International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

Advanced Methodologies Employed in Ensemble of Classifiers: A Survey

Vikas Singh[1], Madhavi Ajay Pradhan[2]

[1]Department of Computer Engineering, AISSMSCOE, Kennedy Road, Pune University, Pune, Maharashtra, India

Abstract: ***If we look a few years back, we will find that ensemble classification model has outbreak many research and publication in the data mining community discussing how to combine models or model prediction with reduction in the error that results. When we ensemble the prediction of more than one classifier, more accurate and robust models are generated. We have convention that bagging, boosting with neural network etc. are the most popular method of combining different models and are realized in many data mining software but there are variation and alternative to bagging and boosting. This survey paper will give insight into various newly proposed ensemble classification models based on different methodologies.***

Keywords: Ensemble classifier, bagging, boosting, classification, neural network

1. Introduction

Ensemble learning refers to a collection of methods that learn a target function by training a number of individual learners and combining their predictions. When you can build component classifiers that are more accurate and more importantly, that are independent from each other because uncorrelated errors of individual classifiers can be eliminated through averaging. As we know data is generating day by day in large amount and availability of data both structured and unstructured data, this extensive data availability if well utilized then more accurate analyses may lead to more confident decision making. And better decisions can mean greater operational efficiencies, cost reductions and reduced risk. This paper also examine the working and performance of ensemble classifier on high dimensional , imbalance and microarray data to converge its application in various field science with diversified approach .This paper gives a overview of some latest technique in ensemble classifier like online class imbalance learning which is relatively new in the field of data mining along with this typical high dimensional data classification like microarray data classification along with it active heterogeneous Ensembling is also discussed. This paper is organized as follows Section 2 briefly describe some of the latest methodology that that are used in ensemble predictive classifier for binary and multiclass classification problem Section 3 discuss the findings and Section 4 gives concluding remarks

2. Advanced Methodologies Used In Ensemble Classifier

2.1 Extended Space Forest

Mehnet and Ersoy [1] in Classifier Ensemble With The Extended Space Forest proposed extended space forest algorithm for construction of decision tree. Proposed algorithm (ENS) uses bagging (BG) [2] , ,random subspace(RSs) [3] , random forest [4](RFs)and rotation forest [5] and boosting [7] as ensemble. The major highlight of this methodology is that it add new feature set to the original dataset .This task accomplished trough permuting and pairing original feature set .The major advantage of this methodology comes with the choice of ensembling different classifier and base learner. This methodology has adopted principle component analysis (PCA) [6] that is why only numeric data can be analyzed.

Given:

E= { x_p , y_p } =[X Y] , where p={1....N}

Where X is an N*d matrix containing the training set and Y is an N- dimensional column vector. Feature set containing the class labels. D is the number of features. N is the number of training samples.

Initialization:

Choose the ensemble size T , the ratio of number of new feature to number of original feature set K , the feature generation operator OP , the base learner model L and the ensemble algorithm ENS.

Table 1: The genetic algorithm for building an extended space forest**.**

Training: For i = 1:T
1. Create new feature (EX_i) by using randomly paired original features.
Generate 2*K random permutation of the original feature indices. Concatenate them and store in C_i .(C_i have 2*K*d indices)
J=1
For z = 1:2*K*d step by 2
Create j^{th} new feature applying OP to$C_i(z)^{th}$ and $C_i(z+1)^{th}$ features of X matrix.
J=j+1
EndFor
Construct the new training set (E_i) by concatenating the matrix X (original feature) and EX_i(the new features) as E_i=[X EX_iY]
Train $L_i$with E_i according to Ensemble algorithm (ENS)
EndFor
Testing :
For i=1: t
1. Extend the Feature space of the test sample (x) by using the feature pairs in C_i
2. Classify the extended sample with L_i.
EndFor
Combine the base learners' decision by combination rule of the ensemble algorithm ENS.

In this algorithm T is the number of base learners and L_T is the base learner, E_i is the extended training set for L_i ,C_i consist of feature pair indices is used in generation of E_i and ENS ensemble algorithm .New training feature set are obtained from the original feature set and each base learner is trained with different training set. In extended space forest new feature set is generated using linear and non linear transformations etc.. And also some feature generating operator which are given in

Table: 1 the parameters used are as follows
Let there be **d** number of feature then,
W1: d*d dimension matrix generated from uniform distribution on (-1 ,1)
W2: d*d dimensional matrix with zero mean Gaussian distribution.
Dot products of w1 and w2 numbers generated from uniform distribution (-1,-1).The operator like sum, difference, comparison, divide, multiply etc are used. These operators are applied two paired original feature using two original features.

