

# A Computational Model Based on Human Performance for Fluid Management in Critical Care

Anqi Li<sup>1</sup>, Michael Lewis<sup>2</sup>, Christian Lebiere<sup>1</sup>, Katia Sycara<sup>1</sup>, Shehzaman Salim Khatib<sup>1</sup>,  
Yuqing Tang<sup>1</sup>, Matthew Siedsma<sup>3</sup>, Don Morrison<sup>1</sup>

**Abstract**—Computational simulation is one of the most important ways of reproducing the dynamic responses of a Cyber Physical System using a model of the system. The simulation discovers areas of differential system performance and allows linking such performance back to system characteristics. In the medical domain, patient simulators are used to train physicians in patient management. One critical question is how to verify these systems under realistic human (physician) input. This requires the creation of realistic human models that would be able to capture human cognitive and decision abilities and limitations. Verification of such an overall physician-patient model would result in two advantages: (a) since physicians realistically would not give all possible inputs to the system, verification could be more efficient and (b) the verification may uncover areas of poor human performance. In this paper, we describe our methodology and results in creating a computational model of human fluid management in critical care, based on human experiments.

## I. INTRODUCTION

Computational simulation is one of the most important ways of reproducing the dynamic responses of a Cyber Physical System using a model of the system. The simulation discovers areas of differential system performance and allows linking such performance back to system characteristics. Computational simulations are the most important tools in trying to design, understand and control complex systems. One of the most important complex Cyber Physiochemical Systems is the physiology of the human body. Currently physicians train on Wizard of Oz virtual patients, mannequins, standardized patient or medical serious games for standardized, repeated practice as well as specific feedback [1], [2], [3], [4], [5]. High fidelity, physiology based simulations are currently rare but will become more prevalent since the need for them is being increasingly expressed. One of the most important challenges in the design and deployment of such systems is making them amenable to effective human control. This requirement is complicated by the nonlinear dynamics of the human body and the large degree of individual variations in response to diseases and treatments. Since the attending physicians also have

individual variations in treatments strategies and styles as well as cognitive limitations and capabilities, the problem of creating robust and safe models of the physician-human body simulations becomes extremely challenging. Formal and validated models of human and simulation systems, where the human element is modeled realistically, would be beneficial not only because these models would provide guarantees of performance, but also because they may uncover parts of the control space where human performance can deteriorate to unacceptable levels. Cognitive modeling based on cognitive architectures such as ACT-R [6], [7] has existed for many years. However, the resulting models are not in a mathematical form that is amenable to the techniques of formal verification. One way of meeting this challenge is to create an analytic model of physician control performance based on human performance data. Such an analytic model is cognitively compatible by construction, and because of its mathematical nature, is in the appropriate form for formal verification. In the case of a physician control of a patient simulation, the analytic model can be integrated with the patient model so that the overall system can be verified.

The paper is organized as follows. Section II presents the experimental task and the patient simulator. Section III presents the experimental procedure, Section IV presents the results of the human experiments and Section V presents the analytic model and results. Section VI presents conclusions and future work.

## II. EXPERIMENTAL TASK AND APPARATUS

### A. Background

The experimental task was for the participants to treat a simulated patient that had lost large amount of fluid or blood. Fluid management is done in critical care via the administration of different types of fluids to replenish lost fluids and bring the vital signs of the patient back to normal range. Abnormal conditions can result from loss of blood (hemorrhagic shock) or fluid (hypovolemic shock). Liquid solutions administered intravenously (IV) are then required to reestablish normal cardiovascular operation. Available solutions are classified as either crystalloids or colloids. Crystalloid solutions such as Normal Saline (NS) and colloid solutions such as Albumin can be administered to replace a simple loss of fluids. When serious internal bleeding occurs, blood products such as Packed Red Blood Cells (PRBC) should be given to replace the lost blood.

<sup>1</sup>The authors are with Carnegie Mellon University, Pittsburgh, Pennsylvania, USA. Email: {anqill, skhatib}@andrew.cmu.edu, {katia, yuqing.tang}@cs.cmu.edu, {cl,dfm2}@cmu.edu.

<sup>2</sup>The authors are with the Department of Information Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania, USA. Email: ml@sis.pitt.edu.

