Integrating Static and Time-Series Data in Deep Recurrent Models for Oncology Early Warning Systems

Dingwen Li
dingwenli@wustl.edu
Washington University in St. Louis
McKelvey School of Engineering
St. Louis, Missouri, USA

Patrick Lyons
plyons@wustl.edu
Washington University in St. Louis
School of Medicine
St. Louis, Missouri, USA

Jeff Klaus
jeff.klaus@bjc.org
Barnes-Jewish Hospital
St. Louis, Missouri, USA

Brian Gage
bgage@wustl.edu
Washington University in St. Louis
School of Medicine
St. Louis, Missouri, USA

Marin Kollef
kollefm@wustl.edu
Washington University in St. Louis
School of Medicine
St. Louis, Missouri, USA

Chenyang Lu
lu@wustl.edu
Washington University in St. Louis
McKelvey School of Engineering
St. Louis, Missouri, USA

ABSTRACT

Machine learning techniques have shown promise in predicting clinical deterioration of hospitalized patients based on electronic health record (EHR). However, building accurate early warning systems (EWS) remains challenging in practice. EHRs are heterogeneous, comprising both static and time-series data. Moreover, missing values are prevalent in both static and time-series data, and the missingness of certain data can be correlated to clinical outcomes. This paper proposes a novel approach for integrating static and time-series clinical data in deep recurrent models through multimodal fusion. Furthermore, we exploit the correlation of static and time-series data through cross-modal imputation in an integrated recurrent model. We apply the proposed approaches to a dataset extracted from the EHR of 20,700 hospitalizations of adult oncology patients in a research hospital. The experiments demonstrate the proposed approaches outperform the state-of-the-art models in terms of predictive accuracy in generating early warnings for clinical deterioration. A case study further establishes the efficacy of the predictive model for early warning systems under realistic clinical settings.

CCS CONCEPTS

• **Computing methodologies** → *Neural networks*; • **Applied computing** → **Health informatics**.

KEYWORDS

data mining, healthcare, recurrent neural networks, imputation

ACM Reference Format:
Dingwen Li, Patrick Lyons, Jeff Klaus, Brian Gage, Marin Kollef, and Chenyang Lu. 2021. Integrating Static and Time-Series Data in Deep Recurrent Models for Oncology Early Warning Systems. In *Proceedings of the 30th ACM International Conference on Information and Knowledge Management (CIKM '21), November 1–5, 2021, Virtual Event, QLD, Australia.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3459637.3482441

CIKM '21, November 1–5, 2021, Virtual Event, QLD, Australia

ACM ISBN 978-1-4503-8446-9/21/11.
https://doi.org/10.1145/3459637.3482441

1 INTRODUCTION

Clinical deterioration of hospitalized patients has been a significant challenge in healthcare. Cancer patients are at particularly high risk for clinical deterioration. There were over 18 million cancer cases around the world in 2018, and 9.6 million people died because of cancer in the same year [14]. Most cancer cases cause hospitalizations due to their devastating effects on the patients' health. Oncology inpatients are at risk of life-threatening clinical deterioration: 6.4% of oncology inpatients have at least one ICU transfer, and 2.7% of them die on the hospital wards according to a recent study [26]. Clinical deterioration is often presaged by abnormal vital signs, lab values or test results, which leads to the development of early warning systems (EWS) [2] based on electronic health record (EHR). With the help of EWS, clinicians can identify decompensating patients in advance and then provide timely interventions to save their lives.

However, it remains challenging to accurately predict deterioration events in EWS. To improve predictive performance a machine learning model needs to exploit heterogeneous data in EHR comprising both static and time-series variables. Static variables are usually collected at the time of hospital admission. Examples include demographics and comorbidity diagnoses (ICD codes from previous hospital admissions). Time-series variables are collected repeatedly during the patient's hospital stay. Examples include vital signs, lab values, cultures, medications, and procedures. As static and time-series variables often make complementary contributions to predicting clinical events, a machine learning model must effectively incorporate both types of variables to maximize its accuracy.

Traditional clinical machine learning models rely on feature engineering techniques to extract statistical features from time-series data in order to integrate them with static variables in a machine learning model [16, 19, 23, 33]. Recent advances in recurrent neural networks (RNN) provide a powerful approach to model time-series data by exploiting the temporal information without the need for feature engineering. However, original RNN models were designed for time-series inputs only, which cannot exploit the static data available in EHR that can be highly informative [4, 5, 37]. Existing

models for integrating heterogeneous inputs are to concatenate the learned representations from each input [6, 9, 12, 13, 39] or to feed time-series inputs together with the learned representation of static inputs into RNN at each time step [20, 34]. However, as shown in our experiments on a real-world dataset of oncology inpatients, these approaches have limitations in exploiting heterogeneous clinical data for EWS.

