Classifier Performance Measures in Multi-Fault Diagnosis for Aircraft Engines

Weizhong Yan, Kai Goebel

Information & Decision Technology Lab
GE Global Research Center
One Research Circle,
Niskayuna, NY 12309
{yan, goebelk} @crd.ge.com

James C. Li

Department of Mechanical Engineering,
Aeronautical Engineering & Mechanics
Rensselaer Polytechnic Institute
Troy, NY 12181
lic3@rpi.edu

ABSTRACT

Classifier performance evaluation is an important step in designing diagnostic systems. The purposes of performing classifier performance evaluation include: 1) to select the best classifiers from the several candidate classifiers, 2) to verify that the classifier designed meets the design requirement, and 3) to identify the need for improvements in the classifier components. In order to effectively evaluate classifier performance, a classifier performance measure needs to be defined that can be used to measure the goodness of the classifiers considered. This paper first argues that in fault diagnostic system design, commonly used performance measures, such as accuracy and ROC analysis are not always appropriate for performance evaluation. The paper then proposes using misclassification cost as a general performance measure that is suitable for binary as well as multi-class classifiers, and -most importantly- for classifiers with unequal cost consequence of the classes. The paper also provides strategies for estimating the cost matrix by taking advantage of fault criticality information obtained from FMECA. By evaluating the performance of different classifiers considered during the design process of an engine fault diagnostic system, this paper demonstrates that misclassification cost is an effective performance measure for evaluating the performance of multi-class classifiers with unequal cost consequence for different classes.

Keywords: Classifier performance, performance measures, aircraft engine, fault diagnosis, ROC

1. INTRODUCTION

Typical diagnostic system design involves several steps, including data preprocessing, feature extraction and selection, classifier design, and classifier performance evaluation, as shown in Figure 1. Classifier performance evaluation is an indispensable step in diagnostic system design. This is because the same classifier performs differently from application to application, i.e., classifier performance is problem specific[1]. Given that no single classifier is always superior over others for all applications, common practice for designing classifier for a given problem, therefore, involves experimenting with many different classifiers, comparing their performance, and selecting the classifier (individual or combined) with the best performance. Obviously, in this design practice, classifier performance evaluation is essential. Performing classifier performance evaluation is required not only for selecting the best classifiers from the several candidate classifiers, but also for verifying that the classifier designed meets the design requirement and for identifying the need for improvements in the classifier components.

Classifier performance generally refers to both computational performance and classification performance. In this paper, however, we limit our study to classification performance only. In order to effectively evaluate classifier performance, a classifier performance measure has to be defined. A classifier performance measure is a single index that measures the goodness of the classifiers considered. Depending on the design or application requirements, different problems may call for different performance measures to ensure that the classifiers considered can be properly compared and selected. Given a

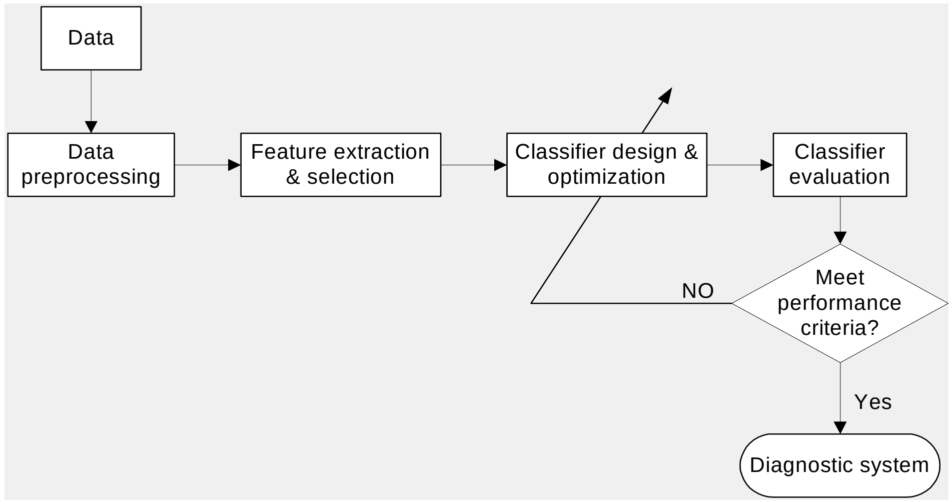

Figure 1: Typical fault diagnostic system design process

problem at hand, it is not always trivial to define a good performance measure so that the classifier performance can be accurately measured. In fact, defining/identifying a proper performance measure can be difficult for the cases when more comprehensive classifiers are required for the diagnostic system of a complex system.

