

Combining Multiple Classifiers to Quantitatively Rank the Impact of Abnormalities in Flight Data

E. Smart^{a,*}, D. Brown^a, J. Denman^b

^a*Institute of Industrial Research, Floor 7, Mercantile House, Hampshire Terrace, University of Portsmouth, Portsmouth, Hampshire, PO1 2EG, United Kingdom.*

^b*Flight Data Services (FDS) Ltd, 189-199 West Street, Fareham, Hampshire, PO16 0EN, United Kingdom*

Abstract

This paper presents a novel two phase method that combines one class support vector machine classifiers using combination rules to quantitatively assess the degree of abnormality at various heights during individual aircraft descents and also over the whole descent. Whilst classifiers have been combined before in the literature with success, it is the first time they have been applied to the problem of analysing the act of descending of commercial jet aircraft. The method is tested on artificial Gaussian data and flight data from an industrial partner, Flight Data Services Ltd, the world's leading flight data analysis provider, with promising results.

Keywords: Aircraft landing guidance, Artificial intelligence, Fault diagnosis, Support Vector Machines

1. Introduction

The introduction and development of one class classifiers and novelty detection methods has increased the potential for more accurate fault detection systems. Such systems are very important in a world that has become increasingly automated. As such, the topics of fault diagnosis [1, 2, 3, 4], detecting mechanical failure [5, 6, 7] and condition monitoring [8, 9] are important within the research community. The faults themselves can potentially have catastrophic consequences such as loss of the machine, loss of revenue and even death. Therefore there is an increasing demand for diagnostic systems that can not only detect faults but also identify when the system is moving from normal operation to abnormal (or unusual) operation.

It is difficult to create an efficient binary classifier with a good generalisation ability because whilst there is usually plenty of normal data, faulty data can be expensive to acquire or very rare, thus creating a class imbalance. To investigate this problem, one class classifiers are specifically designed to model normal data and detect outliers or abnormal data. Expert knowledge of a particular problem or operation can be invaluable as it can be used to identify useful features and also provide details of likely faults.

Identifying a gradual or sudden change from normality to abnormality has great value in the field of condition monitoring. In the field of flight safety, knowing at what point the aircraft started to move towards an unsafe state can help flight safety officers identify the reasons and advise other flight crew in similar situations.

Furthermore, such a system can not only detect changes from normality to abnormality, but also changes within normality itself. An abnormality can be an actual fault or merely unusual

*Corresponding author. Tel: +44 (0) 23 9284 4448, Fax: +44 (0) 23 9284

5192

Email addresses: edward.smart@port.ac.uk (E. Smart),

david.j.brown@port.ac.uk (D. Brown),

john.denman@flightdataservices.co.uk (J. Denman)

URL: <http://www.port.ac.uk/research/iir/> (E. Smart)

or different behaviour from the system. This could result from changes to the aircraft due to maintenance or new instructions for the approach to a specific aircraft.

In this paper we consider how fault detection methods can be applied to aircraft descending and landing. Section 2 looks at methods for novelty detection and how they have been applied to real world problems. Section 3 introduces the one class Support Vector Machine (SVM). Section 4 introduces combination rules as methods for summarising data. Section 5 introduced the proposed two phase method. Section 6 details the experiments on artificial and flight data and their results. Section 7 discusses the method and results and section 8 contains the conclusion.

eps

March 18, 2011

2. Novelty Detection Research

There has been a lot of research in novelty detection and the identification of pre-cursors for abnormalities. A comprehensive review of novelty detection methods can be found in [10, 11]. Broadly speaking there are two main areas; statistical methods and neural network/SVM type approaches. Statistical methods often aim to estimate the probability density function of the normal class and then test if unknown objects belong to this distribution. Such an estimation can be made via parametric methods, such as a Mixture of Gaussians classifier, or non-parametric methods such as a Parzen windows classifier. Disadvantages of these methods are that there can be a lot of parameters to optimise and they do depend on a large and representative training set. Furthermore, they find it hard to incorporate abnormal data in the training set and their presence can distort the model.

Neural network and SVM type approaches do not directly estimate the underlying probability distribution and they do not

make any a priori assumptions about the data. A disadvantage with neural networks is that during training, they may get trapped within local minima whereas the SVM method guarantees a global minimum. An additional advantage of SVMs is that they are not adversely affected by having abnormal data in the training set. Furthermore their presence can improve the model. SVMs are also examples of the Structural Risk Minimisation (SRM) paradigm [12], providing a trade off between model complexity and training set error. For these reasons, the one class SVM [13] will be used in this paper to detect abnormalities.

Cannas et al. [14] looks at detecting and predicting disruptions in JET. Such pulses can cause damage to the system. Each pulse in the training set is clustered using a Self Organising Map (SOM) and then trained with using an SVM. Using the SVM decision function output, data falling outside of two bands is regarded as disruptive. It is important for the novelty detection system to detect precursors to these disruptive pulses and they should lie between 1 second and 100 milliseconds before the event. The system reports an 83% success rate with no missed disruptions and only 10% premature alarms.

Perdisci et al. [15] also uses the one class SVM in order to detect anomalies in computer network traffic. The difficulty of this task can be seen by the fact that the system should be able to detect unseen attacks. Previous systems used payload statistics but it has been shown that such systems can be evaded by a crafted mimicry attack. To tackle this problem, they propose an ensemble of SVM classifiers which makes a classification based on a majority vote, on the premise that with several models of normality, a constructed mimicry attack will be hard to achieve. The SVMs are trained on different descriptions of the payload and the ensemble reports a very high AUC (area under the receiver operator characteristic curve) value. This demonstrates that multiple classifiers can be used to make a decision

on the overall impact of the parts.

Gardner et al. [16] uses a one class SVM with a sliding window to detect seizures from electroencephalography recordings (EEGs). The benefits of the method are that it is not patient specific and it does not require training on seizure data which can be very difficult to obtain. It achieves a sensitivity of over 95%, highlighting the ability of the one class SVM to detect unseen events.

Abnormalities in combustion were looked at by Clifton [17]. Note that in this paper, a positive SVM decision function value denotes abnormality whereas in this paper, it denotes normality. A one class SVM was trained on data from each of the three combustion chambers to identify the moment that the overall combustion became unstable. To this end, a mean, product, maximum and a minimum combination rule was used on the three classifier outputs. It was found that the mean and maximum rules were the most effective in detecting precursors to unstable combustion but it was also stated that all four rules could provide useful information. Clifton showed that this approach was very accurate in identifying the first signs of unstable combustion.

Combination rules were also studied by Tax et al. [18] for the purposes of improving classification by combining the outputs from multiple classifiers. Rather than training one classifier on a large dataset, there can be advantages to training different classifiers on different parts of the dataset and combining their information to classify the data. It is found that the mean rule is best when the posterior probabilities are not well estimated.

