EasyChair Preprint

№ 15468

Enhancing Clinical Time Series Forecasting with Latent Class Integration in Variational Recurrent Neural Networks

Ruiying Chen and Qingang Tang

November 25, 2024

Enhancing clinical time series forecasting with latent class integration in variational recurrent neural networks

1st Ruiying Chen
Maths Department
University College London
London, United Kingdom
zcahrc3@ucl.ac.uk

2nd Qingang Tang
Statistics Department
University of Oxford
Oxford, United Kingdom
corp3481@ox.ac.uk

Abstract*—Forecasting clinical time series data plays a vital role in healthcare by facilitating early interventions and enhancing patient outcomes. Traditional approaches such as linear and logistic regression, and Recurrent Neural Networks (RNN) have been extensively explored. In addition, generative models like Variational Autoencoders (VAE) have been utilized to handle uncertainty and variability in time series data. However, these methods face challenges in capturing complex temporal dependencies and disease-specific characteristics. We propose two novel variational recurrent neural networks (VRNN) based methods, incorporating patient similarity (VRNN-I+) and latent disease classification (VRNN-II+), to enhance predictive performance in clinical forecasting. The first approach (VRNN-I+) enhances the VRNN model by incorporating temporal data from similar patients as additional domain knowledge, aiming to improve the model's ability to predict patient outcomes. The second approach (VRNN-II+) transitions from using a standard Recurrent Neural Network (RNN) to a Long Short-Term Memory (LSTM) network, while introducing the disease class as a hidden latent variable within the model to capture complex dependencies in the data. The dataset used in this study is derived from Medical Information Mart for Intensive Care-IV (MIMIC-IV), which provided more comprehensive and up-to-date patient records compared to the widely used MIMIC-III dataset. The preprocessing steps were also described accompanying the dataset. To evaluate the effectiveness of incorporating supplementary information for time series prediction, we used the root mean square error (RMSE) as a metric. We selected three subjects to assess the one-step-ahead prediction accuracy, and our results clearly demonstrate that incorporating additional domain knowledge about the patients significantly improves the accuracy of the VRNN (by 15.4%) for clinical time series forecasting.

***Index Terms*—VRNN, VRNN-I+; VRNN-II+**

I. Introduction

It has been estimated that 20-40% of global health system spending is wasted due to inefficiency [1]. Problems of controlling quality, hospital management, and unnecessary care increase the burden on patients and the government [2]. Thus, the expenditure and quality of care have considerable space to improve [3], while the blackness of advanced technology have been considered as one of the fundamental issues.

By 2015, more than 80% of hospitals had implemented at least a basic Electronic Health Record (EHR) system [4]. The widespread adoption of EHR has led to significant improvements [5][6]. Most EHRs collect both continuous and categorical data from patients during their Intensive Care Unit (ICU) stay, enabling comprehensive analysis and monitoring. The multitude of data available in EHRs make them well-suited for high-dimensional analyses [7], including phenotype classification, risk-of-mortality prediction, length-of-stay prediction, etc.

According to the review paper [8], early prediction models primarily used generalized linear models, along with Bayesian methods, random forests, and regularized regression, often evaluated with the c-statistic for model discrimination. In recent years, research on clinical time series diagnostic code prediction has used phenotyping, feedforward networks, LSTM networks and temporal convolutional networks. In addition to multitask learning, phenotyping was formulated as multi-label classification, using neural networks to capture comorbidities in hidden layers implicitly. Others attempted to jointly solve multiple related clinical tasks, including predicting mortality and length of stay. However, none of this work addressed problem settings where sequential or temporal structure varies across tasks [9]. Previous works mainly involved LSTM for clinical time series prediction, a model used to process sequential data, primarily used in language analysis [10]. Our model is heavily based on SRNN, VRNN and its extension which, to the best of our knowledge, has yet to be used with additional domain specific knowledge in the clinical data.