Fault diagnostic systems have proved to be an effective technology in improving the reliability and in reducing the operating costs of various mechanical systems. There are great benefits for increasing the reliability/performance of fault diagnostic systems. However, as the systems become increasingly complex, designing a reliable fault diagnostic system becomes more difficult. Frequently, reasonable amount of design efforts result in only a small improvement in the performance of the diagnostic system designed. However, to the owner of the underlying system, even a small amount of improvement in reliability/performance of the diagnostic system can result in significant benefits. To uncover the subtle performance difference between one design and another, the performance measure used for classifier evaluation needs to be better defined to accurately represent the classifier performance. Diagnostic systems are typically designed to detect and isolate several different faults with different fault criticality. Consequently, classifier evaluation should take into account the difference between classifiers that have different misclassification costs for individual faults. In this paper, we illustrate that the commonly used performance measures (accuracy and ROC analysis) are not always adequate for evaluating the performance of typical fault diagnostic systems. To overcome this problem, a more general performance measure, misclassification cost is utilized. In addition, this paper provides means for estimating the cost matrix required by the misclassification cost measure based on fault criticality information.

The reminder of the paper is organized as follows: Section 2 reviews commonly used classifier performance measures. Limitations of each of the measures are highlighted. Section 3 details the misclassification cost measure that is suitable for evaluating the performance of multi-class classifiers with unequal cost consequence of the classes. Emphasis of this section is on developing the strategy for estimating the cost matrix based on fault criticality. Section 4 demonstrates how the misclassification cost measure is used for evaluating the performance of the classifiers for an aircraft engine fault diagnosis system. Section 5 offers conclusions.

2. COMMON CLASSIFIER PERFORMANCE MEASURES

Common measures used for classifier performance evaluation include overall classification rate (accuracy) and ROC analysis. Descriptions of the two measures are given in this section. Advantages and limitations of the respective measures are discussed.

2.1. Overall classification rate

The most common measure of classifier performance is the overall classification rate (or alternatively its equivalent term, *error rate,* that is equal to one minus the overall classification rate). The overall classification rate, also called accuracy, is defined as the ratio of number of cases that are correctly classified over the total number of cases. Let CM be the M-by-M confusion matrix (where M is the number of classes), then the overall classification rate (OCR) is expressed as:

$$OCR = \frac{1}{N}\sum_{i=1}^{M} CM(i,i) \tag{1}$$

where, N is the total number of test cases.

This single performance measure is fairly easy to compute. It is also suitable for all kinds of classifiers. The underlying assumption of the overall classification rate, however, is that the classification errors for all classes have equal cost consequences. Since this assumption rarely holds, the overall classification rate is often not an appropriate measure of classifier performance[2]. Additional limitations of the overall classification rate as a performance measure include that it is sensitive to the unequal class size and that it is not reflect the performance of the classifier across the entire range of possible decision thresholds[3].

Some efforts have been made to make the accuracy measure useable for unequal error cost cases. For example, Breiman et al[4] have proposed stratifying the classes based on the target cost and class distribution so that maximizing accuracy on the transformed data corresponds to minimizing costs on the target data. However, this strategy only works for 2-class problems and requires precise "true" class distribution, which are not available for most of real world problems.

2.2. ROC Analysis

ROC (short for Receiver Operating Characteristic or Relative Operating Characteristic) analysis is an established method of measuring diagnostic performance in various domains, especially in medical imaging studies. Originated from the field of signal detection to depict tradeoffs between hit rate and false alarm rate[5], ROC analysis and its associated indices have recently been extended for use in evaluating performance of 2-class classifiers [3,6]. The ROC space is a coordinate system that