The system described in this paper is similar to the method in Clifton's paper, but there are subtle differences. Clifton's method is trying to detect the first time an anomaly occurs whereas when analysing flight data, if possible, all anomalies and their impact should be detected. To this end, the first

phase consists of classifiers considering snapshot data at different heights during the descent. In Clifton's paper, the classifier output was assessed using combination rules. Also, there was no interest in comparing several sets of combustion tests to find the most abnormal combustion. This ability is useful to analyse multiple descents. To achieve this, the second phase of the approach consists of one classifier analysing the outputs of the classifiers in phase one and ranking the occurrences by a single novelty score. The method is fully described in section 5.

The novelty of the proposed method is that it quantifies the degree of normality/abnormality at selected points during the descent. By modelling each height during the descent via a one class SVM and using the difference between the decision function value and the computed threshold, analysts can identify the points where abnormalities occur and how abnormal they are. In addition, the method has the ability to rank multiple occurrences of the descents using another SVM which compares the all the outputs for an individual task. This unique feature enables the analyst to identify those descents that had the greatest overall novelties so that action can be taken to remedy any problems. Furthermore, such scores could be plotted over time to identify persistent deviations from the airline's standard operating procedures.

3. Support Vector Novelty Detection

One Class Support Vector Machine (OCSVM) [13, 19] is a powerful novelty detection method based on the support vector machine [20, 21].

Consider 'normal' training data $x_1, x_2, \dots, x_l \in \mathbb{R}^n$. Let ϕ be the mapping $\phi : \mathbb{R} \rightarrow F$ into some feature dot product space F . Let $k(x, y) = (\phi(x), \phi(y))$ be a positive definite kernel which operates on the mapping ϕ . In this paper, the kernel used is the Gaussian kernel, $k(x, y) = \exp(-\|x - y\|^2 / 2\sigma^2)$, as it suppresses growing distances in larger feature spaces. Here, σ is

the width parameter associated with the Gaussian kernel. The data is mapped into the feature space via the kernel function and is separated from the origin with maximum margin. The decision function is found by minimising the weighted sum of the support vector regulariser and the empirical error term depending on a margin variable ρ and individual error terms ξ_i ,

$$\begin{aligned} \min_{w \in F, \xi \in \mathbb{R}^l, \rho \in \mathbb{R}} \quad & \frac{1}{2} \|w\|^2 + \frac{1}{\nu l} \sum_{i=1}^l \xi_i - \rho, \\ \text{subject to} \quad & (w \cdot \phi(x_i)) \geq \rho - \xi_i, \\ & \xi_i \geq 0, \end{aligned} \quad (1)$$

where w is a weight vector in F and ν is the fraction of the training set to be regarded as outliers. Using Lagrangian multipliers, $\alpha_i, \beta_i \geq 0$, with constraints and setting the derivatives of those multipliers with respect to w equal to zero leads to

$$w = \sum_{i=1}^l \alpha_i \phi(x_i), \quad (2)$$

$$\sum_{i=1}^l \alpha_i = 1, \quad (3)$$

$$\alpha_i + \beta_i = \frac{1}{\nu l}. \quad (4)$$

The dual problem is formulated to give

$$\begin{aligned} \min_{\alpha \in \mathbb{R}^l} \quad & \sum_{i,j=1}^l \alpha_i \alpha_j k(x_i, x_j), \\ \text{subject to} \quad & \sum_{i=1}^l \alpha_i = 1, \\ & 0 \leq \alpha_i \leq \frac{1}{\nu l}. \end{aligned} \quad (5)$$

Solutions for the dual problem yield parameters w_0, ρ_0 where

$$w_0 = \sum_{i=1}^{N_s} \alpha_i \phi(s_i), \quad (6)$$

$$\rho_0 = \frac{1}{N_s} \sum_{j=1}^{N_s} \sum_{i=1}^{N_s} \alpha_i k(s_i, x). \quad (7)$$

Here, N_s is the number of support vectors and s_i denotes a support vector. The decision function is given by

$$f(x) = \text{sgn}(w \cdot \phi(x) - \rho_0) \quad (8)$$

$$= \text{sgn} \left(\sum_{i=1}^{N_s} \alpha_i k(s_i, x) - \rho_0 \right). \quad (9)$$

The 'abnormality' detection function is then given by

$$g(x) = \rho_0 - \sum_{i=1}^{N_s} \alpha_i k(s_i, x). \quad (10)$$

The user has to choose the appropriate kernel, with its associated parameters for the problem. However, rather than choosing an error penalty C as via the classical SVM method, one chooses a value for ν which is the fraction of the training set to be classified as outliers. The software used for this classifier is LIBSVM for Matlab version 2.91 [22], a well established SVM program.

The one class SVM is ideally suited to this type of problem (see the end of section 2). The abnormality detection function takes values depending on the distance between the data point and the boundary threshold. It is positive if the data point is inside the decision boundary (normal) and negative if it lies outside (abnormal). The function is bounded above by the minimum distance from the centroid of the hypersphere to the boundary but is not bounded below as a data point could be an infinite distance from the boundary. In practise typical positive values are between 0 and 5 and typical negative values are between 0 and -50, depending on the level of abnormality.

4. Combination Rules

In this paper, a Combination Rule (CR) is regarded as a statistic summarising a set of data; in this case, the set of SVM decision values for an occurrence of a descent. For n heights, the combination rules used are as follows

1. Sum $\Sigma = \sum_{i=1}^n x_i$
2. StDev $\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2}$,
3. Max $Ma = \max_{i=1}^n (x_i)$,
4. Min $Mi = \min_{i=1}^n (x_i)$,
5. NumNeg = $\left(\sum_{i=1}^n i \right)$ where $x_i < 0$,
6. SumNeg $\Sigma^- = \left(\sum_{i=1}^n x_i \right)$ where $x_i < 0$,

$$7. \text{ SumPos } \Sigma^+ = \left(\sum_{i=1}^n x_i \right) \text{ where } x_i \geq 0,$$

$$8. \text{ Ratio Pos/Neg} = \ln \left| \frac{\Sigma^+ + 1}{\Sigma^- - 1} \right|.$$

The sum rule has been chosen because it is able to assess the impact of positive and negative values in an additive way and thus is able to quantitatively assess the quality of an occurrence of a descent. The standard deviation is a measure of the spread of the data and it is anticipated that a high standard deviation will be indicative of an abnormal descent. It is probable that the minimum will be more useful than the maximum value as it directly measures the most abnormal value. The number of negatives measures how many of the heights returned negative values and, along with the sum of the negative values, should be a good measure of the degree of abnormality of the descent. The sum of the positive values should provide a measure of the normality of the descent and the ratio rule, similar to the sum rule, should be able to consider the descent as a whole.

