A New Objective Function for Ensemble Selection in Random Subspaces

Albert Hung-Ren Ko, Robert Sabourin, and Alceu de Souza Britto, Jr.

Abstract— **Most works based on diversity suggest that there exists only weak correlation between diversity and ensemble accuracy. We show that by combining the diversities with the classification accuracy of each individual classifier, we can achieve a strong correlation between the combined diversities and the ensemble accuracy in Random Subspaces.**

I. INTRODUCTION

Different classifiers usually make different errors on different samples, which means that, by combining classifiers, we can arrive at an ensemble that makes more accurate decisions [1, 7, 9, 10, 14]. In order to have classifiers with different errors, diverse classifiers are grouped together into what is known as an Ensemble of Classifiers (EoC). There are several methods for creating diverse classifiers, among them Random Subspaces [2], Bagging and Boosting [13]. The Random Subspaces method creates various classifiers by using different subsets of features to train them. Because problems are represented in different subspaces, different classifiers develop different borders for the classification.

Given a pool of different classifiers, ensemble selection in meant to select classifiers to construct the best ensemble [1, 5, 8, 9]. Obviously, a correct criterion is one of the most crucial elements in selecting pertinent classifiers [1, 5, 8, 9]. It is considered that, each classifier is required to have different errors, so that they will be corrected by the opinions of the whole group [1, 5, 13, 14, 32]. The need for reducing the correlation among individual classifiers has been well demonstrated [15, 18]. This property is regarded as the diversity of an ensemble. But, there is no universal definition of diversity, and therefore a number of different diversity measures have been proposed [1, 2, 4, 8]. Even with so many different diversity measures, clear correlations between ensemble accuracy and diversity measures cannot be found [1, 9, 13]. Given the challenge of using diversity for ensemble selection, here are the key questions that need to be addressed:

1) Is there a correlation between the diversity and ensemble accuracy?
2) Which is the best diversity measure for observing such a correlation?
3) Is there any effect on such a correlation, e.g. from the number of classifiers?
4) Can diversity be effective for ensemble selection?

To answer these questions, we derive compound diversity functions by combining diversities and the classification accuracies of individual classifiers, and we show that with such functions there are strong correlations between the diversity measures and ensemble accuracy in Random Subspaces.

II. DILEMMA OF THE AMBIGUITY TOWARDS THE ENSEMBLE ACCURACY

Theoretically, the concept of ensemble diversity can be understood via bias-variance decomposition [3, 9, 11, 16, 17]. Here, we adopt the framework established in [9]. Given the features of a certain sample $x \in X$, assume that we have a classifier f trained with a particular dataset X, the expectation of the output of the classifier can be written as $E(f(x))$. For convenience, we denote $E\{f\}$ instead of $E(f(x))$. If the correct value of the output is r, then we can write the bias of the classifier f as:

$$bias(f) = E\{f\} - r \tag{1}$$

and the variance of the classifier f can be written as:

$$var(f) = E\{(f - E\{f\})^2\} \tag{2}$$

Now, the mean square error (MSE) of this classifier f can be exactly represented by its variance and bias:

$$E\{(f-r)^2\} = (E\{f\} - r)^2 + E\{(f - E\{f\})^2\} \tag{3}$$

$$or\ MSE\{f\} = bias(f)^2 + var(f) \tag{4}$$

This form can be further decomposed into bias-variance-covariance [9, 10]. For L classifiers, the averaged bias of the ensemble members is defined as:

$$\bar{b} = \frac{1}{L}\sum_i^L (E\{f_i\} - r) \tag{5}$$

Then, the averaged variance of the ensemble members will be:

$$\bar{v} = \frac{1}{L}\sum_i^L (E\{(f_i - E\{f_i\})^2\}) \tag{6}$$

and the averaged covariance of the ensemble members will be:

$$\bar{c} = \frac{1}{L(L-1)}\sum_i^L \sum_{j\neq i}^L E((f_i - E\{f_i\})(f_j - E\{f_j\})) \tag{7}$$

If we decompose the mean square error for this ensemble of L classifiers, we get:

Albert Hung-Ren Ko is with LIVIA, École de Technologie Supérieure, University of Quebec, Montreal, Quebec, H3C 1K3, Canada (email: albert@livia.etsmtl.ca).