<sup>3</sup>The authors are with Department of Critical Care Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania, USA. Email: siedsmamp@upmc.edu.

Corresponding author: Katia Sycara (katia@cs.cmu.edu)

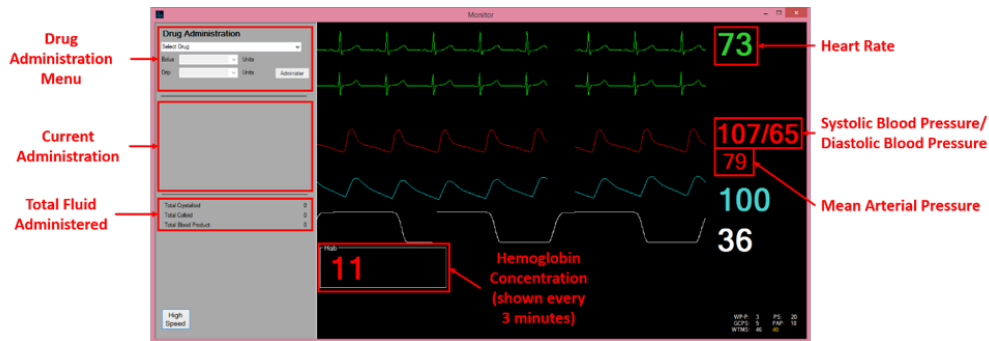


Fig. 1: Patient Monitor - The graphical user interface used by subjects to interact with the simulator. On the left side is the “Drug Administration” menu and on the right side are the vital signs with their corresponding wave forms. PRBC and Albumin are given at a fixed rate (400 ml/min) in a discrete amount called a bolus. Normal Saline (NS) is given continuously at a rate of infusion that needs to be specified by the participant. Multiple fluids can be administered at the same time. Once started, the parameter for a fluid administration cannot be changed but it can be canceled.

### B. The Patient Simulator

The patient computational simulation that was used was created by Dr. Joseph Rinehardt of UC Irvine. The simulator was designed to investigate acute resuscitation in a variety of hemorrhage scenarios. It is capable of accurately modeling heart rate (HR), systolic blood pressure (SBP), diastolic blood pressure (DBP), mean arterial pressure (MAP), cardiac output (CO), and various dynamic parameters of fluid responsiveness, e.g. pulse pressure variation (PPV), systolic pressure variation (SVV), or plethysmograph variability (DeltaPOP) over a wide range of blood volumes. Like most simulation models, the elements are simplifications of the true physiology. Despite these simplifications, however, the simulator aims to faithfully preserve the relationships between the elements themselves and the final simulator outputs. Regarding the evolution of dynamic parameters of fluid responsiveness and their relationship with hemorrhage and fluid responsiveness status, the simulator uses data from published studies [8], [9], [10]. The model begins with a baseline state that includes a height in inches, weight in kilos, HR, SBP, DBP, central venous pressure (CVP), left ventricular enddiastolic volume (LVEDV), and left-ventricular end-systolic volume (LVESV). From these starting values, baseline values for other parameters are calculated. Additionally, two hidden and randomized parameters, Patient Stability and Patient Fluctuations influence how rapidly patient physiology changes in response to a stimulus, e.g. blood loss. The simulator represents a non-linear complex system that the participants must control. More details on the simulator, including the equations and variables used, can be found in the appendix of [11].

The participants get information about the state of the simulated patient and perform control actions, namely administration of different types of fluids via the patient monitor interface. (See Fig. 1). The monitor shows HR, SBP, DBP, MAP, blood oxygenation, and respiration rate of the patient. The “Drug Administration” menu at the left of the monitor can be used to select a fluid and display labels showing current administration and total fluid administered. Please note that the interface does not show either the rate of fluid loss or the fluid volume of the patient.