In this paper, we propose two novel VRNN-based methods and demonstrate their effectiveness in increasing forecasting accuracy through experimental results. Although both methods build on VRNN, they incorporate disease diagnosis as supplementary information in distinct ways. The first method seeks to learn the time series patterns for a specific patient by leveraging time series data from similar patients, effectively learning a group-based time series model. The second method introduces a latent categorical variable to identify the disease class.

The structure of the paper is as follows: we (i) present the basic concepts of SRNN and VRNN; (ii) introduce the VRNN-

I+ and VRNN-II+ models, each incorporating a class variable as a latent factor within the model; (iii) describe the MIMIC-IV dataset and outline the data preprocessing steps; (iv) explain how the similarity variable is integrated into the VRNN-I+ model; and (v) compare VRNN-I with VRNN-I+ and VRNN-II+. For evaluation, we selected three random study subjects and plot the predicted versus ground truth values for one-step-ahead predictions.

Building on the limitations of existing clinical time series forecasting models, this paper aims to enhance predictive forecasting by proposing two ways of implementing supplementary information into VRNN-I models(VRNN-I+, VRNN-II+). These models leverage the strengths of LSTM, VAE architectures, and utilize ICD-9 disease codes for patient grouping, thereby enhancing the prediction results. This research contributes to hospital management by supporting phenotype classification, risk of mortality prediction and length of stay forecasting, as well as reducing documentation error, and providing valuable references for doctors. Additionally, it offer a refined approach for analyzing clinical data for scientists and medical researchers. For instance, computational phenotyping research aids to discovering and categorizing new subtypes and identifying specific phenotypes to improve classification under existing disease boundaries and definitions.

II. Methods

This section first introduces SRNN, which is designed to handle sequential data by incorporating stochastic latent variables into a recurrent structure, thereby modeling uncertainty in time-series data. Then, we present an approach called VRNN-I+, which extends the basic SRNN by utilizing the VAE's probabilistic latent space, enabling for a more expressive representation of the hidden structure in sequential data. Finally, we introduce another approach called VRNN-II+, which integrates a LSTM networks into both the encoder and decoder networks of a VAE model to enhace its sequential modeling capabilities[11][12][13]. This approach adds an additional continuous latent variable w, into the model to represent the probability of being in a class. Note that this differs from the VRNN-I+ model, where the latter inserts extra patients with similar temporal features (decided via KNN approach) to enhance the prediction.

A. Stochastic Recurrent Neural Networks (SRNN)

In brief, SRNN combines RNN with a state space model (SSM)both of which are commonly used in time series modeling. Again, let us assume the time series $x_{1:T} = (x_1, x_2, \cdots, x_T)$ may depend on $u_{1:T} = (u_1, u_2, \cdots, u_T)$ As shown in Fig.1, RNN and SSM have similarities, such as incorporating information from the sequence $x_{1:t---1}$ is integrated into the latent state d_t or z_t. In comparison: RNN has strong nonlinear fitting capabilities, but its hidden states are deterministic; meanwhile, SSM's stochastic state transitions are better suited for modeling uncertainty, though its inference process is usually simpler. Can we combine the strengths of both approaches? This is the motivation of

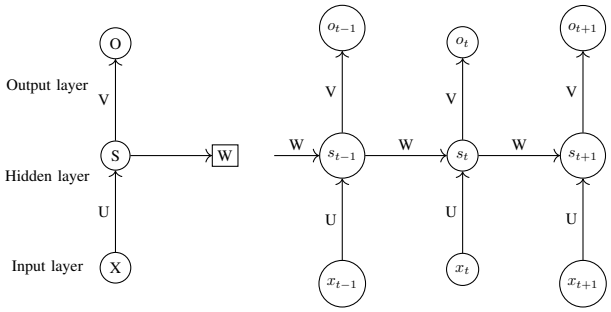

Fig. 1. Recurrent Neural Network (RNN) architecture: Left hand side is the structure for one time step and right hand side is the overview of RNN

SRNN, which turns the stochastic state transitions in SSM into nonlinear ones, while preserving the gating activation mechanism of RNN. For further details on SRNN, please refer to [14].

