Journal of Soft Computing and Information Technology (JSCIT)
Vol. 5, No. 1, Spring 2016 ISSN: 2383-1006

Using Curve Fitting in Error Correcting Output Codes

Maryam Haddadi[1], Maliheh Ahmadi[2], Mohammad Reza Keyvanpour[3], and Noushin Riahi[4]
1-Master of Artificial Intelligence, Computer Engineering Department, Alzahra University,
2-Master of Artificial Intelligence, Computer Engineering Department, Alzahra University,
3-Associative Professor, Computer Engineering Department, Alzahra University,
4-Associative Professor, Computer Engineering Department, Alzahra University,
[1]m.haddadi@student.alzahra.ac.ir, [2] ahamdi@student.alzahra.ac.ir, [3]keyvanpour@alzahra.ac.ir,
[4]nriahi@alzahra.ac.ir

***Abstract*—The Error Correcting Output Codes (ECOC) represent any number of the binary classifiers to model the multiclass problems successfully. In this paper, we have used Curve Fitting as a binary classifier in ECOC algorithm to solve multiclass classification problems. Curve Fitting is a classifier based on a nonlinear decision boundary that separates two pattern classes by the curves of the best fit, and arriving at optimal boundary points between two classes. Since we need a coding and a decoding strategy to design an ECOC system, this paper gives five coding and eight decoding strategies of ECOC and compares the results of Curve Fitting with Adaboost classification and Nearest Mean Classifier (NMC). This evaluation has been performed on different data sets of UCI machine learning repository. The results indicate that One-versus-one, ECOC-ONE coding and LAP, BDEN decoding having the best results in contrast with another coding and decoding strategies and Curve Fitting is a good base classifier in ECOC, also it is comparable with the other ECOC approaches.**

***Keywords*—Classifier, Coding, Curve Fitting, Decoding, Error Correcting Output Codes (ECOC).**

I. Introduction

Machine learning investigates automatic techniques to make accurate predictions based on past observations. There are various multiclass classification techniques [1]: Support Vector Machines [2], [3] multiclass Adaboost [4], [5], decision trees [6], Mixture of Experts [7], ECOC [8], etc. although organizing a highly accurate multiclass prediction rule is definitely a difficult task, An alternative approach is to use a set of relatively simple suboptimum classifiers and to decide a combination strategy that combines together the outcomes. A usual way to deal with Multiclass classification problem is by war of a divide and - conquer approach. In this scope, ECOC has been implemented with successful results. The ECOC technique can be broken down into two different phases: decoding and encoding. Given a set of classes, the coding phase creates a code word for each class based on various binary problems. The decoding phase makes a classification determination for a given test sample based on the value of the output code. Many coding designs have been proposed to codify an ECOC coding matrix, acquiring successful results [9]. In [10] authors reformulated the ECOC models from the perspective of multi-task learning, where the binary classifiers were learned in a common subspace of data. This novel model could be considered as an adaptive generalization of the traditional ECOC framework. It simultaneously optimized the representation of data as well as the binary classifiers. More importantly, it built a bridge between the ECOC frameworks and multitask learning for multi-class learning problems. Authors in [11] investigated the behavior of the ECOC approach on two image vision problems: logo recognition and shape classification using Decision Tree and Adaboost as the base learners. The results showed that the ECOC method can be used to improve the classification performance in comparison with the classical multiclass approaches.

One of the most well-known characteristics of the ECOC is that it makes better the generalization efficiency of the base classifiers [12], [13]. Furthermore, the ECOC technique has proven to be able to reduce the error caused by the bias and the variance of the base learning algorithm [14]. In addition to many binary classifiers proposed as a base classifier in ECOC, For example, the decision-tree methods, such as C4.5 and CART can build trees whose leaves are labeled with binary values, most artificial neural network algorithms , such as the perceptron algorithm and the error Back Propagation (BP) algorithm [15], Fisher Linear Discriminant Analysis (FLDA), Discrete Adaboost, Linear support vector machine (SVM) [16], [17], SVM with Radial Basis Function kernel (RBF), Nearest Mean Classifier (NMC) [18], that are best suited to learning binary functions. In [19] the generalization ability of ECOC SVMs was discussed. ECOC SVMs with optimum coding matrices were selected by experiment, and applied to remote sensing image classification.

In this paper, Curve Fitting has been used as a base classifier in ECOC. Curve Fitting is a binary classifier which is based on a nonlinear decision boundary that segregates two classes with high-accuracy [20]. The decision boundary is reached by sampling the twodimensional feature spaces and getting to optimum boundary points between the two classes. This algorithm can apply either interpolation where an exact fit to the data is needed, or smoothing in which a "smooth" function is constructed that approximately fits the data. To demonstrate the efficiency of Curve Fitting, eight data sets have been selected from UCI machine learning repository. These data sets have multidimensional feature space; by using PCA algorithm the feature space have been reduced in two. We compared the results of Curve Fitting with Adaboost classification and Nearest Mean Classifier (NMC) on five coding and eight decoding strategies of ECOC.

This paper is organized as follows: Section 2, 3 presents the Curve Fitting algorithm and compact ECOC design. Section 4 evaluates the novel methodology on different data sets. Finally, Section 5 concludes the paper.

II. CURVE FITTING

The purpose of this classifier involves reaching the nonlinear decision border that distinguishes the two classes.

