Influence of Unstable Patterns in Layered Cluster Oriented Ensemble Classifier

Ashfaqur Rahman
Intelligent Sensing and Systems Lab, CSIRO
Hobart, Australia
Email: ashfaqur.rahman@csiro.au

Brijesh Verma
Central Queensland University
Rockhampton, Australia
Email: b.verma@cqu.edu.au

***Abstract*— In this paper, we have investigated the influence of cluster instability on the performance of layered cluster oriented ensemble classifier. The final contents of clusters in some clustering algorithms like *k*–means depend on the initialization of clustering parameters like cluster centres. Layered cluster oriented ensemble classifier is based on this philosophy where the base classifiers are trained on clusters generated at multiple layers from random initialization of cluster centres. As the data is clustered into multiple layers some patterns move between clusters (unstable patterns). This instability of patterns brings in diversity among the base classifiers that in turn influences the accuracy of the ensemble classifier. There is thus a connection between the instability of the patterns and the accuracy of layered cluster oriented ensemble classifier. The research presented in this paper aims to find this connection by investigating the influence of unstable patterns on the overall ensemble classifier accuracy as well as diversity among the base classifiers. We have provided results from a number of experiments to quantify this influence.**

***Keywords*—ensemble classifiers, committee of experts, multiple classifier systems, cluster oriented ensemble classifier**

I. INTRODUCTION

Ensemble of classifiers [1][2] refers to a collection of classifiers that learns the class boundaries within a data set simultaneously. The individual classifiers are commonly called base classifiers whereas the combination is called ensemble classifiers, mixture of experts, and multiple classifier systems. The combination of classifiers aims to achieve better classification accuracy compared to their base counterparts. In order for the ensemble to perform better than individual classifiers it is required that (i) the individual classifiers are accurate, and (ii) the errors made by individual classifiers are complementary i.e. the base classifiers are diverse among them [1][3]–[6].

Aiming to achieve diversity a number of ensemble classifier generation methods are found in the literature. Among different categories of ensemble classifiers the one that is relevant to our research creates diverse base classifiers by manipulating the training set. The idea is to train base classifiers on different subsets of the data. In bagging [7]–[13] the subsets are randomly drawn (with replacement) from the training set. Boosting [14]–[20] is a hierarchical process where the first subset is created by randomly drawing patterns from the training set. The patterns that are not correctly classified by the current classifiers are given more importance by the classifiers in the next passes. AdaBoost [16] is a more generalized version of boosting.

Another commonly used practice for generating ensemble classifiers uses clustering to partition data and trains base classifiers on clusters [21]–[29]. Layered cluster oriented ensemble classifier [31] trains base classifiers on non–overlapping clusters of a data set. In order to bring diversity among the base classifiers, clustering is done at multiple layers where the cluster centres are randomly initialized at each layer. As the cluster contents are dissimilar at different layers, the base classifiers are trained on different training subsets of the data. This brings in diversity among the base classifiers. When the data is clustered into multiple layers, some patterns move from one cluster to another. We call them unstable patterns. Patterns that do not move between clusters at different layers are called stable patterns. These unstable patterns contribute to altered compositions of clusters at different layers and thus diversify the base classifiers. What remains a question is to what extent the unstable patterns influence accuracy and diversity of layered cluster oriented ensemble classifier. Motivated by this fact, we aim to investigate the following research questions regarding the layered cluster oriented ensemble classifier: (i) influence of percentage of unstable patterns on diversity and accuracy on layered cluster oriented ensemble classifier, and (ii) the relative contribution of stable and unstable patterns towards total ensemble classifier accuracy.

The paper is organized as follows: Section 2 presents layered cluster oriented ensemble classifier. The quantification process of unstable and stable patterns is presented in Section 3. Experimental setup including data sets and base classifier configurations is presented in Section 4. Experimental results and discussion are presented in Section 5. Section 6 concludes the paper.

II. LAYERED CLUSTER ORIENTED ENSEMBLE CLASSIFIER

Clustering algorithm like k–means has a property that the final contents of the cluster depend on the initialization of the cluster centres for a given value of k [30]. This is evidenced from the following example where the data set in Figure 1(a) is divided into three clusters using k–means clustering algorithm with random initialization of the cluster centres. This process is repeated three times and the final cluster centres for these three passes are presented in Figure 1(b) to Figure 1(e). Note that the cluster centres are changing. This indicates that some patterns in the data set are changing clusters at different passes whereas some remains fixed in their original cluster. Let's call each pass a layer. Now consider training base classifiers on the clusters in all the three layers. As the cluster centres in all the three layers are different, the training sets of all the base classifiers are also different. This brings in diversity among the base classifiers and we followed this basic principle to create layered cluster oriented ensemble classifier [31].

The ensemble classifier generation and prediction method used in layered cluster oriented ensemble classifier is presented in Figure 2. During learning the data set is partitioned into N clusters based on random initialization of the cluster centres using k–means clustering algorithm. The process is repeated L times to create L independent partitioning of the data set into N clusters. Base classifiers are now trained on the clusters at each layer. As the clusters are composed of different patterns at different layers, the training set for each base classifier is different and this brings in diversity among the base classifiers of different layers. During prediction, the nearest cluster is first identified at each layer for a test pattern. The corresponding base classifier in each layer provides a decision of the pattern. The L decisions are then merged using majority fusion to obtain the final classification verdict.

Layered cluster oriented ensemble classifier is shown to improve accuracy and diversity [31] with increasing number of layers. However the following research issues remained open that we have investigated in the research presented in this paper:

- How many patterns do change clusters as the data sets are partitioned into multiple clusters? The patterns which change clusters are called unstable patterns whereas those remaining within their clusters are called stable patterns.
- How much influence do the unstable patterns have on the overall accuracy of the ensemble classifier?
- Which patterns do contribute more towards total accuracy? Stable or the unstable patterns?

