HEALTH MONITORING OF DIMENSIONAL GAS TURBINE ENGINE (ENGDATA) USING ISOMAP DATA-BASED ANALYSIS APPROACH

ISSN: 2141 – 3290

JERRY E. E.[1], IMOH, J. E.[2], UDUAK, A. U.[2] AND EDWARD, N. U.[2]

[1]*Department of Turbo Machinery*
(Reliability and Maintenance),
Exxon Mobile, QIT, Eket, AKS, Nigeria
[2]*Department of Computer Science, University of Uyo, Uyo, Nigeria*

ABSTRACT: Maintenance of complex engineering systems such as Gas Turbine Engine (GTE) has posed a serious challenge to systems engineers, as this affects the GTE subsystems and entire system reliability and performance. Monitoring the health of a system is part of the predictive maintenance approach that seeks to extend the reliability and life of the system. In this study, we develop data-based model that monitors health condition of Gas Turbine Engines. This model analyses engine signal characteristics and extracts features based on Isometric Feature Mapping (ISOMAP) data-based analysis approach. We employ Nearest-neighbour classification technique to reduce dimensional data for fault diagnosis of the GTE. The systems data were classified into three classes namely; good GTE, average GTE and bad GTE. The data-based model was visualised in two-dimension using scatter plot, and their performance evaluated using M-fold cross validation. The model is implemented in Matlab and C++ programming tool. Our model gears towards minimising the downtime of gas turbine engines, improving the safety of plant operations, enhancing system reliability and availability, and reducing the operating cost of the system.

INTRODUCTION

Gas turbine engines have proven to be very efficient and are widely used in many industrial and engineering systems. In most cases, areas of application of gas turbine engines are safety critical which require very high reliability and availability of these systems. Gas Turbine Engine generates a huge amount of data, from the numerous measured variables, such as temperature, pressure, flow, vibrations, velocities, etc. To maintain high system reliability and availability, critical system parameter variables such as engine vibration, bearing temperature, lube oil pressure, etc, must be continuously monitored for prompt detection of deviation from normal operation values.

Isometric Feature Mapping technique is used to transform the high dimensional data space to low-dimensional data space but still preserve the local structure and information contained in the original data variables. The lower dimensional presentation of the data produced by ISOMAP facilitates faults detection and diagnosis when appropriate classification techniques are applied.

To design a system for high reliability means, increasing the cost of the system and its complexity [Ghoshal et al, 1999]. More so, monitoring, control and protection subsystems of the Gas Turbine Engines further add more cost and complexity to the overall system. The application of a classical maintenance approaches has been proven over the years, to be unsuitable for Engineering systems such as Gas turbine engines [Kadirkamanathan, 2008] [Isermann, 2006]. The health state of a GTE is determined by its functional state or characteristics of the parameter variables. Depending on the characteristics of these parameter variables, the GTE health state can be in a particular state.

This study focuses on maintaining high reliability and availability of GTE, by reducing or eliminating erratic failure of the system, extending the phase 2 – the operating period to T_N, thus enhancing the reliability of the turbine engine during wear out phase, through predictive maintenance strategies. To achieve our objective, we continually monitored the system operating condition to detect any changes in the normal operating condition of the Gas turbine engine.

As a consequence of technology advances, a significant increase in the system's complexity and sophistication has made Engineering system maintenance to be formidable challenges to the manufacturers and the end users of these indispensible systems. Indeed, the cost of maintenance over the lifetime of these systems far exceeds the manufacturing cost [Ghoshal, 1999]. Gas turbine engines are a versatile, cost-effective source of electricity, mechanical power and propulsion. They are the power sources of choice in many applications such as mechanical drive of machinery and for electrical power generation. Due to high demand and the critical nature of services provided by systems driven by gas turbine engines, there is a need to ensure continuous operability, high reliability and availability of these systems. This can only be achieved by developing Gas turbine engine inspection and maintenance methods that are measurably compatible with the complexity and sophistication of these systems. The application of a classical maintenance approaches has been proven over the years, to be unsuitable for Engineering systems such as Gas turbine engines due to the certain factors such as, high cost of maintenance, low productivity, reliability of gas turbine engine is high and fault conditions are rare and often not seen before, etc.

