AFRL-SA-WP-TR-2013-0019

Computational Gene Mapping to Analyze Continuous Automated Real-Time Vital Signs Monitoring Data

Lynn Stansbury, MD, MPH; Catriona Miller, PhD; Shiming Yang, PhD; Peter Hu, PhD; Hegang Chen, PhD; Deborah Stein, MD, MPH; Raymond Fang, Col, USAF, MC

September 2013

Final Report
for June 2012 to June 2013

Air Force Research Laboratory
711th Human Performance Wing
School of Aerospace Medicine
Air Force Expeditionary Med Skills Inst
C-STARS Baltimore
2510 Fifth St.
Wright-Patterson AFB, OH 45433-7913

NOTICE AND SIGNATURE PAGE

AFRL-SA-WP-TR-2013-0019 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

//SIGNATURE//	//SIGNATURE//
COL RAYMOND FANG, USAF, MC	COL BENJAMIN A. HARRIS, USAF, MC
Director, C-STARS Baltimore	Chair, Air Force Expeditionary Med Skills Inst

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From – To)*
23 Sep 2013	Final Technical Report	June 2012 – June 2013

4. TITLE AND SUBTITLE

Computational Gene Mapping to Analyze Continuous Automated Real-Time Vital Signs Monitoring Data

5a. CONTRACT NUMBER
FA8650-12-2-6D08

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Lynn Stansbury, MD, MPH; Catriona Miller, PhD; Shiming Yang, PhD; Peter Hu, PhD; Hegang Chen, PhD; Deborah Stein, MD, MPH; Raymond Fang, Col, USAF, MC

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
USAF School of Aerospace Medicine
Air Force Expeditionary Medical Skills Institute
C-STARS Baltimore
2510 Fifth St.
Wright-Patterson AFB, OH 45433-7913

8. PERFORMING ORGANIZATION REPORT NUMBER

AFRL-SA-WP-TR-2013-0019

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSORING/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Distribution A: Approved for public release; distribution is unlimited. Case Number: 88ABW-2013-5295, 13 Dec 2013

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This project explored machine learning "gene mapping" algorithms (MLA) as possible analytic platforms for advanced, forward-deployable patient care instrumentation. A patient database was developed to 1) identify physiologic data collected electronically by continuous automated critical care patient monitoring, 2) derive clinically useful vital signs (VS) "features" of potential use in predicting patient functional outcome after severe brain trauma, 3) define outcomes, 4) train and test machine learning algorithms, and 5) cross-validate and finalize results. The computer-based experimental work developed in three stages: 1) feature selection by conventional univariate methodology, 2) feature selection using logistic regression, and 3) feature selection using other procedures that could modulate the limitations of the first two approaches. Stage 1 results suggested that long-term patient outcomes may be predictable using MLA and VS features gathered within the first 12 hours of critical care. Stage 2 results were much stronger but showed signs of "overfitting," an inherent pitfall of feature selection using logistic regression. Stage 3 results confirmed Stage 1 and 2 results with much better correlations, particularly for early (<6 weeks post discharge) and late (3-6 months) patient functional outcomes after severe brain trauma and using only the first 12 hours of continuous monitoring data.

15. SUBJECT TERMS

Continuous automated patient monitoring, vital signs, forward-deployed, outcomes, severe brain trauma, gene mapping, algorithms

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Col Raymond Fang
U	U	U	SAR	21	19b. TELEPHONE NUMBER *(include area code)*

This page intentionally left blank.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1.0 SUMMARY

The overall aim of this project was to explore the kinds of computer-based machine learning algorithms (MLA) used in "gene mapping" as possible analytic platforms for advanced and ultimately forward-deployable patient care instrumentation capable of assisting in the transport, triage, and care of casualties with traumatic brain injury (TBI).

From sequential admissions to our regional adult neuro-trauma referral center, part of the R Adams Cowley Shock Trauma Center in Baltimore, MD, a baseline study group of 191 adult (>17 years) patients was identified with TBI severe enough to require intracranial pressure (ICP) monitoring and continuous electronic data collection of all vital signs (VS) of interest (including ICP and ICP-derived indices) that was stored for at least the first 12 hours of hospital-based critical care. Short-, intermediate-, and long-term outcomes were identified for these patients.

This patient-care dataset was then further developed to 1) define patient outcomes of interest after severe TBI, 2) derive clinically useful VS "features" (akin to the identification of amino acid sequences of interest as "genes" in classic gene-mapping techniques) of potential use in predicting these outcomes, 3) train and test MLAs, and 4) cross-validate and finalize results.

In the first stages of work, 588 VS features of potential utility were identified. These were derived from eight patient VS waveforms and digital inputs routinely collected in the neuro-trauma intensive care setting and linked to various established clinical thresholds, time frames, and other aspects of dose. These features were then sub-selected for application in the MLAs using three different approaches. The first approach used conventional univariate methodology to select VS features with potential to predict outcomes. This method is sensitive and commonly used, but can miss critical interactions between features. The results of this approach with the study data were not strong, but were published in 2012 and did suggest that high quality, electronically dense, continuous automated data collected in the first 12 hours of hospital-based critical care do have potential to predict long-term functional outcome after severe TBI. The second approach used multivariate logistic regression (MLR) for feature selection, another commonly used method. Features identified in this way produced much stronger correlations with prediction of functional outcome. However, feature selection using MLR often "overfits" the model to the dataset. An overfitted model may not perform well when faced with novel, dynamic, incoming real-time data, which are characteristic of clinical data in field situations. Data processing and analysis platforms for field-ready instrumentation must be able to cope with such data, which tend to be qualitatively and quantitatively quite different from the static pool of patient-care data used in experimental modeling, even when appropriate "testing" and "training" procedures are used. The third approach explored several novel weighting procedures aimed at optimizing selectivity while remaining open to a wider range of potentially useful features than does MLR. These approaches included recursive feature elimination, greedy pairs algorithm, lasso for 10 features, lasso for 20 features, and elastic net. Unlike other potential alternative novel approaches, these approaches tend to be computationally efficient and have good potential for miniaturized, field-ready systems. Results using these additional approaches confirmed the overall results of the first two approaches and had strong correlations for early (<6 weeks post discharge) and late (3-6 months) patient functional outcomes after severe TBI.