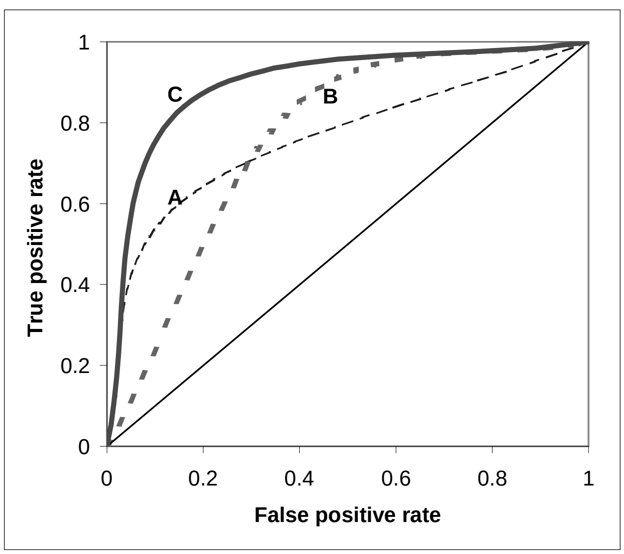

Figure 2: Typical ROC curves

is used for visualizing classifier performance. In ROC space, the true positive rate, TPR, is plotted on the Y-axis and the false positive rate, FPR, is plotted on the X-axis, where the TPR is commonly referred to as "sensitivity" while (1-FPR) is called "specificity". A point in ROC space corresponds to a (FPR, TPR) pair of a classifier. By varying the parameters of the classifier, a series of points are obtained and a ROC curve is generated in ROC space by connecting these points. Typical ROC curves are shown in Figure 2 where the three ROC curves represent three different classifiers. Classifiers with ROC curves located in the upper-left corner in ROC space are better because they represent classifiers that have lower false positive rate and higher true positive rate than the classifiers below them.

ROC curves are a valuable technique for visualizing classifier behavior over a range of decision rules, therefore ROC curves are thus frequently used for selecting a suitable operating point, or a decision threshold, for the task at hand. However, when used for comparing or ranking classifiers based on their performance, evaluation based on ROC curves becomes more involved when the curves overlap. Use Figure 2 as an example where the three ROC curves correspond to three different classifiers A, B, and C. One can easily conclude that classifier C is better than or at least as good as the other two classifiers for all possible cost and class distributions since curve C dominates others in all range. Determining which of the two classifiers (A and B) is better, on the other hand, would not be so straightforward unless a specific performance requirement is given. For example, given the maximum number of false positive rate, one would simply draw a vertical line at the specified maximum FPR, and rank the classifiers based on TPR at the intersection of ROC curves with this vertical line. Similarly, one would use a horizontal line to rank the classifiers if the maximum true positive rate is given.

When either the cost distribution or the class distribution is completely unknown and one wants to use a simple, single quantitative index to represent the entire ROC curve, the *area under the curve* (AUC) [6] is the most popular performance measure for ranking or comparing 2-class classifiers. The AUC provides a measure of performance that is not sensitive to the prior probability of class occurrence and measures the performance of the classifier across the entire range of decision thresholds [2].

As noted above, all ROC analysis and its associated performance measures were developed for *2-class* problems only, which is somewhat of a drawback of ROC analysis. This drawback restricts the ROC analysis from much wider applications.

3. MISCLASSIFICATION COST AS PERFORMANCE MEASURE

A general classifier performance evaluation poses the following requirements on performance measures:

- First, performance measures need to be more accurate in order to uncover the subtle difference between one classifier and another. This is because small reliability improvement of diagnostic systems can bring significant benefits to the owner of the underlying system.
- Second, performance measures need to be able to handle performance evaluation of multi-class classifiers. Most diagnostic systems are required to detect and isolate several different possible faults. With each of the classifier outputs representing one fault, most of diagnostic systems are multi-class classifiers.
- Last, performance measures need to be able to represent criticality differences between faults, featuring in that the consequence cost of each fault is different.

These requirements make the commonly used performance measures discussed in the previous section inappropriate or inaccurate at best for evaluating fault diagnostic systems. More specifically, using the overall accuracy measure for evaluating classifiers with different consequence cost of each fault would not be accurate due to the underlying assumption of equal cost for misclassifying all faults. On the other hand, since ROC analysis and its associated indices work for 2-class classifiers only, using the ROC analysis as the performance measure for multi-class classifiers would be too complex and would lose visualization merits of ROC curves. To address these problems, we use *misclassification cost* [7] as a general classifier performance measure for evaluating fault diagnostic systems. The misclassification cost (MC) is defined as the product of each element of the normalized confusion matrix and the corresponding element of the cost matrix and summing the results, expressed in Equation 2 as follows:

$$MC = \sum_{i,j} \overline{cm}(i,\, j) \cdot C(i,\, j) \tag{2}$$

where, $\overline{cm}(i,\, j) = CM\,(i,\, j) \big/ \sum CM\,(i)$ is the normalized confusion matrix; the cost matrix and will be discussed later in this section.