The supervision of engineering system in normal operation is usually performed by limit-checking or threshold-checking of some measurable output variables Y(t), such as pressure, temperature, shaft displacement/vibration, speed, liquid level, etc. This means, one checks if the quantities are within a tolerant zone, $Y_{min} < Y(t) < Y_{max}$. An alarm is raised if the tolerance zone is exceeded. In the case of a critical situation, the monitoring function automatically initiates an appropriate counteraction usually the system is commanded to be in a fail-safe state, which is an emergency shutdown. These classical methods of system monitoring and automatic protection are suitable for overall supervision of the entire system. The advantage of the classical limit-value-based supervision method is their simplicity and reliability for steady state situations. But it only reacts after a large sudden fault or a long-lasting gradually increasing fault has occurred [Isermann, 2006].

METHODOLOGY

The data used for ISOMAP analysis are pre-processed. That is the EngData set are split into Training data set (200 by 98), and Test data set (100 by 98). The ISOMAP MATLAB function/code are downloaded from the internet site – http://isomap.stanford.edu/. The geodesic distances computed based on Floyd's algorithm, which produces the best performance in Matlab.

Selecting the optimal parameter value for ISOMAP algorithm is very critical, for example, the value of the reduced dimension and the 'K' value. If K is too large, the local neighbourhood will include the data points from other branches of the manifold, which may lead to errors in the final embedding. Also if too small K is used, it may lead to discontinuity, thereby casing manifold into a large number of cluster that is disconnected.

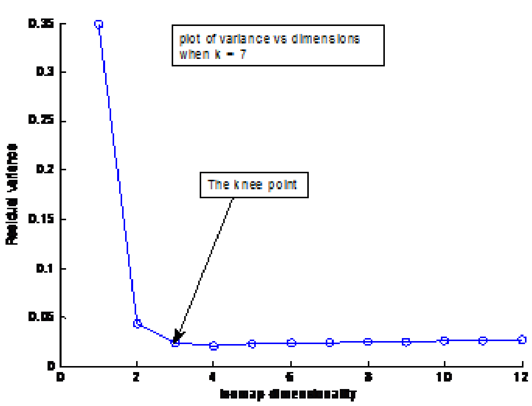
In this paper, we employ the following parameters: Low dimensionality = 12, K values for neighbourhood graph = 7 and 8 (any value of K below 7 is too small for this study, as it leads to discontinuity) for the analysis of GTE Training Data Using ISOMAP Technique.

Data Visualisation in 2-dimensional space with ISOMAP is performed with the following data classification. The Labels or the Training target of the EngData set is classified into three classes, namely; Class 1: good GTE with blue cross sign, Class 2: average GTE with green circle, and Class 3: bad GTE with red plus sign. 1-NN classification for GTE fault diagnosis using ISOMAP is performed on the EngData with the total number of test labels = 100, total number of classified cases = 65, and total number of misclassification = 35.

M-Fold Cross-Validation of the Training Data set (ISOMAP) is carried out with 10-fold cross validation of the Training data set. This is used to evaluate the accuracy of the nearest-neighbour classification of the GTE data analyzed with ISOMAP and the results are shown in Table 5.

MODEL EXPERIMENT

From the selected data, the graph of residual variance is plotted against Isomap dimensionality with the value of K = 7 and 8 as shown in Figures 1 and 2. 2- Dimensional Isomap is also embedded with neighborhood graph with the value of K being 7 and 8 as presented in Figures 3 and 4 respectively.