We have found that MLA algorithms, particularly recursive feature elimination and elastic net, using weighted feature selection from the first 12 hours of continuous neuro-trauma intensive care monitoring can predict long-term functional outcomes after TBI and have potential

to be used in analytic platforms for advanced, field-ready patient care and decision-assist instrumentation.

2.0 INTRODUCTION

This report details the results of an effort to explore, develop, and test machine learning algorithms (MLA) of potential use as future analytic platforms for advanced, field-ready decision-assist instrumentation of use in the triage, transport, and monitoring of casualties with severe traumatic brain injury (TBI).

3.0 BACKGROUND

3.1 Trauma Epidemiology

Traumatic brain injury is the most common cause of emergency care admission and trauma-related death in the U.S. civilian population [1] and a major cause of death and disability in combat casualties [2]. Because of the frequency of TBI and its fatality rate and profound impact on survivors' quality of life, much research has focused on the development of early-warning decision-assist systems that can maximize the potential for timely therapeutic interventions to improve long-term clinical outcomes. Ideally, these systems would also be sufficiently reliable, robust, and miniaturizable for field deployment. Such systems will depend on sophisticated computer-based analytic platforms. Identification and testing of such platforms are a priority.

3.2 Machine Learning Algorithms in the Analysis of Large Patient Databases

Computer-based, high-information-throughput techniques have been used for years to perform micro-array gene mapping and have derived useful information out of vast streams of data [3-5]. These techniques assess the significance of repeated sequences of amino acids in DNA ("genes") in relation to tumors and tumor response to chemotherapy. These techniques have important potential for interpreting the huge quantities of raw, real-time, automated electronic clinical monitoring data generated by modern critical care—most of which is now wasted—and for integrating these data into individualized, real-time, valid, and useful critical care bedside instrumentation.

3.3 Preliminary Studies

The study team previously demonstrated the superiority of automated versus manual vital signs (VS) data collection and processing systems in providing data on patients with severe TBI and the power of calculating a pressure-times-time "dose" (PTD/D) of intracranial pressure (ICP) and cerebral perfusion pressure (CPP) [6,7]. Using receiver operating characteristic (ROC) techniques, prognostic algorithms were developed correlating VS-related features derived from routine neuro-trauma intensive care electronic monitoring with 30-day mortality and Glasgow Outcome Score-Extended (GOSE) [8] at 3 and 6 months. These algorithms were then incorporated into real-time two-dimensional graphic displays of ongoing calculations of Shock Index (SI=systolic blood pressure (SBP)/heart rate (HR)] and brain trauma index

[BTI=(CPP/ICP)*time]. This prototype patient monitoring video display system is now deployed on a translational basis throughout the R Adams Cowley Shock Trauma Center (STC) in Baltimore, MD (Figure 1, far right, upper and lower panels, respectively).

