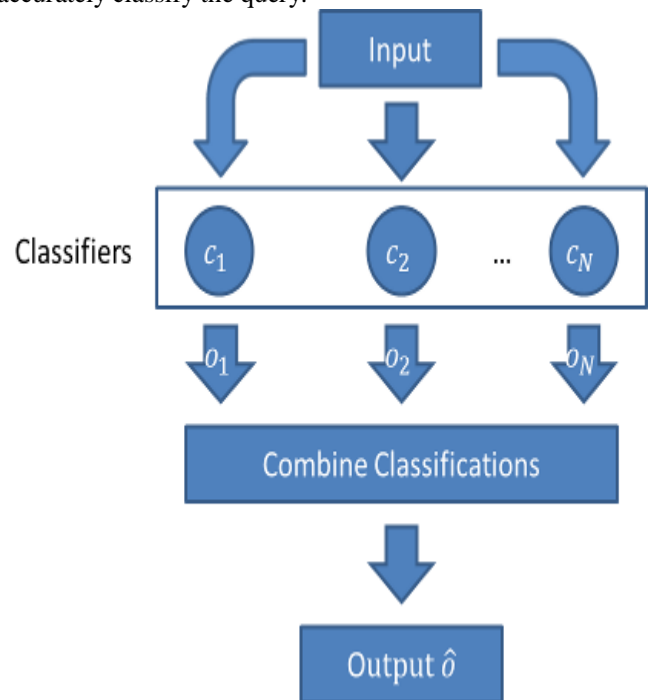


Figure.1 A homogeneous ensemble of classifier.

Heterogeneous ensemble in our system consist of multiple instances of C4.5, Neural network multi-layer perceptron[6] and naive bytes within a single ensemble this ensemble is responsible for classification of the test set through k-fold cross validation. The traditional ensemble can be seen in figure 1. The classification ensemble proposed is an active learning approach in which the algorithm automatically selects the most informative example from the presented dataset. Along with heterogeneous ensembling an algorithm[13] is implemented in this experiment which will adapt the ratio of different classifier types in a heterogeneous ensemble during training. Any other classifier can also be used in this proposed system.

Table 1:Creating new feature (NewFea) using “+” operator.

| Operator Name | Equation |
|---------------|------------------|
| Sum | NewFea=feaX+feaY |

I. RELATED WORKS

We have gone across many feature selection and feature generation schemes like Bagging[2] which creates new training data set for the principle learner by re-sampling and subset formation using different techniques [12] [5][4] the original data set with replacement.

Some of the approach use different combination of feature subspace as in Random Subspaces[4][8]. This method operates in two ways. In this first form, each base learner, is trained with a distinct feature subspace of the original training data set. But, only decision trees can be used as base learner in the second form, at each node of the decision trees, a randomly selected feature subspace is changed. The two forms are very similar to each other in terms of performance. RF [14] is a statistical algorithm that is used to cluster the points of data in the given groups of function. When the data set is large and/or there are many variables it becomes difficult to cluster the data because not all variables can be taken into account, therefore the algorithm can also give a certain chance that a data point may or may not belongs in a certain group.

Of the absolute set of data a subset is taken for obtaining training set. The algorithm makes the clusters the data in groups and subset of groups. If a line would be drawn between the data points in a subgroup, and lines that connect subgroups into group etc. At each split or node in this cluster or tree variables are chosen at random by the program to judge whether data points have a close relationship or not. The program makes may or generates multiple trees i.e. a grown forest consisting nodes. Each tree is different because for each partition in a tree, variables are chosen at random. Then the rest of the data set (not the training set) is used to predict which tree in the forests makes the best classification of the data points (in the data set the right classification is known). This generated tree with the highest accuracy and predictive power i.e. the accuracy of classification is shown as output by the algorithm.

In RF[14] each base learner is trained with slightly rotated original training data set. The permutation matrix is calculated for each principal learner by bootstrapping samples from the training data and from the classes. This method works only with numeric features. In many cases the data set may have features of other types apart from numeric features then these features are transformed to numeric representation. These methods adopt the approach of combining results using majority voting.

To generate new features from original feature set is not new idea HO[4] suggested if the number of features in the feature set are less than to increase redundancy between features we employ random forest. Breimen[10] also proposed feature set generation using linear combination of the feature in his paper of RF.