Fig, 1: Residual Variance vs Isomap dimensionality with K = 8

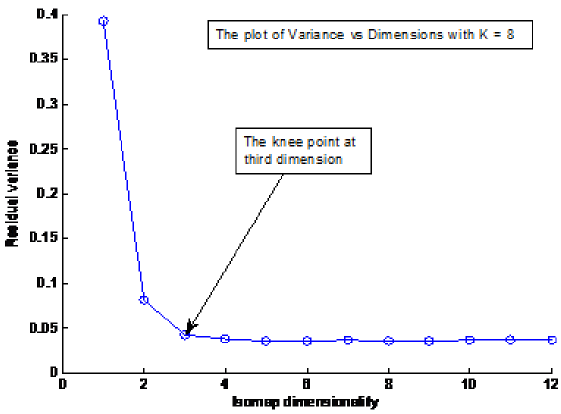


Fig. 2: Residual Variance vs Isomap dimensionality with K = 7

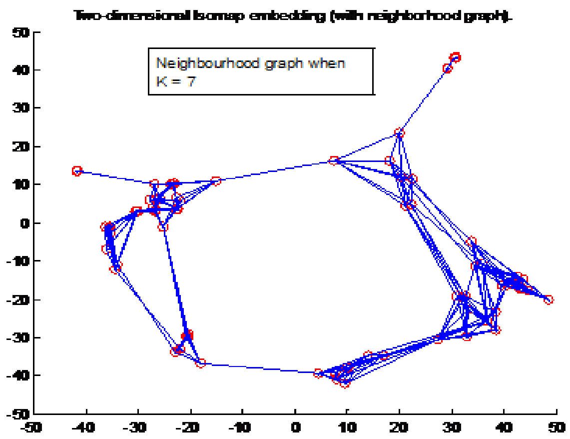


Figure 3: Neighbourhood graph when k = 7

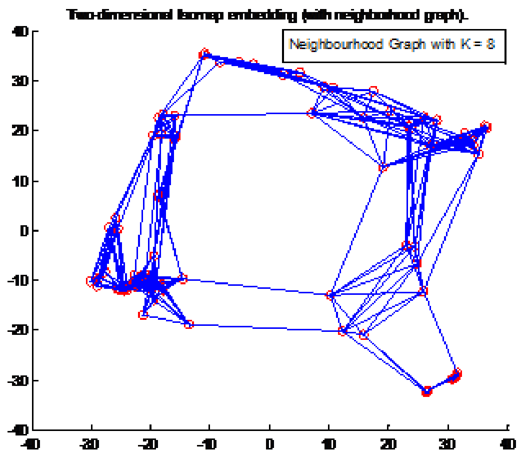


Figure 4: Neighbourhood graph when k = 8

From Figures 1 and 2, the two variance graphs show that ISOMAP can identify the intrinsic dimensionality as indicated at the knee points on the two graphs.

The two-dimensional plots give data visualisation in 2-dimensional space with the value k = 7 and 8 respectively. These are shown in Figures 5 and 6.