Operator Name	Equation
Sum	NewFea=feaX+feaY
Difference	NewFea=feaX-feaY
Comp	NewFea=feaX>feaY
Divide	NewFea=feaX/feaY
Divide tanh	NewFea=tanh(feaX/feaY)
Multiply	NewFea=feaX*feaY
Multiply tanh	NewFea=tanh(feaX*feaY)
Single	NewFea=feaX>w1
Two linear	NewFea=w1*feaX+w2*feaY
Uniform random matrix	NewFeaS=fea*W1
Gaussian random matrix	NewFeaS= fea*W2
Mix	NewFea= (feaX/feaY) or (feaX-feaY)

Figure1: Feature Generating Operator

Random sorting is in practice for generation of new feature as shown in Table.1 two original features are used to generate third new feature by using mahtematical operators.For example if there are d number of features then d/2 new features will be generated.

2.2 Resampling based Ensemble

A proposed system by Shou wang and minku [8] in ensemble methods for online class imbalance learning authors proposed two learning algorithm that addresses the problem of online learning and class imbalance learning . This algorithm WEOB1 AND WEOB2 proves its effectiveness against data streams having very skewed class distribution . We see how the two algorithm over sampling based online bagging (OOB) and under sampling based UOB [9] is improved further effectively and how the new algorithm administer imbalanced static and dynamic data stream [10].The author also discusses the scope of updating the old data with new one in the case of dynamic data(online bagging)[11] . Now we will see OOB with adaptive weight WEOB1 [8] first and then UOB with adaptive weight WEOB2 [8] with taking into consideration the major changes that has been introduced in OOB and OUB[6].Now we look at this algorithm in step by step manner.The algorithm takes an ensemble M as input base learner and training example (x_t, y_t) and current class size

$$w^{(t)} = (w_+^{(t)} , w_-^{(t)}) \quad (1)$$

Where,
t – The training step
$w_+^{(t)}$ - The size of positive class .
$w_-^{(t)}$ - The size of negative class .
y^t –The value of **y**
K- The number of times this example is used for training

Table 2: .*Alogorithm for Adaptive weight over* and under sampling

If
y^t=+1 and $\begin{cases} w_+^t < & w_-^t \text{ for OOB} \\ w_+^t > & w_-^t \text{ for OUB} \end{cases}$
Then
Poison ratio ***k*** is set to : (w_-^t / w_+^t)
Else if
y^t=-1 and $\begin{cases} w_-^t < & w_+^t \text{ for OOB} \\ w_-^t > & w_+^t \text{ for OUB} \end{cases}$
Then
Poison ratio k is set to: (w_+^t / w_-^t)
Else
set $k \sim$ Poisson ($\gamma = 1$)
Update
f_m K times.
End

The major changes from OOB to WEOB1 is that in OOB if the training example in OOB belongs to the minority class OOB increases the value of K i.e. the number of times the example is used for training . Similarly if the training example belongs to majority class then UOB decreases the value of K [9], K follows Poisson distribution. Whereas in WEOB1and WEOB2 [8] the value of K is the ratio of minority class to majority class if the majority class is smaller than minority class at the current time for WEOB1 and ratio of majority to minority in WEOB2 if majority class if minority class is smaller than minority class. There are many advantages of improved OOB AND OUB version like time decaying class size [9], which estimates the imbalance status without storing old data and adaptively decides re-sampling rate. Other major advantage choice of choosing ensemble method's and also re-sampling of data is algorithm independent.

2.3 Thiknov Regularized Least Square

In Microarray data classification Using the Spectral feature based TLS Ensemble algorithm Zhan and wang [10] proposed an Ensemble algorithm Tikhnov Regularized Least Square (TLS). TLS is employed for cancer classification using gene expression (A process by which information from a gene is used in the synthesis of functional gene product) data.

The authors also proposed two types of dictionaries namely Singular Value Decomposition based eigenassays [11] (SVD) and independent component analysis based eigenassays [12] (ICA) which will be used in TLS algorithm

We will further study algorithm and proposed framework. Firstly we will look at TLS classifier with frame work then the two dictionaries SVD and ICA .The data is microarray data which has sparse representation (SR)[13] for sparse signal reconstruction. Sparse representation based classification [14] (SRC) is applied to SR. The principal of SRC is that when sufficient numbers of training examples from same class are available then test sample can be characterized by using only training samples from same class.