Heterogeneous ensemble of one instance of chosen type was proposed by Zenko[11] in this approach it first build the large library of classifier from the available set of different type and then selects heterogeneous ensemble by selecting members from library.

II. SYSTEM COMPONENTS

A. Architecture

First we are trying to to generate extended space data set using (Extended space forest) ESF [5] algorithm which adds new features to the original ones. This approach generates new features by using optimal space extending operator(SEO), one of the SEO [5] is applied on pair of random combination of original feature of original data set to generate new feature. There are total twelve Space Extending Operators[5],we in our experiment used addition operator for extending the data set attributes after extended space data set is generated then rotation forest creates sub set of the data set[14] fig. 2 shows block level architecture of MIHC-ES.

The obtained new feature are concatenated to original data set .Table. 1 shows addition as SEO used for creating new feature in this experiment. We found twelve such mathematical operator in our study[5] of the optimal space Extending operators. In Table.1 Sum as a SEO is used which uses some feature X_i and Y_i of the original data set $E=\{X_p, Y_p\}_{p=1..N}=[X Y]$ and fig.2 shows the basic blocks of classifier system which consists component ordered from top to bottom in terms of their execution. ESF[5] is at the top most position it is employed to perform the task orgf extending the given dataset and generate new dataset E_i to conclude execution. Second layer is Rotation Forest(RF)[14] which transforms the data set without any loss of information. RF creates different subset of instances, classes and features .At the time of training classifier(s) each of the classifier(s) is trained with different subset of original training set. Advantage of using rotation forest for achieving diversity are well known and established. Rotation Forest conform low computation and low storage[15].The third layer of this system is Stably Sized adaptive Heterogeneous set of classifier(s)[12] or committee of experts of heterogeneous type with single or multiple instances in one ensemble . We are expecting generation of small decision trees to reduce the complexity constraint. A diagrammatic representation of ESF is shown in fig.2 it shows the interrelation ship between random subspaces and ESF as shown in various steps, the generation of extended spaces through permuting the features and then applying the operation on randomly paired attributes these operation are shown in figure where new feature are obtainedthrough applying operation ,average kappa mean is used to access which operator is best Execution of RF and ESF can be depicted in figure 3.

After completion of ES(extended space) generation and random subspaces we move to classifier ensemble module. This module is different from conventional ensemble in choosing the classifier types. In this experiment we used HE(Heterogeneous Ensemble) we rather than ensembling multiple instances of single base classifier, we choose multiple instances of multiple classifier types. This is the first phase before the second phase which is adaptation phase, based on the diversity and accuracy obtained after several iteration the overall size of the ensemble will not either expand or shrink but instances of each classifier type in the ensemble changes this will help us to find most suitable classifier ratio in the ensemble based on accuracy evaluation.An adaptation set is also employed which is used

for adapting the configuration of the ensemble. The HE operates in two phases in the first phase of the training iteration each classifier in initial ensemble makes prediction of class label on each of the first chunk of data instances in the training pool and in second phase data instances that causes maximum disagreement among the ensemble member is choose for querying the label. This data instance along with the label is then added to the training set. Finally classification module consist of updated configuration of the ratio of different classifier instances in the ensemble to which test set is applied for query labeling. We will have more detail view on each component in the section below.

