Applying Multiple Complementary Neural Networks to Solve Multiclass Classification Problem

Pawalai Kraipeerapun and Somkid Amornsamankul

***Abstract*—** In this paper, a multiclass classification problem is solved using multiple complementary neural networks. Two techniques are applied to multiple complementary neural networks which are one-against-all and error correcting output codes. We experiment our proposed techniques using an extremely imbalance data set named *glass* from the UCI machine learning repository. It is found that the combination between multiple complementary neural networks and error correcting output codes provides better performance when compared to the combination between multiple complementary neural networks and one-against-all. Its performance is also better than using a single k-class neural network, a single k-class complementary neural networks, multiple binary neural networks based on error correcting output codes, and multiple binary neural networks based on one-against-all.

***Keywords*—** multiclass classification, feedforward backpropagation neural network, complementary neural networks, error correcting output codes, one-against-all, one-against-one, *p*-against-*q*

I. Introduction

In recent years, there are many different kinds of classifier used to solve multiclass classification problem in various areas. For example, neural network was applied to classify car seat fabrics [1] and used to solve multiclass classification based on the progressive and repeated sampling technique proposed in [2]. C4.5 was used to classify Leukemia data [3] and used to test three UCI data sets [4]. Support Vector Machine was used to classify land cover [5], remote sensing image [6], and real time gas classification [7].

In this study, we apply neural network as the based classifier since it is one of the most widely used classifiers used to solve multiclass classification problems. In order to model pattern classes, several techniques can be used such as one-against-all (OAA), one-against-one (OAO), and P-against-Q (PAQ) [8]. In OAA technique, each pattern belonging to class i is trained against all other classes j where $j \neq i$. For example, in a 3-class pattern classification, patterns belonging to class 1 are trained against class 2 and 3. Patterns belonging to class 2 are trained against class 1 and 3, and patterns belonging to class 3 are trained against class 1 and 2. In OAO technique, patterns belonging to class i are trained against class j of other patterns where $j \neq i$. For an example of 3-class pattern, three neural networks are trained. In the first network, patterns belonging to class 1 are trained against patterns belonging to class 2. The second network is trained based on patterns belonging to class 1 and 3 whereas the last network is trained using patterns belonging to class 2 and 3. The number of neural network used in this technique is $k(k-1)/2$ where k is the total number of classes. In PAQ technique, patterns are trained using P classes against Q classes of other patterns. Error-correcting output code (ECOC) [9] is an example of this technique, in which each class is mapped to a codeword which is a binary string of l bits, where l can be greater or less than the number of classes. These three techniques can be modeled using multiple binary neural networks; however, the OAA and PAQ can also be implemented using a single k-class neural network.

In our previous researches [10], [11], [12], instead of using traditional neural networks, we applied complementary neural networks (CMTNN) to those three techniques: OAA, OAO, and PAQ. CMTNN consist of a pair of neural networks in which both networks have the same parameter values and they are trained using the same input; however, their target outputs are complement to each other, which are called truth target and falsity target. For example, if the truth target is "1101" then the falsity target will be "0010". if neural network is trained using the truth target then we get the truth output. In contrast, if neural network is trained using the falsity target then we get the falsity output. The final output is calculated based on both truth and falsity outputs.

We have implemented OAA and PAQ using a single CMTNN with multiple outputs [10], [12]. Figure 1 shows the OAA and PAQ modeling based on a single CMTNN with multiple outputs. A pair of neural networks: truth and falsity neural networks are created. The truth network applies the truth target to predict multiple truth outputs whereas the falsity network applies the falsity target to predict multiple falsity outputs which are the complement of the truth outputs. The number of outputs is the length of codeword. The differences between OAA and PAQ are the length of the codeword and the aggregation technique used to combine the truth and falsity outputs. The length of codeword created for OAA is equal to the number of classes in which the codeword for the i-th class of the truth target has a bit at the i-th bit position equal to 1 and the rest is equal to 0. The codeword created for PAQ is set up based on ECOC in which, in our previous experiment, the length of codeword is greater than the number of classes. Both truth and falsity outputs are involved in the aggregation for both OAA and PAQ; however, Hamming distance is also used to aggregate the outputs of PAQ.

Multiple CMTNNs have been used to implement OAO [11].

P. Kraipeerapun is with Department of Computer Science, Faculty of Science, Ramkhamhaeng University, Bangkok, Thailand (email: pawalai@yahoo.com)

S. Amornsamankul (corresponding author) is with Department of Mathematics, Faculty of Science, Mahidol University, and is with Centre of Excellence in Mathematics, CHE, Bangkok, Thailand (email: somkid.amo@mahidol.ac.th)

Figure 2 shows the OAO modeling based on multiple CMTNNs for 3-class pattern classification. The training input is split into several sets in which each set consists of patterns containing two classes. The truth and falsity targets are created for each set and trained to predict the truth and falsity output. All outputs are voted based on both truth and falsity outputs obtained from each set to get the final result.

