



Published in final edited form as:

IEEE Trans Biomed Eng. 2013 January ; 60(1): 235–239. doi:10.1109/TBME.2012.2210042.

Reducing False Intracranial Pressure Alarms using Morphological Waveform Features

Fabien Scalzo, David Liebeskind, and Xiao Hu

Abstract

False alarms produced by patient monitoring systems in intensive care units (ICU) are a major issue that causes alarm fatigue, waste of human resources, and increased patient risks. While alarms are typically triggered by manually adjusted thresholds, the trend and patterns observed prior to threshold crossing are generally not used by current systems. This study introduces and evaluates a smart alarm detection system for intracranial pressure signal (ICP) that is based on advanced pattern recognition methods. Models are trained in a supervised fashion from a comprehensive dataset of **4791** manually labelled alarm episodes extracted from **108** neurosurgical patients. The comparative analysis provided between spectral regression, kernel spectral regression, and support vector machines indicates the significant improvement of the proposed framework in detecting false ICP alarms in comparison to a threshold-based technique that is conventionally used. Another contribution of this work is to exploit an adaptive discretization to reduce the dimensionality of the input features. The resulting features lead to a decrease of **30%** of false ICP alarms without compromising sensitivity.

Index Terms

patient monitoring; false alarm; intensive care unit; ICP; brain injuries; smart alarm; supervised learning

I. Introduction

Bedside monitors are omnipresent in intensive care units (ICU) of modern hospitals. They allow for monitoring of key clinical variables of the patient and for alarms to be set to trigger sounds when abnormal values are detected to quickly attract the attention of the nurse. Although the use of alarm-based patient monitoring systems is potentially life-saving and improves the efficiency of treatment, it is currently far from being implemented optimally. It has been acknowledged that most of alarms produced by the systems are false [3], [8], [10], [15]. As few as 15% of alarms have been found to be clinically relevant [13]. False alarms are typically generated due to noise and artifacts in the signals or by alarming criteria that are too generic. They not only distract bedside clinicians but also cause alarm fatigue and distrust in the device. These factors may contribute to true alarms to be missed which in turn would place patient safety in jeopardy. There is a clear need to design intelligent monitoring systems that would reduce the number of false alarms using specific patterns observed in continuously recorded signals of the patient.

Most of research efforts to reduce false alarms have focused on signal processing aspects of alarm generation [1], [16]. Recent studies [7], [9] have emphasized the potential interest in using contextual and trending information of clinical variables in identifying critical events. Drawing from these observations, we investigated whether the use of pattern recognition methods based on machine learning models can help reduce the number of false alarms in comparison with a threshold based approach as currently implemented in monitoring systems. In the ICU, patients treated for traumatic brain injuries (TBI) are continuously

monitored with electrocardiogram (ECG), arterial blood pressure (ABP), and intracranial pressure (ICP). Specific thresholds can be set manually to generate different types of alarms associated with various degrees of severity. This study targets alarms related to intracranial hypertension (IH), a known cause of complications for TBI. ICP alarms are typically generated when the ICP is above 20mmHg for a few seconds and therefore does not exploit specific morphological patterns and trends that may occur in the ICP waveform of the patient. It has been shown [12] that changes occur in the morphology of the ICP waveform prior to IH episodes. Existing treatments for IH include drainage of cerebrospinal fluid (CSF), osmotherapy, hyperventilation, sedation, and diuretics. This study provides a comparative analysis between different types of regression models trained from morphological time-series features extracted from the ICP prior to the time of the alarm. Reducing the burden of false ICP alarms using machine learning and feature extraction techniques can be seen as a proof of concept for reducing false alarms triggered by other physiological signals (e.g. ECG).

II. Methods

A. Patient Population and Data Acquisition

The dataset of ICP signals and alarms originates from the University of California, Los Angeles (UCLA) Medical Center and its usage in the present retrospective study was approved by the local institutional review board committee (IRB). This study includes 154 patients admitted for various conditions with known risk of IH. The majority of the patients (108 patients) was treated for brain injuries (TBI, subarachnoid hemorrhage (SAH), and intracerebral hemorrhage (ICH)). A total of 63, 954 ICP alarms were recorded from bedside monitors between 8/2010 and 10/2011 in the 24-bed UCLA NeuroICU. For these patients, the proportion of ICP alarms among all monitor alarms ranges from 0.17% to 88.7% with a median value of 14.4%. The total number of ICP alarms per patient ranges from 2 to 2620 with a median value of 146. The average median of inter-alarm intervals is 37.1 ± 140.7 minutes. ICP signals were recorded continuously at a sampling rate of 240 Hz using ventriculostomy systems.