Robert Sabourin is with LIVIA, École de Technologie Supérieure, University of Quebec, Montreal, Quebec, H3C 1K3, Canada (email: robert.sabourin@etsmtl.ca).

Alceu de Souza Britto, Jr. is with PPGIA, Pontifical Catholic University of Parana, PR 80215-901, Curitiba, Brazil (email: alceu@ppgia.pucpr.br).

$$MSE(L) = E\{((\frac{1}{L}\sum_i^L f_i) - r)^2\} \tag{8}$$

$$= \bar{b}^2 + \frac{1}{L}\bar{v} + \frac{L-1}{L}\bar{c} \tag{9}$$

To determine the link between $MSE(L)$ and the ambiguity, which measures the amount of variability among classifier outputs in ensembles, we need to apply ambiguity decomposition. It has been proved [12] that, at a single data point, the quadratic error of the ensemble f_{ens} is guaranteed to be less than or equal to average quadratic error of the individual classifiers [12]:

$$(f_{ens} - r)^2 = \sum_i^L w_i(f_i - r)^2 - \sum_i^L w_i(f_i - f_{ens})^2 \tag{10}$$

where w_i is the weight of classifier f_i in the ensemble, and $0 \leq w_i \leq 1$. If every classifier f_i has the same output, then the second term is 0, and f_{ens} would be equal to the average quadratic error of the individual classifiers. Note that the ensemble function is a convex combination ($\sum_i^L w_i = 1$):

$$f_{ens} = \sum_i^L w_i f_i \tag{11}$$

For the $MSE(L)$ of this ensemble of classifiers, suppose that every classifier has the same weight, i.e. $\forall i, w_i = \frac{1}{L}$, so f_{ens} is merely the average function of all individual classifiers $f_{ens} = \bar{f}$. Consequently the ambiguity decomposition can be written as:

$$(\bar{f} - r)^2 = \frac{1}{L}\sum_i^L (f_i - r)^2 - \frac{1}{L}\sum_i^L (f_i - \bar{f})^2 \tag{12}$$

Note that its expectation is exactly eq.8 and eq.9:

$$E\{\frac{1}{L}\sum_i^L (f_i - r)^2 - \frac{1}{L}\sum_i^L (f_i - \bar{f})^2\} = \bar{b}^2 + \frac{1}{L}\bar{v} + \frac{L-1}{L}\bar{c} \tag{13}$$

The ambiguity is the second term on the left-hand side in eq.13, and it can be written as [12]:

$$E\{(\frac{1}{L}\sum_i^L (f_i - \bar{f})^2)\} \tag{14}$$

$$= \frac{1}{L}\sum_i^L E\{(f_i - E\{f_i\})^2\} - E\{(\bar{f} - E(\bar{f}))^2\} \tag{15}$$

$$= \bar{v} - var(\bar{f}) = \bar{v} - \frac{1}{L}\bar{v} - \frac{L-1}{L}\bar{c} \tag{16}$$

The first term of the left-side in eq.13 is the sum of averaged bias and averaged variance of classifiers:

$$E\{\frac{1}{L}\sum_i^L (f_i - r)^2\} = \bar{b}^2 + \bar{v} \tag{17}$$

As stated in [9], the term $\bar{v}$, the average variance, exists in both the ambiguity part and the non-ambiguity part of $MSE(L)$. This means that we cannot simply maximize the ambiguity without affecting the bias component of $MSE(L)$. When we try to maximize the ambiguity among classifiers, we actually maximize the difference between its variance $\bar{v}$ and its covariance $\bar{c}$. If the term $\bar{v}$ increases, the non-ambiguity part of $MSE(L)$ will increase too.

III. PROPOSED COMPOUND DIVERSITY FUNCTIONS

Even though the ambiguity among classifiers is not a guarantee of ensemble accuracy, it has been shown that this ambiguity is a necessary condition for an ensemble to achieve a high degree of accuracy [1, 7, 10]. To compensate for the ambiguity dilemma, and to use its intrinsic property to reduce $MSE(L)$, we propose compound diversity functions using the diversity measures in a pairwise fashion to estimate ensemble accuracy. First, suppose that we have an ensemble with 2 classifiers f_i, f_j, and that classifiers f_i and f_j have the recognition rates a_i and a_j respectively, and the diversity d_{ij} is measured between them. With only two classifiers, we get $L = 2$ in eq.6 and eq.7. As a result, the ambiguity between f_i and f_j is exactly half of the difference between their variance and covariance in eq.16:

$$amb_{ij} = \frac{1}{2}(\bar{v} - \bar{c}) \tag{18}$$

$$= \frac{1}{4}(E\{(f_i - E\{f_i\})^2\} + E\{(f_j - E\{f_j\})^2\} - 2 \cdot E\{(f_i - E\{f_i\}) \cdot (f_j - E\{f_j\})\}) \tag{19}$$

