



HHS Public Access

Author manuscript

Pediatr Crit Care Med. Author manuscript; available in PMC 2016 November 01.

Published in final edited form as:

Pediatr Crit Care Med. 2015 November ; 16(9): e332–e339. doi:10.1097/PCC.0000000000000560.

Using Time Series Analysis to Predict Cardiac Arrest in a Pediatric Intensive Care Unit

Curtis E Kennedy, MD, PhD [Assistant Professor, Section of Critical Care Medicine],
Department of Pediatrics, Baylor College of Medicine, Texas Children's Hospital, Houston, TX

Noriaki Aoki, MD, PhD [Associate Professor],
The University of Texas School of Biomedical Informatics, Texas Children's Hospital, Houston, TX

Michele Mariscalco, MD [Regional Dean], and
University of Illinois College of Medicine at Urbana

James P Turley, RN, PhD [Associate Professor]
The University of Texas School of Biomedical Informatics, Texas Children's Hospital, Houston, TX

Abstract

Objectives—To build and test cardiac arrest prediction models in a pediatric intensive care unit, using time series analysis as input, and to measure changes in prediction accuracy attributable to different classes of time series data.

Methods—A retrospective cohort study of pediatric intensive care patients over a 30 month study period. All subjects identified by code documentation sheets with matches in hospital physiologic and laboratory data repositories and who underwent chest compressions for two minutes were included as arrest cases. Controls were randomly selected from patients that did not experience arrest and who survived to discharge. Modeling data was based on twelve hours of data preceding the arrest (reference time for controls).

Measurements and Main Results—103 cases of cardiac arrest and 109 control cases were used to prepare a baseline data set that consisted of 1025 variables in four data classes: multivariate, raw time series, clinical calculations, and time series trend analysis. We trained 20 arrest prediction models using a matrix of five feature sets (combinations of data classes) with four modeling algorithms: linear regression, decision tree, neural network and support vector machine. The reference model (multivariate data with regression algorithm) had an accuracy of 78% and 87% area under the receiver operating characteristic curve (AUROC). The best model (multivariate + trend analysis data with support vector machine algorithm) had an accuracy of 94% and 98% AUROC.

Contact Information: Curtis Kennedy, MD, PhD, 6621 Fannin WT 6-006, Houston, TX 77030, ckennedy@bcm.edu, 832 826 6230.
Noriaki Aoki (Deceased)
James P Turley (Retired)

Disclosures: The authors have no financial disclosures or conflicts of interest

Conclusions—Cardiac arrest predictions based on a traditional model built with multivariate data and a regression algorithm misclassified cases 3.7 times more frequently than predictions that included time series trend analysis and built with a support vector machine algorithm. Although the final model lacks the specificity necessary for clinical application, we have demonstrated how information from time series data can be used to increase the accuracy of clinical prediction models.

Keywords

pediatric intensive care units; projections and predictions; machine learning; cardiac arrest; clinical decision support systems

Introduction

Children admitted to pediatric intensive care units (PICU) are inherently unstable, and change rapidly between states of improvement and deterioration. When deteriorations happen, bedside caregivers must detect them, assess their potential impact to the patient, and intervene if necessary. Each of these functions is operator-dependent, so patients receive different levels of service at different points in time. Despite continuous monitoring of their vital signs and high staffing ratios, thousands of children suffer cardiac arrests in PICUs every year (1-4). Many arrests are preceded by serial deteriorations in vital signs (5-8), suggesting that progressive shock may contribute to some arrests, though not all since some arrests occur suddenly and without warning. The shock state can be a result of insufficient oxygen or fuel availability, or of inadequate cardiac output.

In an ideal environment, a cardiac arrest attributable to undertreated shock could be averted by early recognition of deterioration and timely intervention before the arrest occurred. The goal of this study is to focus on the recognition aspect of the problem by using time series analysis as a way to encode deterioration (a time dependent phenomenon) and evaluating its utility as an input into prediction models for cardiac arrest in a PICU.

Most scoring tools in clinical use are based on models built using multivariate data structures(9-14), for which single values represent a variable of interest. If time dependent phenomena are included in these models, they are usually limited to quasi-multivariate abstractions such as min/max or magnitude of change between two time points(15). The tools are typically based on regression algorithms, which perform well in smaller data sets but have not proven robust in large data sets where the number of variables outweigh the number of training examples. This problem has been addressed in genetic microarray analysis(16,17), where more sophisticated algorithms such as neural networks and support vector machines have proven robust(18). In this study our aims are to determine whether or not adding time series analysis results as input into cardiac arrest prediction models increase their predictive accuracy, and to determine whether or not modeling algorithm choice influences predictive accuracy.

Methods

Setting

This is a retrospective cohort study of patients admitted to a tertiary care PICU that provides care to over 2000 cases annually and serves a referral population of over 4 million people. Patients treated range in age from infants to the early 20s, with a median age of 5 years. Disease classes treated include medical, surgical, trauma, cancer, and bone marrow and solid organ transplantation. Patients with operative cardiac conditions are cared for in a separate cardiovascular ICU that is not part of this study. The study protocol was approved by our institutional review board. We identified 103 cases of cardiac arrest that occurred in the PICU by reviewing code sheets that were generated when patients received acute, intensive resuscitation between July 2006 and December 2008. Criteria for inclusion as an arrest case were: 1. event location in PICU; 2. first cardiac arrest in the PICU; 3. external cardiac massage for at least two minutes; and 4. able to be matched with records from the hospital's data repository and physiologic monitor database. We also identified 109 control cases from patients admitted to the PICU by random selection from the following three categories: 1. first six hours of admission (representing the most consistent point in time when patients experience rapid change and receive multiple interventions); 2. day of maximum severity of illness occurring after 24 hours of hospitalization (representing deteriorations after admission); and 3. random point in time (representing baseline noise in PICU physiology). Criteria for inclusion as a control case included: 1. did not experience a cardiac arrest in the PICU; 2. survived to discharge (to exclude deteriorations in Do Not Attempt Resuscitation patients); 3. had matching data in the physiologic monitor database and the data repository; and 4. selected by random number generation (to keep the case:control ratio roughly equal in order to satisfy modeling algorithm assumptions).