B. Variational Recurrent Neural Network(VRNN-I)

1) Model Generation: The VRNN-I [15] architecture merges elements of both VAE and RNN, with a VAE integrated at each time step t. However, the VAE accounts the temporal structure present in sequential data by conditioning on the RNN's state variable h_{t-1}. Unlike a typical VAE, the prior for the latent random variable is no longer modeled by a standard Gaussian distribution, but instead follows a different distribution:

$$\mathbf{z}_t \sim \mathcal{N}(\boldsymbol{\mu}_{0,t}, \text{diag}(\boldsymbol{\sigma}_{0,t}^2)) \tag{1}$$

And the parameters are determined using the following equations:

$$[\boldsymbol{\mu}_{0,t}, \boldsymbol{\sigma}_{0,t}] = \boldsymbol{\psi}_\tau^{prior}(\mathbf{h}_{t-1}). \tag{2}$$

Furthermore, the generating distribution will be conditioned not only on $\mathbf{z}_t$ but also on $\mathbf{h}_{t-1}$ such that:

$$\mathbf{x}_t|\mathbf{z}_t \sim \mathcal{N}(\boldsymbol{\mu}_{x,t}, \text{diag}(\boldsymbol{\sigma}_{x,t}^2)) \tag{3}$$

where

$$[\boldsymbol{\mu}_{x,t}, \boldsymbol{\sigma}_{x,t}] = \boldsymbol{\psi}_\tau^{dec}(\boldsymbol{\psi}_\tau^z z_t, h_{t-1}). \tag{4}$$

$\mu_{x,t}$ and $\sigma_{x,t}$ represent the parameters of the generating distribution. Functions $\boldsymbol{\psi}_\tau^{prior}$, $\boldsymbol{\psi}_\tau^{dec}$ can be any highly flexible function such as neural networks. Both $\boldsymbol{\psi}_\tau^x$ and $\boldsymbol{\psi}_\tau^z$ are implemented as neural networks, designed to extract features respectively from x_t and z_t. Our analysis shows that these feature extractors are vital for learning complex sequences. The RNN updates its hidden state using the recurrence equation:

$$\boldsymbol{h}_t = f_\theta(\boldsymbol{\psi}_\tau^x(x_t), \boldsymbol{\psi}_\tau^z(z_t), \boldsymbol{h}_{t-1}) \tag{5}$$

The parameterization of the generative model results in and is motivated by the factorization:

$$p(\boldsymbol{x} \leq T, \boldsymbol{z} \leq T) = \prod_{t=1}^{T} p(\boldsymbol{x}_t|\boldsymbol{z}_{\leq t}, \boldsymbol{x}_{\leq t}) p(\boldsymbol{z}_t|\boldsymbol{x}_{<t}, \boldsymbol{z}_{<t}) \tag{6}$$

2) Model Inference: In a similar manner, the approximate posterior is not only a function of x_t, but also a function of h_{t-1}, as shown below:

$$z_t|x_t \sim \mathcal{N}(\mu_{z,t}, \text{diag}(\sigma^2_{z,t})), \tag{7}$$

where $\mu_{z,t}$ and $\sigma_{z,t}$ represent the parameters of the approximate posterior. It is observed that the encoding and decoding processes of the approximate posterior depend on the hidden state h_{t-1} of the RNN. Additionally, conditioning on h_{t-1} leads to the following factorization:

$$q(z_{\leq T}|x_{\leq T}) = \prod_{t=1}^{T} q(z_t|x_{\leq t}, z_{<t}). \tag{8}$$

3) Learning: The objective function changes to the following time-step lower bound using (6) and (8):

$$\mathbb{E}_{q(z_{\leq T}|x_{\leq T})}[\sum_{t=1}^{T}(-KL\left(q(z_t|x_{\leq t}, z_{<t})||p(z_t|x_{\leq t}, z_{<t})\right) + \log p(x_t|z_{\leq t}, x_{\leq t})]. \tag{9}$$

As in a standard VAE, we jointly learn the generative model and inference model by maximizing the variational lower bound with respect to their parameters.