If we use $L = 2$ in eq.9 and replace $\frac{1}{2}(\bar{v} - \bar{c})$ by amb_{ij}, we can write $MSE(2)$ as:

$$MSE(2) = \bar{b}^2 + \frac{1}{2}(\bar{v} + \bar{c}) = amb_{ij} + \bar{b}^2 + \bar{c} \tag{20}$$

As a result of this decomposition, there are basically two $MSE(2)$ terms, the first being the ambiguity of the ensemble, and the second being the sum of the averaged covariance and the averaged bias of individual classifiers. Using the eq.18, we can also write the above equation as:

$$MSE(2) = \bar{b}^2 + \bar{v} - \frac{1}{2}(\bar{v} - \bar{c}) = \bar{b}^2 + \bar{v} - amb_{ij} \tag{21}$$

where $amb_{ij} = \frac{1}{2}(\bar{v} - \bar{c})$. Suppose that the individual classifiers have a similar $MSE(f)$, one could obtain an proportional approximation of $(\bar{b}^2 + \bar{v})$ by the error rates of individual classifiers:

$$(\bar{b}^2 + \bar{v}) \approx \gamma((1 - a_i) \cdot (1 - a_j))^{\frac{1}{2}} \tag{22}$$

Suppose that $d_{ij} \propto amb_{ij}$, given that $0 \leq d_{ij} \leq 1$. The term d_{ij} will have a high correlation with $amb_{ij} = \frac{1}{2}(\bar{v} - \bar{c})$, and the approximation of $\frac{1}{2}(\bar{v} - \bar{c})$ can be written as:

$$(\bar{v} - \bar{c}) \approx \delta \cdot d_{ij} \tag{23}$$

If we assume that classifier-error and classifiers-diversity have similar weights in $MSE(2)$:

$$|\bar{b}^2 + \bar{v} - \frac{1}{2}(\bar{v} - \bar{c})| \leq \epsilon(\bar{b}^2 + \bar{v} + \frac{1}{2}(\bar{v} - \bar{c})), \epsilon \ll 1 \tag{24}$$

then the value $MSE(2)$ can be approximated as the product of the error rates of each classifier and their pairwise diversity. Given $0 \leq d_{ij} \leq 1$, we have $0 \leq 1 - d_{ij} \leq 1$, and we define an index for the proportional approximation of $MSE(2)$ as:

$$\widetilde{MSE}_{ij} \equiv (1 - d_{ij}) \cdot ((1 - ai) \cdot (1 - aj))^{\frac{1}{2}} \tag{25}$$

For multiple classifiers, we can regard them as a network of classifier-pairs (Fig.1). Consequently, the approximation of $MSE(L)$ will depend on the individual classifier errors and the respective diversity between classifiers. Given the number of selected classifiers $L \geq 2$, and we have $\widetilde{MSE}(L) \sim (\prod_{i=1}^{L}(1-a_i))^{\frac{1}{L}}(\prod_{i,j=1,i\neq j}^{L}(1-d_{i,j}))^{\frac{1}{L\times(L-1)}}$. Fig.1 shows the mechanism of such an approximation. Each circle represents the error rate of each individual classifier, and each line represents the diversity among them.

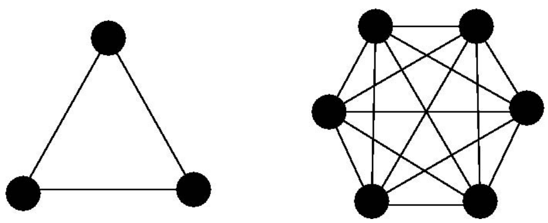

Fig. 1. The relationship between the error rate of individual classifiers (circles) and the diversity between them (lines).