### III. EXPERIMENTAL PROCEDURE

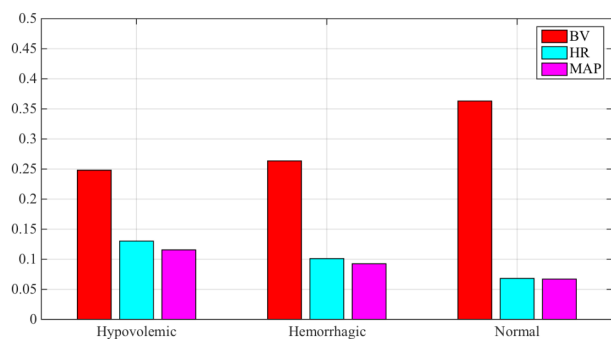
Thirty six participants, all students from the Pittsburgh area were recruited. The experiment consisted of 3 types of episodes, hemorrhagic, hypovolemic and normal. Each subject was given 6 experimental scenarios, namely 2 hemorrhagic, 2 hypovolemic and 2 normal cases. The trials were run in 2 blocks, each containing one scenario from each category. Within each block, the scenarios were permuted. This gave a total of 36 different scenario orderings. We ran 36 subjects, each of whom saw a unique order of the 2-block permuted scenarios. The participants were given background information about the different patient scenarios and trained for about 10 minutes in the use of the simulator. Then they were shown a training scenario, specified to be hemorrhagic, and were allowed to administer fluids and see the effects of their administration on the simulated patient. The subjects could see the variation in patient parameters on the simulation GUI but could not see blood volume (BV), the parameter being controlled.

### IV. HUMAN EXPERIMENT RESULTS

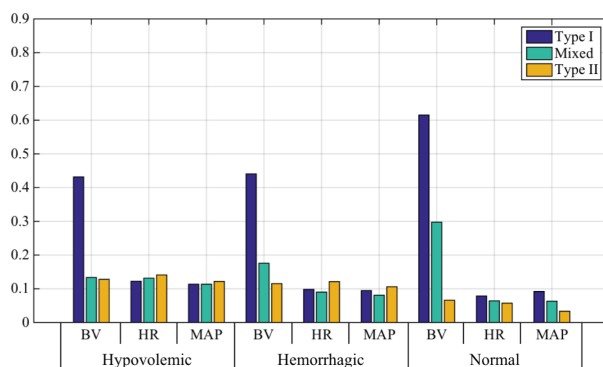
The subject performance metric was normalized root mean square (RMS) error from baseline of the displayed patient parameters, namely HR, MAP, and the hidden variable BV. A main effect of Episode was found for all three dependent measures (HR  $F(2, 32) = 70.66, p < .001$ , MAP  $F(2, 32) = 43.9, p < .001$ , BV  $F((2, 32) = 8.66, p = .001$ ). Order effects were noted for HR,  $F(1, 33) = 17.87, p < .001$  and MAP,  $F(1, 33) = 36.93, p < .001$  but not BV. A significant interaction between Episode $\times$ Strategy was found for all three measures (HR  $F(4, 66) = 6.6, p < .001$ , MAP  $F(4, 66) = 8.73, p < .001$ , BV  $F(4, 66) = 4.69, p = .002$ ) while Order $\times$ Strategy interactions were found for HR,  $F(2, 33) = 3.13, p = .057$  and BV,  $F(2, 33) = 3.13, p = .057$  but not MAP. Only MAP showed an Episode $\times$ Order interaction,  $F(2, 32) = 5.69, p = .008$ .

#### A. Episodes

For HR and MAP there were differences between each of the episode types: HR  $F(2, 32) = 98.7, p < .001$ , with contrasts ( $p < .001$ ) found between each of the scenario

