Proceedings of the Federated Conference on
Computer Science and Information Systems pp. 125–131

ISBN 978-83-60810-22-4

Identification of Patient Deterioration in Vital-Sign Data using One-Class Support Vector Machines

Lei Clifton*, David A. Clifton*, Peter J. Watkinson†, and Lionel Tarassenko*

*Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Oxford, UK
{lei.clifton, david.clifton, lionel.tarassenko}@eng.ox.ac.uk
†Nuffield Department of Anaesthetics, University of Oxford, Oxford, UK

***Abstract*—Adverse hospital patient outcomes due to deterioration are often preceded by periods of physiological deterioration that is evident in the vital signs, such as heart rate, respiratory rate, etc. Clinical practice currently relies on periodic, manual observation of vital signs, which typically occurs every 2-to-4 hours in most hospital wards, and so patient deterioration may go unidentified. While continuous patient monitoring systems exist for those patients who are confined to a hospital bed, the false alarm rate of conventional systems is typically so high that the alarms generated by them are ignored. This paper explores the use of machine learning methods for automatically identifying patient deterioration, using data acquired from continuous patient monitors. We compare generative and discriminative techniques (a probabilistic method using a mixture model, and a support vector machine, respectively). It is well-known that parameter tuning affects the performance of such methods, and we propose a method to optimise parameter values using "partial AUC". We demonstrate the performance of the proposed method using both synthetic data and patient vital-sign data collected from a recent observational clinical study.**

***Index Terms*—support vector machine, novelty detection, one-class classification, parameter optimisation, partial AUC.**

I. Introduction

A. Detecting Patient Deterioration

ADVERSE events in acutely ill hospital patients may occur when their physiological condition is not recognised or acted upon early enough [1]. Clinical guidance in the UK [2] recommends the regular observational recording of certain vital signs (such as heart rate, HR, measured in beats per minute; respiration rate, RR, measured in breaths per minute; blood oxygen saturation, SpO_2, measured as a percentage; and systolic blood pressure, SysBP, measured in mmHg), combined with the use of early warning score (EWS) systems. The latter involve the clinician applying univariate scoring criteria to each vital sign in turn (e.g., "score 3 if heart rate exceeds 140 beats per minute"), and then escalating care to a higher level if any of the scores assigned to individual vital signs, or the sum of all such scores, exceed some threshold.

This current standard of care has a number of disadvantages. (i) The early-warning scores assigned to each vital sign, and the thresholds against which the scores are compared, are mostly determined heuristically[1]. (ii) EWS systems are used with periodic observation of vital signs, which may be made as infrequently as once every 12 hours in some wards. Patients may deteriorate significantly between observations. (iii) There is a significant error-rate associated with manual scoring, especially in the high-workload setting of a high-dependency clinical ward. (iv) Each vital sign is treated independently and correlations between vital signs are not taken into account.

This paper addresses these four disadvantages by evaluating automated systems, which use novelty detection algorithms.

B. Novelty Detection for Patient Monitoring

Novelty detection, or one-class classification, involves construction of a model of normality using examples of "normal" system behaviour, and which then classifies test data as either "normal" or "abnormal" with respect to that model. This technique is particularly applicable to the monitoring of high-integrity systems, such as jet engines, manufacturing processes, or human patients.

Monitoring high-integrity systems is difficult due to the variability between individual systems of the same system type (such as different human patients of the same demographic background). The few examples of "abnormal" system behaviour that may exist for some population are often inapplicable to the analysis of previously-unseen individuals. For example, a heart rate of 50 beats per minute may be indicative of considerable physiological abnormality in one hospital patient, while it may be entirely normal for a fitter patient of the same age and background.

Finally, high-integrity systems typically exhibit a high degree of structural complexity, and can often comprise many sub-systems that interact in a non-linear manner. Thus, the potential space of "abnormality" is extremely large, and so the large resultant number of failure modes is often poorly understood. For example, the exact response of a particular human's physiology to a given failure mode (such as, for example, deterioration leading to myocardial infarction) will vary significantly between patients, and what data exist are insufficient for constructing accurate models of these failure states, typically being derived from a small number of patients, with differing co-morbidities, lifestyles, etc.

Novelty detection avoids such problems by modelling the "normal" mode of operation of the system, which is often well-understood because most high-integrity systems function

[1]However, a large evidence base of vital-sign data was used to construct the EWS proposed in [3], which is currently undergoing clinical validation in its own study.

"normally" most of the time, and then looking for deviations from that normal model.

This is appropriate for the monitoring of physiological condition in patients, because sufficient data exist from "stable" patients such that a model of the well-understood "normal" state of these patients may be constructed. Physiological deterioration may then be detected as being corresponding departures in the vital signs from that "normal" state.