B. Retrospective Alarm Annotation

One-hour ICP/ECG waveform segments surrounding each of the 63, 954 ICP alarms were retrieved from the database. An expert researcher was presented randomly selected segments of ICP through a dedicated annotation software created in our research laboratory and was asked to label them using the following criterion: an alarm is a false positive if there was no CSF drainage to treat ICP elevation within 15 minutes following the alarm. The expert marked the segment as “noise” if the quality of the ICP recording was not satisfying or if the ICP recording was stopped at any time during the 30 minutes prior to the time of the alarm.

CSF drainage could be visualized by noticing a sudden loss of pulsatile ICP and drop of mean ICP. The expert annotation effort was challenged by total blinding of clinical context of these alarms. In addition, situations were encountered: 1) no CSF drainage following an alarm that was associated with a rising ICP trend or obvious ICP elevation; 2) CSF drainage was activated in response to an alarm associated with no apparent acute increase of ICP prior to the alarm. Those cases were skipped by the expert and were not used in the rest of this study as we are not sure whether they are true or false. Therefore, because of this limitation, the expert was able to annotate 1739 true ICP alarms and 3052 false ICP alarms from 108 patients. Other medical interventions are used at our center to manage ICP but not to treat acute ICP elevation, as considered in this study. For the vast majority of cases, CSF drainage is the first line of defense. True alarms could only have been missed if these

interventions were given without first resorting to CSF drainage. Given the common practice at UCLA, such instances should be rare.

C. Alarm Detection Framework

The proposed alarm detection framework relies on a regression model that is learned in a supervised fashion from a set of labelled training segments of ICP waveforms. Once the model has been trained, it can be used to detect alarms on new patients. The following subsections describe how the raw ICP waveforms are processed to extract morphological features.

1) Analysis of ICP Waveforms—The morphology of the ICP waveform holds essential information about cerebral volume compensatory mechanisms and is related to several cerebrovascular pathophysiologicals. Therefore, characterizing the distribution of ICP waveform features for the segment preceding an ICP alarm may be useful for validating it. Morphological Clustering and Analysis of ICP Pulse (MOCAIP) algorithm [6], [11] is applied to process the 20 min ICP waveform segment extracted prior to each alarm. Tracking is performed in real-time through a probabilistic graphic model that characterizes the interdependence among the position of peaks within a pulse and those between consecutive pulses. Whereas the original algorithm [11] only tracks the latency and elevation of the three peaks, we extend it to track the 24 morphological metrics (Fig. 1) by adding a random variable in the model for each additional metric.

2) Conditional Discretization of Morphological Features (CDF)—In contrast to directly using time-series of MOCAIP metrics as input vectors, we evaluate if the use of a supervised dimensionality reduction algorithm [14] producing features that are invariant to different pace and initial state of ongoing ICP crisis may improve the detection of false alarms. Drawing from the fact that the mean ICP level is one of the influencing factors of the waveform morphology [12], the algorithm independently process each of the 24 metrics m_i , $i \in [1, 24]$ together with the mean ICP ($mICP$) using an adaptive discretization based on their joint occurrence frequency accumulated across the 20-min segment. This leads to 24 subspaces whose dimensions vary and are automatically determined by the algorithm [14]. The concatenation of these subspaces is used as our input feature and termed CDF.

D. Regression-based Detection model

A regression model $y = f(x)$ is used as alarm detection method and maps the input features $x \in X$ extracted from the 20 min segment of ICP prior to the alarm to its label $y \in Y$. A comparison is provided between Spectral Regression (SR-DA) [2], Kernel Spectral Regression (SR-KDA) [2], and Support Vector Machine (SVM) [4]. SR-KDA has been shown in the literature to successfully capture nonlinear relationships in a wide variety of problems. The performance of SR-KDA was at least on par with state-of-the-art techniques such as AdaBoost and decision trees while offering more efficient training of the model. To the best of our knowledge, however, it has never been used in the context of alarm detection.