It is noted that the sum rule is a linear combination of the sum of the negatives rule and the sum of the positives rule. This is not advisable in general when choosing features but for this problem, a brute force method will be used which considers all possible combinations of rules in order to find the best subset of rules.

The motivation behind introducing the combination rules was the concern that by using the raw-values, large abnormalities for one stage could distort the overall score. Furthermore, if the descent was represented by a large number of heights, the 2nd phase feature space would have a large number of dimensions. This would mean there would need to be training data numbering at least an order of magnitude larger than the number of heights. For some applications this could be hard to satisfy. By using the combination rules, the dimension of the second phase feature space will remain small and large amounts of training data will not be needed.

5. Proposed Method

In this section, the proposed two phase method is detailed. The first phase is designed to assess the level of abnormality at each height in the descent by training a one class SVM on snapshot data from each height. The magnitude of the SVM output measures the degree of normality/abnormality for positive/negative outputs respectively. Each occurrence of a descent can now be represented as a feature vector consisting of the SVM output at each height in the descent.

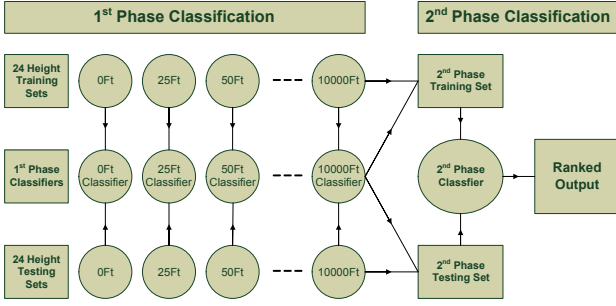
The second phase of the method looks at all of the individual descents and determines which is the most abnormal overall. Two methods of representing individual descents are compared. The first considers each descent represented by the raw SVM outputs. The second method calculates the combination rules (see section 4) from the SVM outputs and these represent the descent. A new SVM is trained on individual descents represented by these feature vectors of combination rules or the raw values. The combination rules (CR) and raw values (RV) that were created from data in the training sets for phase 1 also form the training set for phase 2. The same process occurs for the testing set descents. The SVM outputs can then be ranked in ascending order, i.e. with the negative values ranked highest.

In summary, phase 1 of the method is concerned with identifying abnormalities at individual heights for individual descents. Phase 2 is concerned with assessing the impact of these abnormalities over the whole of the descent and comparing this to other descents to see if there is any significant difference. Figure 1 shows the method in block diagram form. A more detailed description can be found in bullet point form in section 5.1.

5.1. Method Details

- Using relevant domain knowledge, identify the airport for which the descents are being made into and select a suit-

Figure 1: Block Diagram of Proposed Method



able number of heights.

- Create training and testing sets for each of the heights.
- Train a one class SVM for each height.
- For each descent in the training and testing sets, form a feature vector containing the combination rules computed from the SVM output from each height.
- Feature vectors formed from SVM outputs from training data form a new training set for the second phase. Likewise for feature vectors created from testing data.
- Train a new one class SVM on the training set of feature vectors and test it using the corresponding testing set.
- The SVM output ranks the descents by their level of overall abnormality.

6. Experiments

6.1. Overview

In this section, the proposed method is tested on artificial data and real world flight data. The artificial data is used to establish if the combination rules provide better performance than using the raw outputs and if so, which combination rules achieved the best results.

The flight data is obtained from jet aircraft making a descent and landing on the same runway. The data has been extracted from data obtained from that airline's Flight Data Monitoring

(FDM) program. For more details regarding FDM, see references [23, 24, 25]. Each aircraft is equipped with a flight data recorder and data from the flights can be downloaded and analysed for exceedances. The dataset used consists of a number of descents onto the same runway over a 13 month period from June 2007 to June 2008. Flight data experts have analysed each descent and identified those which are regarded as abnormal (or unusual). For full details, see section 6.4.

6.2. Error Metrics

To assess the performance of the classifiers in this paper, the standard confusion matrix will be utilised, where True Positive (TP) denotes the percentage of correctly identified normal descents, True Negative (TN) denotes the percentage of correctly identified abnormal descents, False Positive (FP) denotes the percentage of incorrectly identified normal descents and False Negative (FN) denotes the percentage of incorrectly identified abnormal descents.

The Balanced Error Rate (BER) [26] is a useful measure of the impact of misclassification on both classes and is given by

$$\text{BER} = (\text{FP} + \text{FN})/2. \quad (11)$$

It is a very useful error metric in one class classification problems where there is an imbalance between positive and negative examples. Consider an example with 90 positive examples and 10 negative examples and a classifier that predicts all examples are positive. The accuracy is 90% and the error is only 10% which gives the impression of a strong model. The BER however is 50%, highlighting the fact that the classifier is very poor at detecting negative examples.

6.3. Artificial Data

To test the proposed method, Gaussian data was generated along with outliers to identify how well descents with outliers

Table 1: Parameter Values for one class SVM.

Phase	Name	Values
1	σ_{P1}	10^i for $i = -3, -2, -1, 0, 1$
1	ν_{P1}	0.01, 0.05, 0.1
2	σ_{P2}	10^i for $i = -3, -2, -1, 0, 1$
2	ν_{P2}	0.01, 0.05, 0.1

Table 2: Parameter Details for the Artificial Data Experiment

Parameter	Values
Number of Features	2, 5, 10
Number of Sample Points	10, 20, 50
Covariance Matrix Scale Multiplier	1.25, 1.5, 2

could be detected. Artificial outlier data for a one class classification problem can be difficult to generate because there is a danger that the outlier data is too distant from the target class. This means that the problem becomes a two class classification problem and so the outlier model is too unrealistic. This problem is avoided by using the same distribution for the target class and the outlier class but multiplying the covariance matrix by small numbers to generate outliers, thus ensuring that the majority of outliers are close to the target class.

For this experiment the artificial data consists of ten, twenty or fifty sample points. A training set for an arbitrary sample point consists of 250 Gaussian data points with a mean vector containing all zeros. The covariance matrix is the identity matrix. The testing set for the same sample point consisted of another 150 Gaussian data points generated in the same manner and 50 outlier Gaussian data points. The outlier data is generated from the data in the training set by multiplying the covariance matrix by a scale factor of 1.25, 1.5 or 2. The number of features is 2, 5 or 10. Full details of the parameters used in the experiment are found in table 2. The software used to generate the artificial data was the DD.Tool box for Matlab [27] created by David Tax who is an expert on one class classification. The one class SVM used is LIBSVM for Matlab version 2.9 [22], one of the best SVM implementations.

Parameter ranges for the one class SVM classifiers can be found in table 1.