III. STABLE/UNSTABLE PATTERNS

A. Philosophy

We define *unstable* patterns as the ones that change clusters when a data set is partitioned into say N clusters at L different layers. The patterns that do not change their clusters are called *stable* patterns. The simplest approach on identifying unstable patterns would be to list the cluster labels of the patterns obtained from successive execution of the clustering algorithm (Figure 3(a)) and identifying the cases where the cluster labels change. In the example shown in Figure 3(a) the fourth pattern changes cluster in successive runs and can be considered unstable. A problem with this approach of identification is that the successive executions of clustering algorithm are independent and thus the cluster labels may not remain the same in two consecutive runs. This is portrayed in Figure 3(b). Note that the cluster labels have been swapped in two consecutive runs. The patterns in fact have not changed clusters and are stable. The earlier detection approach will declare all the patterns as unstable patterns.

It is thus required to establish an association between the cluster labels first. This can be done by computing a *co–occurrence matrix* between the cluster labels. Given that there are two clusters, the dimension of the co–occurrence matrix will be 2×2. The entry (i, j) in a co–occurrence matrix refers to the number of patterns that has label i in Pass1 and label j in Pass 2. The computation can be explained from the co–occurrence matrices presented in Figure 3(d)–Figure 3(f) that are obtained from the tables presented in Figure 3(a)–Figure 3(c). In Figure 3(a) all the four patterns in cluster 1 belong to cluster 1 at consecutive passes. The entry $(1,1)$ in Figure 3(d) is thus 4. Similarly five patterns in cluster 2 belong to cluster 2 at consecutive passes in Figure 3(a). The entry $(2,2)$ in Figure 3(d) is thus 5. Only one pattern in cluster 2 belongs to cluster 1 at consecutive passes in Figure 3(a). The entry $(2,1)$ in Figure 3(d) is thus 1.The co–occurrence matrices in Figure 3(e) and Figure 3(f) are computed in a similar way.

Once we have computed the co–occurrence matrix we are in a position to find association between the cluster labels in successive runs and identify unstable patterns. The column label j corresponding the *maximum* entry along the i–th row refers to highest overlapping of cluster i with cluster j in successive run. Patterns that belong to these two associated clusters are considered stable whereas the patterns contributing to the other entries of the i–th row are considered unstable. For example, consider the co–occurrence matrix in Figure 3(e) obtained from the table in Figure 3(b). The maximum entry along row 1 occurs along column 2 in Figure 3(e). Cluster 1 in Pass 1 is thus associated with cluster 2 in Pass 2. Similarly Cluster 2 in Pass 1 is thus associated with cluster 1 in Pass 2. All the patterns are stable in between Pass 1 and 2 in Figure 3(b). Similarly in Figure 3(a) Cluster 1 in Pass 1 is thus associated with Cluster 1 in Pass 2 and Cluster 2 in Pass 1 is thus associated with Cluster 2 in Pass 2 as implied by the co–occurrence matrix in Figure 3(d). Only one pattern that falls out of this association is pattern 4 that belongs to Cluster 2 in Pass 1 and Cluster 1 in Pass 2. Pattern 4 is thus unstable pattern. In an identical fashion pattern 3 in Figure 3(c) is unstable as implied by the co–occurrence matrix in Figure 3(f).

0.67874	0.18687	0.8909	0.91719	0.16565
0.75774	0.48976	0.95929	0.28584	0.60198
0.74313	0.44559	0.54722	0.7572	0.26297
0.39223	0.64631	0.13862	0.75373	0.65408
0.65548	0.70936	0.14929	0.38045	0.68921
0.17119	0.75469	0.25751	0.56782	0.74815
0.70605	0.27603	0.84072	0.075854	0.45054
0.031833	0.6797	0.25428	0.05395	0.083821
0.27692	0.6551	0.81428	0.5308	0.22898
0.046171	0.16261	0.24352	0.77917	0.91334
0.097132	0.119	0.92926	0.93401	0.15238
0.82346	0.49836	0.34998	0.12991	0.82582
0.69483	0.95974	0.1966	0.56882	0.53834
0.3171	0.34039	0.25108	0.46939	0.99613
0.95022	0.58527	0.61604	0.011902	0.078176
0.034446	0.22381	0.47329	0.33712	0.44268
0.43874	0.75127	0.35166	0.16218	0.10665
0.38156	0.2551	0.83083	0.79428	0.9619
0.76552	0.50596	0.58526	0.31122	0.004634
0.7952	0.69908	0.54972	0.52853	0.77491

(a) Data Set

0.62711	0.69754	0.35251	0.26225	0.44242
0.22378	0.39715	0.36581	0.61692	0.78605
0.57503	0.38262	0.79528	0.54459	0.26673

(b) Cluster centres at pass 1

0.60835	0.548	0.60121	0.15016	0.22097
0.48696	0.59632	0.26704	0.52223	0.7675
0.36865	0.31425	0.74763	0.71177	0.36909

(c) Cluster centres at pass 2

0.66191	0.61545	0.48528	0.25087	0.41541
0.44898	0.35164	0.79542	0.7848	0.20249
0.22378	0.39715	0.36581	0.61692	0.78605

(d) Cluster centres at pass 3

Figure 1: Variability of final cluster centres when a data set is clustered into three clusters using *k*–means clustering algorithm based on random initialization of the cluster centres.