1) Spectral Regression—SR-DA [2] is a recently proposed method to solve discriminant analysis (DA) as a regularized regression problem,

$$\alpha = \operatorname{argmin}_{\alpha} \sum_{i=1}^n (\alpha^T x_i - y_i)^2 + \delta \|\alpha\|^2 \quad (1)$$

SR-DA formulates the regularized problem as follows,

$$\alpha = (XX^T + \delta I)^{-1} X^T y \quad (2)$$

where I is the identity matrix, α is the eigenvector, and $\delta > 0$ the regularization parameter. Interestingly, this formulation can be solved efficiently using a Cholesky decomposition,

$$r = \text{chol}(XX^T + \delta I) \quad (3)$$

$$\alpha = r \setminus (r^T \setminus (X^T y)). \quad (4)$$

2) Kernel Spectral Regression—SR-KDA [2] generalizes SR-DA to utilize a kernel projection of the data and obtain nonlinearity. Input data samples $x \in X$ are projected onto a high-dimensional space via a Gaussian kernel K , and class labels y to obtain vectors α ,

$$K(i, j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (5)$$

where σ is the standard deviation of the kernel.

Similarly to SR-DA, SR-KDA uses a Cholesky decomposition from the regularized positive definite matrix K and class labels y to obtain vectors α ,

$$r = \text{chol}(K + \delta I) \quad (6)$$

$$\alpha = r \setminus (r^T \setminus y). \quad (7)$$

3) Support Vector Machines—SVM [4] is a supervised machine learning technique that aims at finding the optimal separating hyperplane that minimizes the misclassification rate on the training set, while maximizing the sum of distances of the training samples from this hyperplane.

E. Experimental Setup

The experiments proposed in this study first aim at evaluating if the use of machine learning methods (Section II-D) improves the alarm detection accuracy (and by extension reduces false alarms) in comparison to a threshold-based method. The second purpose is to investigate if the proposed feature encoding via conditional distributions (CDF) improves the performance in comparison with raw morphological metrics extracted from ICP waveforms using MOCAIP tracking.

Using the dataset of 4791 samples (section II-A), a 10-fold cross-validation (CV) is performed to compare three regression methods; SR-DA, SR-KDA, and SVM. The parameters of these three models are optimized at each iteration by using an inner 10-fold CV excluding the patient to be tested at the current iteration, which is a recommended practice to avoid model overfitting and conduct fair comparison. The outer 10-fold CV is executed for five independent runs. For each run, the area under the curve (AUC) is computed from the receiver operating characteristic (ROC) curve. The average AUC and standard deviation are calculated across the five runs and used as measure of performance.

Although differences in AUCs can be used to rank different models, they may not be statistically significant. To verify if the differences between SR-DA, SR-KDA, and SVM and between raw morphological features and CDF features models are statistically significant on our dataset, the 95% confidence interval associated with each AUC is computed using DeLong *et al* [5] method and significance tests are performed using a binomial exact test between AUCs obtained for the different models. For the threshold-based method, the ROC curve is generated by increasing the threshold from 20 to 80mmHg. Therefore, for this model it was not possible to obtain a confidence interval or a standard error since the predicted output is binary and does not change between different runs.

The AUC is a global measure of performance that reflects both sensitivity and specificity. However, a desirable property of an alarm detection system is to offer a very high sensitivity (i.e. true positive rate, $TPR = TP/(TP+FN)$) so that a minimum number of true alarms are missed. To quantify such a property, we compute the false positive rate (i.e. 1-specificity) at three given TPRs, {90%, 95%, 97.5% }.

III. Results

The average and standard deviation of each AUC are reported in the third row of Table I. The AUC of the threshold-based method is 55.9, while the improvement in terms of AUC for morphological metrics and CDF features is respectively $64.1 \pm .9$ to 80.6 ± 2.6 for SVM, 55.2 ± 7.1 to 79.4 ± 3.7 for SR-DA, and 69.2 ± 1.2 to 85.9 ± 1.1 for SR-KDA. The corresponding 95% confidence interval is reported in the fourth row in Table I. Statistical tests indicates that the performance of all the models were significantly different (p -value < 0.01), except for Threshold-based versus SR-DA+metrics, and SR-DA+CDF versus SVM+CDF. The latter indicates that when SR-DA is combined with CDF features it can be considered equivalent to the SVM model. When considering the computational cost in training the models, there is an advantage of using SR-DA over SVM, which is typically more time consuming. Average ROC curves are illustrated in Fig. 2 shows the significant improvement obtained by SR-KDA over SR-DA and threshold, when used with CDF features.

