Accepted for publication at the IEEE EMBC 2024, Orlando, Florida, July 15-19, 2024.

# Bayesian Variational Autoencoders for Out-of-Distribution Detection in Physiological Modeling: A Case Study in Fluid Therapy

Elham Estiri
*College of Aeronautics and Engineering*
*Kent State University*
Kent, OH, USA
eestiri@kent.edu

Hossein Mirinejad*
*College of Aeronautics and Engineering*
*Kent State University*
Kent, OH, USA
hmiri@kent.edu

***Abstract*— Uncertainty quantification is crucial in modeling critical care systems, where external factors such as clinical disturbances significantly impact decision-making. This study employs Bayesian variational autoencoders (BVAEs) to quantify inherent randomness in clinical data (aleatoric uncertainty) and detect uncertainty in the biases and weights of the neural network model (epistemic uncertainty). Focusing on fluid therapy, the proposed BVAE models aim to detect hemorrhage incidents through out-of-distribution (OoD) data detection. The models' ability to self-identify OoD scenarios not only provides a measure of confidence in their predictions but also highlights areas where additional data collection could enhance performance. Simulation results show promising outcomes, particularly in identifying hemorrhage through increased model uncertainty in OoD scenarios.**

**Keywords— Bayesian variational autoencoder, uncertainty modeling, out-of-distribution (OoD) detection, active hemorrhage, fluid therapy.**

## I. Introduction

Fluid therapy is a medical treatment to restore intravascular volume in critical care scenarios, such as hemorrhagic shock and severe dehydration [1-2]. Optimal fluid management is a complex task due to intricate interactions between patient hemodynamic variables and drug infusion. Over-aggressive fluid infusion may lead to severe complications such as sepsis, underscoring the need for timely and prompt fluid management [3].

In recent years, there has been a growing interest in utilizing computational methods and novel modeling algorithms to develop automated medication dosing tools in critical care scenarios [3-10]. In [5], a novel model-free reinforcement learning approach was proposed to control the mean arterial pressure (MAP) in response to fluid infusion. In [6], a radial basis function-Galerkin (RBF-Galerkin) [7] optimal control method was used to automatically adjust fluid infusion dosages in hypovolemic scenarios. In [11], a computer-assisted mechanism was suggested for the continuous adjustment of fluid dosages.

Conventional physiological models often struggle with overconfidence in predictions, lacking the ability to handle medical decision-making's complexity and uncertainty, which compromises their generalizability and predictability. In critical care scenarios, these uncertainties may arise from several factors, such as the interaction of multiple medications in comorbidities or the occurrence of clinical disturbances during the treatment [12]. These factors represent out-of-distribution (OoD) scenarios, situations unseen by the model during training but encountered in testing. Without specific training to recognize these scenarios, the model may make decisions for which it has not been adequately prepared [13], leading to improper hemodynamic predictions and dosing recommendations.

Uncertainty quantification can detect OoD scenarios and provide an estimate of variability and unpredictability in physiological systems. In the context of fluid therapy, external factors such as active hemorrhage may be seen as a clinical disturbance influencing treatment strategies. The hemorrhage represents an OoD scenario, and its proper identification proves highly beneficial in designing effective resuscitation regimens, and this work aims to address it.

This work presents a novel uncertainty quantification method in physiological modeling using Bayesian variational autoencoders (BVAEs). The BVAE unifies autoencoder learning, variational inference, and Bayesian neural network (BNN) concepts in a single framework [14]. It offers several advantages, including generative modeling, robustness to overfitting, precise control over latent space, and the ability to quantify uncertainties. Unlike conventional neural networks (NNs) that estimate fixed values for weights, BNNs model weights as a distribution. The BVAE not only has the power of a variational autoencoders (VAE) to quantify inherent randomness in data (aleatoric uncertainty) but can also detect uncertainty in the biases and weight of the NN model (epistemic uncertainty) [15]. Focusing on fluid therapy in hemorrhagic scenarios, we develop BVAE models for robust uncertainty quantification (OoD detection). The ability of our models to quantify uncertainties not only provides a measure of the model's confidence in its predictions but also highlights

This material is based upon work supported by the National Science Foundation CAREER Award under Grant No. 2340139 and by the National Science Foundation Engineering Research Initiation (ERI) Award under Grant No. 2138929.

areas where additional data collection could enhance the model's performance. This feature bolsters the model's utility and reliability in real-world critical care applications. To the best of the authors' knowledge, this work is pioneering in considering both epistemic and aleatoric uncertainties in critical care scenarios.