In the case where the diversity measures represent the ambiguity, we combine the diversity measures with the error rates of each individual classifier:

$$\widehat{div_{amb}} = (\prod_{i=1}^{L}(1 - a_i))^{\frac{1}{L}} (\prod_{i,j=1,i\neq j}^{L}(1 - d_{i,j}))^{\frac{1}{L\times(L-1)}} \tag{26}$$

Apparently we have $\frac{L\times(L-1)}{2}$ diversity measures on different classifier-pairs. $\widehat{div_{amb}}$ is an estimation of the likelihood of common error being made by all classifiers, and we expect $\widehat{div_{amb}}$ to have negative correlation with ensemble accuracy. If the diversity measures represent the similarity, the proposed compound diversity function should be:

$$\widehat{div_{sim}} = (\prod_{i=1}^{L}(1 - a_i))^{\frac{1}{L}} (\prod_{i,j=1,i\neq j}^{L} d_{i,j})^{\frac{1}{L\times(L-1)}} \tag{27}$$

where $d_{i,j}$ should be interpreted as the similarity between f_i and f_j in this case. So, $\widehat{div_{sim}}$ ought to mean the likelihood of a common error being by all the classifiers. We expect negative correlation between the $\widehat{div_{sim}}$ and ensemble accuracy. The proposed compound functions are based on diversity measured in a pairwise manner, ensembles with fewer classifiers are more likely to be favored in the ensemble selection. So functions with various number of classifiers shall be rescaled by:

$$\widehat{div_{amb}} = \frac{L}{L-1}(\prod_{i=1}^{L}(1 - a_i))^{\frac{1}{L}} (\prod_{i,j=1,i\neq j}^{L}(1 - d_{i,j}))^{\frac{1}{L\times(L-1)}} \tag{28}$$

$$\widehat{div_{sim}} = \frac{L}{L-1}(\prod_{i=1}^{L}(1 - a_i))^{\frac{1}{L}} (\prod_{i,j=1,i\neq j}^{L} d_{i,j})^{\frac{1}{L\times(L-1)}} \tag{29}$$

TABLE I

UCI DATA FOR ENSEMBLES OF CLASSIFIERS. C: NUMBER OF CLASSES; TR: NUMBER OF TRAINING SAMPLES; TS: NUMBER OF TEST SAMPLES; FEAT: NUMBER OF TOTAL FEATURES. RS: CARDINALITY OF FEATURES FIXED IN RANDOM SUBSPACES.

Database	C	Tr	Ts	Feat	RS
Wisconsin Breast-Cancer	2	284	284	30	5
Satellite	6	4435	2000	36	4
Image Segmentation	7	210	2100	19	4
Letter Recognition	26	10007	9993	16	12

IV. CORRELATIONS BETWEEN DIVERSITY AND ENSEMBLE ACCURACY

We carried out experiments on problems extracted from UCI machine learning repository with Random Subspaces (See Table 1). Random Subspaces ensemble creation method was used because we need to have a large diversity among classifiers to better observe the effect of diversity on ensemble accuracy.

For each of 4 databases, for each of 3 classification algorithms, 18 classifiers were generated as the pool for base classifiers. Three different classification algorithms used in our experiments, including Naive Bayesian Classifiers (NBC), Quadratic Discriminant Classifiers (QDC), and 5-Layer Neural Network Classifiers (NNC) with Back-Propagation [6]. Ensembles were composed from $3 \sim 15$ classifiers, we evaluated 13 different numbers of classifiers for ensembles. 50 ensembles were constructed with the same number of selected classifiers for each database, we repeated this process 30 times to obtain a reliable evaluation. The simple majority voting rule is used as the fusion function. A total of $3 \times 4 \times 13 \times 50 \times 30 = 0.234$ million ensembles should be evaluated.

We measured ensemble accuracy correlation on 10 traditional diversity measures, including the disagreement measure (DM), the double-fault (DF), Kohavi-Wolpert variance (KW), the interrater agreement (INT), the entropy measure (EN), the difficulty measure (DIFF), generalized diversity (GD), coincident failure diversity (CFD), Q-statistics (Q), and the correlation coefficient (COR) [1, 2, 4, 8], as well as on 10 respective proposed compound diversity functions (eq.26 & eq.27). They are also compared with the Mean Classifier Error (ME) of individual classifiers.