Rows 5, 6, 7 of Table I reports the False Positive Rate (FPR) at for three given True Positive Rates (TPR); 90%, 95%, and 97.5%. In contrast with the average AUCs, Threshold-based method performs better than the regression techniques (SR-DA, SR-KDA, SVM) trained from raw metrics. However, for each TPR, significant improvements are obtained using CDF features as input to the models. A major result of this study is that the SR-KDA reduces the FPR from 66.7 to 31 ± 2.9 at a TPR of 90%, from 69.9 to $42 \pm .8$ at a TPR of 95%, and from 72.5 to $53 \pm .3$ at 97.5%. This means that at a similar TPR, when SR-KDA is combined with CDF features, it can reduce false alarms by about 30%.

IV. Discussion

The experiments have demonstrated that significant improvements in terms of AUC (up to 30%) can be obtained by SR-KDA and SVM models in comparison with a threshold-based approach. Similarly, reduction of false alarms by the best of our models in terms of FPR was 27% at a TPR= 95%. This suggests that specific patterns in the morphology of the ICP waveform may be related to true ICP alarms. In addition, the use of CDF features provides an improvement over the use of raw ICP metrics. Further work is needed to optimize the models so that the number of true alarm missed is minimized. This can be done by defining an objective function that would, for example, maximize the portion of the AUC falling in a given TPR interval, such as $TPR = [90\%, 100\%]$.

It can be hypothesized that some of the good performance in detecting true alarms can be attributed to the fact that the proposed features can capture the known “rounding” of the pulse associated with higher ICP elevation. In the case where an alarm is triggered by an artifactual shift, it is less likely that the pulse did not contain such a change.

While ICP alarms account for about 15% of the total of alarms in the neuro ICU, several other critical types of alarms are generated. The results obtained by our pattern recognition framework are encouraging. A similar approach could be used to identify false alarms in other signals such as ECG and ABP. Although the type of feature will inevitably be different due to the specific nature of the signal and may add an additional layer of complexity, supervised models trained over time-series of those signals seem to be a promising approach as well.

A limitation of this study is that the annotation was only done on a subset of 4791 alarms. Integration of the unlabeled alarms within the model is desirable. SR-DA and SR-KDA can accommodate with such a requirement and be trained with a semi-supervised strategy. We plan to investigate if the models can be trained in an incremental way using active learning strategies that would identify the most relevant episodes to annotate, thus minimizing the time required for annotation. In addition, we will also explore if the use of patient-specific informations can help to further reduce the number of false alarms; it is possible that the clinical context may help to improve the alarm detection accuracy.

V. Conclusion

The trend and morphological properties of the ICP waveform hold essential properties about future elevation of the mean ICP. Current monitoring systems do not exploit this information and are based on a threshold to detect critical values and therefore produce a large percentage of false alarms. This study has introduced a framework to detect false ICP alarms using spectral regression models that learn predictive patterns in the morphology of ICP. Results demonstrate the significant improvement of the model in detecting false ICP alarms and its potential to be applied in bedside monitor.

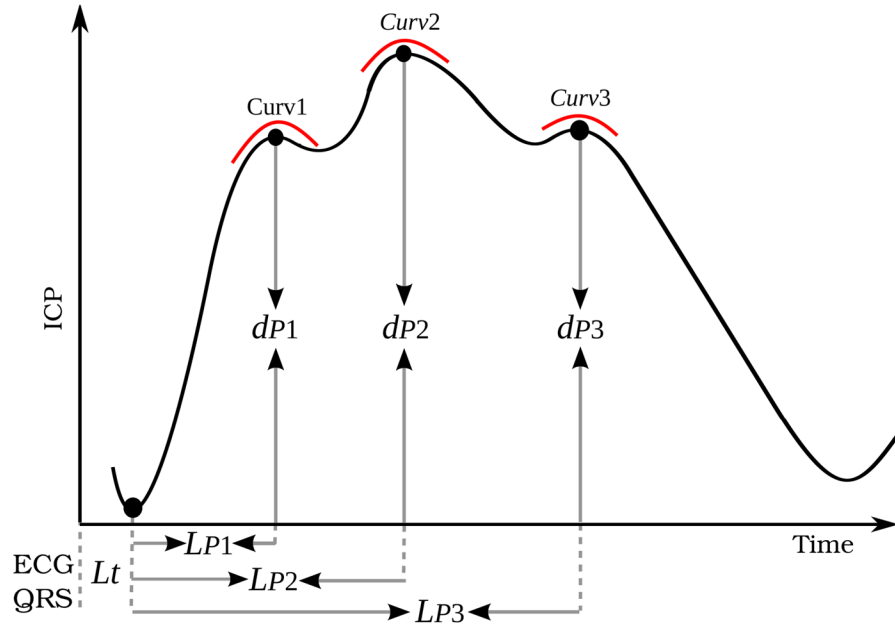
Acknowledgments

The present work is partially supported by NS066008, NS076738, and BIRC to Xiao Hu.