## II. Methodology

In this section, the structure and training methodology of the BVAE is explained. BVAE is a powerful tool for modeling probability distribution of dataset and estimating uncertainty [14]. The BVAE used in this work has a single encoder and multiple decoders, as shown in Fig. 1. However, it could be extended to include multiple encoders, if needed. In fact, the BVAE is a VAE with a network constructed by BNN serving as the decoder. Therefore, weights of the decoder are treated as a distribution rather than fixed values. It can be viewed as a variational encoder and an ensemble of decoders, with the output being an average over different decoders.

In the BVAE framework, the inputs maps into a probability distribution over latent space. After sampling from the probability distribution, the data passes through a Bayesian network to model uncertainty by proposing a probability distribution for each weight. Within this framework, the decoder parameters $\beta$ and the latent state $x$ are considered as random variables with corresponding posteriors $p(\beta \mid D)$ and $p(y \mid x, \alpha)$, respectively, where D is the dataset, y represents the output reconstructed from the decoder, and $\alpha$ denotes the fixed parameters associated with the encoder. The primary objective encompasses optimizing $\alpha$ and simultaneously estimating two posteriors.

The estimation of posteriors $p(\beta \mid D)$ and $p(y \mid x, \alpha)$ involves optimizing a reconstruction loss and a negative log likelihood over posteriors, thereby integrating both latent space $x$ and model parameters $\beta$. Both posteriors are intractable, necessitating the imposition of priors over $x$ and $\beta$ to estimate them by constraining them to their prior. To quantify the integration over latent variables $x$ and decoder parameters $\beta$, the following marginal likelihood is introduced to emphasize the comprehensive nature of the model's assessment of both latent variables and model parameters:

$$p(y \mid D) = \int\int p(y \mid x, \beta) p(x) dx \, p(\beta \mid D) d\beta \tag{1}$$

where $\beta$ is generated by sampling from $p(\beta \mid D)$, $x$ is generated by sampling form $p(x)$, and the system outputs are produced by sampling from $p(y \mid x, \beta)$.

The loss function for the BVAE model comprises three components:

1. *Kullback-Leibler (KL) Divergence 1 (KL1):* Representing the KL Divergence between the approximate posterior of the latent variables and the prior, this term is formulated as:

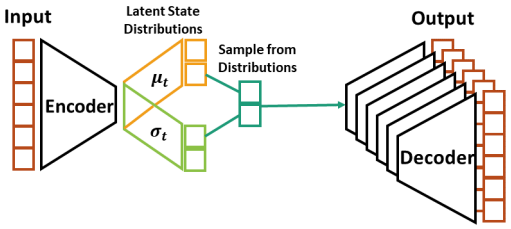


Fig. 1. Bayesian Variational Autoencoder (BVAE) Framework

$$KL1 = D_{KL}(q(x \mid y) \,||\, p(x)) \tag{2}$$

where $q(x \mid y)$ is the approximate posterior of the latent variables given input data, and $p(x)$ is the prior. This term encourages the distribution of the latent variables to be close to the prior (usually a standard Gaussian).

2. *KL Divergence 2 (KL2):* This term is the KL divergence between the approximate posterior and the prior of the weights in the BNN and is formulated as:

$$KL2 = D_{KL}(q(\beta \mid D) \,||\, p(\beta)) \tag{3}$$

where $q(\beta \mid D)$ is the approximate posterior of the weights and $p(\beta)$ is the prior. This term encourages the distribution of weights to be close to the prior.

3. *Negative Log-Likelihood (NLL):* This term measures how well the BVAE is able to reconstruct the input data. It is often the mean squared error (MSE) or binary cross-entropy (BCE) between the original input and the reconstructed output. It is formulated as:

$$NLL = -\log p(y \mid x, \beta) \tag{4}$$

where $p(y \mid x, \beta)$ is the likelihood of the data, given the latent variables and the weights.

The total loss function is constructed as:

$$L_{BVAE} = KL1 + KL2 + NLL \tag{5}$$

to ensure that the BVAE learns a useful latent space (through KL1), the BNN learns a distribution over weights that explain the data (through KL2), and the reconstructed output is as close as possible to the original input (through NLL).