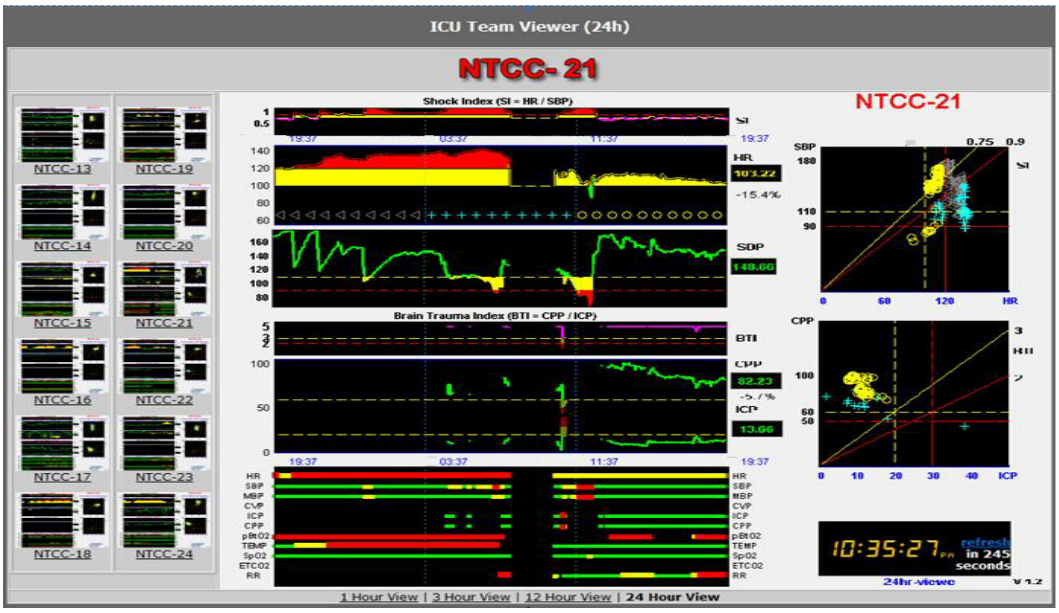

Figure 1. Real-Time Bedside and Telemetric Critical Care Monitoring Display

The BTI graph in the bottom right hand corner of Figure 1 allows for the tracking and visual display of head-injury status. Data point clusters in the left upper quadrant (ICP<20 mmHg and CPP>60 mmHg) are associated with the best outcomes, left lower and right upper quadrants with relatively poorer outcomes, and the lower right quadrant (ICP≥20 mmHg and CPP<60 mmHg) with significantly worse outcomes. This display allows clinicians to track and monitor shifts in patients' status over the previous 12 and 24 hours in a single real-time display linked to predicted outcome rather than just conventional single-parameter threshold readouts. As well as the two indices noted, SI and BTI, VS thresholds of interest in this work were SBP, mean arterial pressure (MAP), HR, ICP, CPP, and oxygen saturation (SpO_2).

4.0 METHODS

4.1 Data Sources: Patient Selection

This work was undertaken as part of the protocol approved by the University of Maryland School of Medicine Human Research Protections Office for intensive monitoring after severe TBI. Included were adult patients (older than 17 years) admitted to the R Adams Cowley STC, Baltimore, Maryland, with Glasgow Coma Score (GCS) <9 and a clinically determined requirement for ICP monitoring. The nature of these patients' injuries precluded personal informed consent; therefore, informed consent was secured from a legally authorized representative prior to study inclusion and from the patient as soon as and if that became possible. Patients with severe multitrauma (more than one non-head abbreviated injury score >3) were excluded.

4.2 Data Sources: Patient Records

The demographics, mechanism of injury, injury scoring data, admission VS, and laboratory data on all trauma patients admitted to the STC are recorded by our trauma registry. Outcome measures available through the registry include in-hospital mortality, length of stay (LOS) in the hospital and intensive care unit (ICU), and discharge disposition (home with or without additional services or various levels of extended care). At 3 and 6 months, GOSE of survivors was assessed in structured phone interviews by an experienced trauma clinical research coordinator. A GOSE score between 1 and 4 was defined as a "poor functional outcome," while a GOSE of 5-8 was considered a "good functional outcome." All outcome data were reviewed and assessed by the principal investigator.

4.3 High-Resolution Automated Data Collection

VS data collection for this project was initiated when an ICP monitoring device was placed in either the trauma resuscitation unit or the ICU. The analyses for this study—feature selection, modeling, and cross-validation steps—were done using National Cancer Institute free software BRB-ArrayTools, Version 4.2. Details of the electronic data capture, storage, and data-point assembly procedures used in general by this study team to construct vital signs signal sequences for analysis have been published previously [9] and are summarized here. All ICU patient monitors at the STC are networked to capture incoming electronic data every 6 seconds. Data are then compressed and transferred to a centralized VS data recorder server through a secured intranet. Potential artifacts and defined extreme outliers are filtered via a moving median window process. ICP readings distorted by periodic drainage are corrected using the piecewise cubic Hermite interpolation method (Matlab 7.7 R2008b, Mathworks, Natick, MA). Together, these processes discard less than 1% of data points. Five- minute means are calculated as noted above. Data are reviewed by a physician to ensure clinical validity.

4.4 Identification of Critical Time-and-Threshold Vital Signs Signal Sequences

The general approach to the use of advanced MLA for this work was similar to that used for microarray studies. However, the source data for feature selection ("gene" identification) were not the kinds of biomolecular samples addressed by the current MIAME (Minimum Information About a Microarray Experiment) standards, but were virtual constructs from electronic data points summarized from routine ICU VS monitoring. From electronic VS data, recorded, compressed, filtered, and stored as above, we developed machine learning "features," potential discriminator vital signs signal (VSS) sequences, via the following steps.
We focused on the following categories of VS, as previous work noted above suggested they would prove most useful:

- Brain trauma/vascular-pressure-related: ICP, CPP, SBP, MAP, BTI
- Cardiac/shock-related: HR, SBP, Shock Index (SI=HR/SBP)
- Perfusion-related: SpO_2

These VSS were then characterized via conventional clinical thresholds: ICP>20 and >30 mmHg; CPP<50, <60, and >100 mmHg; SBP<90, <100, <110, and <120 mmHg; MAP<60 and <70 mmHg; BTI<1.67, <2.0, and <3.0; HR>100, >110, and >120 bpm; SI>0.7, >0.8, >0.9, and >1.0; and SpO_2 <88% and <90%. Maximum, minimum, and mean ICP and CPP PTD/Ds were also characterized. Finally, the various threshold sequences were linked to time, that is, periods and proportions of time for VS and index segments above or below defined limits: greater than, less than, or equal to 5, 10, 15, 20, 25, 30, 45, and 60 minutes (Figure 2).