References

1. Aboukhalil A, Nielsen L, Saeed M, Mark RG, Clifford GD. Reducing false alarm rates for critical arrhythmias using the arterial blood pressure waveform. *J Biomed Inform.* Jun.2008 41:442–451. [PubMed: 18440873]
2. Cai, D.; He, X.; Han, J. ICCV. 2007. Spectral Regression for Efficient Regularized Subspace Learning.
3. Chambrin MC, Ravaux P, Calvelo-Aros D, Jaborska A, Chopin C, Boniface B. Multicentric study of monitoring alarms in the adult intensive care unit (ICU): a descriptive analysis. *Intensive Care Med.* Dec.1999 25:1360–1366. [PubMed: 10660842]
4. Chang, C-C.; Lin, C-J. LIBSVM: a library for support vector machines. 2001. <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
5. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics.* Sep.1988 44:837–845. [PubMed: 3203132]
6. Hu X, Xu P, Scalzo F, Vespa P, Bergsneider M. Morphological Clustering and Analysis of Continuous Intracranial Pressure. *IEEE Trans Biomed Eng.* 2009; 56(3):696–705. [PubMed: 19272879]

7. Laramée CB, Lesperance L, Gause D, McLeod K. Intelligent alarm processing into clinical knowledge. *Conf Proc IEEE Eng Med Biol Soc.* 2006:6657–6659. vol. Suppl. [PubMed: 17959478]
8. Lawless ST. Crying wolf: false alarms in a pediatric intensive care unit. *Crit Care Med.* Jun.1994 22:981–985. [PubMed: 8205831]
9. Leite CR, Sizio GR, Neto AD, Valentim RA, Guerreiro AM. A fuzzy model for processing and monitoring vital signs in ICU patients. *Biomed Eng Online.* 2011; 10:68. [PubMed: 21810277]
10. Meredith C, Edworthy J. Are there too many alarms in the intensive care unit? An overview of the problems. *J Adv Nurs.* Jan.1995 21:15–20. [PubMed: 7897067]
11. Scalzo F, Asgari S, Kim S, Bergsneider M, Hu X. Bayesian Tracking of Intracranial Pressure Signal Morphology. *Artif Intell Med.* Feb; 2012 54(2):115–123. [PubMed: 21968205]
12. Scalzo F, Hamilton R, Asgari S, Kim S, Hu X. Intracranial Hypertension Prediction using Extremely Randomized Decision Trees. *Med Eng Phys.* 2012 in press.
13. Siebig S, Kuhls S, Imhoff M, Gather U, Scholmerich J, Wrede CE. Intensive care unit alarms—how many do we need? *Crit Care Med.* Feb.2010 38:451–456. [PubMed: 20016379]
14. Tsai C-J, Lee C-I, Yang W-P. A discretization algorithm based on class-attribute contingency coefficient. *Inf Sci.* Feb; 2008 178(3):714–731.
15. Tsien CL, Fackler JC. Poor prognosis for existing monitors in the intensive care unit. *Crit Care Med.* Apr.1997 25:614–619. [PubMed: 9142025]
16. Zong W, Moody GB, Mark RG. Reduction of false arterial blood pressure alarms using signal quality assessment and relationships between the electrocardiogram and arterial blood pressure. *Med Biol Eng Comput.* Sep.2004 42:698–706. [PubMed: 15503972]



Group		MOCAIP Metrics
Amplitude	Value	$dP_1, dP_2, dP_3, diasICP, mICP$
	Ratio	$dP_2/dP_1, dP_3/dP_1, dP_3/dP_2$
Latency	Value	L_T, L_1, L_2, L_3
	Ratio	$L_2 - L_1, L_3 - L_1, L_3 - L_2$
Curvature	Value	$Curv_1, Curv_2, Curv_3, Curv_m$
	Ratio	$Curv_2/Curv_1, Curv_3/Curv_1, Curv_3/Curv_2$
Slope		$(P_1 - diasICP)/L_1$
Decay		L_x where $dP_x = 0.37dP_3$

Fig. 1. Illustration of the 24 morphological metrics extracted from the configuration of the three peaks of the ICP pulse detected using MOCAIP.