C. Overview

This paper considers the one-class support vector machine (SVM), which is a commonly-used method of performing novelty detection. Its formulation is briefly recapped in section II, where disadvantages arising from the setting of its parameters are discussed. Following discussion on the topic of evaluating classifiers in section III, a method of optimising parameter values in a one-class SVM is described in section IV, in light of those evaluation methods. The method is illustrated using simulated data in section V and patient vital-sign data, collected from an observational clinical study, in section VI. Limitations of the method, and potential future extensions, are discussed in section VII.

II. One-Class SVMs

The one-class SVM is a frequently-employed method of performing novelty detection, and it has been applied to many such problems, including jet engine condition monitoring [4], signal segmentation [5], and fMRI analysis [6], among many others, a review of which may be found in [7].

A. Formulation

This paper considers the one-class SVM formulation proposed by [8], in which a quantity l of d-dimensional data $\{\mathbf{x}_1, \ldots, \mathbf{x}_l\} \in \mathbb{R}^d$ are mapped into a (potentially infinite-dimensional) feature space $\mathbb{F}$ by some non-linear transformation $\mathbf{\Phi}$: $\mathbb{R}^d \rightarrow \mathbb{F}$. A kernel function k provides the dot product between pairs of transformed data in $\mathbb{F}$:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{\Phi}(\mathbf{x}_i) \cdot \mathbf{\Phi}(\mathbf{x}_j) \tag{1}$$

A Gaussian kernel allows any data-point to be separated from the origin in $\mathbb{F}$ [8], hence is chosen for us in the work described by this paper:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / 2\sigma^2\right) \tag{2}$$

where σ is the width parameter associated with the Gaussian kernel.

The decision boundary is a hyperplane in feature space $\mathbb{F}$, found by minimising the weighted sum of a support vector-type regulariser and an empirical error term depending on an overall margin variable ρ and individual errors ξ_i,

$$\min_{w \in F,\ \xi \in \mathbb{R}^l,\ \rho \in \mathbb{R}} \quad \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}\xi_i - \rho \tag{3}$$

$$\text{subject to} \quad w \cdot \mathbf{\Phi}(\mathbf{x}_i) \ \geq \ \rho - \xi_i, \quad \xi_i \geq 0 \tag{4}$$

where w is a weight vector in the feature space, and C is a user-specified penalty parameter, with a larger C corresponding to a higher penalty to errors [9].

The decision function in feature space $\mathbb{F}$ is:

$$z(\mathbf{x}) = w_o \cdot \mathbf{\Phi}(\mathbf{x}) - \rho_0 \tag{5}$$

with parameters

$$w_o \ = \ \sum_{i=1}^{N_s} \alpha_i \mathbf{\Phi}(\mathbf{s}_i) \tag{6}$$

$$\rho_o \ = \ \frac{1}{N_s}\sum_{j=1}^{N_s}\sum_{i=1}^{N_s} \alpha_i k(\mathbf{s}_i, \mathbf{s}_j), \tag{7}$$

where $\mathbf{s}_i$ are the support vectors, of which there are N_s, and where k is the Gaussian kernel defined in (1). We note that $w_o \in \mathbb{F}$, $\rho_o \in \mathbb{R}$, and that α_i are Lagrangian multipliers used to solve the dual formulation, more details of which may be found in [8] and which are not reproduced here. The "abnormal" data (i.e., those outside the single, "normal" training class) take negative values of $z(\mathbf{x})$, while "normal" data take positive values.

We note in passing that this approach is typically employed in favour of the one-class formulation proposed in [10], [11], in which a hypersphere of minimum radius is found to enclose the data in $\mathbb{F}$. The interested reader is directed to a useful tutorial for this latter method in [12].

III. Receiver Operating Characteristic (ROC) curves

The performance of a two-class decision rule can be summarised in a receiver operating characteristic (ROC) curve, which plots the true-positive rate on the vertical axis against the false-positive rate (FPR) on the horizontal axis, as the decision threshold varies [13]. Although an ROC curve gives a more thorough evaluation of classifier performance than a confusion matrix, it is difficult to compare two ROC curves. One possible comparison is to consider the area-under-the-ROC-curve (AUC), which integrates the FPR over varying thresholds. AUC is independent of a fixed decision threshold, and is invariant to prior class probabilities [14]. AUC represents the probability that a randomly chosen positive observation is correctly classified, and therefore a higher value of AUC indicates better separation between the two classes [14], [15].

For the novelty detection approach taken by this work, we label the "normal" data as "negative" cases, for ROC analysis, and the "abnormal" data as "positive" cases. Most practical novelty detection systems require low FPRs, and so we are most interested in the ROC curve for low values of FPR when evaluating the performance of a novelty detector. (Its performance at higher FPRs is irrelevant, and possibly confounding, because these represent choices of decision threshold that would never be used in practice.)