In the Table 2, we show the correlations between original diversity measures and ensemble accuracy, and the correlation between compound diversity functions and ensemble accuracy. First, we observe that ME has an apparent correlation with ensemble accuracy (Table 2). Furthermore, it shows that, in general, compound diversity functions give better results than the original diversity measures. Of all the diversity measures, Q, COR, INT and DIFF are not stable. By contrast, DM, DF, KW, EN, GD and CFD are quite reliable. We found that the strongest correlation with ensemble accuracy on the

TABLE II

CORRELATION BETWEEN ENSEMBLE ACCURACY AND: (A) MEAN CLASSIFIER ERROR; (B) THE AVERAGE OF PURE DIVERSITY MEASURES; (C) THE PROPOSED COMPOUND DIVERSITY FUNCTIONS. THE SIGN ↑ MEANS THAT THE POSITIVE CORRELATION IS FORESEEABLE, AND THE SIGN ↓ INDICATES THAT NEGATIVE CORRELATION IS EXPECTED.

	Breast Cancer	Satellite	Image Segmentation	Letter Recognition
ME (↓)	-0.44 ± 0.13	-0.58 ± 0.07	-0.61 ± 0.01	-0.47 ± 0.05
Original Diversity Measures	**Breast Cancer**	**Satellite**	**Image Segmentation**	**Letter Recognition**
DM (↑)	-0.02 ± 0.08	0.08 ± 0.04	-0.19 ± 0.08	-0.06 ± 0.04
DF (↓)	-0.39 ± 0.10	-0.12 ± 0.06	-0.47 ± 0.08	-0.38 ± 0.04
KW (↑)	-0.02 ± 0.12	0.08 ± 0.04	-0.19 ± 0.04	-0.06 ± 0.10
INT (↓)	-0.36 ± 0.13	-0.08 ± 0.08	0.00 ± 0.10	-0.03 ± 0.05
EN (↑)	-0.02 ± 0.08	0.08 ± 0.04	-0.19 ± 0.08	-0.06 ± 0.08
DIFF (↓)	0.24 ± 0.09	-0.13 ± 0.12	0.55 ± 0.07	0.14 ± 0.04
GD (↑)	0.29 ± 0.09	0.08 ± 0.06	0.35 ± 0.08	0.14 ± 0.07
CFD (↑)	0.30 ± 0.09	0.08 ± 0.07	0.36 ± 0.09	0.15 ± 0.07
Q (↓)	-0.17 ± 0.14	-0.08 ± 0.08	0.11 ± 0.08	0.05 ± 0.07
COR (↓)	-0.36 ± 0.13	-0.08 ± 0.08	0.01 ± 0.10	-0.03 ± 0.10
Compound Diversity Functions	**Breast Cancer**	**Satellite**	**Image Segmentation**	**Letter Recognition**
DM (↓)	-0.64 ± 0.12	-0.46 ± 0.11	-0.43 ± 0.11	-0.44 ± 0.10
DF (↓)	-0.49 ± 0.08	-0.47 ± 0.03	-0.51 ± 0.02	-0.49 ± 0.02
KW (↓)	-0.54 ± 0.12	-0.53 ± 0.11	-0.76 ± 0.11	-0.50 ± 0.10
INT (↓)	-0.24 ± 0.09	-0.05 ± 0.08	-0.10 ± 0.10	-0.15 ± 0.12
EN (↓)	-0.64 ± 0.12	-0.46 ± 0.11	-0.43 ± 0.11	-0.44 ± 0.10
DIFF (↓)	-0.33 ± 0.12	-0.29 ± 0.11	0.07 ± 0.02	-0.12 ± 0.06
GD (↓)	-0.46 ± 0.08	-0.50 ± 0.07	-0.60 ± 0.07	-0.49 ± 0.12
CFD (↓)	-0.43 ± 0.08	-0.46 ± 0.07	-0.53 ± 0.07	-0.45 ± 0.12
Q (↓)	-0.34 ± 0.11	-0.24 ± 0.10	-0.12 ± 0.11	-0.44 ± 0.11
COR (↓)	-0.25 ± 0.09	-0.05 ± 0.08	-0.10 ± 0.10	-0.15 ± 0.12

minimum number of classifiers, i.e. when ensembles were constructed with only 3 classifiers. But this correlation could decrease to nearly 0 when the number of classifiers is close to the total number of classifiers available in the pool.

