ORIGINAL RESEARCH **Open access**

A Variational Recurrent Neural Network with Stochastic Attention for Imputation of Irregularly Sampled ICU Time Series under Non-Random Missingness

Minh Tran[1*], Duc Pham[1]

Abstract

Intensive care unit (ICU) data consist of high-frequency multivariate time series, including vital signs, laboratory results, and hemodynamic variables, which are crucial for clinical decision-making and predictive modeling. However, these data are frequently incomplete due to monitor interruptions, clinical workflows, and selective measurement, with missing rates ranging from 20% to over 80% depending on the variable. Missingness in ICU time series is often not random, as sicker patients tend to be monitored more frequently, creating a missing not at random (MNAR) mechanism. Conventional imputation methods such as mean filling, interpolation, and multiple imputation assume random missingness and therefore introduce bias and distort clinical signals under MNAR conditions. We propose a variational recurrent neural network (VRNN) with stochastic attention to impute ICU time series under MNAR settings. The framework integrates latent state modeling of physiological dynamics, stochastic attention over observed measurements using Gumbel-Softmax sampling, and a missingness pattern encoder that explicitly models the observation process. An imputation decoder generates probabilistic estimates of missing values conditioned on latent states, attention context, and missingness structure. This framework enables uncertainty-aware and potentially unbiased imputation in ICU time series by jointly modeling physiological dynamics and missingness mechanisms. It combines variational inference and stochastic attention to address systematic bias in conventional approaches, with future work needed to validate performance on real-world ICU datasets.

Keywords Intensive care unit, Uncertainty quantification, Variational recurrent neural network, Stochastic attention, Missing not at random, Time series imputation

*Correspondence:
Minh Tran
minh.tran@gmail.com
[1] Department of Intelligent Healthcare Analytics, Ho Chi Minh City University of Technology, Ho Chi Minh City, Vietnam

Introduction

Intensive care unit monitoring systems generate multivariate time series at varying frequencies, with vital signs such as heart rate, blood pressure, respiratory rate, oxygen saturation, and temperature recorded every minute to every hour alongside less frequent laboratory measurements and medication administrations [1, 2]. These high-dimensional temporal data streams form the foundation for clinical prediction models targeting sepsis onset, acute kidney injury, mortality risk, and hemodynamic decompensation, yet their utility is fundamentally compromised by pervasive missing values. Empirical analyses of ICU databases including MIMIC-IV and the eICU Collaborative Research Database have documented missingness rates ranging from 20% for frequently charted vital signs to over 80% for intermittent laboratory tests, with missingness patterns that correlate strongly with patient acuity and clinical workflow [3, 4]. The scale and structure of

these missing data present a methodological challenge that standard preprocessing approaches fail to address adequately.

The missingness patterns observed in ICU time series are predominantly non-random, corresponding to the missing not at random mechanism in Rubin's taxonomy, where the probability of observing a value depends on the unobserved value itself [5, 6]. When nurses measure blood pressure more frequently during hypotensive episodes, the resulting data exhibit a systematic relationship between the severity of hypotension and the density of blood pressure observations, with normotensive periods selectively under-sampled. Similarly, patients with deteriorating respiratory status receive more frequent oxygen saturation checks, creating an informative missingness structure where gaps in the record convey clinically meaningful information about patient stability. This dependence of measurement frequency on unobserved physiology violates the assumptions underlying standard imputation approaches and creates a structural challenge that requires explicit modeling of the selection mechanism [7, 8].

Standard imputation methods applied to ICU time series—including mean imputation, median imputation, forward filling of the last observation carried forward, linear interpolation between adjacent measurements, and multivariate approaches such as multiple imputation by chained equations—implicitly or explicitly assume that missingness occurs completely at random or at random conditional on observed data [9, 10]. When these assumptions are violated, as they routinely are in critical care settings, the resulting imputed values systematically distort the distribution of physiological parameters, attenuating the very clinical signals that prediction models seek to detect. Forward filling, for example, propagates values measured during periods of instability into subsequent stable periods, artificially inflating the apparent severity of illness, while mean imputation shrinks extreme values toward the population average, blunting the discriminative features that distinguish deteriorating patients from stable ones [11, 12]. These biases propagate through downstream prediction pipelines, degrading model calibration and potentially leading to missed or delayed clinical alerts.