C. VRNN with latent class (VRNN-II)

First, we describe how class variables are embedded into VRNN-II model. The ICD-9 code is designed to group similar diseases into 1 single class. Unlike the VRNN-I framework, this study groups patients with similar status together, summarizes all ICD-9 disease types into five classes (more details can be found in the Model implementation and evaluation section). Next, we add an additional continuous latent variable w which follows a multinomial distribution with parameter θ representing the probability of data belonging to each category. The joint distribution considered is as follows:

$$p_\theta(X, z, w) = p_\theta(X|z, w)p_\theta(z)p_\theta(w) \tag{10}$$

where X is the observed data, z and w are latent variables. We also assumed that z and w are independent variables.

During training, we assumed both X and w are given while during inference, only X is given. This is due to the VAE's limitation in inferring discrete latent variables. Instead, our approach model them as class probabilities rather than treating them as discrete categorical variables.

We would like to learn the entire posterior of class probabilities: $p_\theta(w|X)$. Following the derivation from [16], which builds on the modification of [17]: we constructed a variational lower-bound on the log of the marginal likelihood

$$p_\theta(X) = \int p_\theta(X|z, w)p_\theta(z)p_\theta(w)\ \partial w \partial z \tag{11}$$

with

$$\begin{aligned}&\log p_\theta(X) - \mathcal{D}[q_\phi(z, w \mid X) \parallel p_\theta(z, w \mid X)] \\ =&\mathbb{E}_{(z,w)\sim q_\phi(z,w|X)}[\log p_\theta(X \mid z, w)] \\ &-\mathcal{D}[q_\phi(z, w \mid X) \parallel p_\theta(z, w)]\end{aligned} \tag{12}$$

Then follow [14], let us rewrite the right-hand side as:

$$\begin{aligned}&\mathbb{E}_{(z,w)\sim q_\phi(z,w|X)}[\log p_\theta(X \mid z, w)] \\ &-\mathbb{E}_{w\sim q_\phi(w|X)}[\mathcal{D}\left[q_\phi(z \mid X, w) \parallel p_\theta(z)\right]] \\ &-\mathcal{D}[q_\phi(w \mid X) \parallel p_\theta(w)] = -\mathcal{L}_{VAE}(X)\end{aligned} \tag{13}$$

where $p_\theta(z)$ and $p_\theta(w)$ are priors, while $q_\phi(w|X)$ and $q_\phi(z|w, X)$ are variational approximations to the true posteriors. Same as [17], we aim to optimize the marginal likelihood, which can be achieved via maximizing (13).

Furthermore, we ensure that $\mathcal{D}[q_\phi(z, w \mid X) \parallel p_\theta(z, w \mid X)]$ is small as possible. We assume that the true class is accessible to m, is denoted as $\tilde{w}$. This setup allows us to minimize the categorical cross-entropy loss between $q_\phi(w|X)$ and $\tilde{w}$ [16]. We end up with the final objective, denoted as ($\mathcal{L}$), is therefore as follows:

$$\mathcal{L}(X) = \mathcal{L}_{VAE}(X) + \alpha\mathbb{E}_{w\sim q_\phi(w|X)}\left[\mathcal{L}_c(w; \tilde{w})\right] \tag{14}$$

$\mathcal{L}_c(w; \tilde{w})$ represents the categorical cross-entropy loss between a sampled $w \sim q_\phi(w|X)$ and the true class $\tilde{w}$.

III. Date and Model

A. Data Description

The "Medical Information Mart for Intensive Care (MIMIC-IV)" [18] is a modern electronic health record dataset that spans a decade of patient admissions from 2008 to 2019. MIMIC-IV offers comprehensive data for retrospective clinical studies and critical care operations, featuring high granularity on aspects such as re-admissions, length of stay, prescriptions, caregivers, and diagnoses and procedures (ICD-9). As shown in Fig.3, MIMIC-IV separates patient information in the Emergency Department (ED) and Intensive Care Unit (ICU). We focus on exclusively on ICU data, supported by with demographic data, as ICU data contains longer time-series inputs and less missing data (further details are provided in the following subsections).