The misclassification cost defined in Equation 2 has been used by Margineantu and Dietterich[8] for designing cost-sensitive classifiers. One can see from Equation 2 that the overall classification rate or accuracy is a special case of the misclassification cost. That is, when the cost matrix has a value of 1 on its diagonal terms and zeros on all off-diagonal terms, the misclassification cost becomes accuracy. In that sense, the misclassification cost measure is a general form of the accuracy measure. The most significant advantages of the misclassification cost measure is that it can be used for multi-class classifiers and can take care of classifiers with different costs for different classes through proper definition of the cost matrix.

As shown in Equation 2, two elements of misclassification costs are the normalized confusion matrix and the cost matrix. While the normalized confusion matrix can be readily calculated from the confusion matrix of the classifier, precisely determining a cost matrix can be more difficult, requiring domain expertise inputs or involvement of experts in different areas from design to practice to field engineering. In this paper, we focus on strategies of estimating the cost matrix based on the fault criticality information from FMECA of the underlying system so that misclassification cost can be used as a measure to evaluate the classifiers used for fault diagnostic systems.

A cost matrix is a matrix where each cell, $C(i, j)$, represents the cost incurred for misclassification, i.e., when a case is predicted to be in class *j* when in fact it belongs to class *i*. Based on this definition, it is obvious that all diagonal cells of a cost matrix should have a zero value. We know that for a typical fault diagnosis problem, different faults have different consequence. For example, in aircraft engine fault diagnosis, a fan fault will cause more severe damage to the engine than a variable bleed valve (VBV) leak does. Hence, misclassifying a critical fault (fan fault) as a non-critical fault (VBV fault) should have different cost consequence from misclassifying a non-critical fault as a critical fault. Generally, cost of misclassifying fault “a” as fault “b” should not be the same as that of misclassifying fault “a” as fault “c” if the criticalities of faults “b” and “c” are different. Capturing this difference into performance measures is the key for better evaluation of classifier performance. Thus, a full cost matrix is generally a non-symmetric matrix.

In modern mechanical system design, Failure Modes, Effects and Criticality Analysis (FMECA) is used as a tool to gain an initial measure of system reliability. The outcome of FMECA includes a list of failure modes of the system and associated failure effects and failure criticality. The failure/fault criticality from FMECA provides valuable information and can be used for estimating the cost matrix.

Given the fault criticality ranking and based on the notion that a more critical fault typically results in more costly effects, one could derive a *heuristic* cost matrix such that the numbers of all cells in the same row are constant by dividing the i^{th} ranking score by the sum of the ranking scores. However, such obtained cost matrix is not different from weighting each class based on its fault criticality, which is still not able to differentiate the misclassification costs between two different faults.

In order to estimate a full cost matrix, we propose to capture two basic notions: 1) the cost of misclassifying i^{th} fault as j^{th} fault is different from that of misclassifying j^{th} fault as i^{th} fault if i and j are different; and 2) the cost of misclassifying i^{th} fault as j^{th} fault is higher if the ordered criticality ranking of j^{th} fault is further away from that of i^{th} fault.

Table 1: Fault list and ranking by criticality

Faults	**F1**	**F2**	**F3**	**…**	**Fn**
Ranking	R1	R2	R3	…	Rn

Consider *n* faults and their corresponding ordered criticality rankings as shown in Table 1. Let R_i and R_j be the fault criticality rankings for i^{th} and j^{th} faults, respectively. We then define $d_{ij} = R_i - R_j$ as the distance that measures how far apart the two faults are in the criticality ranking. Such defined distance also represents the “degree of misclassification” when i^{th} fault is misclassified as j^{th} fault. Similar to confusion matrices, distance or “degree of misclassification” between each pair of faults can be represented in a matrix as shown in Table 2.

Table 2: Matrix representing degree of misclassification

		Predicted Faults			
		F1	**F2**	**….**	**Fn**
True faults	**F1**	*0*	d_{12}	….	d_{1n}
	F2	d_{21}	*0*	….	d_{2n}
	:	:	:	….	:
	Fn	d_{n1}	d_{n2}	….	*0*

Based on the definition of d_{ij}, the value of d_{ij} can be positive or negative. While a positive value of d_{ij} means that criticality ranking for i^{th} fault is higher than that for j^{th} fault, i.e., i^{th} fault is more critical than j^{th} fault, a negative value of d_{ij} means that i^{th} fault is less critical than j^{th} fault.