This paper proposes a conceptual framework that addresses the fundamental challenge of MNAR imputation in ICU time series through the integration of variational recurrent neural networks with stochastic attention mechanisms. The framework jointly models the underlying physiological dynamics via a VRNN that learns probabilistic latent state transitions across irregular time intervals, the observation selection process via a missing pattern encoder that captures informative missingness relationships, and the imputation task via a decoder that generates predictive distributions with calibrated uncertainty [13, 14]. The remainder of this manuscript develops the theoretical foundations of the framework, beginning with background on missing data mechanisms and existing imputation approaches, proceeding through the architectural components and their integration, and concluding with evaluation strategies and limitations [15].

Table 1 clarifies how the proposed framework differs conceptually from conventional ICU imputation strategies by treating missingness as an informative clinical process rather than an ignorable preprocessing problem.

Table 1. Conceptual Differentiation of ICU Time-Series Imputation Approaches Under Non-Random Missingness

Imputation approach	Assumption about missingness	Treatment of ICU measurement frequency	Main risk under MNAR	What the proposed framework adds
Mean or median imputation	Missingness is treated as ignorable	Measurement gaps are not interpreted clinically	Shrinks abnormal physiology toward population averages	Preserves uncertainty and avoids deterministic distributional compression
Forward filling / last observation carried forward	Recent observed value remains clinically representative	Dense measurement during instability may be overextended into later periods	Propagates crisis-period values beyond their valid temporal window	Uses latent dynamics and time gaps to model physiological evolution probabilistically

Linear interpolation	Physiological change is smooth between observations	Missingness pattern is ignored	Masks abrupt deterioration or intervention-related shifts	Allows nonlinear latent transitions and uncertainty expansion across long gaps
MICE / standard multiple imputation	Missingness is MAR conditional on observed variables	Measurement decisions are only indirectly represented	Biased when observation probability depends on unobserved physiology	Explicitly models the selection process linking latent state and observation probability
Deterministic neural imputation	Missingness may be encoded as input features	Mask can be used but often without a missingness mechanism	Learns correlations without distinguishing physiology from observation process	Separates latent physiological dynamics, stochastic attention, and MNAR selection modeling
Proposed VRNN with stochastic attention	Missingness may be MNAR and clinically informative	Measurement density is modeled as part of the data-generating process	Residual bias remains if MNAR assumptions are misspecified	Produces distributional imputations with calibrated uncertainty and explicit MNAR-aware structure

Background

Missing data mechanisms

The statistical literature classifies missing data mechanisms into three categories that describe the relationship between the probability of missingness and the underlying data values [16]. Missing completely at random occurs when the probability of a value being missing is independent of both observed and unobserved data, as might arise from random sensor disconnection or accidental data deletion during transfer. Missing at random occurs when the missingness probability depends on observed data but remains conditionally independent of the unobserved values given the observed ones, a mechanism that can be addressed through likelihood-based methods and multiple imputation when the relevant conditioning variables are available [17]. Missing not at random describes the most challenging scenario, where the probability of missingness depends directly on the unobserved values themselves, creating a fundamental identifiability problem that requires untestable assumptions or auxiliary data to resolve. In clinical contexts, MNAR mechanisms are pervasive, arising whenever clinical judgment about a patient's state influences the decision to obtain measurements [18, 19].

MNAR in intensive care

The intensive care environment generates MNAR missingness through the clinical decision-making processes that govern patient monitoring and assessment [20]. Hypotension prompts more frequent blood pressure measurements, creating a sampling density that tracks inversely with the unobserved blood pressure trajectory and systematically underrepresents normotensive periods. Desaturation episodes trigger increased oxygen saturation monitoring and arterial blood gas sampling, producing missingness patterns that correlate with the severity of respiratory compromise. The Glasgow Coma Scale score, which requires active patient assessment, is measured more frequently when patients exhibit altered mental status, with periods of normal neurological function documented less intensively. These patterns of informative measurement extend to laboratory testing, where electrolyte panels, complete blood counts, and coagulation studies are ordered more frequently during clinical instability, generating data that are fundamentally non-representative of the patient's entire ICU stay [21]. The consequence of ignoring this mechanism is that imputed values are systematically biased toward the physiological states observed during periods of clinical concern, distorting the true distribution of patient physiology.