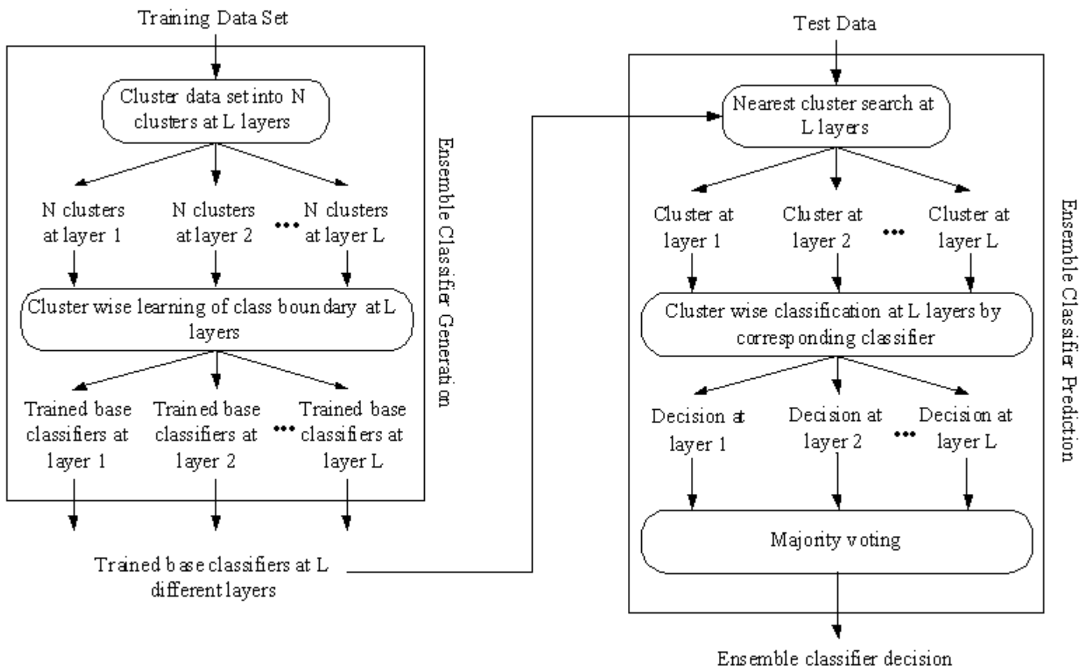

Figure 2: Ensemble classifier generation and prediction model in layered cluster oriented ensemble classifier considering *L* layers.

Pattern	Pass 1	Pass 2
1	1	1
2	2	2
3	2	2
4	**2**	**1**
5	2	2
6	1	1
7	2	2
8	1	1
9	2	2
10	1	1

(a)

Pattern	Pass 1	Pass 2
1	1	2
2	2	1
3	2	1
4	2	1
5	2	1
6	1	2
7	2	1
8	1	2
9	2	1
10	1	2

(b)

Pattern	Pass 1	Pass 2
1	1	2
2	2	1
3	**2**	**2**
4	2	1
5	2	1
6	1	2
7	2	1
8	1	2
9	2	1
10	1	2

(c)

	1	2
1	4	
2	1	5

(d)

	1	2
1		4
2	6	

(e)

	1	2
1		4
2	5	1

(f)

Figure 3: Computation of co–occurrence matrix and identification of unstable patterns.

B. Computation of Stable/Unstable Patterns

We have developed the following approach to identify unstable patterns. Let there are T patterns in the data. The cluster number of the pattern i in layer l is expressed as $\theta_{i,l}$ where $1 \le i \le T$ and $1 \le l \le L$. Given two layers $l_1$and l_2 we need to identify the clusters that have maximum overlapping. We do this by computing cluster co–occurrence matrix φ between two layers $l_1$and l_2. The dimension of φ is $N \times N$ where N is the total number of clusters. The entries of φ are defined as

$$\varphi_{l_1,l_2}(c_1, c_2) = \sum_{i=1}^{T} \upsilon(\theta_{i,l_1}, c_1) \times \upsilon(\theta_{i,l_2}, c_2) \qquad (1)$$

where

$$\upsilon(\theta_{i,l}, c) = \begin{cases} 1 \; if \; \theta_{i,l} \; equals \; c \\ 0 \; otherwise \end{cases} \qquad (2)$$

and $1 \le c_1, c_2 \le N,\; 1 \le l_1, l_2 \le L$. Given cluster c_1 at layer l_1, the corresponding cluster c_d at layer l_2 is identified as

$$c_d = \arg\max_c \varphi_{l_1,l_2}(c_1, c) \qquad (3)$$

where $1 \le c \le N$. (c_1, c_d) will form a cluster correspondence pair ρ_1. Given N clusters there will be N correspondence pairs namely $\rho_{l_1,l_2} = \{\rho_1, \rho_2, \dots, \rho_N\}$ between layers l_1 and l_2. A pattern i is considered to be unstable between layers l_1 and l_2 if

$$(\theta_{i,l_1}, \theta_{i,l_2}) \notin \rho_{l_1,l_2} \qquad (4)$$

Let the set of unstable patterns between layers l_1 and l_2 is expressed as δ_{l_1,l_2}. The complete set of unstable patterns is computed as

$$\bigcup_{\mathrm{l_1 \in L}} \bigcup_{\mathrm{l_2 \in L, l_2 \neq l_2,}} \delta_{l_1,l_2} \qquad (5)$$

where $1 \le l_1, l_2 \le L$. Patterns that do not belong to the category of unstable patterns are identified as *stable* patterns.

Let the training patterns in the data set be represented by $\Gamma = \{(\boldsymbol{x}_1, t_1), (\boldsymbol{x}_2, t_2), \dots, (\boldsymbol{x}_{|\Gamma|}, t_{|\Gamma|})\}$ where each pattern is described by a vector of *n* continuous valued features $\boldsymbol{x}_j =< x_{j1}, x_{j2}, \dots, x_{jn} >$ and a class label t_j with $t_j \in \{class_1, class_2, \dots, class_{N_{class}}\}$. A layer is denoted by l and the K clusters at layer l are denoted by $\Omega_{l,1}, \Omega_{l,2}, \dots, \Omega_{l,K}$ where $1 \le l \le N_{layers}$.