Intuitively, the matrix representing the "degree of misclassification" should be directly related to the misclassification cost matrix. Therefore, we compute the cost matrix, C_{ij}, in terms of "degree of misclassification", d_{ij}, as follows.

$$C_{ij} = \begin{cases} \dfrac{d_{ij}}{\sum_i R} \cdot S & \text{for } d_{ij} > 0 \\ -m \cdot \dfrac{d_{ij}}{\sum_i R} \cdot S & \text{for } d_{ij} \le 0 \end{cases} \tag{3}$$

where $\sum R$ in the denominator of Equation 3 is the sum of the values of the criticality rankings and is used for normalization purpose. The factor *m*, $\{m \le 1\}$ used for $d_{ij} < 0$ case in Equation 3 captures the notion that misclassifying a less critical fault as a more critical fault is typically less costly than misclassifying a critical fault as a less critical fault. For classifier performance evaluation, only relative values of cost matrix matter, i.e., scaling a cost matrix with a constant will not change the classifier evaluation results. Therefore, the relationship between C_{ij} and d_{ij} is unique, but can be scaled. The particular scaling is performed with the domain-specific constant cost scaling parameter, *S*, in Equation 3.

4. EVALUATING AN AIRCRAFT ENGINE FAULT DIAGNOSTIC SYSTEM

In this section, the misclassification cost measure is used to evaluate the performance of an aircraft engine fault diagnostic system. The cost matrix estimation based on fault criticality discussed in the previous section is also validated through the engine fault diagnostic system application.

Aircraft engine fault diagnostic systems have proved to be an effective technology in improving the reliability and safety of aircraft. Because the full potential of the engine fault diagnostic systems has not yet been fully exhausted, strategies/technologies on improving the performance of engine fault diagnostic systems are needed. To that end, several studies have recently been undertaken using soft computing technologies to enhance the performance of engine fault diagnostic systems. The IMATE project [9] is such an example and is used in this paper as the baseline.

Three individual classifiers are considered here. They are 1) multiple-layer feed-forward neural network (NN), 2) binary decision tree (CART), and 3) Fuzzy neural network (FNN). Neural network has been one of the most popular classifiers used for various applications. Fuzzy neural networks are hybrid systems that integrate fuzzy inference system with neural networks, thus take advantage of the benefits of transparency of fuzzy sets and learning capacity of neural networks. CART (classification and regression tree) is a popular non-parametric approach for both classification and regression. Like other tree-based classifiers, CART is a recursive and iterative procedure and has been considered as a powerful, fast, and flexible classification tool. We favor the three classifiers over others because of their reasonable performance under the engine fault diagnosis application based on our preliminary study.

The three classifiers are designed and their structure is individually optimized such that the performance of the individual classifiers is maximized. The design and optimization of the individual classifiers are based on a complete set of simulated engine performance data. The simulation that is based on a real-time, nonlinear engine model to covers six different faults, five different levels of engine deterioration and random engine initial quality.

The six engine hot gas path faults considered for a commercial engine are fan damage (FAN), compressor fault (COMP), high-pressure turbine fault (HPT), low-pressure turbine fault (LPT), customer discharge pressure fault (CDP), and variable bleed valve leak (VBV). The relative criticality ranking of the six faults provided by the service engineers is shown in Table 3. A criticality ranking value of zero is assigned for the "no fault" (NF) class and included in Table 3.

Table 3: Relative fault criticality ranking

Faults	NF	FAN	COM	HPT	LPT	CDP	VBV
Ranking	0	0.9	0.8	0.8	0.8	0.1	0.1

Table 4: Matrix representing degree of misclassification

			Predicated Faults						
			NF	FAN	COM	HPT	LPT	CDP	VBV
			0	0.9	0.8	0.8	0.8	0.1	0.1
True Faults	NF	0	0	-0.9	-0.8	-0.8	-0.8	-0.1	-0.1
	FAN	0.9	0.9	0	0.1	0.1	0.1	0.8	0.8
	COM	0.8	0.8	-0.1	0	0	0	0.7	0.7
	HPT	0.8	0.8	-0.1	0	0	0	0.7	0.7
	LPT	0.8	0.8	-0.1	0	0	0	0.7	0.7
	CDP	0.1	0.1	-0.8	-0.7	-0.7	-0.7	0	0
	VBV	0.1	0.1	-0.8	-0.7	-0.7	-0.7	0	0

