



Predicting delayed cerebral ischemia after subarachnoid hemorrhage using physiological time series data

Soojin Park¹ · Murad Megjhani¹ · Hans-Peter Frey¹ · Edouard Grave² · Chris Wiggins³ · Kalijah L. Terilli¹ · David J. Roh¹ · Angela Velazquez¹ · Sachin Agarwal¹ · E. Sander Connolly Jr.⁴ · J. Michael Schmidt¹ · Jan Claassen¹ · Noemie Elhadad²

Received: 7 April 2017 / Accepted: 15 March 2018 / Published online: 20 March 2018
© Springer Science+Business Media B.V., part of Springer Nature 2018

Abstract

To develop and validate a prediction model for delayed cerebral ischemia (DCI) after subarachnoid hemorrhage (SAH) using a temporal unsupervised feature engineering approach, demonstrating improved precision over standard features. 488 consecutive SAH admissions from 2006 to 2014 to a tertiary care hospital were included. Models were trained on 80%, while 20% were set aside for validation testing. Baseline information and standard grading scales were evaluated: age, sex, Hunt Hess grade, modified Fisher Scale (mFS), and Glasgow Coma Scale (GCS). An unsupervised approach applying random kernels was used to extract features from physiological time series (systolic and diastolic blood pressure, heart rate, respiratory rate, and oxygen saturation). Classifiers (Partial Least Squares, linear and kernel Support Vector Machines) were trained on feature subsets of the derivation dataset. Models were applied to the validation dataset. The performances of the best classifiers on the validation dataset are reported by feature subset. Standard grading scale (mFS): AUC 0.58. Combined demographics and grading scales: AUC 0.60. Random kernel derived physiologic features: AUC 0.74. Combined baseline and physiologic features with redundant feature reduction: AUC 0.77. Current DCI prediction tools rely on admission imaging and are advantageously simple to employ. However, using an agnostic and computationally inexpensive learning approach for high-frequency physiologic time series data, we demonstrated that our models achieve higher classification accuracy.

Keywords Subarachnoid hemorrhage · Random kernels · Time series · Machine learning · Critical care

1 Introduction

The intensive care unit (ICU), and the Neurologic ICU (NICU) in particular, collects a myriad of data about their patients. Some are physiologic, and some are clinical. In addition, time is of the essence to detect adverse events that arise as secondary complications. In this paper, we focus

on patients with subarachnoid hemorrhage (SAH), one of the most common disease entities treated in the NICU [1, 2]. Our interest is in predicting the secondary complication of delayed cerebral ischemia (DCI) from vasospasm (VSP).

SAH is a devastating illness and a major public health burden, estimated as 14.5 per 100,000 persons in the United States alone [3, 4]. Poor outcomes occur after survival from the initial aneurysm rupture with 15% mortality and 58% functional disability; of which 26% have persistent dependence [5]. Additionally, as many as 20% of patients have global cognitive impairment contributing to poor functional status [6]. Thereby, SAH is associated with a substantial burden on health care resources, most of which are related to long-term care for functional and cognitive disability [7]. Much of the resulting functional and cognitive disability is due to DCI from VSP [7–11]. VSP refers to the narrowing of cerebral blood vessels triggered after a ruptured aneurysm due to the unusual presence of blood surrounding the vessel. It will occur in 30% of SAH patients [12, 13] (up to

✉ Soojin Park
spark@columbia.edu

¹ Department of Neurology, Columbia University, 177 Fort Washington Ave, 8 Milstein – 300 Center, New York, NY, USA

² Department of Biomedical Informatics, Columbia University, New York, NY, USA

³ Department of Applied Physics and Applied Mathematics, Columbia University, New York, NY, USA

⁴ Department of Neurosurgery, Columbia University, New York, NY, USA

54% for SAH patients in coma [14]). In its extreme, severe VSP precludes blood flow to brain tissue, resulting in stroke. DCI is defined as the development of new focal neurological signs or decrease of ≥ 2 points on the Glasgow Coma Scale (GCS), lasting for > 1 h, or the appearance of new infarctions on CT or MRI [15, 16]. The underlying pathophysiology is VSP, and other causes are thus excluded.

For a syndrome with subtle symptoms and time sensitivity, it is helpful to be accurate in prediction so clinicians can remain vigilant for detection. Clinicians use a static instrument called the Modified Fisher Scale (mFS) to predict likelihood of DCI, based on the volume and pattern of blood on the initial brain computed tomography scan (CT) [17–19]. Resource planning and monitoring intensity are scripted around this prediction. Prevention, detection, and management of secondary complications generate a large health care burden for SAH patients [20, 21]. For the higher risk SAH patients, the first 10–14 days are occupied by efforts to detect subtle examination changes that suggest VSP (highest risk: post-bleed days 4–12) [22], and arrange urgent interventions to prevent permanent injury. The only noninvasive tool supported by guidelines to potentially identify asymptomatic VSP is the transcranial Doppler (TCD), which can have poor sensitivity and negative predictive values, and is subject to technician availability and poor interrater reliability [23–29]. Asymptomatic VSP left unchecked may progress to symptomatic VSP, which is dependent on the consciousness of the patient and quality and availability of expertise in the complex and diurnal environment of the NICU. Discharging patients from the ICU at low risk for DCI can result in significant cost savings [30].

Existing predictive models of DCI and VSP after spontaneous SAH are non-dynamic and while they may help risk-stratify patients, they can lack accuracy and precision when applied to individuals. The initial head CT assessment of blood thickness and distribution has spawned 3 grading scales to assess the likelihood of the development of DCI [19], angiographic VSP [18], or symptomatic VSP [31]. DCI is thought to be a more meaningful outcome than symptomatic VSP, especially in patients with severe SAH whose neurologic exam may be limited thus allowing for deterioration to go unrecognized. Such grading scales are performed on static assessments of radiology at admission and are associated with differential odds ratios of outcome [19]. They are not precise predictors of DCI for individual patients. Efforts have been made to improve this early prediction without additional monitoring with moderate results, by combining risk scores [32], incorporating baseline features such as clinical condition and age [33], or assessment of autoregulation [34]. Few efforts have explored time series physiological data for the early prediction of DCI.

In prior proof of concept work (SP) [35], a hypothesis-driven approach to angiographic VSP classification using

24–48 h summary statistics of passively collected electronic health record data (cerebrospinal fluid drainage volume, mean arterial blood pressure, heart rate (HR), intracranial pressure, sodium and glucose) performed with a moderately favorable AUC of 0.71. The raw data used in that study was low frequency (hourly at best) and extracted features summarized over 24 or 48 h. This result was encouraging that EMR and physiologic data could allow risk stratification for future events. The question remains whether increased precision can be achieved with use of higher frequency data.