Fig. 2. structure of MIMIC-IV constructed in this paper

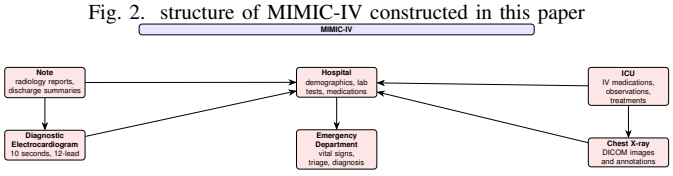

Fig. 3. MIMIC-IV follows a modular structure. Modules are linked by identifiers including subject_id, hadm_id, and deidentified date and time. Example content of each module is shown.

B. Data preprocessing

Gupta et al. [19] contains extensive data processing pipeline for MIMIC-IVyielding optimal result with MIMIC-IV 2.2. Consequently, this paper adhered to this version of MIMIC-IV.

1) Outlier removal: During the preprocessing, we cleaned the original dataset by removing outliers, allowing us to impute missing entries. Following the approach in [19], we identified statistically extreme values, unlike the method in [20], where only values labeled 999999 and the negative (infeasible) values were removed as in [13]. We customized the two-tailed threshold to 95th percentile, which means all extreme values outside this percentile range should be removed.

2) Missing values: Missing data can lead to issues such as distorted patterns, trends, and biases in parameter estimation [21]. We adopt the approach of [22], excluding variables with more than 20% missing data (table I. For variables with less than 5% missing data, missing values in continuous variables that exhibit a normal distribution were imputed using the mean specific to the patient group. For continuous variables with skewed distributions, the median was used as replace missing values [23]. For variables with more than 5% missing data, we implemented a method adapted from [9], a benchmark from MIMIC-III. This approach imputted missing values using the most recent measurement when available, or a predefined 'normal' value otherwise (for further details, see [9]).

TABLE I
PERCENTAGE OF MISSING VALUES FOR VARIOUS VARIABLES

Variables	Percentage of missing values
Heart rate	4.23%
Respiratory rate	4.89%
MBP	4.39%
SBP	7.81%
DBP	7.81%
AOS	10.93%
Glucose	12.68%
Temperature	19.54%

3) Time-series representation: The following continuous data were extracted from MIMIC-IV ICU: heart rate (HR), systolic blood pressure (SBP), diastolic blood pressure (DBP), mean blood pressure (MBP), respiratory rate (RR), body temperature, Arterial O2 Saturation (AOS) and Glucose. Next, we applied the pipeline from [19], these variables were organized into uniform 4-hour intervals for consistency. Since the time-series data selected is typically recorded every 4 hours, we opted for a 4-hour interval. If a patient has multiple entries within a 4-hour period, the mean value was computed and replaced by those entries with a single value for consistency. Conversely, if there are no measurements within a 4-hour slot, we addressed this as a missing value issue and resolved it using the method outlined in [9]. Since the average stay in the ICU is 11 days (with a standard deviation of 13), we limited the selection to patients who stayed in the ICU for more than 10 days. For each binned time interval, the pipeline assigned a feature value of 1 if it is recorded during that interval, and 0 otherwise. This paper ended up with a matrix of size 60 × 8 for each stay and 51080 stay records. Additionally, since the data were extracted from department where all subjects have relatively severe health issues, facing a survivorship bias problem [24]. This means the inference model may show higher accuracy for severe patients.

C. Computation of similarity variable X_t^{rel}(VRNN-I+)

MIMIC-IV contains a lot of supplementary information such as age, gender, diseases and medication etc. We extracted age, disease diagnosis (ICD-9), and medication as extra domain information. With apporximately 13000 diagnoses in ICD-9, which is sufficient to provide domain specific knowledge to the model, we represented the patient's disease as a vector of size 13,000, where a "1" indicates a confirmed diagnosis for the patient, and "0" indicates otherwise. In addition, we included their temporal data as previously described to facilitate clustering patients who exhibit remarkably similar behaviors. Using a K-nearest neighbor approach, we grouped similar patients with similar diseases based cosine similarity. This study tests the values of k for 3,4,5,6 and found that 4 is the best option. Thus, all the results mentioned in the next section are computed with $k = 4$ and $x_t^{rel} \in \mathbb{R}^{32}$.

