Analyzing Fetal Heart Rate Patterns via Latent Representations with Sequential Variational AutoEncoder (SeqVAE)

Mahdi Shamsi[1], Aditi Lahiri[2], Yvonne Wu[3], Lawrence Gerstley[2], Michael Kuzniewicz[2], Cornet Marie-Coralie[3], Emily Hamilton[1,4], Robert Kearney[1], Philip Warrick[1,4]

[1] McGill University, Montreal, Canada
[2] Kaiser Permanente Northern California, Oakland, USA
[3] University of California, San Francisco, USA
[4] PeriGen Inc., Montreal, Canada

Abstract

Fetal heart rate (FHR) monitoring is a non-invasive method for assessing fetal well-being during labor, providing vital insights into potential fetal distress and the risk of hypoxic-ischemic encephalopathy (HIE). However, traditional clinical methods for interpreting FHR signals are often subjective. In this paper, we propose a novel Sequential Variational Autoencoder (SeqVAE) model, inspired by the Variational Recurrent Neural Network (VRNN), to learn latent representations of FHR signals. The SeqVAE model is designed to capture the temporal dependencies within the FHR signals while maintaining a probabilistic framework, resulting in smoother and more informative latent representations compared to the VRNN. These latent representations are then utilized for the classification of FHR signals into healthy and HIE cases. Our experiments demonstrate that the SeqVAE model outperforms the VRNN in generating latent representations that align closely with significant physiological events in FHR signals.

1. Introduction

Hypoxic-ischemic encephalopathy (HIE) is a severe neonatal condition caused by insufficient oxygen and blood flow to the infant's brain during the perinatal period. This condition can lead to neonatal death or long-term neurodevelopmental impairments. Early detection and intervention are essential to mitigate the severity of HIE. Cardiotocography (CTG) records fetal heart rate (FHR) and maternal uterine pressure (UP) for continuous non-invasive monitoring of intrapartum fetal well-being during labor. Clinicians use FHR monitoring as a standard of care to assess fetal distress, which can signal hypoxia and the potential for subsequent HIE. Fetal heart rhythm and variability, usually described in terms of events such as baseline, accelerations and decelerations, provide valuable insight about the fetus's response to labor, which is driven by maternal uterine contractions [1]. Specific FHR patterns, such as late decelerations and reduced variability, have been linked to hypoxic events [2]. However, the subjective nature of FHR interpretation leads to significant variability among clinicians in assessing CTG recordings, and inter- and intra-observer agreement is low. Additionally, these evaluations often tend to be very unspecific, making it difficult to draw precise conclusions about fetal health. Machine-learning approaches have recently gained significant attention for processing and analysis of FHR signals. Machine-learning and deep-learning algorithms can detect patterns and information in biomedical signals that are not visually apparent [3]. In this context, unsupervised algorithms for learning latent representations of time series data have been extensively explored and applied across various domains, including biomedical informatics [4]. These algorithms are designed to learn meaningful, compact representations of the input data that capture essential features and underlying patterns within the time series. These can be used for downstream tasks such as classification, anomaly detection and clustering [5]. Variational Autoencoders (VAEs) [6] are unsupervised algorithms that introduce a probabilistic framework for learning latent representations in a latent space parameterized by mean and variance vectors. Building on the VAE foundation, Variational Recurrent Neural Networks (VRNNs) introduce a dynamic advancement that enables VRNNs to model the evolving nature of time-series data. [7]. VRNNs extend the capabilities of VAEs by incorporating the temporal dependencies inherent in biomedical signals, providing a more comprehensive encoding of the evolving patterns in the FHR signal. This paper proposes a Sequential VAE (SeqVAE) model that used a modified VRNN architecture to improve the learning of time-series latent representations. The SeqVAE architecture, depicted in Fig-