Variational recurrent neural networks

Variational recurrent neural networks extend the variational autoencoder framework to sequential data by introducing a latent random variable at each time step that conditions both the generation of observed data and the transition to the subsequent latent state [22]. The joint distribution over a sequence of observations and latent states factorizes as the product of a prior over the initial latent state, transition distributions that evolve the latent state forward in time, and emission distributions that generate observations from the corresponding latent states. Learning proceeds by maximizing the evidence lower bound, which decomposes as the sum over time steps of the expected log-likelihood of observations given latent states minus the Kullback-Leibler divergence between the approximate posterior and the prior at each step. The approximate posterior is parameterized by a recurrent neural network that processes observations forward in time, enabling the latent state to summarize all information available up to the current time point. In the context of irregularly sampled time series, the VRNN can incorporate time intervals between observations into its transition dynamics, learning to propagate uncertainty forward during longer gaps in the observation stream [23, 24].

Attention mechanisms for missing data

Attention mechanisms provide a flexible framework for weighting the contribution of different observations to a prediction or imputation task based on their learned relevance [25]. When applied to time series with missing values, attention architectures must address the dual challenge of identifying which time points contain informative measurements and accounting for the missingness context in which those measurements occur. Deterministic attention approaches compute scalar relevance scores for each historical observation and aggregate their contributions through a weighted sum, with missing values typically masked or replaced with a learned embedding. Recent work has explored stochastic attention mechanisms that sample attention weights from a probability distribution, capturing uncertainty about which observations are most relevant for a given imputation target. These approaches can incorporate missingness indicators as additional inputs to the attention computation, enabling the model to learn relationships between missingness patterns and the informativeness of available observations [26]. The integration of stochastic attention with variational sequence models offers a principled approach to handling the joint uncertainty from missing values and latent state estimation.

Framework Overview

High-level architecture

The proposed framework processes multivariate ICU time series as a sequence of observation vectors accompanied by binary missingness masks indicating which variables were recorded at each time step [27]. The architecture comprises four interconnected modules that operate within a variational inference framework: the VRNN backbone maintains a probabilistic latent state that captures the patient's underlying physiological trajectory, the stochastic attention mechanism computes observation weights that adaptively aggregate information from available measurements, the missing pattern encoder learns the parameters governing which values are likely to be observed given the latent state, and the imputation decoder generates predictive distributions for missing values by conditioning on the latent state, attended context, and estimated missingness parameters. The VRNN processes the time series forward, updating its latent state distribution at each time step using both the observed values at the current step and the attention-weighted summary of relevant historical information. The missing pattern encoder takes the latent state as input and outputs parameters of a Bernoulli distribution over observation indicators, capturing the MNAR mechanism where measurement probability depends on the unobserved or partially observed physiological state.

Figure 1 presents the proposed hierarchical architecture for MNAR-aware ICU time-series imputation using a variational recurrent neural network, stochastic mask-aware attention, explicit selection modeling, and uncertainty-calibrated predictive decoding.