The nonlinear decision border is reached applying the following steps:
(1) The Curves of the best fit correlated to the pattern vectors of the two classes are achieved applying the statistical techniques.
(2) Within the region bounded the feature space is sampled by the curves of best-fit [14].
(3) Along each sample line, the pattern vectors of the two classes are analyzed, and optimal border point is arrived at.
(4) The different border points are then connected to give a rough decision border.
(5) The rough decision border which is removed roughness to obtain the nonlinear decision border which offers as discriminate function to divide the two classes [13].

A. Feature Space Sampling

The best curve among different options such as line, higher order polynomial, exponential, logarithmic, etc. it can be selected by analyzing the mean remaining error for each case and selecting the curve that gives the least value of mean remaining error [15].

In the region bounded the feature space is sampled by the two curves of the best fit along the y direction. The sampling range [a b] along the x direction is reached using:

a =Min (x direction) =Min (Min (Class1), Min (Class2))
b =Max (x direction) =Max (Max (Class1), Max (Class2))

The sampling interval is selected based using the expression $\delta_L=1/10^n$, where n is the number of decimal places that are used in the representation of a feature.

B. Optimum Border Points

The feature vectors are analyzed along each sampling line and the optimum point of separation of two classes is to be discovered. The selection of an optimum border point under different conditions, as given by below:
(1) In the sampling line when no feature vectors are existent, the mean value of the two best-fit curve functions, analyzed at a point in the x direction.
(2) When feature vectors are membership to both classes in a sampling line, the border point will be the mean of the maximum remaining vectors of the two curves of the best fit, falling inside the region bounded by them.
(3) When feature vectors being a part of only one class are existent in a sampling line, the border point will be the mean of the maximum remaining vector of the existent class and the value of the best-fit curve function.

C. Nonlinear Decision Border

The border points when connected together forms a rough decision border. This decision border can be smoothened by fitting a cubic spline curve to the border points [16]. The equation of the curve thus achieved is the discriminated function that distinguishes the two classes.

III. ERROR CORRECTING OUTPUT CODES

Creating a code word for each of the N_c classes is the foundation of the ECOC structure. Arrangement the code words as rows of a matrix we determine the "coding matrices" M, where $M \in \{-1,1\}^{N_c \times n}$, being n the code length. From point of view of learning, matrix M shows n binary learning problems (dichotomies), each correlated to a matrix column. Combining classes in sets, each dichotomy determines a section of classes (coded by +1, -1 according to their class membership). Applying the n trained binary classifiers, a code is achieved for each data point in the test set. This code is compared to the base code words of each class determined in the matrix M, and the data point is allocated to the class with the "closest" code word. The matrix values can be extended to the ternary cases $M \in \{-1, 0, 1\}^{N_c \times n}$, showing that a particular class is not considered (gets 0 value) by a given dichotomy. We need a coding and a decoding strategy to design an ECOC system. When the ECOC technique was first advanced it was considered that the ECOC code matrices should be planned to have certain properties to conclude well. A good error-correcting output code for a k-class problem should satisfy that rows; columns (and their complementary) are well-separated from the rest in terms of Hamming distance. Most of the discrete coding strategies up to now are based on pre-designed problem independent code word construction satisfying the need of high separability between rows and columns. These strategies contain one-versus-all that uses N_c dichotomies, random techniques, with estimated length of 10 $\log_2 (N_c)$ bits per code for Dense random and 15 $\log_2 (N_c)$ for Sparse random [17] and one-versus-one with $N_c (N_c-1)/2$ dichotomies [18]. The last one mentioned has obtained high popularity showing a better accuracy in comparison to the other commented strategies. These traditional coding strategies are based on a prior division of subsets of classes independently of the problem to be used.

Originally, the decoding step was based on error-correcting precepts under the supposition that the learning task can be modeled as a communication problem, in which class information is transmitted over a channel [19].

The decoding strategy corresponds to the problem of distance estimation between the code word of the new example and the code words of the trained classes. Concerning the decoding strategies, two of the most standard techniques are the Euclidean distance in Equation (1) and the Hamming decoding distance in Equation (2).

$$d_j = \sqrt{\sum_{i=1}^{n}\left(x_i - y_i^j\right)^2} \tag{1}$$

$$d_j = \sum_{i=1}^{n} | x_i - y_i^j | /2 \tag{2}$$

Where d_j is the distance to the row class j, n is the number of dichotomies (and thus, the components of the code word), and x and y are the values of the input vector code word and the base class code word, respectively.

If the minimum Hamming distance between any pair of class code words is d, then any [(d-1)/2], errors in the unique dichotomies result can be corrected, since the nearest code word will be the correct one. In Fig. 1, an example of a coding matrix M for a one-versus-all toy problem is shown.

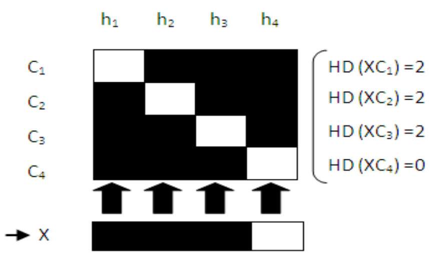

Fig. 1. Coding matrix M for four classes one-versus-all toy problem. New test sample with code word X is classified to class c4 of minimal distance using the Hamming distance

The problem has four classes, and each column shows its correlated dichotomy. The dark and white regions are coded by -1 and 1, respectively. The first column h_1 represents the training of {c_1} vs {c_2, c_3, c_4}, and so on. A new test input is evaluated using dichotomies h_1… h_4, and its code word X is decoded using the Hamming distance (HD) between each row of M and X. Finally, the new test input is classified by the class of minimum distance (c_4, in this case).