Equipment

Physiologic monitor data was measured on GE Solar 8000M patient monitors and archived by Excel Medical BedMaster Software (version 1.3) configured to log data every one minute. Modeling and analyses were conducted in MATLAB (R2007a), using The Spider programming environment (version 1.71).

Data Aggregation

Data from code sheets, the hospital's data repository, and the physiologic monitor database were merged to create the initial data set. For arrest cases, we defined the reference time in the monitor database as the initial deterioration's worst measurement, ranked by heart rate, then pulse oximeter, and finally blood pressure criteria. We assumed time synchronization between the data repository and monitor database. Reference time assignment for control cases were randomly assigned from within their designated block of time. Time series data were constrained to a 12 hour prearrest window.

Preprocessing and Data Class Assignment

Outlier removal and imputation of missing data elements were performed using a limit-based, carry-forward strategy. When no values were present from which to carry forward, a normal value for the field was imputed.

We defined multivariate data as either a single measurement (for lab data) or the first measurement preceding the reference time (for time series data). We defined raw time series data as all measurements preceding the one classified as multivariate. Raw time series data consisted of 60 high-resolution data elements (every minute for one hour prior to the reference point) and 12 low-resolution data elements (hourly averages for 12 hours prior to the reference time). Trend analysis data consisted of means, slopes, and intercepts for 5, 10, 15, and 60 minute prearrest epochs; and also included ratios of means between each epoch and more distal epochs: 5/(10,15,60); 10/(15,60); and 15/60.

Given the model's theoretical basis on a progressive shock state, we explicitly encoded two clinical calculations (CC) as a separate data class: the shock index (SI = heart rate / systolic blood pressure) and an oxygen delivery index (ODI = heart rate * pulse pressure * hemoglobin * % oxygen saturation) were calculated for each set of vital signs. The shock index has been shown to correlate with severity of shock(19), and the oxygen delivery index (ODI) is based on the oxygen delivery equation used in hemodynamic calculations(20). Here, we substituted heart rate and pulse pressure as surrogate variables for cardiac output, since it is not directly measured.

We created five feature sets to use as model inputs, using various combinations of the four data classes described above. To prepare variables for modeling, each variable was transposed to a 0:1 range by min-max normalization. Data was then split: 67% into a training / internal validation set and 33% into a holdout testing / external validation set.

Modeling

We developed 20 models from a matrix of five feature sets and four modeling algorithms. (Figure 1) Feature sets included: 1. pure multivariate (MV: traditional approach as reference / no time series elements); 2. multivariate + raw time series (TS: to measure effects of unaltered time series measurements); 3. multivariate + raw time series + clinical calculations (CC: to measure effects of clinical calculations); 4. multivariate + trend analysis (TRD: to measure effects of trend analysis without raw time series measurements); and 5. All classes combined (ALL: to measure effects of all data classes in combination). Modeling algorithms included: 1. linear regression (LR: traditional approach as reference); 2. j48 decision tree (DT: advanced algorithm / readable model); 3. neural network (NN: advanced algorithm / black box); and 4. support vector machine (SVM: advanced algorithm / black box). For each of the 20 models, we performed 10-fold cross validation to determine mean performance characteristics and their associated variance. We generated one representative model for each combination of feature set + modeling algorithm using the training data set. We estimated external validity in the holdout / testing data set by comparing each model's predictions to the actual categories (arrest vs. control). Performance measures included accuracy (ACC) and area under the receiver operating characteristic curve (AUROC). Since

data sets contained a balanced ratio of cases:controls, performance measures relying on positive predictive value in their determination were not performed.

To further isolate and validate the effects of trend calculations on model performance, we repeated the above procedures after eliminating variables measured only once (i.e., not subject to trend calculations).

To assess for overfitting effects we performed two methods of automated variable selection: SVM weighting (SVMW) and recursive feature elimination (RFE), constraining models to a range of 15 to 50 variables in 5 variable increments. We used the dominant modeling algorithm (SVM) to repeat model training and testing for the resultant matrix of eight input feature sets (of 15, 20, 25, 30, 35, 40, 45 and 50 variables) for each variable selection tool (SVMW and RFE).

Our final step was to qualitatively assess the results we obtained in the steps above, to determine if any variables were conserved between the various modeling algorithms.

Results

We identified 103 cases of initial PICU cardiac arrest events with corresponding data in the physiologic monitor database and the data repository. 109 controls were randomly selected from their respective categories. Lab data had 0.47% outliers, and physiologic data had 1.18% outliers. Imputation accounted for 16.8% of lab data and for 1.7% of physiologic data.

Significant differences in mean values between arrest and control cohorts were present for: 1) 16 of 20 (80%) multivariate variables; 2) 413 of 497 (83%) raw time series variables; 3) 155 of 288 (54%) clinical calculations; and 4) 182 of 220 (83%) trend analysis results. Figure 2 shows the differences in heart rate, oxygenation, and systolic blood pressure for arrest versus control cases. With the notable exception of respiratory rate, all vital signs in the arrest category demonstrated a drop in mean value starting as many as 20 minutes from the arrest, with more drastic drops occurring in the 5-minute prearrest window.

Internal measures of model accuracy, using 10-fold cross validation, demonstrated a baseline accuracy of $66 \pm 4\%$ (MV+LR model). The TRD+SVM model demonstrated the best performance, with an accuracy of $79 \pm 4\%$ ($p < 0.0001$ vs. baseline). Only models that included trend analysis data demonstrated significant increases in accuracy when compared to baseline. When all data elements were used (ALL+SVM), accuracy dropped to $73\% \pm 5\%$ ($p = 0.02$ vs. TRD+SVM), suggesting a moderate overfitting effect.