IV. Experimental Setup

A. Evaluation Criteria

The previous section mentions the steps followed to compute the total number of stable and unstable patterns. Let N_u be the total number of unstable patterns and N_s be the total number of stable patterns. The total number of patterns N in a data set is thus $N = N_u + N_s$. Let N_c represents the correctly classified instances and N_{cu} and N_{cs} be the correctly classified unstable and stable patterns respectively such that $N_c = N_{cu} + N_{cs}$. The total accuracy of the ensemble classifier is computed as $100 \times N_c/N$. The contribution of unstable and stable patterns towards total accuracy is computed as $100 \times N_{cu}/N$ and $100 \times N_{cs}/N$ respectively. Note that the percentage of unstable and stable patterns $100 \times N_u/N$ and $100 \times N_s/N$ change with number of layers in layered cluster oriented ensemble classifier.

Given L layers the base classifiers are trained on clusters at these layers. A pattern belongs to a unique cluster in each layer. A total of L decisions are thus obtained on a pattern in layered cluster oriented ensemble classifier. We have computed diversity among the base classifiers using Kohavi–Wolpert (KW) variance [32]. Given a set of $|\Gamma|$ examples $\{(\boldsymbol{x}_1, t_1), (\boldsymbol{x}_2, t_2), \dots, (\boldsymbol{x}_{|\Gamma|}, t_{|\Gamma|})\}$ in a data set, the diversity among the base classifiers is computed as

$$KW = \frac{1}{|\Gamma| \times L} \sum_{j=1}^{|\Gamma|} \left(\sum_{l=1}^{L} D_l(\boldsymbol{x}_j)\right) \times \left(L - \sum_{l=1}^{L} D_l(\boldsymbol{x}_j)\right) \qquad (6)$$

where L is the number of layers, and D_l is set as

$$D_l(\boldsymbol{x}_j) = \begin{cases} 1 \quad if \; \boldsymbol{x}_j \; classified \; correctly \; in \; layer \; l \\ \\ 0 \quad if \; \boldsymbol{x}_j \; classified \; incorrectly \; in \; layer \; l \end{cases} \qquad (7)$$

B. Data Sets

We have used the benchmark data sets in the experiments. The data sets are compiled from the UCI Machine Learning Repository [33] as used in contemporary research works [10][18][34] on ensemble classifiers. The summary of twenty data sets used in the experiments is presented in Table 1. We have used 10–fold cross validation approach for reporting the results.

C. Parameter Settings

k–means clustering algorithm with random cluster centre initialization is used in the experiments. Multi Layer Perceptron (MLP) is used in the experiments as the base classifier. Single hidden layer is used in the network with five hidden units . The number of units in the output layer is set equal to the number of classes. Tan sigmoid activation function is used in the MLP nodes. The weights are learned using a backpropagation learning algorithm. The MLPs were trained with the following parameter setting: (a) Learning rate = 0.01, (b) Momentum = 0.4, (c) Epochs i.e. # of iterations = 25, and (d) RMS goal = 0.00001. All the experiments were conducted using MATLAB.

Table 1: Data sets used in the experiments.

Dataset	# instances	# attributes	# classes
Breast Cancer	699	9	2
Diabetes	768	8	2
Ecoli	336	7	8
German	1000	20	2
Glass	214	10	7
Ionosphere	351	33	2
Iris	150	4	3
Liver	345	6	2
Parkinsons	197	23	2
Satellite	6435	36	6
Segment	2310	19	7
Sonar	208	60	2
Spam	4601	57	2
Spect	267	23	2
Thyroid	215	5	3
Transfusion	748	5	2
Vehicle	946	18	4
Vowel	528	13	11
Waveform	5000	21	3
Wine	178	13	3

V. Experimental Results and Discussion

In this section we present the following results: (i) influence of layers on percentage of unstable patterns, (ii) relationship between percentage of unstable patterns and diversity among the base classifiers, (iii) contribution of stable/unstable patterns towards total accuracy, and (iv) influence of stable/unstable patterns on total accuracy. We present the graphs on these results on the first 9 data sets while the best parameters discussed at the end of this section are shown on all the 20 data sets.

When a data set is partitioned multiple times into *k* clusters based on random initialization, some patterns change clusters at different passes. It is evidenced in Figure 1 where the cluster centres are different at different passes. We have investigated the number of patterns that change clusters. The process of counting the total number of unstable patterns when a data set is partitioned into a certain number of layers is presented in (1)–(5). Figure 4 represents the percentage of unstable patterns with the change of number of layers for the data sets in Table 1. Note that the trend line representing the change in all the graphs has a positive slope. This indicates the fact that the percentage of unstable patterns increases with increasing number of layers. The amount of change is different for different data sets with the highest being on *German* data set (74.54% with ten layers). Note that increase in unstable patterns implies reduction in stable patterns.

As the unstable patterns change clusters at different layers, the composition of clusters also change at different layers. The base classifiers are trained on clusters at different layers and the training set for the base classifiers are thus also dissimilar at different layers. This brings in diversity among the base classifiers. Addition of more layers cause more changes among the cluster contents and thus increases diversity among the base classifiers. This is evidenced from Figure 5 where the diversity is plotted against the percentage of unstable patterns as the number of layers change. The trend line in all cases shows a positive change. The increase in unstable patterns thus increases diversity among the base classifiers (Figure 5). The change in diversity is lowest for the *Breast Cancer* data set (0.009 at ten layers) and highest for the *Liver* data set (0.128 at ten layers).