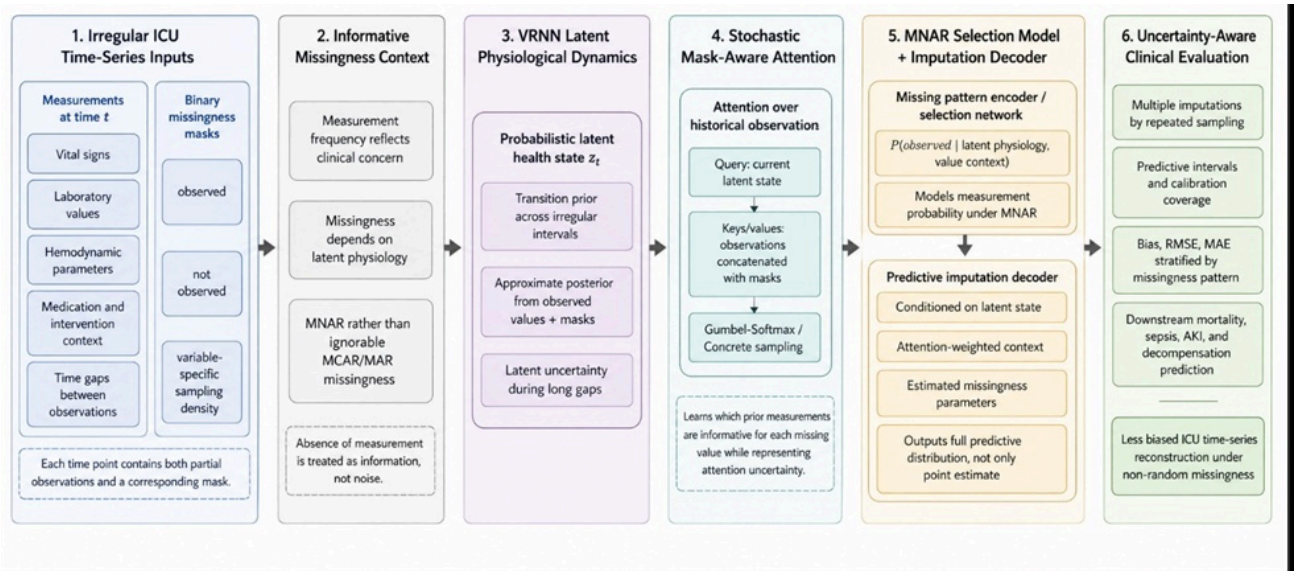

Figure 1. Hierarchical architecture of an MNAR-aware variational recurrent neural network with stochastic attention for ICU time-series imputation

Core assumptions

The framework rests on several assumptions that define its scope of applicability and guide its architectural design [28]. Temporal ordering is maintained across the sequence, with the latent state at each time step depending on the history of states and observations up to that point, an assumption that aligns with the natural progression of physiological processes in critical care. The missingness mechanism is assumed to depend on the latent health state that the VRNN learns to represent, operationalizing the clinical intuition that measurement decisions are driven by an assessment of patient status that may not be fully captured in the observed data. Sufficient observed data must be available for training the variational inference procedure, as the model must learn both the physiological dynamics and the selection mechanism from the joint distribution of observations and missingness patterns. The framework does not assume a specific parametric form for the physiological dynamics, instead relying on the expressive capacity of neural network function approximators within the VRNN architecture.

Design principles

Three design principles guide the architectural choices embodied in the framework [29]. Unbiased imputation under MNAR requires that the model explicitly represents the selection mechanism and integrates it into the imputation process, avoiding the systematic distortion that arises when missingness is assumed ignorable. The VRNN and stochastic attention components address this principle by conditioning imputations on the estimated probability of observation and adaptively weighting information from time points with different missingness patterns. Uncertainty quantification constitutes the second principle, with the variational formulation providing a natural framework for representing both the uncertainty in the latent state trajectory and the uncertainty in the imputed values through the predictive posterior distribution. The probabilistic nature of the imputation decoder supports multiple imputation, enabling downstream analyses to propagate imputation uncertainty through to their conclusions. The third principle emphasizes modularity and extensibility, with the VRNN, attention, missing pattern encoder, and imputation decoder designed as separable components that can be independently modified or replaced as methodological advances emerge in each area.

