Classification of Aircraft Maneuvers for Fault Detection

Nikunj C. Oza, Kagan Tumer, Irem Y. Tumer, and Edward M. Huff

Computational Sciences Division
NASA Ames Research Center
Mail Stop 269-3
Moffett Field, CA 94035-1000
{oza,itumer,kagan,huff}@email.arc.nasa.gov

Abstract. Ensemble classifiers tend to outperform their component base classifiers when the training data are subject to variability. This intuitively makes ensemble classifiers useful for application to the problem of aircraft fault detection. Automated fault detection is an increasingly important problem in aircraft maintenance and operation. Standard methods of fault detection assume the availability of data produced during all possible faulty operation modes or a clearly-defined means to determine whether the data represent proper operation. In the domain of fault detection in aircraft, the first assumption is unreasonable and the second is difficult to determine. Instead we propose a method where the mismatch between the actual flight maneuver being performed and the maneuver predicted by a classifier is a strong indicator that a fault is present. To develop this method, we use flight data collected under a controlled test environment, subject to many sources of variability. In this paper, we experimentally demonstrate the suitability of ensembles to this problem.

1 Introduction

Ensembles have been shown to improve the generalization performance of many types of classifiers in many real world pattern recognition problems (e.g., [3]). The improvement tends to increase as the variability among the classifiers in the ensemble increases [10]. This property of ensembles intuitively makes them useful in understanding aircraft data, which is subject to considerable variability. In this paper, we discuss the results of applying ensemble methods to aircraft data for fault detection.

A critical aspect of operating and maintaining aircraft is detecting problems in their operation in flight. This allows maintenance and flight crews to fix problems before they become severe and lead to significant aircraft damage or a crash. Fault detection systems are becoming a standard requirement in most aircraft [1, 8]. However, most systems produce too many false alarms, mainly due to an inability to compare real behavior with modeled behavior, making their reliability questionable in practice [7]. Other systems assume the availability of data produced during all possible faulty operation modes [1, 4, 8]. Because of the

Table 1. Conceptual open loop model illustrating assumed causal relationships.

Flight $\longrightarrow$ Maneuver (M)	Aircraft $\longrightarrow$ Attitude (A)	Physical $\longrightarrow$ Input (I)	Internal $\longrightarrow$ Response (R)	Measured Output (O)
•Fwd. Flight	•Radar Alt.	•Engine Torque	•*[Tooth*	•Vibration
•Side Flight	•Airspeed	•Engine Speed	*Bending]*	- x axis
•Fwd. Climb	•Climb Rate	•*[Mast Lifting]*	•*[Backlash]*	- y axis
•Fwd. Descent	•Heading	•*[Mast Bending]*	•*[Friction]*	- z axis
•Hover	•Bank	•*[other]*	•*[Heat]*	•*[Temp]*
•Hover Turn	•Pitch		•*[other]*	•*[Noise]*
•Coord. Turn	•Side Slip		•*[DAMAGE]*	•*[other]*
•*[other]*	•*[other]*			

highly safety-critical nature of the aircraft domain application, most fault detection systems must function well even though fault data are non-existent and the set of possible faults is unknown. Models are often used to predict the effect of damage and failures on otherwise healthy (baseline) data. However, while models are a necessary first start, the modeled system response often does not take operational variability into account, resulting in high false-alarm rates [5, 7].

In this paper, we use in-flight aircraft data that were collected as part of a research effort to understand the sources of variability present in the actual flight environment, with the purpose of reducing the high rates of false alarms [5, 9]. That work described aircraft operation conceptually according to the open-loop causal model shown in Table 1. We assume that the maneuver being performed (M) influences the observable aircraft attitudes (A), which in turn influence the set of possibly observable physical inputs (I) to the transmission. The physical inputs influence the transmission in a variety of ways that are not typically observable (R); however, they influence outputs that can be observed (O).