There is an extensive literature regarding robust feature extraction from physiological time series data for outcome prediction. Approaches can be broadly classified as either hypothesis driven or data driven. Hypothesis driven approaches have focused primarily on temporal data abstraction that relies on knowledge-based symbolic representations of clinical states, either by a priori threshold setting or interval changes [36, 37], summary statistics [35, 38–40], or template matching [41]. Hypothesis driven feature extraction can be effective in prediction but requires domain expertise in designing metafeatures, and may introduce a bias [38]. Data driven or learning approaches extract meaningful features directly from the labeled data without a priori hypothesis [42–51]. The data-driven approach of featurization via random kernels [52, 53] has shown promise in the field of image classification [54]. Random kernels, when convolved with unknown images, extract features that are frequency selective and translation invariant, characteristics that are also desired when processing temporal physiologic data. In our approach, we apply random kernels to extract features from high frequency temporal physiologic data that maximally classify for DCI.

2 Patients and methods

2.1 Study population

Consecutive patients with spontaneous SAH admitted to the Columbia University Medical Center NICU between August 1996 and December 2014 were prospectively enrolled in an observational cohort study of SAH patients designed to identify novel risk factors for secondary injury and poor outcome. The study was approved by the medical center Institutional Review Board. In all cases, written informed consent was obtained from the patient or a surrogate. SAH secondary to perimesencephalic bleeds, trauma, AVM, and patients < 18 years old were not enrolled in the study. Starting in 2006, physiologic data was acquired using a high-resolution acquisition system (BedmasterEX; Excel Medical Electronics Inc, Jupiter, FL, USA) from General Electric Solar 8000i monitors (Port Washington,

NY, USA; 2006–2013) or Philips Intellivue MP70 monitors (Amsterdam, The Netherlands; 2013–2014) at 0.2 Hz.

Exclusion criteria for this project were the following: (1) absence of physiologic monitoring data (before 2006), (2) VSP or DCI before post bleed day (PBD) 3, and (3) patients missing all candidate features. The targeted classification outcome was DCI, defined as development of new focal neurologic signs or deterioration of consciousness for > 1 h or appearance of new infarctions on imaging due to VSP [16]. This was adjudicated by consensus among the treating neurointensivists during a weekly meeting, as part of the observational cohort study.

2.2 Data analysis

Data analysis, and model building were performed using custom software developed in Matlab 2016a (Mathworks, Natick, MA) and Python (<http://www.python.org>). Figure 1 shows a flowchart of the data processing and analysis. The input to the model was physiologic data sampled at 0.2 Hz limited to the PBD 0–3. The target classification outcome was DCI (beyond PBD 3).

2.2.1 Baseline candidate features and outcomes

The following baseline characteristics, grading scales, and outcomes were prospectively recorded at admission: age, sex, worst Hunt-Hess grade in first 24 h (HH), mFS, admission GCS, length of stay, timing of DCI, mortality, and Modified Rankin Scale (MRS). HH grade was dichotomized into low grade (1–3) and high grade (4–5). MFS was dichotomized into low grade (0–2) and high grade (3–4). MRS was dichotomized into good outcome (0–3) and poor outcome (4–6). Baseline features and outcomes were compared for patients with DCI versus no DCI. Baseline features were also compared for the derivation versus validation dataset.

Frequency comparisons for categorical variables were performed by Fisher exact test. Two-group comparisons of continuous variables were performed with the Mann–Whitney *U* test. All statistical tests were two-tailed and a *p*-value < 0.05 was considered statistically significant.

2.2.2 Physiologic feature extraction using random kernels

While 0.2 Hz physiological data was available, we remained agnostic about the optimal scale or sampling rate for DCI

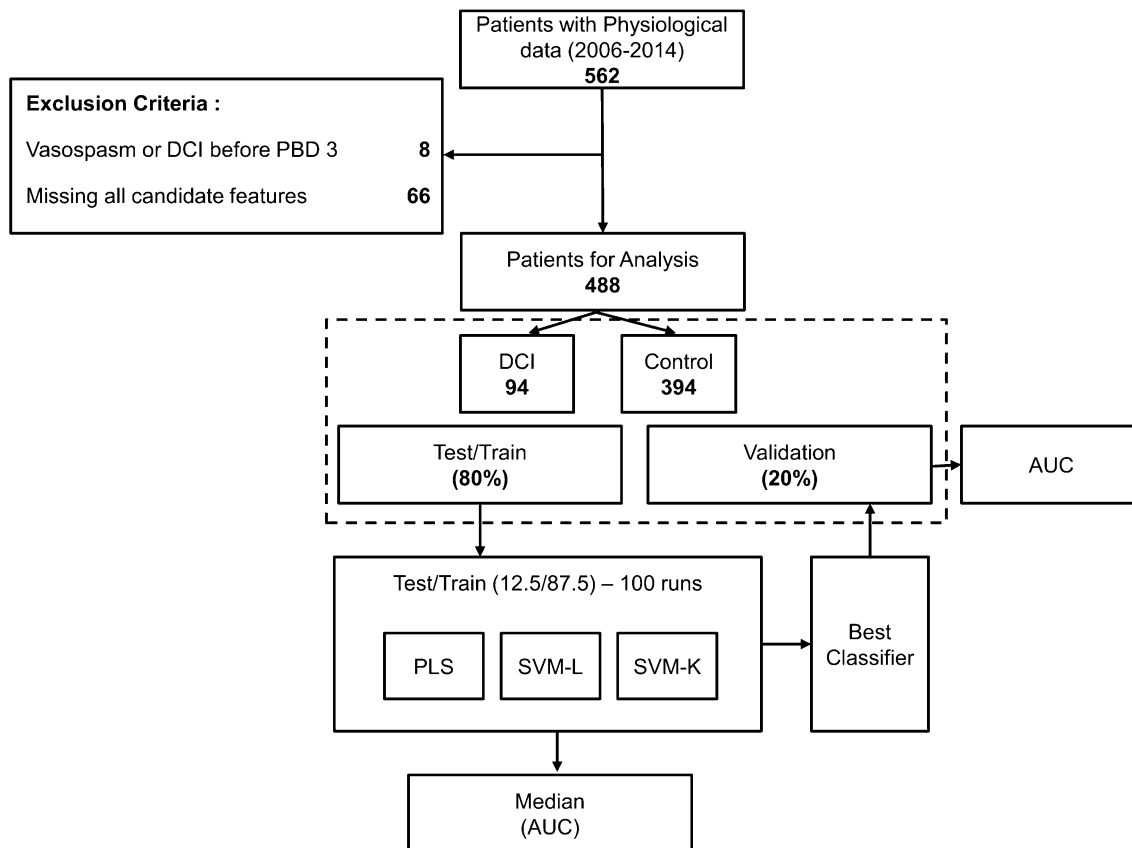


Fig. 1 Overview of the approach

classification. Five universally available ICU variables [HR, respiratory rate (RR), systolic blood pressure (SBP), diastolic blood pressure (DBP), and oxygen saturation (O2)] were downsampled (ds) from 0.2 Hz to 1, 5, 10, 20, 60 min, 2, and 4 h. Downsampling was computed as medians, which dealt with erroneous data [55]. Data was truncated to PBD 0–3 and zero-padding was performed for missing data as a pre-processing step. We selected random kernels [52–54] to be applied and assessed for maximal convolution as shown in Fig. 2. In particular, we first generated random kernels or filter h of size k by sampling values from the normal distribution $N(0,1)$. Next, given a time series $X \in R^{1 \times t}$, convolving filter $h \in R^{1 \times k}$, where $k \leq t$, the feature f_i for series X and filter h is given by $\max(f * h)$, where $*$ denotes the valid convolution. Valid convolution means that f is applied only at each position of x such that f lies within x . In other words, we performed convolutions only when the contiguous data length was twice the length of kernel.