With the increase in number of layers the percentage of unstable patterns goes up whereas the some of stable patterns goes down. Both unstable and stable patterns contribute towards the total accuracy obtained on a data set using layered cluster oriented ensemble classifier. It is thus worth investigating the nature of contribution towards accuracy by unstable and stable patterns. Figure 6 represents the contribution made by unstable and stable patterns towards total accuracy as the number of layers changes. It can be observed that the contribution of unstable patterns increases whereas that of stable patterns decreases with increasing number of layers for all the data sets. In case of majority of the data sets the contribution of unstable patterns is more than stable patterns after a certain number of layers (e.g. 5 layers for *Breast Cancer*, 9 layers for *Ecoli*). Stable patterns contribute more than unstable patterns at higher number of layers for some data set like *Ionosphere*.

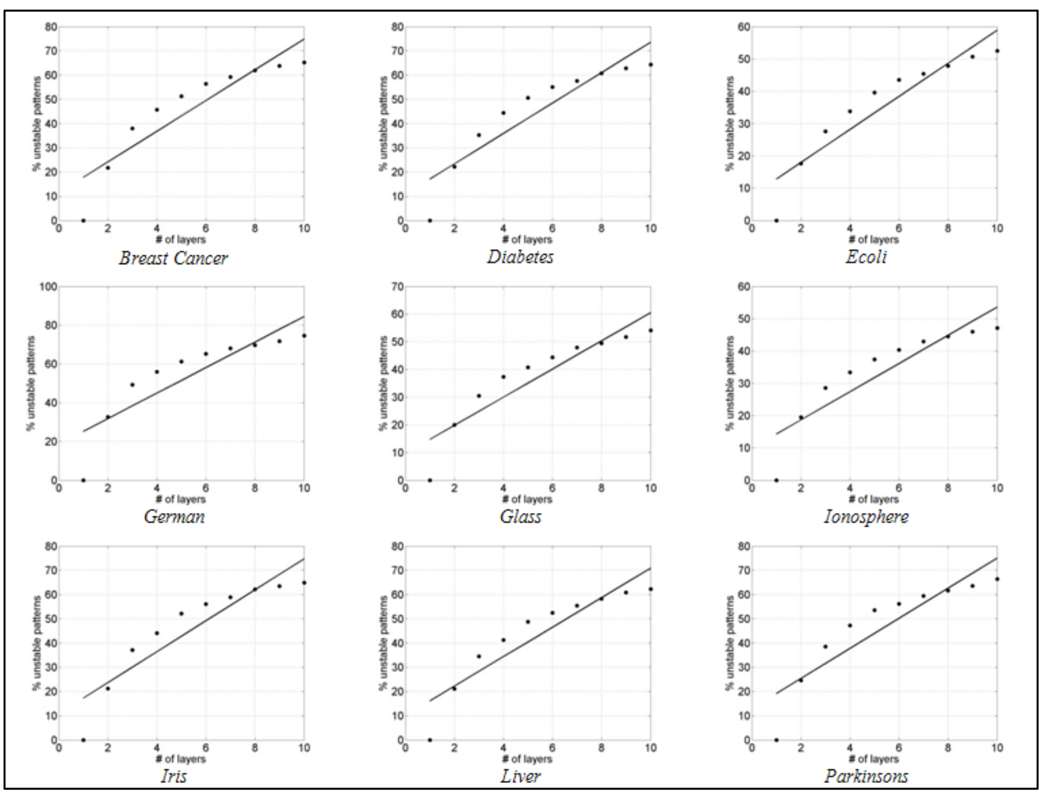

Figure 4: Influence of layers on % of unstable patterns

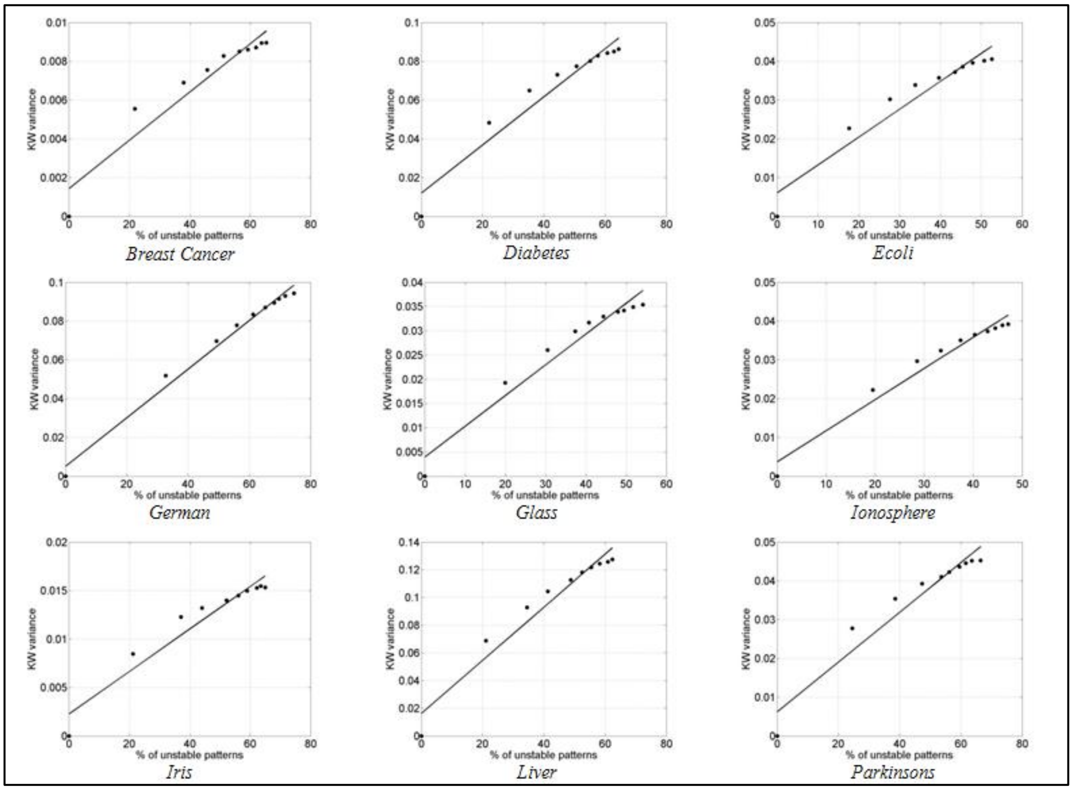

Figure 5: Influence of pattern instability on diversity as the number of layers change.