Our approach to fault detection in aircraft depends fundamentally on the assumption that the nature of the relationships between the elements M, A, I, R, and O described above change when a fault materializes. As mentioned earlier, the many approaches that try to model only the set of possible outputs (O) and indicate the presence of a fault when the actual outputs do not match the model do not account for operational variability. Also, the output space is often too complicated to allow faithful modeling and measuring differences between the model and actual outputs. This latter difficulty remains even if one attempts to model the output as a function of the flight maneuver or other influence due to noise, wind, and other conditions. Approaches to fault diagnosis (e.g., [12]) attempt to predict either normal operation or one of a designated set of faults. As stated earlier, this is not possible in the aircraft domain because the set of possible faults is unknown and fault data are non-existent. In this work, we create a system diagrammed in Figure 1. We create classifiers that predict the flight maneuver (M) as a function of other available data such as the outputs (O). The data that we use contain the actual maneuver, but in general, this may be calculated using pilot input and/or attitude data. We propose that mismatches

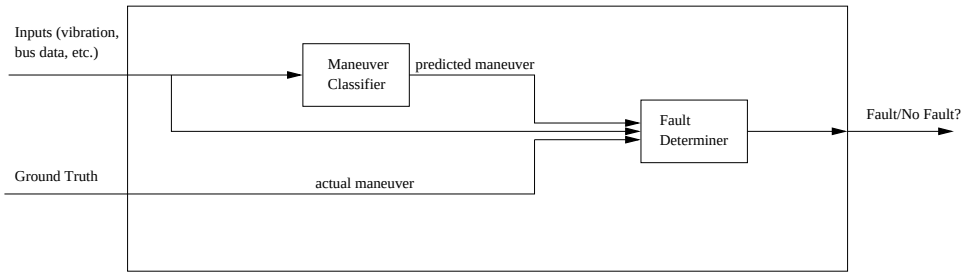

Fig. 1. Online Fault Detection System Block Diagram.

between the predicted maneuver and the actual maneuver being performed is a strong indicator for the presence of a fault.

In order for our method to have a low false-alarm rate, we need a maneuver classifier with the highest performance possible. In addition to using Multilayer Perceptrons (MLPs) and Radial Basis Function (RBF) networks, we use ensembles [2, 10] of MLPs and RBF networks. We have also identified sets of maneuvers (e.g., three different hover maneuvers) that are similar enough to one another that misclassifications within these groups are unlikely to imply the presence of faults. Additionally, we smooth over the predictions for small windows of time in order to mitigate the effects of noise.

In the following, section 2 discusses the aircraft under study and the data generated from them. We discuss the ensemble methods that we used and the associated data preparation that we performed in section 3. We discuss the experimental results in section 4. We summarize the results of this paper and discuss ongoing and future work in section 5.

2 Aircraft Data

The data used in this work were collected from two helicopters: an AH1 Cobra and OH58c Kiowa [5]. The data were collected by having two pilots each fly two designated sequences of steady-state maneuvers according to a predetermined test matrix [5]. It uses a modified Latin-square design to counterbalance changes in wind conditions, ambient temperature, and fuel depletion. Each of the four flights consisted of an initial period on the ground with the helicopter blades at flat pitch, followed by a low hover, a sequence of maneuvers drawn from the 12 primary maneuvers (e.g., high-speed forward flight), a low hover, and finally a return to ground. Each maneuver was scheduled to last 34 seconds in order to allow a sufficient number of cycles of the main rotor and planetary gear assembly to apply the signal decomposition techniques used in the previous studies [5].

Summary matrices were created from the raw data by averaging the data produced during each revolution of the planetary gear. The summarized data consists of 31475 revolutions of data for the AH1 and 34144 revolutions of data for the OH58c. Each row, representing one revolution, indicates the maneuver being performed during that revolution and the following 30 quantities: Rev-