IV. EVALUATION

Before the results are displayed, we investigate our validation methodology concerning the data, comparatives, measurements, and experiments.

- Data: The data which we have used for the experiments are eight multiclass data sets from UCI machine learning repository data sets. The UCI is a collection of databases, domain theories, and data generators that are used by the machine learning community for the empirical analysis of machine learning algorithms [20]. This eight datasets are: Iris, Ecoli, Glass, Balance, Wine, Hayes, Teacher assistant and Vowel.
- Comparatives: For the comparatives, Hamming Decoding (HD), Euclidean Decoding (ED), Inverse Hamming Decoding (IHD), Attenuated Euclidean Decoding (AED), Linear Loss-based decoding (LLB) and Exponential Loss-based decoding (ELB), Laplacian Decoding (LAP), β-Density Distribution Decoding (BDEN) have been applied. Furthermore, all the decoding strategies are used over the state-of-the-art ECOC coding: one-versus-one [21], one-versus-all [22], dense random [23], sparse random [23], DECOC [24], Forest-ECOC [25] and ECOC-ONE designs.

- Measurements: To measure the performance of the different strategies we apply ten-fold cross-validation.

Cross-validation is a computer intensive technique, using all available examples as training and test examples. It mimics the use of training and test sets by repeatedly training the algorithm K times with a fraction 1/K of training examples left out for testing purposes [ ].
The base classifiers used for the experiments are Curve Fitting, Adaboost and NMC.

- Experiments: We evaluate the classification of UCI data sets.

Definition 4.1: Decoding bias is the value introduced by comparison of two code words on locations including the zero symbols, this means that being the magnitude of value proportional to the number of zero locations.

Definition 4.2: A dynamic range bias pertains to the difference between the ranges of values related to the decoding procedure of each code word.

Definition 4.3: A general decoding decomposition to express decoding strategies is explained as follows:

$$D = \sum_{k \in L_a} a_k + \sum_{i \in L_b} b_i + \sum_{j \in L_c} c_j \tag{3}$$

Where L_a, L_b and L_c are the sets of indexes of a code word related to the zero locations, Matches on {-1, +1} values, and mismatches on {-1, +1} values, respectively.

Let $|L_a| = \alpha$, $|L_b| = \beta$ and $|L_c| = \gamma$ be the number of zeros, number of matches between two code words, and number of mismatches between two code words, respectively.

As a zero symbol means that the corresponding classifier is not trained over a class, respecting the decision of this classifier to evaluate the similarity of the new test example to that class does not make sense.

Hypothesis 4.1: The bias caused by a zero location applying a special decoding strategy should be zero (b=0).

In addition, we discuss that to acquire comparable results between the classes' code words; each code word of the coding matrix M should take values in the same dynamic range. The Dynamic Range (DR) related to each code word is defined as follows:

$$DR = [\min(s_1, s_2), \max(s_1, s_1)], s_1 = \sum_{i \in L_b} |b_i|, s_2 = \sum_{j \in L_c} |c_j| \qquad (4)$$

If S_1 and S_2 are constant allocated factors for all the code words, the dynamic range is preserved for all classes, and the decoding measures are comparable.

Hypothesis 4.2: S_1 and S_2 should be constant allocated factors for all the code words.

Based on the prior hypothesis, we explain four classes of decoding strategies in Table I.

TABLE I

TYPES OF DECODING STRATEGIES

	b≠0	b=0
Different dynamic ranges	Type 0	Type I
Same dynamic ranges	Type II	Type III

Definition 4.4: A decoding strategy is Type 0 if the bias created by the zero symbol is higher than zero (b>0), and the dynamic ranges between code words are different.

A decoding strategy is of Type I if the bias created by the zero symbol is null (b=0), and the dynamic ranges between code words are different.

Definition 4.6: A decoding strategy is of Type II if the bias created by the zero symbol is higher than zero (b>0), and the dynamic ranges between code words are the same.

Definition 4.7: A decoding strategy is of Type III if the bias created by the zero symbol is null (b=0), and the dynamic ranges between code words are the same.

Note that none of the decoding strategies demonstrated, belongs to Type II and Type III strategies since the dynamic ranges differ for various number of locations coded by zero. Only the BDEN decoding variants normalize the dynamic ranges to work in the same dynamic ranges for all code words.

Based on the classes of decoding strategies and with the use of discrete or continuous outputs of the classifiers, six different types of decoding are shown in Table II. The Laplacian decoding (LAP) has also been contained as the simplest selection of Type III strategies. Some strategies, such as ED, AED and LB (LLB, ELB) can also be applied in both discrete and continuous domains. Note that none of the decoding strategies belongs to Type II strategies since it does not exist a method that maintain the dynamic range for all code words at same time that includes bias for the zero symbol.

TABLE II

DECODING STRATEGIES GROUPED BY TYPE AND DISCRETE/CONTINUOUS DOMAINS

Class	Discrete	Continuous
Type 0	HD,IHD,ED	-
Type I	AED	LB
Type III	BDEN, LAP	-

Based on the present formulation, our working hypothesis is that when the decoding strategies avoid the bias produced by the zero symbols and all the code words work in the same dynamic range, the performance of the ECOC designs is improved same as Type III. Therefore, we apply the decoding strategies on different coding designs and we test their behavior over different multiclass data sets.

Finally, the mean ranking positions grouping the techniques in their respective classes are shown in Table III. One can observe that the ranking performance in all cases is better when satisfying the decoding properties. Besides, the Type III strategies obtain results statistically significantly better than the rest of the strategies.