Table 5: Cost Matrix

		Predicted Faults						
		NF	Fan	Comp	HPT	LPT	CDP	VBV
True Faults	NF	0	23.1	20.6	20.6	20.6	2.6	2.6
	Fan	25.7	0	2.9	2.9	2.9	22.9	22.9
	Com	22.9	2.6	0	2.9	2.9	20.0	20.0
	HPT	22.9	2.6	2.9	0	2.9	20.0	20.0
	LPT	22.9	2.6	2.9	2.9	0	20.0	20.0
	CDP	2.9	20.6	18.0	18.0	18.0	0	2.9
	VBV	2.9	20.6	18.0	18.0	18.0	2.9	0

Table 4 is the matrix representing the degree of misclassification derived from the fault criticality ranking based on the strategies discussed in Section 3. The final cost matrix calculated from the degree of misclassification using Equation 3 is shown in Table 5, with m=0.9 and S=100. Since the critical ranking for engine faults considered in this paper contains three faults that have the same criticality, the distance or degree of misclassification between each pair of these three faults becomes zero (Table 4) based on our distance definition in the previous section. We assign a minimum distance value of 0.1

for these cells to account for the fact that any misclassification has cost even when the misclassified fault has the same criticality.

Table 6 contains the confusion matrices of different classifiers considered. These confusion matrices are the outputs of the designed classifiers under the test data set. To demonstrate the effectiveness of the misclassification cost as performance measure, the confusion matrices of three different neural network classifiers are included in Table 6. The three neural network classifiers designated as NN20, NN25, and NN30 have the same number of input and output nodes. However, the number of hidden nodes is 20, 25, and 30, respectively. Evaluating the performance of the three NN classifiers is necessary for determining the optimal NN structure.

Given the confusion matrices of the classifiers and the cost matrix (Table 5) estimated from the fault criticality, we can compute the misclassification costs for each classifier based on Equation 2 in Section 3. It is noted that prior to computing the misclassification costs, the cell values of the confusion matrices are normalized by dividing each cell in a row of the confusion matrix by the total number of the test samples in that row. With the computed misclassification costs, the performance of classifiers can be ranked. Table 7 summarizes the misclassification costs and the corresponding ranking for all of the five classifiers considered. For comparison purpose, Table 7 also includes the overall classification rate (accuracy) that is computed based on Equation 1 and the performance ranking based on the overall classification rate.

Examining the performance evaluation results in Table 7, we can see that the misclassification costs for the five classifiers considered are different, with the NN25 classifier being the lowest and the FNN classifier being the highest in misclassification cost. The performance of the five classifiers thus can be uniquely ranked based on their calculated misclassification costs. Additionally, for this specific application, the misclassification cost measure gives us the similar ranking as the overall accuracy does, except for the top two classifiers (NN25 and NN30) where the accuracy measure fails to tell the difference. This is mainly because the classifiers coincidently make less number of misclassification for the more critical fault classes. Had the fault criticality ranking been in a different order, the results would have been completely different. For example, if a new fault criticality ranking is given as R={0.0, 0.95, 0.01, 0.1, 0.01, 0.01, 0.2}, the misclassification cost measure will give us a different performance ranking for the same five classifiers as shown in Table 8, while the accuracy as well as the ranking based on the accuracy measure for the five classifiers keep unchanged.

5. CONCLUSIONS

Classifier performance evaluation is an important step in designing fault diagnostic systems and a proper performance measure is the key for accurate classifier evaluation. Due to the characteristics (multi-fault with different fault criticality) of typical fault diagnostic systems, commonly used classifier performance measures are not appropriate for evaluating fault diagnostic systems. This paper identifies a more general performance measure, misclassification cost for fault diagnostic systems application. To circumvent the difficulties of precisely determining the cost matrices, this paper introduces a strategy for estimating the cost matrix based on fault criticality information derived from FMECA.

The application of the proposed misclassification cost measure for evaluating the classifiers for an aircraft engine fault diagnostic system demonstrates that the misclassification cost can be an effective and general performance measure for evaluating classifiers that are multi-class and have unequal cost consequence of the different classes. The same application also demonstrates that the strategy proposed in this paper for estimating the full cost matrix from the fault criticality information is effective.