We therefore consider *partial AUC*, to restrict evaluation of the classifier only over those ranges of decision threshold that

are likely to be used in practice. Partial AUC is defined as the integral area between two false-positive rates [16]. Unlike AUC, whose maximum value is always 1, partial AUC depends on the two chosen false-positive rates, over which the ROC curve is integrated. We will use this partial AUC metric for optimisation of SVM parameters.

IV. Parameter optimisation for a one-class SVM

A. Choosing appropriate parameter values

For the case of a Gaussian kernel $k(\mathbf{x}_i, \mathbf{x}_j)$, it is important to choose an appropriate value for the bandwidth parameter σ. Larger values of σ result in smoother decision boundaries, which therefore tend to exhibit lower variance (i.e., better ability to generalise to previously-unseen data), at the expense of increased bias (i.e., under-fitting the "normal" data space, as represented by the "normal" training data). Conversely, smaller values of σ provide decreased bias (i.e., a closer fit to the "normal" data space, as represented by the "normal" training data), but at the expense of increased variance (i.e., they are less able to generalise successfully to previously-unseen data). The "optimal" value for σ will depend on the distribution of the particular dataset under consideration,and it is not usually obvious how one should choose the value of σ. Typically, a cross-validation exercise is performed, where one uses a validation set as an estimation of how well the system will perform in practice, when presented with previously-unseen test data.

For a Gaussian kernel $k(\mathbf{x}_i, \mathbf{x}_j)$, the quantity $-\log k(\mathbf{x}_i, \mathbf{x}_j)$ is the Euclidean distance between two observations scaled by a factor $1/2\sigma^2$. Based on this link between σ and Euclidean distance, we propose the following method to determine an appropriate value for σ in our discriminative case, adapted from a similar method proposed in [17] for selecting σ when estimating pdfs for probabilistic inference within a generative framework:

1) First, we calculate the local average Euclidean distance Δ_i of K nearest neighbours from each observation in the training set, where $K = \sqrt{l}$,

$$\Delta_i = \frac{1}{K} \sum_{j \in \mathcal{D}} \|\mathbf{x}_i, \mathbf{x}_j\|, \quad \forall i = 1 \ldots l \tag{8}$$

where $\mathcal{D}$ is the set of K nearest neighbours for $\mathbf{x}_i$.

2) Next, the global average distance Δ_G is found by averaging Δ_i over all the training data, $\Delta_G = l^{-1} \sum_i \Delta_i$. The value of Δ_G provides a guide for the range of σ, where we define $\sigma = \kappa \times \Delta_G$, where κ is a linking constant between the value of σ and the global average distance Δ_G of any dataset. Therefore, κ provides a guide for the appropriate value of σ, which is independent of l. Once an appropriate value of κ is chosen for one dataset, it provides a good starting point for another dataset with similar dynamics (e.g., for another patient vital-sign dataset, allowing the value of κ to be reused from previous analyses, when the dataset has changed).

The other parameter to optimise in a one-class SVM is ν, which will defined as follows. The support vector constraints [9] in terms of the penalty parameter C from (3) are

$$\sum_i \alpha_i = 1, \quad 0 \leq \alpha_i \leq C. \tag{9}$$

allowing us to state[2] that $1/l \leq C \leq 1$. Also, C may be written

$$C = \frac{1}{\nu l} \tag{10}$$

so we have $1/l \leq \nu \leq 1$. Therefore, ν and C take values in the same range.

The parameter ν serves as an upper bound on the proportion of training observations that lie on the wrong side of the hyperplane, and is also a lower bound on the fraction of support vectors among normal training data [8]; i.e. $\nu \leq N_s/l$. Parameter ν is used in this investigation instead of C, due to its clear meaning, as described above; the value of C can be easily recovered using (10).

Based on the above discussion, we will optimise parameters (κ, ν) in section IV-B for the SVM approach taken in this paper.

B. Parameter optimisation using partial AUC

Combining the suggestions of the previous two sections, we take the following approach to optimising (κ, ν).

STEP 1: Choose a pair of parameter values (κ, ν).

STEP 2: Use the chosen (κ, ν) to train a one-class SVM, which is dependent on a training set of "normal" data.

STEP 3: Use the resulting SVM to classify a validation dataset, which comprises both "normal" and "abnormal" data in equal quantity.

STEP 4: Compute partial AUC, using the validation results obtained in the previous step.

STEP 5: Repeat **STEP 1-4** using different values of (κ, ν), typically using a grid search. Choose the (κ, ν) with the maximum partial AUC.

We assume the presence of *some* examples of "abnormal" behaviour, which are placed within the validation set for the purposes of parameter optimisation. However, as noted previously, these are likely to be small in quantity compared with the number of "normal" observations, and hence the training set is entirely comprised of "normal" data, and a one-class approach is taken.