The rank shows the average position of each technique. Note that all strategies with results not statistically significant from the top one are considered also as the first choice. We can see that our method is very competitive when compared to the other methods.

TABLE III

RANKING POSITIONS OF DECODING STRATEGIES ON THE UCI EXPERIMENTS GROUPED BY TYPE

	Curve Fitting	Adaboost	NMC
Type I	***4.36***	4.79	4.93
Type 0	6.87	6.68	***6.12***
Type III	***3.27***	3.55	4.28

One-versus-one and ECOC-ONE coding strategies, which are known to obtain the best results because, one reason which causes we have a good result is the coding strategies have the most discriminated dichotomizes. One-versus-all coding, in general, is one of the poorest choices for learning with ECOC. However, it is still used because of the small number of dichotomizes involved.

Some of the comparative results on UCI data sets are shown in Fig. 2-8.

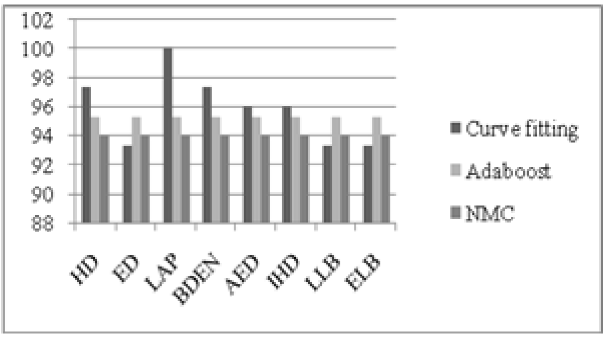

Fig. 2. One-vs-one coding with all decoding on Iris data set

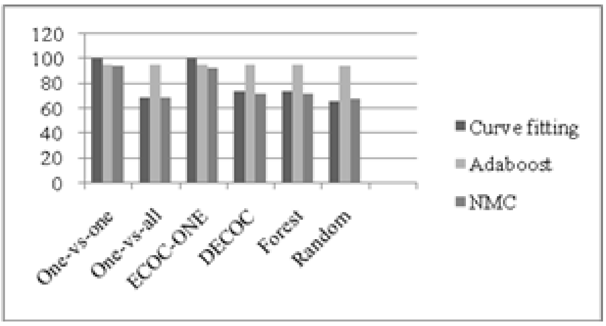

Fig. 3. ECOC-ONE coding with all decoding on Iris data set

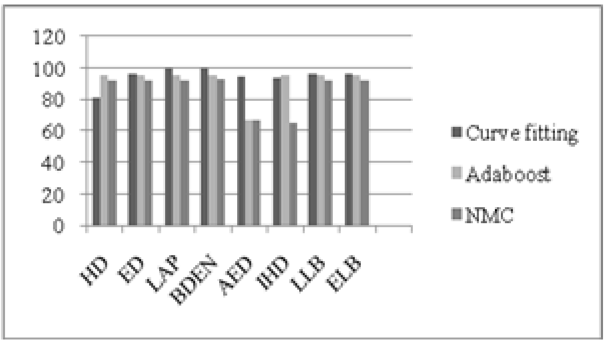

Fig. 4. LAP decoding with all coding on Iris data set

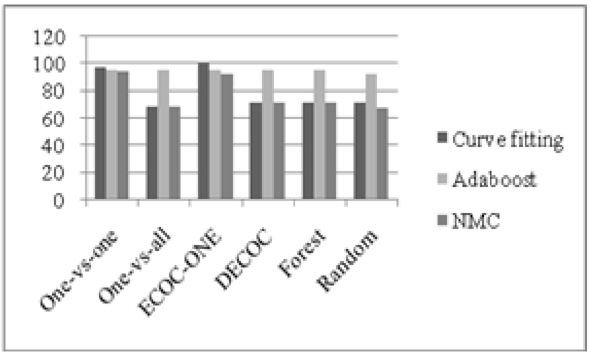

Fig. 5. BDEN decoding with all coding on Iris dataset

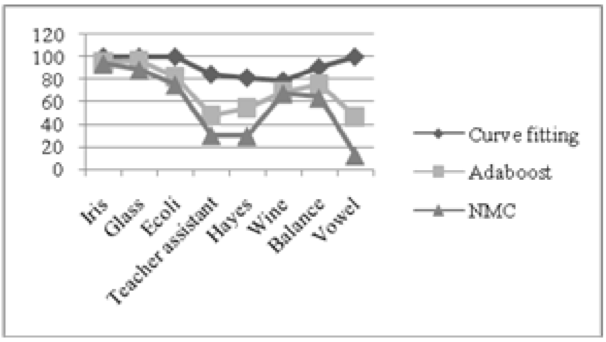

Fig. 6. One-vs-one coding with LAP decoding on 8 data sets

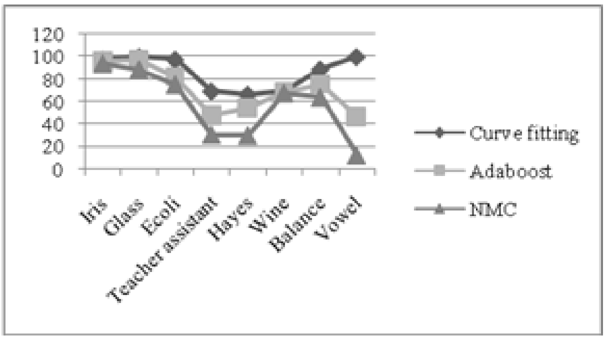

Fig. 7. One-vs-one coding with BDEN decoding on 8 data sets

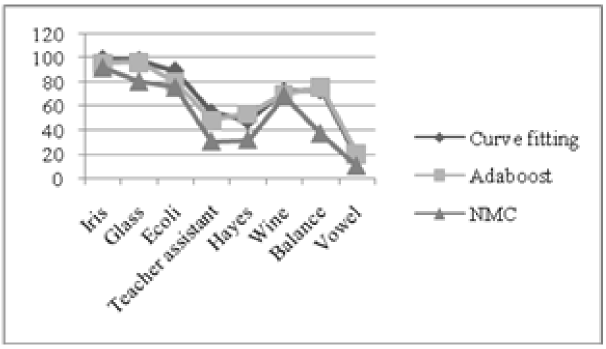

Fig. 8. ECOC-ONE coding with LAP decoding on 8 data sets

V. CONCLUSION

Although the results of Adaboost and Curve Fitting show that Curve Fitting approach is comparable with the other ECOC approaches and in some cases Curve Fitting having better results than Adaboost, it cannot be considered that Curve Fitting is significantly better than Adaboost. It is caused by the fact that Adaboost is a relatively strong classifier and it is able to fit better the problem boundaries. when the results of Curve Fitting has been compared with the NMC, it has been shown that the results of the Curve Fitting approach are significantly better for most of the cases because NMC depends on statistical properties such as mean and when the scatter of data set is high, it doesn't give good results, but Curve Fitting is independent of statistical properties. One-versus-one and ECOC-ONE coding strategies have the best results. One reason which causes we have a good result is the coding strategies