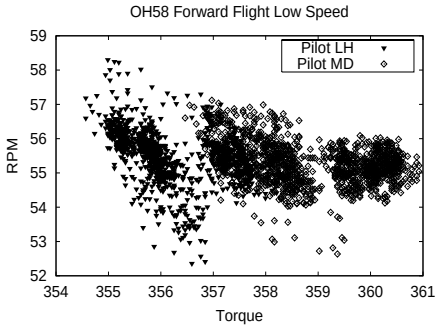

Fig. 2. OH58 Maneuver A (Forward Flight Low Speed) Pilot-Separated

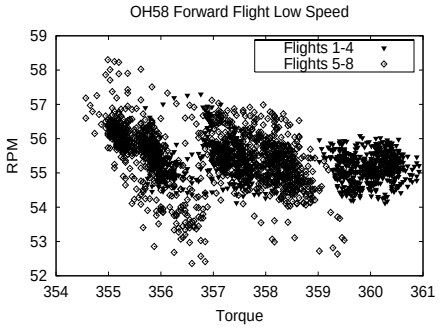

Fig. 3. OH58 Maneuver A (Forward Flight Low Speed) Flight-Separated

olutions per minute of the planetary gear, torque (mean, standard deviation, skew, and kurtosis), and vibration data from six accelerometers (root-mean-square, skew, kurtosis, and a binary variable indicating whether signal clipping occurred). For the AH1, also available were the mean and standard deviation values for the following attitude data from a 1553 bus: altitude, speed, rate of climb, heading, bank angle, pitch, and slip.

3 Methodology

Sample torque and RPM data from one maneuver separated by pilot and by flights are shown in Figures 2 and 3, respectively. The highly-variable nature of the data and differences due to different pilots and different times of day when the aircraft were flown, are clearly visible and make this a challenging classification problem. We chose multilayer perceptrons (MLPs) with one hidden layer and radial basis function (RBF) networks as base classifiers. We also constructed ensembles of each type of classifier and ensembles consisting of half MLPs and half RBF networks, because ensembles have been shown to improve upon the performances of their base classifiers, particularly when the correlations among them can be kept low [10, 11]. In particular, we use averaging ensembles (the output of the ensemble is the average of the outputs of the base classifiers) because of its combination of simplicity and high performance relative to many more sophisticated methods [2].

We created data sets for each of the two aircraft by combining its 176 summary matrices. This resulted in 31475 patterns (revolutions) for the AH1 and 34144 for the OH58. Both types of classifiers were trained using a randomly-selected two-thirds of the data (21000 examples for the AH1, 23000 for the OH58) and were tested on the remainder for the first set of experiments. Each aircraft's complete set of available inputs (described in section 2) was used.

We calculated the *confusion matrix* of every classifier we created. Entry (i, j) of the confusion matrix of a classifier states the number of times that an exam-

Table 2. Sample confusion matrix for OH58 (MLP).

True Class	Predicted Class													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	693	0	7	6	79	0	0	0	0	0	0	0	0	0
2	0	679	0	0	0	0	0	0	0	0	0	0	47	0
3	55	1	568	64	31	6	0	11	9	1	11	7	0	3
4	26	0	43	691	15	0	0	3	0	0	0	2	0	1
5	196	0	68	41	412	0	0	0	0	0	2	16	0	0
6	0	0	0	0	0	719	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	1079	0	0	0	0	0	0	0
8	0	9	22	16	0	0	0	**748**	**177**	**97**	11	6	3	0
9	0	1	1	6	0	0	0	**172**	**381**	**162**	4	7	6	0
10	0	4	1	6	0	0	0	**186**	**170**	**376**	0	8	13	0
11	4	0	15	4	3	0	0	2	1	0	**494**	**217**	0	0
12	3	0	7	6	4	0	0	2	1	0	**200**	**531**	0	0
13	0	63	0	0	0	0	0	4	1	0	0	0	712	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	685

ple of class i is classified as class j. The confusion matrices (see Table 2 for an example of a confusion matrix—entry $(1, 1)$ is in the upper left corner), indicate that particular maneuvers were continually confused with one another. In particular, the three hover maneuvers (8-Hover, 9-Hover Turn Left, and 10-Hover Turn Right) were frequently confused with one another and the two coordinated turns (11-Coordinated Turn Left and 12-Coordinated Turn Right) were also frequently confused (the counts associated with these errors are shown in bold in Table 2.) The maneuvers within these groups are similar enough that misclassifications within these groups are unlikely to imply the presence of faults. Therefore, for the second set of experiments, we recalculated the classification accuracies allowing for these misclassifications. In section 4, we refer to the results of these experiments as "Post-Consolidated" because the class consolidation was performed after the learning.

For our third set of experiments, we consolidated these two sets of maneuvers in the data before running the experiments. That is, we combined the hover maneuvers into one class and the coordinated turns into one class, yielding a total of 11 possible predictions instead of the original 14. In section 4 we refer to these as "Pre-Consolidated" results since the consolidation was done before the learning. We expected the performance to be best for this third set of experiments because, informally, the classifiers do not have to waste resources distinguishing among the two sets of similar maneuvers.