A commonly-employed alternative which uses only "normal" data [12], [18] is to vary the SVM parameters until some fixed value of the false-positive classification rate is achieved (e.g., 0.05) when presented with the training set of "normal" examples. However, as demonstrated in [4], the overall expected performance of the one-class SVM can be improved by setting parameters by taking into account any available examples of "abnormal" data that may be available, even if they are few in comparison to the number of "normal"

[2]where the lower constraint arises because, in the worst case, we have all training data as support vectors and $N_s = l$, and therefore $C \geq 1/l$ in order for $\sum_i \alpha_i = 1$. The upper constraint arises because $\alpha_i \leq C$.

training data. Therefore, we adopt the latter approach in our formulation, and include any available "abnormal" data in our validation set, described in the algorithm above.

V. Illustration Using Synthetic Data

The optimisation method described in the previous section is illustrated in this section using bivariate artificial data. We compare results obtained with the SVM to a commonly-used generative method of novelty, that of the Gaussian mixture model (GMM). The latter technique has been used for novelty detection in many applications, the details of which are surveyed in [7]. It is of particular interest to the investigation described by this paper, because previous work in performing novelty detection in patient vital-signs has used a generative approach, comprising a mixture of Gaussian kernels [19], [20].

A. Methodology

The dataset employed in this section comprises 200 "normal" and 200 "abnormal" data, with distributions that are concave in support, as shown in figure 1. For the purposes of evaluating novelty detection algorithms, 50% of both datasets were held back for testing. 40% of the "normal" data are used for training, with the remaining 10%, and all remaining "abnormal" data, used for validation.

For the SVM, the training and validation procedures (in which parameter optimisation occurs) proceed as described in section IV-B.

The GMM is defined by the pdf

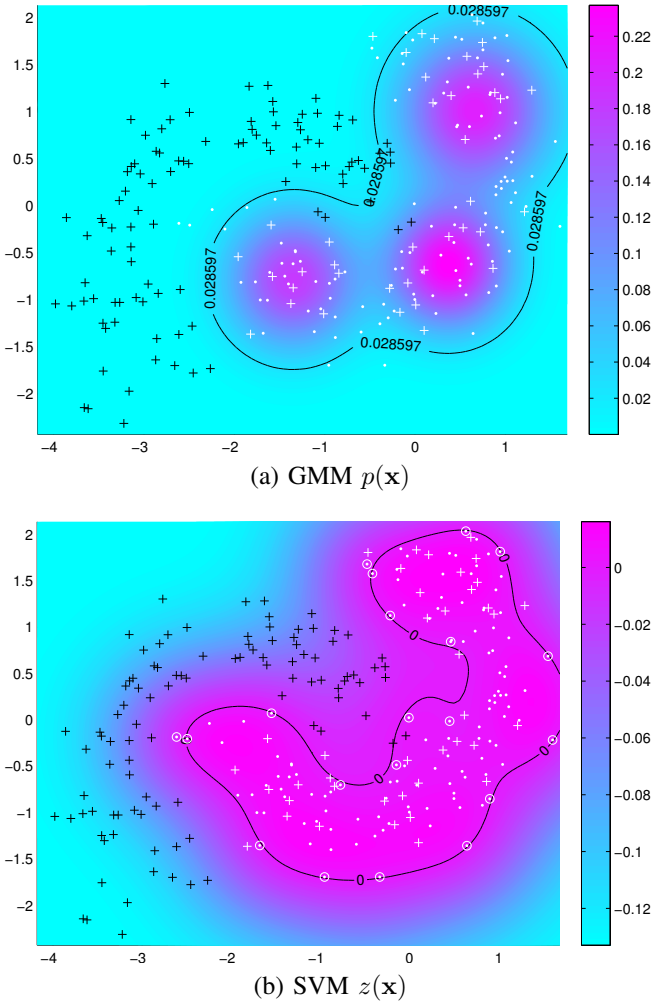
$$p(\mathbf{x}) = \sum_{i=1}^{M} \pi_i \; p(\mathbf{x}|\boldsymbol{\theta}_i) \tag{11}$$

which is comprised of M component distributions, each of which has a prior probability π_i and a likelihood $p(\mathbf{x}|\boldsymbol{\theta}_i) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, where $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ have their usual meanings of the centre and covariance matrix for multivariate Gaussian i, respectively. The training procedure finds appropriate values for the quantities shown in (11); in this case, the maximum likelihood estimates were determined using expectation maximisation [21].

Novelty detection with a GMM is typically performed by setting a decision threshold on the pdf $p(\mathbf{x})$, which, as with the decision boundary of the SVM, may be set (i) using only "normal" data, whereby one finds that threshold that yields some pre-determined false-positive rate, or (ii) using a validation set that contains any available "abnormal" examples. In order to allow a direct comparison with the SVM, the latter approach was taken for setting the GMM decision threshold, and partial AUC was again used as the optimisation metric, the minimum of which (with respect to the validation set) yields the "optimal" value of the decision threshold on the pdf $p(\mathbf{x})$.