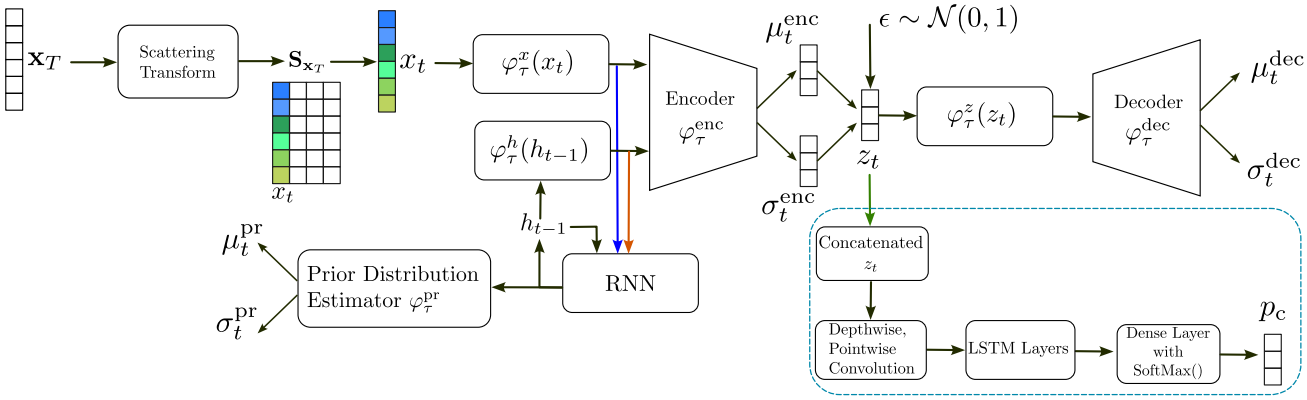

Figure 1: Sequential VAE model architecture with classification head shown within the blue dashed box.

ure 1, enhances the architecture of encoder, decoder and RNN blocks to improve the capture of temporal dynamics in the latent representation while maintaining the probabilistic framework of VAEs. In this work, we apply SeqVAE to FHR signals to learn robust latent representations that capture informative features. These learned representations are used to train a supervised classifier of healthy and HIE records.

2. Methods

Data This study used a comprehensive fetal monitoring database constructed for the investigation of perinatal hypoxic-ischemic encephalopathy (HIE) [8]. This database includes over 250,000 de-identified electronic fetal monitoring (EFM) records linked to rich clinical data about the infant, mother, pregnancy, labor, and birth outcomes. The database covers births from 15 Kaiser Permanente hospitals in Northern California between 2011 and 2019. The subset of the dataset used in this paper comprised two groups: 1) 24,295 singleton vaginal births with healthy outcomes determined with blood pH $>$ 7 and a base deficit of $<$ 10 mmol/L; 2) 418 infants who developed HIE, characterised by the presence of acidosis along with neurological signs of encephalopathy. Each FHR recording, sampled at 4Hz, was segmented into 20-minute epochs. We used 101,369 high-quality epochs from 24295 healthy records to pre-train the SeqVAE model to learn the latent representation of FHR signals. 418 HIE records and a subset of 418 healthy records excluded from pre-training, were used for training the classifier.

Preprocessing Step To extract robust features from raw FHR signals, we apply the wavelet scattering transform, which provides representations that are translation-invariant and stable to deformation [9]. The zeroth-order and first-order scattering coefficients were extracted and used as the input features to the SeqVAE model. $\mathbf{S}_{\mathbf{x}_T}$ denotes the wavelet scattering transform of the multi-channel input signal with sampling rate of 4 Hz, $\mathbf{x}_\mathbf{T}$, with length T. After the scattering transform, the resulting coefficients, x_t, are decimated by a factor of 16, reducing the sampling rate to 0.25 Hz, while retaining the higher-frequency band information in the lower scales. In Figure 1, x_t denotes the scattering transform coefficients at time step t at the reduced sampling rate.