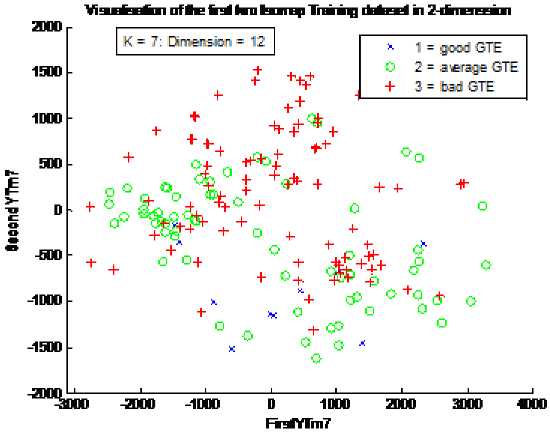


Figure 5

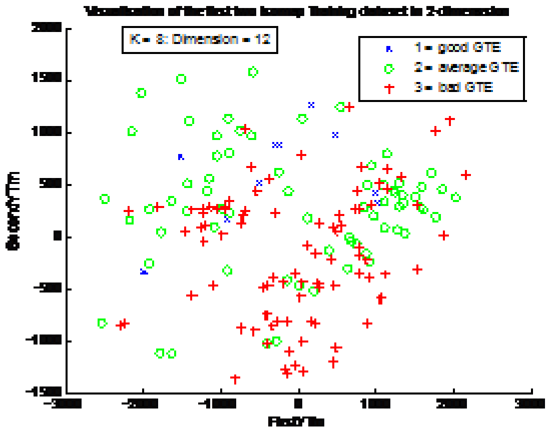


Figure 6

Figure 5: Visualisation of the result from ISOMAP analysis of GTE data set in two-dimension with K = 7

Figure 6: Visualisation of the result from ISOMAP analysis of GTE data set in two-dimension with K = 8

Table 1 shows the output of 1-NN classification for GTE fault diagnosis using ISOMAP. Due to nonlinear nature of the data under analysis, and the fact that nonlinear method of dimensionality reduction technique has been used to analyze the GTE data, 1-NN classification of the analyzed data does not give an impressive classification. The result indicates that the error rate or percentage misclassification of these data is very high as can be seen in the Table 1.

The result indicates the following performance; the total number of test labels = 100, Total number of classified cases = 65, Total number of misclassification = 35, Thus performance of the classification is 65%, And error rate of percentage misclassification is 35%

Table 1 Nearest-neighbours classification of Training data using ISOMAP

S/ N	Test Labe ls	NN Predi ct	Classified/ Misclassifi ed (CL / MCL)	S/ N	Test Labe ls	NN Predi ct	Classified/ Misclassifi ed	S/ N	Test Labe ls	NN Predi ct	Classified/ Misclassifi ed
1	2	2	CL	35	2	3	MCL	69	3	3	CL
2	3	3	CL	36	2	2	CL	70	2	2	CL
3	3	3	CL	37	1	2	MCL	71	2	2	CL
4	3	3	CL	38	1	2	MCL	72	3	3	CL
5	3	3	CL	39	1	2	MCL	73	3	3	CL
6	3	3	CL	40	2	3	MCL	74	1	2	MCL
7	3	3	CL	41	3	3	CL	75	1	2	MCL
8	2	3	MCL	42	3	3	CL	76	1	2	MCL
9	3	3	CL	43	2	3	MCL	77	1	2	MCL
10	3	3	CL	44	3	3	CL	78	1	3	MCL
11	3	3	CL	45	3	3	CL	79	2	2	CL
12	3	3	CL	46	2	2	CL	80	2	3	MCL
13	3	2	MCL	47	2	2	CL	81	2	3	MCL
14	3	3	CL	48	2	2	CL	82	2	3	MCL
15	3	2	MCL	49	2	2	CL	83	2	3	MCL
16	3	3	CL	50	1	2	MCL	84	2	2	CL
17	3	3	CL	51	2	3	MCL	85	2	3	MCL
18	3	3	CL	52	2	2	CL	86	2	2	CL
19	3	2	MCL	53	1	2	MCL	87	3	3	CL
20	3	2	MCL	54	2	2	CL	88	3	3	CL
21	2	2	CL	55	2	2	CL	89	3	3	CL
22	2	2	CL	56	2	2	CL	90	1	2	MCL
23	2	2	CL	57	2	2	CL	91	2	3	MCL
24	2	2	CL	58	2	2	CL	92	1	2	MCL
25	2	2	CL	59	3	3	CL	93	2	2	CL
26	2	2	CL	60	3	2	MCL	94	1	2	MCL
27	2	2	CL	61	2	2	CL	95	2	2	CL
28	2	2	CL	62	2	2	CL	96	2	3	MCL
29	2	3	MCL	63	1	2	MCL	97	1	2	MCL
30	3	3	CL	64	2	2	CL	98	2	2	CL
31	3	3	CL	65	2	3	MCL	99	2	3	MCL
32	3	3	CL	66	3	3	CL	10	2	2	CL
33	3	3	CL	67	3	3	CL	Percentage Classified = 65%			
34	2	2	CL	68	3	3	CL	Percentage Misclassified = 35%			