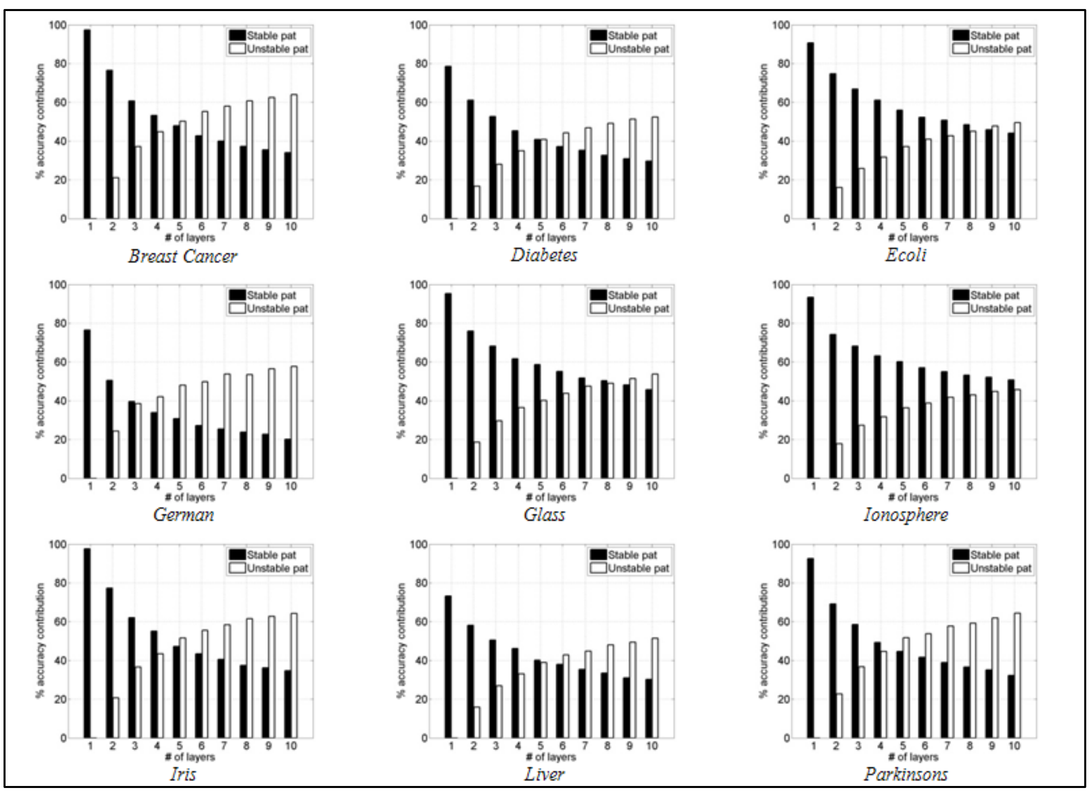

Figure 6: Accuracy contribution by stable and unstable patterns

The combination of contributions from stable and unstable patterns constitutes the total accuracy for a data set. It is observed from Figure 6 that the contribution of stable patterns decreases whereas that of unstable pattern increases. A bar graph showing the average change in contribution made by unstable and stable patterns across the layers is presented in

Figure 7 for all the data sets. Note that the average change in contribution made by unstable patterns is more than that made by stable patterns towards total accuracy. This result is overall increase of the total accuracy. The change of accuracy against percentage of unstable patterns is presented in Figure 8. Note that the trend of change has a positive slope for all data sets. This implies that increasing percentage of unstable patterns increase accuracy in general.

Table 2 reports the number of layers at which the maximum classification accuracy is obtained and the corresponding percentage of unstable patterns and diversity. The percentage of unstable patterns at which the maximum accuracy is achieved is greater than zero for all data sets. Note that the maximum accuracy is achieved at either nine or ten layers for all the twenty data sets. It is worth mentioning that percentage of unstable patterns is also high at nine/ten layers as evidenced from Figure 4. This implies that instability of patterns increases classification accuracy. However, the percentage of unstable patterns at which the maximum accuracy is achieved depends on the data set. *Satellite* data set achieves maximum accuracy (92.45%) at low percentage of unstable patterns (29.55%). The maximum accuracy for *Vowel* data set (98.12%) is obtained with high percentage of unstable patterns (75.20%).

VI. CONCLUSION

In this paper we have investigated the influence of unstable patterns on layered cluster oriented ensemble classifier. We have defined a co–occurrence matrix based approach to identify unstable patterns and observed their influence on accuracy and diversity. Based on the experimental results we can conclude the following:

(i) Unstable patterns move between clusters and this brings diversity among the base classifiers in layered cluster oriented ensemble classifier. With increasing number of layers the percentage of unstable patterns increases that in turn increases diversity among the base classifiers,

(ii) Both stable and unstable patterns contribute towards total accuracy. With increasing number of layers the contribution of unstable patterns towards total accuracy increases and that of stable patterns decreases. However the contribution made by unstable patterns is more than that of stable patterns. This is why the total accuracy increases with increasing percentage of unstable patterns.