Now we will look at the TLS algorithm and framework in brief firstly we will how spectral feature image is represented. Let a **p×n** matrix and X^0 denotes gene expression with **p** gene and **n** samples. The element X_{ij}^0 the i^{th} row of r_i of X^0 denotes the expression profile of i^{th} gene and the j^{th} Columns c_j is the snapshot of j_{th} assay (protein to be analyzed). Images can have arbitrary row and column for simplicity images are reshaped as square images. Each sample from X^0 is from **1D** from signal to **2D** image. In **fig.2** shows transformation from **gene** sample **c** to image **I.** If the sample is not long enough then inverse of symmetry filling is performed which is described in figure 2.

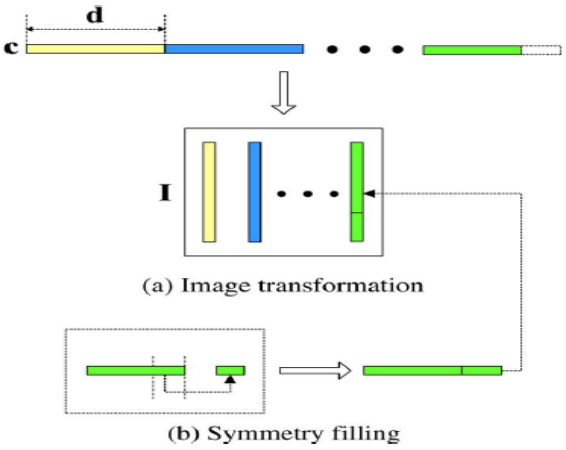

Figure 2: Image Transformation and symmetry filling

Moving further now we will look at the Dictionary Extraction Of Spectral Feature Images. The whole TLS model can be said to really on Dictionary Extraction. here the authors have presented two types of spectral feature based dictionaries : eigenassays computed through singular value decomposition (SVD) and eigenassays obtained via independent component analysis (ICA). In SVD matrix can be represented as

$X_i = U^i A_i$ [10]

where ,

A_i- is considered as coefficient having enough samples for the training class

U^i -is subspace spanned by columns of U^i which will represent any sample belonging to the i^{th} of A_i.

The 2D matrix can be converted to 1D matrix by inverse process of fig.1. Whereas the second method ICA -based eigenassays [10] is formulated as

$X_i^T = A_i S_i^T$,

Where,

X_i^T- is a linear mixture of m_i statistically independent basis snapshot of eigenassays $s_{\tilde{j}}$. These two method as given in figure 3 are used for extracting a special feature based dictionary from gene expression data.

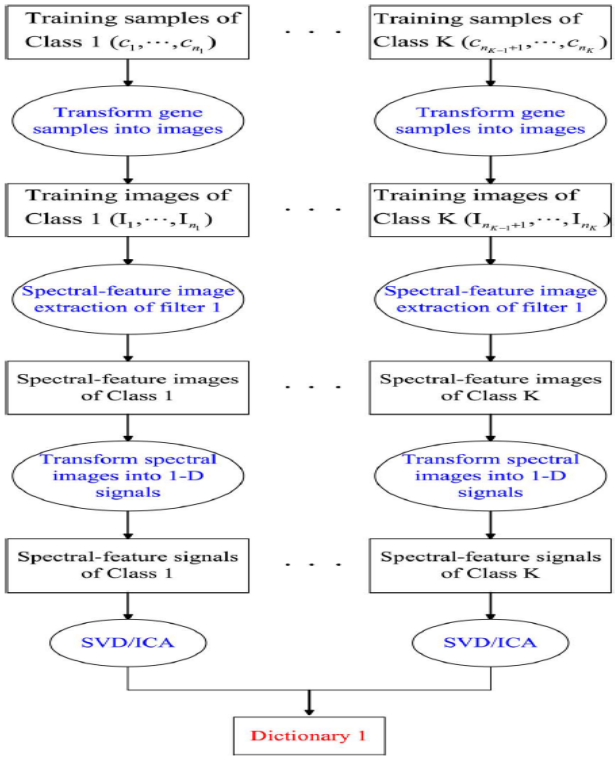

Figure 3: Extraction of a Special Feature Based Dictionary

Now we will look at TLS ensemble model, TLS algorithm is governed by the equation **z** = $\emptyset\alpha$+**v**, where $\emptyset$ is the dictionary formed by either SVD/ ICA and Z is spectral feature which can be modelled as linear representation of the fragments of $\emptyset$.