We selected 20 random kernels for each of five variables (var: HR, RR, SBP, DBP, O2), for each varying kernel length (kl; 2, 5, 10, 20) and for 5 s data and each downsampling period (ds; 1, 5, 10, 20, 60, 120, 240 min). For 5 s data, we used larger size kernel lengths (kl; 20, 60, 120, 180) with the assumption that smaller kernel lengths (10–100 s total range) would result in clinically irrelevant features which

would cause the models to over-fit the dataset. This resulted in 3200 candidate random kernel derived physiological features. As the number of kernels increases, the computational complexity to generate the random feature increases. Increasing the number of kernels per ds/kl/var combination beyond 20 did not affect our analysis, and thus was chosen to maximize performance while minimizing computational complexity.

Physiologic data was limited to the first 4 days after aneurysm rupture to limit the influence of clinical treatment in response to suspected VSP or DCI [22].

2.2.3 Feature selection and model building

Minimal Redundancy Maximal Relevance (mRMR) [56–58] was applied to identify the most relevant features for classification. MRMR selects the features that maximize the mutual information *between* features and target class, and minimizes mutual information *among* the features. The features are ranked based on the greedy search that maximizes the Mutual Information Difference Criterion (MID) or Mutual Information Quotient Criterion (mRMR-Q). Let $S \in \{x_1, \dots, x_n\}$ be the set of features and h be the target class (in our case DCI vs. non DCI) then the features are ranked as

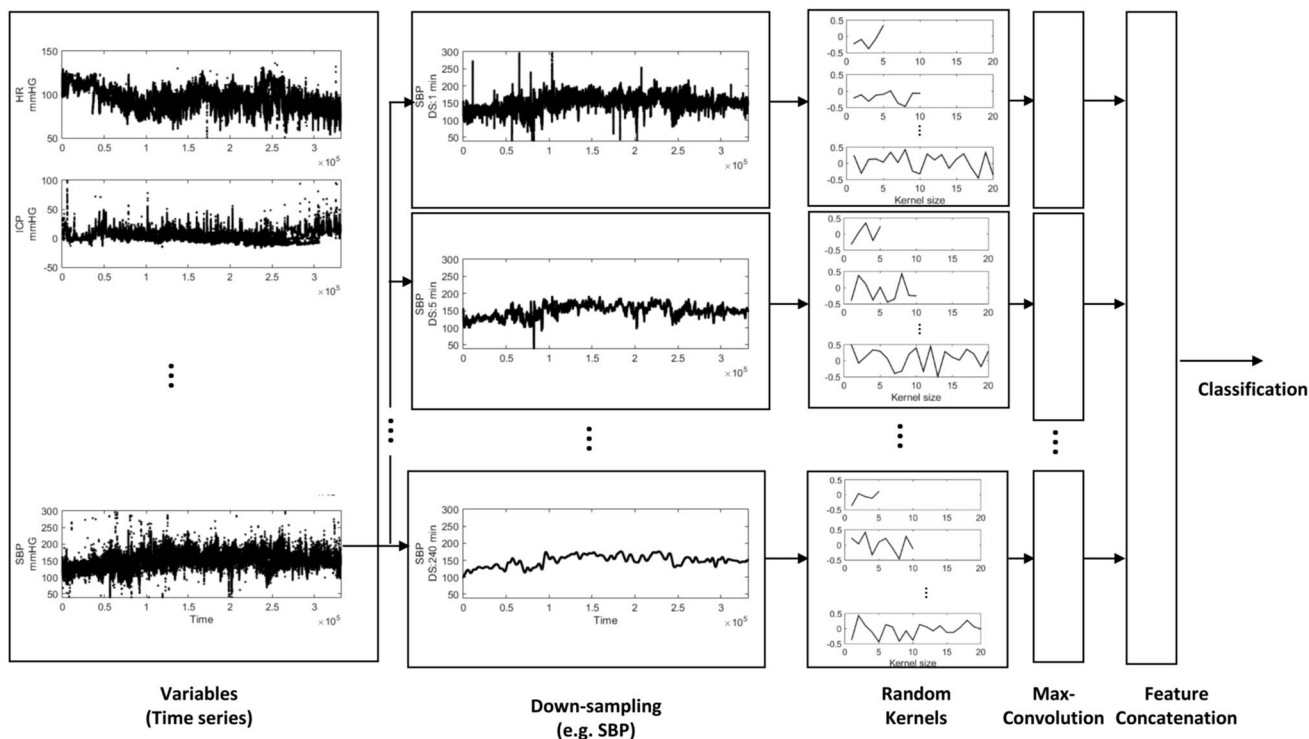


Fig. 2 Feature extraction from physiologic time series data. 20 randomly chosen kernels were applied and assessed for maximal convolution, for each varying kernel length (kl; 2, 5, 10, 20) and for each

downsampling period (ds; 1, 5, 10, 20, 60, 120, 240) and for each of five variables (var; HR, RR, SBP, DBP, O2). This resulted in a convolution matrix of 2800 candidate features

$$\text{MID} : \max_{x_i \in S} \left[I(x_i, h) - \frac{1}{|S|} \sum_{x_j \in S} I(x_i, x_j) \right],$$

$$\text{MIQ} : \max_{x_i \in S} \left[\frac{I(x_i, h)}{\frac{1}{|S|} \sum_{x_j \in S} I(x_i, x_j)} \right],$$

where $I(x_i, h)$ is the information gain between the feature x_i and target class h . The first ‘k’ ranked features are then used to learn the classifier. This simplifies the model, reduces training times, and enhances the generalizability of the classification model.

We used mRMR in combination with linear and kernel based Support Vector Machines (SVM-L and SMV-K) classifiers [59, 60], as well as Partial Least Squares (PLS) regression [61] for combined feature selection and classification. The mRMR feature selection criteria identified the top 1600 (50%) features from 3200 physiologic and 5 baseline demographics/scales. These features were then used by the classifiers to learn the model. PLS regression performs a principal component analysis on all feature vectors first and then applies a least squares regression using those components that explain the most variance. Weighted SVM was utilized to account for the imbalance in classification categories (i.e. fewer DCI vs. non-DCI in any consecutive SAH dataset).

All classifiers were trained for binary outcome of DCI presence/absence. We trained our models on single variables as well as collections of variables, to identify the optimal combination of features that was most informative (i.e. maximally performing) for classifying DCI.