In this paper, we explore novel RNN-based approaches that incorporate both static and time-series data, thereby combining the power of RNN and the benefits of heterogeneous data in EHR. To effectively integrate heterogeneous clinical inputs, we first propose a *multi-modal fusion* approach that uses the representation learned from static variables as the initial hidden state of RNN. To exploit the correlations between static and time-series variables, we design an end-to-end model called *CrossNet* that integrates multi-modal bidirectional LSTM and *cross-modal imputation*. CrossNet is able to learn the classification objective, i.e., predicting deterioration events, while accurately imputing the missing values in the input data across static and time-series variables.

We apply the proposed approaches to a large dataset of adult oncology patients hospitalized in Barnes-Jewish Hospital. The objective is to generate early warning alerts that can predict a patient's deterioration as a composite of ICU transfer or ward death. In the evaluation, we show the multi-modal fusion can be integrated with state-of-the-art recurrent models used for clinical prediction, e.g., m-RNN [42], GRU-D [5], BRITS [4] and LGnet [37] and yields consistent performance gains over two common representation fusion approaches. Moreover, we demonstrate the effectiveness of CrossNet to further boost the predictive performance through cross-modal imputation.

The main contributions of our work are four-fold: (1) we present a multi-modal fusion technique that can integrate static and time-series clinical variables in deep recurrent models; (2) we propose the CrossNet model that combines multi-modal fusion and cross-modal imputation to exploit the correlations across static and time-series data for clinical early warning; (3) we demonstrate the efficacy of the proposed approaches on a large EHR dataset including 20,700 hospitalizations of adult oncology patients while also showing the generality of our approach on the MIMIC-III dataset [17]; (4) we provide a case study that establishes the advantages of our CrossNet model for EWS under realistic clinical settings on oncology hospital wards.

2 RELATED WORK

2.1 Clinical Models for Time Series

There has been significant work on clinical predictive models focusing on time-series data. Mao et al. employ feature engineering of time-series data and logistic regression to predict clinical deterioration in the ICU [30]. ForecastICU predicts ICU admission based on physiological data streams using a Bayesian belief system [40]. Lipton et al. [21] empirically evaluate LSTM models to predict diagnostic outcomes based on multi-variate time series in EHR. Liu et al. [22] propose a personalized predictive framework for multivariate clinical time series to support the prediction of patients' future physiological signals. UA-CRNN discovers irregularity of clinical time series and incorporates uncertainty in the generated regular data for predicting mortality risks [36]. Recent studies, e.g., GRAM [10], RETAIN [7], Dipole [27], MIME [11], KAME [28], DoctorAI [8] and HiTANet [24], propose attention mechanisms that can be trained jointly with RNN to predict clinical outcomes based on sequential medical codes from hospital visits. Other RNN approaches, e.g., m-RNN [40], GRU-D [5], BRITS [4] and LGnet [37], propose integrated models for data imputation and outcome prediction based on the clinical time series.

While previous works have shown promise in exploiting time-series data to predict clinical outcomes, none of the aforementioned models integrate static variables and time-series variables. As static variables (e.g., demographics and comorbidities) are commonly available in EHR and can be strongly correlated to clinical outcomes, ignoring static variables can negatively impact the predictive performance.

2.2 Integrating Static and Time-Series Data

Traditional clinical predictive models [16, 19, 23, 33] integrate static and time-series data by extracting statistical features from the time-series data (thereby removing the temporal dimension), so that the extracted features can be directly combined with static features. However, feature engineering can be labor-intensive, and the calculation of various statistical features, especially second-order statistical features, inevitably introduces large computational complexity in the preprocessing stage. Moreover, feature engineering methods are domain-specific, where prior knowledge is used to guide the feature engineering design.

RNN models provide an effective approach to exploit time-series data without the need for feature engineering. There exist two approaches to integrate static variables and time-series variables in RNN models. Most of prior studies [6, 9, 12, 13, 15, 39] *concatenate* the representations learned separately from static and time-series variables. However, concatenating the learned representations of static and time-series variables at later stages does not take into account the fact that the two modalities are usually correlated. Recent clinical models [20, 34] employ a *static-repeat* approach that feed time-series inputs together with the learned representation of static inputs into RNN at each time step. This approach has the advantage of exploiting the correlation among static variables and time-series variables to guide the RNN training. However, this approach may make the RNN parameters fit to the noise, which yields sub-optimal results [18, 38]. In our work, we explore multi-modal fusion as a new approach to integrate heterogeneous clinical inputs, which yields better performance than concatenating the learned representations and the static-repeat approach. Our approach is inspired by computer vision methods [18, 29, 38] designed to generate image captions based on image as the static input and descriptive sentences as the sequential input. We propose to adapt the multi-modal fusion approach for integrating the static and time-series clinical variables in RNN models to predict clinical outcomes. In our clinical predictive models, we use the output of the model trained on the static variables as the initial input to the recurrent layer for time-series variables. This integrated approach effectively leverages knowledge learned from static variables (available upon hospital admission) to provide the context for training RNN model for time-series data.