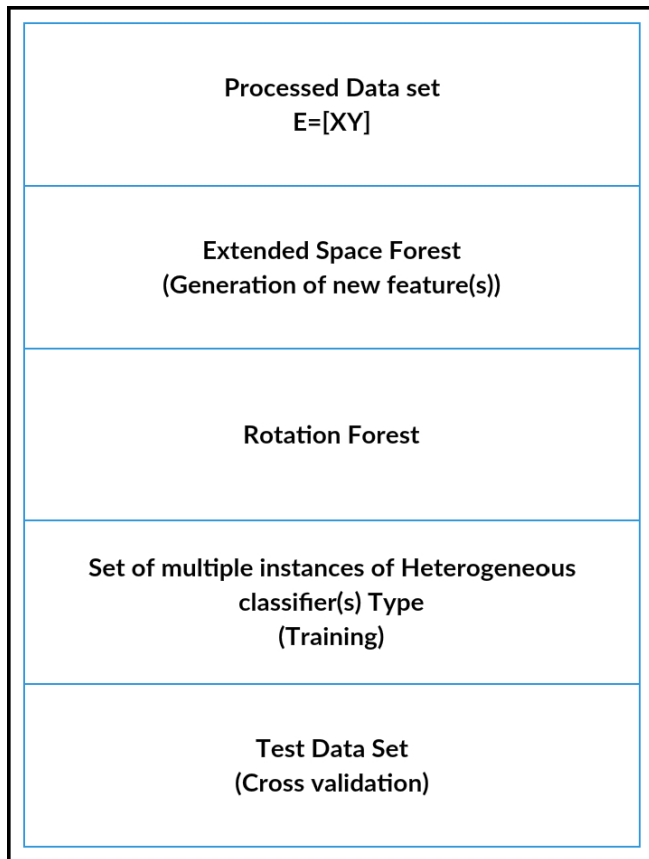


Figure 1: Layered architecture of the classification system

B. Extended Space Forest and Rotation Forest

The ESF arrangement is responsible for generating the training set for the HE which will have newly generated features in addition to feature set of the original dataset i.e. a extended training set is generated. The algorithm works as follows:

Given:

Original Dataset $E = \{ X_p, Y_p \} = [XY]_{p=1 \dots N}$.

$X = [N*d]$ matrix containing records with attribute values,

$N =$ Number of Records in data set,

$d =$ Numbers of features in data set,

$Y = N$ dimensional column vector consisting class label(multi-class),

Ensemble size is given by T , K is the ratio of new features to the original features in the data set. We also have a set SEO which has several OP and the heterogeneous Ensemble Model L_j . For each heterogeneous learner in T we will create new features EX_i by randomly pairing the original

features. We will generate $2*K$ random permutation of the original feature set. We concatenate all these random permutation and store them in C_i, C_i will have $2*K*d$ indicates. Instance along with the label is then added to the training set. Finally classification module consists of updated configuration of the ratio of classifier in the ensemble to which test set is applied for query labeling. We will have more detail view on each component in the section below. Now for each permuted feature set stored in C_i , generate $new_{j_{th}}$ features by applying OP to $C_i(z_{th}), C_i(z+1)_{th}$ features of X matrix for each matrix for $z = 1 \dots 2*k*d$ indices. The mentioned steps are applied iteratively for generating more new features. Construct the new training set (XEX, Y) by concatenating matrix X (original features) and EX_i the (new features).

To create L different subsets of data set E , feature vector stores $\hat{F} = [X+2*k*d]$ indices from extended data set is directed towards creating diversity. Each new extracted feature set $M = \left\lfloor \frac{(X * 2 * k * d)}{L} \right\rfloor$ which is linear combination of original feature[15].

. Then the "rotation matrix", R , used to transform a sample extracted using bootstrap T of the original training set into a new training set $(T' = TR)$ is sparse.

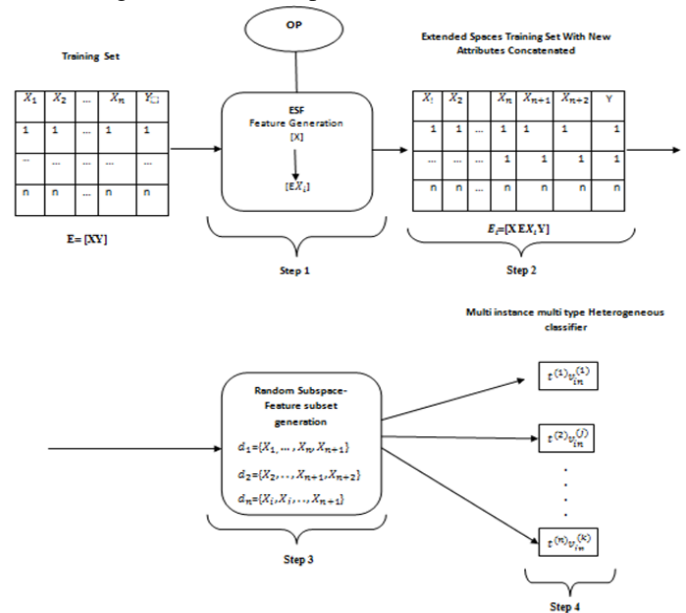


Figure 3. Step by step execution of ESF and rotation forest(step 4).