B. Results

Figure 1 shows the output of both GMM and SVM novelty detectors when presented with the previously-unseen test dataset. The upper plot shows $p(\mathbf{x})$ for the GMM, where the decision threshold on the pdf is shown as a black contour



(a) GMM $p(\mathbf{x})$

(b) SVM $z(\mathbf{x})$

Fig. 1. (a) GMM output $p(\mathbf{x})$ using the "optimal" parameter values determined using validation; "normal" training data are shown by white $\{\cdot\}$, and "normal" and "abnormal" test data are shown by white $\{+\}$ and black $\{+\}$, respectively. (b) SVM output $z(\mathbf{x})$ using the "optimal" parameter values determined using validation; labelling scheme as above, with support vectors circled.

on the pdf, which describes the locus of the "normal" training data (shown as white dots). The test data are shown as crosses in white and black for "normal" and "abnormal" classes, respectively. It may be seen that the training and validation procedure resulted in $M = 3$ component distributions being used, where these distributions were constrained to have isotropic covariance matrices (although each may be assigned a different determinant during training).

The lower plot shows $z(\mathbf{x})$ for the SVM, as defined in (5), where the decision threshold (which occurs at $z(\mathbf{x}) = 0$ for a SVM) is shown as a black contour. The symbols used are the same as those for the GMM plot, described above, where training examples that are support vectors ($\alpha_i > 0$) are circled in white.

Table I shows results obtained from both GMM and SVM when applied to the previously-unseen test data. Defining true-

TABLE I
NOVELTY DETECTION PERFORMANCE OF GMM AND SVM APPLIED TO TEST DATA, AT "OPTIMAL" THRESHOLD. ONE STANDARD DEVIATION ON THE RESULT IS SHOWN FOR EACH, USING THE RESULTS OF 50 EXPERIMENTS, IN WHICH EACH EXPERIMENT INVOLVES RANDOM SELECTION OF NEW TEST DATA FROM THOSE AVAILABLE, WITH EQUAL NUMBERS OF TEST DATA DRAWN FROM THE "NORMAL" AND "ABNORMAL" CLASSES.

Classifier	Accuracy	Partial AUC	Sensitivity	Specificity
GMM	0.95 ± 0.01	0.28 ± 0.01	0.95 ± 0.01	0.93 ± 0.03
SVM	0.96 ± 0.01	0.30 ± 0.01	0.99 ± 0.01	0.88 ± 0.01

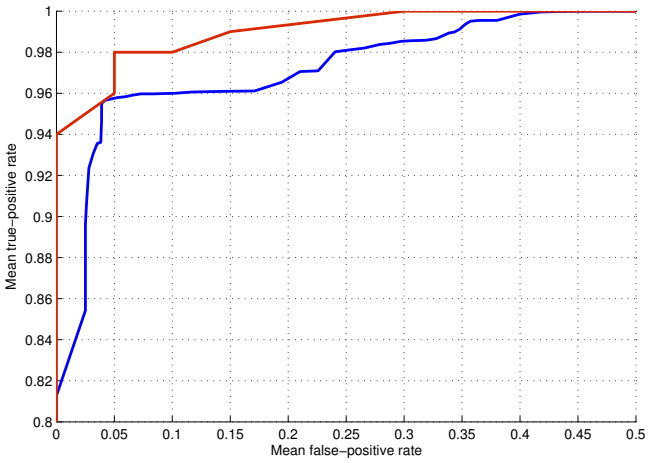


Fig. 2. ROC curve for results obtained using GMM and SVM to classify previously-unseen artificial test data, shown in blue and red, respectively. The mean of 50 experiments has been shown at each point on the ROC curve, where each experiment involves random selection of new test data from those available, with equal numbers of test data drawn from the "normal" and "abnormal" classes.

positive, true-negative, false-positive, and false-negative to be TP, TN, FP, and FN, respectively, then *accuracy* is defined to be (TP + TN) / (TP + TN + FP + FN), *sensitivity* is TP / (TP + FN), and *specificity* is TN / (TN + FP). It may be seen that both methods perform similarly with this simple bivariate example, with the SVM performing marginally better overall, as shown by slightly higher accuracy and AUC results. It achieves this with a higher sensitivity, at its "optimal" threshold, at the cost of a lower specificity. This is confirmed in figure 2, which shows an overall higher ROC curve for the SVM than the GMM.

We conclude that the training and optimising procedures for both techniques result in stable, usable parameter configurations, and hence we now extend our analysis to consider higher-dimensional patient vital-sign data, as acquired from a recent clinical study.

VI. PATIENT VITAL-SIGN MONITORING

This section reports results obtained from evaluating both GMM and the proposed SVM method for patient vital-sign monitoring.