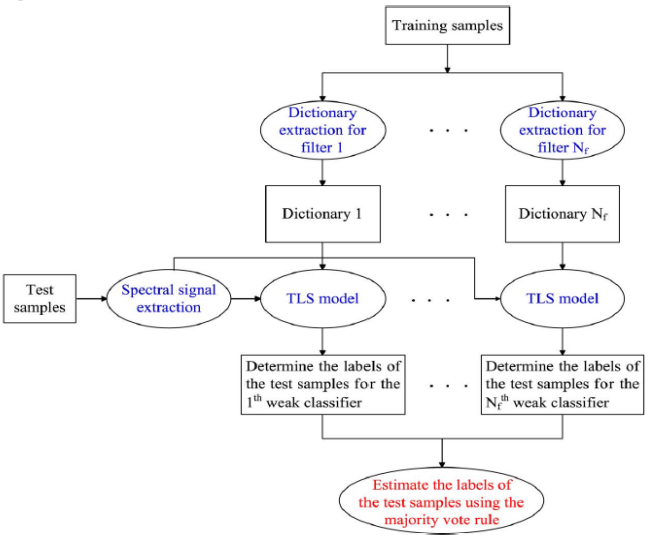

Figure 4: Framework for TLS Ensemble Algorithm

The training samples are fed to dictionary extraction filter and spectral signal are extracted from the test samples as shown in Figure 4 which then are delivered to the model which determines the labels for the test samples.

2.4 Adaptive Heterogeneous Ensembling

Zhenyu and Xindong in Active Learning Thorough Adaptive Heterogeneous Ensembling proposed Adaptive

Heterogeneous Ensemble framework (AHE) [15] framework. Many studies have already shown that blend of heterogeneous ensemble overtake homogenous ensemble in terms of classification accuracy like bagging boosting, RSM etc. [**16**]. There are three variant of AHE algorithm first is stably sized for AHE(SSAHE) and second is variably sized AHE (VSAHE)[**17**]. The difference between AHE and SSAHE is the adaptation phase where the ratio of instances of each classifier type is adapted. Whereas the in VSAHE, adaptation phase the overall ensemble size is changed along with the number of instances of each type .First we will look at AHE algorithm. Adaptive heterogeneous ensemble consists of multiple classifier instances of multiple types. We will look at the algorithm in step by step manner

- **Input**-Following input are provided to ensemble framework

D_{pool}– Training pool,
D_{te}– The testing set,
D_{ad} - The adaptation set,
d_{in}-The Initial training set,
T={t^1, t^2,….t^N }- The initial ensemble of size M
V_{in} ={ v_{in}^1 , v_{in}^2 …. v_{in}^N}-Initial number of instances of classifier
$t^{(i)}$ - Classifier type in the ensemble,
W- Window size,
S- Stopping criteria

- **Initialization** –The training set D_{tr} = D_{in} , the ensemble size m=M , V=V_{in}, train the initial ensamble C={$c^{(1)}$, c^2,… … …, $c^{(M)}$} on D_{tr} , each classifier type $t^{(i)}$ has $v^{(i)}$ instances.

Table 3: AHE Algorithm

Algorithm

while not S:
1 for each data instance d_i from D_{pool} in the current window:
record the vote entropy of C as VE_i;
t = $argmax_i$VE_i;

query the label of d_t;
add d_t with the acquired label to D_{tr};
2 through adaptation update parameters of the ensemble such as V or m;
3 train a new ensemble C on D_{tr} according to the current V ;
C = C';
end while;

Output: The ensemble C trained according to latest *V*.

In the above algorithm as shown in Table: 3 the training is made available to algorithm in streaming manner. The new data point in each iteration is chosen by dividing the training set into chunks of equal size. AHE starts with an initial heterogeneous ensemble, where each classifier type is started with same number of instances. This algorithm has three iteration phases. In the first phase of the iteration the initial ensemble makes prediction on the each of the first chunks of data instances of the training pool. The data instances that causes maximum disagreement is among the ensembles member is chosen for ensemble quarrying. This dataset with its label is added to training set. The second phase is adaptation phase in which each in which number of instances of each classifier type is updated .In the third phase of the iteration , the current classifier are discarded and new and new ensembles is trained on the updating training set. Stably sized adaptive heterogeneous ensemble is the first variant of AHE framework, all the steps and input remains the same except the second phase of training iteration, variants of the current ensemble is made. We will look at the algorithm now which will replace the step 2 of previous algorithm.