Table 3: Best BER for the Artificial Data Experiment with 10 Sample Points

Number of Features	Cov. Matrix Multiplier	RV BER	CR BER
2	1.25	35%	28.5%
2	1.5	29.5%	25%
2	2	19%	10.5%
5	1.25	32%	27.5%
5	1.5	23%	20%
5	2	10.5%	5.5%
10	1.25	26.5%	20.5%
10	1.5	18.5%	10.5%
10	2	3.5%	0%

6.3.1. Results

This section shows the BERs for each of the artificial experiments and compares the results if the raw values or the combination rules are used as the inputs for the second phase SVM. Tables 3, 4 and 5 show the results for 10, 20 and 50 sample points respectively. In all cases the combination rules produce a superior BER by a significant margin. This is perhaps because the combination rules are better able to describe the 'shape' of the graphed phase 1 SVM outputs which enables the phase 2 SVM to obtain better results. Furthermore, there is likely to be some redundancy in using all the SVM outputs (raw values) as not all sample points may be significant. This is likely in this instance due to the fact that the artificial abnormal data is abnormal at every sample point. In general, for both methods, classification improves with more sample points though this could be due to the same reason.

6.4. Flight Data Experiment

In this section, the proposed method is tested on real data from jet aircraft descending and preparing to land. Snapshot data is taken from certain heights in the descent, which of course all aircraft must pass through if they are to land. The data set consists of 1,518 descents by one operator onto the same runway at the same airfield. All the descents were analysed by flight data experts to identify any that had significant abnormalities and 63 were found. These 63 descents, along with another 240 normal descents selected at random form the testing set whilst the remaining 1,215 descents form the training set.

Existing methods of flight data analysis are event based (see chapter 5 of [25]). The event based approach involves checking to see whether aircraft parameters exceed given limits, for example, if the airspeed exceeds a fixed limit between predefined heights then this is regarded as an event. Exceedances are graded by three levels, level 1 being a minor exceedance and level 3 a major exceedance. An example of an event is the following: Pitch angle low during final approach between 1000ft and 100ft. Level 1 limit = -2 deg, level 2 limit = -4 deg and level 3 limit (most serious) = -6 deg. Level 3 events are regarded as severe and are reported to the flight safety officer. Level 2 and 1 events are used for statistical purposes.

For the descent there are 24 sample points, corresponding to heights that the aircraft descends through during the descent. The heights are 10000ft, 9000ft, 8000ft, 7000ft, 6000ft, 5000ft, 4000ft, 3500ft, 3000ft, 2500ft, 2000ft, 1500ft, 1000ft, 750ft, 500ft, 400ft, 300ft, 200ft, 150ft, 100ft, 75ft, 50ft, 25ft and 0ft.

In this experiment the number of heights is varied to determine if this has any impact on the results, along with all possible non empty subsets of the set of all 8 combination rules.

Table 1 contains the one class SVM parameter ranges for the experiment. The number of features for the snapshot data at

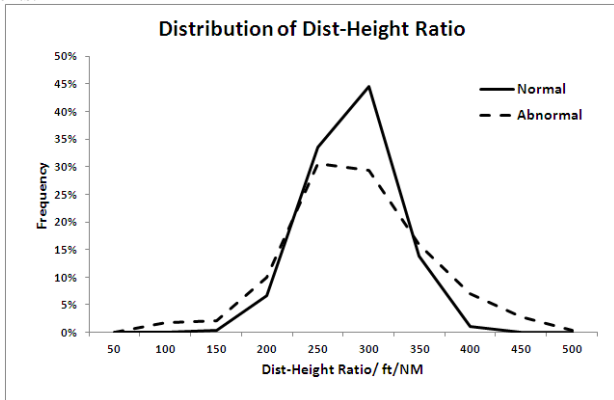
Table 4: Best BER for the Artificial Data Experiment with 20 Sample Points

Number of Features	Cov. Matrix Multiplier	RV BER	CR BER
2	1.25	33%	31%
2	1.5	21.5%	18.5%
2	2	7.5%	5.5%
5	1.25	29.5%	24.5%
5	1.5	17%	8%
5	2	6%	1%
10	1.25	20%	12%
10	1.5	11.5%	1.5%
10	2	4%	0%

Table 5: Best BER for the Artificial Data Experiment with 50 Sample Points

Number of Features	Cov. Matrix Multiplier	RV BER	CR BER
2	1.25	28%	22.5%
2	1.5	12.5%	5%
2	2	9%	0.5%
5	1.25	25.5%	15%
5	1.5	9%	4.5%
5	2	7.5%	0%
10	1.25	16.5%	4%
10	1.5	13.5%	0%
10	2	8.5%	0%

Figure 2: Distance-Height Ratio Distribution for Normal and Abnormal Descents.



each height varies between 5 and 12. It is known that a good choice for σ is the reciprocal of the number of features and the range chosen reflects this. 10 fold cross validation was used during training.

In the results section of this experiment, there is reference to the term Descent Abnormality Profile (DAP). This is a chart plotted for an individual descent for which the raw SVM output from each of the phase 1 classifiers is plotted against the heights that those SVMs represent. This enables the user to immediately identify any regions of abnormality and the heights at which they occur at.

6.4.1. Features

Table 6 shows all the features used in the flight data. These were selected via expert knowledge. Note that not all features are used at each height.

In this section, some of the features used in this dataset are analysed. Figure 2 shows the differences in the distributions of the Distance-Height ratio for the normal and abnormal descents. All heights for which this parameter was used are included in this chart and range from 10000ft to 2500ft. The majority of the normal data is contained between 250Ft/NM and 350Ft/NM whereas the abnormal data has a larger proportion of its values in the higher and lower regions.

Figure 3 shows the differences in the Indicated Airspeed dis-

Figure 3: Indicated Airspeed Distribution for Normal and Abnormal Descents.

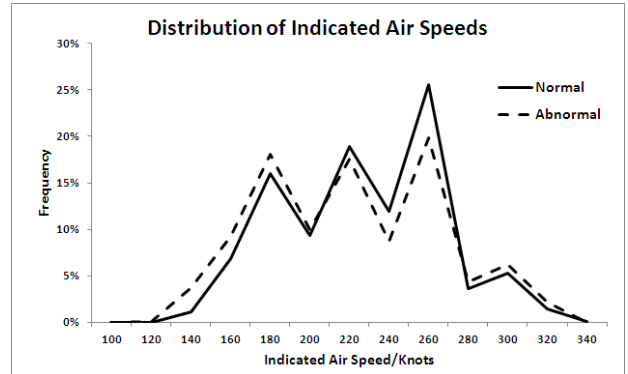
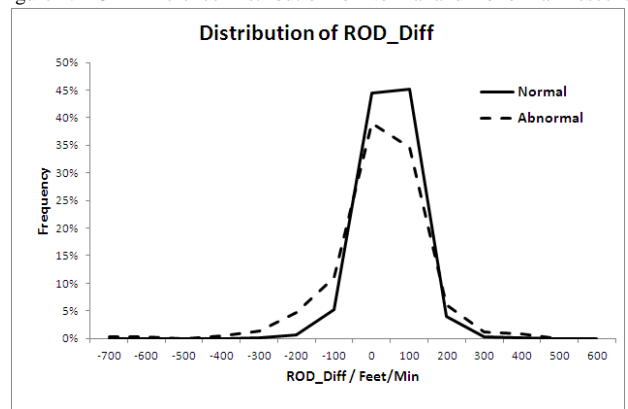


Figure 4: ROD Difference Distribution for Normal and Abnormal Descents.



tributions for the normal and abnormal descents. The spiky nature of the middle part of the chart highlights that the aircraft for the most part of the descent aim to fly at specific speeds. Though it is not as distinct as the Distance-Height chart, there is a larger proportion of the abnormal data in the higher and lower regions. An important point to note is that many of the data points on the abnormal descents will be normal; thus making it harder to detect overall differences.