TABLE II
NUMBER OF CLINICAL OBSERVATIONS IN THE TRAINING, VALIDATION, AND TEST SETS

	Train	Validate	Test
Normal	1,240	65	65
Abnormal	0	65	65

A. Dataset

We consider vital-sign data acquired from patients in a "step-down unit" (SDU), which is a level of acuity lower than that of the intensive care unit (ICU). There is a significant need for effective novelty detection systems in such wards, because patient deterioration can go unnoticed by clinical staff, leading to adverse patient outcomes. Existing patient monitors generated univariate alarms whenever vital signs exceed some pre-defined threshold, and often go unheeded due to the high false-positive rate of such alarms, where [22] reported results of a study in which it was deemed that 84% of alarms from conventional continuous patient monitors were false alarms.

The dataset used for the work described by this section comprises measurements of heart rate, respiratory rate, blood oxygen saturation, and systolic blood pressure, acquired once every four hours by ward staff (as is common practice in most SDU-level wards in the UK and the US) at the Oxford Cancer Hospital, Oxford, UK. 1,500 such clinical observations $\mathbf{x}_i \in \mathbb{R}^4$ were acquired from 19 patients, who were recovering from upper gastro-intestinal surgery.

B. Methodology

130 of the clinical observations were deemed by clinicians to be sufficiently "abnormal" that the patient would require clinical review. The remaining 1,370 were thus classified as being "normal". The available "abnormal" data are insufficient to train a multi-class classifier, being small in comparison with the number of "normal" data, and therefore the novelty detection approach is justified for this clinical application.

Table II shows how the 1,500 observations may be assigned to each of the training, validation, and test sets. The available examples of abnormality must be split between the validation set (to enable parameter optimisation, as described in section IV-B) and the test set (to allow evaluation of the results); therefore, the 130 examples of abnormality are split equally between validation and test data. Similar numbers of "normal" data are required for each of the validation and test sets; the remainder of the "normal" data are placed into the training set, as shown in the table.

The split between the training, validation, and test sets was performed randomly. In order to test the variability of the results to this random partitioning, 50 experiments were performed, each experiment containing a different random partition of the data between the three sets, and each experiment requiring retraining, revalidation, and retesting, in order to obtain fully independent results for each experiment.

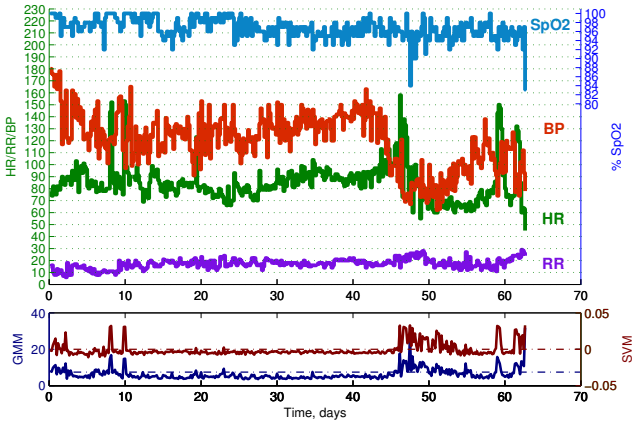


Fig. 3. The upper plot shows time-series of vital signs for an exemplar patient, showing HR, BR, SpO_2, and BP in green, purple, blue, and red, respectively, with time (in hours) shown on the horizontal axis. The lower plot shows novelty scores derived from GMM output $-\log p(\mathbf{x})$ and SVM output $z(\mathbf{x})$ on the same time-base, in blue and red, respectively. Horizontal lines in the lower plot show the decision thresholds for the GMM and SVM in blue and red, respectively.

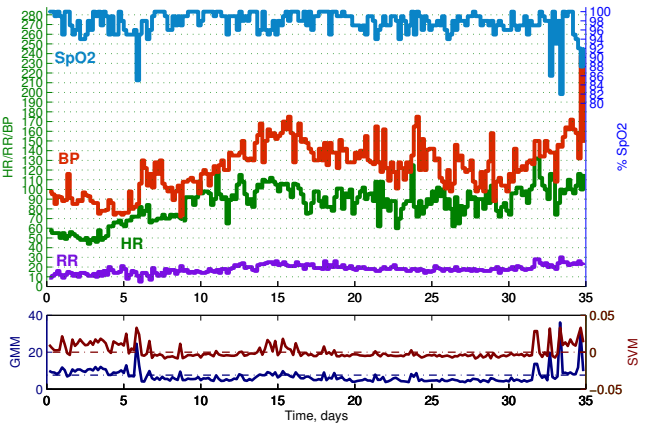


Fig. 4. The upper plot shows time-series of vital signs for a second exemplar patient, showing vital signs and novelty detection output as in the first example.

TABLE III
NOVELTY DETECTION PERFORMANCE OF GMM AND SVM APPLIED TO TEST CLINICAL DATA, AT "OPTIMAL" THRESHOLD. ONE STANDARD DEVIATION ON THE RESULT IS SHOWN FOR EACH.