2.2.3.1 Weighted support Vector Machines and imbalanced class sizes Given a training set $\{\{x_1, y_1\}, \dots, \{x_N, y_N\}\}$, where $x_i \in R^N$ and $y_i \in \{-1, 1\}$, then the SVM problem can be formulated as

$$\min_{w, \epsilon, b} \left\{ \frac{1}{2} \|w\|^2 + \delta_{ic} * C \sum_{i=1}^n \epsilon_i \right\}$$

subject to

$$y_i(w \cdot x_i - b) \geq 1 - \epsilon_i; \epsilon_i \geq 0$$

where $w \cdot x_i - b$ is the hyperplane that separates the two classes, and $\epsilon_i = \max(0, 1 - y_i(w \cdot x_i - b))$ is the slack variable (a means for relaxing the constraint by considering points for which our constraint can fail). C is the trade-off parameter that controls the slack variable; a small C allows constraints to be easily ignored and a large C makes constraints hard to ignore. For traditional SVM, $\delta_{ic} = 1$,

meaning the optimization function penalizes equally for all data points. For cases with imbalanced class sizes, this biases the classifier in favor of the class with a larger sample size. To overcome that issue we have added a penalty term δ_{ic} , that is set based on the total number of samples for a given class, setting higher values for the class with smaller size. We used LibSVM [62] for the SVM and weighted SVM classification.

2.3 Internal validation and validation strategy

The cohort was randomly split 80/20%, while maintaining proportional targeted outcome (DCI). 80% were used to train/test models, and considered the primary derivation dataset. For internal validation of our models, we performed cross-validation of the derivation data with a 12.5% hold-out set; the hold-out set was proportional to the training data set for percentage of targeted outcome. The discriminative performance is described by an area under the receiver operating characteristic curve (AUC). The median value of AUC is reported, over 100 runs. 20% of the cohort were not involved in model training, and used exclusively for testing the classification accuracy of our models. Classification accuracy of our models on the validation test set is reported as AUC, with 95% confidence intervals (CI). An overview of the analytical approach is illustrated in Fig. 1.

3 Results

From August 1996 to December 2014, 1595 SAH patients were enrolled in SHOP. 562 SAH patients with physiologic data were available from May 2006 to December 2014. 8 had VSP or DCI identified before PBD 3, 66 were missing all candidate features leaving a total of 488 subjects included into the study (Fig. 1).

Table 1 displays the baseline features, grading scales, and outcomes of subjects with and without DCI. DCI was found in 94 subjects (19.3%) in the entire cohort; 75 (19.2%) of the derivation set and 19 (19.4%) of the test set. Patients with DCI were more frequently women (82 vs. 65% $p=0.001$). None of the grading scales were found to be significantly different between these two groups (HH $p=0.27$; MFS $p=0.69$; GCS $p=0.16$). Length of stay was significantly longer in the DCI group (18.8 ± 6.6 vs. 9.4 ± 7.1 days, $p < 0.0001$). When DCI occurred, it occurred on day 7.1 (± 2.6 days). No DCI was associated with significantly more mortality (19.8 vs. 8.5%, $p=0.0098$); notably 76.9% of mortality in the no DCI group occurred before mean day of DCI onset (7.1). MRS at 3 months favored good outcome for patients without DCI ($p=0.0236$).

Table 1 Baseline features, grading scales, and outcomes of subjects with and without DCI

Baseline Features	DCI (n=94)	No DCI (n=394)	<i>p</i> value ^a
Age (years) (mean ± SD)	52.5 ± 12.7	55.8 ± 15.1	0.018
Sex, female (%)	77 (81.9)	255 (64.7)	0.001
Hunt Hess grade (%)			0.27
1–3	57 (60.6)	263 (66.8)	
4–5	37 (39.4)	131 (33.3)	
Modified Fisher Scale (%)			0.69
0–2	71 (75.5)	287 (72.8)	
3–4	23 (24.5)	107 (27.2)	
Glasgow Coma Scale n (%)			0.16
3–7	28 (29.8)	122 (31.0)	
8–12	18 (19.2)	43 (10.9)	
13–15	48 (51.1)	229 (58.1)	
Length of stay (days) (mean ± SD)	18.8 ± 6.6	9.4 ± 7.1	< 0.0001
Timing of DCI (days) (mean ± SD)	7.1 ± 2.6	NA	
Mortality, n (%)	8 (8.5)	78 (19.8)	0.0098
Modified Rankin Scale at 3-months (%)	n=56 (57.2)	n=232 (58.2)	0.0236
0–3	41 (73.2)	201 (86.6)	
4–6	15 (26.8)	31 (13.4)	

DCI delayed cerebral ischemia, SD standard deviation

^a*p* values were calculated using Fisher exact test for frequency comparisons of categorical variables and Mann–Whitney *U* test for two-group comparisons of continuous variables. All statistical tests were two-tailed

Table 2 Baseline features, grading scales, and outcomes of the derivation and validation groups

	Derivation dataset (n=390)	Validation dataset (n=98)	<i>p</i> value ^a
DCI (%)	75 (19.2)	19 (19.4)	1.0
Age, years (mean ± SD)	55.6 ± 15.1	53.5 ± 13.2	0.19
Sex, female (%)	264 (67.7)	68 (69.4)	0.81
Hunt Hess Grade (%)			0.72
1–3	254 (65.1)	66 (67.4)	
4–5	136 (34.9)	32 (32.7)	
Modified Fisher Scale (%)			0.20
0–2	281 (72.1)	77 (78.6)	
3–4	109 (28.0)	21 (21.4)	
Glasgow Coma Scale (%)			0.20
3–7	128 (32.8)	22 (22.5)	
8–12	45 (11.5)	16 (16.3)	
13–15	217 (55.6)	60 (61.2)	
Length of stay (days) (mean ± SD)	11.1 ± 7.9	11.6 ± 7.8	0.52
Timing of DCI (days) (mean ± SD)	7.2 ± 2.7	6.8 ± 1.9	0.85
Mortality (%)	72 (18.5)	14 (14.3)	0.376
Modified Rankin Scale at 3 months (%)	220/390 (56.4)	68/98 (69.3)	0.185
0–3	181 (82.3)	61 (89.7)	
4–6	39 (17.7)	7 (10.3)	

DCI delayed cerebral ischemia, SD standard deviation

^a*p* values were calculated using Fisher exact test for frequency comparisons of categorical variables and Mann–Whitney *U* test for two-group comparisons of continuous variables. All statistical tests were two-tailed

Table 2 displays the baseline features, grading scales, and outcomes of the derivation and validation groups that were randomly selected while maintaining proportional targeted outcome. There was no significant difference found between the two groups.

3.1 Model performance

The median AUC of 100 runs of cross-validation (with 12.5% hold-out set) is presented in Table 3. Among the baseline characteristics, sex (AUC 0.59) performed slightly better than age (AUC 0.57, PLS). HH (AUC 0.58, PLS) and GCS (AUC 0.58, SVM-K) achieved better accuracy for DCI prediction than MFS (AUC 0.53, SVM-L). By combining baseline characteristics and grading scales (age, sex, HH, mFS, GCS), a PLS classifier performed better than the individual features with an AUC of 0.64. A combination of all random kernel derived physiological features using a PLS

classifier achieved an AUC of 0.71, while an SVM-L classifier achieved 0.72. Adding baseline characteristics and grading scales did not improve the performance.