Figure 4 shows the differences in the Recommended Rate of Descent (ROD) distribution for the normal and abnormal descent. This data ranges from 1000ft to 100ft. For a given aircraft groundspeed, it is the recommended rate of descent in order to land on the correct part of the runway. Like figure 2, most of the normal data is between -100 and 200 feet per minute, the centre part of the graph. There is a greater proportion of abnormal data in the tails of the distribution.

Table 6: List of features used.

Name	Unit	Typical Range	Description
V-Vref	Knots	-5 to 50	Difference between airspeed and reference landing speed
IAS	Knots	110 to 300	Indicated Airspeed
IVV	Feet/Min	-4000 to 0	Rate of Descent
Pitch	Degrees	-2 to 5	Angle of aircraft relative to the horizon
Glideslope Deviation	Dots	-3 to 3	Deviation in the vertical from optimum landing path
Localiser Deviation	Dots	-3 to 3	Deviation in the horizontal from optimum landing path
Flap	Degrees	0 to 30	Flap setting
Landing Gear	No Units	0 or 1	Landing gear deployment
Speedbrake	No Units	0 or 1	Speedbrake deployment
Engine Speed	No Units	30 to 70	Percentage of nominal maximum speed
Ratio of height to distance to landing	Feet/NM	200 to 400	Height divided by track miles to landing
Difference between IVV and Recommended Rate of Descent (ROD)	Feet/Min	-300 to 300	Difference between actual descent rate and recommended descent rate

6.4.2. Results

Table 7 shows the main results for the flight data experiment. As expected, by increasing the number of heights, the BER for both approaches falls. The raw value method and the combination rule method produce similar results with the combination rule approach having a lower BER at 10 heights. A two-tailed t-test was also used to analyse the data. The combination rule approach has a significantly lower average BER with 10 heights ($t(18) = 5.2807$, $p < 0.05$) though with 13 heights the raw value approach BER is significantly lower ($t(18) = 4.9906$, $p < 0.05$). However for 24 heights, there is no significant difference between the two methods ($t(18) = 1.6612$, $p < 0.05$). What is interesting to note is that the best rule set for all experiments contains rule 4 and 8 (minimum and ratio). This is logical given that they make statements about the abnormal regions of the DAPs and therefore should be valuable in describing them.

Of the 63 descents which have abnormalities, only 19 were detected by traditional flight data analysis methods using the event based parameter exceedance approach (see Appendix B of [25] for details). There were no false positives and the de-

Table 7: Best BER for the Flight Data Experiment

Number of Heights	Best Combination Rule	RV BER (SD)	CR BER (SD)
10	(2,4,5,6,7,8)	9.6% (0.7%)	7.9% (0.7%)
13	(1,4,6,8)	4.9% (0.8%)	6.4% (0.4%)
24	(4,5,8)	3.0% (0.5%)	3.4% (0.6%)

tection rate was 30%. This gives a BER of 35% which is much higher than the results for the raw value and combination rule methods in table 7.

6.5. Analysis of Select Descents of the Flight Data

In this section, 3 descents are studied to highlight how the method provides information about the degree of abnormality at various heights.

6.5.1. Descent 1 - Very steep descent.

The large negative region on figure 5 is caused by the very steep descent of the aircraft. At 10000ft, the aircraft has just 24NM track miles to go compared to the average value of 40NM. This leads to high rates of descent, high airspeeds and heavy speedbrake usage. Furthermore, at 2500ft and 2000ft,

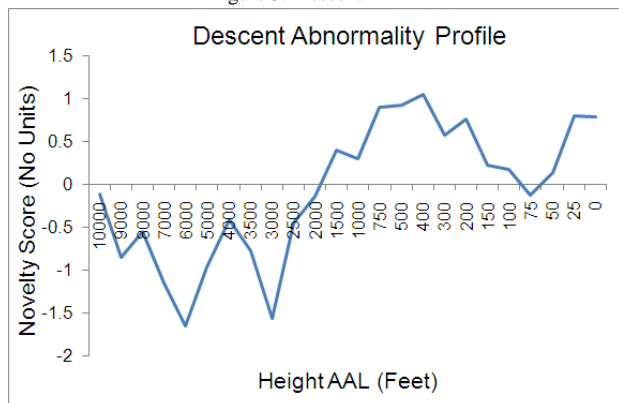
Table 8: A Sample of Points of Interest on Descent 1.

Height	Parameter	Parameter Value	Parameter Percentile	Average Parameter Value
10000	DISTRAT	23.75	0	40.05
9000	DISTRAT	20.36	0	36.07
8000	DISTRAT	18.03	0	32.22
7000	DISTRAT	15.54	0	27.22
6000	DISTRAT	13.03	0	22.71
3000	IVV	-3002	0	-973.36
2500	IVV	-2368	1	-875.33
2000	IVV	-1926	1	-798.40
1500	IVV	-1507	1	-823.33

Table 9: Event List Descent 1.

Event Name	Severity Level	Height
High Descent Rate >2000ft	1	2935
High Speed 500-50ft	1	286

Figure 5: Descent 1 DAP



the aircraft has the speedbrakes deployed but with more than 10 degrees of flap set, which is prohibited in the airline’s Standard Operation Procedure (SOP). However, the aircraft manages the descent well as seen by the largely positive region of flight after 1000ft. This is an example where a potentially unsafe approach has been corrected and the lack of high severity level events shows this. Nonetheless a flight safety officer would be interested in this descent as it may indicate a wider problem. See table 9 for a list of events.

Table 11 shows some of the heights of interest for this flight.

Table 10: A Sample of Points of Interest on Descent 2.

Height	Parameter	Parameter Value	Parameter Percentile	Average Parameter Value
500	V-Vref	30.16	100	8.336
500	RODDIFF	-139.9	5	4.87
500	Pitch	-2.1	0	1.871
400	V-Vref	27.16	99	8.239
400	Pitch	-2.5	0	2.002
300	V-Vref	27.16	100	7.914
300	RODDIFF	-229.2	1	2.87
300	Pitch	-0.7	1	2.073

Table 11: Event List Descent 2.