Classifier	Accuracy	Partial AUC	Sensitivity	Specificity
GMM	0.92 ± 0.03	0.25 ± 0.02	0.92 ± 0.04	0.92 ± 0.04
SVM	0.95 ± 0.01	0.28 ± 0.02	0.98 ± 0.01	0.92 ± 0.03

C. Results

An example of the application of the techniques to patient vital-sign data is shown in figures 3 and 4.

The first example shows a patient who enters the ward, following surgery, in a state of physiological derangement, as shown by elevated BP (at around 180 mmHg). The patient begins to stabilise, but, after 10 days, episodes of tachycardia (elevated HR, reaching 150 bpm) may be seen, with a corresponding increase in RR from 10 breaths/min to 20 breaths/min. The patient then stabilises again, but deteriorates significantly after 45 days, showing periods of prolonged hypotension (decreases in BP below 60 mmHg), with corresponding tachycardia (HR reaching 160 bpm), and desaturations (SpO_2 decreasing to 84%).

The output of the GMM and SVM are shown beneath, along with their decision thresholds. For the purposes of visualisation, the output of the GMM, which is a density $p(\mathbf{x})$, has been scaled into a novelty score $-\log p(\mathbf{x})$ such that it takes high values for data with low density; i.e., "abnormal" data will take high novelty scores. The SVM output needs no such transformation, as it takes positive values when data on the "non-normal" side of the decision boundary are presented.

It may be seen from the figure that both the SVM and the GMM increase in value during the initial post-surgical period of abnormality, during the transient period around 10 days, and in the final period of deterioration. The similarity of the GMM and SVM output is not accidental, as the $-\log p(\mathbf{x})$ scaling of the GMM output makes it a comparable score to the SVM, because the SVM asymptotically approaches the level sets on the pdf in its tails [23].

The second example shows a patient who is similarly unstable at the start of their admission to the Cancer Hospital ward, following surgery. This patient exhibits bradycardia (low HR, decreasing to 40 bpm). After 5 days, periods of apnoea are evident (low BR, decreasing below 10 breaths/min) with corresponding decreases in SpO_2. Periods of abnormal physiology occur after around 15 days, with transient hypertension (increases in BP over 160 mmHg). Finally, a period of extreme deterioration may be seen at the end of the patient stay, with extreme hypertension (BP exceeding 200 mmHg) and corresponding extreme desaturation (SpO_2 decreasing below 82%).

It may be seen that, again, both the GMM and SVM scores increase above their respective decision thresholds for these periods of abnormality.

Table III shows the overall results for both GMM and SVM after 50 experiments. Unlike the bivariate case considered in section V, there is a significant difference between the results obtained from each method. The SVM achieves higher accuracy and partial AUC, as before, but matches the specificity of the GMM (0.92) while improving on the sensitivity (from 0.92 to 0.98).

This is confirmed by the ROC plots shown in figure 5, in which it may be seen that the ROC curve for the SVM is higher than that for the GMM throughout most of the interval on the horizontal axis.

VII. CONCLUSIONS AND DISCUSSION

We have proposed a method for optimising SVM parameters in a natural manner, considering ν and κ, which, we have argued, have a more intuitive interpretation that the conventional C and σ parameters (assuming that a Gaussian kernel is used).

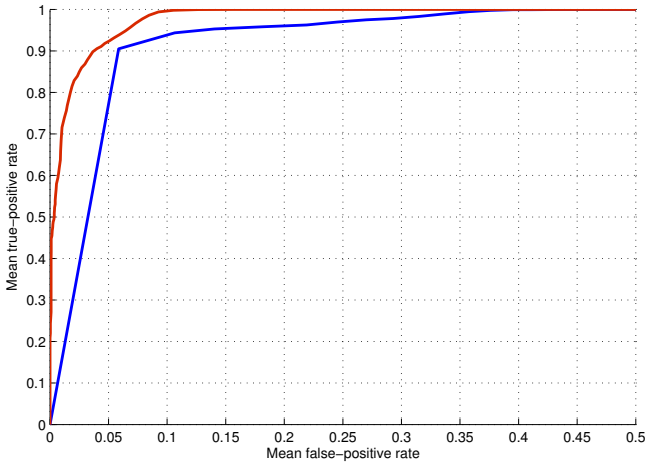


Fig. 5. ROC curve for results obtained using GMM and SVM to classify previously-unseen clinical test data, shown in blue and red, respectively. The mean of 50 experiments has been shown at each point on the ROC curve, where each experiment involves random selection of new test data from those available, with equal numbers of test data drawn from the "normal" and "abnormal" classes.

While we have demonstrated the method for one-class SVMs, they are equally applicable to two- and multi-class SVMS.