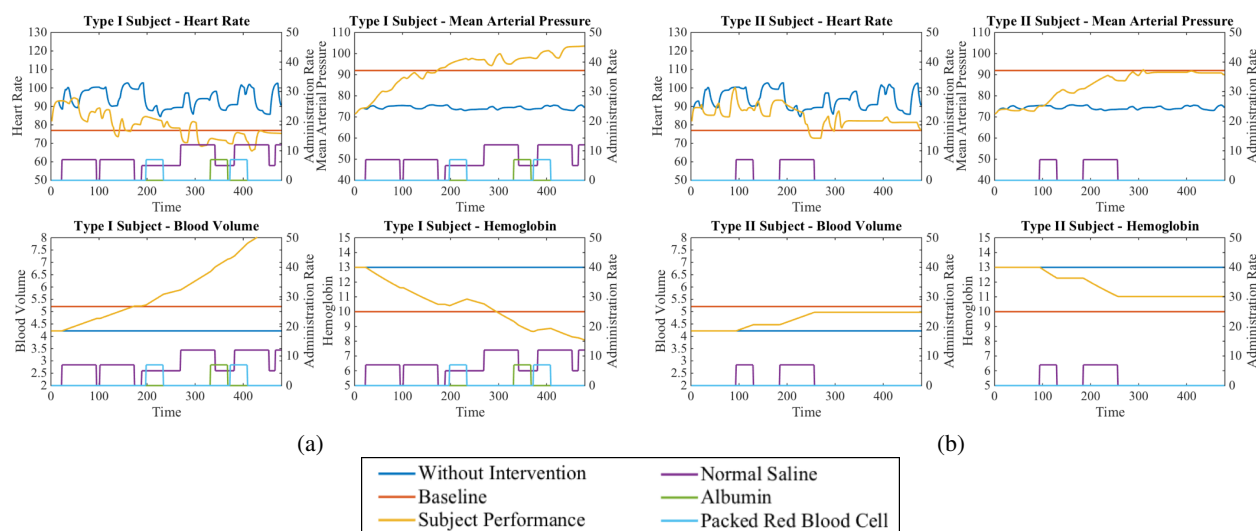


(a)



(b)

Fig. 2: (a) RMS error for blood volume, heart rate, and mean arterial pressure for hypovolemic or hemorrhagic shock and normal episodes. Note that errors on displayed parameters (HR & MAP) decrease between shock and normal conditions while the inferred parameter, BV, increases for normal scenarios. (b) RMS errors by episode, measure, and strategy. Note the poor performance of Type I users on regulating blood volume.



(a)

(b)

Fig. 3: Subjects are classified as Type I and Type II by strategy. (a) One trial for a typical Type I subject. (b) The same trial for a typical Type II subject. Note that Type I subjects are more prone to administration of large amount of fluids at the same time, and apply frequent intervention, while Type II subjects are more conservative. Mixed Type subjects act like Type I in some scenarios (especially normal scenarios) and act like Type II in other scenarios

types. Similar results held for MAP  $F(2, 31) = 57.7$ ,  $p < .001$  with contrasts ( $p < .001$ ) again found between each of the scenario types. For blood volume,  $F(2, 32) = 9.58$ ,  $p = .001$ , however, contrasts showed no difference between conditions with blood loss (hypovolemia and hemorrhagic shock) although both (hypovolemia  $p = .014$ , hemorrhagic  $p < .001$ ) differed from the normal condition.

### B. Order

RMS error was reduced for MAP,  $F(1, 33) = 46.93$ ,  $p < .001$ , between first and second encounters with an episode type. Significant improvements were observed within each episode type. While there was no overall order effect for HR, significant reductions were found within each scenario type although the Episode $\times$ Order interaction did not reach significance. No order effect overall or within scenario type was found for BV.

### C. User Strategy

Since the task is basically a control task, we decided to study the effectiveness of the subjects' control strategies with respect to their effect on visible (heart rate and mean arterial pressure) and non-visible but important (blood volume) parameters. Subjects were categorized by their strategies. Type I subjects are more aggressive in terms of intervention. In contrast, Type II subjects are more conservative, they only do intervention for less a half of time and seldom administer large amount of fluids. We also observe subjects that act like Type I in some cases (especially normal cases), and like Type II in others. These subjects are categorized as Mixed Type. In particular, the classification was based on (a) the total length of gap between administrations, i.e., the total amount of time that the subject is not giving significant amount of fluid ( $\leq 100$  ml/min), and (b) the amount of time that subject is giving large amount of fluid (twice the rate of bolus). If a