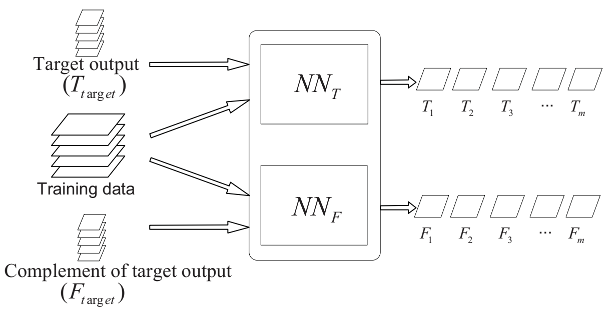

Fig. 1. OAA and PAQ modeling based on single CMTNN

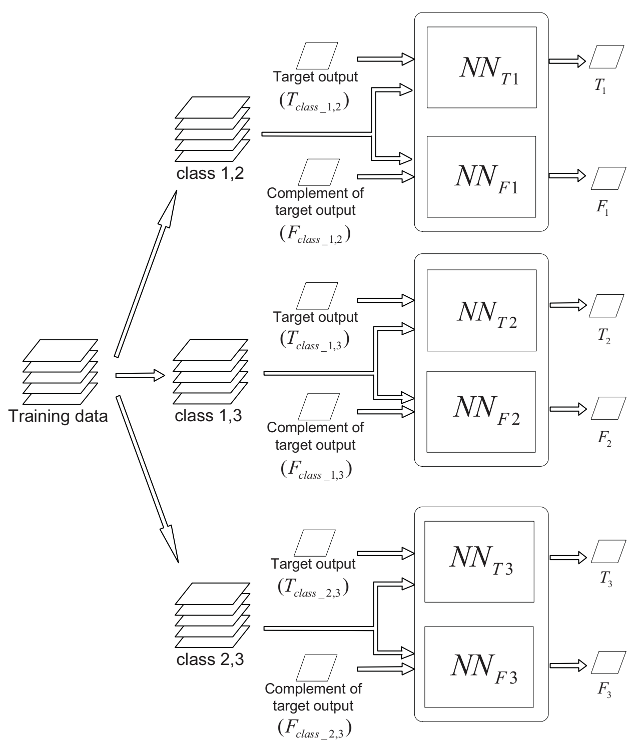

Fig. 2. OAO modeling based on multiple CMTNNs (for 3-class pattern)

It was found that the multiclass classification based on truth and falsity outputs obtained from CMTNN provide better average results when compared to average results obtained from traditional backpropagation neural network (BPNN) for OAA, OAO, and PAQ [10], [11], [12].

From our previous experiments, we have applied only a single CMTNN to OAA and ECOC techniques. In this paper, we aim to solve multiclass classification problem based on OAA and ECOC techniques using multiple CMTNNs. In order to see the performance, the results obtained from multiple CMTNNs based on ECOC are compared to results obtained from multiple CMTNNs based on OAA, multiple binary neural networks based on ECOC, multiple binary neural networks based on OAA, a single k-class neural network, and a single k-class complementary neural networks.

We test our proposed technique using *glass* data set from UCI machine learning repository. In recent years, this data set has been widely used to test several classification algorithms in many areas of application. The algorithms used in some of those applications can be described as follows.

Kurban and Beşdok [13] compared four training algorithms of radial basis function (RBF) neural networks, which are artificial bee colony, genetic (ABC), kalman filtering, and gradient descent algorithms. They tested their technique using iris, wine, and glass data sets from UCI. They found that the artificial bee colony algorithm provided better learning than the others.

Zang and Zhang [14] applied softmax discriminant classifiers (SDC) to several data sets including UCI, handwritten digit and alphabet, and IDA. Their technique was compared to NN, PCA, LDA, SRC, and LRC. All their experimental results obtained from SDC were found to give better performance.

Wen et al. [15] used the relative transformation (RT) to find the nearest neighbors for the k nearest neighbors (KNN) classifiers and local mean classifiers (LMC). They experimented their algorithms using benchmark data sets and compared results to traditional KNN and LMC in which their results provided better classification accuracies.

Huang et al. [16] proposed extreme learning machine (ELM) to solve binary and multiclass classification as well as regression problems. They tested their technique using 36 data sets from UCI and statlib. They found that ELM provides better scalability and achieve similar or much better generalization performance at faster learning speed than traditional support vector machine and least square support vector machine.

Nikolaidis et al. [17] used ten data sets from UCI to test their algorithm which is the instance seriation for prototype abstraction algorithm (ISPA), a data condensation method that generates a new set of prototypes. They compared the results to three other pruning methods: ICF, ICPL, HMN and found that their algorithm provided competitive classification results.

Ouchi et al. [18] tested their proposed efficient construction methods of rectangle greedy covers (RGC) based on ten UCI data sets. They found that the result was comparable to other methods such as RSM, logistic regression, SVM, and so on.