## III. Results and Discussion

In this study, we utilized a BVAE approach to detect the incidence of hemorrhage as an unexpected external factor (OoD detection) during fluid therapy. By integrating the power of VAE and BNN in a unified framework, the proposed BVAE provides a robust and effective method for capturing uncertainty in physiological modeling. We employed two separate datasets: one for training and the other for testing the BVAE model. For training, we used a clinical dataset from a study conducted at the Department of Anesthesia, South Hospital, Stockholm, Sweden [16], which did not include active hemorrhage cases. For testing the BVAE, we used an

animal dataset with active hemorrhage from a study at the Resuscitation Research Laboratory, University of Texas Medical Branch [17].

In the clinical study [16], 10 individual human subjects were infused with the Ringer's Acetate, a crystalloid fluid, at a volume rate of 25 ml/kg over a period of 30 minutes. Following the infusion, an observation period of 150 minutes was maintained. Notably, there was no pre-infusion hemorrhage volume during the study. Blood pressure was measured every 5 minutes during the study. We computed the MAP from the systolic and diastolic blood pressure measurements using the following formula [18]:

$$MAP = \frac{Systollic\ pressure + 2 \times Diastolic\ pressure}{3} \tag{6}$$

In the animal study [17], 10 sheep were subjected to a hemorrhage rate of 25 ml/kg within the initial 15 minutes, after which the hemorrhage was stopped, except at times t = 52 and t = 72 minutes where additional hemorrhages of 5 ml/kg were introduced for two minutes for each subject. These instances represented unexpected bleeding during transition to the hospital. Fluid resuscitation with Lactated Ringer's solution was initiated 30 minutes after the commencement of the study. Throughout the 180-minute period, MAP measurements were recorded every 5 minutes. The model input data included fluid infusion and hemorrhage rates, while the output data consisted of the corresponding MAP values.

The training dataset consisted of 1800 MAP samples from human subjects, all of whom experienced no hemorrhage. The diversity in human subjects aimed to capture a wide range of physiological variations. To assess the model's robustness in the presence of active hemorrhage, we conducted tests using 180 samples from the animal data, specifically chosen to represent active hemorrhage as an OoD scenario during treatment.

The BVAE architecture employed in this work consisted of two layers of conventional NN for the encoder, comprising 10 and 5 nodes, and two layers of BNN for decoder, with 5 and 10 nodes, respectively. The training and optimization process involved minimizing loss function (2) using Adam optimizer with a learning rate of 0.001 over 100 epochs of training. These configurations were empirically determined through an iterative search process, balancing computational feasibility and model performance. To validate the model's performance, we employed the Early Stopping methodology [19]. The dataset was split into training and validation sets, with the validation set consisting of 30% of the training data. At the end of the training process, the model achieved a MSE of 0.086 on the validation set.

Fig. 2 shows the results from testing the model on an individual animal subject. Fig. 2a shows the treatment scenarios of infusion and hemorrhage, while Fig. 2b compares predicted and real MAP values, highlighting the areas of uncertainty. Notably, Fig. 2b demonstrates increased uncertainty in the model's predictions during hemorrhage events, as indicated by a wider uncertainty boundary affecting MAP values. This heightened uncertainty indicates the model's ability to detect OoD scenario, i.e., hemorrhage incidences, not encountered during training. To analyze the model's performance and accuracy, three performance metrics were employed: root mean square error (RMSE), mean absolute error (MAE), and median absolute percentage error (MDAPE). The results shown in Table I suggest the model's effectiveness in capturing the MAP dynamics. Additionally, we performed a linear regression test [20] on the dataset to highlight the significant linear correlation between MAP and the hemorrhage profile. This analysis implies that the observed higher uncertainty during sudden drops in MAP values can be attributed to the impact of hemorrhage events on MAP.

The model's ability to detect OoD scenarios provides a measure of the model's confidence in its predictions. In addition, it highlights areas where additional data collection could improve the model's performance. This feature, therefore, enhances the model's utility and reliability in real-world applications, particularly in critical areas such as fluid resuscitation.

The incorporation of uncertainty quantification in our modeling approach plays a pivotal role in advancing the design of optimal fluid management strategies in critical care scenarios. By precisely capturing uncertainty through BVAE, we can offer a subtle understanding of the inherent variability and unpredictability in healthcare scenarios. This detailed uncertainty information serves as a crucial input for developing model-based reinforcement learning (MBRL) control algorithms. In reinforcement learning (RL), the agent adapts its strategy by recognizing the need for increased exploration in the environment, when confronted with higher uncertainty in model's output. Leveraging this insight, our future direction involves utilizing the MBRL agent to dynamically adjust fluid management policies in response to higher uncertainty, actively exploring and gaining more knowledge about the evolving physiological state. This adaptive response enhances fluid management strategies, promising potential for future research and application in dynamic clinical settings.