Clinical data acquired from a recent observational study of Cancer Hospital patients have been used to demonstrate that automated methods can be used to identify patient deterioration. Existing methods, based on generative mixture distributions have been shown to be outperformed by SVM-based novelty detection on the preliminary data considered so far in this study. This is perhaps unsurprising, given that the SVM minimises its objective function so as to result in a small number of misclassifications in the high-dimensional space of the vital signs. In comparison, generative methods, while offering more functionality than a discriminative method, typically exhibit higher misclassification rates.

The on-going clinical study will result in further data on which to confirm these preliminary findings.

ACKNOWLEDGEMENTS

LC was supported by the Overseas Research Students Award Scheme, provided by the UK Government, and later by the NIHR Biomedical Research Centre Programme, Oxford. DAC was funded by the Centre of Excellence in Personalised Healthcare funded by the Wellcome Trust and EPSRC under grant number WT 088877/Z/09/Z. The authors wish to thank Sarah Vollam and Deborah Evans for the collection of clinical data used in this investigation.

REFERENCES

[1] National Patient Safety Association, "Safer care for acutely ill patients: Learning from serious accidents," NPSA, Tech. Rep., 2007.
[2] National Institute for Clinical Excellence, "Recognition of and response to acute illness in adults in hospital," NICE, Tech. Rep., 2007.
[3] L. Tarassenko, D. Clifton, M. Pinsky, M. Hravnak, J. Woods, and P. Watkinson, "Centile-based early warning scores derived from statistical distributions of vital signs," *Resuscitation*, no. DOI:10.1016/j.resuscitation.2011.03.006, 2011.
[4] P. Hayton, L. Tarassenko, B. Scholkopf, and P. Anuzis, "Support vector novelty detection applied to jet engine vibration spectra," in *Proc. NIPS*, Denver, US, 2000, pp. 946–952.
[5] A. Gretton and F. Desobry, "On-line one-class support vector machines. an application to signal segmentation." in *Proc. IEEE ICASSP*, Hong-Kong, China, 2003.
[6] D. R. Hardoon and L. M. Manevitz, "fMRI analysis via one-class machine learning techniques," in *Proc. 19th International Joint Conference on Aritifical Intelligence (IJCAI)*, Edinburgh, UK, 2005, pp. 1604–1605.
[7] M. Markou and S. Singh, "Novelty detection: A review - part 2: Neural network based approaches," *Signal Processing*, vol. 83, no. 12, pp. 2499–2521, 2003.
[8] B. Scholkopf, J. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Computation*, vol. 13, no. 7, pp. 1443–1471, 2001.
[9] C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol. 2, pp. 121–167, 1998.
[10] D. M. J. Tax and R. P. W. Duin, "Data domain description using support vectors," in *Proc. ESAN99*, Brussels, 1999, pp. 251–256.
[11] D. Tax and R. Duin, "Support vector domain description," *Pattern Recognition Letters*, vol. 20, pp. 1191–1199, 1999.
[12] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*, 1st ed. Cambridge, UK: Cambridge University Press, 2004.
[13] A. R. Webb, *Statistical Pattern Recognition*, 2nd ed. Chichester, England: John Wiley and Sons Ltd., 2002.
[14] A. P. Bradley, "The use of the area under the ROC curve in the evaluation of machine learning algorithms," *Pattern Recognition*, vol. 30, no. 7, pp. 1145–1159, 1997.
[15] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under the receiver operating characteristic (ROC) curve," *Radiology*, vol. 143, no. 1, pp. 29–36, 1982.
[16] S. H. Park, J. M. Goo, and C. H. Jo, "Receiver Operating Characteristic (ROC) Curve: Practical Review for Radiologists," *Korean Journal of Radiology*, vol. 5, no. 1, pp. 11–18, 2004.
[17] C. M. Bishop, "Novelty detection and neural network validation," *Proceedings of IEE Conference on Vision and Image Signal Processing*, vol. 141, no. 4, pp. 217–222, 1994.
[18] B. Scholkopf and A. Smola, *Learning with Kernels*, 1st ed. Cambridge, USA: MIT Press, 2002.
[19] L. Tarassenko, A. Hann, and D. Young, "Integrated monitoring and analysis for early warning of patient deterioration," *British Journal of Anaesthesia*, vol. 98, no. 1, pp. 149–152, 2007.
[20] A. Hann, "Multi-parameter monitoring for early warning of patient deterioration," Ph.D. dissertation, University of Oxford, 2008.
[21] C. M. Bishop, *Pattern Recognition and Machine Learning*. Berlin: Springer-Verlag, 2006.
[22] C. Tsien and J. Fackler, "Poor prognosis for existing monitors in the intensive care unit," *Critical Care Medicine*, vol. 25, no. 4, pp. 614–619, 1997.
[23] R. Vert and J. Vert, "Consistency and convergence rates of one-class svms and related algorithms," *Journal of Machine Learning Research*, vol. 7, pp. 817–854, 2006.