External validation in the test data set (unseen during model training) demonstrated better performance than internal measures. The best performance was again seen in the TRD+SVM model: with 94% accuracy (baseline accuracy for MV+LR was 78%). Similarly, AUROC increased from an 87% baseline to 98% in the TRD+SVM model. Performance characteristics across the matrix of 20 models is given in Table 1. The AUROC for the reference model (MV+LR) vs. best model (TRD+SVM) is shown in Figure 3.

Repeated training and external validation in data excluding single measurement variables (where trend calculations could not be performed) resulted in an increase in logistic regression performance without substantial effect on other modeling algorithm performance. Trend features again served to improve model performance, shown in Table 1, with accuracy increasing from 83% (MV+LR) to 94% (TRD+SVM), and AUROC increasing from 89% (MV+LR) to 98% (TRD+SVM).

Feature reduction with RFE demonstrated peak accuracy at 15 variables: 88% accuracy with 86% AUROC. SVMW demonstrated peak accuracy at 35 variables: 95% accuracy with 96% AUROC.

No variables were conserved across all models. The best model (TRD+SVM) used 51% trend analysis variables and 49% multivariate variables. Figure 4 shows a plot of the top 35 variables selected by SVM weighting, providing the best visual discrimination of variables separating arrest cases from control subjects.

Discussion

The mortality rate of inpatient pediatric cardiac arrest is generally reported to be in excess of 60%, with high disability rates in survivors (21-27). Antecedents to cardiac arrest identified in the literature have been described primarily in the context of patients in acute care units who deteriorate to the point of requiring transfer to an ICU or arresting (28-30). These studies suggest that deteriorations often are detectable hours before an arrest occurs and that patients often are evaluated beforehand but fail to receive treatment that could possibly prevent the event. Scoring tools have been developed and deployed to help Medical Emergency Teams (METs) objectively assess patients for risk of having life-threatening deteriorations (9,31). However, their target population is one of relatively healthy patients, and their purpose is to differentiate a patient who is sick from one who is healthy. ICUs contain a population of patients already determined to be sick, so scoring tools that have proven useful in an acute care setting are unable to identify the patients who are most likely to suffer cardiac arrest in an intensive care setting. Clinicians use data from earlier points in time to interpret new data and determine their implications. Tools that can perform these interpretations automatically are needed.

Physiologic monitors connected to patients are not good tools for identifying patients at risk of having cardiac arrest. They are plagued by such high rates of false alarms that nurses frequently ignore them(32). Several techniques relevant to this study have been employed to help reduce false alarms (33-36). Our approach combines the usefulness of multiple channels and multiple measurements, but also explicitly characterizes the outcome of interest (cardiac arrest) in terms of measureable risk factors based on a physiologic model (progressive shock).(37)

Deteriorations are frequent events in PICUs, and the vast majority do not result in cardiac arrest. Determining which deteriorations are associated with higher risk than others of progressing to arrest becomes one of clinical intuition. Objective scoring tools that accurately and continuously screen patients for arrest risk can help provide a decision

support based safety net that can present caregivers with automated assessments, help them perceive and interpret changes they may not otherwise be aware of, and help them to decide which deteriorations have the highest risk. In the same spirit that early warning scores have helped prevent death and disability in acute care units of the hospital, intensive care based warning scores could prevent death and disability in a population that is at much higher baseline risk.

We were surprised that the trend analysis model (without the raw time series data) outperformed models that included raw time series data, and that it even outperformed models whose variables were selected by automated feature selection tools. This is likely attributable to an overfitting effect in the raw time series data, since downsampling into trend data substantially reduces the variable:case ratio. We were also surprised that the external validation performance exceeded that of the internal 10-fold cross-validation. We believe this is most likely due to a “doubly restricted” set of targets available to the algorithms during internal validation: models trained for external validation used all training data, whereas internal cross-validation required additional cases to be held out during each train/test cycle.

This being the first case of use, there are many limitations to the study. First and foremost, although we have demonstrated a significant improvement in baseline performance by adding time series analysis results as model inputs, the incidence of cardiac arrest is so low that employing the current model would still result in an unacceptably high false alarm rate. A tool that continuously monitors risk for cardiac arrest requires a specificity of roughly 99.99% in order to alarm once for every 7 patient-days. Even at that level, the false alarm : true alarm ratio would be close to 10:1. This is an important consideration for future work in continuous risk prediction models.

A second limitation to this study is the superficial nature of comparisons. We only included variables that relate to physiologic and laboratory findings, and were unable to include a number of desirable variables because they were not available electronically: comorbid conditions, prior ICU admissions, medications (including vasoactive infusions), transfusions, etc. Also, we only included trend analysis from the time series data. Numerous other analyses could be performed on the time series data and included as additional classes of data. Specifically, heart rate variability measures from the frequency domain of time series analysis could be used in this way. We only included four modeling algorithms of dozens that are available, and we limited our scope to using default model parameters because the number of permutations would otherwise be too large. It is unlikely that the optimum modeling algorithm is the SVM with its default parameters.

We focused this study on permutations of data classes used to train models, modeling algorithms, and feature selection. We were not able to include a sensitivity analysis of optimum resolutions in the prearrest timeframe. We selected a 60 minute high resolution + 12 hour low resolution prearrest window based on intuition, and it is possible that different sized windows could improve predictive accuracy. Similarly, our data set contained data up to one minute prior to the arrest. We therefore do not know precisely how fast performance will deteriorate as data is serially restricted to prearrest windows that exclude the two to ten

minutes (or more) of data immediately adjacent to the arrest. Our control selection focused on culling data from different phases of care (peri-admission, maximum point of illness, and random point in time). Due to project constraints relating to much of the data predating implementation of our electronic medical record, we could not control for age, gender, diagnosis, or other similar factors. An important consequence of this limitation is that the absolute performance characteristics are not applicable in a real-world setting. However, our study was focused on comparing relative differences between time series based and multivariate based models, and to that end we feel this limitation was acceptable.