Event Name	Severity Level	Height
Pitch Low 1000-100ft	1	568
High Speed 500-50ft	3	284
Low Pitch at Touchdown	3	20
G Landing	1	0

6.5.2. Descent 2 - High speed event

At 10000ft the aircraft is 60NM from the runway at an air-speed of 207kts. The average track miles to landing is 40NM and the average indicated airspeed is 275kts. From the available evidence the aircraft chose a shallow descent because of high winds. Once the aircraft reaches a height of around 750ft, the airspeed begins to increase and the pitch angle becomes negative. See table 11 for a list of events.

Table 10 shows some of the heights of interest for this flight. The slightly negative region shown on the DAP (see figure 6) resulted from the aircraft descending earlier than usual and at a slower than average indicated airspeed. Whilst this is not unsafe, it is unusual. However, the main point of interest is after 500ft. At 1000ft the aircraft satisfies the criteria for a stable approach but from 500ft, the airspeed has increased rapidly and the pitch angle is negative. The impact of these parameters is visible on the DAP.

Figure 6: Descent 2 DAP

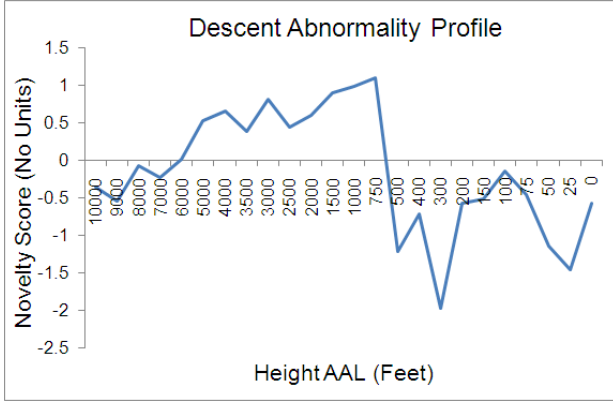


Figure 7: Descent 3 DAP

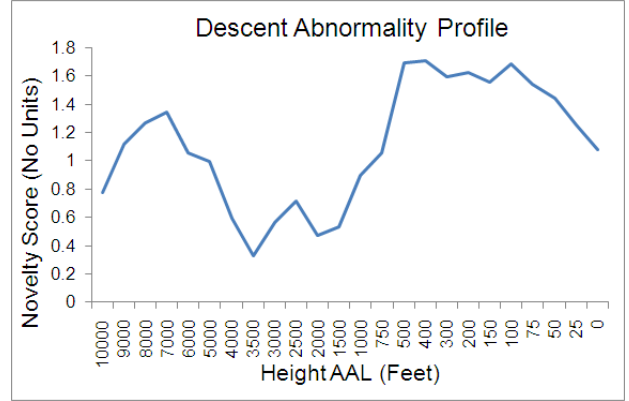


Table 12: A Sample of Points of Interest on Descent 3.

Height	Parameter	Parameter Value	Parameter Percentile	Average Parameter Value
2000	IAS	145	1	172
2000	Flap	25	98	8.52
1500	IAS	130	2	156

6.5.3. Descent 3 - Normal descent

This descent is smooth with an airspeed and a rate of descent typical for this approach. Landing gear and flaps are deployed at typical heights and by 1500ft, the aircraft is established on the Instrument Landing System (ILS) with a normal speed. By 1000ft, the aircraft's airspeed is around $v_{ref} + 8$ kts with a rate of descent appropriate for its groundspeed. The approach power is set and flap 30 (landing flap) has been chosen.

Table 12 shows some of the heights of interest for this flight. The DAP (see figure 7) is such that all data points are positive, highlighting that the descent has been conducted with the majority of parameters at the different heights within normal ranges. Table 12 shows that some of the parameters at certain heights were abnormal but their impact on the whole descent was not enough to make the descent itself abnormal or for that matter, any of the heights.

7. Discussions of Method and Results

7.1. Discussion of Method

The benefits of such a method are that firstly abnormalities can be detected and individual descents can be ranked against others. Phase 2 outputs can be stored and over time, this historical information could prove very valuable in assessing any changes in the act of descending. It could be affected by very cold or very hot weather or there could be differences between early descents and later descents resulting from changes to the standard operating procedures. All of this information can be utilised to improve the responses of the maintenance teams and the flight safety officers. The ability of the one class SVM to interpret abnormal data in the training set is important because whilst it may be possible to obtain faults with which to test on, there might be examples of unusual data which contains no faults but they are still different to normal approach conditions. The ability to handle this unseen data is very important. Another benefit is that the magnitude and sign of the SVM output allows descents to be directly compared and this can be useful in assessing any differences over a period of time. There are also very few parameters to be optimised. If the Gaussian kernel is used then a grid search can be used to optimise σ_{P1} , σ_{P2} , ν_{P1} and ν_{P2} . The P1 parameters are those of the individual SVMs for phase 1 and the P2 parameters are those for the single SVM

in phase 2.

There are however some disadvantages. Firstly, it is not intuitively clear how to choose the number of heights. However, the results show that more heights led to a smaller BER. It is clear that significant domain knowledge will be needed in order to understand which heights could be informative and the features necessary to record at each height.

7.2. Discussion of Results

Looking at the results for the artificial data experiment (see tables 3 to 5), it is clear that the combination rules provide the lowest BER compared to the raw values. This is in part because the combination rules are better able to describe the shape of the profile created from the phase 1 SVM outputs for each occurrence. Furthermore, some of the raw values are likely to be redundant due to the artificial nature of the data in that the abnormal occurrences contain outlier data at every sample point; thus making it easier for the classifier to successfully discriminate between normal and abnormal occurrences.

It is also unlikely that abnormal occurrences of a descent will be abnormal at every sample point, thus making them harder in principle to detect.

In the flight data experiment (see table 7), the BERs for the raw value method and the combination rule method are closer and in fact there is little difference between the two methods as the combination rule method is significantly lower with 10 heights, the raw value method is significantly lower with 13 heights and there is no significant difference with 24 heights.

The results for the artificial data set (see tables 13 to 15) are illuminating in terms of the best combination rules to choose. For the experiments with just 2 features at each sample point, the rules appear roughly the same number of times. This suggests that with so little discriminative information, it is a difficult classification problem, which is also confirmed by the high BER values. When 5 features are used, rule 2 (standard de-

viation) appears the least often. This is probably because all the normal occurrences have nearly all positive sample point outputs whereas the abnormal occurrences will have nearly all negative outputs. Hence the standard deviation may be similar between the two sets. Rules 1 and 8 (sum and ratio) appear the most often. This could be due to the fact that they incorporate information about the normal (positive outputs) and the abnormal (negative outputs) into a single figure whereas rule 3 (max rule) for example only makes a statement about the normal sample points. For 10 features, rules 2, 4 and 6 (standard deviation, min, positive sum) appear the least often. The fact that rule 4 appears the least often is unusual given that a low minimum is an indication of an abnormality or unusual behaviour. However, given that the outlier data will be similar for each sample point (since they were generated from the same Gaussian parameters), the minimum may not stand out very much.