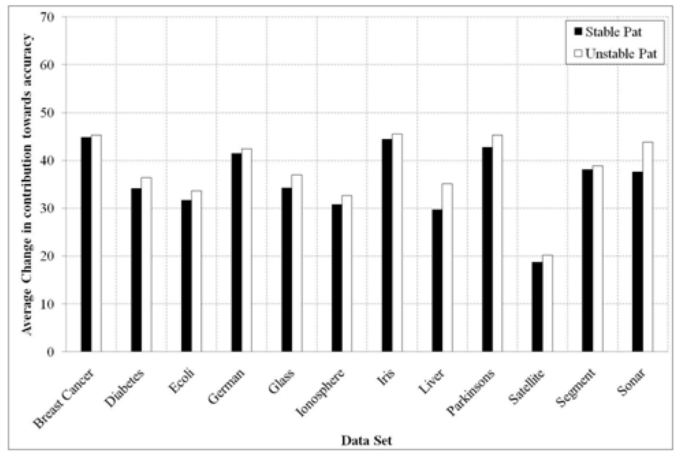

Figure 7: Average change in contribution towards total accuracy made by stable and unstable patterns

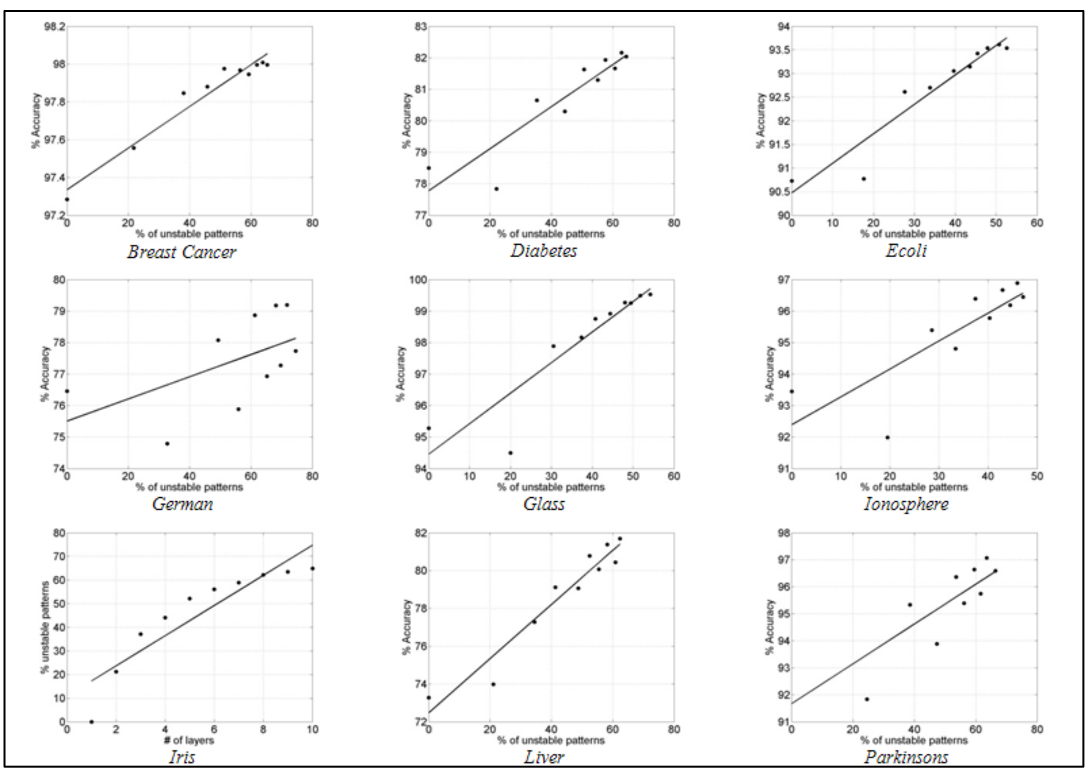

Figure 8: Influence of pattern instability on total accuracy

(iii) Instability of patterns can be controlled by number of layers. However, the percentage of unstable patterns at which the maximum accuracy is obtained depends on the data set. It is thus required to tune the number of layers for finding the percentage of unstable patterns that achieves maximum accuracy.

In future we aim to investigate the influence of incorporating fuzzy clustering in layered cluster oriented ensemble classifier.

Table 2: Number of layers that obtain maximum accuracy. The percentage of unstable patterns and diversity are reported corresponding to maximum accuracy.

Data Set	No. of layers	Percentage unstable patterns	Diversity KW variance	Max Accuracy
Breast Cancer	9	63.78	0.0089	98.01
Diabetes	9	62.89	0.0851	82.16
Ecoli	9	50.68	0.0401	93.62
German	9	71.68	0.0929	79.19
Glass	10	54.16	0.0354	99.53
Ionosphere	9	45.97	0.0390	96.89
Iris	10	64.93	0.0153	99.04
Liver	10	62.32	0.1276	81.69
Parkinsons	9	63.57	0.0452	97.07
Satellite	9	29.55	0.0342	92.45
Segment	10	56.79	0.0123	98.83
Sonar	10	62.70	0.0872	97.86
Spam	9	55.83	0.0278	95.36
Spect	10	58.06	0.0635	85.91
Thyroid	9	53.56	0.0145	99.44
Transfusion	10	54.74	0.0416	81.03
Vehicle	10	59.23	0.0809	94.30
Vowel	10	75.20	0.0711	98.12
Waveform	10	52.00	0.0780	94.54
Wine	10	64.69	0.0145	99.83

REFERENCES

[1] R. Polikar, "Ensemble based systems in decision making," IEEE Circuits and Systems Magazine, vol. 6, no. 3, pp. 21–45, 2006.

[2] T. G. Dietterich, "Ensemble Methods in Machine Learning," Int. Workshop on Multiple Classifier Systems, pp. 1–15, 2000.

[3] T. Windeatt, "Accuracy/Diversity and ensemble MLP classifier design," IEEE Trans. on Neural Networks, vol. 17, no. 5, pp. 1194–1211, 2006.

[4] H. Zouari, L. Heutte, and Y. Lecourtier, "Controlling the diversity in classifier ensembles through a measure of agreement," Pattern Recognition, vol. 38, pp. 2195–2199, 2005.