Finally, we used the knowledge that a helicopter needs some time to change maneuvers. That is, two sequentially close patterns are unlikely to come from different maneuvers. To obtain results that use this "prior" knowledge, we tested the classifiers from the previous experiments on sequences of revolutions by averaging the classifiers' outputs on a window of examples surrounding the current one. In one set of experiments, we averaged over windows of size 17 (8 revolu-

Table 3. OH58c Single Revolution Test Set Results.

Base Type	N	Single Rev	Corr	Post-Run Consolidated	Corr	Pre-Run Consolidated	Corr
MLP	1	79.789 ± 0.072	–	92.709 ± 0.055	–	93.566 ± 0.060	–
	4	81.997 ± 0.065	0.4193	93.820 ± 0.044	0.4118	94.422 ± 0.038	0.4443
	10	82.441 ± 0.045	0.4193	94.015 ± 0.028	0.4133	94.672 ± 0.032	0.4395
	100	82.771 ± 0.016	0.4199	94.133 ± 0.011	0.4139	94.672 ± 0.032	0.4374
RBF	1	75.451 ± 0.103	–	89.305 ± 0.080	–	90.460 ± 0.169	–
	4	75.817 ± 0.048	0.7164	89.485 ± 0.047	0.7046	90.912 ± 0.056	0.5877
	10	75.871 ± 0.040	0.7185	89.498 ± 0.034	0.7058	90.987 ± 0.032	0.6009
	50	75.908 ± 0.016	0.7162	89.506 ± 0.011	0.7058	91.018 ± 0.014	0.6028
MLP/ RBF	2	80.190 ± 0.079	0.3687	92.834 ± 0.065	0.3176	93.777 ± 0.046	0.2905
	4	80.946 ± 0.059	0.4352	93.189 ± 0.042	0.3997	94.097 ± 0.048	0.3788
	10	81.406 ± 0.043	0.4574	93.403 ± 0.039	0.4273	94.348 ± 0.025	0.3941
	100	81.543 ± 0.020	0.4681	93.463 ± 0.017	0.4392	94.457 ± 0.011	0.4056

tions before the current one, the current one, and 8 revolutions after the current one) which corresponds to about three seconds of real time. Because the initial training and test sets were randomly chosen from the original data sequence, this averaging could not be performed on the test set alone. Instead it was performed on the full data set for both helicopters. In order to isolate the benefits of window averaging, we also compute the errors of the single-revolution classifiers on this full dataset.[1]

4 Experimental Results

In this section we describe the experimental results that we have obtained so far. We first discuss results on the OH58 helicopter. In Table 3, the column marked "Single Rev" shows the accuracies of individual networks and ensembles of various sizes on the summary matrices randomly split into training and test sets. We only present results for some of the ensembles we constructed due to space limitations and because the ensembles exhibited relatively small gains beyond 10 base models. MLPs and ensembles of MLPs outperform RBFs and ensembles of RBFs consistently. The ensembles of MLPs improve upon single MLPs to a greater extent than ensembles of RBF networks do upon single networks, indicating that the MLPs are more diverse than the RBF networks. This is corroborated by the fourth column (marked "Corr")[2] which shows that the av-

[1] We performed this windowed averaging as though the entire data were collected over a single flight. However, it was in fact collected in stages, meaning that there are no transitions between maneuvers. We show these results to demonstrate the applicability of this method to sequential data obtained in actual flight after training the network on "static" single revolution patterns.

[2] Each correlation in this paper is the average of the correlations of every pair of base classifiers in the ensemble. We calculate the correlation of a pair of classifiers as the number of test patterns that the two classifiers agree on but misclassify, divided by

Table 4. OH58c Full Data Set Results.