It shows that the output of nearest-neighbour classification when ISOMAP dimensionality technique was applied to the training data set of the GTE. The percentage of classified cases is 65% and the percentage of misclassified cases is 35%. Good turbine engine is donated by 1, average turbine engine is donated by 2 and bad turbine engine is donated by 3.

Table 2 NN classification result of Test data set with ISOMAP when K = 7

	PREDICTED CLASSIFICATION			
		Good GTE (class 1)	Average GTE (class 2)	Bad GTE (class 3)
KNOWN CLASSIFICATION	Good GTE (class 1)	0 (0%)	14	1
	Average GTE (class 2)	0	33 (67.75%)	15
	Bad GTE (class 3)	0	5	32 (86%)

Total number of test cases = 100, Total number of Good GTE = 15; percentage of good GTE classification = 0%, Total number of Average GTE = 48; percentage of average GTE, classification = 68.75%, Total number of Bad GTE = 37; percentage of bad GTE classification = 86%.

Table 2 shows that no good GTE out of 15 were classified as good GTE, 14 good GTE out of 15 were classified as average GTE and one good GTE was classified as bad GTE. Though no bad GTE was classified as good GTE, 14 good GTE being classified as average GTE and one good GTE have made The ISOMAP analysis very unreasonable for safety critical system such as GTE, and it will impact on the availability as well as productivity of the system. This poor result can be attributed to the nature of the data set.

The distribution of GTE data set among Good, Average and Bad (the three classes) in the EngData Training set are shown in Table 3. This makes data classification very difficult especially when ISOMAP technique is applied. Table 4 shows the NN classification result of test data set with ISOMAP when K = 8.

Table 3: Distribution of EngData set among the three classes

EngData Set	Good GTE	Average GTE	Bad GTE
Training Data set	10	75	115
Test Data set	15	48	37

Table 4 NN classification result of Test data set with ISOMAP when K = 8

	PREDICTED CLASSIFICATION			
		Good GTE (class 1)	Average GTE (class 2)	Bad GTE (class 3)
KNOWN CLASSIFICATION	Good GTE (class 1)	1	14	0
	Average GTE (class 2)	3	34	11
	Bad GTE (class 3)	0	7	30

Total number of test cases = 100, Total number of Good GTE = 15; percentage of good GTE classification = 6.7%, Total number of Average GTE = 48; percentage of average GTE classification = 70.8%, Total number of Bad GTE = 37; percentage of bad GTE classification = 81%. With K = 8, (though, even number is not a good choice for K), the classification gives a slightly good result as no good GTE was classified as bad GTE and no bad GTE was classified as good GTE. It is still not generally good approach because 14 out of 15 good GTE were classified as average GTE.

The cross validation of the training data set (Table 5) gives an impressive result with isomap. The total misclassification recorded is 26.5% instead of 35% error rate recorded when nearest neighbour classification was used.

Table 5: Result of 10-fold Cross-validation of GTE Training Data set - ISOMAP.