[5] T. Windeatt, "Diversity/accuracy and ensemble classifier design," Int. Conf. on Pattern Recognition, vol. 3, pp. 454–457, 2004.

[6] L. I. Kuncheva and C. J. Whitaker, "Measures of Diversity in Classifier Ensembles and Their Relationship with the Ensemble Accuracy," Machine Learning, vol. 51, no.2, pp. 181–207, 2003.

[7] L. Breiman, "Bagging predictors," Machine Learning, vol. 24, no. 2, pp. 123–140, 1996.

[8] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32, Oct. 2001.

[9] L. Breiman, "Pasting small votes for classification in large databases and on-line," Machine Learning, vol. 36, pp. 85–103, 1999.

[10] G. Martínez-Muñoz and D. Hernández-Lobato, and A. Suarez, "An analysis of ensemble pruning techniques based on ordered aggregation," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 31, no. 2, pp. 245–259, 2009.

[11] L. Chen and M. S. Kamel, "A generalized adaptive ensemble generation and aggregation approach for multiple classifiers systems," Pattern Recognition, vol. 42, pp. 629–644, 2009.

[12] L. Nanni and A. Lumini, "Fuzzy bagging: a novel ensemble of classifiers," Pattern Recognition, vol. 39, pp. 488–490, 2006.

[13] S. Eschrich and L. O. Hall, "Soft partitions lead to better learned ensembles," Proc. Annual Meeting of the North American Fuzzy Information Processing Society (NAFIPS), pp. 406–411, 2002.

[14] R. E. Schapire, "The strength of weak learnability," Machine Learning, vol. 5, no. 2, pp. 197–227, 1990.

[15] H. Drucker, C. Cortes, L. D. Jackel, Y. LeCun, and V. Vapnik, "Boosting and other ensemble methods," Neural Computation, vol. 6, no. 6, pp. 1289–1301, 1994.

[16] Y. Freund and R. E. Schapire, "Decision-theoretic generalization of on-line learning and an application to boosting," Journal of Computer and System Sciences, vol. 55, no. 1, pp. 119–139, 1997.

[17] N. García–Pedrajas, "Constructing ensembles of classifiers by means of weighted instance selection," IEEE Trans. on Neural Networks, vol. 20, no. 2, pp. 258–277, 2009.

[18] J. J. Rodriguez and J. Maudes, "Boosting recombined weak classifiers," Pattern Recognition Letters, vol. 29, pp. 1049–1059, 2008.

[19] D. Parikh and R. Polikar, "Ensemble based incrimental learning approach to data fusion," IEEE Trans. on Systems, Man and Cybernetics, vol. 37, no. 2, pp. 437–450, 2007.

[20] M. D. Muhlbaier, A. Topalis, and R. Polikar, "Learn++.NC: Combining ensemble of classifiers with dynamically weighted consult-and-vote for efficient incremental learning of new classes," IEEE Trans. on Neural Networks, vol. 20, no. 1, pp. 152–168, 2009.

[21] L. Rokach, O. Maimon, and I. Lavi, "Space Decomposition In Data Mining: A Clustering Approach", Int. Symp. On Methodologies For Intelligent Systems, pp. 24–31, 2003.

[22] J. Xiuping and J. A. Richards, "Cluster-space classification: a fast k-nearest neighbour classification for remote sensing hyperspectral data," IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data, pp. 407–410, 2003.

[23] L. I. Kuncheva, "Cluster-and-selection method for classifier combination," Int. Conf. on Knowledge-Based Intelligent Engineering Systems & Allied Technologies (KES), 185–188, 2000.

[24] L. I. Kuncheva, "Switching between selection and fusion in combining classifiers: An experiment," IEEE Trans. on Systems, Man and Cybernetics, vol. 32, no. 2, pp, 146–156, 2002.

[25] B. Tang, M. I. Heywood, and M. Shepherd, "Input partitioning to mixture of experts," Int. Joint Conf. on Neural Networks, pp. 227–232, 2002.

[26] G. Nasierding, G. Tsoumakas, A. Z. Kouzani, "Clustering based multi-label classification for image annotation and retrieval," IEEE Int. Conf. on Systems, Man and Cybernetics, pp. 4514–4519, 2009.

[27] M. J. Jordan and R. A. Jacobs, "Hierarchical mixtures of experts and the EM algorithm," Neural Computation, vol. 6, no. 2, pp. 181–214, 1994.

[28] H. Parvin, H. Alizadeh, and B. Minaei–Bidgoli, "Using clustering for generating diversity in classifier ensemble," Int. Journal of Digital Content Technology and its Applications, vol. 3, no. 1, pp. 51–57, 2009.

[29] R. Neumayer, "Clustering based ensemble classification for spam filtering," Proc. Workshop on Data Analysis (WDA), pp. 11–22, 2006.

[30] W.D. Mulder, S. Schliebs, R. Boel, and M. Kuiper, "Initialization dependence of clustering algorithms," Int. Conf. on Neural Information Processing (ICONIP), pp. 615–622, 2008.

[31] A. Rahman and B. Verma, "Novel Layered Clustering-Based Approach for Generating Ensemble of Classifiers," IEEE Transactions on Neural Networks, Vol. 22, No.5, pp.781–792, May 2011.

[32] R. Kohavi and D. H. Wolpert, "Bias plus variance decomposition for zero-one loss functions," Proc. Int. Conf. on Machine Learning, pp. 275–283, Bari, Italy.

[33] UCI Machine Learning Database, http://archive.ics.uci.edu/ml/, accessed on 6th October 2009.

[34] A. H. R. Ko, R. Sabourin, A. de S. Britto, and L. Oliveira, "Pairwise fusion matrix for combining classifiers," Pattern Recognition, vol. 40, pp. 2198–2210, 2007.