Huaizhen et al. [19] proposed decision basing representative data and discretization (DBRDD), a combination between partitioning around mediod (PAM) and discretization preprocess. The PAM is used to build training sets which will later be discreted using discrete algorithm of the combination of boolean logic and rough set theory. They tested their technique using seven data sets from UCI and found that their technique produced higher classification accuracy and used a smaller training set than applying the representative data based decision tree ensemble (RDDTE) based on only PAM.

Masoudnia et al. [20] proposed an evidence-based mixture of experts (ME) by using Dempster-Shafer theory of evidence to improve dynamic combination of neural networks classifiers trained based on different initial random weights. They found that their method tested on five data sets from UCI gave better classification rate when compared to basic ME and static combining of neural network based on D-S theory.

In order to see the performance of our proposed technique, we also compare our results to the results obtained from algorithms described in previous paragraphs. Although they have different environments, our purpose is to investigate the possible results obtained from other experiments compared to our experiments.

The next two sections describes the proposed concepts of combining multiple CMTNNs to OAA and ECOC for multiclass classification problem. Data set used in this study and results obtained from our experiments are explained in section IV. Finally, conclusions and future works are presented in section V.

II. The Proposed Multiple Complementary Neural Networks Based on One Against All

In the OAA modeling for k-class pattern classification, each pattern belonging to class i is trained against all other classes q where $q = 1, 2, 3, ..., k$ and $q \neq i$ [8]. In order to apply multiple CMTNNs to OAA, k pairs of neural networks are created in which each pair is set up for a single CMTNN. Figure 3 shows our proposed OAA modeling based on multiple CMTNNs ($m = k$).

Each CMTNN consist of a pair of neural networks. Let NN_{T_i} be the truth neural network and NN_{F_i} be the falsity neural network created for class i where $i = 1, 2, 3, ..., k$. Both NN_{T_i} and NN_{F_i} have the same architecture and parameter values. They are trained using the same training input data $x_{j, j=1,2,3,...,n}$ where n is the total number of input patterns in the training phase. However, they are trained using different target outputs. The NN_{T_i} is trained based on the truth target $T_{class\_i}(x_j)$ to predict degree of the truth output $T_i(x_j)$. The truth target $T_{class\_i}(x_j)$ is set to 1 if the input pattern x_j belongs to class i, otherwise $T_{class\_i}(x_j)$ is set to 0. The NN_{F_i} is trained based on the falsity target $F_{class\_i}(x_j)$ to predict degree of the falsity output $F_i(x_j)$. The $F_{class\_i}(x_j)$ is computed as the complement of the $T_{class\_i}(x_j)$ which can be written as $F_{class\_i}(x_j) = 1 - T_{class\_i}(x_j)$. Therefore, if $T_{class\_i}(x_j)$ is 1 then $F_{class\_i}(x_j)$ is set to 0. All CMTNNs in our OAA modeling apply the same training input; however, they are trained independently using different parameter setting in order to get high quality output for each CMTNN.

In the test phase, the unknown k-class input pattern $y_{j, j=1,2,3,...,p}$ is fed to all NN_{T_i} and NN_{F_i}, where p is the total number of unknown input. Let $T_i(y_j)$ be the truth output and $F_i(y_j)$ be the falsity output obtained from NN_{T_i} and NN_{F_i}, respectively. In order to get the final output for the input pattern y_j, all outputs $T_i(y_j)$ and $F_i(y_j)$ where $i = 1, 2, 3, ..., k$ must be aggregated. Our proposed aggregation technique is described below.

step 1: For each input pattern y_j,
if $T_i(y_j) > F_i(y_j)$ then $O_i(y_j)$ = 1, otherwise $O_i(y_j)$ = 0.

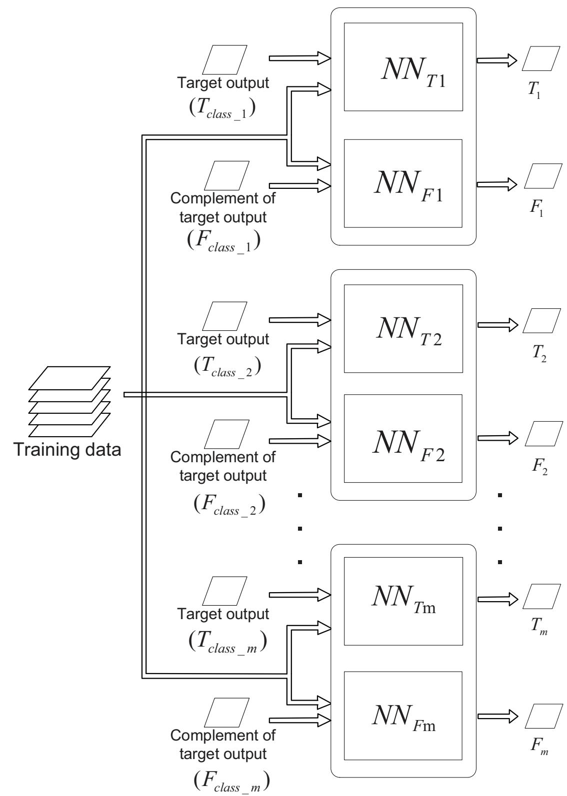

Fig. 3. The proposed OAA and ECOC modeling based on multiple CMTNNs

step 2: For each input pattern y_j, if there is only one output having the value 1, that is $O_i(y_j) = 1$, and all other $O_{t \neq i}(y_j) = 0$ where $t = 1, 2, 3, ..., k$ then the input pattern y_j is classified as class i, otherwise goto step 3.