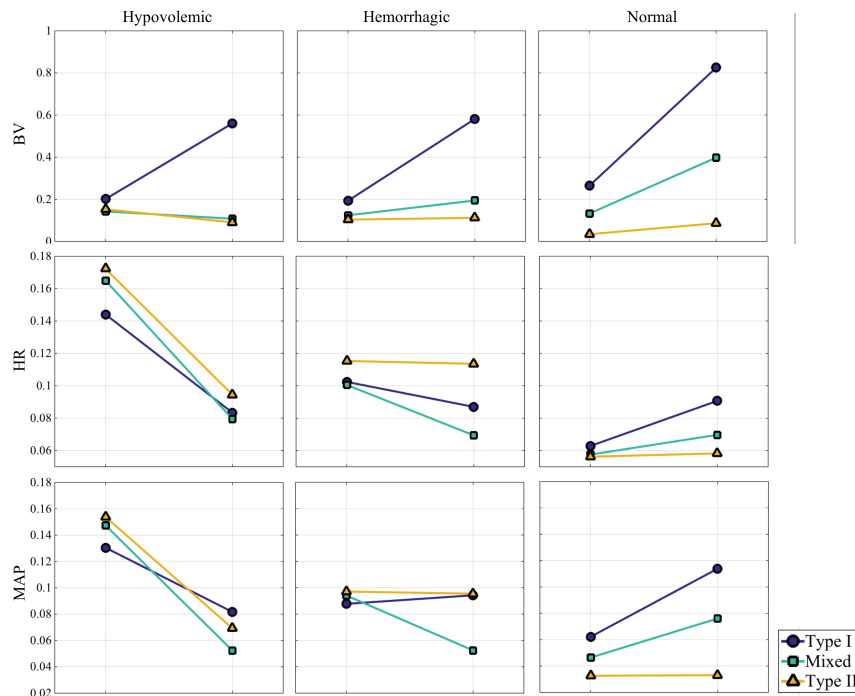


Fig. 4: Subject performance for hypovolemic or hemorrhagic shock and normal episodes where each episode is evaluated based on blood volume (BV), heart rate (HR) and mean arterial pressure (MAP) in two-time split.

subject is giving significant amount of fluids for less than 1/2 of time in the simulation, and giving large amount of fluids for less than 1/4 of time, then the subject is labeled Type II for that scenario. Each subject is classified based on the number of Type I and Type II scenario they score. One trial for a typical Type I subject, and a typical Type II subject is shown in Fig. 3. The Type I subject shown in Fig. 3a gives interventions throughout the trial and administers large amount of fluids for considerable time. The Type II subject, however, only administers relatively small amount of fluids for two times.

A main effect for user strategy was found for all three measures BV  $F(2, 32) = 8.66, p = .001$ , HR  $F(2, 32) = 70.66, p < .001$ , MAP  $F(2, 32) = 43.9, p < .001$ . Examining the effect of user strategy within episode types, however, found no effect for HR or MAP in hypovolemic episodes while pairwise comparisons of error in BV found differences favoring Type II  $p = .012$  and Mixed  $p = .025$  strategies over Type I. A more complex pattern of effects was found for hemorrhagic and normal episodes with Type II users doing less well than others in managing HR (Type II > Mixed  $p = .001$  or Type I  $p = .01$ ) and MAP (Type II error > Mixed  $p = .018$ ) in hemorrhagic episodes but better in normal episodes (HR Type I > Type II  $p = .004$  or Mixed  $p = .024$ , MAP Type I error > Mixed  $p = .018$  > Type II  $p = .016$ ). Type II users did better in managing BV in both these conditions with comparisons showing Type I > Type II  $p = .029$  or Mixed  $p = .04$  for hemorrhagic episodes and Type I > Type II  $p < .001$  or Mixed  $p = .016$  in the normal condition.

For BV measurement, Type I is always worse in the second half than the first. Type II is better in the second half only in the hypovolemic cases, because they overcome the initial deficit and then stop. Mixed Type subjects are usually between Type I and Type II, except in the hypovolemic cases, where they are comparable to Type II. For the other measures (HR and MAP), there is a strong effect of episodes. For normal cases, Type II is still better than Type I (and Mixed Type in the middle) because they start in balance and Type II is better at staying there while Type I over-administers. For hemorrhagic cases, Type II is worse than Type I because Type I is better at aggressively addressing the initial deficit and continuing to administer fluid to combat the bleeding. For hypovolemic cases, all types are much better in the second half because they have addressed the initial deficit. Type I is relatively worse in the second half though because of their over-correction.