While the current study focuses on hemorrhagic scenarios, the modular nature of the BVAE framework suggests potential applicability to diverse patient populations and medical conditions within critical care. To fortify its versatility, we will explore model's performance in different scenarios for the same subject, ensuring its robustness to intra-patient variability. Furthermore, we aim to evaluate its adaptability by testing model across different subjects, providing insights into its responsiveness to inter-patient variability. In addition, ongoing research and collaboration are essential to validate and customize the model for different healthcare settings, ensuring its robustness and effectiveness in enhancing clinical decision-making.

The current study primarily focused on fluid management strategies in hemorrhagic scenarios. However, the effects of multiple drug infusions, which are often a combination of critical care interventions, have not been incorporated into the models. Future research could explore the integration of multiple drugs and their interactions to provide a more comprehensive understanding of the complex physiological dynamics during treatment.

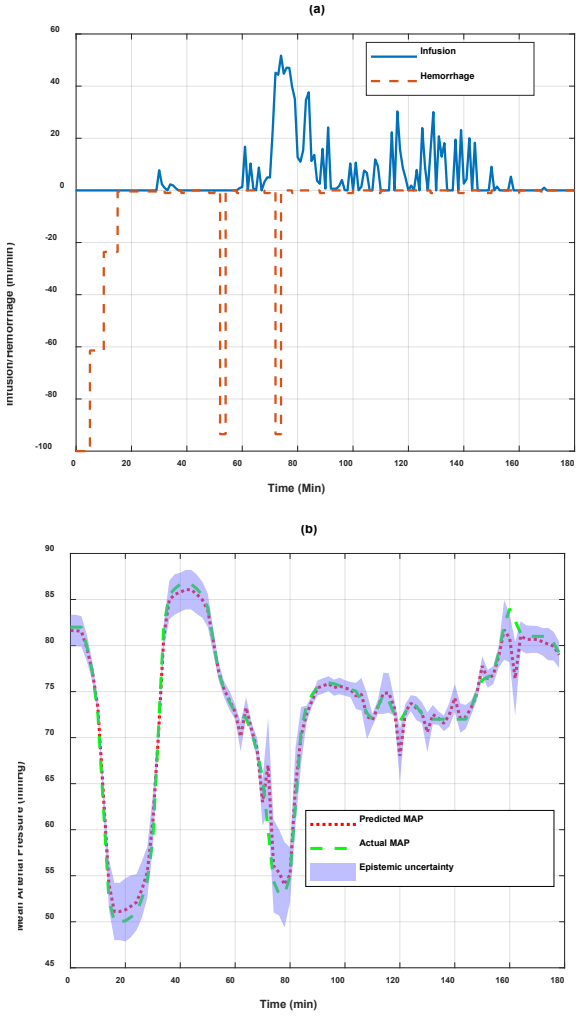


Fig. 2. (a) Infusion and hemorrhage profiles of an animal subject, (b) Comparing predictied MAP with actual data, along with uncertainity bounds

TABLE I. PERFORMANCE METRICS FOR ALL SUBJECTS

| | **RMSE (%)** | **MAE (%)** | **MDAPE (%)** |
|---|---|---|---|
| MEAN | 1.77 | 1.20 | 0.91 |
| STD | 0.44 | 0.29 | 0.30 |

This study emphasizes the significance of uncertainty modeling in fluid therapy, primarily centering around the impact of hemorrhagic events. Incorporating additional hemodynamic endpoints, such as blood volume and cardiac output, could potentially enhance the applicability of the proposed approach. Future directions may involve expanding the model to consider a broader range of hemodynamic variables, enabling a more comprehensive evaluation of fluid management strategies across various clinical scenarios.

## IV. CONCLUSION

In this study, we developed BVAE models to capture uncertainties in physiological systems, focusing specifically on fluid therapy in hemorrhagic scenarios. The simulation outcomes were promising, particularly in detecting hemorrhages by assigning increased uncertainty to OoD scenarios. We plan to conduct a comprehensive comparison with leading modeling techniques and perform an in-depth statistical analysis to further affirm the effectiveness of our methodology. Upon refining and validating our model's accuracy, we aim to develop an automated fluid management system that enhances decision-making in critical care.