Feature reduction with mRMR achieved the best classification performance with an AUC of 0.77 (PLS). In the case of the PLS classifier, the weights indicate the discriminative power of the features in separating the two classes. Figure 3a shows the PLS weights of these features; Fig. 3b shows the kernels corresponding to a demonstrative selection of the top ten features selected by the PLS model. The kernel associated with the feature O2 with downsampling (DS) 1 and kernel length (KL) 120 was given the highest weight among the 1600 features selected by the mRMR. The kernel displays the time varying characteristics for different variables and highlights the need for capturing high frequency data at different scales (downsampling rate).

The classification accuracy of the derived models on the unseen validation test set was found to be similar to

Table 3 Model performance in derivation and validation datasets

Feature extraction	Derivation dataset (median AUC of 100 runs)			Validation dataset [AUC (95% CI)]		
	PLS	SVM-L	SVM-K	PLS	SVM-L	SVM-K
Age	0.57	0.56	0.55	0.61 (0.49–0.73)	0.61 (0.49–0.73)	0.65 (0.54–0.76)
Sex	0.59	0.59	0.59	0.64 (0.52–0.76)	0.64 (0.52–0.76)	0.64 (0.52–0.76)
Hunt Hess Scale	0.58	0.54	0.53	0.57 (0.45–0.69)	0.6 (0.48–0.72)	0.56 (0.44–0.68)
Modified Fisher Scale	0.51	0.53	0.48	0.54 (0.42–0.66)	0.58 (0.46–0.7)	0.62 (0.5–0.74)
Glasgow Coma Scale	0.41	0.43	0.58	0.58 (0.46–0.7)	0.63 (0.51–0.75)	0.67 (0.56–0.78)
Combined baseline and scales	0.64	0.60	0.53	0.6 (0.48–0.72)	0.63 (0.51–0.75)	0.53 (0.41–0.65)
Diastolic BP	0.66	0.66	0.53	0.57 (0.45–0.87)	0.54 (0.45–0.87)	0.47 (0.31–0.75)
Systolic BP	0.57	0.62	0.53	0.64 (0.35–0.78)	0.7(0.41–0.84)	0.47 (0.31–0.75)
HR	0.61	0.59	0.50	0.68 (0.4–0.83)	0.62(0.37–0.81)	0.5 (0.28–0.72)
SpO2	0.54	0.55	0.49	0.58 (0.33–0.76)	0.6(0.33–0.77)	0.57 (0.27–0.71)
RR	0.58	0.60	0.49	0.66 (0.36–0.8)	0.67 (0.38–0.82)	0.58 (0.27–0.71)
Combined physiologic	0.71	0.72	0.50	0.68 (0.51–0.91)	0.74 (0.52–0.92)	0.5 (0.28–0.72)
Baseline and physiologic	0.71	0.71	0.50	0.69 (0.51–0.91)	0.75 (0.5–0.91)	0.5 (0.28–0.72)
mRMR (baseline + physiologic)	0.77	0.75	0.50	0.74 (0.58–0.96)	0.77 (0.55–0.94)	0.5 (0.28–0.72)

AUC area under the receiver operating characteristic curve, PLS partial least squares, SVM-L and -K support vector machine-linear and -kernel, BP blood pressure, HR heart rate, RR respiratory rate, mRMR Minimum Redundancy Maximal Relevance. All physiologic features are random kernel derived (diastolic and systolic BP, HR, SpO2, RR)

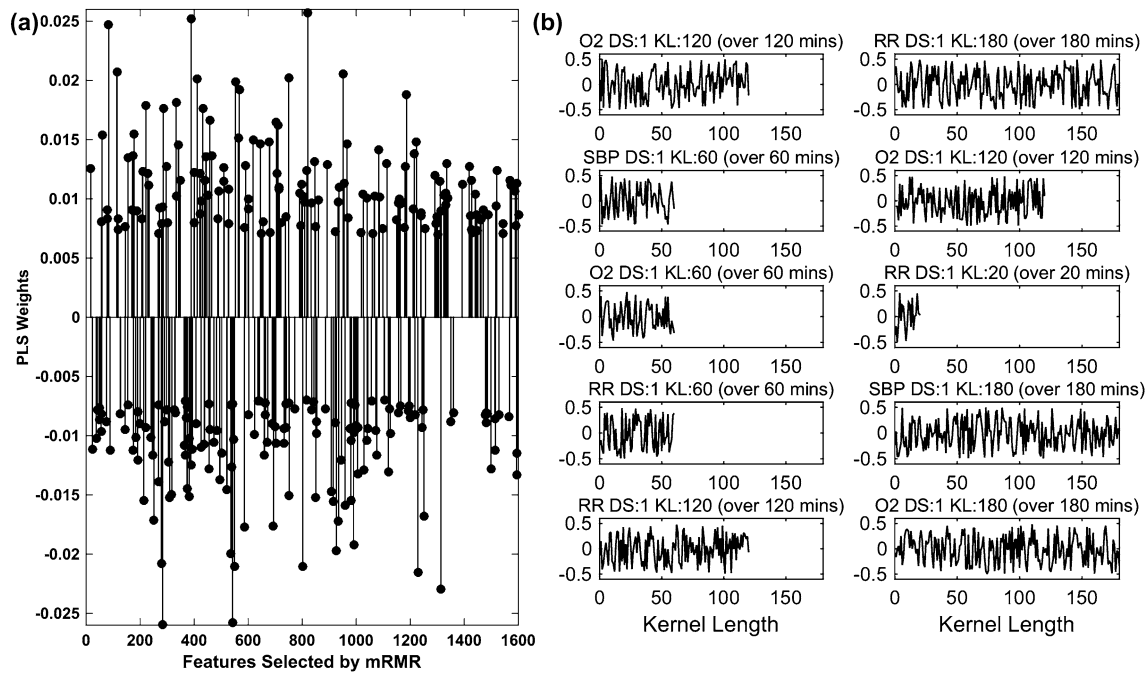


Fig. 3 Features selected by Minimum Redundancy–Maximum Relevance (mRMR) and classification by partial least squares (PLS). **a** A very large candidate feature set was reduced by mRMR, selecting for the top 50% of least redundant and most informative features. A PLS classifier was trained on these features, the weights are shown.

b For demonstration, the kernels for the top 10 weighted features selected by PLS are visualized, showing the time varying characteristics captured by the random kernels. Kernel length is represented on the x-axis