The strategies taken by Type I and II subjects reveal differences in their internal model of the simulated patient when they are taking the experiments. Although in terms of visible parameters, i.e. vital sign(s), all type subjects are doing comparably well, for blood volume measurement which is more relevant in fluid management tasks, Type II subjects achieve significantly better performance than Type I subjects. In an experiment with a limited number of medical Fellows, we observed that (a) the Fellows acted like subjects of Type II, (b) they intervened a bit later than the naive subjects, and (c) their intervention resulted in a smoother trajectory of the vital signs being brought to the normal range for the patient. One of the reasons is that the inner model of

practitioners is much more sophisticated and therefore they are more conservative in terms of interventions.

## V. NEURAL NETWORK MODEL

Modeling of the Cyber Physical System in critical care scenario is a challenging task in two respects. First, the human body, by its nature, is an extremely complex system with nonlinear dynamics and numerous parameters that vary among individuals. Second, the decision process for physicians to give treatment is highly complicated with individual variations as well as cognitive limitations.

Recurrent Neural Networks (RNN) is a natural generalization of feedforward neural networks to sequences. In recent research, RNNs have been proven to work in various sequence prediction tasks such as language translation, speech recognition and image captioning [12], [13], [14], [15]. The capacity of recurrent neural networks in sequence prediction for complicated nonlinear dynamic system can be utilized to model our critical care Cyber Physical System, consisting of the patient model and the physician control actions for patient fluid management. Moreover, since recurrent neural networks are built using weights and activations, which is in an explicit mathematical form, it is amenable for further verification of the Cyber Physical System. In this section, we will introduce our preliminary RNN, or more specifically, Long Short Term Memory (LSTM) RNN architecture [16] for modeling the critical care system.

### A. Long Short Term Memory (LSTM) Model

We define states  $x$  of our model at time step  $t$  as  $s_t$ , which consists of the Heart Rate (HR), Mean Arterial Pressure (MAP), Hemoglobin (Hgb), and Blood Volume (BV) at that time. We also define actions  $a$  at time step  $t$  as  $a_t$ , consisting of amount of fluids and Packed Red Blood Cell (PRBC) given per minutes at time step  $t$ . Note that we merge the administration of Normal Saline and Albumin as fluids, since we do not clearly differentiate these two fluids both in experiments and in practice in our scenarios.

Given a sequence  $\mathbf{x} = \{x_1, \dots, x_T\}$  of states, a standard RNN computes the hidden vector sequence  $\mathbf{h} = \{h_1, \dots, h_T\}$  and output sequence  $\mathbf{y} = \{y_1, \dots, y_T\}$  by iterating the following equations from  $t = 1$  to  $T$ ,

$$\begin{aligned} h_t &= \mathcal{H}(W_{xh}x_t + W_{hh}h_{t-1} + b_h) \\ y_t &= W_{hy}h_t + b_y \end{aligned} \quad (1)$$

where the  $W$  terms denote weight matrices, the  $b$  term denotes bias vectors and  $\mathcal{H}$  is the hidden layer function.

However, instead of a single hidden unit for a standard RNN model, a LSTM unit [16] consists of four parts, *input gate*, *forget gate*, *output gate* and *cell*. For a LSTM model, the hidden layer function  $\mathcal{H}$  is the following composite function,

$$\begin{aligned} i_t &= \text{sigm}(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \\ f_t &= \text{sigm}(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \\ c_t &= f_t c_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \\ o_t &= \text{sigm}(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_{t-1} + b_o) \\ h_t &= o_t \tanh(c_t) \end{aligned} \quad (2)$$

where  $i, f, o, c$  represents input gate, forget gate, output gate and cell respectively. This structure helps LSTM models to avoid the long-term dependency problem and increase their capability to predict sequences.

The Cyber Physical System consists of two separate models, (a) the patient model that describes how the human body reacts to blood loss and fluid administration, and (b) the subject model that predicts subject decision of fluid management. For the patient model, the goal is to predict states at current time step  $s_t$ , based on previous states and actions. Similarly, for the subject model, we want to predict the action at current time step  $a_t$ , based on current and previous states, and previous actions. The LSTM model is trained on a sequence of states  $x_t = (s_t, a_{t-1})$  and outputs  $y_t = a_t$ . These LSTM models can be used for analysis on the viability and reachable set in the Cyber Physical System, and therefore determining safe actions and safe patient status in verification of the system.