Table 4: Algorithm for stably sized AHE

Algorithm

* 2 for each T^t in T:
c^r is a random classifier of type t^t ;
record the accuracy of C \ c^r on D_{ad} as ACC_C;
end for;
record the accuracy of C on D_{ad} as ACC_C;;
p = $\arg max_i$ ACC_i;
q = $\arg max_i$ ACC_i;
if ACC_p> ACC_C:
v^p = t^p – 1;
v^q = t^q + 1;
else: remain the current V ;
end if;

For each classifier type, one member of its type is randomly chosen to be taken out of the whole ensemble. All reduced ensembles, as well as the original ensemble, are then tested on the adaptation set for their accuracies If one variant achieves the highest accuracy, then the number of classifiers of its corresponding type is decreased by one, because it is expected to increase the accuracy of the ensemble. If one variant achieves the lowest accuracy, its corresponding type is increased by one. If no variant is more accurate than the original ensemble, then the current configuration is retained. The choice of searching all the variants of the current ensemble that has one less classifier is to keep the search effort manageable. Variably Sized Adaptive Heterogeneous Ensemble: The Variably-sized Adaptive Heterogeneous Ensembles is different from the Stably-sized Adaptive Heterogeneous Ensembles because in its adaptation phase the overall ensemble size is changed along with the number of instances of each type.

Algorithmically In its adaptation phase, two subsets of ensemble variants are created. In the first subset, each classifier is taken out of the original ensemble. In the second subset, a new instance is created for each classifier type, and then added to the current ensemble.

Now, we will look at the algorithm and infer our deduction all phase remained same as above but adaption phase will have approach of chaining the ensemble size in accordance to the training data.

Table 5: Algorithm for variably sized AHE

Algorithm

type(*i*) returns the corresponding classifier type for the i_{th} classifier;

* 2 for each $c^{(i)}$ in C:
record the accuracy of C \ $c^{(i)}$ on D_{ad} as $ACC_i^{(-)}$;
end for;
for each t^i in T:
c^r is a new instance of t(i) trained on D_{tr};

$C' = C^{r} + C$;
record the accuracy of C^{i} as $ACC_i^{(+)}$
end for;
record the accuracy of C on D_{val} as ACC_c;
p = type($argmax_i ACC_i^{(-)}$)
q = $argmax_i ACC_i^{(+)}$
if $ACC_p > ACC_c$ and ACCp > ACCq:

$t_p = t_p - 1$, m = m − 1;
else if $ACC_q > ACC_p$ and $ACC_p > ACC_c$
$t^q = t^q + 1$, m = m + 1;
else: remain the current V ;

end if;

All variants and the current ensemble are tested on the adaptation set for accuracy. If one variant with reduced size achieves the best accuracy, then the number of classifiers of its corresponding type is decreased by one. The numbers of instances of other types remains the same, thereby leading to a reduction of one in the size of the ensemble. If one variant with increased size achieves the best accuracy, then the number of classifiers of its corresponding type is increased by one. The numbers of instances of other types remains the same, thereby leading to an addition of one in the size of the ensemble. If the current ensemble achieves the highest accuracy, the size and internal ratio of classifier types remain unchanged.

3. Findings

In this paper we have reviewed four ensemble methodology used for classification. In first methodology we have seen ENS outperforming other ensemble by creating extended feature set which improves classification accuracy and also generates smaller (simple) base learner. Whereas in second methodology resampling based ensemble outperforms other algorithm in statistical test. Third methodology for classification is proposed in heterogeneous Ensembling method the basic advantage of this method logy is use of algorithmically different type of classifier .the TLS classifier has shown its efficiency and accuracy for microarray data classification, which has scope in biomedical sciences.