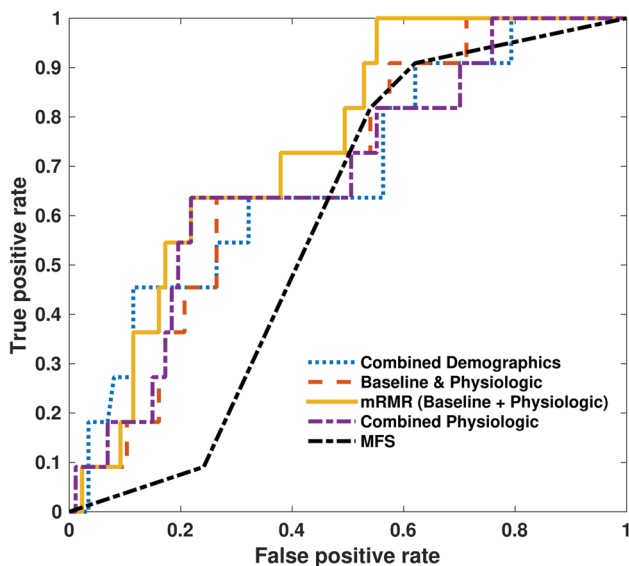


Fig. 4 AUC curves of model performance on validation dataset

the estimated performance of the cross-validated models. AUC curves are shown in Fig. 4. A model based on the traditional grading scale (MFS) achieved an AUC of 0.62 (SVM-L, 95% CI 0.5–0.74). Adding demographics and other

baseline scales did not improve prediction (AUC 0.63, 95% CI 0.51–0.75, SVM-L). Adding random kernel derived physiologic features improved prediction (AUC of 0.75, 95% CI 0.5–0.91, SVM-L), but actually performed the same as random kernel derived physiologic features alone (AUC 0.74, 95% CI 0.52–0.92, PLS). This combined physiologic model (HR, SBP, DBP, RR, and O2 sat) was more predictive than any single variable model (AUC range 0.47–0.70). Feature reduction (to reduce redundancy and maximal relevance) when applied to combined baseline, grading scales, and physiologic data produced the best classification performance with an AUC of 0.77 (95% CI 0.55–0.94, SVM-L).

4 Discussion

Recognizing trends and patterns, and minutely analyzing complex data requires the layered knowledge of clinical experts, but defies rule-based systems. In prior work (SP) using summary statistics of 24 h (lower frequency) data, a Naïve Bayes classifier for angiographic VSP outperformed TCD and exams, and generated an AUC of 0.71 [35]. Here, we show that features extracted from higher frequency temporal data (ranging from 1 min to 4 h) may be superior to lower frequency data in classification of outcome after SAH.

In our approach, we extracted high level features from existing physiologic data, without an a priori hypothesis of what patterns might emerge. To enable validation efforts and generalizability to other datasets and institutions, we focused on universal physiologic ICU variables and typical baseline grading scales pertinent to SAH used in the NICU. In this translational work, we used a random featurization method to extract frequency selective and translation invariant characteristics of time series data. The novel application to time series data required some choices bound by characteristics of the dataset (kernel lengths) and domain (downsampling rates). We tested our method for its discriminative ability for DCI and found that random kernel derived physiologic features outperformed current static grading scales. When combined with grading scales and demographics, our random kernel derived physiologic features predicted DCI with an AUC (0.77, PLS) approaching clinical reliability (threshold of 0.8 [63]). PLS and SVM-L models performed equally well, indicating the sufficient discriminative ability of our feature extraction method.

An effort to show robustness of the model was an internal validation strategy, testing on a separate dataset excluded from model building entirely. Generalizability of a machine learning algorithm, however, assumes that the training dataset is large and diverse enough to be representative. A limitation to this study is the single center approach; there is no publicly available dataset for SAH with similar granularity of physiologic data. Future efforts will include developing complementary SAH cohorts and validating these algorithms.

5 Conclusions

A random kernel featurization and learning approach to physiological time series data prior to peak DCI period shows promise to improve prediction precision. This is a computationally inexpensive and agnostic feature extraction approach for physiologic time series parameters in the ICU (HR, RR, SBP, DBP, O₂ sat). There is a vast pool of candidate features within the EMR with a biological basis for classification ability (i.e. drawn from frequentist statistical studies showing relationship with VSP and DCI in specific SAH cohorts). Future efforts will also draw from this feature pool to further improve the precision of DCI prediction, favoring those candidate features that are obtained for standard clinical care and thus potentially automatable.

Acknowledgements Data Collection (SP, HF, JC, SA, DJR, ESC, JMS, AV, KT), Analysis (SP, HF, EG, MM, CW, NE), Writing (SP, HF, MM), Editing (All).

Funding National Institutes of Health K01-ES026833-02 (SP), National Institutes of Health U54-CA193313-01 (CW). National Science Foundation 1305023 (CW). National Science Foundation 1344668 (CW, NE).

Compliance with ethical standards

Conflict of interest Authors do not have any other disclosures of potential conflicts of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. This study has been approved by the Columbia University Medical Center Institutional Review Board.

Informed consent Informed consent was obtained from all individual participants or their surrogates included in the study.

References

1. Suarez J. PRINCE Neurocritical Care Point-Prevalence Study Preliminary Results Revealed, vol. 9. *Currents: News Magazine of the Neurocritical Care Society*; 2014.
2. Report of World Federation of Neurological Surgeons Committee on a Universal Subarachnoid Hemorrhage Grading Scale. *J Neurosurg.* 1988;68(6):985–6.
3. Shea AM, Reed SD, Curtis LH, Alexander MJ, Villani JJ, Schulman KA. Characteristics of nontraumatic subarachnoid hemorrhage in the United States in 2003. *Neurosurgery* 2007;61(6):1131–7. <https://doi.org/10.1227/01.neu.0000306090.30517.ae> (**discussion 1137–1138**).
4. Qureshi AI, Suri MF, Nasar A, Kirmani JF, Divani AA, He W, Hopkins LN. Trends in hospitalization and mortality for subarachnoid hemorrhage and unruptured aneurysms in the United States. *Neurosurgery* 2005;57(1):1–8 (**discussion 1–8**).
5. Molyneux AJ, Kerr RS, Birks J, Ramzi N, Yarnold J, Sneade M, Rischmiller J, Collaborators I. Risk of recurrent subarachnoid haemorrhage, death, or dependence and standardised mortality ratios after clipping or coiling of an intracranial aneurysm in the International Subarachnoid Aneurysm Trial (ISAT): long-term follow-up. *Lancet Neurol.* 2009;8(5):427–33. [https://doi.org/10.1016/S1474-4422\(09\)70080-8](https://doi.org/10.1016/S1474-4422(09)70080-8).
6. Springer MV, Schmidt JM, Wartenberg KE, Frontera JA, Badjatia N, Mayer SA. Predictors of global cognitive impairment 1 year after subarachnoid hemorrhage. *Neurosurgery* 2009;65(6):1043–50. <https://doi.org/10.1227/01.NEU.0000359317.15269.20> (**discussion 1050–1041**).
7. Roos YB, Dijkgraaf MG, Albrecht KW, Beenen LF, Groen RJ, de Haan RJ, Vermeulen M. Direct costs of modern treatment of aneurysmal subarachnoid hemorrhage in the first year after diagnosis. *Stroke* 2002;33(6):1595–9.
8. Mayer SA, Kreiter KT, Copeland D, Bernardini GL, Bates JE, Peery S, Claassen J, Du YE, Connolly ES Jr. Global and domain-specific cognitive impairment and outcome after subarachnoid hemorrhage. *Neurology.* 2002;59(11):1750–8.
9. Hackett ML, Anderson CS. Health outcomes 1 year after subarachnoid hemorrhage: An international population-based study. The Australian Cooperative Research on Subarachnoid Hemorrhage Study Group. *Neurology.* 2000;55(5):658–62.