2.3 Handling Missing Clinical Data

Missing values introduce both challenges and opportunities in clinical predictive models. Missing values are prevalent in EHR data. The absence of a variable at certain time points and the missing patterns can be informative for clinical predictions [4, 5]. Traditional imputation methods, such as K nearest neighbor (KNN) imputation [3] and Multiple Imputation by Chained Equations (MICE) [1], and more recent generative adversarial network based imputation evaluated on clinical datasets, e.g., GAIN [41] and E^{2}GAN [25], impute missing values to minimize imputation errors. Then a separate machine learning model can be trained on the imputed data to perform the outcome prediction. However, this two-step model training is time consuming and overlook the correlations between missingness of clinical data and predictive outcomes. Recently, RNN-based imputation methods have shown advantages over traditional approaches in terms of predictive performance [4, 40]. Multi-directional Recurrent Neural Networks (m-RNN) [40] employs a dedicated RNN model to impute the missing values in clinical data. However, m-RNN makes the assumption that missing values occur at random and do not correlate with clinical outcomes. GRU-D [5] and BRITS [4] provide integrated RNN models that exploit the correlation between missing values in EHR and clinical outcomes. LGnet [37] adopts similar architecture as GRU-D, but introduces memory units to capture global temporal patterns. The imputation components of these RNN-based methods are integrated into the RNN cell computation, so that the model parameters can be jointly optimized for both predictive and imputation objectives. However, these RNN models are designed for time-series data only and ignore static data. In our work, we incorporate the imputation component of static variables into the recurrent model and explicitly exploit the correlations between static variables and time-series variables to improve predictive performance.

3 METHODOLOGY

3.1 Early Warning for Hospitalized Patients

In hospitals, a variety of data are collected for an inpatient, including static and time-series data. The static data, such as demographics, co-morbidity diagnoses (ICD-9 and ICD-10 codes from previous hospital admissions), and hospitalized location, are collected at the time of hospital admission. The time-series data, such as vital signs and lab values, are collected at multiple time instants during the entire hospital stay. As the time-series data are recorded at different frequencies, the overall data space is sparse, which means that there inevitably exist missing values for those less frequently measured variables. We formalize early warning as a real-time binary classification problem. The inputs of the model are the data obtained in a fixed-size window of X hours. The output of the model is a binary label indicating the onset of a deterioration event Y hours, where "Y hours" is the pre-defined prediction horizon. In our paper, we choose a 72-hour window and a 6-hour prediction horizon to facilitate timely clinical intervention for oncology inpatients. In the following, we present the multi-modal fusion integrating the static and time-series variables and a novel model with cross-modal imputation that can handle the missing values in both static and time-series variables.

3.2 Multi-modal Fusion for Heterogeneous Medical Data

Supposing we have a dataset of N samples, each of which has a D_s-dimensional feature vector $\mathbf{s} \in \mathbb{R}^{D_s}$ derived from the raw static data and a $T \times D_t$ feature matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_T]^{\mathrm{T}}$ derived from the raw time-series data. $\mathbf{x}_t \in \mathbb{R}^{D_t}$ indicates observations at the t-th timestamp with D_t number of variables. We use $\mathbf{y}$ to denote the label.

The architecture of *multi-modal fusion* is illustrated in Figure 1. To integrate static and time-series data, our model leverages knowledge learned from static variables to provide the context for training RNN model for time-series data. In our model the static variables $\mathbf{s}$ are passed through a stack of dense layers before entering the recurrent layer. The operations inside each dense layer are:

$$\tilde{\mathbf{s}}_l = \sigma(\mathbf{W}_l \mathbf{s}_{l-1} + \mathbf{b}_l) \tag{1}$$

$$\mathbf{s}_l = \mathbf{r}_l \odot \tilde{\mathbf{s}}_l, \ \mathbf{r}_l \sim \text{Bernoulli}(p) \tag{2}$$