Variational Recurrent Neural Network

Latent state transition

The VRNN maintains a latent state vector at each time step that summarizes the patient's physiological trajectory up to and including that step, with prior and posterior distributions over this latent state parameterized by neural network functions [1]. The prior distribution over the latent state at time step t, conditioned on the latent state at the previous step, is modeled as a diagonal Gaussian whose mean and log-variance are computed by a transition network that takes the previous latent state as input. This prior captures the model's belief about how the latent physiology evolves during the interval between observations, with the variance increasing during longer gaps to reflect growing uncertainty about the patient's unobserved state. The approximate posterior distribution is parameterized by an inference network that processes both the previous latent state and the current observation vector, incorporating information about missingness by conditioning on the binary mask that indicates which variables are observed. The KL divergence between the posterior and prior at each time step penalizes latent representations that deviate from the expected dynamics, regularizing the learned trajectory while allowing deviations when supported by observed data [2, 3].

Generation model

Conditioned on the latent state at each time step, the generation model produces predictive distributions over all variables in the observation vector, including those that are missing and require imputation [4]. The emission distribution for each variable is parameterized as a Gaussian whose mean and variance are computed by a decoder network that takes the latent state and the attention-weighted context as input. For variables that are observed, the log-likelihood under this distribution contributes to the reconstruction term in the evidence lower bound, driving the latent state to capture features that are predictive of the observed physiology. For variables that are missing, the predictive distribution provides the imputed values and their associated uncertainty, with the mean serving as a point estimate and the variance quantifying confidence in that estimate. The generation model can incorporate variable-specific decoders that account for differences in scale, distributional form, and physiological plausibility across vital signs, laboratory values, and other clinical measurements [5, 13].

Stochastic Attention Mechanism

Missingness mask attention

The missingness mask attention component computes relevance weights for historical observations that incorporate both the temporal relationship between time steps and the pattern of which variables were observed at each step [6]. At each time step t, the attention mechanism receives as input the sequence of observation vectors and corresponding binary masks for all previous time steps, along with an embedding of the current latent state that serves as the query. The key and value vectors for each historical time step are computed by transforming the observation vector concatenated with the missingness mask, enabling the model to learn that an observation of blood pressure carries different information when other vital signs were simultaneously recorded versus when it was the sole measurement. The resulting attention weights represent the learned importance of each historical time point for imputing the current missing values, with the model able to attend selectively to time periods where variables correlated with the missing value were observed. This masking-aware attention architecture avoids the pitfall of treating all observations as equally informative regardless of the completeness of their context [7, 21].

Stochastic attention sampling

Rather than computing deterministic attention weights through a softmax operation, the stochastic attention mechanism samples weights from a Concrete distribution, which provides a continuous relaxation of the discrete categorical distribution that is differentiable with respect to its parameters [8]. The Gumbel-Softmax reparameterization enables gradient-based optimization of the attention parameters while maintaining stochasticity in the weight assignment, capturing uncertainty about which historical observations are most relevant for a given imputation. At each forward pass during training and imputation, the attention weights are sampled from this distribution, with the temperature parameter controlling the entropy of the sampling process and annealing from high to low values during training. This stochastic

formulation acknowledges that multiple attention configurations may be equally plausible given the available evidence, and the variability across samples contributes to the overall uncertainty quantification. The stochastic attention complements the VRNN's variational inference by introducing additional randomness in the information aggregation step that is marginalized through sampling [9, 22].

Missingness mechanism modeling

The missingness mechanism is modeled through a selection network that takes the latent state and, when available, the true observation as input and outputs the probability of each variable being observed at each time step [10]. This formulation corresponds to a selection model in the missing data taxonomy, where the joint distribution of observations and missingness indicators factorizes into the marginal distribution of the complete data and the conditional distribution of the missingness given the complete data. During training, for variables that are observed, the selection network learns to predict the observation indicators from the latent state and the observed values, encouraging the latent state to capture features that explain both the physiology and the measurement process. During imputation, for variables that are missing, the selection network's estimate of the observation probability informs the imputation by downweighting the influence of time points where the missing variable would likely have been observed had it been abnormal. This explicit parameterization of the MNAR mechanism distinguishes the framework from imputation methods that treat missingness as ignorable, addressing the bias that arises when the probability of observation depends on the unobserved value [11, 23].