Classifier Types	C45		NB		KNN		Improvement
Data Sets	mean	std	mean	std	mean	std	
letter	0.71	0.05	0.11	0.04	0.19	0.04	1.068
mfeat-pixel	0.37	0.07	0.2	0.05	0.43	0.07	1.056
isolet	0.58	0.07	0.18	0.04	0.24	0.05	1.05
mfeat-karhunen	0.43	0.08	0.2	0.06	0.37	0.07	1.05
mfeat-zernike	0.31	0.1	0.23	0.08	0.46	0.09	1.047
mfeat-fourier	0.53	0.09	0.25	0.05	0.22	0.06	1.041
segment	0.54	0.09	0.12	0.08	0.34	0.1	1.032
spambase	0.57	0.12	0.15	0.06	0.28	0.1	1.031
mfeat-factors	0.53	0.08	0.16	0.09	0.31	0.09	1.024
pendigits	0.37	0.1	0.06	0.05	0.57	0.09	1.023
optdigits	0.32	0.09	0.15	0.05	0.53	0.08	1.02
splice	0.44	0.09	0.41	0.1	0.15	0.07	1.019
landsat	0.49	0.1	0.08	0.05	0.44	0.09	1.019
page	0.61	0.12	0.2	0.11	0.18	0.11	1.012
waveform(2)	0.34	0.09	0.34	0.07	0.32	0.09	1.009
mus	0.66	0.15	0.15	0.1	0.2	0.16	1.004
madelon	0.34	0.17	0.31	0.13	0.36	0.19	0.996
magic	0.4	0.11	0.21	0.08	0.39	0.11	0.993
average percentage — rank correlation	0.474	-0.75	0.195	-0.65	0.332	-0.75	

Figure 5: Comparison of AHE vs. Other classifiers [15]

4. Conclusion

From above survey we can conclude that ensemble classifier is efficient and classification is more accurate. Also, the algorithms which include the modified bagging and boosting for decision tree generation are more reliable and accurate than the conventional algorithms.

References

[1] Mehmet Faith Amasyali and Okan K.Eesoy ”Classifier Ensemble with the Extended Space Forest” 2014 IEEE.
[2] L. Breiman, “Bagging Predictors,” Machine Learning, vol . 24, no. 2,, 1996.
[3] 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.
[4] L. Breiman, “Random Forests,” Machine Learning, vol. 45, no. 1, pp. 5-32, 2001.
[5] J.J. Rodriguez and C.J. Alonso, “Rotation-Based Ensembles,” Proc. 10th Conf. Spanish Assoc. Artificial Intelligence, pp. 498-506, 2004.
[6] C.-X. Zhang and J.-S. Zhang, “A Novel Method for Constructing Ensemble Classifiers,” Statistics and Computing, vol. 19, no. 3,pp. 317-327, 2009.
[7] Y. Freund and R.E. Schapire, “Experiments with a New Boosting Algorithm,” Proc. 13th Int’l Conf. Machine Learning, pp. 148-156,1996.
[8] Shuo Wang,” Resampling-Based Ensemble Methods for Online Class Imbalance Learning” 2014 IEEE.
[9] S. Wang, L. L. Minku, and X. Yao, “A learning framework for online class imbalance learning,” in IEEE Symposium on Computational Intelligence and Ensemble Learning (CIEL), 2013, pp. 36–45
[10] Zhan-Li Sun and HanWang,” Microarray Data Classification Using the Spectral-Feature-Based TLS Ensemble Algorithm” 2014, IEEE.
[11] P. Maji, “Fuzzy-rough supervised attribute clustering algorithm and classification of microarray data,” *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 41, no. 1, pp. 222–233, 2011.
[12] Z. L. Sun, D. Rajan, and L. T. Chia, “Scene classification using multiple features in a two-stage probabilistic classification framework,” *Neurocomputing*, vol. 73, no. 16–18, pp. 2971–2979, 2010
[13] R. Rigamonti, M. Brown, and V. Lepetit, “Are sparse representations really relevant for image classification?,” in *Proc. 24th IEEE Conf. Computer. Vis. Pattern Recog.*, 2011, pp. 1545–1552.
[14] L. Zhang, M. Yang, and X. Feng, “Sparse representation or collaborative representation:Which helps face recognition?,” in *Proc. Int. Conf.Comput. Vis.*, 2011, pp. 471–478.
[15] Zhenyu Luand Xindong Wu” Active Learning Through Adaptive Heterogeneous Ensembling”2014, IEEE
[16] L. I. Kuncheva and J. J. Rodriguez, “Classifier ensembles with a random linear oracle,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 19(4), pp. 500 – 508, 2007.
[17] I. H. Witten and E. Frank, *Data Mining: Practical machine learning tools and techniques*. San Francisco: Morgan Kaufmann, 2005.