Finally, derangements in physiology leading to a cardiac arrest have different time scales: some patients may deteriorate from progressive shock slowly over hours from a slow bleed while others may deteriorate over seconds from something like an unplanned extubation. We expect that the work presented here is more likely to predict the first case rather than the second – for which bedside monitors already do an adequate job in most cases.

We believe the most meaningful finding in this study is that the trend analysis from the time series data were more important in discriminating cases from controls than were the raw time series data. This finding supports extending the scope of time series analysis to generate other data classes for modeling cardiac arrest in a PICU.

Conclusions

A traditional model using multivariate data with a linear regression algorithm misclassified cases 3.7 times more frequently than one that used time series trend analysis with a support vector machine algorithm. Features based on clinical calculations did not improve model performance. Most vital signs changes occur within five minutes of an arrest, although more subtle drops may be noted as many as 20 minutes before an arrest. We have demonstrated how time series trend analysis can be used to improve the predictive accuracy of a clinical prediction model for cardiac arrest in a PICU. Although these findings have significant potential to improve identification of patients at risk for cardiac arrest, and to prevent death and disability by avoiding their occurrence, further refinements are needed to improve model specificity prior to application in a real-world setting.

Acknowledgments

Financial support: NIH NLM 5 K22 LM008389 Internal research funding by Baylor College of Medicine and Texas Children's Hospital

References

1. Meert KL, Donaldson A, Nadkarni V, et al. Multicenter cohort study of in-hospital pediatric cardiac arrest. *Pediatr Crit Care Med*. 2009; 10:544–553. [PubMed: 19451846]
2. Berg MD, Nadkarni VM, Zuercher M, et al. In-hospital pediatric cardiac arrest. *Pediatr Clin North Am*. 2008; 55:589–604. [PubMed: 18501756]
3. Zideman DA, Hazinski MF. Background and epidemiology of pediatric cardiac arrest. *Pediatr Clin North Am*. 2008; 55:847–859. [PubMed: 18675022]
4. de Mos N, van Litsenburg RRL, McCrindle B, et al. Pediatric in-intensive-care-unit cardiac arrest: Incidence, survival, and predictive factors. *Crit Care Med*. 2006; 34:1209–1215. [PubMed: 16484906]

5. Kause J, Smith G, Prytherch D, et al. A comparison of antecedents to cardiac arrests, deaths and emergency intensive care admissions in Australia and New Zealand, and the United Kingdom--the ACADEMIA study. *Resuscitation*. 2004; 62:275–282. [PubMed: 15325446]
6. Bharti N, Batra YK, Kaur H. Paediatric perioperative cardiac arrest and its mortality: Database of a 60-month period from a tertiary care paediatric centre. *Eur J Anaesthesiol*. 2009; 26:490–495. [PubMed: 19300269]
7. Schein RM, Hazday N, Pena M, et al. Clinical antecedents to in-hospital cardiopulmonary arrest. *Chest*. 1990; 98:1388–1392. [PubMed: 2245680]
8. Reis AG, Nadkarni V, Perondi MB, et al. A prospective investigation into the epidemiology of in-hospital pediatric cardiopulmonary resuscitation using the international Utstein reporting style. *Pediatrics*. 2002; 109:200–209. [PubMed: 11826196]
9. Egddell P, Finlay L, Pedley D. The PAWS score: Validation of an early warning scoring system for the initial assessment of children in the emergency department. *Emerg Med J*. 2008; 25:745–749. [PubMed: 18955610]
10. Hodgetts TJ, Kenward G, Vlachonikolis IG, et al. The identification of risk factors for cardiac arrest and formulation of activation criteria to alert a medical emergency team. *Resuscitation*. 2002; 54:125–131. [PubMed: 12161291]
11. Subbe C, Kruger M, Rutherford P, et al. Validation of a modified early warning score in medical admissions. *QJM*. 2001; 94:521–526. [PubMed: 11588210]
12. Pollack MM, Patel KM, Ruttimann UE. PRISM III: An updated pediatric risk of mortality score. *Crit Care Med*. 1996; 24:743–752. [PubMed: 8706448]
13. Pollack MM, Ruttimann U, Getson PR. Pediatric risk of mortality (PRISM) score. *Crit Care Med*. 1988; 16:1110–1116. [PubMed: 3048900]
14. Leteurte S, Martinot A, Duhamel A, et al. Validation of the paediatric logistic organ dysfunction (PELOD) score: Prospective, observational, multicentre study. *Lancet*. 2003; 362:192–197. [PubMed: 12885479]
15. Tontisirin N, Armstead W, Moore A, et al. Change in cerebral autoregulation as a function of time in children after severe traumatic brain injury: a case series. *Childs Nerv Syst*. 2007; 23(10):1163–9. [PubMed: 17429655]
16. Nadon R, Shoemaker J. Statistical issues with microarrays: Processing and analysis. *Trends Genet*. 2002; 18:265–271. [PubMed: 12047952]
17. Murphy D. Gene expression studies using microarrays: Principles, problems, and prospects. *Adv Physiol Educ*. 2002; 26:256–270. [PubMed: 12443997]
18. Dreiseitl S, Ohno-Machado L. Logistic regression and artificial neural network classification models: A methodology review. *J Biomed Inform*. 2002; 35:352–359. [PubMed: 12968784]
19. Rady MY. The role of central venous oximetry, lactic acid concentration and shock index in the evaluation of clinical shock: A review. *Resuscitation*. 1992; 24:55–60. [PubMed: 1332162]
20. Fuhrman, BP.; Zimmerman, JJ. *Pediatric critical care*. 2nd edition. Mosby, Inc; St. Louis, MO: 1998.
21. Berlot G, Pangher A, Petrucci L, et al. Anticipating events of in-hospital cardiac arrest. *Eur J Emerg Med*. 2004; 11:24–28. [PubMed: 15167189]
22. Tibballs J, Kinney S, Duke T, et al. Reduction of paediatric in-patient cardiac arrest and death with a medical emergency team: Preliminary results. *Arch Dis Child*. 2005; 90:1148–1152. [PubMed: 16243869]
23. Naeem N, Montenegro H. Beyond the intensive care unit: A review of interventions aimed at anticipating and preventing in-hospital cardiopulmonary arrest. *Resuscitation*. 2005; 67:13–23. [PubMed: 16150531]
24. Chen Y. Pediatric in--intensive-care-unit cardiac arrest: New horizon of extracorporeal life support. *Crit Care Med*. 2006; 34:2702–2703. [PubMed: 16983285]
25. Sharek PJ, Parast LM, Leong K, et al. Effect of a rapid response team on hospital-wide mortality and code rates outside the ICU in a children's hospital. *JAMA*. 2007; 298:2267–2274. [PubMed: 18029830]
26. Fineberg SL, Arendts G. Comparison of two methods of pediatric resuscitation and critical care management. *Ann Emerg Med*. 2008; 52:35–40. e13. [PubMed: 18407376]