10. Charpentier C, Audibert G, Guillemin F, Civit T, Ducrocq X, Bracard S, Hepner H, Picard L, Laxenaire MC. Multivariate analysis of predictors of cerebral vasospasm occurrence after aneurysmal subarachnoid hemorrhage. *Stroke* 1999;30(7):1402–8.
11. Dorsch N. A clinical review of cerebral vasospasm and delayed ischaemia following aneurysm rupture. *Acta Neurochir Suppl.* 2011;110(Pt 1):5–6. https://doi.org/10.1007/978-3-7091-0353-1_1.
12. Schmidt JM, Wartenberg KE, Fernandez A, Claassen J, Rincon F, Ostapkovich ND, Badjatia N, Parra A, Connolly ES, Mayer SA. Frequency and clinical impact of asymptomatic cerebral infarction due to vasospasm after subarachnoid hemorrhage. *J Neurosurg.* 2008;109(6):1052–9. <https://doi.org/10.3171/JNS.2008.109.12.1052>.
13. Rabinstein AA, Pichelmann MA, Friedman JA, Piepgras DG, Nichols DA, McIver JI, Toussaint LG 3rd, McClelland RL, Fulgham JR, Meyer FB, Atkinson JL, Wijdicks EF. Symptomatic vasospasm and outcomes following aneurysmal subarachnoid hemorrhage: a comparison between surgical repair and endovascular coil occlusion. *J Neurosurg.* 2003;98(2):319–25. <https://doi.org/10.3171/jns.2003.98.2.0319>.
14. Kirmani JF, Qureshi AI, Hanel RA, Siddiqui AM, Safdar A, Yahia AM, Kim SH, Guterman LR, Hopkins LN. Silent cerebral infarctions in poor-grade patients with subarachnoid hemorrhage. *Neurology* 2002;58(7):A159.
15. Frontera JA, Fernandez A, Schmidt JM, Claassen J, Wartenberg KE, Badjatia N, Connolly ES, Mayer SA. Defining vasospasm after subarachnoid hemorrhage: what is the most clinically relevant definition? *Stroke* (2009);40(6):1963–8. <https://doi.org/10.1161/STROKEAHA.108.544700>.
16. Vergouwen MD, Vermeulen M, van Gijn J, Rinkel GJ, Wijdicks EF, Muizelaar JP, Mendelow AD, Juvela S, Yonas H, Terbrugge KG, Macdonald RL, Diringier MN, Broderick JP, Dreier JP, Roos YB. Definition of delayed cerebral ischemia after aneurysmal subarachnoid hemorrhage as an outcome event in clinical trials and observational studies: proposal of a multidisciplinary research group. *Stroke* 2010;41(10):2391–5. <https://doi.org/10.1161/STROKEAHA.110.589275>.
17. Rosen DS, Macdonald RL. Subarachnoid hemorrhage grading scales: a systematic review. *Neurocrit Care.* 2005;2(2):110–8. <https://doi.org/10.1385/NCC:2:2:110>.
18. Fisher CM, Kistler JP, Davis JM. Relation of cerebral vasospasm to subarachnoid hemorrhage visualized by computerized tomographic scanning. *Neurosurgery* 1980;6(1):1–9.
19. Claassen J, Bernardini GL, Kreiter K, Bates J, Du YE, Cope land D, Connolly ES, Mayer SA. Effect of cisternal and ventricular blood on risk of delayed cerebral ischemia after subarachnoid hemorrhage: the Fisher scale revisited. *Stroke* 2001;32(9):2012–20.
20. Gaieski DF, Mikkelsen ME, Band RA, Pines JM, Massone R, Furla FF, Shofer FS, Goyal M. Impact of time to antibiotics on survival in patients with severe sepsis or septic shock in whom early goal-directed therapy was initiated in the emergency department. *Crit Care Med.* 2010;38(4):1045–53. <https://doi.org/10.1097/CCM.0b013e3181cc4824>.
21. Diringier MN. Management of aneurysmal subarachnoid hemorrhage. *Crit Care Med.* 2009;37(2):432–40. <https://doi.org/10.1097/CCM.0b013e318195865a>.
22. Heros RC, Zervas NT, Varsos V. Cerebral vasospasm after subarachnoid hemorrhage: an update. *Ann Neurol.* 1983;14(6):599–608. <https://doi.org/10.1002/ana.410140602>.
23. Lindgaard KF, Nornes H, Bakke SJ, Sorteberg W, Nakstad P. Cerebral vasospasm diagnosis by means of angiography and blood velocity measurements. *Acta Neurochir.* 1989;100(1–2):12–24.
24. Krejza J, Szydlik P, Liebeskind DS, Kochanowicz J, Bronov O, Mariak Z, Melhem ER. Age and sex variability and normal reference values for the V(MCA)/V(ICA) index. *AJNR* 2005;26(4):730–5.
25. Naval NS, Thomas CE, Urrutia VC. Relative changes in flow velocities in vasospasm after subarachnoid hemorrhage: a transcranial Doppler study. *Neurocrit Care.* 2005;2(2):133–40. <https://doi.org/10.1385/NCC:2:2:133>.
26. Grosset DG, Straiton J, McDonald I, Cockburn M, Bullock R. Use of transcranial Doppler sonography to predict development of a delayed ischemic deficit after subarachnoid hemorrhage. *J Neurosurg.* 1993;78(2):183–7. <https://doi.org/10.3171/jns.1993.78.2.0183>.
27. Sekhar LN, Wechsler LR, Yonas H, Luyckx K, Obrist W. Value of transcranial Doppler examination in the diagnosis of cerebral vasospasm after subarachnoid hemorrhage. *Neurosurgery* 1988;22(5):813–21.
28. Lysakowski C, Walder B, Costanza MC, Tramer MR. Transcranial Doppler versus angiography in patients with vasospasm due to a ruptured cerebral aneurysm: A systematic review. *Stroke* 2001;32(10):2292–8.
29. Harders AG, Gilsbach JM. Time course of blood velocity changes related to vasospasm in the circle of Willis measured by transcranial Doppler ultrasound. *J Neurosurg.* 1987;66(5):718–28. <https://doi.org/10.3171/jns.1987.66.5.0718>.
30. Crobeddu E, Mittal MK, Dupont S, Wijdicks EF, Lanzino G, Rabinstein AA. Predicting the lack of development of delayed cerebral ischemia after aneurysmal subarachnoid hemorrhage. *Stroke* 2012;43(3):697–701. <https://doi.org/10.1161/STROKEAHA.111.638403>.
31. Frontera JA, Claassen J, Schmidt JM, Wartenberg KE, Temes R, Connolly ES Jr, MacDonald RL, Mayer SA. Prediction of symptomatic vasospasm after subarachnoid hemorrhage: the modified Fisher Scale. *Neurosurgery* 2006;59(1):21–7. <https://doi.org/10.1227/01.NEU.0000218821.34014.1B> (**discussion 21–27**).
32. Foreman PM, Chua MH, Harrigan MR, Fisher WS 3rd, Tubbs RS, Shoja MM, Griessenauer CJ. External validation of the Practical Risk Chart for the prediction of delayed cerebral ischemia following aneurysmal subarachnoid hemorrhage. *J Neurosurg.* 2016. <https://doi.org/10.3171/2016.1.JNS152554>.
33. de Rooij NK, Greving JP, Rinkel GJ, Frijns CJ. Early prediction of delayed cerebral ischemia after subarachnoid hemorrhage: development and validation of a practical risk chart. *Stroke* 2013;44(5):1288–94. <https://doi.org/10.1161/STROKEAHA.113.001125>.
34. Calviere L, Nasr N, Arnaud C, Czosnyka M, Viguier A, Tissot B, Sol JC, Larrue V. Prediction of delayed cerebral ischemia after subarachnoid hemorrhage using cerebral blood flow velocities and cerebral autoregulation assessment. *Neurocrit Care.* 2015;23(2):253–8. <https://doi.org/10.1007/s12028-015-0125-x>.
35. Roederer A, Holmes JH, Smith MJ, Lee I, Park S. Prediction of significant vasospasm in aneurysmal subarachnoid hemorrhage using automated data. *Neurocrit Care.* 2014;21(3):444–50. <https://doi.org/10.1007/s12028-014-9976-9>.
36. Sacchi L, Dagliati A, Bellazzi R. Analyzing complex patients' temporal histories: new frontiers in temporal data mining. *Methods Mol Biol.* 2015;1246:89–105. https://doi.org/10.1007/978-1-4939-1985-7_6.
37. Stacey M, McGregor C. Temporal abstraction in intelligent clinical data analysis: a survey. *Artif Intell Med.* 2007;39(1):1–24. <https://doi.org/10.1016/j.artmed.2006.08.002>.
38. Verduijn M, Sacchi L, Peek N, Bellazzi R, de Jonge E, de Mol BA. Temporal abstraction for feature extraction: a comparative case study in prediction from intensive care monitoring data. *Artif Intell Med.* 2007;41(1):1–12. <https://doi.org/10.1016/j.artmed.2007.06.003>.