Base Type	N	Window of 17	Corr	Window 17 Post-Consolidated	Corr	Window 17 Pre-Consolidated	Corr
MLP	1	89.905 ± 0.121	–	96.579 ± 0.066	–	97.586 ± 0.078	–
	4	90.922 ± 0.074	0.5014	96.799 ± 0.026	0.6145	97.635 ± 0.041	0.6258
	10	91.128 ± 0.064	0.5013	96.820 ± 0.018	0.6255	97.729 ± 0.031	0.6067
	100	91.307 ± 0.015	0.5052	97.063 ± 0.140	0.6290	97.695 ± 0.006	0.6086
RBF	1	82.564 ± 0.154	–	92.831 ± 0.103	–	94.611 ± 0.124	–
	4	82.634 ± 0.059	0.7509	92.882 ± 0.047	0.7755	94.548 ± 0.063	0.5870
	10	82.618 ± 0.055	0.7543	92.895 ± 0.043	0.7758	94.517 ± 0.029	0.6001
	50	82.644 ± 0.019	0.7505	92.901 ± 0.013	0.7747	94.524 ± 0.012	0.6072
MLP/ RBF	2	88.674 ± 0.108	0.3652	95.910 ± 0.059	0.3596	97.155 ± 0.045	0.3419
	4	88.895 ± 0.078	0.4520	95.902 ± 0.040	0.4791	97.145 ± 0.067	0.4383
	10	89.140 ± 0.057	0.4788	95.980 ± 0.033	0.5143	97.226 ± 0.032	0.4576
	100	89.320 ± 0.025	0.4937	96.003 ± 0.012	0.5335	97.204 ± 0.009	0.4706
Base Type	N	Single Rev	Corr	Single Rev Post Consolidated	Corr	Single Rev Pre-Consolidated	Corr
MLP	1	82.097 ± 0.072	–	93.539 ± 0.058	–	94.495 ± 0.064	–
	4	84.304 ± 0.049	0.4069	94.622 ± 0.039	0.4019	95.321 ± 0.035	0.4443
	10	84.750 ± 0.043	0.4075	94.805 ± 0.028	0.4029	95.540 ± 0.029	0.4372
	100	85.048 ± 0.012	0.4081	94.922 ± 0.011	0.4036	95.595 ± 0.008	0.4355
RBF	1	76.406 ± 0.099	–	89.680 ± 0.077	–	90.788 ± 0.147	–
	4	76.799 ± 0.040	0.7164	89.872 ± 0.039	0.7142	91.187 ± 0.045	0.6027
	10	76.836 ± 0.033	0.7186	89.902 ± 0.027	0.7162	91.244 ± 0.027	0.6157
	50	76.910 ± 0.011	0.7162	89.948 ± 0.007	0.7143	91.271 ± 0.013	0.6182
MLP/ RBF	2	82.146 ± 0.075	0.3613	93.523 ± 0.061	0.3172	94.587 ± 0.049	0.2883
	4	82.877 ± 0.053	0.4293	93.854 ± 0.041	0.4022	94.876 ± 0.051	0.3783
	10	83.332 ± 0.036	0.4516	94.066 ± 0.029	0.4291	95.089 ± 0.024	0.3948
	100	83.505 ± 0.015	0.4618	94.142 ± 0.015	0.4406	95.163 ± 0.014	0.4076

erage correlations among the base models are much higher for ensembles of RBF networks than ensembles of MLPs. Mixed ensembles perform worse than pure-MLP ensembles and better than pure-RBF ensembles for all numbers of base models. The standard errors of the mean performances decrease with increasing numbers of base models as is normally the case with ensembles. The column marked "Post-Run Consolidated" shows the single revolution results after allowing for confusions among the hover maneuvers and among the coordinated turns, consolidating them into single classes. As expected, the performances improved dramatically. The column "Pre-Run Consolidated" shows the single revolution results on the summary matrices in which the hovers and coordinated turns were consolidated before learning as described in section 3. The performances here were consistently the highest as we had hypothesized. In all these experiments, the improvement due to adding base models to the ensemble increases as the average correlation decreases, as expected.

---

the number of patterns that at least one classifier misclassifies. Note that this is not the posterior-based correlation used in [10, 11].

Table 5. AH1 Results.