SeqVAE Model This model incorporates a VAE at each time step. The latent representation at time step t, denoted as z_t, is conditioned on the input at that time x_t, as well as the previous hidden state h_{t-1}. This allows the model to capture both the current input characteristics and the temporal dependencies encoded in the RNN's hidden state. Prior to the encoder block, the linear neural networks $\varphi_\tau^x(x_t)$ and $\varphi_\tau^h(h_{t-1})$ extract features from the current input sample and the previous hidden state, respectively. These extracted features are then combined and passed to the encoder block, which learns the parameters of the latent distribution, mean μ_t^{enc} and standard deviation σ_t^{enc} vectors. These parameters define the Gaussian distribution from which the latent variable z_t is sampled. The distribution of the latent variable z_t conditioned on the input up to time step t is:

$$q(z_t \mid x_{\leq t}) \sim \mathcal{N}(\mu_t^{\text{end}}, \text{diag}((\sigma_t^{\text{enc}})^2)) \quad (1)$$

$$[\mu_t^{\text{enc}}, \sigma_t^{\text{enc}}] = \varphi_\tau^{\text{enc}}(\varphi_\tau^x(x_t), \varphi_\tau^h(h_{t-1})), \quad (2)$$

Unlike the original VRNN model, the SeqVAE hidden state at time t is computed as a function of both the features extracted from the current input data, $\varphi_\tau^x(x_t)$, and those derived from the previous hidden state, $\varphi_\tau^h(h_{t-1})$. This modification isolates the RNN block from the decoder. Therefore, the hidden state updated as:

$$h_t = f(\varphi_\tau^x(x_t), \varphi_\tau^h(h_{t-1}), h_{t-1}) \quad (3)$$

SeqVAE model decouples the RNN from the latent representation, and the hidden state is only updated with respect

to the current time step and the history of the input sequence. Moreover, the prior of the latent random variable is a function of current time step and previous hidden state:

$$p(z_t|x_{\leq t}) \sim \mathcal{N}(\mu_{0,t}, \mathrm{diag}(\sigma^2_{0,t})), \tag{4}$$

$$[\mu_{0,t}, \sigma_{0,t}] = \varphi^{\mathrm{pr}}_\tau(\varphi^x_\tau(x_t), \varphi^h_\tau(h_{t-1})), \tag{5}$$

where $\varphi^{\mathrm{pr}}_\tau$, is also a linear neural network. The prior distribution $p(z_t|x_{\leq t})$ represents the model's belief about the latent variable z_t before observing the current input and enables the model in a generative mode. Before feeding the z_t to the decoder, a linear neural network, $\varphi^z_\tau(z_t)$, extracts the features of the latent representation. The SeqVAE decoder also differs from the VRNN model. The SeqVAE posterior distribution is not directly influenced by the hidden state of the RNN; instead, it relies only on the latent representation vector z_t. Consequently, SeqVAE decouples the decoder from the RNN, thereby encouraging a smoother latent space for improved stability and temporally coherent representations. The outputs of the decoder μ^{dec}_t and σ^{dec}_t are the mean $\widehat{x}_t$ and variance $\widehat{\sigma}_{x_t}$ (i.e., uncertainty) of the reconstruction at time step t, respectively.

$$x_t \mid z_t \sim \mathcal{N}(\mu^{\mathrm{dec}}_t, \mathrm{diag}((\sigma^{\mathrm{dec}}_t)^2)) \tag{6}$$

$$[\mu^{\mathrm{dec}}_t, \sigma^{\mathrm{dec}}_t] = \varphi^{\mathrm{dec}}_\tau(\varphi^z_\tau(z_t)), \tag{7}$$

Consequently, the loss function of the SeqVAE is:

$$\mathbb{E}_{q(z_{\leq T}|x_{\leq T})}\left[\sum_{t=1}^{T}(-\beta \mathrm{KL}(q(z_t \mid x_{\leq t}) \parallel p(z_t \mid x_{\leq t})) + \log p(x_t \mid z_t))\right] \tag{8}$$

where hyperparameter β regularizes the Kullback–Leibler (KL) divergence term, to balance two competing model objectives: encoder effectiveness in conditioning the latent space on the input in $p(z|x)$ vs. decoder reconstruction fidelity in $\widehat{x}_t$.

Classification The classification head (Figure 1) feeds the concatenated latent representations into a depth-wise and point-wise convolutional block with a kernel size of 1. This block output is passed to a four-layer LSTM with progressively decreasing hidden state sizes of 36, 27, 18, and 9. Finally, a linear layer produces the logits for each class. The model was trained using cross-entropy loss.