step 3: For each input pattern y_j, compute the average output $A_i(y_j) = (T_i(y_j) + (1 - F_i(y_j)))/2$.

- In the case of having more than one outputs having the value 1, the input pattern y_j is classified as class i if $A_i(y_j) = max\{A_t(y_j)\}$ and $O_t(y_j) = 1$.
- In the case of all outputs having the value 0, the input pattern y_j is classified as class i if $A_i(y_j) = max\{A_t(y_j), t = 1, 2, 3, ..., k\}$.

For the OAA based on the traditional multiple binary neural networks, only the truth neural networks are trained based on the truth targets. The output obtained from the truth neural network $T_i(y_j)$ is compared to the threshold value, in which we do not know which threshold value is the best choice; however, 0.5 is the most widely used threshold value for the classification. Similar to our proposed technique, the maximum value of $T_i(y_j)$ where $i = 1, 2, 3, ..., k$ will be used to make the final decision if the decision cannot be made using only the threshold value.

III. The Proposed Multiple Complementary Neural Networks Based on Error Correcting Output Codes

In the ECOC modeling for k-class pattern classification, each class i is mapped to a unique codeword C_i which is a binary string of length l [9]. Codewords are arranged as rows of a matrix M where $M \in \{0,1\}^{k \times l}$. Two properties must be satisfied which are column separation and row separation [9], [21], [22]. For column separation, the columns of the codewords should be uncorrelated. Also, their complementaries should not appear in the matrix. For row separation, codewords should be separated in Hamming distance. In order to apply multiple CMTNNs to ECOC, l pairs of neural networks are created in which each pair is set up for a single CMTNN. Figure 3 shows our proposed ECOC modeling based on multiple CMTNNs ($m = l$). The same model is used for applying multiple CMTNNs to both OAA and ECOC; however, the differences are the characteristics of their target outputs and the number of neural networks. Similar to the integration between multiple CMTNNs and OAA model, each CMTNN is independently trained in order to get the best prediction result. In the test phase, all outputs $T_s(y_j)$ and $F_s(y_j)$ where $s = 1, 2, 3, ..., l$ will be aggregated. The aggregation technique can be described below.

step 1: For each input pattern y_j,
if $T_s(y_j) > F_s(y_j)$ then $O_s(y_j)$ = 1, otherwise $O_s(y_j)$ = 0.
The output code B_j containing $O_s(y_j)$ is created.

step 2: For each input pattern y_j, the output code B_j is compared to the base codewords $C_{i,i=1,2,...,k}$ defined in the matrix M. The input pattern y_j is assigned to the class i whose codeword C_i is closest in Hamming distance to the output code B_j.

For the ECOC based on the traditional multiple binary neural networks, only the truth neural networks are trained and the output $T_s(y_j)$ is compared to the threshold value before considering the Hamming distance.

IV. Experiments

A. Data Set

The benchmark data set named *glass* from UC Irvine Machine Learning Repository [23] is used in our experiment. The glass data set contains 214 instances which are classified into 6 types of glass defined in terms of their oxide content. This data set consists of 9 features (attributes) which are refractive index, sodium, magnesium, aluminum, silicon, potassium, calcium, barium, and iron. The unit measurement of features 2-9 is the weight percent in corresponding oxide. This data set is one of the extremely imbalance benchmark data. The number of instances for each type of glass is shown in Table I below.

TABLE I
THE NUMBER OF INSTANCES FOR EACH CLASS (TYPE) OF GLASS

class1	class2	class3	class4	class5	class6
70	76	17	13	9	29

B. Experimental Methodology and Results

In the experiment, we apply five-fold cross validation to the *glass* data set. For each fold, two groups of classifiers are created. Both groups contain the same types of classifiers which are multiple CMTNNs (T>F), multiple BPNNs (T>0.5), a single CMTNN with multiple outputs (T>F), and a single BPNN with multiple outputs (T>0.5). However, classifiers in the first group are trained based on OAA whereas classifiers in the second group are trained based on ECOC.