have the most discriminated dichotomizers. According to Table III, Type III strategies obtain results statistically significantly better than the rest of the strategies because these strategies avoid the bias produced by the zero symbols and all the code words work in the same dynamic range. The results of different coding and decoding strategies on eight USI data sets are shown in Appendix A.

APPENDIX A

Tables of results on 8 data sets of UCI are shown as follows:

A.1. RESULTS ON IRIS DATA SET

Coding ECOC	Base classifier	Decoding							
		HD	ED	LAP	BDEN	AED	IHD	LLB	ELB
One-Vs-One	CURVE	97.33	93.33	٩٩,٨	97.33	96	96	93.33	93.33
	ADA	95.33	95.33	95.33	95.33	95 .33	95.33	95.33	95.33
	NMC	94	94	94	94	94	94	94	94
One-Vs-All	CURVE	69.33	69.33	69.33	69.33	33.33	62.67	69.33	69.33
	ADA	95.33	95.33	95.33	95.33	33.33	94	95.33	95.33
	NMC	68.67	68.67	69.33	68.67	33.33	64.67	68.67	68.67
ECOC-ONE	CURVE	81.33	96	٩٩,٨	٩٩,٨	94.67	94	96	96
	ADA	95.33	95.33	95.33	95.33	66.67	95.33	95.33	95.33
	NMC	92	92	92	92.67	66.67	64.67	92	92
DECOC	CURVE	72	72	73.33	72	72	72	72	72
	ADA	95.33	95.33	95.33	95.33	95.33	95.33	95.33	95.33
	NMC	71.33	71.33	71.33	71.33	71.33	71.33	71.33	71.33
Forest	CURVE	72	72	73.33	72	72	72	72	72
	ADA	95.33	94.67	95.33	95.33	95.33	95.33	95.33	71.33
	NMC	71.33	71.33	71.33	71.33	71.33	71.33	71.33	71.33
Random	CURVE	61.33	65.33	66	71.33	38	62.67	64	64
	ADA	94.67	61.33	94	92	59.33	95.33	95.33	95.33
	NMC	66	64.67	68	68	48	66.67	73.33	69.33

A.2. RESULTS ON GLASS DATA SET

Coding ECOC	Base classifier	Decoding							
		HD	ED	LAP	BDEN	AED	IHD	LLB	ELB
One-Vs-One	CURVE	96.8	83.12	٩٩,٨	٩٩,٨	13.18	96.8	83.12	78.44
	ADA	96.8	96.8	96.8	96.8	96.8	96.8	96.8	96.8
	NMC	88.57	88.57	88.57	88.57	88.57	88.57	88.57	88.57
One-Vs-All	CURVE	52.84	52.84	52.84	52.84	15.54	43.77	52.84	52.88
	ADA	95.89	95.89	95.89	95.89	15.54	96.34	95.89	95.43
	NMC	55.28	55.28	69.33	52.28	15.54	49.78	55.28	52.28
ECOC-ONE	CURVE	96.32	75.82	98.67	٩٩,٨	20.54	96.8	68.92	59.29
	ADA	94.98	95.43	96.8	94.07	94.48	94.07	94.98	95.89
	NMC	80.26	77.64	80.39	92.67	61.65	85.84	73.10	77.19
DECOC	CURVE	72.21	41.10	82.64	84.03	22.03	81.77	42.01	47.97
	ADA	95.45	95.45	94.55	46.62	18.23	85.91	86.62	95.91
	NMC	57.51	68.46	64.94	31.90	17.77	56.71	68.07	46.37
Forest	CURVE	77.23	54.33	82.25	94.55	26.82	84.00	59.91	49.76
	ADA	95.45	94.67	95.91	75.45	34.18	95.91	95.91	68.07
	NMC	80.32	69.44	69.94	64.46	36.84	70.39	76.30	68.07
Random	CURVE	36.13	31.56	36.15	34.13	16.41	29.29	38.84	32.92
	ADA	95.45	36.13	95.91	94.00	20.95	95.45	95.45	95.91
	NMC	50.28	51.15	50.74	53.87	17.77	52.10	58.90	52.97