3. Results

VRNN vs. SeqVAE Both models were trained under identical conditions, using the same hyperparameters. Specifically, both models utilized an LSTM with four layers with hidden state dimension 90, and a latent dimension of 9. The β was set to 0.0001. As shown in Figure 2, which presents a raw FHR signal alongside its scattering transform, SeqVAE provides a more informative latent representation that aligns with significant FHR events. For each model, Table 1 compares the VAF [10] mean and standard deviation for each scattering transform coefficient, calculated across all test samples. The results demonstrate that both models are able to reconstruct the scattering transform accurately; however the SeqVAE model achieves marginally better VAF. Notably, the VRNN model exhibits a significant drop in fidelity for coefficients 3 and 4, which correspond to the important FHR variability bands at 0.5 Hz and 0.25 Hz. These bands are meaningful for clinicians as they are indicative of FHR variability. In contrast, the SeqVAE model shows less severe drops in these coefficients. While both models achieve accurate reconstruction of the scattering transform, SeqVAE's latent representation is better correlated with the physiological events in the FHR signal. In contrast, the reconstruction performance of the VRNN model is primarily driven by the RNN block on the decoder side.

Classification Results Figure 3 illustrates the test-set performance of the classification model over the last six hours before birth across 10 folds of 10-fold cross-validation. The folds are carefully partitioned to ensure a balanced representation of both healthy and HIE recordings within each fold. These accuracy curves show the model's average performance with confidence bounds in classifying FHR signals as either healthy or HIE. The confidence bounds are determined with respect to the minimum and maximum accuracy across all folds. The accuracy for healthy cases generally remains higher and more stable, with values fluctuating between approximately 0.6 and 0.9 across folds. In classification, it is crucial to maintain a low false positive rate to avoid unnecessary Cesarean sections, which can increase costs and pose additional risks to both mother and infant. In contrast, the accuracy for HIE cases is lower and exhibits greater variability. Notably, there is a trend for improved HIE detection as the time approaches birth. In some folds, this accuracy reaches 0.8 in the final hour before birth. This suggests that the model becomes more effective at identifying HIE as the critical time window approaches birth.

4. Conclusion

In this work, we introduced the Sequential Variational Autoencoder (SeqVAE) model, a novel approach designed to learn latent representation of FHR signals and capture the temporal dynamics for the detection of hypoxic-ischemic encephalopathy (HIE). Our comparative analysis demonstrated that SeqVAE outperforms VRNN in generating meaningful latent features. The classification results indicated a reasonable level of discriminative power, particularly with an increase in accuracy for detecting HIE

Table 1: Reconstruction mean %VAF (± standard deviation) per scattering coefficient for SeqVAE and VRNN models.

Coefficient	0	1	2	3	4	5	6	7	8	9	10
SeqVAE	99.4 (0.860)	97.3 (0.077)	97.8 (0.0215)	95.9 (0.0245)	97.8 (0.0191)	99.3 (0.0047)	99.7 (0.0010)	99.8 (0.0006)	99.8 (0.0008)	99.8 (0.0014)	99.8 (0.0021)
VRNN	99.3 (0.0106)	95.0 (0.0406)	95.1 (0.0292)	91.1 (0.0583)	91.7 (0.0440)	95.0 (0.0260)	99.1 (0.0050)	99.6 (0.0032)	99.6 (0.0045)	99.5 (0.0069)	99.5 (0.0076)

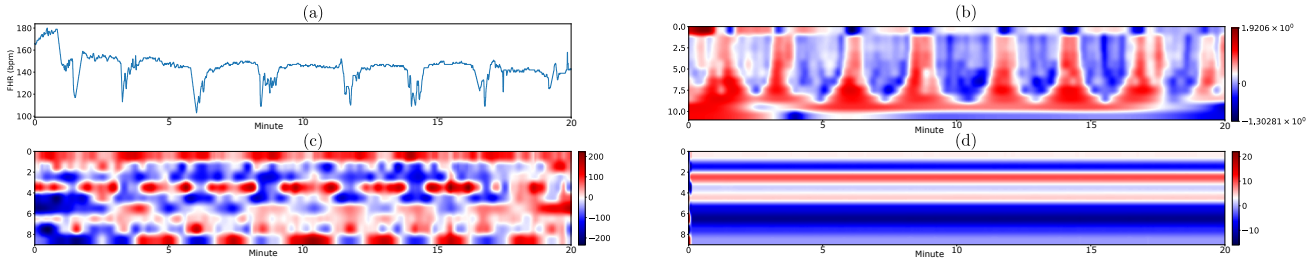

Figure 2: (a) FHR signal (b) True scattering transform (c) SeqVAE latent representation (d)VRNN latent representation.