For each pair of neural networks in multiple CMTNNs, the first neural network is trained using the truth target to predict the truth output. This network is actually a traditional BPNN. The second neural network is trained using the falsity target to predict the falsity output. Both neural networks are created based on the same architecture and parameter values. We apply nine input-nodes, corresponding to nine features of glass. One hidden layer constituting of $2m$ neurons where m is the number of input features is created. Hence, 18 neurons are created in the hidden layer. For a single CMTNN, we apply the same parameter values used in the creation of multiple CMTNN; however, multiple outputs are set instead of using a binary output. Therefore, nine input-nodes, six output-nodes, and 18 neurons created in one hidden layer are set for each single CMTNN.

1) The classifiers trained based on OAA
For each fold in our proposed multiple CMTNNs based on OAA, six pairs of neural networks are created to support six types of glass. They are created separately. Each pair is chosen from twenty pairs of neural networks randomly created based on different initial weight and different order of training input data. In order to select the best pair of neural networks among twenty pairs for each type of glass, we first select the best output obtained from the truth neural network and then create the falsity neural network based on the same architecture of that selected truth neural network to get the CMTNN. Table II shows results based on the best truth neural network. It can be seen that our proposed multiple CMTNNs provides the best average result.

TABLE II
THE PERCENT CORRECT MULTICLASS CLASSIFICATION BASED ON OAA (NNS ARE CHOSEN BASED ON *truth* OUTPUT)

Fold	multiple classifiers		single classifier	
	proposed T>F	T>0.5	T>F	T>0.5
1	70.00	70.00	67.50	62.50
2	78.57	78.57	61.90	76.19
3	74.42	74.42	53.49	62.79
4	77.27	75.00	65.91	75.00
5	73.81	71.43	64.44	73.33
Average	74.81	73.88	62.65	69.96

Instead of first consider the truth neural network, another way of the selection is to consider both truth and falsity neural networks all together. The CMTNN that has the best result is chosen for each class of glass. Table III shows results obtained from neural

TABLE III

THE PERCENT CORRECT OF MULTICLASS CLASSIFICATION BASED ON OAA (NNS ARE CHOSEN BASED ON BOTH *truth* AND *falsity* OUTPUT)

Fold	multiple classifiers		single classifier	
	proposed			
	T>F	T>0.5	T>F	T>0.5
1	72.50	65.00	67.50	62.50
2	78.57	80.95	71.43	57.14
3	74.42	74.42	62.79	55.81
4	77.27	79.55	68.18	11.36
5	80.95	73.81	64.44	73.33
Average	76.74	74.75	66.87	52.03

networks selected based on the best accuracy of the combination of both truth and falsity outputs. It can be seen that the average result obtained from our multiple CMTNNs provide the best results. It also provides better result than the average result obtained from Table II. Furthermore, it can be noticed that the average result obtained from multiple binary neural networks based on the threshold value is better than the average result obtained from multiple binary neural networks in Table II. Hence, it can be concluded that the aggregation of all the best outputs obtained from each truth neural network may not provide the best final result.

Table IV shows the comparison among the best results obtained from both selection techniques. Table V shows the percent improvement of our proposed multiple CMTNNs compared to other techniques based on OAA. From our previous paper [10], a single CMTNN with multiple outputs was found to provide better performance than a single binary neural network with multiple outputs. However, in this paper, it is found that a single binary neural network with multiple outputs provides better performance than a single CMTNN with multiple outputs. These two results are not contradict since the result from [10] is the average output obtained from twenty networks. However, this paper selects the best result among twenty networks. Both conclusions come from different selection techniques. However, we can conclude from this paper that our multiple CMTNNs approach improves classification performance when compared to the existing OAA techniques based on a single k-class neural network, a single k-class complementary neural networks, and multiple binary neural networks.

2) The classifiers trained based on ECOC
In this paper, each element in the codeword matrix is chosen uniformly at random from the set $\{0, 1\}$. We use the codeword of length 15. The matrix is chosen with the largest minimum Hamming distance. Table VI shows the codewords used for the glass data set in our experiment. Therefore, fifteen pairs of neural networks are created to support six types of glass. They are created separately. Each pair is chosen from twenty pairs of neural networks randomly created based on different initial weight and different order of training input data. In order to select the best pair of neural networks among twenty pairs, the same techniques used for OAA in the previous section is applied. Table VII shows results obtained from neural networks selected based on the best truth neural network. Table VIII shows results obtained from neural networks selected based on the best accuracy of the combination of both truth and falsity outputs. Table IX shows the comparison among the best results obtained from both selection techniques. Table X shows the percent improvement of our proposed multiple CMTNNs compared to other techniques based on ECOC. From table IX and X, it can be seen that our proposed multiple CMTNNs based on ECOC provides the best average result when compared to other techniques based on ECOC.