Non-Random Missing Pattern Modeling

Informative missingness

The missingness mask itself carries clinically meaningful information about the patient's underlying health state, and the framework is designed to exploit this informative missingness rather than treating it as a nuisance to be removed [12]. When a blood pressure measurement is absent at a scheduled recording time, this absence may reflect a clinical decision that the patient's cardiovascular status appeared stable enough to defer measurement, or conversely, that the patient was undergoing an intervention that precluded cuff inflation. The missing pattern encoder learns to map these patterns of presence and absence to latent representations that inform the VRNN's state estimation, effectively using the measurement density as an additional physiological signal. This approach aligns with clinical practice, where an experienced clinician interprets the frequency of monitoring as an indicator of concern, and with recent empirical findings that incorporating missingness indicators as features improves prediction model performance in critical care settings [14, 15]. The framework formalizes this clinical intuition through the joint modeling of observations and observation indicators.

Selection model versus pattern mixture model

The framework adopts a selection model factorization, modeling the joint distribution of the complete data and the missingness mechanism as the product of the marginal distribution of the complete data and the conditional distribution of the missingness given the complete data [16, 18]. This factorization directly parameterizes the MNAR mechanism as the probability of observing each variable given its potentially unobserved value and the latent state, enabling the model to learn that extreme physiological values are more likely to be documented. An alternative factorization, the pattern mixture model, would stratify the data by missingness pattern and fit separate imputation models within each stratum, an approach that handles MNAR but requires sufficient observations within each pattern and does not provide a direct parameterization of the selection mechanism. The selection model formulation within the variational framework offers the advantage of sharing statistical strength across missingness patterns through the latent state representation while maintaining an explicit, interpretable parameterization of the relationship between physiology and measurement probability [19, 25].

Imputation and uncertainty

Table 2 consolidates the functional role of each architectural component in connecting MNAR bias reduction, probabilistic uncertainty modeling, and clinically meaningful ICU prediction tasks.

Table 2. Functional Roles of the Proposed Architecture in Reducing MNAR Bias and Propagating Imputation Uncertainty

Architectural component	Primary function	MNAR-specific contribution	Uncertainty contribution	Clinical relevance
Irregular time-series input layer	Represents observed values, time gaps, and binary masks jointly	Preserves the structure of non-random observation patterns	Retains uncertainty introduced by sparse and uneven sampling	Reflects real ICU documentation and monitoring conditions
VRNN latent state module	Learns probabilistic physiological trajectories over time	Represents unobserved health states that may influence measurement decisions	Captures latent-state uncertainty, especially during long gaps	Supports reconstruction of evolving patient physiology
Stochastic mask-aware attention	Selectively weights historical observations and masks	Learns which observations are informative given their missingness context	Samples alternative attention configurations	Avoids treating all observed values as equally reliable evidence
Missing pattern encoder	Models the probability that each variable is observed	Explicitly parameterizes the selection mechanism underlying MNAR	Quantifies uncertainty about the observation process	Links clinical monitoring intensity to latent illness severity
Imputation decoder	Generates predictive distributions for missing values	Conditions imputations on physiology, attention context, and missingness parameters	Produces multiple imputations and predictive intervals	Enables uncertainty-aware downstream prediction
Evaluation layer	Tests accuracy, bias, calibration, and downstream utility	Evaluates performance under controlled MNAR scenarios	Assesses interval coverage and uncertainty propagation	Determines whether imputation improves mortality, sepsis, AKI, and deterioration prediction

Predictive distribution

The framework outputs a full predictive distribution for each missing value rather than a point estimate, capturing plausible values consistent with observed data and learned dynamics [20]. The decoder uses latent states, attention-weighted history, and missingness parameters to produce Gaussian distributions. Multiple imputations are generated via repeated sampling, enabling uncertainty-aware inference using Rubin's rules. Credible intervals provide interpretable uncertainty bounds for decision-making [24, 26].

Uncertainty quantification

Three uncertainty sources are modeled: aleatoric (noise), epistemic (model limitations), and imputation uncertainty under MNAR [27, 28]. These are reflected in different variance components of the predictive distribution. Downstream models can weight uncertain imputations פחות heavily, reducing error propagation. Calibration is assessed via coverage of predictive intervals on masked observed data [29].