A.3. RESULTS ON ECOLI DATA SET

Coding ECOC	Base classifier	Decoding							
		HD	ED	LAP	BDEN	AED	IHD	LLB	ELB
One-Vs-One	CURVE	87.55	70.85	٩٩,٨	97.26	34.97	93.93	70.85	45.00
	ADA	82.07	82.07	82.07	82.07	44.37	82.68	82.07	82.07
	NMC	75.39	75.39	75.39	75.39	37.37	75.39	75.39	75.39
One-Vs-All	CURVE	23.71	23.71	23.71	23.71	6.07	59.60	23.71	24.31
	ADA	74.78	74.78	74.78	74.78	6.07	78.42	74.78	74.78
	NMC	51.66	51.99	51.66	51.99	6.09	73.87	51.99	46.84
ECOC-ONE	CURVE	87.86	67.49	89.76	88.78	52.63	89.38	70.22	54.44
	ADA	79.64	79.34	79.95	80.57	40.38	79.04	82.40	80.27
	NMC	73.58	73.58	75.39	68.73	42.23	75.40	74.77	75.08
DECOC	CURVE	56.25	50.47	68.44	75.39	16.81	56.66	53.82	50.80
	ADA	75.05	77.79	75.68	45.55	16.71	71.13	73.20	77.50
	NMC	71.43	71.14	71.48	56.24	20.63	63.30	53.85	72.36
Forest	CURVE	66.60	59.31	80.24	81.44	21.22	69.34	57.09	53.16
	ADA	87.55	70.85	٩٩,٨	97.26	34.97	93.93	70.85	45.00
	NMC	82.07	82.07	82.07	82.07	44.37	82.68	82.07	82.07
Random	CURVE	75.39	75.39	75.39	75.39	37.37	75.39	75.39	75.39
	ADA	23.71	23.71	23.71	23.71	6.07	59.60	23.71	24.31
	NMC	74.78	74.78	74.78	74.78	6.07	78.42	74.78	74.78

A.4. RESULTS ON TEACHER ASSISTANT DATA SET

Coding ECOC	Base classifier	Decoding							
		HD	ED	LAP	BDEN	AED	IHD	LLB	ELB
One-Vs-One	CURVE	38.96	33.33	84.29	69.12	63.58	63.50	33.33	33.33
	ADA	47.83	47.83	47.83	47.83	47.17	49.12	47.83	47.83
	NMC	30.75	30.75	30.75	30.75	27.63	30.75	30.75	30.75
One-Vs-All	CURVE	31.50	31.50	31.50	31.50	32.71	30.67	31.50	31.50
	ADA	53.46	53.46	53.46	53.46	32.71	46.58	53.46	53.36
	NMC	32.75	33.96	33.96	33.96	32.75	28.79	33.96	33.96
ECOC-ONE	CURVE	39.58	33.96	54.75	45.88	37.79	43.17	33.96	33.96
	ADA	48.38	48.38	48.38	48.38	49.00	48.46	48.38	48.29
	NMC	30.83	30.83	30.83	30.21	28.92	30.83	30.83	30.83
DECOC	CURVE	38.83	38.83	57.83	37.17	38.83	38.83	38.83	38.83
	ADA	49.63	49.63	49.63	49.63	49.63	49.63	49.63	49.63
	NMC	33.88	33.88	33.88	33.88	33.88	33.88	33.88	33.88
Forest	CURVE	38.83	38.83	57.83	35.21	34.04	38.83	38.83	38.83
	ADA	49.63	47.79	49.63	49.63	49.63	49.63	49.63	33.88
	NMC	33.88	33.88	33.88	33.88	33.88	33.88	33.88	33.88
Random	CURVE	38.33	32.63	33.79	33.21	37.75	33.29	33.79	26.92
	ADA	45.88	52.17	45.88	44.63	40.88	49.75	49.63	45.29
	NMC	33.92	33.92	30.79	33.21	33.83	33.38	30.08	32.67

A.6. RESULTS ON WINE DATA SET

Coding ECOC	Base classifier	Decoding							
		HD	ED	LAP	BDEN	AED	IHD	LLB	ELB
One-Vs-One	CURVE	60.78	31.24	78.56	69.12	58.89	40.69	31.24	31.24
	ADA	68.79	68.79	68.79	68.79	43.10	69.90	68.79	68.59
	NMC	67.65	67.65	67.65	67.65	41.41	67.65	67.65	67.65
One-Vs-All	CURVE	39.61	39.61	39.61	39.61	27.32	34.08	39.61	39.61
	ADA	68.21	68.20	68.20	68.20	27.32	67.09	68.20	68.20
	NMC	53.14	53.07	53.10	54.25	27.10	63.24	53.56	53.56
ECOC-ONE	CURVE	43.46	37.91	72.55	52.91	43.46	45.16	37.91	37.91
	ADA	70.46	70.46	70.46	70.46	50.78	71.01	70.46	68.17
	NMC	68.76	68.76	68.76	68.93	68.76	68.76	68.76	68.76
DECOC	CURVE	39.02	39.02	54.15	38.50	39.02	39.02	39.02	39.02
	ADA	66.47	66.47	66.47	66.47	66.47	66.47	66.47	66.47
	NMC	51.93	51.93	51.93	51.93	51.93	51.93	51.93	51.93
Forest	CURVE	39.02	39.02	54.15	36.93	37.45	39.02	39.02	39.02
	ADA	66.47	63.79	66.47	66.47	66.47	66.47	66.47	51.93
	NMC	51.93	51.93	51.93	51.93	51.93	51.93	51.93	51.93
Random	CURVE	41.76	40.20	41.86	40.23	32.94	37.42	37.42	41.24
	ADA	68.14	65.36	64.80	68.17	35.72	68.79	66.47	62.55
	NMC	55.23	59.12	55.78	58.69	35.16	64.80	60.29	50.78