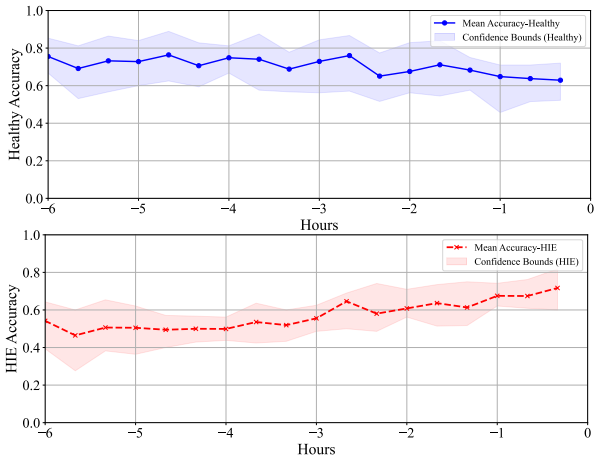

Figure 3: Test set AUROC and accuracy for healthy and HIE classes over time before birth.

cases as the epochs approach the time of birth. Future research should focus on improving the overall model architecture, particularly the classification head, to enhance the model's accuracy and robustness in detecting HIE cases.

Acknowledgments

This study was supported by funding provided by the Bill & Melinda Gates Foundation and the National Institutes of Health.

References

[1] Freeman RK, Garite TJ, Nageotte MP. Fetal heart rate monitoring. Lippincott Williams & Wilkins, 2003.

[2] Larma JD, Silva AM, Holcroft CJ, Thompson RE, Donohue PK, Graham EM. Intrapartum electronic fetal heart rate monitoring and the identification of metabolic acidosis and hypoxic-ischemic encephalopathy. American journal of obstetrics and gynecology 2007;197(3):301–e1.

[3] Signorini MG, Pini N, Malovini A, Bellazzi R, Magenes G. Integrating machine learning techniques and physiology based heart rate features for antepartum fetal monitoring. Computer methods and programs in biomedicine 2020; 185:105015.

[4] Wei R, Mahmood A. Recent advances in variational autoencoders with representation learning for biomedical informatics: A survey. Ieee Access 2020;9:4939–4956.

[5] Eldele E, Ragab M, Chen Z, Wu M, Kwoh CK, Li X, Guan C. Time-series representation learning via temporal and contextual contrasting. arXiv preprint arXiv210614112 2021;.

[6] Kingma DP, Welling M. Auto-encoding variational bayes. arXiv preprint arXiv13126114 2013;.

[7] Chung J, Kastner K, Dinh L, Goel K, Courville AC, Bengio Y. A recurrent latent variable model for sequential data. Advances in neural information processing systems 2015; 28.

[8] Kearney RE, Wu YW, Vargas-Calixto J, Kuzniewicz MW, Cornet MC, Forquer H, Gerstley L, Hamilton E, Warrick PA. Construction of a comprehensive fetal monitoring database for the study of perinatal hypoxic ischemic encephalopathy. MethodsX 2024;12:102664.

[9] Bruna J, Mallat S. Invariant scattering convolution networks. IEEE transactions on pattern analysis and machine intelligence 2013;35(8):1872–1886.

[10] Peterson RA. A meta-analysis of variance accounted for and factor loadings in exploratory factor analysis. Marketing letters 2000;11:261–275.

Address for correspondence:

Mahdi Shamsi

Address: 3775 University - Montreal - QC - H3A 2B4

E-mail: mahdi.shamsi@mcgill.ca