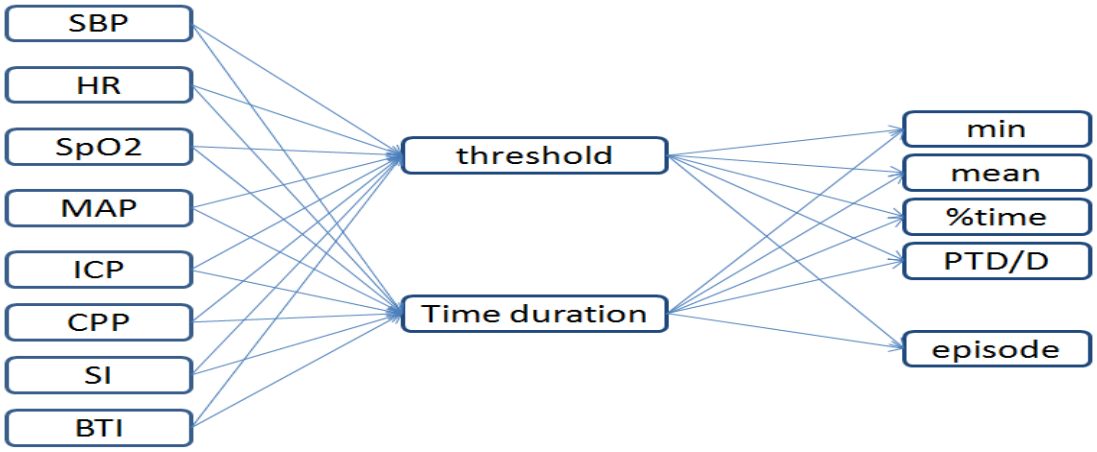

Figure 2. Schematic of Feature Design

To quantify this link, an episode was defined as one count when a value or an index of a VS remained above or below a pre-set threshold for a certain duration. For example, every interval in the first 12 hours of VS data collection where SpO_2 remained at or below 92% for 10 minutes or more was counted as one episode.

This identification and sorting process yielded 588 time-and-threshold variables that could be evaluated for their potential utility as "features" in algorithms examining outcome prediction over the first 12, 24, and 48 hours after ICP monitor placement.

4.5 Feature Selection for Class Prediction

In classic structured machine learning algorithms, features are the variables selected to construct the algorithms. In preliminary testing, features are selected that appear to have the greatest likelihood, when used in the actual algorithm, of supporting correct prediction of selected binary outcomes. They are then reassessed within the algorithm for performance when faced with novel data. Feature selection is often viewed as more important than the specific class prediction model used [10,11]. In the work reported here, three approaches to feature selection, univarite selection, logistic regression, and the "elastic net" method [12], were assessed for performance in supporting the correct prediction of outcome and for potential utility in field applications.

4.5.1 Univariate Feature Selection. A common approach to feature selection is univariate testing of differences in the ability of each variable to correctly identify the outcome classes compared with each other variable. For this work, these outcomes were life or death; being in-hospital at 14 days or in ICU at 14 days, yes/no; and good/bad GOSE at 3 and 6 months. For each outcome, a random variance t-test [3] was used to sequentially compare the performance of the mean representing each potential feature against the mean representing each other potential feature in correctly identifying the selected outcomes, with a p-value of <0.05. This sifting process demonstrated that several of the threshold variables and/or groupings under consideration were not workable. Specifically, 6-month GOSE scores were not available for sufficient numbers of study subjects at the time when this stage of work was undertaken to use these as outcome class labels. Likewise, the first 12 hours after ICP placement provided the earliest potentially clinically useful information (for example, identified increased risk of death well in advance of the event rather than immediately before it occurred).

Using this subset of VSS time-and-threshold variables identified as critical features for the class prediction analysis, six different prediction models were built: compound covariate predictor, linear discriminant analysis, one-nearest-neighbor classifier, three-nearest-neighbor classifier, nearest-centroid classifier, and support vector machines [4,13,14].

4.5.2 Logistic Regression Feature Selection. Using conventional ROC area under the curve (AUC) for prediction of good/bad outcome at 6 weeks and 3, 6, and 12 months after discharge and the pool of features described above, a logistic regression model was built. To test the ability of the regression model to absorb the accrual of new data, training and testing were carried out using a classic, 10-fold-times-10 procedure and 75% of the data as the training set and 25% of the data as the testing set.

4.5.3 Feature Selection Using Various Weighting Methods. The recursive feature elimination (RFE) [15] technique uses a support vector machine as the training algorithm and recursively eliminates irrelevant variables, as measured by certain score functions [11]. The greedy pairs algorithm evaluates genes in pairs and assesses how well a pair in combination distinguishes two experimental classes [16]. (In genetics, these "gene" features are amino acid sequences derived from subject nucleic acids. In our work, as discussed above, these features are time/threshold sequences identified from clinical electronic VS recordings.) The "lasso" method was proposed by Tibshirani to achieve sparse solutions for feature selection [17]. This adds an *l*-1 penalty term expressed as

$$\lambda||w||_1$$

which weights the coefficients of the less useful potential features toward zero. In our work with the lasso method, we set our parameters to identify no more than 10 features (L10) or 20 features (L20). However, Zou and colleagues [12] have shown that the lasso method is limited in that it tends to select one feature from each group of highly correlated variables. These researchers proposed to add an *l*-2 norm penalty term to avoid such limitation. This method is known as "the elastic net" and is expressed as

$$penalty = \alpha_1||w||_1 + \alpha_2||w||_2^2$$

It has the effect of compromising between the overselectivity of the lasso method and the occasional inclusion of physiologically impossible variable coefficients.

For the RFE and greedy pairs algorithms, we used the BRB-ArrayTools [13], a comprehensive analysis tool for microarray. For other feature selection methods, we used the R packages.