D. Model Implementation and Evaluation

The models were implemented with T4 GRU, with all temporal features re-scaled in value between -1 and 1. The training took approximately 3 hours to complete with training to test split of 80/20. Table II shows the implementation details of VRNN-I and VRNN-II model, respectively, where X and Z denoted the dimension of x_t and z_t respectively. This table also includes details on the number of hidden layers, batch size and epoch. The batch size selection was guided by prior research [15]. For the choice of the number of hidden layers and their size, we experimented with many combinations, including setting from previous work[15]. The final choice of the hidden size, number of layers and number of epochs were based on test set performance. However, the optimal value produced the best result in the test set was different from any of the settings in [17][14][15].

TABLE II
IMPLEMENTATION DETAILS OF VRNN-I AND VRNN-II

Model	X	Z	No. Layers	Hidden	Batch	Epoch
VRNN-I	8	2	2	60	100	5
VRNN-II	8	4	2	80	150	5

We focus on assessing the performance of the models for predictions that are less than 8 hours into the future. Accordingly, we trained on the first 232 hours (58 steps) and predicted the last 8 hours (2 steps). The metric used in this paper is Root Mean Square Error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{15}$$

where n denoted the size of test dataset.
For VRNN-II+ model, we grouped different diseases into 5 classes (Heart, Blood Pressure, Kidney, Respiratory, and Others). This provided the model with priori information on potential patient pattern. In the following section, we review the outcomes generated from the experimental setup described earlier.

IV. Results

As shown in Fig.4, we chosed 3 patients randomly and analyzed the average cosine similarity in respect to different k. It is clear that the choice of $k = 4$, as discussed in previous section, appears plausible since all three cases demonstrates high similarity values for the first few related patients only. Furthermore, from table III, the Root Mean Square Error (RMSE) with standard deviation on the test dataset for 1- and 2- step ahead forecasting using VRNN-I model. The result clearly shows that X_t^{rel} (additional domain

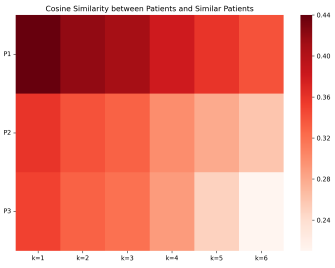

Fig. 4. This heat map visualises the average cosine similarity values between our patients of interest (P1,P2, and P3) and its similar patients in k clusters.

knowledge) improves the forecasting accuracy of the VRNN-I for clinical signals (we observed that VRNN-I+ achieves the lower RMSE in comparison with VRNN-I). This improvement is especially notable for temporal features associated with the set of common diseases between the patients.

TABLE III
RMSE WITH ROUNDED STANDARD DEVIATIONS ON FIRST 2 STEPS AHEAD FORECASTING TASKS ON TEST DATA

Step Size	**VRNN-I**	**VRNN with disease diagnosis (VRNN-I+)**	**VRNN-II with disease diagnosis (VRNN-II+)**
1	0.01220 ± 0.0039	0.01032 ± 0.0024	0.01030 ± 0.0029
2	0.01212 ± 0.0035	0.01034 ± 0.0021	0.01034 ± 0.0020

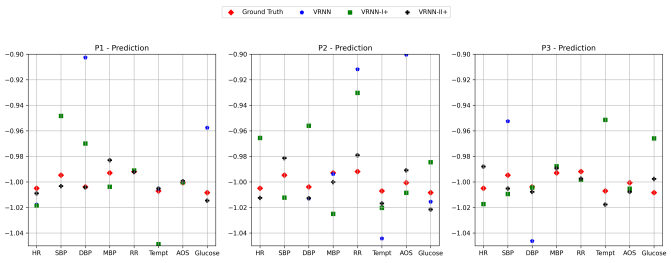

Fig. 5. one step predicted values for all 8 temporal variables

Furthermore, comparing with the result of VRNN-II+, their RMSE values are very similar. Although VRNN-II [14] theoretically should outperform VRNN-I model. There is a minor difference in the result. The conclusion is that the integration method of domain knowledge differs across models, which leads to limited further improvement (we categorized the domain knowledge into 6 categories for VRNN-II+ model which is less detailed compared to the clusters for VRNN-I+ model).