Another point to make is that although there were 8 rules available, the classifier never needed more than 6 rules to achieve the best BER and usually 2 or 3 was enough. In some cases one rule was enough but this is likely to be because of the artificial nature of the data.

It is important to note that there are some important differences between the artificial data and the flight data. Whilst the artificial abnormal occurrences had abnormalities at nearly all of the sample points, this was not the case for the flight data. In fact some normal flights had abnormalities at some heights but in the opinion of the flight data experts, these were not enough to make the label of those descents abnormal. Similarly, descents that were regarded as abnormal often had regions of normal flight. For this reason, it is not surprising that rule 8 is one of the best rules in the 3 experiments (see table 13). It is able to consider the impact of the negative and positive regions in terms of number and magnitude and is therefore ideal for detecting abnormalities. Rule 4 is also prominent because a low

Table 13: Best BER and Combination Rules for the Artificial Data Experiment with 10 Heights

Number of Features	Cov. Matrix Multiplier	Combination Rule	CR BER
2	1.25	(1,2,3,6,7,8)	28.5%
2	1.5	(1,2,3,4), (2,3,4,7), (2,3,4,8)	10.5%
2	2	(3,8)	10.5%
5	1.25	(1,4,6), (4,6,7), (4,6,8)	27.5%
5	1.5	(3,5,7), (3,7,8)	20%
5	2	(3,8)	5.5%
10	1.25	(8)	20.5%
10	1.5	(1,3,5,7)	10.5%
10	2	(1,5), (5,8)	0%

Table 14: Best BER and Combination Rules for the Artificial Data Experiment with 20 Heights

Number of Features	Cov. Matrix Multiplier	Combination Rule	CR BER
2	1.25	(1,6), (4,6), (5,6), (6,7), (6,8)	31%
2	1.5	(7,8)	18.5%
2	2	(1,7)	5.5%
5	1.25	(1,2,8), (5,6,8)	24.5%
5	1.5	(1)	8%
5	2	(3,4,7)	1%
10	1.25	(1,8), (7,8)	12%
10	1.5	(1,3)	1.5%
10	2	(2,5)	0%

minimum will almost certainly indicate the presence of a significant abnormality.

Table 16 shows the average BER for all appearances of each rule for each height experiment with the flight data. For all experiments rule 8 produces the lowest average BER which is consistent for reasons already explained. Rule 3 produces the highest average BER for all experiments which is in contrast with the artificial data results. However, it highlights that even an abnormal flight can have many normal heights (hence a similar maximum to a normal descent) and therefore it struggles to discriminate between the normal and abnormal descents.

Table 15: Best BER and Combination Rules for the Artificial Data Experiment with 50 Heights

Number of Features	Cov. Matrix Multiplier	Combination Rule	CR BER
2	1.25	(1,3,4,5,7)	22.5%
2	1.5	(1,2,3,7,8)	5%
2	2	(1,2,4,5), (2,4,5,7), (4,5,6,7)	0.5%
5	1.25	(1,5,8)	15%
5	1.5	(1,3,7), (1,3,8)	4.5%
5	2	(1), (7), (8)	0%
10	1.25	(1)	4%
10	1.5	(7)	0%
10	2	(7)	0%

Table 16: Best Average BER for each Combination Rule for the Flight Data Experiment. H stands for heights.

CR Number	Average BER (10 H)	Average BER (13 H)	Average BER (24 H)
1	21.6%	19.2%	14.3%
2	21.6%	19.3%	13.8%
3	23.4%	21.0%	16.5%
4	21.5%	19.1%	13.8%
5	21.4%	19.2%	14.2%
6	22.0%	19.7%	14.7%
7	22.3%	20.0%	15.2%
8	21.1%	18.7%	13.6%

Table 17: Average Correlation for each Combination Rule over all SVM parameters for 24 Heights in the Flight Data Experiment

Combination Rule Number	Average Correlation (St Dev)	Average BER (St Dev)
1	0.774 (0.170)	14.3% (9.2%)
2	0.661 (0.287)	13.8% (9.3%)
3	0.301 (0.115)	16.5% (9.8%)
4	0.667 (0.238)	13.8% (8.8%)
5	0.667 (0.201)	14.2% (9.6%)
6	0.687 (0.238)	14.7% (9.7%)
7	0.622 (0.174)	15.2% (9.3%)
8	0.758 (0.188)	13.6% (8.9%)

7.2.1. Analysis of Second Phase Features

Table 17 shows the average degree of correlation between each of the second phase features. The average correlation between rules can be computed by considering the correlations for each permutation of phase 1 SVM parameters. Let values for one rule be denoted by $X = \{x_i | 1 \leq i \leq N\}$ and let the values for the other rule be denoted by $Y = \{y_i | 1 \leq i \leq N\}$. Then the linear correlation between the two variables is given by

$$Correl(X, Y) = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^N (y_i - \bar{y})^2}} \quad (12)$$

It highlights that whilst rule 3 (maximum) has very low correlation with any rule, inclusion of this rule significantly raises the average BER. This demonstrates that it is not a useful rule because even abnormal descents can have high maximums.

Table 18 shows the average BERs and average best BERs when using a certain number of rules. As is perhaps expected, using more rules reduces the average BER but interestingly 6 is the optimum number in terms of best average BER. This reflects the complexity of the dataset. However, due to the high correlation between many of the features, the lowest BER is achieved with only 3 rules (see table 10).

Table 18: Average BER for a set number of Combination Rules used over all SVM parameters for 24 Heights in the Flight Data Experiment

Number of Rules Used	Average BER (St Dev)	Average Best BER (St Dev)
1	23.3% (10.8%)	9.7% (11.6%)
2	17.9% (5.7%)	6.1% (3.7%)
3	15.7% (3.6%)	4.9% (2.0%)
4	14.5% (2.5%)	4.3% (1.2%)
5	13.7% (1.8%)	4.1% (0.8%)
6	13.1% (1.3%)	4.0% (0.6%)
7	12.6% (0.8%)	4.1% (0.4%)
8	12.1% (n/a)	4.3% (n/a)

8. Conclusion

In this paper, a method that demonstrates two different ways to combine one class classifiers to identify abnormalities in aircraft descents and rank multiple descents has been introduced. The results show that on artificial data, using combination rules, rather than the raw SVM outputs, achieves a lower BER for all experimental parameters as shown in 2. It also highlights that there is no optimum set of combination rules to achieve a low BER on artificial data. However, of the combination rules, rules 1, 7 and 8 (sum, sum of negatives and ratio) appear the most often. Rules 1 and 8 are able to assess both positive and negative regions so it is not surprising that they perform well whilst rule 7 is useful given that the abnormal occurrences are designed to be abnormal at each sample point.