27. Topjian AA, Nadkarni VM, Berg RA. Cardiopulmonary resuscitation in children. *Curr Opin Crit Care*. 2009; 15:203–208. [PubMed: 19469022]
28. Kinney S, Tibballs J, Johnston L, et al. Clinical profile of hospitalized children provided with urgent assistance from a medical emergency team. *Pediatrics*. 2008; 121:e1577–1584. [PubMed: 18519463]
29. Akre M, Finkelstein M, Erickson M, et al. Sensitivity of the pediatric early warning score to identify patient deterioration. *Pediatrics*. 2010; 125:e763–769. [PubMed: 20308222]
30. Edwards ED, Powell CVE, Mason BW, et al. Prospective cohort study to test the predictability of the Cardiff and Vale paediatric early warning system. *Arch Dis Child*. 2009; 94:602–606. [PubMed: 18812403]
31. McGaughey J, Alderdice F, Fowler R, et al. Outreach and early warning systems (EWS) for the prevention of intensive care admission and death of critically ill adult patients on general hospital wards. *Cochrane Database Syst Rev*. Jul 18.2007 (3):CD005529. [PubMed: 17636805]
32. Graham KC, Cvach M. Monitor alarm fatigue: Standardizing use of physiological monitors and decreasing nuisance alarms. *Am J Crit Care*. 2010; 19:28–34. [PubMed: 20045845]
33. Orphanidou C, Clifton D, Khan S, et al. Telemetry-based vital sign monitoring for ambulatory hospital patients. *Conf Proc IEEE Eng Med Biol Soc*. 2009; 2009:4650–4653. [PubMed: 19963616]
34. Hravnak M, Edwards L, Clontz A, et al. Defining the incidence of cardiorespiratory instability in patients in step-down units using an electronic integrated monitoring system. *Arch Intern Med*. 2008; 168:1300–1308. [PubMed: 18574087]
35. Ismail F, Davies M. Integrated monitoring and analysis for early warning of patient deterioration. *Br J Anaesth*. 2007; 98:149–150. [PubMed: 17158134]
36. Kunadian B, Morley R, Roberts AP, et al. Impact of implementation of evidence-based strategies to reduce door-to-balloon time in patients presenting with STEMI: Continuous data analysis and feedback using a statistical process control plot. *Heart*. 2010; 96:1557–1563. [PubMed: 20736208]
37. Kennedy CE, Turley JP. Time-series analysis as input for clinical predictive modeling: modeling cardiac arrest in a pediatric ICU. *Theor Biol Med Model*. 2011; 8:40. [PubMed: 22023778]