For analytic purposes, at this point in the overall project, the baseline study group (which then comprised all eligible patients admitted from January 2008 – December 2010) was then subgrouped by hospital survival as Group 1, all of whom survived to discharge, and Group 2, all patients except those who died in-hospital after their families elected to withdraw care). This was done as an attempt to distinguish feature characteristics of those who survived to discharge and on into long-term follow-up and those who did not.

4.6 Cross-Validation of Prediction Models

The literature on methods for developing multivariate predictors of class membership (yes/no membership in an outcome class) is a large one [18,10,11], but the goal is to construct a classifier, the mathematical tool, that will accurately classify incoming individuals not involved in creation of the prediction rule. Estimation of the prediction error of each model requires an approach that will avoid the overfitting inherent in using the same set of data to develop a predictive model and then to test its predictive accuracy—the model will always work best for the data from which it was built. To avoid this tautology and because our sample size was relatively small for this kind of work, we chose cross-validation via a leaving-one-out technique [5]. In this process, the selected testing data for each individual patient are sequentially omitted from the calculations. For each training set with one individual omitted, feature selection is done de novo. From these features, predictive models are built that assess the influence of individuals in the model. Then each model is rated as either correct or incorrect in predicting the outcome class of each individual. This procedure is repeated for each individual, and the mean percentage of correct classification is determined as an assessment of the overall validity of the model.

5.0 RESULTS

5.1 Using Univariate Feature Selection

At the time we did this work, 52 patient datasets were available for analysis. These yielded a total of 589 ICP monitor hours or 353,600 x 6 seconds of continuously collected VS records, which in turn permitted identification of the baseline time-and-threshold features of potential use in prediction of mortality, length of stay, and GOSE at 3 months. (As noted above, although much more data were potentially available, the fields for GOSE at 6 months were insufficiently populated to provide useful features, and the first 12 hours of ICP monitoring appeared to be the most clinically useful.) Of the 588 features that we constructed from these data, univariate analysis associated correct identification of any given outcome with as many as 76 features to as few as 4. In general, those representing ICP or BP over or under given thresholds over time (e.g., ICP >20 mmHg for 20 minutes) provided the best discrimination for outcome. As examples of the information being processed in the prediction analyses, Tables 1 and 2 list the features elected by the univariate analyses for the class prediction modeling for two outcomes—the ability to predict, by 12 hours into care, 3-month GOSE and mortality—and

which provided the best results in the subsequent cross-validation study. In both of these sets of features, cerebral and vascular pressure measurements and indices were more useful in predicting class—outcome—than were measurements of oxygen saturation.

Table 1. Features Selected by Univariate Analysis as Most Likely to Provide Useful Information at 12 Hours into Care Regarding Eventual Outcome = Death

VSS	VSS Threshold Features[a]
VSS A 01	mean ICP ≥30 mmHg, total number of episodes ≥1 h
VSS A 02	mean CPP ≤50 mmHg, PTD/D per day
VSS A 03	mean CPP ≥100 mmHg, total number of episodes = 25-30 min
VSS A 04	mean BTI ≤1.67, total number of episodes ≥1 h
VSS A 05	mean SBP ≤110 mmHg, total number of episodes = 10-15 min
VSS A 06	mean SBP ≤120 mmHg, total number of episodes ≥1 h
VSS A 07	mean HR ≥120 bpm, total number of episodes = 45-60 min
VSS A 08	mean SPO_2 ≤92%, total number of episodes ≥10 min
VSS A 09	mean SPO_2 ≤92%, total number of episodes = 10-15 min
VSS A 10	mean MAP ≤60 mmHg, total number of episodes = 10-15 min
VSS A 11	mean SI ≥0.9, total number of episodes = 45-60 min

[a]mean = 5-min means of every 6-s data collection; BTI = CPP/ICP dose (pressure-times-time); SI = SBP/HR.

Table 2. Features Selected by Univariate Analysis as Most Likely to Provide Useful Information at 12 Hours into Care Regarding Eventual Outcome = GOSE at 3 Months

VSS	VSS Threshold Features
VSS B 01	mean CPP ≤50 mmHg, total number of episodes = 25-30 min
VSS B 02	mean CPP ≤50 mmHg, total number of episodes = 30-45 min
VSS B 03	mean MAP ≤60 mmHg, total number of episodes = 15-20 min
VSS B 04	mean SI[a] ≥0.8, total duration of episodes = 20-25 min

[a]SI = SBP/HR.

Table 3 summarizes the results of the leave-one-out cross-validation using the six prediction models.

Table 3. Mean Percent of Correct Classification Using Various Methods and a Sequential "Leave-One-Out" Strategy

Predicted Outcomes	Compound Covariate Predictor Percent Correct	Diagonal Linear Discriminant Analysis Percent Correct	1-Nearest Neighbor Percent Correct	3-Nearest Neighbors Percent Correct	Nearest Centroid Percent Correct	Support Vector Machines Percent Correct
3-mo GOSE <5	52	54	38	40	52	58
4 days	62	60	56	63	63	62
ICU LOS ≥14 days	71	67	62	71	77	71
Mortality	69	75	87	88	69	81

5.2 Using a Logistic Regression Model

Complete data were available for analysis at the various outcome periods on 113-116 patients. ROC AUC for the logistic regression model were 0.85 at 6 weeks (n=113), 0.88 at 3 months (n=116), 0.90 at 6 months (n=115), and 0.92 at 12 months (n=113). Results for the training run were essentially the same as using all data, but results for the incoming "new" data were reduced by 5 to 10%.