### B. Preliminary Results

Our preliminary model consists of one single-layer LSTM model for patient and one for subject. We split subject data into two halves. The model is trained on data of all subjects for three scenarios, and is tested on subject data for the other three scenarios. Since the scenarios need time to set up, data for the first 10 seconds of each scenario (when simulation is initializing) are removed from training and testing sequences. The model does one-step prediction based on states (HR, MAP, BV, Hgb, etc) and actions that were performed in the last 20 steps (40 seconds) of subject data.

The average prediction errors over all subjects for HR, MAP, Hgb and BV are 5.9 bps, 3.9 mmHg, 0.4 g/dL, and 0.1 L respectively, all of which are less than 10% of subjects baselines. The average prediction errors for fluids and PRBC administration are 35.5 ml/min and 16.4 ml/min, respectively, which are significantly less than the slowest drip of 100 ml/min. Model prediction and ground truth for the subject and scenario depicted in Fig. 3b are shown in Fig. 5. The single layer LSTM model achieves decent performance in one-step prediction, despite some offset on vital signs. Since the training set is relatively small, the LSTM model might easily overfit to the baselines of the first three scenarios, and therefore cause the offset on vital signs.

### C. Discussion

Although neural network models do not explicitly encode probability distribution over output, there has recently been research about representing uncertainty for neural network models. For example, Gal and Ghahramani [17] create a model of dropout<sup>1</sup> Such uncertainty measurement can help the Cyber Physical System model to reflect possible circumstances during critical care, and increase the performance and safety of the overall system.

Besides introducing uncertainty measurement into the current model, there are a few directions for future work. One of

<sup>1</sup>A technique that temporarily removes randomly selected units from a neural network, along with all its incoming and outgoing connections.

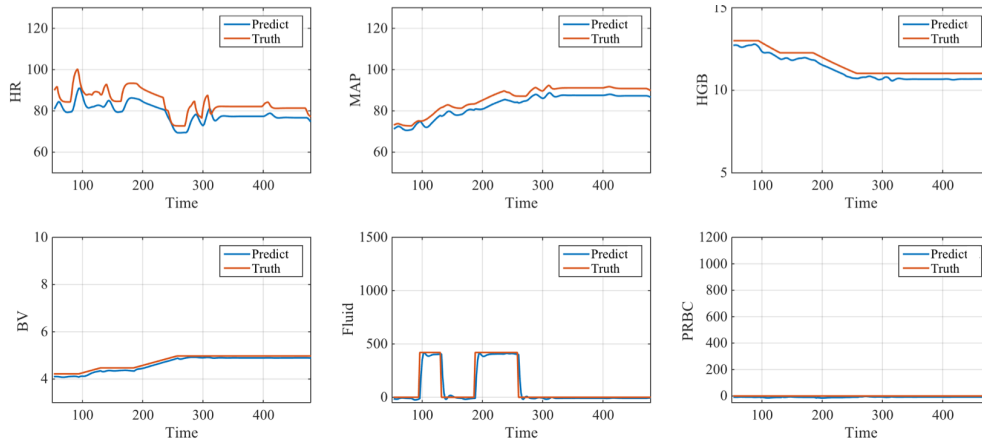


Fig. 5: LSTM model prediction for vital signs and actions for the experiment trial in Fig. 3b. The single layer LSTM model achieves decent performance in one-step prediction, despite some offset on vital signs. The average prediction errors over all subjects for HR, MAP, HgB and BV are 5.9 bps, 3.9 mmHg, 0.4 g/dL, and 0.1 L respectively, and for fluids and PRBC administration 35.5 ml/min and 16.4 ml/min, respectively.

the directions is to increase the number of layers in current LSTM model, since the literature shows that deep LSTM models can out-perform shallow LSTM models [13], [15]. A deeper LSTM model may improve performance for one-step predictions and other more complicated tasks such as multi-step predictions.