TABLE IV

THE PERCENT CORRECT MULTICLASS CLASSIFICATION BASED ON OAA (NNS ARE CHOSEN BASED ON THE BEST OUTPUT)

Fold	multiple classifiers		single classifier	
	proposed			
	T>F	T>0.5	T>F	T>0.5
1	72.50	70.00	67.50	62.50
2	78.57	80.95	71.43	76.19
3	74.42	74.42	62.79	62.79
4	77.27	79.55	68.18	75.00
5	80.95	73.81	64.44	73.33
Average	76.74	75.75	66.87	69.96

TABLE V

THE PERCENT IMPROVEMENT OF MULTIPLE CMTNNS BASED ON OAA COMPARED TO OTHER TECHNIQUES APPLIED IN THIS PAPER

Fold	multiple BPNNs	single classifier	
	T>0.5	T>F	T>0.5
1	3.57	7.41	16.00
2	-2.94	10.00	3.13
3	0.00	18.52	18.52
4	-2.86	13.33	3.03
5	9.68	25.62	10.39
Average	1.32	14.77	9.69

Figure 4 shows the graphical representation of the comparison of the percent corrects obtained from all techniques used in this paper for each fold. Table XI shows the average percent correct obtained from each technique. Table XII shows the percent improvement of our proposed multiple CMTNNs based on ECOC compared to other techniques applied in this paper. It can be conclude that the combination between multiple CMTNNs and ECOC provides the best result when compared to other techniques used in this study.

TABLE VI

CODEWORDS FOR 6-CLASS CLASSIFICATION USED IN THIS STUDY

1	1	1	0	0	1	1	0	1	1	1	1	1	0	1
1	1	1	1	1	1	0	1	1	0	1	1	0	0	1
1	0	0	0	1	0	1	0	1	0	0	1	0	0	1
0	1	0	1	1	0	0	1	1	0	0	1	1	1	0
0	0	1	1	0	1	0	0	1	1	0	1	1	1	0
1	1	0	0	1	0	1	1	1	0	1	0	0	0	1

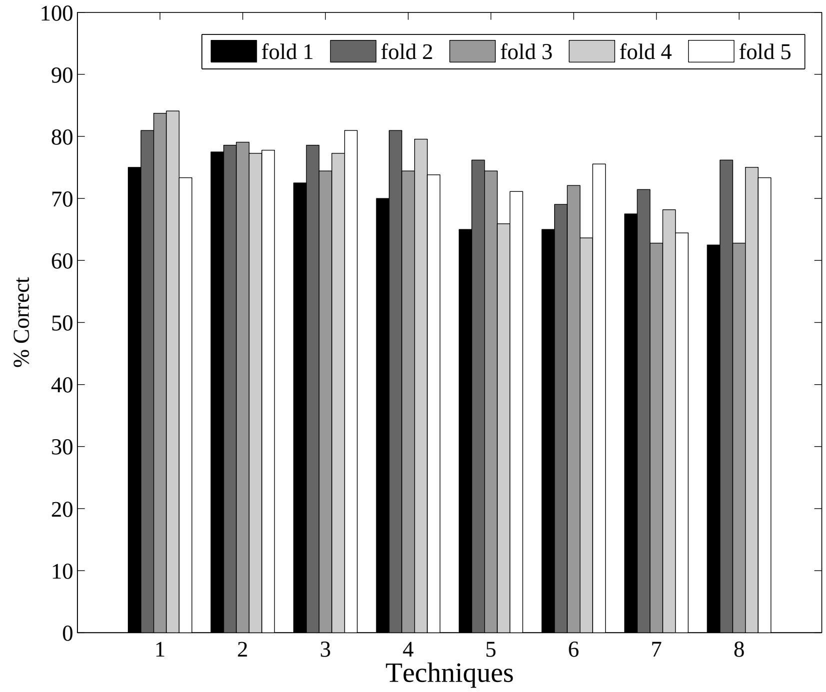

Fig. 4. The percent correct of all techniques used in this paper for each fold (technique: 1. multiple CMTNNs & ECOC, 2. multiple NNs & ECOC, 3. multiple CMTNNs & OAA, 4. multiple NNs & OAA, 5. single CMTNNs & ECOC, 6. single NNs & ECOC, 7. single CMTNNs & OAA, 8. single NNs & OAA

TABLE XI

THE AVERAGE PERCENT CORRECT OF ALL TECHNIQUES APPLIED IN THIS PAPER

multiple classifiers				single classifier			
ECOC		OAA		ECOC		OAA	
T>F	T>0.5	T>F	T>0.5	T>F	T>0.5	T>F	T>0.5
79.42	78.04	76.74	75.75	70.53	69.07	66.87	69.96

TABLE XII

THE PERCENT IMPROVEMENT OF OUR PROPOSED MULTIPLE CMTNNS BASED ON ECOC COMPARED TO OTHER TECHNIQUES APPLIED IN THIS PAPER

multiple classifiers			single classifier			
ECOC	OAA		ECOC		OAA	
T>0.5	T>F	T>0.5	T>F	T>0.5	T>F	T>0.5
1.77	3.49	4.85	12.61	14.99	18.77	13.52

Furthermore, results obtained from our techniques are also compared to results obtained from other techniques using the same glass data set in their experiments. Table XIII shows the comparison of percent corrects of classification accuracy obtained from our techniques and some recent techniques explained in section I. It can be seen that our techniques based on CMTNN provide competitive classification results.