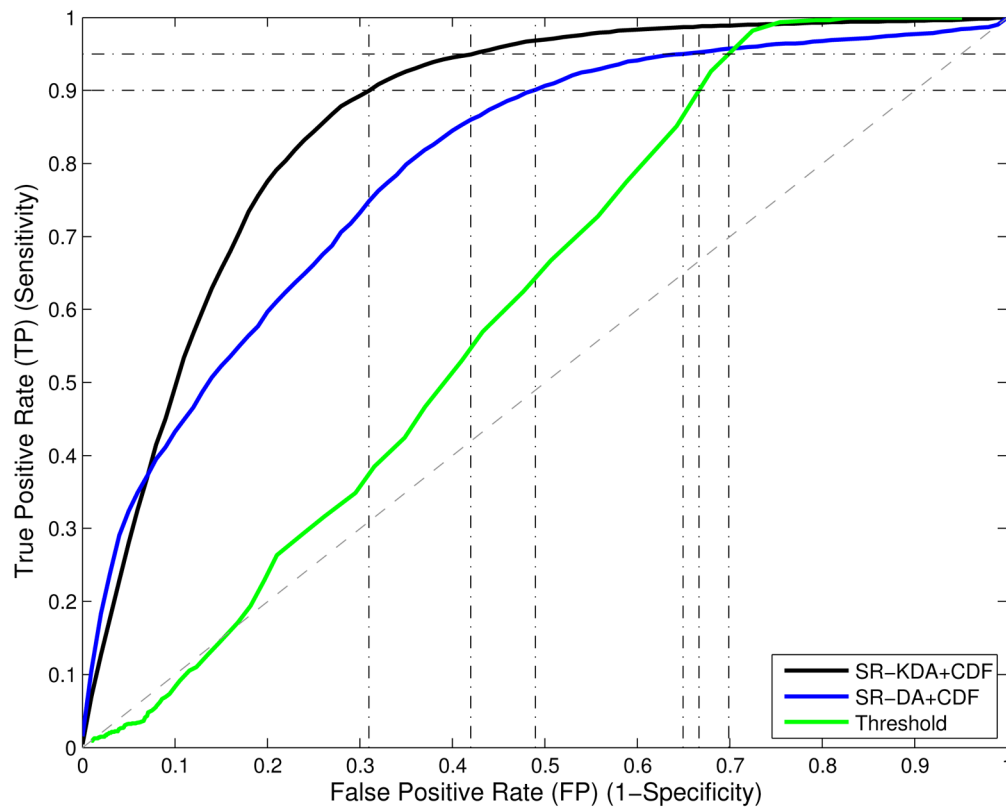


Fig. 2. Average ROC curves computed for SR-KDA, SR-DA, and threshold models. Improvement in terms of AUC between SR-DA (blue) and threshold-based approach (green) is significant (p -value < 0.01). A significant improvement can also be observed between SR-KDA (black) and SR-DA (blue).

\$watermark-text

\$watermark-text

\$watermark-text

TABLE I

Average AUCs, standard deviation, and 95% confidence intervals are reported for each model evaluated using a 10-fold CV. FPR is reported at 3 TPRS; 0.9, 0.95, 0.975.

Feature	Threshold		SR-DA		SVM		SR-KDA	
	mCP		metrics	CDF	metrics	CDF	metrics	CDF
AUC	55.9		55.2±7.1	79.4±3.7	64.1±9	80.6±2.6	69.2±1.2	85.9±1.1
95% CI	-		[53.6, 56.9]	[78.1, 80.7]	[62.5, 65.7]	[79.3, 81.8]	[67.6, 70.7]	[84.8, 86.9]
FPR TPR=9	66.7		82±6.2	49±3.9	75±5	43±3.2	72±1.6	31±2.9
FPR TPR=95	69.9		93±4.5	65±2.3	88±4	66±1.8	82±1.3	42±8
FPR TPR=975	72.5		97±2.1	88±1.6	93±.39	87±1.25	85±.5	53±.3