Partition #1			**Partition #2**			**Partition #3**			**Partition #4**			**Partition #5**			**Partition #6**			**Partition #7**			**Partition #8**			**Partition #9**			**Partition #10**		
TL	Pd	C/M	TL	Pd	C/M	TL	Pd	C/M	TL	Pd	C/M	TL	Pd	C/M	TL	Pd	C/M	TL	Pd	C/M	TL	Pd	C/M	TL	Pd	C/M	TL	Pd	C/M
2	1	M	2	2	C	3	3	C	3	3	C	3	3	C	3	2	M	3	2	M	2	3	M	2	2	C	2	1	M
3	2	M	3	2	M	3	3	C	2	3	M	3	3	C	2	2	C	1	1	C	3	3	C	2	3	M	2	2	C
3	3	C	3	2	M	2	2	C	3	3	C	2	2	C	2	2	C	3	3	C	3	3	C	3	2	M	2	2	C
3	3	C	3	3	C	2	1	M	3	3	C	3	3	C	2	1	M	3	3	C	3	3	C	3	3	C	2	2	C
2	3	M	3	3	C	3	3	C	3	3	C	2	3	M	3	3	C	3	3	C	1	1	C	3	3	C	3	3	C
1	1	C	3	3	C	3	3	C	2	1	M	2	2	C	3	2	M	3	3	C	3	3	C	3	3	C	3	3	C
2	1	M	3	3	C	3	3	C	3	3	C	3	3	C	3	3	C	3	3	C	3	3	C	3	2	M	3	2	M
3	3	C	3	3	C	2	2	C	2	1	M	2	2	C	3	3	C	3	2	M	2	2	C	3	3	C	3	3	C
2	2	C	2	1	M	2	2	C	3	3	C	3	3	C	2	1	M	3	2	M	3	3	C	2	2	C	2	1	M
2	2	C	2	2	C	2	3	M	3	3	C	3	3	C	3	3	C	3	3	C	2	2	C	2	2	C	3	3	C
3	3	C	3	3	C	3	3	C	3	2	M	3	3	C	3	3	C	3	3	C	2	1	M	3	2	M	3	3	C
3	2	M	1	2	M	3	2	M	3	2	M	2	2	C	2	2	C	2	2	C	2	2	C	2	3	M	3	3	C
2	2	C	2	2	C	3	3	C	2	1	M	2	2	C	2	2	C	3	3	C	3	3	C	3	2	M	3	2	M
3	3	C	2	1	M	3	3	C	3	3	C	2	2	C	2	3	M	2	2	C	3	3	C	2	2	C	3	3	C
3	3	C	3	3	C	3	2	M	3	3	C	2	2	C	3	3	C	3	3	C	2	2	C	2	3	M	3	3	C
2	2	C	2	2	C	3	3	C	3	2	M	3	3	C	3	3	C	2	3	M	3	3	C	3	2	M	2	2	C
3	2	M	1	1	C	2	1	M	3	3	C	2	2	C	2	2	C	2	2	C	2	2	C	3	3	C	3	3	C
1	1	C	3	3	C	2	1	M	3	3	C	2	2	C	2	2	C	3	3	C	3	3	C	3	3	C	3	3	C
1	1	C	2	2	C	3	3	C	2	2	C	2	2	C	1	1	C	2	3	M	3	3	C	2	2	C	2	2	C
3	3	C	3	2	M	3	2	M	3	2	M	1	1	C	3	3	C	2	2	C	2	2	C	1	2	M	3	3	C
C = 14; M = 6			C = 14; M = 6			C = 13; M = 7			C = 12; M =8			C = 19; M = 1			C = 15; M = 5			C = 17; M = 3			C = 18; M = 2			C = 11; M = 9			C = 16; M = 4		
Perform = .70			Perform = .70			Perform = .65			Perform = .60			Perform = .95			Perform = .75			Perform = .75			Perform = .90			Perform = .55			Perform = .80		
E_{R1} = 0.30			E_{R2} = 0.30			E_{R3} = 0.35			E_{R4} = 0.40			E_{R5} = 0.05			E_{R6} = 0.25			E_{R7} = 0.25			E_{R8} =.10			E_{R9} = 0.45			E_{R10} = 0.20		