V. CONCLUSION

Instead of using only multiple binary neural networks, we propose multiple CMTNNs to deal with OAA and ECOC since our proposed technique can eliminate vagueness that may occur during the aggregation. From the proposed aggregation technique, we avoid the use of threshold. Outputs obtained from a pair of neural networks, which are complement to each other, are compared in order to increase the confidence in the decision making. Multiple CMTNNs are created and tested based on the glass data set which is an extremely imbalance data set. The results show that the combination between multiple CMTNNs and ECOC provides better performance when compared to the combination between multiple CMTNNs and OAA. Its performance is also better than using a single k-class neural network, a single k-class CMTNNs, multiple binary neural networks based on ECOC, and multiple binary neural networks based on OAA.

Acknowledgements: This research project is supported by Faculty of Science, Mahidol University.

TABLE XIII

THE PERCENT CORRECT OBTAINED FROM PAPERS USING GLASS DATA SET

Ref.	%correct	Algorithm	Data characteristic	
[13]	91.20	RBF + ABC	random train:test 60%:40%	30 runs
[14]	73.2	SDC	random train:test 60%:40%	30 runs
	65.7	NN		
[15]	73.44	LRT-KNN	random train:test 70%:30%	10 runs
	74.26	LRT-LMC		
	69.18	KNN		
	71.97	LMC		
[16]	68.41	ELM	random train:test 142:72	50 runs
[17]	74.72	ISPA	10-fold	30 runs
	74.56	ICF		
	73.97	ICPL		
	79.83	HMN		
[18]	65.00	RGC	10-fold	
	68.12	RSM		
	67.40	SVM		
[19]	87.5	DBRDD + Covering	10-fold	
	72.00	RDDTE + Covering		
[20]	66.82	Evidence-Based ME	5-fold	
	65.42	ME		
	65.42	combining NN based on DS		
This paper	79.42	multiple CMTNNs + ECOC	5-fold	
	76.74	multiple CMTNNs + OAA		
	70.53	CMTNN + ECOC		
	66.87	CMTNN + OAA		

TABLE VII

THE PERCENT CORRECT MULTICLASS CLASSIFICATION BASED ON ECOC (NNS ARE CHOSEN BASED ON *truth* OUTPUT)

Fold	multiple classifiers		single classifier	
	proposed T>F	T>0.5	T>F	T>0.5
1	72.50	77.50	65.00	65.00
2	73.81	76.19	76.19	69.05
3	74.42	76.74	74.42	72.09
4	79.55	77.27	52.27	63.64
5	68.89	77.78	71.11	75.56
Average	73.83	77.10	67.80	69.07

TABLE VIII

THE PERCENT CORRECT OF MULTICLASS CLASSIFICATION BASED ON ECOC (NNS ARE CHOSEN BASED ON BOTH *truth* AND *falsity* OUTPUT)

Fold	multiple classifiers		single classifier	
	proposed T>F	T>0.5	T>F	T>0.5
1	75.00	70.00	65.00	65.00
2	80.95	78.57	76.19	69.05
3	83.72	79.07	74.42	72.09
4	84.09	72.73	65.91	59.09
5	73.33	71.11	71.11	75.56
Average	79.42	74.30	70.53	68.16

REFERENCES

[1] R. Furferi, L. Governi, and Y. Volpe, "Neural Network based classification of car seat fabrics," *Inernational Journal of Mathematical Models and Methods in Applied Sciences*, vol. 5, no. 3, pp. 696–703, 2011.

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

[3] W. QingQiang, L. Hua, and L. K. Hong, "Mixed-sampling Approach to Unbalanced Data Distributions: A Case Study involving Leukemia's Document profiling," *WSEAS Transactions on Information Science and Applications*, vol. 8, no. 9, pp. 356–379, September 2011.

[4] M. M. Mazid, A. B. M. S. Ali, and K. S. Tickle, "Input space reduction for Rule Based Classification," *WSEAS Transactions on Information Science and Applications*, vol. 7, no. 6, pp. 749–759, June 2010.

[5] M. Pal, "Multiclass approaches for support vector machine based land cover classification," *CoRR*, vol. abs/0802.2411, 2008.

[6] H.-N. Qi, J.-G. Yang, Y.-W. Zhong, and C. Deng, "Multi-class SVM based remote sensing image classification and its semi-supervised improvement scheme," in *Proceedings of 2004 Inernational Conference on Machine Learning and Cybernetics*, vol. 5, 2004, pp. 3146–3151.

[7] G. H. Kim, Y. W. Kim, S. J. Lee, and G. J. Jeon, "Multi-class system based on SVM for real-time gas mixture classification," in *Proceedings of SICE Annual Conference 2010*, 2010, pp. 1764–1767.