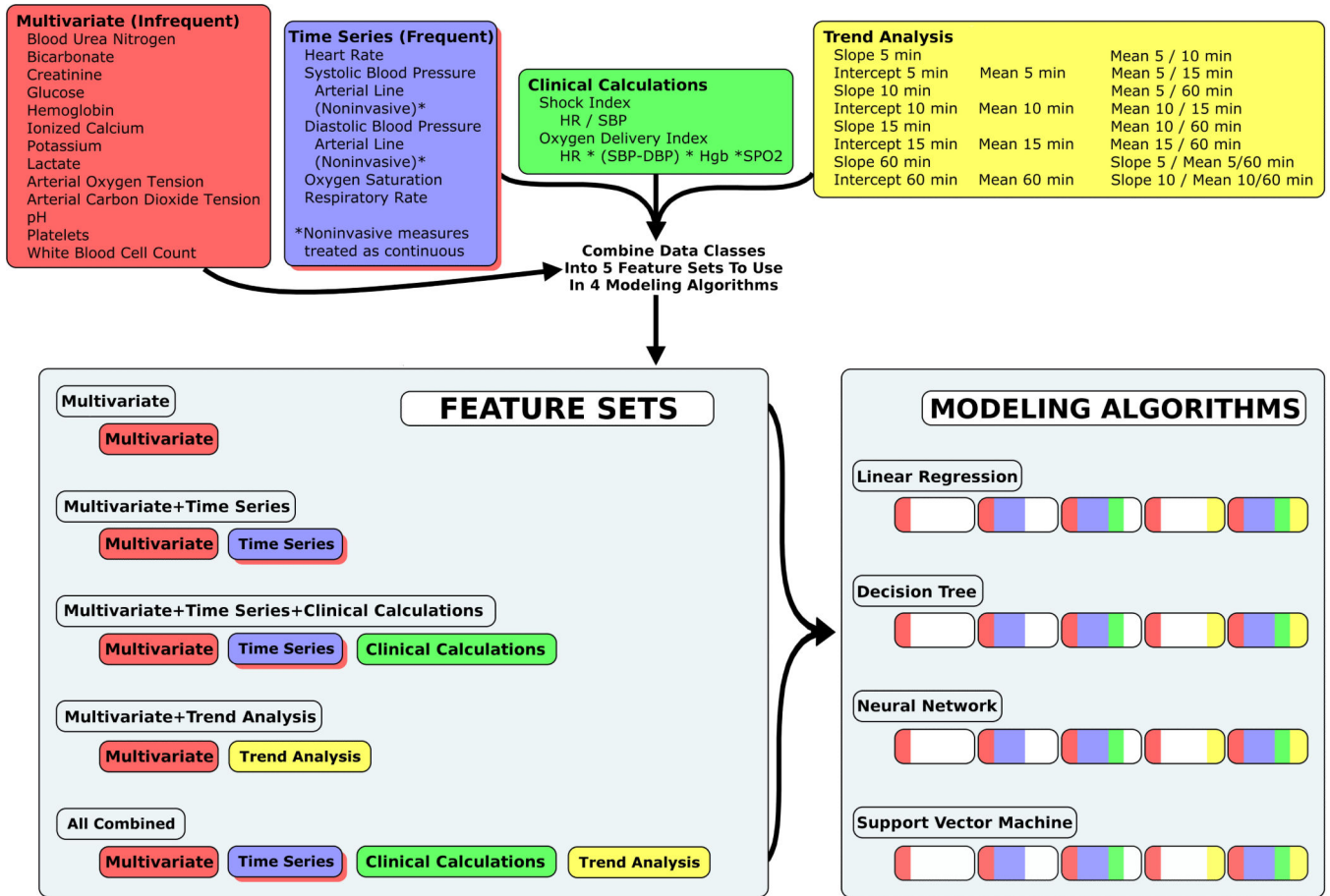


Figure 1. Four classes of data were combined into five feature sets. Each feature set was used to train and test cardiac arrest prediction models using four modeling algorithms. Of the 1025 variables, 20 were multivariate, 497 were raw time series, 288 were clinical calculations, and 220 were time series trend analysis. Differential performance between the Multivariate feature set with the Linear Regression algorithm and other models measured: 1) effects attributable to data class (for a given algorithm); or 2) effects attributable to modeling algorithm (for a given data class).

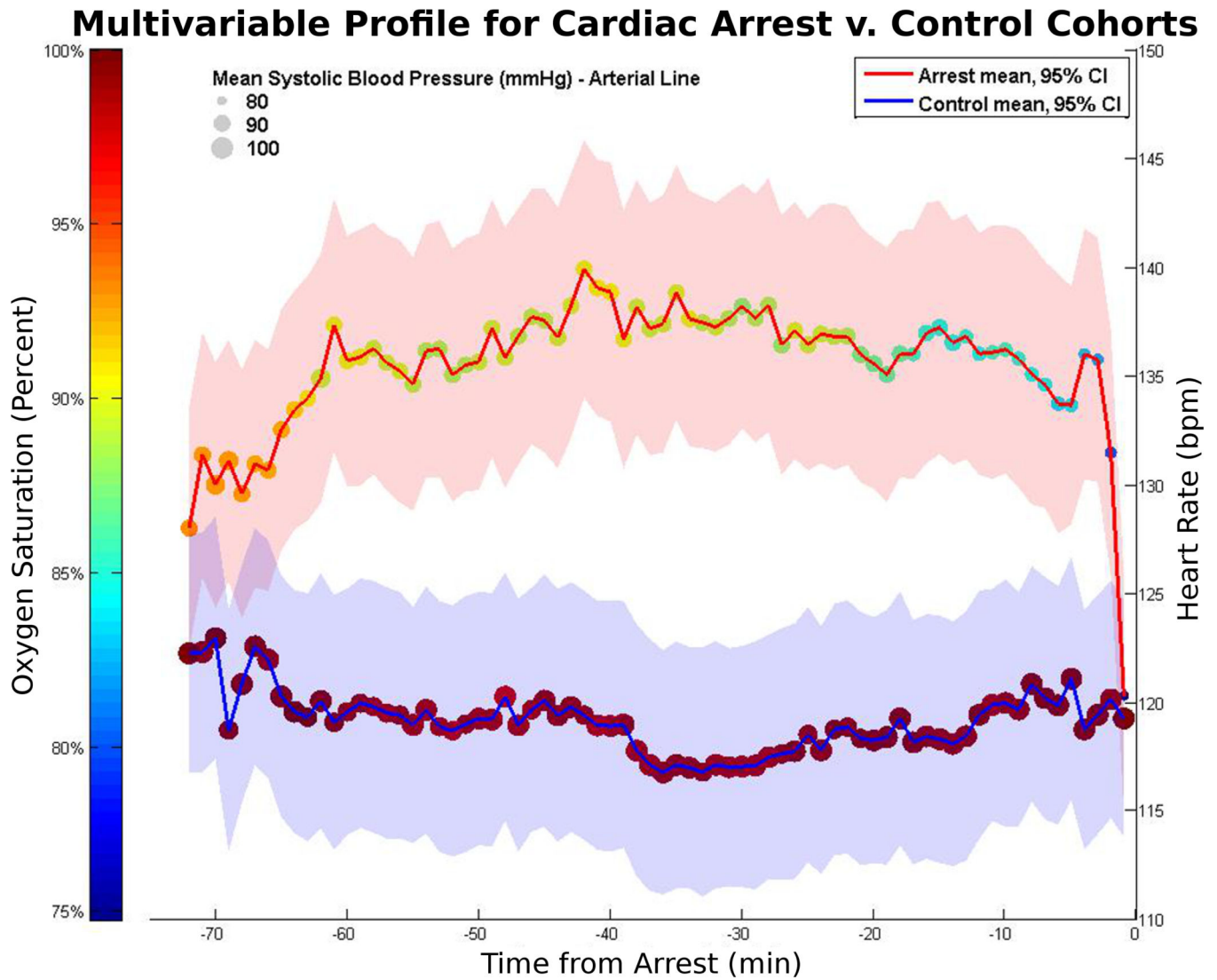


Figure 2.