V. CONCLUSION

Previous published studies suggested that diversity is not unequivocally related to ensemble accuracy, and it is our objective to demonstrate that the implementation of diversity can help in ensemble selection. As we can see in these experiments, there are strong correlations between the proposed compound diversity functions and ensemble accuracy. The result also suggests that DM, KW, EN, GD and CFD are stable for all ensemble creation methods. Performance depends strongly on the accuracy of individual classifiers, but, in general, an equivalent or stronger correlation could be achieved with compound diversity functions, especially with KW.

The result encourages further exploration of the implementation of compound diversity functions, and the pertinence of these functions for use with different searching algorithms. Moreover, it suggests that the problem resides in finding ways to amalgamate diversities and individual classifier errors, rather than allowing diversity measures to select EoCs single-handedly. Another advantage of compound diversity functions is that they can be calculated beforehand, since diversities are measured in a pairwise manner, and error rates are measured on each classifier; thus, for time-consuming searching methods, such as GA or exhaustive searching, ensemble accuracy can be estimated quickly by simply calculating the products of the diversity measures and individual classifier errors, which is much faster than other objective functions.

ACKNOWLEDGMENT

This work was supported in part by grant OGP0106456 to Robert Sabourin from the NSERC of Canada.

REFERENCES

[1] 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

[2] T.K. Ho, "The random space method for constructing decision forests," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 8, pp. 832-844, 1998

[3] G. Brown, J. Wyatt and P. Sun, "Between Two Extremes: Examining Decompositions of the Ensemble Objective Function," *International Workshop on Multiple Classifier Systems (MCS 2005)*, pp. 296-305, 2005

[4] G. Giacinto and F. Roli, "Design of effective neural network ensembles for image classification purposes," *Image and Vision Computing*, vol. 19, no. 9-10, pp. 699-707, 2001

[5] D. Ruta and B. Gabrys, "Classifier Selection for Majority Voting," *International Journal of Information Fusion*, pp. 63-81, 2005

[6] R.P.W. Duin, "Pattern Recognition Toolbox for Matlab 5.0+," available free at: ftp://ftp.ph.tn.tudelft.nl/pub/bob/prtools

[7] E. Pekalska, M. Skurichina and R. P. W. Duin, "Combining Dissimilarity-Based One-Class Classifiers," *International Workshop on Multiple Classifier Systems (MCS 2004)*, pp. 122-133, 2004

[8] D. Partridge and W. Krzanowski, "Software diversity: practical statistics for its measurement and exploitation," *Information and Software Technology*, vol. 39, pp. 707-717, 1997

[9] G. Brown, J. Wyatt, R. Harris and X. Yao, "Diversity Creation Methods: A Survey and Categorisation," *International Journal of Information Fusion*, vol. 6, no. 1, pp. 5-20, 2005

[10] H. Zouari, L. Heutte, Y. Lecourtier and A. Alimi, "Building Diverse Classifier Outputs to Evaluate the Behavior of Combination Methods: the Case of Two Classifiers," *International Workshop on Multiple Classifier Systems (MCS 2004)*, pp. 273-282, 2004

[11] B. E. Geman, S. and R. Dorsat, "Neural Networks and the Bias/Variance Dilemma," *Neural Computation*, no. 4, pp. 1-58, 1992

[12] A. Krogh and J. Vedelsby, "Neural Network Ensembles, Cross Validation, and Active Learning," *Advances in Neural Information Processing Systems*, vol. 7, pp. 231–238, 1995

[13] L. I. Kuncheva, M. Skurichina, and R. P. W. Duin, "An Experimental Study on Diversity for Bagging and Boosting with Linear Classifiers," *International Journal of Information Fusion*, vol. 3, no. 2, pp. 245-258, 2002

[14] J. Kittler, M. Hatef, R. Duin, and J. Matas, "On Combining Classifiers," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 20, no. 3, pp. 226–239, 1998

[15] K. Tumer and J. Ghosh, "Error Correlation and Error Reduction in Ensemble Classifiers", *Connection Science*, vol. 8, no. 3-4, pp. 385–403, 1996

[16] G. James, "Variance and Bias for General Loss Functions," *Machine Learning*, vol. 51, no. 2, pp. 115-135, 2003.

[17] P. Domingos, "A Unified Bias-Variance Decomposition and its Applications," *International Conference on Machine Learning (ICML 2000)*, pp. 231-238, 2000

[18] B. E. Rosen, "Ensemble Learning using Decorrelated Neural Networks", *Connection Science*, vol. 8, no. 3-4, pp. 373–384, 1996