## VI. CONCLUSION

In this paper we presented a methodology and preliminary results for creating an analytic model of fluid management for various types of shock (hemorrhagic and hypovolemic) based on results from human performance on the task. Creating an analytic model that captures realistic human behavior is the first step to the development of a model of fluid management in critical care consisting of a simulated patient and a physician that can be formally verified. In future work we plan to run a large number of Medical Fellows as experimental participants, further refine the neural network model and also proceed with the formal verification of the overall model.

## VII. ACKNOWLEDGEMENTS

The authors would like to thank Dr. Joseph Rinehart for developing the simulator, and Yulong Pei for modifying and adapting the simulator for our human experiment. This work has been funded by NSF awards CNS1329986 and CNS1329762 to Carnegie Mellon University and University of Pittsburgh.

## REFERENCES

- [1] J. H. Barsuk, E. R. Cohen, D. B. Wayne, V. J. Siddall, and W. C. McGaghie, "Developing a simulation-based mastery learning curriculum: Lessons from 11 years of advanced cardiac life support," *Simulation in Healthcare*, vol. 11, no. 1, pp. 52–59, 2016.
- [2] W. C. McGaghie, S. B. Issenberg, E. R. Cohen, J. H. Barsuk, and D. B. Wayne, "Translational educational research: a necessity for effective health-care improvement," *CHEST Journal*, vol. 142, no. 5, pp. 1097–1103, 2012.
- [3] R. H. Blum, J. R. Boulet, J. B. Cooper, and S. L. Muret-Wagstaff, "Simulation-based assessment to identify critical gaps in safe anesthesia resident performance," *The Journal of the American Society of Anesthesiologists*, vol. 120, no. 1, pp. 129–141, 2014.
- [4] R. Wang, S. DeMaria Jr, A. Goldberg, and D. Katz, "A systematic review of serious games in training health care professionals," *Simulation in Healthcare*, vol. 11, no. 1, pp. 41–51, 2016.
- [5] R. P. Scott, L. R. V. Diepen, and R. J. Markert. (2010, June) Simulation update: a review of simulation-based strategies for healthcare, education and training. [Online]. Available: <https://www.simulationinformation.com/>
- [6] J. R. Anderson, *How can the human mind occur in the physical universe?* Oxford University Press, USA, 2007.
- [7] J. R. Anderson and C. J. Lebiere, *The atomic components of thought*. Mahwah, N.J.: Erlbaum, 1998.
- [8] A. Perel, R. Pizov, and S. Cotev, "Systolic blood pressure variation is a sensitive indicator of hypovolemia in ventilated dogs subjected to graded hemorrhage," *Anesthesiology*, vol. 67, no. 4, pp. 498–502, 1987.
- [9] R. Pizov, A. Eden, D. Bystritski, E. Kalina, A. Tamir, and S. Gelman, "Arterial and plethysmographic waveform analysis in anesthetized patients with hypovolemia," *The Journal of the American Society of Anesthesiologists*, vol. 113, no. 1, pp. 83–91, 2010.
- [10] M. Cannesson, Y. Le Manach, C. K. Hofer, J. P. Goarin, J.-J. Lehot, B. Vallet, and B. Tavernier, "Assessing the diagnostic accuracy of pulse pressure variations for the prediction of fluid responsiveness: a gray zone approach," *The Journal of the American Society of Anesthesiologists*, vol. 115, no. 2, pp. 231–241, 2011.
- [11] J. Rinehart, B. Alexander, Y. Le Manach, C. K. Hofer, B. Tavernier, Z. N. Kain, and M. Cannesson, "Evaluation of a novel closed-loop fluid-administration system based on dynamic predictors of fluid responsiveness: an in silico simulation study," *Critical Care*, vol. 15, no. 6, p. 1, 2011.
- [12] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with lstm," *Neural computation*, vol. 12, no. 10, pp. 2451–2471, 2000.
- [13] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Advances in neural information processing systems*, 2014, pp. 3104–3112.
- [14] A. Graves, A.-r. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *2013 IEEE international conference on acoustics, speech and signal processing*. IEEE, 2013, pp. 6645–6649.
- [15] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85–117, 2015.
- [16] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [17] Y. Gal and Z. Ghahramani, "Dropout as a bayesian approximation: Representing model uncertainty in deep learning," *arXiv preprint arXiv:1506.02142*, 2015.