Evaluation Strategy

Imputation metrics

Evaluation includes accuracy (RMSE, MAE), bias, and uncertainty calibration under MNAR conditions [1]. Coverage of predictive intervals assesses reliability, while bias is critical due to clinical implications [2, 5]. Metrics should be stratified by variable type and missingness patterns.

Downstream task evaluation

Imputation quality is ultimately judged by performance in clinical tasks (e.g., mortality, sepsis, AKI prediction) [3]. Models trained on imputed data are compared against standard methods using AUROC and calibration. MNAR-aware methods are expected to perform better for variables linked to illness severity [7, 13]. Robustness of uncertainty for decision-making should also be evaluated.

MNAR benchmarking

Simulation studies introduce controlled MNAR missingness to compare against ground truth [9, 11]. Performance is assessed across varying MNAR strengths and compared to MAR/MCAR methods. Benchmarks on datasets like MIMIC-IV and eICU ensure reproducibility [17, 22].

Limitations

Technical limitations

The framework is computationally intensive due to variational inference and stochastic sampling [12, 23]. MNAR identifiability remains a fundamental issue, as assumptions about missingness cannot be fully verified. Optimization challenges from stochastic components may affect stability [25, 28].

Clinical limitations

Imputed values risk being treated as real data, despite uncertainty [14]. Calibration may degrade under distribution shifts, requiring ongoing validation. Misspecification of the MNAR mechanism can introduce bias, and sensitivity analyses are essential. The framework does not address variables never measured for some patients [18, 24].

Conclusion

This framework addresses MNAR imputation in ICU time series using VRNNs with stochastic attention, jointly modeling data dynamics, missingness, and imputation. It reduces bias by treating missingness as informative and provides full predictive distributions for uncertainty-aware inference.

Key strengths include modeling informative missingness, uncertainty quantification, and a flexible architecture. Limitations include computational cost, identifiability challenges, and reliance on unverifiable assumptions.

Future work involves validation on datasets such as MIMIC-IV and eICU, benchmarking against standard methods, and developing open-source tools to support reproducible, uncertainty-aware clinical analysis.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 25 May 2025 Revised: 09 Jul 2025 Accepted: 19 Jul 2025
Published online: 20 January 2026

References

Che Z, Purushotham S, Cho K, Sontag D, Liu Y. Recurrent neural networks for multivariate time series with missing values. Sci Rep. 2018;8(1):6085.

Cao W, Wang D, Li J, Zhou H, Li L, Li Y. BRITS: bidirectional recurrent imputation for time series. Adv Neural Inf Process Syst. 2018;31:6776-86.

Yoon J, Jordon J, van der Schaar M. GAIN: missing data imputation using generative adversarial nets. In: Proceedings of the 35th International Conference on Machine Learning (ICML); 2018; Stockholm, Sweden. PMLR; 2018. p. 5689-98.

Shukla SN, Marlin BM. Interpolation-prediction networks for irregularly sampled time series. arXiv [Preprint]. 2019;arXiv:1909.07782.

Rubanova Y, Chen RTQ, Duvenaud DK. Latent ordinary differential equations for irregularly-sampled time series. Adv Neural Inf Process Syst. 2019;32:5320-30.

De Brouwer E, Simm J, Arany A, Moreau Y. GRU-ODE-Bayes: continuous modeling of sporadically-observed time series. Adv Neural Inf Process Syst. 2019;32:7379-90.

Fortuin V, Baranchuk D, Rätsch G, Mandt S. GP-VAE: deep probabilistic time series imputation. In: Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics (AISTATS); 2020; Palermo, Italy. PMLR; 2020. p. 1651-61.

Shukla SN, Marlin BM. Multi-time attention networks for irregularly sampled time series. arXiv [Preprint]. 2021;arXiv:2101.10318.

Tashiro Y, Song J, Song Y, Ermon S. CSDI: conditional score-based diffusion models for probabilistic time series imputation. Adv Neural Inf Process Syst. 2021;34:24804-16.