Plotting of one-step-ahead predictions of all three models for patients of interest is in Fig.5. The ground truth values for all temporal features are shown in red. VRNN-I+ and VRNN-II+ outperform VRNN-I on MBP, SBP, glucose, and AOS for all subjects. It is evident that VRNN-II+ provides extremely accurate predictions, which might seem contradictory to the results in Table 5, where VRNN-II+ is not shown to be better than VRNN-I+. However, this is because the randomly selected three patients coincidentally showed more accurate predictions under VRNN-II+. Therefore, the process of adding extra disease diagnoses improve the forecasting accuracy of the VRNN-I model. On average, VRNN-II+ is slightly better than the VRNN-I+ model.

V. Conclusion

This study presents two innovative approaches using Variational Recurrent Neural Networks (VRNN-I+ and VRNN-II+) to enhance the predictive accuracy of clinical time series data derived from the MIMIC-IV dataset. The proposed methods, which incorporate domain-specific knowledge such as patient similarity through the K-nearest neighbor approach and disease classification using ICD-9 codes, successfully address the limitations of existing models like RNNs and stochastic recurrent neural networks in capturing complex temporal dependencies and disease-specific patterns.

Our findings show that the inclusion of additional domain knowledge, particularly in the VRNN-I+ model, significantly improved predictive performance, as demonstrated by lower RMSE values compared to standard VRNNs. Furthermore, the comparison between VRNN-I+ and VRNN-II+ reveals that the latter provides marginal improvements in prediction accuracy, especially in capturing more granular disease-specific information. However, the difference in performance between the two models suggests that the manner of incorporating domain knowledge plays a critical role in determining model effectiveness. The implications of this research are far-reaching in clinical settings. By improving forecasting models for patient outcomes, these approaches offer practical benefits in hospital management, such as better phenotype classification, risk-of-mortality prediction, and length-of-stay predictions. The models also provide a framework for more detailed analysis of clinical data, which could enhance decision-making processes for healthcare professionals and offer new avenues for medical research.

Overall, the integration of temporal patterns with disease-specific latent variables in VRNN models opens new possibilities for improving predictive models in healthcare, ultimately leading to more accurate and actionable insights for patient care. Future work could explore further refinement in domain knowledge integration and the exploration of additional latent variables to enhance model accuracy across broader datasets and clinical applications.

References

[1] World Health Organization. Health systems financing: the path to universal coverage. 2010.

[2] Yip W, Hafez R, Organization WH. Improving health system efficiency: reforms for improving the efficiency of health systems: lessons from 10 country cases. World Health Organization. 2015.

[3] Mutter RL, Rosko MD, Wong HS. Measuring hospital inefficiency: the effects of controlling for quality and patient burden of illness. Health Serv Res. 2008 Dec;43(6):1992-2013. doi: 10.1111/j.1475-6773.2008. 00892. x.

[4] Shickel B, Tighe PJ, Bihorac A, Rashidi P. Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis. IEEE J Biomed Health Inform. 2018 Sep;22(5):1589-1604. doi: 10.1109/JBHI. 2017.2767063.

[5] Choi E, Schuetz A, Stewart WF, Sun J. Medical Concept Representation Learning from Electronic Health Records and its Application on Heart Failure Prediction. arXiv. 2016 Feb; 45. [Online]. Available: http://arxiv.org/abs/1602.03686.

[6] Miotto R, Li L, Kidd BA, Dudley JT. Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records. Scientific reports. 2016;6(no. April):26094.