The experiments on the real world flight data set produced results different to those from the artificial experiment. This is due to the fact that for abnormal descents, abnormalities are unlikely to occur at every height. The results demonstrate that rules 4 and 8 (minimum and ratio) perform well on a difficult dataset. Both are included in the best set of rules for each choice of heights.

Although the combination rule method is no worse statistically than the raw value method, it does have an added advan-

tage. The dimensionality of the feature space for the 2nd phase SVM is equal to the number of heights. This means that for large numbers of heights, the training set would need to contain task instances numbering at least an order of magnitude larger if classifying via the raw-value method. If the combination rule method is used, the dimensionality remains the same which means the training set does not need to be as large.

Section 6.5 demonstrates that the method can accurately identify abnormalities and assess their impact in the form of the DAP which allows the user to easily to assess any points of interest. By collecting historical data, the user can identify if there are any changes at any of the heights and perform the appropriate actions.

For future work, it is intended to investigate whether different heights are needed for different runway approaches and whether there is an optimum number of heights that produces the best BER.

Acknowledgment

This work was supported by EPSRC Industrial CASE Voucher Number 06001600. The authors would like to thank Flight Data Services Ltd for their collaboration with this work.

References

- [1] B. Li, M.-Y. Chow, Y. Tipsuwan, J. Hung, Neural-network-based motor rolling bearing fault diagnosis, *Industrial Electronics, IEEE Transactions on* 47 (2000) 1060–1069.
- [2] S. Nandi, H. Toliyat, Condition monitoring and fault diagnosis of electrical machines—a review, in: *Industry Applications Conference, 1999. Thirty-Fourth IAS Annual Meeting. Conference Record of the 1999 IEEE*, volume 1, IEEE, pp. 197–204.
- [3] F. Filippetti, G. Franceschini, C. Tassoni, P. Vas, Recent developments of induction motor drives fault diagnosis using ai techniques, *Industrial Electronics, IEEE Transactions on* 47 (2000) 994–1004.
- [4] A. Widodo, B.-S. Yang, Support vector machine in machine condition monitoring and fault diagnosis, *Mechanical Systems and Signal Processing* 21 (2007) 2560–2574.
- [5] N. Saravanan, V. Siddabattuni, K. Ramachandran, Fault diagnosis of spur bevel gear box using artificial neural network (ANN), and proximal support vector machine (PSVM), *Applied Soft Computing* 10 (2010) 344–360.
- [6] E. Zio, P. Baraldi, G. Gola, Feature-based classifier ensembles for diagnosing multiple faults in rotating machinery, *Applied Soft Computing* 8 (2008) 1365–1380.
- [7] S. Rajakarunakaran, P. Venkumar, D. Devaraj, K. Rao, Artificial neural network approach for fault detection in rotary system, *Applied Soft Computing* 8 (2008) 740–748.
- [8] S. Bouhouche, M. Yahi, J. Bast, Combined Use of Principal Component Analysis and Self Organization Map for Condition Monitoring in Pickling Process, *Applied Soft Computing* (2010).
- [9] A. Saxena, A. Saad, Evolving an artificial neural network classifier for condition monitoring of rotating mechanical systems, *Applied Soft Computing* 7 (2007) 441–454.
- [10] M. Markou, S. Singh, Novelty detection: a reviewpart 1: statistical approaches, *Signal Processing* 83 (2003) 2481–2497.
- [11] M. Markou, S. Singh, Novelty detection: a reviewpart 2: neural network based approaches, *Signal Processing* 83 (2003) 2499–2521.
- [12] J. Shawe-Taylor, P. Bartlett, R. Williamson, M. Anthony, Structural risk minimization over data-dependent hierarchies, *Information Theory, IEEE Transactions on* 44 (1998) 1926–1940.
- [13] B. Schölkopf, R. C. Williamson, A. J. Smola, J. Shawe-Taylor, J. C. Platt, Support vector method for novelty detection, in: S. A. Solla, T. K. Leen, K.-R. Müller (Eds.), *Advances in Neural Information Processing Systems* 12, MIT Press, Cambridge, MA, 2000, pp. 582–588.
- [14] B. Cannas, R. Delogu, A. Fanni, P. Sonato, M. Zedda, Support vector machines for disruption prediction and novelty detection at JET, *Fusion engineering and design* 82 (2007) 1124–1130.
- [15] R. Perdisci, G. Gu, W. Lee, Using an ensemble of one-class svm classifiers to harden payload-based anomaly detection systems, in: *Data Mining, 2006. ICDM'06. Sixth International Conference on*, IEEE, pp. 488–498.
- [16] A. Gardner, A. Krieger, G. Vachtsevanos, B. Litt, One-class novelty detection for seizure analysis from intracranial eeg, *The Journal of Machine Learning Research* 7 (2006) 1025–1044.
- [17] L. Clifton, H. Yin, Y. Zhang, Support vector machine in novelty detection for multi-channel combustion data, *Lecture Notes in Computer Science* 3973 (2006) 836.
- [18] D. Tax, M. Van Breukelen, R. Duin, Combining multiple classifiers by averaging or by multiplying?, *Pattern recognition* 33 (2000) 1475–1485.
- [19] B. Schölkopf, J. Platt, J. Shawe-Taylor, A. Smola, R. Williamson, Esti-

mating the support of a high-dimensional distribution, *Neural Computation* 13 (2001) 1443–1471.

- [20] B. E. Boser, I. M. Guyon, V. N. Vapnik, A training algorithm for optimal margin classifiers, in: *COLT '92: Proceedings of the fifth annual workshop on Computational learning theory*, ACM, New York, NY, USA, 1992, pp. 144–152.
- [21] V. Vapnik, *Statistical learning theory*, Wiley New York, 1998.
- [22] C.-C. Chang, C.-J. Lin, *LIBSVM: a library for support vector machines*, 2001. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [23] *A guide to flight data monitoring*, Flight Data Services Ltd, 2010. Accessed on 26/01/2010. Available at <http://www.flightdataservices.com>.
- [24] M. Holtom, *Foqa - flight data analysis of aircraft for flight safety*, The Airline Pilots, 2006. Accessed on 13th March 2010. Available at <http://tiny.cc/sv8mn>.
- [25] CAA, *Cap 739 flight data monitoring*, 2003. Last Accessed on 15th November 2010.
- [26] I. Guyon, S. Gunn, A. Ben-Hur, G. Dror, Result analysis of the nips 2003 feature selection challenge, *Advances in Neural Information Processing Systems* 17 (2005) 545–552.
- [27] D. Tax, *Ddtools, the data description toolbox for matlab*, 2009. Version 1.7.3.