Base Type	N	Single Rev Test	Corr	Single Rev Full	Corr	Window of 17	Corr
MLP	1	96.752 ± 0.059	–	96.933 ± 0.060	–	98.344 ± 0.059	–
	4	97.284 ± 0.031	0.4155	97.555 ± 0.025	0.3966	98.757 ± 0.031	0.4052
	10	97.448 ± 0.027	0.4130	97.683 ± 0.013	0.3973	98.779 ± 0.021	0.4105
	100	97.542 ± 0.006	0.4128	97.762 ± 0.008	0.3981	98.861 ± 0.006	0.4055
RBF	1	95.669 ± 0.059	–	95.743 ± 0.067	–	96.662 ± 0.102	
	4	95.946 ± 0.029	0.6462	96.063 ± 0.032	0.6369	96.988 ± 0.042	0.6668
	10	95.911 ± 0.023	0.6561	96.042 ± 0.026	0.6456	96.968 ± 0.028	0.6764
	50	95.946 ± 0.009	0.6538	96.067 ± 0.005	0.6321	97.003 ± 0.008	0.6735
MLP/ RBF	2	97.040 ± 0.054	0.3120	97.231 ± 0.055	0.2933	98.256 ± 0.064	0.2313
	4	97.318 ± 0.025	0.3698	97.502 ± 0.028	0.3539	98.482 ± 0.034	0.3148
	10	97.429 ± 0.018	0.4040	97.570 ± 0.018	0.3899	98.475 ± 0.028	0.3577
	100	97.521 ± 0.011	0.4160	97.659 ± 0.008	0.3978	98.553 ± 0.005	0.3739

The top half of Table 4 shows the results of performing the windowed averaging described in the previous section in the column marked "Window of 17." The columns "Window 17 Post-Consolidated" and "Window 17 Pre-Consolidated" give the results allowing for the confusions mentioned earlier. The bottom half of the table gives the full set errors of the single-revolution classifiers. We can clearly see the benefits of windowed averaging, which serves to smooth out some of the noise in the data.

Table 5 gives all the AH1 results. The column marked "Single Rev" shows the results with the AH1 summary matrices randomly split into training and test sets. The next column has the results of the same single-revolution classifiers on the full data set (training and test combined). The final column gives the results of the windowed averaging method. We do not present the results of the second and third set of experiments (with maneuver consolidation) because they ranged from 99.404% to 100%. The AH1 results are substantially better than the OH58 results. We expected this because the AH1 is a heavier helicopter, so it is less affected by conditions that introduce noise such as high winds. With the AH1 pure-MLP ensembles always outperform mixed ensembles when using windowed averaging. However, in the single-revolution case, the mixed ensembles outperform the pure ensembles for small numbers of base models but perform worse than the MLP ensembles for larger numbers of base models. This is also true with maneuvers consolidated; however, all these performances are very high. Once again, we can see that ensembles of MLPs outperform single MLPs to a greater extent than ensembles of RBFs outperform single RBFs, so the RBFs are not as different from one another. The average correlations among the base models are consistent with this. Because of this, it does not help to add large numbers of RBF networks to an MLP ensemble. The standard errors of the mean performances tend to decrease with increasing numbers of base models just as with the OH58.

Table 6. AH1 Bus and Non-Bus Results

Inputs	Single Rev	Single Rev Consolidated	Window of 17	Window of 17 Consolidated
All	96.752 ± 0.059	99.843 ± 0.032	98.344 ± 0.059	99.737 ± 0.028
Bus	90.380 ± 0.110	95.871 ± 0.091	91.209 ± 0.126	96.027 ± 0.086
Non-Bus	87.884 ± 0.228	93.731 ± 0.171	92.913 ± 0.355	96.110 ± 0.236
$P(agree)$	79.523 ± 0.247	90.063 ± 0.202	85.609 ± 0.320	93.393 ± 0.247

On the AH1, the hover maneuvers were frequently confused just as they were on the OH58, but the coordinated turns were not confused. Taking this confusion into account boosted performance significantly. The windowed averaging approach did not always yield improvement when allowing for the maneuver confusions, but helped when classifying across the full set of maneuvers. However, in all cases when windowed averaging did not help, the classifier performance was at least 99.6%, so there was very little room for improvement.