39. Saria S, Rajani AK, Gould J, Koller D, Penn AA. Integration of early physiological responses predicts later illness severity in preterm infants. *Sci Transl Med*. 2010;2(48):48ra65. <https://doi.org/10.1126/scitranslmed.3001304>.
40. Mayer CC, Bachler M, Hortenhuber M, Stocker C, Holzinger A, Wassertheurer S. (2014) Selection of entropy-measure parameters for knowledge discovery in heart rate variability data. *BMC Bioinform*. 15(Suppl 6):S2. <https://doi.org/10.1186/1471-2105-15-S6-S2>.
41. Dua S, Saini S, Singh H. Temporal pattern mining for multivariate time series classification. *J Med Imag Health Inform*. 2011;1(2):164–9. <https://doi.org/10.1166/jmih.2011.1019>.
42. Lehman LW, Adams RP, Mayaud L, Moody GB, Malhotra A, Mark RG, Nemati S. A physiological time series dynamics-based approach to patient monitoring and outcome prediction. *IEEE J Biomed Health Inform*. 2015;19(3):1068–76. <https://doi.org/10.1109/JBHI.2014.2330827>.
43. Schulam P, Wigley F, Saria S. Clustering longitudinal clinical marker trajectories from electronic health data. Applications to phenotyping and endotype discovery. In: AAAI, Citeseer; 2015. pp. 2956–64.
44. Nemati S, Adams R. Supervised learning in dynamic bayesian networks. *Neural information processing systems (NIPS) workshop on deep learning and representation learning*, Montreal; 2014.
45. Luo Y, Xin Y, Joshi R, Celi L, Szolovits P. Predicting ICU Mortality risk by grouping temporal trends from a multivariate panel of physiologic measurements. In: AAAI, Phoenix; 2016. pp. 42–50.
46. Saria S, Koller D, Penn A. Learning individual and population level traits from clinical temporal data. In: *Proceeding of neural information processing systems (NIPS), predictive models in personalized medicine workshop*; 2010.
47. Lipton ZC, Kale DC, Wetzell RC. (2015) Phenotyping of clinical time series with LSTM recurrent neural networks. *arXiv preprint arXiv:151007641*.
48. Lipton ZC, Kale DC, Elkan C, Wetzell R. (2015) Learning to diagnose with LSTM recurrent neural networks. *ArXiv e-prints 1511*.
49. Kale DC, Gong D, Che Z, Liu Y, Medioni G, Wetzell R, Ross P An examination of multivariate time series hashing with applications to health care. In: *Data Mining (ICDM), 2014 IEEE international conference on, 2014*. IEEE, pp 260–269.
50. Bahadori MT, Kale DC, Fan Y, Liu Y. Functional subspace clustering with application to time series. In: *ICML, Lille; 2015*. pp. 228–37.
51. Marlin BM, Kale DC, Khemani RG, Wetzell RC Unsupervised pattern discovery in electronic health care data using probabilistic clustering models. In: *Proceedings of the 2nd ACM SIGHIT international health informatics symposium, 2012*. ACM, pp 389–98.
52. Rahimi A, Recht B. Random features for large-scale kernel machines. *NIPS*. 2007;3(4):5.
53. Rahimi A, Recht B. Weighted sums of random kitchen sinks: replacing minimization with randomization in learning. *Adv Neural Inform Process Syst*. 2008;885:1313–20.
54. Saxe A, Koh PW, Chen Z, Bhand M, Suresh B, Ng AY. (2011) On random weights and unsupervised feature learning. Paper presented at the proceedings of the 28th international conference on machine learning (ICML-11).
55. Johnson AE, Ghassemi MM, Nemati S, Niehaus KE, Clifton DA, Clifford GD. Machine learning and decision support in critical care. *Proc IEEE Inst Electr Electron Eng* (2016);104(2):444–66. <https://doi.org/10.1109/JPROC.2015.2501978>.
56. Peng H, Long F, Ding C. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans Pattern Anal Mach Intell*. 2005;27(8):1226–38. <https://doi.org/10.1109/TPAMI.2005.159>.
57. Ding C, Peng H. Minimum redundancy feature selection from microarray gene expression data. *J Bioinform Comput Biol*. 2005;3(2):185–205.
58. Peng HC, Ding C, Long FH. Minimum redundancy—maximum relevance feature selection. *IEEE Intell Syst*. 2005;20(6):70–1.
59. Huang YM, Du SX. (2005) Weighted support vector machine for classification with uneven training class sizes. In: *Proceedings of 2005 international conference on machine learning and cybernetics, Guangzhou, vol. 1–9*, pp 4365–69.
60. Cortes C, Vapnik V. Support-vector networks. *Mach Learn*. 1995;20(3):273–97. <https://doi.org/10.1007/bf00994018>.
61. Geladi P, Kowalski BR. Partial least-squares regression: a tutorial. *Anal Chim Acta*. 1986;185:1–17.
62. Chang CC, Lin CJ. LIBSVM: a library for support vector machines. *ACM T Intel Syst Tec*. 2011;2(3):27. <https://doi.org/10.1145/1961189.1961199>.
63. Steyerberg EW, Pencina MJ, Lingsma HF, Kattan MW, Vickers AJ, Van Calster B. Assessing the incremental value of diagnostic and prognostic markers: a review and illustration. *Eur J Clin Invest*. 2012;42(2):216–28. <https://doi.org/10.1111/j.1365-2362.2011.02562.x>.