5.3 Using Various Weighting Techniques in Feature Selection

By the time this portion of the work was done, the baseline study group comprised 191 patients. The subgroups for analysis comprised 148 patients in Group 1 and 176 patients in Group 2. Table 4 summarizes the demographic, admission, and hospital LOS for these patients by group.

Table 4. Demographic and Basic Admission Injury Scoring Data for All Patients and Patients Grouped by Outcome[a]

Demographic/Admission Data[b]	All	Group 1	Group 2
Age (yr), mean (±SD)	41.7 (18.5)	40.4 (18.0)	39.2 (17.2)
Males, n (%)	149 (78.0)	90 (72.0)	118 (76.6)
Admission GCS, mean (±SD)	6.9 (3.7)	6.9 (3.6)	7.1 (3.7)
Neuro GCS, mean (±SD)	6.8 (2.7)	6.9 (2.6)	7.0 (2.8)
LOS, total days, median (IQR)	15.8 (14.2)	15.6 (14.0)	15.8 (14.0)
Marshall, mean (±SD)	2.6 (0.8)	2.6 (0.8)	2.5 (0.8)

[a]Group 1 excluding all hospital deaths and Group 2 excluding only those deaths that occurred after a family decision to withdraw care.
[b]SD = standard deviation; IQR = interquartile range.

Figure 3 summarizes modeling and cross-validation results for Group 1 (excluding all hospital deaths, n=148) and Group 2 (excluding only deaths due to familial decision to withdraw care, n=176) using univariate selection and the various weighting procedures. The model prediction performance shows significant difference between RFE in the univariate discrimination-based selection method and the multivariate, feature-weighting selection methods. With multiple variable logistic regression, multivariate selection methods give more favorable selections by combining features to optimize multivariate performance. The five multivariate feature selection methods generated different subsets of features, with sizes ranging from 9 to 36.

With those selected features, we built simple logistic regression models with respect to GOSE outcomes of <6 weeks (early), 6 weeks to 3 months (mid), 3-6 months (late), and >6 months (long). The overall prediction performances ranged from 0.60 to 0.90, expressed as AUROC. For the lasso-based methods, the AUROCs located compactly between 0.70 and 0.85.

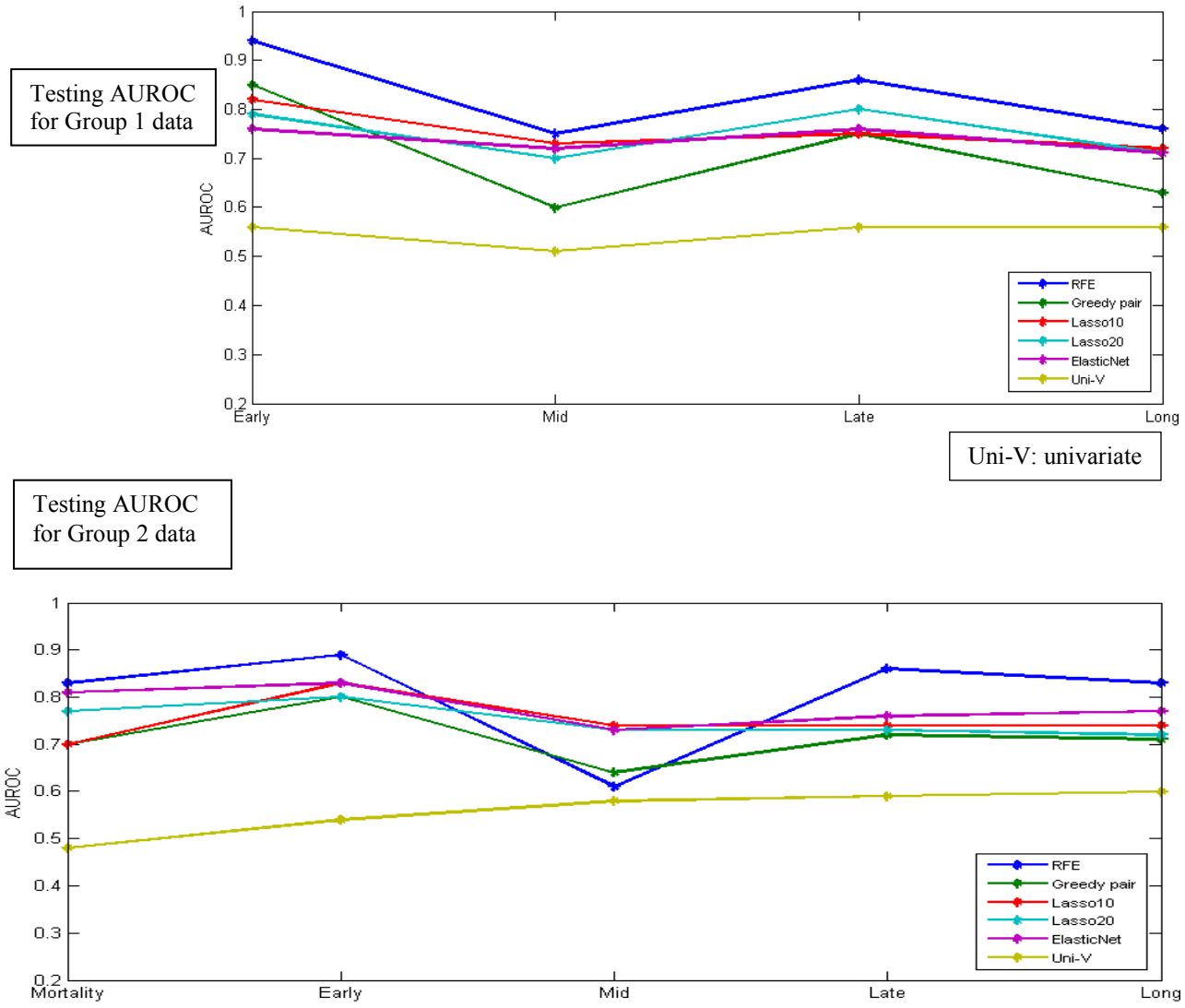

Figure 3. Performance of Selected Features in Logistic Model

As examples of the information being processed in this and in the univariate prediction analyses, Tables 5 through 8 list the five features most frequently selected by all of the selection methods for the four GOSE assessment periods. Again, cerebral and vascular pressure measurements and indices appear to be the most useful in predicting outcome.

Table 5. Five VS Features Most Frequently Selected by All Feature Selection Methods in Predicting Early (<6 Weeks Post Discharge) GOSE Using the First 12 Hours of Data

No.	Feature[a]
1	0 – 12, SI mean, ≥0.9, Episode of 20-25 min
2	0 – 12, CPP mean, ≥100, Episode of 20-25 min
3	0 – 12, SI mean, ≥0.8, Episode of 20-25 min
4	0 – 12, MAP mean, ≤60, Episode of 5-10 min
5	0 – 12, MAP mean, ≤60, Episode of 5 min

[a]Mean = 5-min means of every 6-s data collection; CPP = MAP - ICP; SI = HR/SBP.

Table 6. Five VS Features Most Frequently Selected by All Feature Selection Methods in Predicting Mid-Term (6 Weeks to 3 Months after Discharge) GOSE Using the First 12 Hours of Data

No.	Feature[a]
1	0 – 12, MAP mean, ≤60, Episode of 5-10 min
2	0 – 12, SBP mean, ≤90.00, PTD/D
3	0 – 12, SI mean, ≥0.7, Episode of 30-45 min
4	0 – 12, MAP mean, ≤60, Episode of 5 min
5	0 – 12, CPP mean, ≤50, Episode of 30-45 min

[a]Mean = 5-min means of every 6-s data collection; CPP = MAP - ICP; SI = HR/SBP.

Table 7. Five VS Features Most Frequently Selected by All Feature Selection Methods in Predicting Late (3 to 6 Months after Discharge) GOSE Using the First 12 Hours of Data

No.	Feature[a]
1	0 – 12, CPP mean, ≤50, Episode of 30-45 min
2	0 – 12, BTI mean, ≤3, Episode of 45-60 min
3	0 – 12, SI 5 mean, ≥0.8, Episode of <30 min
4	0 – 12, SI 5 mean, ≥0.7, Episode of 30-45 min
5	0 – 12, ICP mean, ≥20, Episode of 15-20 min

[a]Mean = 5-min means of every 6-s data collection; CPP = MAP - ICP; BTI = CPP/ICP; SI = HR/SBP.

Table 8. Five VS Features Most Frequently Selected by All Feature Selection Methods in Predicting Long-Term (6 Months or More after Discharge) GOSE Using the First 12 Hours of Data

No.	Feature[a]
1	0 – 12, ICP mean, ≥20, Episode of <10 min
2	0 – 12, BTI mean, ≤3, Episode of 45-60 min
3	0 – 12, CPP mean, ≥100, Episode of 30-45 min
4	0 – 12, SI mean, ≥0.8, Episode of 30-45 min
5	0 – 12, SI mean, ≥0.8, Episode of <45 min

[a]Mean = 5-min means of every 6-s data collection; CPP = MAP - ICP; BTI = CPP/ICP; SI = HR/SBP.

6.0 DISCUSSION

The work described here demonstrates the utility of machine learning algorithms for the modeling of long-term patient outcomes after severe TBI and explores the contributions of various methods for feature selection. The univariate methodology provided general support for the notion that features with reasonable clinical relevance can contribute to these models, that the outcomes with which they can be correlated are also clinically relevant, and that data very early in the course of neurocritical care after severe TBI are able to provide information about long-term outcome.

Although logistic regression modeling using these same features probably contributes the least to the long-term goal of optimizing an analytic platform for advanced clinical instrumentation, this modeling step did confirm the utility of the features themselves.

Of the three feature selection methods tested, the various techniques that allow for weighting of selections (Figure 3) provided the most promising results both in terms of predictive power and likelihood of utility as part of an analytic package for translation of this body of work into field-ready clinical tools.

Of interest in reviewing the five features most frequently selected by all of the feature selection methods is that, for the earlier follow-up periods, up to the first 3 months after discharge, vascular (and by inference, vascular volume) features predominate. In contrast, for the later follow-up periods, although vascular features are still "popular," those dependent on ICP move to the fore. Any form of inference based on this work is premature, but those findings are at least clinically plausible and are worth keeping in mind as this work progresses and as techniques to assess and monitor ICP noninvasively, including during extraction and transport, develop, are proven, and become available.

7.0 CONCLUSIONS

Classic machine learning tools have utility for modeling long-term clinical outcomes after severe TBI and have good potential as analytic platforms for field-deployable advanced patient monitoring and decision-assist instrumentation, particularly when coupled with appropriate feature selection software.

We have already integrated the analytic functions developed in this work into the video display system shown in Figure 1. In addition, we are using the findings of this work to focus selection of relevant VS features for four additional projects funded by or in consideration by the U.S. Air Force. "Fit to Fly" (FA8650-12-2-6D09) is examining the correlation between biomarker cytokines and intracranial hypertension and other adverse VS-related events in 6-hour intervals from the time of admission through the first 72 hours. "Noninvasive Intracranial Pressure Monitoring Using Advanced Machine Learning Techniques" (FA8650-11-2-6D06) is a transition step study toward development and testing of a field-ready noninvasive ICP monitor. "Comparison of Automated and Manual Recording of Brief Episodes of Intracranial Hypertension and Cerebral Hypoperfusion and their Association with Outcome after Severe Traumatic Brain Injury" (FA8650-11-2-6142) is a closely related study. It aims to identify any consequences of negative exacerbations of ICP and CPP occurring in the 60-minute "unmeasured" intervals between manual recordings of ICP and CPP in patients with free-draining cerebral catheters. A third closely related study that builds on this work and on the three projects discussed in this section, "A Prospective Study of the Use of First 12-Hour Intracranial

Pressure Data to Provide Long-Term Prognosis after Severe Traumatic Brain Injury," is under review at the 711th Human Performance Wing. This study will test the algorithms developed in the current work against the prospective incoming data from the "Fit to Fly" project described above.

8.0 REFERENCES

1. Dutton RP, Stansbury LG, Leone S, Kramer E, Hess JR, Scalea TM. Trauma mortality in mature trauma systems: are we doing better? An analysis of trauma mortality patterns, 1997-2008. J Trauma 2010; 69(3):620-6.
2. Eastridge BJ, Mabry RL, Seguin P, Cantrell J, Tops T, Uribe P, et al. Death on the battlefield (2001-2011): implications for the future of combat casualty care. J Trauma Acute Care Surg 2012; 73(6 Suppl 5):S431-7.
3. Wright GW, Simon RM. A random variance model for detection of differential gene expression in small microarray experiments. Bioinformatics 2003; 19(18):2448-55.
4. Dudoit S, Fridlyand J, Speed TP. Comparison of discrimination methods for the classification of tumors using gene expression data. J Am Stat Assoc 2002; 97(457):77-87.
5. Simon R, Radmacher MD, Dobbin K, McShane LM. Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification [Commentary]. J Natl Cancer Inst 2003; 95(1):14-8.
6. Kahraman S, Dutton RP, Hu P, Xiao Y, Aarabi B, Stein DM, et al. Automated measurement of "pressure times time dose" of intracranial hypertension best predicts outcome after severe traumatic brain injury. J Trauma 2010; 69(1):110-8.
7. Stein DM, Hu PF, Brenner M, Sheth KN, Liu KH, Xiong W, et al. Brief episodes of intracranial hypertension and cerebral hypoperfusion are associated with poor functional outcome after severe traumatic brain injury. J Trauma 2011; 71(2):364-73.
8. Jennett B, Snoek J, Bond MR, Brooks N. Disability after severe head injury: observations on the use of the Glasgow Outcome Scale. J Neurol Neurosurg Psychiatry 1981; 44(4):285-93.
9. Kahraman S, Dutton RP, Hu P, Stansbury L, Xiao Y, Stein DM, et al. Heart rate and pulse pressure variability are associated with intractable intracranial hypertension after severe traumatic brain injury. J Neurosurg Anesthesiol 2010; 22(4):296-302.
10. Vapnik VN. Statistical learning theory. New York: John Wiley and Sons; 1998:526.
11. Molinaro AM, Simon R, Pfeiffer RM. Prediction error estimation: a comparison of resampling methods. Bioinformatics 2005; 21(15):3301-7.
12. Zou H, Hastie T. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 2005; 67(2):301-20.
13. Simon RM, Korn EL, McShane LM, Radmacher MD, Wright GW, Zhao Y. Design and analysis of DNA microarray investigations. New York: Springer; 2003:169,183.
14. Radmacher MD, McShane LM, Simon R. A paradigm for class prediction using gene expression profiles. J Comput Biol 2002; 9(3):505-11.
15. Ramaswamy S, Tamayo P, Rifkin R, Mukherjee S, Yeang CH, Angelo M, et al. Multiclass cancer diagnosis using tumor gene expression signatures. Proc Natl Acad Sci U.S.A. 2001; 98(26):15149-54.
16. Bø TH, Jonassen I. New feature subset selection procedures for classification of expression profiles. Genome Biol 2002; 3(4):research0017.1.

17. Tibshirani R. The lasso method for variable selection in the Cox model. Stat Med 1997; 16(4):385-95.
18. Duda RO, Hart PE, Stork DG. Pattern classification, 2nd ed. New York: John Wiley & Sons; 2001:95.

LIST OF ABBREVIATIONS AND ACRONYMS

AUC	area under the curve
BTI	brain trauma index
CPP	cerebral perfusion pressure
GCS	Glasgow Coma Scale
GOSE	Glasgow Outcome Scale-Extended
HR	heart rate
ICP	intracranial pressure
ICU	intensive care unit
IQR	interquartile range
LOS	length of stay
MAP	mean arterial pressure
MLA	machine learning algorithm
MLR	multivariate logistic regression
PTD/D	pressure-times-time dose
RFE	recursive feature elimination
ROC	receiver operating characteristic
SpO_2	oxygen saturation
SBP	systolic blood pressure
SD	standard deviation
SI	shock index
STC	Shock Trauma Center
TBI	traumatic brain injury
VS	vital signs
VSS	vital signs signals