[7] Maddox TM, Rumsfeld JS, Payne PRO.. Questions for artificial intelligence in health care. JAMA 2019; 321 (1): 31–2

[8] Benjamin A Goldstein, Ann Marie Navar, Michael J Pencina, John P A Ioannidis, Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review, Journal of the American Medical Informatics Association, Volume 24, Issue 1, January 2017, Pages 198-208, https://doi-org.libproxy.ucl.ac.uk/10.1093/jamia/ocw042

[9] H. Harutyunyan, H. Khachatrian, D. C. Kale, G. V. Steeg, and A. Galstyan, "Multitask learning and benchmarking with clinical time series data," arXiv preprint arXiv:1703.07771, 2017.

[10] Singh, Rachit. Sequential discrete latent variables for language modeling. Diss. 2018.

[11] Justin Bayer and Christian Osendorfer. Learning stochastic recurrent networks. arXiv preprint arXiv:1411.7610, 2014.

[12] Junyoung Chung, Kyle Kastner, Laurent Dinh, Kratarth Goel, Aaron C. Courville, and Yoshua Bengio. A recurrent latent variable model for sequential data. arXiv, 1506.02216, 2015.

[13] Marco Fraccaro, Søren Kaae Sønderby, Ulrich Paquet, and Ole Winther. Sequential neural models with stochastic layers. In Advances in Neural Information Processing Systems, pages 2199-2207, 2016.

[14] Fraccaro, Marco, et al. "Sequential neural models with stochastic layers." Advances in neural information processing systems 29 (2016).

[15] J. Chung, K. Kastner, L. Dinh, K. Goel, A. C. Courville, and Y. Bengio, "A recurrent latent variable model for sequential data," in Advances in neural information processing systems, pp. 2980-2988, 2015.

[16] Hennig, Jay A., Akash Umakantha, and Ryan C. Williamson. A classifying variational autoencoder with application to polyphonic music generation. arXiv preprint arXiv: 1711.07050 (2017).

[17] D. P. Kingma and M. Welling, Auto-encoding variational bayes, arXiv preprint arXiv:1312.6114, 2013.

[18] Johnson, Alistair, et al. "Mimic-iv." PhysioNet. Available online at: https://physionet. org/content/mimiciv/1.0/ (accessed August 23, 2021) (2020): 49-55.

[19] Gupta, M., Gallamoza, B., Cutrona, N., Dhakal, P., Poulain, R., & Beheshti, R. (2022, November). An extensive data processing pipeline for mimic-iv. In Machine Learning for Health (pp. 311-325). PMLR.

[20] Aishwarya Mandyam, Elizabeth C Yoo, Jeff Soules, Krzysztof Laudanski, and Barbara E Engelhardt. COP-E-CAT: cleaning and organization pipeline for ehr computational and analytic tasks. In Proceedings of the 12th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics, pages 1-9, 2021.

[21] I. Pratama, A. E. Permanasari, I. Ardiyanto and R. Indrayani, A review of missing values handling methods on time-series data, 2016 International Conference on Information Technology Systems and Innovation (ICITSI), Bandung, Indonesia, 2016, pp. 1-6, doi: 10.1109/ICITSI.2016.7858189.

[22] Sun, Yiwu & He, Zhaoyi & Ren, Jie & Wu, Yifan. Prediction model of in-hospital mortality in intensive care unit patients with cardiac arrest: a retrospective analysis of MIMIC - database based on machine learning. 2023. 10.21203/rs.3.rs2551943/v1.

[23] Zhang Z. Missing data imputation: focusing on single imputation. Ann Transl Med. 2016 Jan;4(1):9. doi: 10.3978/j.issn.23055839. 2015.12.38. PMID: 26855945; PMCID: PMC4716933.

[24] Elston DM. Survivorship bias. J Am Acad Dermatol. 2021 Jun 18:S0190-9622(21)01986-1. doi: 10.1016/j.jaad.2021.06.845. Epub ahead of print. PMID: 34153385.