A.5. RESULTS ON HAYES DATA SET

Coding ECOC	Base classifier	Decoding							
		HD	ED	LAP	BDEN	AED	IHD	LLB	ELB
One-Vs-One	CURVE	49.62	29.56	81.32	66.15	49.01	52.58	29.56	29.56
	ADA	54.73	54.73	54.73	54.73	50.44	55.44	54.73	54.73
	NMC	30.11	30.11	30.11	30.11	30.82	30.11	30.11	30.11
One-Vs-All	CURVE	27.31	27.31	27.31	27.31	51.59	35.16	27.31	27.31
	ADA	52.53	52.53	52.53	52.53	21.59	52.58	52.53	52.53
	NMC	36.65	32.36	32.42	32.36	26.59	30.11	32.36	32.36
ECOC-ONE	CURVE	40.93	40.99	48.19	32.97	57.64	43.90	40.99	40.99
	ADA	54.01	54.01	54.01	54.01	42.58	55.44	54.01	55.33
	NMC	32.25	32.25	32.25	29.56	32.97	32.25	32.25	32.25
DECOC	CURVE	33.90	33.90	43.96	29.67	33.90	33.90	33.90	33.90
	ADA	61.15	61.15	61.15	61.15	61.15	61.15	61.15	61.15
	NMC	36.59	36.59	36.59	36.59	36.59	36.59	36.59	36.59
Forest	CURVE	33.90	33.90	43.96	33.08	26.83	33.90	33.90	33.90
	ADA	61.15	55.99	61.15	61.15	61.15	61.15	61.15	36.59
	NMC	36.59	36.59	36.59	36.59	36.59	36.59	36.59	36.59
Random	CURVE	33.19	38.90	36.76	30.93	27.25	33.90	34.62	33.90
	ADA	55.99	58.55	56.70	58.24	38.74	61.70	58.85	58.13
	NMC	36.65	35.93	31.65	35.93	31.54	36.70	38.08	35.93

A.7. RESULTS ON BALANCE DATA SET

Coding ECOC	Base classifier	Decoding							
		HD	ED	LAP	BDEN	AED	IHD	LLB	ELB
One-Vs-One	CURVE	72.22	62.70	90.48	88.57	64.76	68.41	62.70	62.70
	ADA	75.40	75.40	75.40	75.40	65.71	75.40	75.40	75.40
	NMC	63.97	63.97	63.97	63.97	63.97	63.97	63.97	63.97
One-Vs-All	CURVE	66.19	66.19	66.19	66.19	45.71	66.19	66.19	66.19
	ADA	75.56	75.56	75.56	75.56	45.71	73.33	75.56	75.56
	NMC	54.04	58.52	52.34	57.69	45.80	54.13	56.03	56.03
ECOC-ONE	CURVE	70.48	67.78	73.45	74.44	74.44	70.48	67.78	67.87
	ADA	75.87	75.78	75.87	75.87	74.44	75.87	75.87	75.87
	NMC	37.30	37.30	37.30	37.30	69.37	65.40	37.30	37.30
DECOC	CURVE	65.35	65.35	80.92	65.19	65.35	65.35	65.35	65.35
	ADA	73.62	73.62	73.62	73.62	73.62	73.62	73.62	73.62
	NMC	67.73	67.73	67.73	67.73	67.73	57.73	67.73	67.73
Forest	CURVE	65.35	65.35	65.35	65.35	80.92	64.13	65.35	65.35
	ADA	73.62	73.62	73.62	73.78	73.62	73.62	73.62	67.73
	NMC	67.73	67.73	67.73	67.73	67.73	67.73	67.73	67.73
Random	CURVE	48.36	51.33	48.36	50.27	27.37	55.78	47.37	55.82
	ADA	73.46	72.50	72.99	74.07	56.96	71.87	72.35	73.61
	NMC	43.87	58.16	51.86	51.39	43.93	58.37	59.43	56.73

A.8. RESULTS ON VOWEL DATA SET

Coding ECOC	Base classifier	Decoding							
		HD	ED	LAP	BDEN	AED	IHD	LLB	ELB
One-Vs-One	CURVE	99.70	90.91	99.8	99.8	35.76	99.70	90.91	71.82
	ADA	46.67	46.67	46.67	46.70	37.27	44.85	46.67	46.97
	NMC	12.73	12.72	12.73	12.71	10.00	12.73	12.76	12.73
One-Vs-All	CURVE	16.36	16.36	16.36	16.36	9.09	11.52	16.36	20.61
	ADA	15.45	15.45	15.45	15.45	9.09	14.24	15.45	14.85
	NMC	12.12	12.12	12.12	12.12	9.09	9.39	12.12	10.91
ECOC-ONE	CURVE	16.97	13.94	16.67	12.12	9.09	15.76	16.97	16.06
	ADA	18.84	21.52	20.30	19.09	9.09	14.55	14.55	16.36
	NMC	13.94	13.94	10.30	10.91	9.09	12.42	10.91	11.52
DECOC	CURVE	98.18	88.48	98.97	98.97	39.39	95.15	89.09	66.97
	ADA	44.24	45.45	45.76	45.77	37.88	42.73	45.15	46.06
	NMC	12.42	12.73	13.03	13.12	10.61	13.33	12.73	12.73
Forest	CURVE	57.27	29.39	26.67	59.39	7.88	60.61	29.09	18.18
	ADA	12.39	29.70	24.85	19.39	9.70	19.09	26.97	29.09
	NMC	13.94	13.03	15.76	11.82	7.88	12.12	13.64	11.82
Random	CURVE	87.58	43.94	99.70	89.39	7.88	83.94	49.70	26.36
	ADA	34.85	28.18	30.61	24.85	6.36	26.67	34.55	27.27
	NMC	13.94	15.15	13.94	12.73	6.97	15.76	13.33	12.12