C. Ensembling of Heterogeneous Classifier

In heterogeneous ensembling[13] of classifier for active learning we in our system will have C4.5, Naive Bayes and Neural Network multi-layer perceptron instances within a single ensemble. The ensemble will consist of multiple instances of multiple classifier type. Rotation forest [14] is used for creating multiple instances of the each classifier in the ensemble the reason behind choosing rotation forest method is that by rotating the feature subspace there is no requirement for reasonably large number of training instances. We also in our proposed system are employing stream based selective sampling for active learning in which a sequence of data instances are drawn from an active learner to decide whether to query the label or not. This system considers an optimal ensembles such that it is constructed with the best ratio of classifier type. The heterogeneous classifier is depicted in fig .4 , shows an

ensemble consisting of three classifier having multiple instances of heterogeneous type. N number of such ensemble can be formed with different ratio of instances of each classifier type, thus infer to achieve high diversity within the ensemble. For example three heterogeneous classifier having multiple instances of C4.5, neural network(MLP) and Naive Bayes classifier. The algorithm for heterogeneous ensemble:

HE is an iterative algorithm(Algorithm)[12], we made certain changes to original algorithm ,it starts with an heterogeneous ensemble which is . At each iteration one dataset is chosen for labelling and added to the training set, for better performance the properties of the ensemble are adaptable.

The stepwise execution of the algorithm is as follows:

Given: initial training set T , initial test set D_{te} ,
 $T = \{t^1, t^2, t^3, \dots, t^N\}$, $M, V_{in} = \{v_{in}^1, v_{in}^2, v_{in}^3, \dots, v_{in}^N\}$, S, M

Where ,

D_{te} -The testing set,

T -A list of classifier,

M -Initial Ensemble size\

V_{in} -Initial number of instances of classifier type,

W -Window size,

S -Stopping criteria ,

$D_{in} = T$ training set generated by RF.

Initialization: The training set $D_{tr} = D_{in}$, ensemble size $m = M$, $V = V_{in}$.

Algorithm : Ensembling of Classifier

Step 1: Randomly choose the classifier type with instances for Initial ensemble $C = \{c^1, c^2, \dots, c^M\}$ in which each Classifier t^i has v^i instances in C .

Step 2 :Set the Stopping criteria S .

Step 3: Record the vote Entropy for each data instance in t^1 in T in the current window.

Step 4.Store the ensemble configuration with Maximum Entropy .

Step 5. Query the label d_t

Step 6: Add d_t to D_{tr} with acquired label.

Step 7: for each t_i in T , C^r is the random no. of classifier instances of type t^i ,

ACC_i records the accuracy of current C ,

ACC_c records the accuracy of C_r ,

$P = \arg \max_i ACC_i$;

$q = \arg \min_i ACC_c$;

if $ACC_p > ACC_c$;

$v^p = t^p - 1$;

$v^q = t^q + 1$;

else remain the current V ;

end if;

Step 8: Train new ensemble on the set according to current size of the ensemble obtained from step 7 .

Step 9: Repeat step 3 to 8 till condition S is fulfilled.

In size adaptation phase the size of the ensemble is unchanged but instances of each classifier type may change in order to achieve high accuracy. In the adaptation phase two subset of the ensemble variety are created we will use some variation and simplification of algorithm [12].for example we choose an heterogeneous ensemble of C4.5 and Naive Bayes with three instances of each classifier binned

with a single ensemble as default then we create two variant of the default ensemble in which one instance of each classifier type is increase and other is decreased and in the second variant instances of latter classifier type in the ensemble decreased and other type is increase i.e. we record the accuracy of default ensemble then to obtain the first variant we will decrease one instance of C4.5 and increase one instance of the Naive Bayes and there accuracy are compared the variant with the highest accuracy is recorded and the classifier achieving the same would increase it's instances in the ensemble simultaneously decreasing the instance of classifier having lowest accuracy.