Mean values for heart rate (position on y-axis with shaded 95% SEM), oxygen saturation (dot color), and systolic blood pressure (dot size) are shown for arrest (red line) and control (blue line) subjects over the span of 12 hours. Left to right, the first 12 values represent hourly averages, and the last 60 values represent minute-by-minute measurements. Heart rate indicators occurred very close to the arrest event (acute drop 1-2 minutes beforehand). Pulse oximetry and blood pressure indicators started noticeable trends downward at 15-20 minutes beforehand.

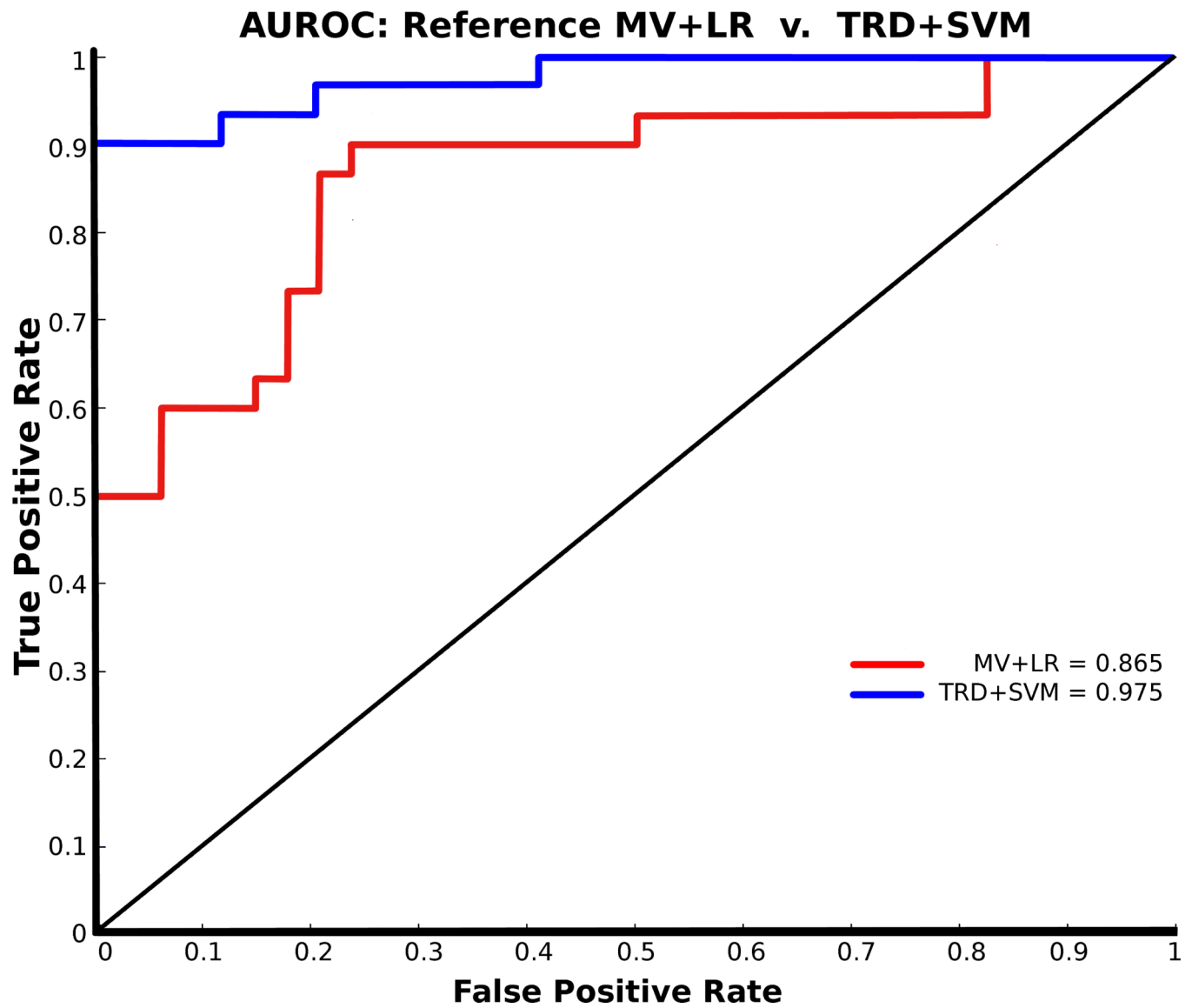


Figure 3.

Area under the receiver operating characteristic curve (AUROC) for reference model (MV+LR: 0.865) (red) and for best model (TRD+SVM: 0.975) (blue). The performance increase is attributable to the combined effects of adding trend analysis and using the SVM algorithm. Performance suffered if either change was made in isolation.