[8] G. Ou and Y. L. Murphey, "Multi-Class Pattern Classification using Neural Networks," *Pattern Recognition*, vol. 40, no. 1, pp. 4–18, 2007.

[9] T. G. Dietterich and G. Bakiri, "Solving Multiclass Learning Problems via Error-Correcting Output Codes," *Journal of Artificial Intelligence Research*, vol. 2, pp. 263–286, 1995.

[10] P. Kraipeerapun, C. C. Fung, and K. W. Wong, "Uncertainty Assessment using Neural Networks and Interval Neutrosophic Sets for Multiclass Classification Problems," *WSEAS Transactions on Computers*, vol. 6, no. 3, pp. 463–470, March 2007.

[11] ——, "Quantification of Vagueness in Multiclass Classification Based on Multiple Binary Neural Networks," in *Proceedings of the Sixth International Conference on Machine Learning and Cybernetics*, August 2007, pp. 140–144.

[12] S. Amornsamankul, J. Promrak, and P. Kraipeerapun, "Solving Multiclass Classification Problems using Combining Complementary Neural Networks and Error-Correcting Output Codes," *International Journal of Mathematics and Computers in Simulation*, vol. 5, no. 3, pp. 266–273, 2011.

TABLE IX

THE PERCENT CORRECT MULTICLASS CLASSIFICATION BASED ON ECOC (NNS ARE CHOSEN BASED ON THE BEST OUTPUT)

Fold	multiple classifiers		single classifier	
	proposed			
	T>F	T>0.5	T>F	T>0.5
1	75.00	77.50	65.00	65.00
2	80.95	78.57	76.19	69.05
3	83.72	79.07	74.42	72.09
4	84.09	77.27	65.91	63.64
5	73.33	77.78	71.11	75.56
Average	79.42	78.04	70.53	69.07

TABLE X

THE PERCENT IMPROVEMENT OF MULTIPLE CMTNNS BASED ON ECOC COMPARED TO OTHER TECHNIQUES APPLIED IN THIS PAPER

Fold	multiple BPNNs	single classifier	
	T>0.5	T>F	T>0.5
1	-3.23	15.38	15.38
2	3.03	6.25	17.24
3	5.88	12.50	16.13
4	8.82	27.59	32.14
5	-5.71	3.12	-2.94
Average	1.77	12.61	14.99

[13] T. Kurban and E. Beşdok, "A Comparison of RBF Neural Network Training Algorithms for Inertial Sensor Based Terrain Classification," *Sensors*, vol. 9, pp. 6312–6329, 2009.

[14] F. Zang and J. she Zhang, "Softmax Discriminant Classifier," in *Proceedings of 2011 Third International Conference on Multimedia Information Networking and Security*, 2011, pp. 16–19.

[15] G. Wen, J. Wei, Z. Yu, J. Wen, and L. Jiang, "Relative Nearest Neigbors for Classification," in *Proceedings of the 2011 International Conference on Machine Learning and Cybernetics*, Guilin, July 2011, pp. 773–778.

[16] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme Learning Machine for Regression and Multiclass Classification," *IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics*, vol. 42, no. 2, pp. 513–529, April 2012.

[17] K. Nikolaidis, E. Rodriguez, J. Y. Goulermas, and Q. Wu, "Instance Seriation for Prototype Abstraction," in *Proceedings of 2010 IEEE Fifth International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA)*, September 2010, pp. 1351–1355.

[18] K. Ouchi, A. Nakamura, and M. Kudo, "Efficient Construction and Usefulness of Hyper-Rectangle Greedy Covers," in *Proceedings of 2011 IEEE International Conference on Granular Computing*, 2011, pp. 533–538.

[19] Y. Huaizhen, L. Linghua, X. Wei, and L. Lei, "A Classification Method Based on PAM Algorithm and Discrete Preprocess," in *Proceedings of 2010 International Conference on Intelligent Computing and Integrated Systems (ICISS)*, October 2010, pp. 530–533.

[20] S. Masoudnia, M. Rostami, M. Tabassian, A. Sajedin, and R. Ebrahimpour, "Evidence-Based Mixture of MLP-Experts," in *Proceedings of the 2010 International Joint Conference on Neural Networks (IJCNN)*, July 2010, pp. 1–7.

[21] O. Pujol, P. Radeva, and J. Vitrià, "Discriminant ECOC: A Heuristic Method for Application Dependent Design of Error Correcting Output Codes," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 28, no. 6, pp. 1007–1012, June 2006.

[22] X. Zhang, L. Liang, and H.-Y. Shum, "Spectral Error Correcting Output Codes for Efficient Multiclass Recognition," in *Proceedings of 2009 IEEE 12th International Conference on Computer Vision (ICCV)*, Beijing, China, 29 September – 2 October 2009, pp. 1111–1118.

[23] A. Asuncion and D. Newman, "UCI Machine Learning Repository," 2007. [Online]. Available: http://www.ics.uci.edu/∼mlearn/MLRepository.html