III. EXPERIMENTAL SETTING AND DATA SETS

This system works on real and categorical values the expected result is based on data set chosen from UCI[6] repository and other reliable sources[7].

Many Data set were short listed for analysis on the system out of which 5 dataset are pre-adapted and fed serially to system as per requirement specification .As this system is not compatible with data set having missing values therefore all the missing values of data set are filled by default value results can be seen in Table 9.2

Table 2 : Data sets

| Data sets | Size | Features | Classes | Iteration | Window Size |
|----------------|------|----------|---------|-----------|-------------|
| Iris | 150 | 4 | 3 | 250 | 30 |
| Glass | 214 | 10 | 7 | 450 | 50 |
| Liver Disorder | 345 | 7 | 3 | 240 | 40 |
| Kidney | 1250 | 6 | 2 | 360 | 70 |

The initial ensemble size is nine with three classifier instances of each base classifier C4.5,Naive Bayes and MLP(multi layer perceptron).The RF is implemented for creating the subset of the data set(s)[4].Algorithmically the number of instances of ensemble would change but the overall ensemble size remains the same. The ratio of New features to original feature(s) is kept K which is described in previous sections to generate new feature set containing $2 * k * d$ indices. The results for changed ensemble configuration recorded can be seen in table .4.Training time for this system will be more but testing is fast.

A) Vote Entropy: In this paper we have measured the necessary measurement using formulas given below,we have taken entropy measurement for accessing predictiondisagreement [8] among the ensemble member it measure by

$$VE = - \sum_{j=1}^M \frac{L_j}{M} \log \frac{L_j}{M}$$

Where,

L_j : Number of votes for the j_{th} output\

M : Committee size\

B) Matching Matrix: Matching Matrix evaluates the performance related measure like TPR, TNR, accuracy etc. major guiding formula are listed below,\

a) Sensitivity (TPR): Also Known as true positive rate evaluated as

$$TPR = TP / (TP + FN)$$

b) Specificity (TNR): Also known as true negative rate is measured as

$$TNR = \frac{TN}{FP+TN}$$

C) Accuracy (ACC): It is used for determining the correctness of prediction by classifier measure as

$$ACC = \frac{TP+TN}{FP+FN}$$

Where,

TP- Correct selection of true instances

TN- Correct rejection of false instances

FP- Incorrect selection of false instances

FN- Correct selection of false instances

D) Stopping Criteria

In HE number of iteration serves as the stopping criteria S determined by formula

$$\text{Number of iteration}(S) = \frac{\text{Number of data instances}}{\text{Window size}}$$

E) Improvement:

Improvement matrix [11] is used to indicate how much ESAHE system outperforms the homogeneous system. For each data set improvement is defined as:

$$\text{Improvement} = \frac{\text{Mean accuracy of ESAHE}}{\text{Mean accuracy of homogeneous methods}}$$

IV. EXPERIMENTAL RESULTS

This section introduces the data set used for MIHC-ES and then reports the experimental result reported by stably sized adaptive heterogeneous ensemble are analysed and compared with Multi instance heterogeneous ensemble with extended space forest(MIHC-ES) see(table. 3)are compared and tabulated and accuracy based graphical analysis is carried on three out of five pre-adapted data set characterised in table. 2.

A. Performance of classification system

From Table .3 it can be observed that the system achieves highest accuracy when training data is 60-70% of the original original data set also we can see that after extending the space MIHC_ES performs better than ensemble of homogeneous classifier C4.5 which is 64 to 66% appx. which achieves better accuracy than Naive Bayes classifier on the same data set, whereas ensemble of MLP beats C4.5 in terms of accuracy having 62 to 78% appx. accuracy but this performance is further improved with application of ESF in table .4 we can observe that Adaptive heterogeneous ensemble (AHE)under-performs when ESF is not included , evidently binding ESF with AHE algorithm improves the performance which records the highest accuracy in the experiment with achieving lowest 95% to the highest of 99.7 % which is good rational proving efficiency of combined ESF and AHE with rotation forest (MIHC_ES)over other ensemble taken in the experiment. We can observe from the graph that MIHC_ES outperforms other ensemble on some data set which lack in feature fusion of ESF with RF directs more improved results on such data sets..With default ensemble size six(table 3) MIHC_ES clearly out performs HE with single instance of each classifier type i.e C4.5 , Naive Bayes and MLP.All the data set taken in this experiment suffer from low dimension of feature space with maximum of seven attributes , so we can deduce that implementing data transformation operation such sa RF and ESF increase the classification accuracy.