Table 6: Confusion Matrices

Classifiers		Predicted Faults						
		NF	Fan	Comp	HPT	LPT	CDP	VBV
CART	NF	2074	0	126	106	347	96	56
	Fan	0	2781	24	0	0	0	0
	Comp	112	2	2577	36	55	22	1
	HPT	110	0	93	2419	179	3	1
	LPT	394	0	70	309	1864	146	22
	CDP	77	0	11	1	105	2539	72
	VBV	114	0	22	1	31	90	2547
FNN	NF	1460	0	174	316	463	278	114
	Fan	0	2805	0	0	0	0	0
	Comp	264	170	1995	72	159	134	11
	HPT	155	0	60	2314	241	35	0
	LPT	432	0	75	908	914	459	17
	CDP	69	0	6	0	157	2496	77
	VBV	167	0	32	2	34	78	2492
NN20	NF	2450	11	73	51	131	50	39
	Fan	4	2750	32	0	1	4	14
	Comp	84	56	2582	29	18	15	21
	HPT	60	7	97	2509	121	6	5
	LPT	188	5	34	162	2364	45	7
	CDP	83	5	21	2	27	2595	72
	VBV	47	6	14	7	4	52	2675
NN25	NF	2462	5	52	39	164	71	12
	Fan	0	2718	63	1	1	22	0
	Comp	97	31	2590	21	29	36	1
	HPT	56	11	88	2499	142	8	1
	LPT	151	1	27	141	2450	31	4
	CDP	68	9	12	1	37	2654	24
	VBV	59	2	16	0	9	113	2606
NN30	NF	2483	4	45	37	166	48	22
	Fan	1	2771	29	4	0	0	0
	Comp	96	30	2594	33	19	24	9
	HPT	62	16	93	2493	137	1	3
	LPT	180	6	36	153	2375	47	8
	CDP	81	5	16	0	23	2654	26
	VBV	68	21	12	0	4	91	2609

Table 7: Performance evaluation results

Classifiers	Based on misclassification cost measure		Based on accuracy measure	
	Cost	Ranking	Accuracy	Ranking
CART	13.03	4	85.57	4
FNN	22.54	5	73.73	5
NN20	7.02	3	91.29	3
NN25	6.61	1	91.57	1
NN30	6.72	2	91.57	1

Table 8: Performance evaluation results for a different fault criticality ranking

Classifiers	Based on misclassification cost measure		Based on accuracy measure	
	Cost	Ranking	Accuracy	Ranking
CART	6.29	2	85.57	4
FNN	14.25	5	73.73	5
NN20	6.68	3	91.29	3
NN25	6.75	4	91.57	1
NN30	5.83	1	91.57	1

REFERENCES:

1. S. Weiss and C. Kulikowski, (1991), *Computer Systems that Learn: Classification and Prediction Methods from Statistics, Neural Nets, Machine Learning, and Expert Systems*, Morgan Kaufmann Publishers, San Mateo, CA.
2. Provost, F. and Fawcett T. (1997), "Analysis and visualization of classifier performance: Comparison under imprecise class and cost distributions", Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining (KDD-97), pp43-48.
3. Downey, T.J., Meyer, D.J., Price, R.K., and Spitznagel, E.L., (1999), "Using the receiver operating characteristic to asses the performance of neural classifiers", IJCNN '99 – International Joint Conference on Neural Networks, Vol.5, pp3642-3646.
4. Breiman, L, Friedman, J.H., Olshen, R.A., and Stone, C.J., (1984), *Classification and regression trees*, Chapman & Hall/CRC, Florida.
5. Egan, J.P. (1975), *Signal detection theory and ROC analysis*, Series in Cognition and Perception. New York: Academic Press.
6. Bradley, A.P., (1997), "The use of the area under the ROC curve in the evaluation of machine learning algorithms", Pattern Recognition, Vol.30, No.7, pp1145-1159.
7. D. Michie, D.J. Spiegelhalter, and C.C. Taylor (Eds.) (1994), *Machine Learning, Neural and Statistical Classification*, Ellis Horwood, New York, NY.
8. Margineantu, D.D. and Dietterich, T.G. (2000), "Bootstrap methods for the cost-sensitive evaluation of classifiers", Processings of ICML-2000, pp583-590
9. Ashby, M. J. and Scheuren, W. J. (2000), "Intelligent maintenance advisor for turbine engines", proceedings of 2000 IEEE Aerospace Conference, Vol. 6, pp211-219.