TL = TEST LABELS, Pd = PREDICTION, C = CLASSIFIED
M = MISCLASSIFIED PERFORM = CROSS-VALIDATION PERFORMANCE
E_R = ERROR RATE

$$\text{Average } E_R = \frac{1}{M}\sum_{i=1}^{M} E_{Ri}$$

$$\text{Average } E_R = \frac{1}{10}\,(0.30 + 0.30 + 0.35 + 0.40 + 0.05 + 0.25 + 0.25 + 0.10 + 0.45 + 0.20)$$

Average E_R = 0.265 = 26.5%

RESULT AND DISCUSSION

The effectiveness or performance of ISOMAP depends on the nature of the data set. ISOMAP give a better result for manifolds of moderate dimensionality, since the estimates of manifold distance for a given graph size degrades as the dimensionality increases. The data set whose

classes or features are sparsely distributed without defined uniformity, such as engineering data obtained from practical systems, may not give a better result when analysed using ISOMAP.
The performance of ISOMAP can be evaluated using nearest neighbour classification of the test data set and cross validation of the training data set. In this paper, the performance of ISOMAP is seriously affected by the choice of neighbourhood factor, k for the algorithm. This may be due to the nature of the data set. The neighbourhood factor above 8 gives a comparatively bad result while a value of k below 7 leads to discontinuity and the Y.index (which contains the indices of the points embedded), produced is less than 98 indices. When k = 6 or 5 was used, the Y.index was 35 and k = 3 gave much lower indices. This made the ISOMAP analysis limited to only neighbourhood factor values. That is 7 or 8.

Nearest-neighbour classification of the ISOMAP Test data set with k =7 did not give an impressive result. The percentage of the misclassification was 35% and that of classified cases was 65%, as can be seen on the in Table 2. Above all, Table 4 shows that no good GTE out of 15 were classified as good GTE, 14 good GTE out of 15 were classified as average GTE and one good GTE was classified as bad GTE. Though no bad GTE was classified as good GTE, 14 good GTE being classified as average GTE and one good GTE have made The ISOMAP analysis very unreasonable for safety critical system such as GTE, and it will impact on the availability as well as productivity of the system. This poor result can be attributed to the nature of the data set. The distribution of GTE data set among Good, Average and Bad (the three classes) in the EngData Training set are shown in Table 5. This makes data classification very difficult especially when ISOMAP technique is applied.

ISOMAP generated a two-dimensional embedding with a neighbourhood graph which give a visual information or characteristic of the data set. This is helpful in studying the geometrical structure of the GTE data. Also, the ISOMAP analysis preserves information contained in the data and the local structure of the data.

With k = 8, ISOMAP was achieve 6.7% of good GTE was classified as good GTE, 70.8% of the average GTE was classified as average GTE and 81% of bad GTE was classified as bad GTE. No bad GTE was classified as good GTE and no good GTE was classified as bad GTE. This achievement is reasonably good as no it is important in safety critical systems such as GTE. But the system availability and productivity is affected as over 93% of good GTE was classified as average GTE.

The cross validation of the training model of the data base using ISOMAP also recorded an impressive result; that is 73.5% of the training data model was classified while only 26.5% of the training data model was misclassified.

REFERENCES

Ghoshal, S., Roshan Shrestha, Anindya Ghoshal, Venkatsh Malepati, Somnath Deb, Krishna pattipati and David Kleinman, (1999)“An Integrated Process For System Maintenance, Fault Diagnosis and Support”, Invited Paper in Proc. IEEE Aerospace Conf., Aspen, Colorado.

Isermann, R. (2006). “Fault-Diagnosis Systems – An Introduction from Fault Detection to Fault Tolerance”, " Springer, Berlin.

Kadirkamanathan, V., (2008) “ACS 6304 – Health Care Monitoring”, Department of Automatic Control & Systems Engineering, University of Sheffield, 21 – 25 January.

Tenenbaum, J. B. “Mapping a Manifold of Perceptual Observations”, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139.