REFERENCES

[1] V. Vapnik, The nature of statistical learning theory, Springer-Verlag, 1995.

[2] J. Friedman, T. Hastie, R. Tibshirani, Additive logistic regression: a statistical view of boosting, The Annals of Statistics vol. 38 (2) 1998, pp. 337–374.

[3] L. Breiman, J. Friedman, R. Olshen, C. Stone, Classification and Regression Trees, Wadsworth, 1984.

[4] R.A. Jacobs, M.I. Jordan, S.E. Nowlan, G.E. Hinton, Adaptive mixture of experts, Neural Computation 3, 1991, pp 79–87.

[5] T. G. Dietterich and G. Bakiri, Solving multi class learning problems via error correcting output codes, J. of Artificial Intelligence Research 2, 1995, pp263-286.

[6] S. Escalera, O. Pujol, and P. Radeva. Boosted landmarks of contextual descriptors and forest-ecoc: A novel framework to detect and classify objects in clutter scenes. *PatternRecognition Letters*, vol. 28, no. 13, 2007, pp. 1759–1768.

[7] E. Allwein, R. Schapire, Y. Singer, Reducing multi class to binary: A unifying approach for margin classifiers, Journal of Machine Learning Research vol. 1, 2002, pp. 113–141.

[8] T.Windeatt, R. Ghaderi, Coding and decoding for multiclass learning problems, Information Fusion vol. 4, no. 1, 2003, pp. 11–21.

[9] T. Dietterich, E. Kong, Error-correcting output codes corrects bias and variance, in: P. of the 21th ICML (Ed.), S. Prieditis and S. Russell, 1995, pp. 313–321.

[10] Thomas G. Dietterich, GhulumBakiriSolvingMulti class Learning Problems via Error-Correcting Output Codes, Journal of Arti_cial Intelligence Research 2, 1995, pp. 263,286.

[11] Sergio Escalera, OriolPujol, and PetiaRadeva, Error-Correcting Output Coding for Chagasic Patients Characterization, 19th International Conference on Pattern Recognition, vol. 17, 2008, pp. 1-4.

[12] Sergio Escalera, David M.J. Tax, OriolPujol, PetiaRadeva, and Robert P.W. Duin, Multiclass Classification in Image Analysis via Error-Correcting Output Codes, H. Kwa´snicka& L.C. Jain (Eds.): Innovations in Intell. Image Analysis, SCI 339, springerlink.com _c Springer-Verlag Berlin Heidelberg, 2011, pp. 7–29.

[13] Ganjeshvaidyanathan.s.DR.BibbasKar,DR.N.Kumaravel,A Curve Fitting approach to separation of nonlinear separable pattern classes, applied to chromosome classification, IEEE international conference on signal processing, 2008.

[14] C James; Truncating criteria for polynomial Curve Fitting ; 1971, J.Phys.D:Appl.Phys.4, pp 357-363.

[15] Draper, N. R; smith,H. Applied regression Analysis; 3rdedition.Johnwiley& Sons, 1998.

[16] White; Micromaths: Cubic Spline Curve Fitting, mathematics applications, 1986, pp 39-45.

[17] E. Allwein, R. Schapire, Y. Singer, Reducing multi class to binary: A unifying approach for margin classifiers, Journal of Machine Learning Research, vol. 1, 2002, pp. 113–141.

[18] T. Hastie, R. Tibshirani, Classification by pairwise grouping, The annals of statistics, vol. 26, no. 5, 1998, pp. 451–471.

[19] T. Dietterich, G. Bakiri, Error-correcting output codes: A general method for improving multi class inductive learning programs, in: A. Press (Ed.), 9th CAI, 1991, pp. 572–577.

[20] A Asuncion and D.J Newman UCI machine learning repository. http://www.ics.uci.edu/~mlearn/MLRepository.html, 2007, University of California, Irvine,school of Information and Computer Science.

[21] A.Rocha, S. Goldenstein, Multiclass from Binary: Expanding One-vs-All, One-vs-One and ECOC-based Approaches, IEEE Transaction on neural networks and learning system, 2013, pp. 1757–1772.

[22] Martine Pelikan, David E. Goldberg, and Erick Cantu-Pez. Learning machines, In McGraw-Hill, 1965.

[23] E. Allwein, R. schapire, and Y, Singer, Reducing multi class to binary: A unifying approach for margine classifiers, volume 1, 2002, pp, 113-141.

[24] O. PujolM.Rosales, and P.Radeva. Intravascular ultrasound images vessel characterization using Adaboost, Functional Imaging and Modeling of the Heart, 2003, pp. 242-251.