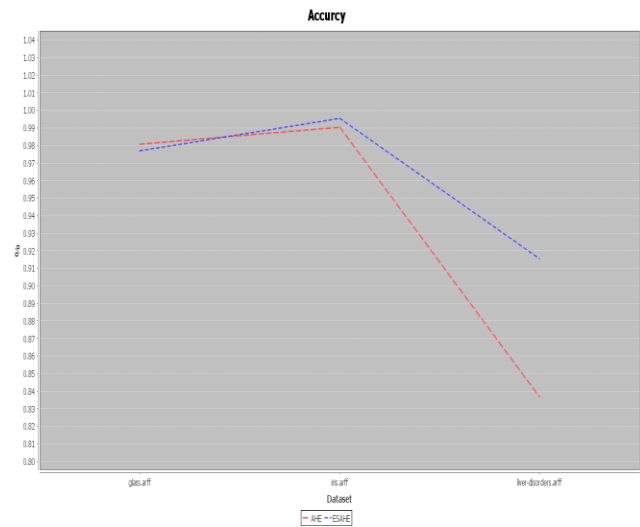


Figure 3. Graph showing accuracy comparison of heterogeneous ensemble(HE) without ESF and RF to multi instance heterogeneous ensemble (MIHC_ES) on three data set.

V. CONCLUSION AND FUTURE SCOPE

In This Experiment , a new system is derived for creating ensemble with different kinds of heterogeneous classifier are implemented .These classifier ensemble is self-adaptable for choosing the best ensemble configuration and has component for feature set generation using the original data set .

This system is useful in handling complex data, fusion of extended space forest system with adaptive heterogeneous system of classifier ensemble imparts the user with benefit of highly diverse analysis of data set with ensemble consisting three of the best classier C4.5, neural network and naive bayes binned within a single ensemble with multiple instances of each classifier.

This system overcome the problem in analysing data set with very few feature, extended space forest is especially integrated in ESAHE as a component that this system does not suffer from low dimension of feature set as it is capable of generating new feature thus, the feature set as it is capable of generating new features thus, the data set under analysis can be extended .From the experiment it is evident that high classification accuracy on data set with very few feature set such as lung cancer, urine , kidney stone therefore we can conclude that this system can be used in bio medical analysis and research. The further enrichment of the systemcould be to add component such as analyser for the object of arbitrary type like modern automated methods to deal with drastically increasing complex data from industries and our system is able to analyse these data.

Moreover this system can be enhanced top a dynamic system that can handle elemental dynamics and relationship between the features of data set also more research can be done towards fast and transparent data analysis.

Table 3: Classifier ensemble performance comparison homogeneous ensemble of classifier C4.5 , Naïve Bayes and

MLP with three instances in each run and heterogeneous ensemble SSHE and MIHC_ES over five data set.