5 Discussion

In this paper, we presented an approach to fault detection that contains a subsystem to classify an operating aircraft into one of several states, with the idea that mismatches between the predicted and actual state is a strong indicator that a fault is present. The classifier predicts the maneuver given vibration data and other available data and compares that prediction with the known maneuver. Through experiments with two helicopters, we demonstrated that the classifier predicts the maneuver with good reliability, especially when using ensemble classifiers. These results show great promise in predicting the true maneuver with high certainty, enabling effective fault detection. Future work will involve applying this approach to "free-flight data", where the maneuvers are not static or steady-state, and transitions between maneuvers are recorded.

We are currently constructing classifiers using different subsets of the available data as inputs. For example, for the AH1, we have constructed some classifiers that use only the bus data as input and others that use only the vibration data. We hypothesize that disagreement among these classifiers that use different sources of information may indicate the presence of a fault. For example, if the vibration data-based classifier predicts that the aircraft is flying forward at high speed but the bus data-based classifier predicts that the aircraft is on the ground, then the probability of a fault is high. Table 6 shows the results of training 20 single MLPs on these data using the same network topology as for the other MLPs trained on all the AH1 data. They performed much worse than the single MLPs trained with all the inputs presented at once. The last line in the table indicates the percentage of maneuvers for which the two types of classifiers agreed. We would like these agreement probabilities to be much higher because none of our data contains faults. However, simpler uses of the bus data may lead to better performance. For example, if a vibration data-based classifier

predicts forward flight, but the bus data indicate that the altitude is zero, then the probability of a fault is high. We did not need a classifier that uses all the bus data to draw this conclusion. We merely needed to know that a zero altitude is inconsistent with a forward flight. We plan to study the bus data in detail so that we may construct simple classifiers representing knowledge of the type just mentioned and use them to find inconsistencies such as what we just described.

References

1. Robert Campbell, Amulya Garga, Kathy McClintic, Mitchell Lebold, and Carl Byington. Pattern recognition for fault classification with helicopter vibration signals. In *American Helicopter Society 57th Annual Forum*, 2001.
2. Thomas G. Dietterich. Ensemble methods in machine learning. In J. Kittler and F. Roli, editors, *First International Workshop on Multiple Classifier Systems*, pages 1–15. Springer Verlag, Berlin, 2000.
3. Thomas G. Dietterich. An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. *Machine Learning*, 40:139–158, Aug. 2000.
4. Paul Hayton, Bernhard Schölkopf, Linel Tarassenko, and Paul Anusiz. Support vector novelty detection applied to jet engine vibration spectra. In Todd K. Leen, Thomas G. Dietterich, and Volker Tresp, editors, *Advances in Neural Information Processing Systems-13*, pages 946–952. Morgan Kaufmann, 2001.
5. Edward M. Huff, Irem Y. Tumer, Eric Barszcz, Mark Dzwonczyk, and James McNames. Analysis of maneuvering effects on transmission vibration patterns in an AH-1 cobra helicopter. *Journal of the American Helicopter Society*, 2002.
6. J. Kittler. Combining classifiers: A theoretical framework. *Pattern Analysis and Applications*, 1:18–27, 1998.
7. D.A. McAdams and I.Y. Tumer. Towards failure modeling in complex dynamic systems: impact of design and manufacturing variations. In *ASME Design for Manufacturing Conference*, volume DETC2002/DFM-34161, September 2002.
8. Sunil Menon and Rida Hamza. Machine learning methods for helicopter hums. In *Proceedings of the 56th Meeting of the Society for Machinery Failure Prevention Technology*, pages 49–55, 2002.
9. I.Y. Tumer and E.M. Huff. On the effects of production and maintenance variations on machinery performance. *Journal of Quality in Maintenance Engineering*, 8(3):226–238, 2002.
10. K. Tumer and J. Ghosh. Error correlation and error reduction in ensemble classifiers. *Connection Science, Special Issue on Combining Artificial Neural Networks: Ensemble Approaches*, 8(3 & 4):385–404, 1996.
11. K. Tumer and J. Ghosh. Linear and order statistics combiners for pattern classification. In A. J. C. Sharkey, editor, *Combining Artificial Neural Nets: Ensemble and Modular Multi-Net Systems*, pages 127–162. Springer-Verlag, London, 1999.
12. V. Venkatasubramanian, R. Vaidyanathan, and Y. Yamamoto. Process fault detection and diagnosis using neural networks—i. steady-state processes. *Computers and Chemical Engineering*, 14(7):699–712, 1990.