Zhang X, Zeman M, Tsiligkaridis T, Zitnik M. Graph-guided network for irregularly sampled multivariate time series. arXiv [Preprint]. 2021;arXiv:2110.05357.

Du W, Côté D, Liu Y. SAITS: self-attention-based imputation for time series. Expert Syst Appl. 2023;219:119619.

Biloš M, Sommer J, Rangapuram SS, Januschowski T, Günnemann S. Neural flows: efficient alternative to neural ODEs. Adv Neural Inf Process Syst. 2021;34:21325-37.

Zargar S. Introduction to sequence learning models: RNN, LSTM, GRU. Dep Mech Aerosp Eng N C State Univ. 2021;37988518.

Agor J, Özaltın OY, Ivy JS, Capan M, Arnold R, Romero S. The value of missing information in severity of illness score development. J Biomed Inform. 2019;97:103255.

Sharafoddini A, Dubin JA, Maslove DM, Lee J. A new insight into missing data in intensive care unit patient profiles: observational study. JMIR Med Inform. 2019;7(1):e11605.

Brinton DL, Ford DW, Martin RH, Simpson KN, Goodwin AJ, Simpson AN. Missing data methods for intensive care unit SOFA scores in electronic health records studies: results from a Monte Carlo simulation. J Comp Eff Res. 2021;11(1):47-56.

Sisk R, Sperrin M, Peek N, van Smeden M, Martin GP. Imputation and missing indicators for handling missing data in the development and deployment of clinical prediction models: a simulation study. Stat Methods Med Res. 2023;32(8):1461-77.

Morid MA, Sheng ORL, Dunbar J. Time series prediction using deep learning methods in healthcare. ACM Trans Manag Inf Syst. 2023;14(1):1-29.

Luo Y, Szolovits P, Dighe AS, Baron JM. 3D-MICE: integration of cross-sectional and longitudinal imputation for multi-analyte longitudinal clinical data. J Am Med Inform Assoc. 2018;25(6):645-53.

Beaulieu-Jones BK, Lavage DR, Snyder JW, Moore JH, Pendergrass SA, Bauer CR. Characterizing and managing missing structured data in electronic health records: data analysis. JMIR Med Inform. 2018;6(1):e11.

Johnson AEW, Bulgarelli L, Shen L, Gayles A, Shammout A, et al. MIMIC-IV, a freely accessible electronic health record dataset. Sci Data. 2023;10(1):1.

Pollard TJ, Johnson AEW, Raffa JD, Celi LA, Mark RG, Badawi O. The eICU Collaborative Research Database, a freely available multi-center database for critical care research. Sci Data. 2018;5(1):180178.

Hyland SL, Faltys M, Hüser M, Lyu X, Gumbsch T, et al. Early prediction of circulatory failure in the intensive care unit using machine learning. Nat Med. 2020;26(3):364-73.

Harutyunyan H, Khachatrian H, Kale DC, Ver Steeg G, Galstyan A. Multitask learning and benchmarking with clinical time series data. Sci Data. 2019;6(1):96.

Sheikhalishahi S, Balaraman V, Osmani V. Benchmarking machine learning models on multi-centre eICU critical care dataset. PLoS One. 2020;15(7):e0235424.

Kowsar I, Rabbani SB, Samad MD. Attention-based imputation of missing values in electronic health records tabular data. In: 2024 IEEE 12th International Conference on Healthcare Informatics (ICHI); 2024; Orlando, FL, USA. IEEE; 2024. p. 177-82.

Kazdaghli S, Kerenidis I, Kieckbusch J, Teare P. Improved clinical data imputation via classical and quantum determinantal point processes. eLife. 2024;12:RP89947.

Digitale J, Franzon D, Pletcher MJ, McCulloch CE, Gennatas ED. Methods for addressing missingness in electronic health record data for clinical prediction models: comparative evaluation. JMIR Med Inform. 2025;13(1):e79307.

Qian L, Ellis HL, Wang T, Wang J, Mitra R, Dobson R, et al. How deep is your guess? a fresh perspective on deep learning for medical time-series imputation. IEEE J Biomed Health Inform. 2025.