Visualization: SVM Features Weighted > 0.5

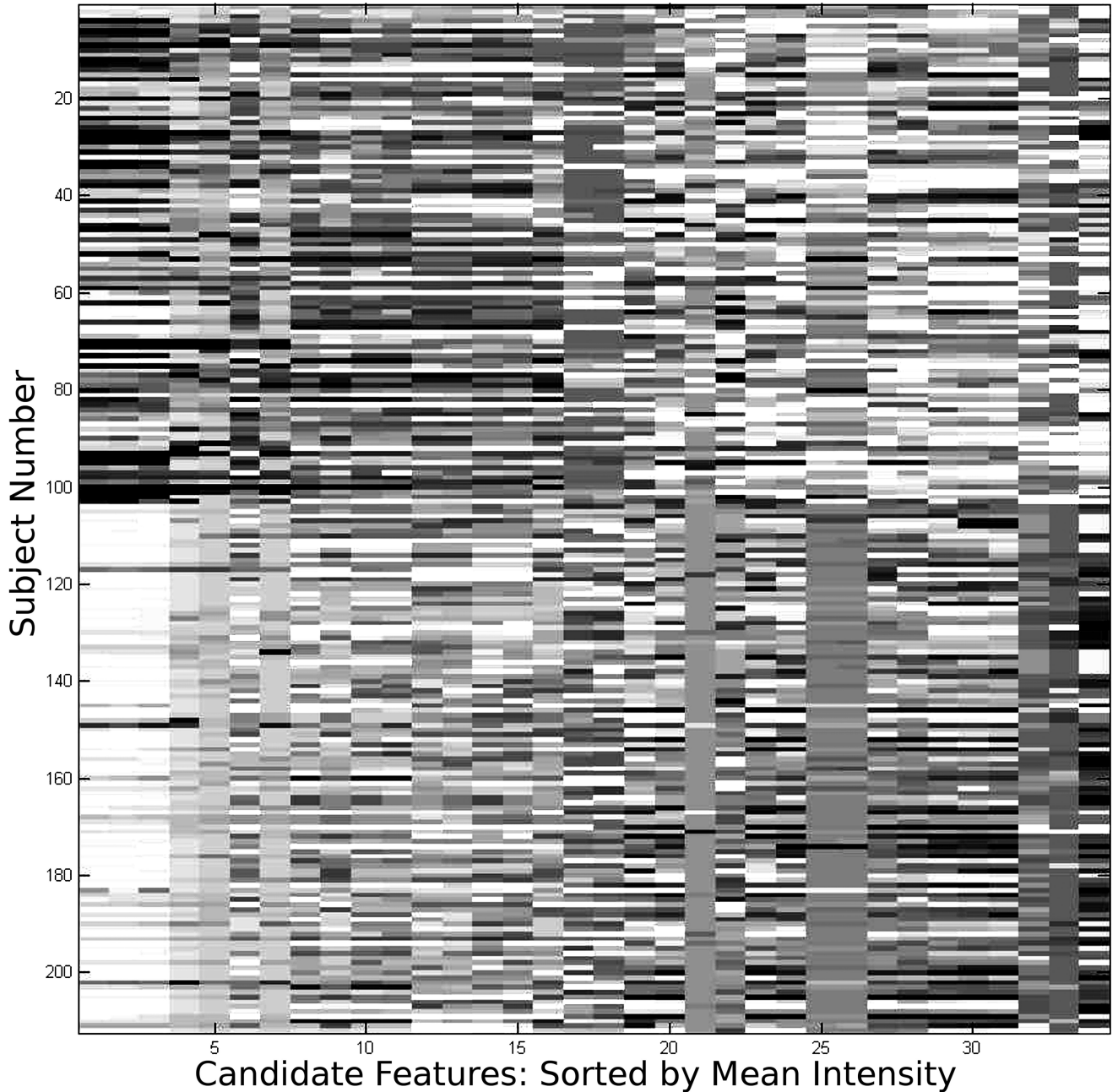


Figure 4.

Top 35 variables determined by support vector machine weighting. The columns are arranged in descending order of mean variable intensity of the arrest cases. However, the order does not reflect the weight of the variable in predicting arrest. Visually, the first four variables are highly correlated, and represent SPO2 values preceding the arrest: at t minus 1, 2, 7, and 10. These raw SPO2 values did not rank in the top 5 weights, but two of the top five weights were trend calculations involving SPO2: 5 minute intercept, and 60 minute

slope. Visually, arrest cases tend to transition from dark to light, whereas control cases tend to transition from light to dark.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Measures of model performance in validation data set. Mean performance of modeling algorithm (rows) or feature set (columns) is listed outside the table, with the best measure in **boldface**. Inside the table, reference performance for MV feature + LR algorithm is shown in the top left, underlined. Models using trend calculations without raw time series features (the TRD data class) along with the SVM algorithm demonstrated the best overall performance. Removing single measurement variables not subject to trend calculations did not significantly impact model performance.

Matrix of Model Performance in Validation Data						
All Available Relevant Variables						
Accuracy	MV	TS	CC	TRD	ALL	
LR	<u>78%</u>	69%	58%	77%	69%	70%
DT	72%	88%	80%	81%	86%	81%
NN	67%	63%	67%	72%	66%	67%
SVM	77%	89%	88%	94%	88%	87%
	73%	77%	73%	81%	77%	
AUROC	MV	TS	CC	TRD	ALL	
LR	<u>87%</u>	72%	60%	72%	80%	74%
DT	74%	90%	79%	83%	88%	83%
NN	82%	83%	80%	92%	79%	83%
SVM	82%	95%	96%	98%	97%	94%
	81%	85%	79%	86%	86%	

Variables Constrained to Trend Calculation Compatible						
Accuracy	MV	TS	CC	TRD	ALL	
LR	83%	78%	----	87%	84%	83%
DT	81%	81%	----	81%	86%	82%
NN	64%	56%	----	72%	64%	64%
SVM	73%	84%	----	94%	91%	86%
	75%	75%	----	84%	81%	
AUROC	MV	TS	CC	TRD	ALL	
LR	89%	79%	----	87%	83%	85%
DT	81%	81%	----	80%	86%	82%
NN	79%	74%	----	92%	82%	82%
SVM	83%	91%	----	98%	97%	92%
	83%	81%	----	89%	87%	

Abbreviations:

AUROC = area under the receiver operating characteristic curve.

Data Classes:

MV = multivariate

TS = multivariate + time series

CC = multivariate + time series + clinical calculations

TRD = multivariate + trend analysis

ALL = all 1025 variables - 17

Model classes:

LR = linear regression

DT = decision tree

NN = neural network

SVM = support vector machine

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript