### Title

Prediction of Recovery From Severe Hemorrhagic Shock Using Logistic Regression




### Authors

Lucas, Alfredo

Williams, Alexander T

Cabrales, Pedro



### Publication Date

2019

### DOI

10.1109/jtehm.2019.2924011

IEEE Journal of Translational Engineering in Health and Medicine

Received 24 January 2019; revised 13 June 2019; accepted 16 June 2019. Date of publication 1 July 2019; date of current version 23 July 2019.

*Digital Object Identifier 10.1109/JTEHM.2019.2924011*

# Prediction of Recovery From Severe Hemorrhagic Shock Using Logistic Regression

**ALFREDO LUCAS, ALEXANDER T. WILLIAMS, AND PEDRO CABRALES**

Department of Bioengineering, University of California at San Diego, La Jolla, CA 92092, USA

CORRESPONDING AUTHOR: P. CABRALES (pcabrales@ucsd.edu)

This work was supported by NIH grants from the Heart Lung and Blood Institute under Grant R01-HL126945 and Grant R01-HL138116.

**ABSTRACT** This paper implements logistic regression models (LRMs) and feature selection for creating a predictive model for recovery form hemorrhagic shock (HS) with resuscitation using blood in the multiple experimental rat animal protocols. A total of 61 animals were studied across multiple HS experiments, which encompassed two different HS protocols and two resuscitation protocols using blood stored for short periods using five different techniques. Twenty-seven different systemic hemodynamics, cardiac function, and blood gas parameters were measured in each experiment, of which feature selection deemed only 25% of the them as relevant. The reduced feature set was used to train a final logistic regression model. A final test set accuracy is 84% compared to 74% for a baseline classifier using only MAP and HR measurements. Receiver operating characteristics (ROC) curve analysis and Cohens kappa statistics were also used as measures of performance, with the final reduced model outperforming the model, including all parameters. Our results suggest that LRMs trained with a combination of systemic hemodynamics, cardiac function, and blood gas parameters measured at multiple timepoints during HS can successfully classify HS recovery groups. Our results show the predictive ability of traditional and novel hemodynamic and cardiac function features and their combinations, many of which had not previously been taken into consideration, for monitoring HS. Furthermore, we have devised an effective methodology for feature selection and shown ways in which the performance of such predictive models should be assessed in future studies.

**INDEX TERMS** Hemorrhagic shock, logistic regression, critical care, cardiovascular function.

## I. INTRODUCTION

Trauma remains a major source of morbidity and mortality in the United States and world-wide. The World Health Organization estimates that over 5.8 million people die each year because of injuries. This accounts for 10% of the world's deaths, 32% more than the number of fatalities that result from malaria, tuberculosis and HIV/AIDS combined [1]. In the United States, traumatic injury is the fifth most frequent cause of death, with roughly 10% of the population suffering from some type of traumatic injury in any given year [2]. After a major trauma, early and accurate assessment of shock state is necessary to provide appropriate interventions to decrease morbidity and mortality. Currently, specific triage criteria such as blood pressure, respiratory status, shock index, and mechanism of injury are used to categorize trauma patients and prioritize emergency and trauma responders. Early resuscitation includes control of bleeding, restoration of circulating blood volume, blood pressure, and restitution of oxygen carrying capacity. Therapy is generally guided by the rate of bleeding and changes in hemodynamic parameters such as, systolic blood pressure (SBP), hear rate (HR), and blood chemistry and gases. Significant improvements in HR and MAP may occur during resuscitation such that the indices approach "normal" limits, but the organism in shock may have persistent hypoperfusion. Inadequate resuscitation with persistent hypoperfusion can result in higher mortality rates from HR.

The normal ratio of HR to SBP is generally $< 0.7$. This ratio is elevated in the setting of acute hypovolemia and circulatory failure and is referred to as the shock index (SI) [3], [4]. The SI has been demonstrated to be a useful guide for diagnosing early acute hypovolemia in the presence of normal HR and blood pressure [5]. It has also been used as a marker for severity of injury and poor outcome in trauma patients [6], [7], as an indication of ongoing hemorrhage during gastrointestinal bleeding [8], and as an early signal

of rupture and intra-abdominal hemorrhage during ectopic pregnancy [9]. Measurement of the SI may be more useful in predicting early shock than either the HR or the SBP alone and has been shown to correlate with other indices of end-organ perfusion such as central venous oxygen saturation and arterial lactic acid concentration [10].

Predictors for survival after HS have been explored extensively. Significant interest has been placed in parameters such as lactate, which is are commonly associated with organ damage after trauma and has been shown to correlate with mortality during HS [3]–[5], [10]–[12]. Other parameters such as HR and MAP are also commonly used for predicting mortality during HS, often in the form of combined parameters such as the SI [6]–[8]. However, these parameters do not necessarily lead to a holistic description of the physiology during HS, and with the advent of novel therapeutics, blood substitutes and blood storage techniques, a more comprehensive description during the state of shock might be necessary for proper survival estimation. An alternative could involve the incorporation of functional cardiovascular data, as has been shown before [9], however, the problem now becomes how to incorporate all of these parameters into an effective predictive model able to forecast outcome.

Machine learning and logistic regression models (LRMs) can provide a potential avenue for incorporating multiple parameters. The use of machine learning for hypovolemic shock and HS recovery prediction has been explored extensively in the literature [5], [13]–[19]. Nonetheless, many of these models rely on small sample sizes and limited parameters. Furthermore, they often use neural networks or Support Vector Machines (SVMs) with radial basis functions, both of which suffer from interpretability and effective feature relevance estimation, yielding very little information about the usefulness of the features used in the model. Alternatively, LRMs are often much simpler and allow for effective feature importance estimation by using the corresponding feature weights. The use of LRMs with multiple features in HS recovery prediction has been explored previously by only a few studies [16]. However, this study does not incorporate cardiovascular parameters and is based on a single experimental group.

In this study we developed a LRM to assess the ability of different parameters measured during shock in classifying successful HS recovery in rats, as determined by infused volume. Twenty-seven (27) different systemic hemodynamics, cardiac function, and blood gas parameters were measured from 61 experiments of HS resuscitation, with 5 different experimental groups, each consisting of different blood storage techniques. Different groups were used to simulate potential variability in resuscitation fluid. A Monte Carlo-like approach was used for feature selection, and three different classifiers, one with all the features, one with a reduced set of features and one with only MAP and HR, were created. Performance was assessed through accuracy, the Cohen's kappa statistic and receiver operating characteristics (ROC). This study aimed at identifying physiologically relevant parameters that could be predictive of HS survival, primarily in combination with LRM, but with potential to extrapolate to other ML learning and regression models.

## II. METHODS

The data used for this study was acquired over a series of hemorrhagic shock experiments at the University of California, San Diego from 2016 to 2018. In summary, the experiments cover a total of two HS protocols, two resuscitation protocols and 5 different types of stored blood used for recovery. What follows is a brief description of the animal preparation and the experimental protocols used. The retrospective use of data from multiple experiments allows to create a model that adapts to the inherent variability present in clinical cases of HS, where depending on the specifics, different treatment routes will take place. In the proposed methodology, no labeling is made based on the HS shock or resuscitation protocol, or on the infused blood storage type, creating a model that is agnostic to the specifics of the HS.

### A. ANIMAL PREPARATION

Studies were performed in 200-250g male Sprague-Dawley rats (Harlan Laboratories, Indianapolis, IN). Animal handling and care followed NIH Guide for Care and Use of Laboratory Animals and all protocols were approved by the UC San Diego Institutional Animal Care and Use Committee. Animals were initially anesthetized with isoflurane (Draegerwerk AG, Luebeck, Germany) at 5% to induce deep anesthesia. Anesthetic was subsequently reduced to 2.5% for the remainder of the experiment, and animals were placed in the supine position on a heating pad to maintain core body temperature at $37^oC$. Catheters were placed in the left femoral artery and left jugular vein. Animals were then instrumented with a 2F pressure-volume (PV) conductance catheter (SPR858, Millar Instruments, TX) inserted into the left ventricle (LV) via the right common carotid artery. For certain experimental protocols, a tracheostomy was performed, and animals were mechanically ventilated (TOPO ventilator, Kent Scientific, Torrington, CT).

### B. HEMORRHAGIC SHOCK RESUSCITATION PROTOCOL

After cessation of surgery, animals recovered for 30 minutes before baseline measurements were taken. Both HS and resuscitation procedures were performed under 2.5% isoflurane. HS was induced via one of two methods: 1) 50% of blood volume is removed over 30 minutes and hypovolemia is maintained for 30 minutes, 2) 40% of blood volume is removed over 30 minutes and hypovolemia is maintained for 30 minutes, followed by additional 5% hemorrhages every 15 minutes until plasma lactate is within a range of 8-12 mMol. Resuscitation was then implemented via one of two methods: 1) An additional 25% of blood volume is removed (150 uL/min) while 50% of the test blood is simultaneously infused IV (300 uL/min), followed by additional transfusions to maintain the animals MAP at 90% of the pre-hemorrhage MAP, 2). Blood is transfused IV (300 uL/min) to

IEEE Journal of Translational Engineering in Health and Medicine

**TABLE 1. Description of animal placement.**

| Blood Proccessing | Surgical Status | Hemorrhage Protocol | Resuscitation Protocol | N |
|---|---|---|---|---|
| Fresh blood | Not-Ventilated | 2 | 2 | 7 |
| 1-week old conventional | Ventilated | 1 | 1 | 7 |
| 1-week old Anaerobic | Ventilated | 1 | 1 | 14 |
| 3 week old conventional | Ventilated | 1 | 2 | 6 |
| | Not-Ventilated | 2 | 2 | 8 |
| 3 week old anaerobic | Ventilated | 1 | 2 | 13 |
| | Not-Ventilated | 2 | 2 | 6 |

**TABLE 2. Measured parameters.**

| Parameter Name | Parameter Abreviation | Units |
|---|---|---|
| Cardiac Output | CO | $\mu L/min$ |
| Stroke Work | SW | $(mmHg)(\mu L)$ |
| Stroke Volume | SV | $\mu L$ |
| Ejection Fraction | EF | % |
| Arterial Elastance | Ea | $mmHg$ |
| Heart Rate | HR | $BPM$ |
| Mean Arterial Pressure | MAP | $mmHg$ |
| Systemic Vascular Resistance | SVR | $mmHg/\mu L/min$ |
| End Systolic Pressure | Pes | $mmHg$ |
| End Diastolic Pressure | Ped | $mmHg$ |
| Isovolemic Contraction (contractility) | dp/dtmax | $mmHg/s$ |
| Isovolemic Relaxation | dp/dtmin | $mmHg/s$ |
| Isovolemic Contraction End Diastolic Volume Ratio | dp/dt/Ved | $mmHg/s/\mu L$ |
| Hematocrit | hct | % |
| Plasma Lactate | lactate | $mmol/L$ |
| Plasma Glucose | Glucose | $mg/dL$ |
| Hemoglobin | HB | $g/dL$ |
| Arterial Saturation | A | % |
| Venous Saturation | V | % |
| Plasma Hemoglobin | Plasma | $g/dL$ |
| Potassium Concentration | K+ | $mEq/L$ |
| Sodium Ion Concentration | Na+ | $mEq/L$ |
| Calcium Ion Concentration | Ca++ | $mEq/L$ |
| Chloride Ion Concentration | Cl- | $mEq/L$ |
| pH | pH | |
| Arterial Partial Pressure of Oxygen | PO2 | $mmHg$ |
| Arterial Partial Pressure of Carbon Dioxide | PCO2 | $mmHg$ |

bring the animal to and maintain MAP at 90% of the baseline MAP. The 5 different types of blood transfused were: 1) Fresh blood, 2) 1-week old conventionally stored blood, 3) 1-week old anaerobically stored blood, 4) 3-week old conventionally stored blood, 5) 3-week old anaerobically stored blood. All stored blood was processed following AABB guidelines. The use of multiple experimental protocols serves to test the robustness of the proposed approach by attempting to simulate the large variability in HS cases present in clinical settings. Table 1 outlines the surgical group, hemorrhage protocol, resuscitation protocol, blood processing method, and sample size for each combination. All parameters measured are presented in Table 2.

### C. GROUP ASSIGNMENTS

In order to treat recovery prediction as a binary classification problem, two groups were created based on the percent resuscitation at the end of recovery (*%Res 60*) of each animal, regardless of experimental procedure or infused blood type. This quantity corresponds to the cumulative percentage of the animal's original blood volume that was infused at the end of the recovery period to reach the goal blood pressure. The threshold was set at $T = 50\%$ (equivalent to approximately 5 units of blood), where animals with a *%Res 60* of more than 50% were classified as one group (class 0), while those with a *%Res 60* of less than 50% were classified as a second group (class 1). Animals belonging to class 0 are deemed to have had a successful recovery from HS, while those belonging to class 1 are deemed to have had an unsuccessful recovery.

The groups were further divided into training and testing sets, with the training set containing 70% of the animals (n = 42), and the test set 30% (n = 19). The ratio between animals belonging to the two recovery groups was kept consistent between training and test set.

### D. FEATURE PREPARATION

Throughout this study the term measured parameter, or just a parameter, is used to describe a quantity that was measured in an experiment. The term feature is used to describe a parameter measured at a given timepoint, which is used as an input to the classification model. The feature names used in this study are presented in a *parameter_timepoint* format. Furthermore, since this study includes data from experiments performed under different protocols, only parameters measured in all experiments at the same timepoints were used as features. The only exception was the timepoint **ShEnd**, which represents the end of shock (i.e. the timepoint right before infusion), which happened at different times in some protocols.

An overview of the feature selection and training pipeline for the proposed model is shown in Figure 1. With 31 measured parameters across multiple timepoints during both shock and resuscitation, a total of 240 features were available per animal. Since we are interested in predicting recovery using parameters obtained during the shock period, all features measured during resuscitation were discarded, resulting in a feature vector with 69 features per animal.

Prior to training the LRM, each feature vector was normalized to have a unit L2-norm. This prevents features with a larger magnitude from dominating over smaller, but potentially equally, or more informative, features.

### E. LOGISTIC REGRESSION MODEL

The open source Python library *scikit-learn* [20] was used, together with Python 3.6 to apply the LRMs to the data. Logistic regression is a predictive model that uses the standard logistic function where the exponential parameter is determined by a linear combination of features [21]. The weight of each feature in the linear combination is determined by minimizing a cost function. In this study, an L2-norm cost function was chosen, as it performs superior to an L1-norm if the number of irrelevant features is kept low [22], which is in line with the objectives of this study.

### F. FEATURE SELECTION

A Monte Carlo-like approach was used to extract the most informative features to be used in the final model. A total of 400 logistic regression classifiers were trained using different random subsets of data containing 30 instances from the training set. The absolute value of the weights, or coefficients, for each of the features in the classifier were summed across all 400 iterations and ranked according to their cumulative weight.

To select the optimal number of features, each feature was removed iteratively, and the performance of a classifier trained with the reduced number of features was assessed.

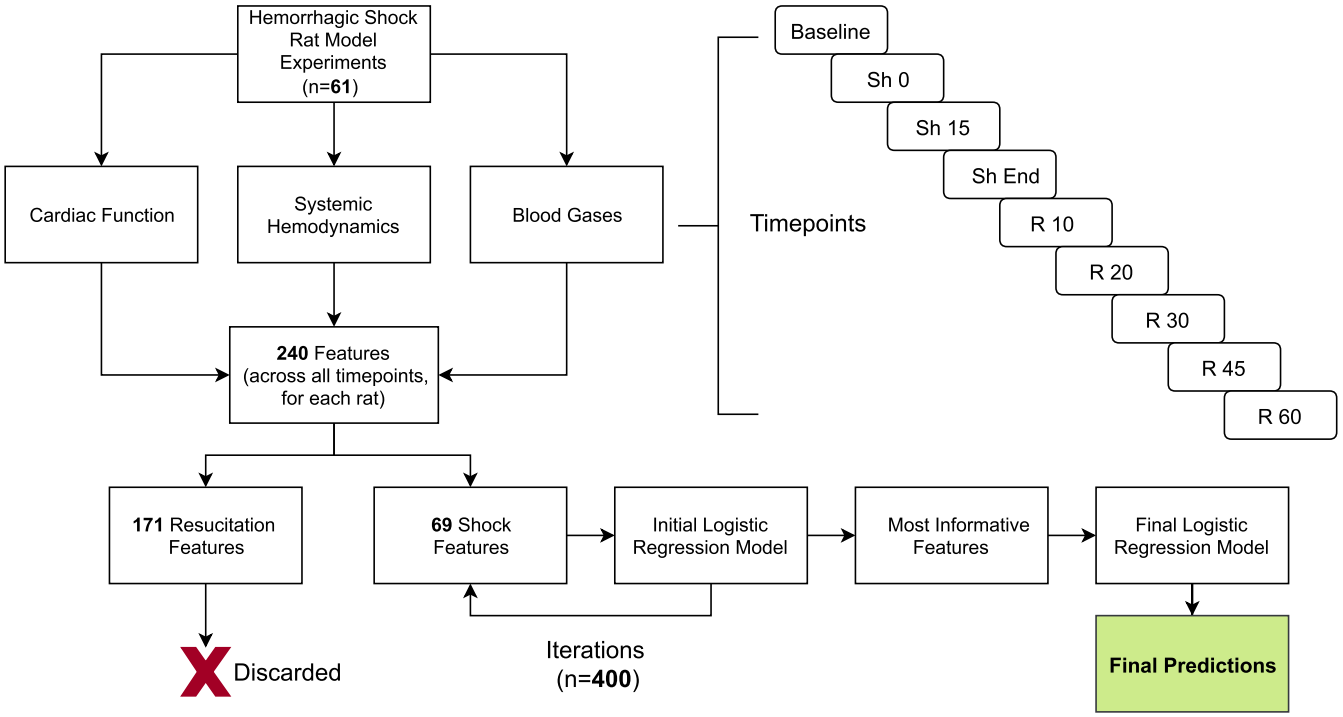


**FIGURE 1.** **Diagramatic representation of the feature extraction and model training pipeline. Timepoints legend: Sh: Shock, R: Resuscitation; numbers represent the time in minutes since the beginning of shock or resuscitation. Sh End corresponds to the end of shock, and its time since the beginning of shock varies with each experiment.**

For this approach, a single random subset of 30 instances from the training set was used for training the classifier in each iteration. 11 instances, also from the training set, were used for testing each classifier. The accuracy of each classifier was assessed by the ratio of correct predictions to total predictions in the reduced testing subset (a more appropriate metric of performance is introduced later for the final model). Furthermore, the same subset of features was used for each new classifier to ensure the accuracy could be compared. Each feature was iteratively removed, in order, from smallest cumulative weight to largest cumulative weight. The reduced feature set that yielded the largest accuracy was then chosen to train a final classifier using the full training set and tested on the full test set.

A baseline classifier was created by using the heart rate (HR) and the mean arterial pressure (MAP) at all available shock timepoints as features (a total of 8 features). These two parameters were chosen because they are often said to be good predictors of HS recovery [6]–[8]. The training and testing instances of this classifier were the same as those for the final classifier.

### G. EVALUATION OF PERFORMANCE

A total of three classifiers were assessed for performance. The first classifier was trained with all available features and will be referred to as the *Initial Classifier*. The second classifier was trained with the reduced feature set, obtained after the feature selection process described above. This classifier will be referred to as the *Final Classifier*. Finally, the performance of the baseline classifier described in the previous section will also be addressed. This classifier will simply be referred to as the *Baseline Classifier*. The training and testing instances in all three classifiers were the same, what changed was the features used by each classifier.

It has been shown that performance assessment in machine learning models can vary significantly depending on the type of classification, the type of data and the way the metrics are computed [23]. To obtain a comprehensive understanding of performance, each classifier was assessed in three ways. First, the accuracy, as measured by the ratio between the number of correct predictions and the total number of predictions was determined. The accuracy in the training set was measured by taking the average of all accuracy scores after a 10-fold cross-validation. The accuracy in the test set was determined through a single accuracy score. Second, the Receiver Operating Characteristics (ROC) of the classifiers in both the training and test data were assessed, and the area under the ROC curve (AUC) was determined [24]. Since the training set is expected to perform well, as it is predicting on the same data it was trained on, the ROC curve is only shown for comparison and as a validation that training took place. The optimal threshold and its performance based on the ROC analysis of the classifiers on the test set was also determined using the Younden's J statistic. This metric calculates the tradeoff between the true positive rate and the false positive rate at a specific threshold of the ROC curve, and is calculated as shown in Equation 1

$$J = TPR - FPR \tag{1}$$

IEEE Journal of Translational Engineering in Health and Medicine

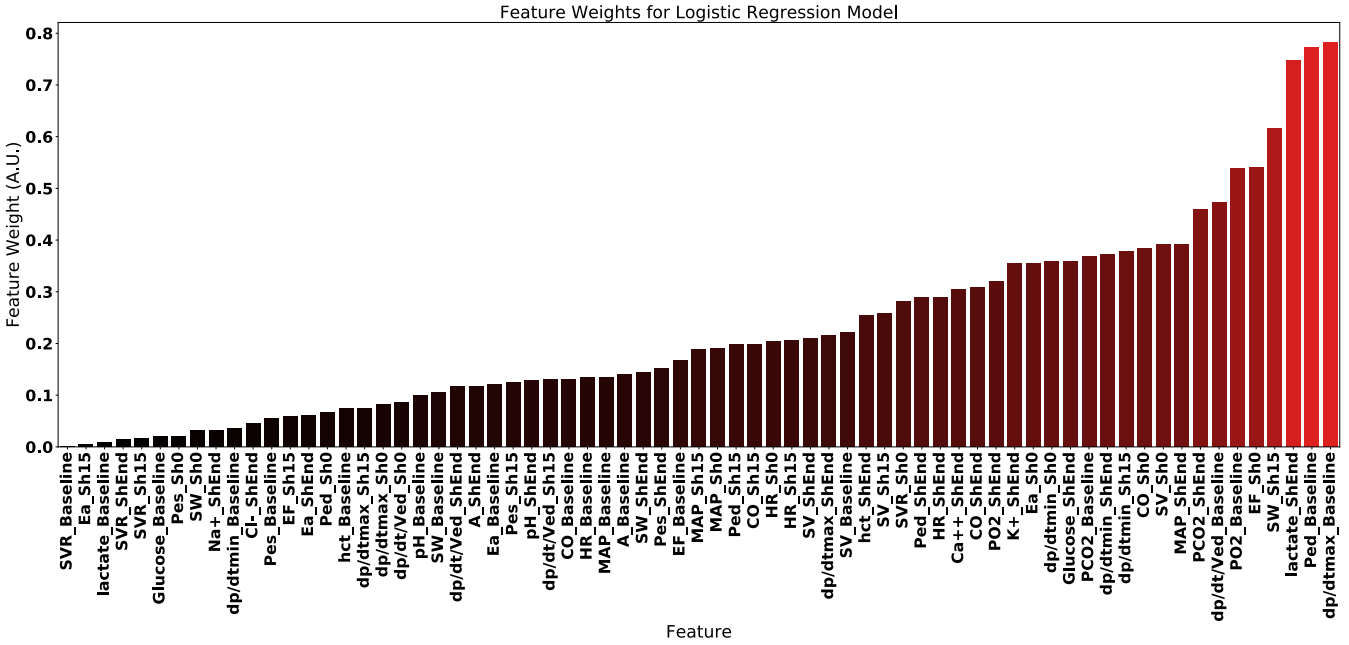


**FIGURE 2.** **Representative distribution of feature weights in a single trained logistic regression model using all original features.**

where TPR corresponds to the true positive rate and FPR to the false positive rate of the predictions [25]. The value of the J statistic ranges between 0 and 1, where a value of 0 is a classifier that is indistinguishable from random chance and a value of 1 is a perfect classifier. Finally, the Cohen's kappa coefficient of each classifier in the test data was assessed. The Cohen's kappa coefficient is a metric of performance similar to accuracy, except it takes into consideration the possibility of agreement occurring by chance [26], [27]. In binary classification tasks this proves particularly useful, since unbalanced classes can results in misleadingly large accuracies [28], [29]. The Cohen's kappa coefficient was computed according to Equation 2

$$\kappa = \frac{p_o - p_e}{1 - p_e} \tag{2}$$

where $p_o$ corresponds to the previously described accuracy, and $p_e$ is the expected accuracy as a result of chance. The expected accuracy is calculated as shown in Equation 3

$$p_e = \frac{\sum \hat{y}_0 \times \sum y_0 + \sum \hat{y}_1 \times \sum y_1}{\left(\sum y_0 + \sum y_1\right)^2} \tag{3}$$

where $\hat{y}_0$ is a binary prediction vector containing ones in the positions where the classifier predicted an instance belonging class 0, and $y_0$ a ground truth binary vector containing ones in the positions of the actual instances of class 0. Similarly, $\hat{y}_1$ and $y_1$, but in this case for class 1. The values of the kappa coefficient range from 0 to 1, with a similar interpretation to that of the J statistic.

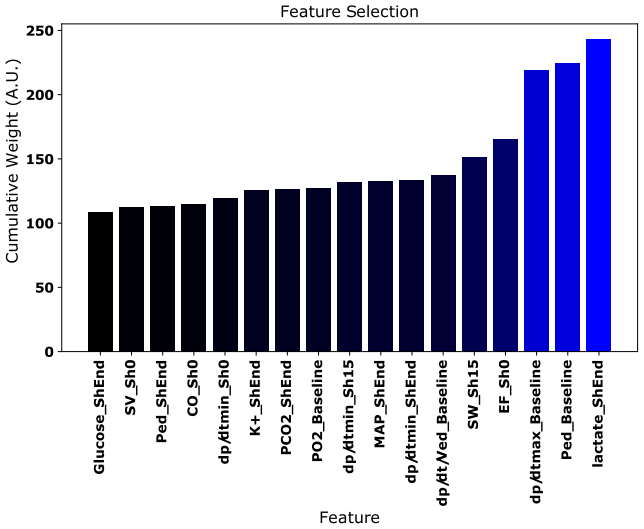


**FIGURE 3.** **Cumulative weights of the chosen reduced feature set after 400 iterations.**

## III. RESULTS

Animals (n = 61) were entered into the HS/resuscitation studies. Within their surgical groups (ventilated and non-ventilated), animals were similar at baseline. Animals were excluded from the LRM if cardiac function data was not collected (n = 1). In approximately 38% of the animals (23 out of 61) the resuscitation required more than 50% of the original blood volume to recover.

### A. FEATURE SELECTION

The resulting weights after training a single classifier with all 69 available shock features are shown in Figure 2.

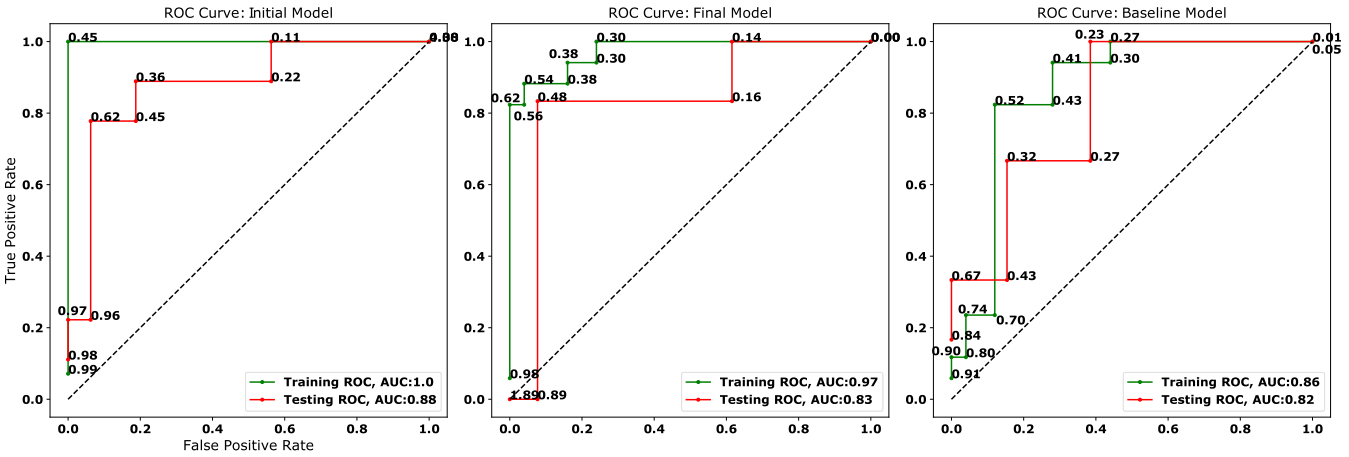


**FIGURE 4.** **ROC curves for the classifiers. Each number on the curve represents the threshold for that specific point. The green line represents the ROC curve for the classifier's performance on the training set. The red line represents the performance of the classifier on the test set. The dashed black line represents random chance performance. The area under the curve (AUC) for each curve is shown in the figure legend.**

**TABLE 3.** **Performance metrics.**

| Performance Metric | Initial Classifier | Final Classifier | Baseline Classifier |
|---|---|---|---|
| Mean Cross-Validation Accuracy | $0.769 \pm 0.088$ | $0.833 \pm 0.114$ | $0.736 \pm 0.056$ |
| Test Set Accuracy | 0.840 | 0.842 | 0.737 |
| Kappa | 0.652 | 0.617 | 0.296 |

The weights appear to be exponentially distributed, supporting the claim that there are weights that contribute significantly more to the model than others.

The feature subset with the best accuracy in the training set had a total of seventeen (17) optimal features, which corresponds to approximately 25% of all available features (Figure 3). The corresponding cumulative weights after 400 iterations are also shown, and an exponential behavior is observed. The distribution of the timepoints of the relevant features was: seven (7) parameters measured at the end of shock (*ShEnd*), followed by four (4) parameters measured at baseline, four (4) parameters measured right after hemorrhage was completed (*Sh0*), and only two (2) parameters measured at 15 minutes into shock (*Sh15*). Since some parameters appeared in more than one timepoint, despite having 17 chosen features, only 14 parameters were used. The parameter with the largest number of useful timepoints was the isovolemic relaxation (*dp/dtmin*), appearing relevant in all 3 shock timepoints.

### *B. CLASSIFIER PERFORMANCE*

A summary of the accuracy and Cohen's kappa values for each of the classifiers is shown in Table 3. The baseline classifier had the lowest performance in all three metrics, while the final classifier had the largest mean cross-validation accuracy and test set accuracy, and the initial classifier had the largest kappa value. The initial classifier had values of cross-validation accuracy, test set accuracy and kappa of $0.769 \pm 0.088$, 0.840 and 0.652 respectively. The final classifier had values of cross-validation accuracy, test set accuracy and kappa of $0.833 \pm 0.114$, 0.842 and 0.617 respectively. Finally, the baseline classifier had values of cross-validation accuracy, test set accuracy and kappa of $0.736 \pm 0.056$, 0.737 and 0.296 respectively.

The ROC curves for each of the classifiers are shown in Figure 4. The initial classifier had the largest values for the AUCs in both training and test set, with values of 1.0 and 0.88 respectively. The final classifier had AUC values for the training and test set of 0.97 and 0.83, slightly below those of the initial classifier. Finally, the baseline classifier had the lowest AUCs, with values of 0.86 and 0.82 for training and test set respectively. The optimal threshold for the initial model was 0.62, with TPR $=$ 0.78, FPR $=$ 0.06 and J $=$ 0.72. The final model had an optimal threshold of 0.54, with TPR $=$ 0.83, FPR $=$ 0.08 and J $=$ 0.76. Finally, the baseline model had an optimal threshold of 0.23 with TPR $=$ 1, FPR $=$ 0.38 and J $=$ 0.61.

## IV. DISCUSSION

The methods presented in this study allowed to successfully determine a subset of useful parameters for classifying rats into two groups based on the infused volume required to achieve successful recovery after HS. Furthermore, the results suggest that the methodology proposed in this study can combine data from multiple experimental protocols and still achieve a classification performance, as measured by multiple metrics, that is above that of the designated baseline classifier. This contrasts with similar studies where a single protocol is maintained for all animals [16], therefore our results serve as a lower limit for the performance new models should be able to achieve. Finally, while many of the chosen features have been independently associated with HS in one way or another, to our knowledge, there has not been a comprehensive study that showed their relative usefulness when they are combined. The discussion that follows presents a physiologically relevant explanation of the validity of these features and why they might have played a role in the performance of the model.

Taking a closer look at the chosen features, physiologically relevant explanations for their usefulness can be formulated.

IEEE Journal of Translational Engineering in Health and Medicine

The choice of lactate concentration at the end of shock as the most relevant parameter is no surprise. Lactate has been shown to be a successful marker for disease mortality in critically ill patients [30]. Furthermore, the use of lactate as a predictor for recovery from HS has been studied extensively [3]–[5], [12]. The decreased oxygen delivery resulting from the loss of blood in HS promotes anaerobic metabolism, resulting in hyperlactataemia as the metabolic flux through glycolysis is increased [30], [31]. The presence of the glucose concentration and the potassium ion concentration as the only other two relevant metabolites is also consistent with other experimental and clinical observations. Glucose has been studied as a marker for monitoring the severity of HS [32], [33]. Glucose metabolism is tightly related to that of lactate, and during conditions of increased anaerobic metabolism, such as that expected during HS, greater glucose utilization and lower energy production is expected [31]. Lactate itself can be converted to glucose in the liver and kidneys through the Cori cycle, exemplifying the direct relationship between the two [34]. Furthermore, the decreased MAP observed during HS, and also selected as an informative feature, triggers catecholamine secretion due to baroreceptor stimulation, promoting glycogenolysis, and subsequent changes in glucose concentration [32], [35]. Finally, glycolysis has been linked to Na+/K+-ATPase activity in smooth muscle and other tissues, and changes in Na+/K+-ATPase activity have been studied in conditions of HS [36], [37], which might explain the presence of K+ concentration as one of the informative features. Furthermore, increased catecholamine production, such as the one expected due to the low MAP, has also been associated with increased Na+/K+-ATPase activity [38], [39]. The tight association of these metabolic parameters with the state of shock is a good validation for their choice. Furthermore, the fact that these metabolic parameters were all chosen at the same time point, end of shock, when the metabolic stress is expected to be at its peak, further validates their choice.

The presence of the two measured blood gases in the selected parameter list is also physiologically consistent. The use of PCO2 for monitoring HS has been explored before [40], [41]. The expected changes in PCO2 during HS can be attributed to respiratory or metabolic imbalances. Changes in the PCO2 are often associated with changes in the metabolic pH balance by action of the bicarbonate-CO2 buffer system [42], [43]. Changes in lactate concentration, as suggested by the previous paragraph, are expected to produce changes in the pH balance, inducing action of the bicarbonate-CO2 buffer system. Furthermore, the fact that the relevance of the PCO2 is at the end of shock, like the lactate and the other metabolites described, might further suggest the association between them. The PO2 value, on the other hand, appears to be relevant at baseline and not at the end of shock. This is a novel finding, and it could be attributed to the fact that a baseline PO2 saturation might be indicative of the inherent oxygen carrying capacity of the animal, which is essential for recovery and survival.

There was also a significant number of cardiac function parameters present in the chosen feature subset, and they range across all available timepoints. The usefulness of these cardiac function parameters for predicting survival from HS has been explored before [9]. However, these results that some of these parameters are relevant not only during and after shock, but also at the baseline of the experiment. The presence of a significant number of baseline features might suggest an inherent predisposition to successful or unsuccessful resuscitation from HS based on the baseline status. Since the animal inclusion criteria is based primarily on weight and MAP, cardiac function parameters are not controlled for at baseline. This allows for variations in cardiac function parameters, which were likely detected by the model and used for recovery prediction.

The state of the animal right after hemorrhage (Sh0) also seems to be relevant. Three of the four parameters selected at this timepoint are directly related to the amount of blood ejected by the animal's heart. In fact, both the ejection fraction and the cardiac output incorporate information from the stroke volume [44], so their joint appearance at the same timepoint might suggest that the model is not arbitrarily choosing features, but instead, it is taking advantage of the pre-existing relationships between the measured parameters to make the predictions. Physiologically, effective blood pumping at the moment right after hemorrhage is critical for maintaining effective oxygen delivery. Since this is the timepoint of the experiment at which the animal's blood volume is at its lowest, the heart's ability to eject blood, as measured by these parameters, will be essential for survival.

The performance results of the different classifiers are also promising. Both initial and final classifiers had superior performance than the baseline classifier in all metrics. The predictions of this baseline classifier were not entirely random, however, since $\kappa > 0$ and the ROC analysis shows indeed some predictive value. Nevertheless, the performance was suboptimal and inferior to that of the other two classifiers. Since the baseline classifier was solely created using MAP and HR parameters, this might imply that using only these two features is insufficient for this type of regression analysis, despite their popular use in shock predictive metrics such as the shock index (SI) [6]–[8]. Furthermore, since the data used in this study involved multiple experiments, it is likely that nuances between the different groups could not be conveyed appropriately by solely using MAP and HR.

Another important observation is the similarity between the performance of the initial classifier and the final classifier suggests that the feature selection process did not cause significant loss of information. In fact, the superior cross-validation accuracy of the final classifier as compared to the initial classifier might suggest an improved generalizability of the reduced final classifier. In terms of test set performance, accuracy values were almost identical between the two classifiers. However, the kappa value of the initial classifier was slightly superior to that of the final classifier. The discrepancy between the test set accuracy and the kappa value

is likely due to the final classifier correctly predicting more instances of the more populous class, making the contribution due to chance more significant.

The ROC analysis of the classifiers also sheds light on their predictive value. Overall, the AUCs of the initial classifier were the greatest, followed by the final classifier and then the baseline classifier.The ideal use of a model as the one proposed, however, would require a minimization of the false positive rate. Since positive predictions indicate an unsuccessful recovery from shock, from a translational perspective, incorrectly determining a positive prediction can be dangerous and impractical. The J statistic attempts to quantify this tradeoff between the TPR and the FPR. A closer look at the ROC values might suggest that the final classifier is actually a better classifier than the initial, even when the latter has a larger AUC. At their respective optimal thresholds, the J statistic of the final classifier is larger than that of both baseline and initial classifier. Furthermore, even with a FPR of 0.08 compared to 0.07 in the initial classifier, the J statistic shows that the TPR of the final classifier compensates for this slightly larger FPR, suggesting an overall better classifier. Finally, the AUC of 0.82 for the baseline classifier can misleadingly suggest similar performance as that of the final classifier. However, with a J statistic of 0.61 at its optimal threshold, the baseline classifier's performance is suboptimal compared to that of the other two.

Features such as lactate, glucose, ionic imbalance and blood gases are readily measured in a clinical setting, often from a single blood sample. The other cardiovascular features, however, involve relatively complex measurements of cardiac function, which either require invasive catheter placement or cardiac echocardiogram (echo) measurements. Recent developments in automated left-ventricular function characterization from 2 and 3-dimensional echocardiogram measurements have been successful [45]–[47]. These automated approaches allow for the measurement, in a matter of seconds, of a wide array of left-ventricular functional parameters including EF, SV, and cardiac strain measurements that can be related to contractility and relaxation. Furthermore, protocols for efficient cardiac echo acquisitions in emergency settings have been proposed previously [48]. Coupling efficient echo acquisition techniques with automated cardiac function characterization algorithms can allow for efficient, and non-invasive, measurement of cardiac parameters that might be predictive of recovery in a HS setting. However, actual coupling of echo-based measurements with the proposed model would require integration of software and hardware, as well as skilled personnel, which can be hard to achieve in low resource settings. As future studies assess the validity of the proposed parameters and model, the practicality of such an integrated approach should also be assessed.

While emphasis should be placed in verifying the feasibility of the proposed parameters acquired in a clinical setting, future studies should also assess the actual accessibility of the proposed parameters. For the easily accessible parameters, such as vital signs and blood-based measurements, future implementations should aim at exploring their HS recovery predictive ability by using more intricate machine learning based models such support vector classifiers (SVCs). These models might be able to further exploit the non-linear relationships between the parameters in order to come up with equivalent predictions, without the need of the harder to obtain measurements. Another alternative might be to modify the current model such that the weighting of the easily accessible parameters is higher under certain conditions. Therefore, in scenarios where the harder to obtain parameters are inaccessible, the model is still capable of carrying out the predictions.

A clear limitation of this study is the sample size. Even though the potential to generalize was demonstrated by the performance of the classifiers in cross-validation and in the test set, increasing the sample size of the training set is expected to produce more generalizable results, and might even lead to some changes in the small weighted chosen features. Furthermore, studies in rats are not directly translatable to clinical scenarios. However, ultimate purpose of the methodology proposed in this study is not to explicitly show which parameters to use, but to both inform about potential parameters that could be useful for monitoring HS in clinical settings, and to show a way in which such parameters could be determined in clinical studies. In addition, we do not yet fully understand how these features would change in light of different interventions, or how the model predictions could dictate intervention. Future studies should address how specific interventions, such as particular drugs, can lead to parameter changes that might modify the prediction outcome of the model. Finally, it is important to remember that experimental data is idealized with respect to real clinical data, and despite attempts to include variability by including multiple experimental protocols, clinical applications of the proposed model will likely require substantially more training data as well as small modifications to the model. The use of retrospectively collected human data could be used to create a more clinically applicable model, however, currently measured parameters do not include many of the parameters described in this study. Therefore, we hope that studies like this encourage the use and study of these parameters in clinically relevant scenarios of HS.

## V. CONCLUSION

The results presented in this study show a successful implementation of logistic regression and feature selection for creating a predictive model for HS recovery classification in multiple experimental rat animal protocols. We have shown the predictive ability of novel features and their combinations, many of which had not previously been taken into consideration for monitoring HS. Furthermore, we have devised an effective methodology for feature selection and shown ways in which the performance of such predictive models should be assessed in future studies. Our results shall serve as a lower limit for the performance newer studies should overcome.

IEEE Journal of Translational Engineering in Health and Medicine

## REFERENCES


[1] *Injuries and Violence: The Facts*. World Health Organization, Geneva, Switzerland, 2010.

[2] D. R. Spahn and R. Rossaint, "Coagulopathy and blood component transfusion in trauma," *BJA, Brit. J. Anaesthesia*, vol. 95, no. 2, pp. 130–139, Aug. 2005.

[3] S. Kushimoto *et al.*, "Lactate, a useful marker for disease mortality and severity but an unreliable marker of tissue hypoxia/hypoperfusion in critically ill patients," *Acute Med. Surgery*, vol. 3, pp. 293–297, Oct. 2016.

[4] U. Krishna, S. P. Joshi, and M. Modh, "An evaluation of serial blood lactate measurement as an early predictor of shock and its outcome in patients of trauma or sepsis," *Indian J. Crit. Care Med.*, vol. 13, no. 2, pp. 66–73, 2009.

[5] J. Y. Choi, W. H. Lee, T. K. Yoo, I. Park, and D. W. Kim, "A new severity predicting index for hemorrhagic shock using lactate concentration and peripheral perfusion in a rat model," *Shock*, vol. 38, pp. 635–641, Dec. 2012.

[6] A. Singh, S. Ali, A. Agarwal, and R. N. Srivastava, "Correlation of shock index and modified shock index with the outcome of adult trauma patients: A prospective study of 9860 patients," *North Amer. J. Med. Sci.*, vol. 6, pp. 450–452, Sep. 2014.

[7] Y.-C. Liu *et al.*, "Modified shock index and mortality rate of emergency patients," *World J. Emergency Med.*, vol. 3, no. 2, pp. 114–117, 2012.

[8] R. Schroll *et al.*, "Accuracy of shock index versus ABC score to predict need for massive transfusion in trauma patients," *Injury*, vol. 49, pp. 15–19, Jan. 2018.

[9] S. Chatpun and P. Cabrales, "Cardiac systolic function recovery after hemorrhage determines survivability during shock," *J. Trauma*, vol. 70, pp. 787–793, Apr. 2011.

[10] P. Manikis, S. Jankowski, H. Zhang, R. J. Kahn, and J.-L. Vincent, "Correlation of serial blood lactate levels to organ failure and mortality after trauma," *Amer. J. Emergency Med.*, vol. 13, pp. 619–622, Nov. 1995.

[11] E. Bolvardi *et al.*, "The role of lactate clearance as a predictor of organ dysfunction and mortality in patients with severe sepsis," *Materia Socio-Medica*, vol. 28, pp. 57–60, Feb. 2016.

[12] B. Soller, F. Zou, M. D. Prince, M. A. Dubick, and J. L. Sondeen, "Comparison of noninvasive pH and blood lactate as predictors of mortality in a swine hemorrhagic shock with restricted volume resuscitation model," *Shock*, vol. 44, pp. 90–95, Aug. 2015.

[13] B. J. P. van der Ster *et al.*, "Support vector machine based monitoring of cardio-cerebrovascular reserve during simulated hemorrhage," *Frontiers Physiol.*, vol. 8, p. 1057, Jan. 2018.

[14] C. A. Rickards *et al.*, "Are you bleeding? Validation of a machine-learning algorithm for determination of blood volume status: Application to remote triage," *J. Appl. Physiol.*, vol. 116, pp. 486–494, Jan. 2014.

[15] V. A. Convertino *et al.*, "Use of advanced machine-learning techniques for noninvasive monitoring of hemorrhage," *J. Trauma*, vol. 71, p. S25–S32, Jul. 2011.

[16] S. B. Choi, J. Y. Choi, J. S. Park, and D. W. Kim, "ATLS hypovolemic shock classification by prediction of blood loss in rats using regression models," *Shock*, vol. 46, no. 1, pp. 92–98, 2016.

[17] S. B. Choi, J. S. Park, J. W. Chung, S. W. Kim, and D. W. Kim, "Prediction of ATLS hypovolemic shock class in rats using the perfusion index and lactate concentration," *Shock*, vol. 43, no. 4, pp. 361–368, Apr. 2015.

[18] K. H. Jang *et al.*, "Comparison of survival predictions for rats with hemorrhagic shocks using an artificial neural network and support vector machine," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. /Sep. 2011, pp. 91–94.

[19] K.-A. Kim, J. Y. Choi, T. K. Yoo, S. K. Kim, K. Chung, and D. W. Kim, "Mortality prediction of rats in acute hemorrhagic shock using machine learning techniques," *Med. Biol. Eng. Comput.*, vol. 51, pp. 1059–1067, Sep. 2013.

[20] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Oct. 2011.

[21] D. W. Hosmer and S. Lemeshow, *Applied Logistic Regression* (Wiley Series in Probability and Statistics), 2nd ed. New York, NY, USA: Wiley, 2000.

[22] A. Y. Ng, "Feature selection, L1 vs. L2 regularization, and rotational invariance," in *Proc. 21st Int. Conf. Mach. Learn. (ICML)*, Banff, AB, Canada, 2004, p. 78.

[23] Y. Jiao and P. Du, "Performance measures in evaluating machine learning based bioinformatics predictors for classifications," *Quant. Biol.*, vol. 4, pp. 320–330, Dec. 2016.

[24] R. Kumar and A. Indrayan, "Receiver operating characteristic (ROC) curve for medical researchers," *Indian Pediatrics*, vol. 48, pp. 277–287, Apr. 2011.

[25] E. F. Schisterman, N. J. Perkins, A. Liu, and H. Bondell, "Optimal cut-point and its corresponding youden index to discriminate individuals using pooled blood samples," *Epidemiology*, vol. 16, pp. 73–81, Jan. 2005.

[26] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," *Biometrics*, vol. 33, pp. 159–174, Mar. 1977.

[27] A. DiCenso, G. Guyatt, and D. Cliska, Eds., *Evidence-Based Nursing: A Guide to Clinical Practice*. St. Louis, MO, USA: Elsevier Mosby, 2005, ch. 30.

[28] M. Hossin and M. N. Sulaiman, "A review on evaluation metrics for data classification evaluations," *Int. J. Data Mining Knowl. Manage. Process*, vol. 5, pp. 1–11, Mar. 2015.

[29] R. Ranawana and V. Palade, "Optimized precision—A new measure for classifier performance evaluation," in *Proc. IEEE Int. Conf. Evol. Comput.*, Vancouver, BC, Canada, Jul. 2006, pp. 2254–2261.

[30] J. Bakker, M. W. N. Nijsten, and T. C. Jansen, "Clinical use of lactate monitoring in critically ill patients," *Ann. Intensive Care*, vol. 3, p. 12, May 2013.

[31] B. Levy, "Lactate and shock state: The metabolic view," *Current Opinion Critical Care*, vol. 12, p. 315, Aug. 2006.

[32] D. Holzrichter, A. Burk, U. Korn, and R. Burk, "The rise of blood sugar as parameter for the degree of severity of hemorrhagic shock in the rabbit," *Arch. Orthopaedic Traumatic Surg.*, vol. 102, pp. 73–77, Nov. 1983.

[33] J. G. Strawitz, H. Hift, A. Ehrhardt, and D. W. Cline, "Irreversible hemorrhagic shock in rats: Changes in blood glucose and liver glycogen," *Amer. J. Physiol.-Legacy Content*, vol. 200, pp. 261–263, Feb. 1961.

[34] X. M. Leverve, "Energy metabolism in critically ill patients: Lactate is a major oxidizable substrate," *Current Opinion Clin. Nutrition Metabolic Care*, vol. 2, pp. 165–169, Mar. 1999.

[35] I. D. A. Johnston, "The Endocrine Response to Trauma," in *Advances in Clinical Chemistry*, vol. 15, O. Bodansky and A. L. Latner, Eds. Amsterdam, The Netherlands: Elsevier, Jan. 1972, pp. 255–285.

[36] F. D. McCarter *et al.*, "Concurrent reduction of glycogenolysis, glycolysis, and NA(+)/K(+) pump activity after hemorrhagic shock," *Current Surg.*, vol. 57, p. 639, Nov. 2000.

[37] F. A. Luchette, L. A. Friend, C. C. Brown, R. K. Upputuri, and J. H. James, "Increased skeletal muscle $Na^+$, $K^{+-ATPase}$ activity as a cause of increased lactate production after hemorrhagic shock," *J. Trauma Acute Care Surg.*, vol. 44, pp. 796–803, May 1998.

[38] J. H. James, C. H. Fang, S. J. Schrantz, P. O. Hasselgren, R. J. Paul, and J. E. Fischer, "Linkage of aerobic glycolysis to sodium-potassium transport in rat skeletal muscle. Implications for increased muscle lactate production in sepsis," *J. Clin. Invest.*, vol. 98, pp. 2388–2397, Nov. 1996.

[39] J. H. James *et al.*, "Stimulation of both aerobic glycolysis and $Na^+$-$K^+$-ATPase activity in skeletal muscle by epinephrine or amylin," *Amer. J. Physiol.-Endocrinol. Metabolism*, vol. 277, pp. E176–E186, Jul. 1999.

[40] Y. Nakagawa *et al.*, "Sublingual capnometry for diagnosis and quantitation of circulatory shock," *Amer. J. Respiratory Crit. Care Med.*, vol. 157, pp. 1838–1843, Jun. 1998.

[41] G. A. A. M. Cammarata *et al.*, "Buccal capnometry for quantitating the severity of hemorrhagic shock," *Shock*, vol. 31, pp. 207–211, Feb. 2009.

[42] P. C. Fernandez, R. M. Cohen, and G. M. Feldman, "The concept of bicarbonate distribution space: The crucial role of body buffers," *Kidney Int.*, vol. 36, pp. 747–752, Nov. 1989.

[43] E. J. M. Campbell, "Blood gas measurements in clinical practice," *Postgraduate Med. J.*, vol. 37, pp. 10–17, Jan. 1961.

[44] E. P. Widmaier, H. Raff, K. T. Strang, and A. J. Vander, *Vander's Human Physiology: The Mechanisms of Body Function*, 13th ed. New York, NY, USA: McGraw-Hill, 2014.

[45] V. H. Rigolin, "Automated Cardiac Volumetric Analysis," *JACC, Cardiovascular Imag.*, vol. 9, pp. 783–784, Jul. 2016.

[46] W. Tsang *et al.*, "Transthoracic 3D echocardiographic left heart chamber quantification using an automated adaptive analytics algorithm," *JACC, Cardiovascular Imag.*, vol. 9, pp. 769–782, Jul. 2016.

[47] J. Zhang *et al.*, "Fully automated echocardiogram interpretation in clinical practice: Feasibility and diagnostic accuracy," *Circulation*, vol. 138, pp. 1623–1635, Oct. 2018.

[48] M. K. Hall *et al.*, "The '5Es' of emergency physician–performed focused cardiac ultrasound: A protocol for rapid identification of effusion, ejection, equality, exit, and entrance," *Acad. Emergency Med.*, vol. 22, no. 5, pp. 583–593, 2015.