## REFERENCES


[1] E. Estiri and H. Mirinejad, "Precision dosing in critical care: application of machine learning in fluid therapy," in *Proc IEEE Conf. Digital Health (ICDH)*, Chicago, IL, July 2023, pp. 348-352.

[2] E. Estiri and H. Mirinejad, "Model-free reinforcement learning for automated fluid administration in critical care," *arXiv preprint,* arXiv:2401.06299, 2024.

[3] H. Mirinejad, B. Parvinian, M. Ricks, Y. Zhang, S. Weininger, J. O. Hahn, and C. G. Scully, "Evaluation of fluid resuscitation control algorithms via a hardware-in-the-loop test bed," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 2, pp. 471-481, Feb. 2020.

[4] E. Estiri and H. Mirinejad, "Robust nonlinear state Space model identification for hemorrhage resuscitation," in *Proc IEEE Conf. Biomed. Health Inform. (BHI)*, Pittsburgh, PA, Oct. 2023, pp. 1-4.

[5] E. Estiri and H. Mirinejad, "Closed-loop conrtol of fluid resuscitation using reinforcement learning," *IEEE Access*, Dec. 2023.

[6] J. Grant and H. Mirinejad, "An optimal control approach for automated fluid resuscitation," *arXiv preprint,* arXiv:2312.06521, Dec. 2023.

[7] H. Mirinejad, A. E. Gaweda, M. E. Brier, J. M. Zurada, and T. Inanc, "Individualized drug dosing using RBF-Galerkin method: case of anemia management in chronic kidney disease," *Comput. Methods Programs Biomed.*, vol. 148, pp. 45-53, 2017.

[8] H. Mirinejad and T. Inanc, "Individualized anemia management using a radial basis function method, in *Proc. IEEE Great Lakes Biomedical Conf. ,* Milwaukee, WI, May 2015.

[9] H. Mirinejad and T. Inanc, "RBF method for optimal control of drug administration in the anemia of hemodialysis patients," in *Proc Northeast Bio. Conf. (NEBEC)*, Troy, NY, Apr. 2015.

[10] H. Mirinejad, "A radial basis function method for solving optimal control problems," Ph.D. Dissertation, University of Louisville, Louisville, KY, USA, May 2016.

[11] R. Eshel, E. J. Snider, and G. Avital, "Computer-assisted fluid therapy," *Curr. Anesthesiol. Rep.*, vol. 13 pp. 41-48, 2023.

[12] E. Estiri and H. Mirinejad, "Autoencoder learning and variational Gaussian inference for predicting mean arterial pressure in fluid resuscitation," in *IEEE SMC*, Honolulu, HI, Oct. 2023.

[13] D. Ruhe, G. Cina, M. Tonutti, D. de Bruin, and P. Elbers, "Bayesian modelling in practice: Using uncertainty to improve trustworthiness in medical applications," *arXiv preprint,* arXiv:1906.08619, 2019.

[14] E. Daxberger and J. M. Hernández-Lobato, "Bayesian variational autoencoders for unsupervised out-of-distribution detection," *arXiv preprint*, arXiv:1912.05651, 2019.

[15] N. Stahl, G. Falkman, A. Karlsson, and G. Mathiason, "Evaluation of uncertainty quantification in deep learning," *Information Proc. Management Uncert. Knowledge-Based Sys.*, May 2020, pp. 556-568.

[16] D. Dorbin and R. G. Hahn, "Volume kinetics of Ringer's solution in hypovolemic volunteers," *Anesthesiology*, vol. 90, pp. 81–91, 1999.

[17] G. C. Kramer, M. P. Kinsky, D. S. Prough, J. Salinas, J. L. Sondeen, M. L. Hazel-Scerbo, and C. E. Mitchell, "Closed-loop control of fluid therapy for treatment of hypovolemia," *J. Trauma Acute Care Surg.*, vol. 64, no. 4, pp. S333-S341, 2008.

[18] P. Salvi, "Mean arterial pressure," *Pulse Waves*: *How Vascular Hemodynamics Affects Blood Pressure,* pp. 3-7, 2012.

[19] X. X. Wu and J. G. Liu, "A new Early Stopping algorithm for improving neural network generalization," in *Intern. Conf. Intelligent Comput. Techn. Autom.*, Changsha, China, 2009, pp. 15-18.

[20] A. K. Kuchibhotla, L. D. Brown, A. Buja, and J. Cai, "All of linear regression," *arXiv preprint,* arXiv, 1910.06386, 2019.