| Ensemble classifiers performance sheet | | | | | | |
|---|-----------------------|-----------------------------|-------------|------------------------|--------------------|-----------------|
| Classifier | Data Set | Classifier instances | Mean | <u>Std dev.</u> | Improvement | Accuracy |
| C4.5 | <i>Liver disorder</i> | 3 | .56 | .66 | 1.06 | .640 |
| | <i>Urine</i> | 3 | .54 | .34 | .987 | .627 |
| | <i>Glass</i> | 3 | .58 | .44 | 1.01 | .940 |
| | <i>Iris</i> | 3 | .32 | .56 | .987 | .560 |
| | <i>kidney</i> | 3 | .39 | .54 | .996 | .658 |
| Naïve Bayes | <i>Liver disorder</i> | 3 | .27 | .96 | .867 | .651 |
| | <i>Urine</i> | 3 | .87 | .76 | .895 | .698 |
| | <i>Glass</i> | 3 | .62 | .44 | .971 | .499 |
| | <i>Iris</i> | 3 | .27 | .29 | .567 | .960 |
| | <i>kidney</i> | 3 | .56 | .27 | .567 | .564 |
| MLP | <i>Liver disorder</i> | 3 | .71 | .56 | 1.04 | .627 |
| | <i>Urine</i> | 3 | .44 | .54 | 1.67 | .827 |
| | <i>Glass</i> | 3 | .66 | .58 | 1.23 | .546 |
| | <i>Iris</i> | 3 | .34 | .32 | 1.01 | .787 |
| | <i>kidney</i> | 3 | .44 | .39 | .786 | .786 |
| HE | <i>Liver disorder</i> | 1,1,1 | .96 | .39 | .545 | .987 |
| | <i>Urine</i> | 1,1,1 | .76 | .56 | .735 | .698 |
| | <i>Glass</i> | 1,1,1 | .44 | .87 | .743 | .964 |
| | <i>Iris</i> | 1,1,1 | .39 | .62 | 1.05 | .985 |
| | <i>kidney</i> | 1,1,1 | .56 | .27 | 1.34 | .658 |
| MIHC_ES | <i>Liver disorder</i> | 2,1,3 | .74 | .44 | 1.04 | .956 |
| | <i>Urine</i> | 2,1,3 | .67 | .66 | .991 | .932 |
| | <i>Glass</i> | 3,2,1 | .29 | .34 | 1.23 | .932 |
| | <i>Iris</i> | 2,1,3 | .27 | .44 | 1.06 | .936 |
| | <i>kidney</i> | 3,1,2 | .45 | .29 | 1.70 | .997 |

REFERENCES

- [1] L. Rokach, "Taxonomy for Characterizing Ensemble Methods in Classification Tasks: A Review and Annotated Bibliography," *Computational Statistics Data Analysis*, vol. 53, no. 12, pp. 4046-4072, 2009.
- [2] L. Breiman, "Bagging Predictors," *Machine Learning*, vol. 24, no. 2, pp. 123-140, 1996. L.S. Jacobs and C.P. Bean, "Fine particles, thin films and exchange anisotropy," in *Magnetism*, vol. III, G.T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271-350.
- [3] O.T. Yildiz and E. Alpaydin, "Omnivariate Decision Trees," *IEEE Trans. Neural Networks*, vol. 12, no. 6, pp. 1539-1546, Nov. 2001. R. Nicole, "Title of paper with only first word capitalized," *J. Name Stand. Abbrev.*, in press.
- [4] T.K. Ho, "The Random Subspace Method for Constructing Decision Forests," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 20, no. 8, pp. 832-844, Aug. 1998. M. Young, *The Technical Writer's Handbook*. Mill Valley, CA: University Science, 1989.
- [5] M.F. Amasayali and k Ersoy, "Classifier Ensemble with Extended Space Forest" *IEEE Trans. Knowledge and Data Engineering*, vol. 26 March 2014
- [6] H. Chen and X. Yao, "Multiobjective neural network ensembles based on regularized negative correlation learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22(12), pp. 1738 – 1751, 2010.
- [7] Springer support data (sep, 2011)," Retrieved from "www.springeropen.com/about/supportingdata..
- [8] J. Du and C. X. Ling, "Asking generalized queries to domain experts to improve learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22(6), pp. 812–825, 2010.
- [9] M. Lindenbaum, S. Markovitch, and D. Rusakov, "Selective sampling for nearest neighbor classifiers," *Machine Learning*, vol. 54(2), pp. 125–152, 2004.
- [10] L. Breiman, J. Friedman, R. Olshen, and C. Stone, "Classification and Regression Trees". Chapman Hall, 1984.
- [11] S. Dzeroski and B. Zenko, "Is combining classifiers with stacking better than selecting the best one?," *Machine Learning*, vol. 54(3), pp. 255–273, 2004.
- [12] "Confusion Matrix", (sep, 2015) Retrieved from "en.wikipedia.org/wiki/Confusion_matrix.
- [13] Zhenyu LU and Xindong Wu and C. Bongard, "Active Learning Through Adaptive Heterogeneous Ensembling" *IEEE Trans. Knowledge and Data Engineering*.
- [14] An Experimental Study on Rotation Forest Ensembles. (2010, October 13). Retrieved from <http://pages.bangor.ac.uk/~mas00a/papers>.