Eq. (1) represents the calculations in the dense layer l, where $\sigma(\cdot)$ is the activation function, e.g., ReLU. Eq. (2) is an abstract representation of the dropout, where $\odot$ is the element-wise multiplication and Bernoulli(p) denotes Bernoulli distribution. We find it is helpful to add dropouts in between dense layers to prevent overfitting. Let $\mathbf{s}_L$ be the representation generated by the last dense layer. $\mathbf{s}_L$ will be used as the initial hidden state of the latter recurrent layer. Similar to the usage of RNN for natural language processing, we use RNN to learn a representation for sequential time-series data. Throughout the Methodology section, we use Long Short Term Memory (LSTM) as the RNN. Similar derivations can be done for other RNN models, such as Gated Recurrent Unit (GRU). To demonstrate its generality, we evaluated multi-modal fusion in combination with different RNN models (LSTM, m-RNN, GRU-D, BRITS and LGnet) in our experiments.

The calculation of hidden state $\mathbf{h}_t$ at each recurrent step t can be expressed as:

$$\mathbf{h}_0 = \tanh(\mathbf{U}\mathbf{s}_L) \tag{3}$$

$$\mathbf{h}_t = \begin{cases} \text{RNNCell}(\mathbf{h}_{t-1}, \mathbf{x}_t) & t > 1 \\ \text{RNNCell}(\mathbf{h}_0, \mathbf{x}_1) & t = 1 \end{cases} \tag{4}$$

where $\mathbf{h}_0$ is a linear transformation of $\mathbf{s}_L$ with hyperbolic tangent to ensure $\mathbf{h}_0$ has the same range as $\mathbf{h}_t$, $t \geq 1$, RNNCell represents all the operations and gate functions in the LSTM cell and $\mathbf{x}$ is the time-series input. The parameters of dense layers and the recurrent

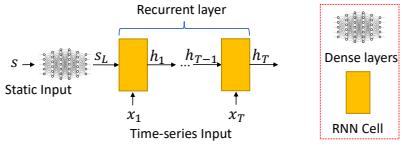

Figure 1: Multi-modal fusion

layers can be jointly learned via back-propagation. When optimizing the parameters of dense layers for static input, the gradient w.r.t. parameter $\mathbf{W}_{l^*}$ at the l^*-th dense layer is a product of all the

Jacobian matrices from T recurrent steps and $L - l^*$ dense layers:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{W}_{l^*}} = \frac{\partial \mathcal{L}}{\partial \hat{\mathbf{y}}} \frac{\partial \hat{\mathbf{y}}}{\partial \mathbf{h}_T} \prod_{t=1}^{T} \frac{\partial \mathbf{h}_t}{\partial \mathbf{h}_{t-1}} \frac{\partial \mathbf{h}_0}{\partial \mathbf{s}_L} \prod_{l=l^*+1}^{L} \frac{\partial \mathbf{s}_l}{\partial \mathbf{s}_{l-1}} \frac{\partial \mathbf{s}_{l^*}}{\partial \mathbf{W}_{l^*}} \tag{5}$$

where we leave out $\hat{\mathbf{y}}, \mathbf{y}$ in loss $\mathcal{L}$. However, the gradient vanishing problem could happen if $\frac{\partial \mathbf{h}_t}{\partial \mathbf{h}_{t-1}}$ is close to zero and T is large, which hampers the learning of dense layer parameters. Thus, we recommend choosing RNN models targeting the vanishing gradient problem, e.g., LSTM or GRU, which can easily let the gradients from the recurrent layers back-propagate to the dense layers.

Our approach is inspired by image captioning approaches based on the image as the static input and descriptive sentences as the sequential input. As indicated by [18, 38], the learned representation of static input is only input once to the RNN. As empirically verified by [18, 38] in image caption problem and our own experiments on the EHR datasets, feeding the output of dense layers once at $t = 1$ yields better results than feeding it at each recurrent step (static-repeat) which may cause the network to exploit noise in the static variables and overfit [38]. The comparison of multi-modal fusion and static-repeat can be found in Section 4.3.

3.3 CrossNet

Recent RNN models ((e.g., GRU-D [5], BRITS [4] and LGnet [37])) have been highly effective in exploiting missingness in clinical data for predicting outcomes by integrating imputation and prediction. However, the existing RNN models do not handle static variables (e.e., comorbidities) that are commonly available and predictive of clinical outcomes. There exist correlations *across* heterogeneous clinical variables including correlations among static variables, correlations among time-series variables and correlations across static and time-series variables. As an example of cross-modal correlation, resting heart rates are correlated with age [31]. It can be particularly beneficial to exploiting cross-modal correlations when time-series data have high missing rates.

We now introduce *CrossNet*, an integrated RNN-based model for clinical predictions based on static and time-series data. CrossNet has several salient features. (1) It employs *multi-modal fusion* to integrate static and time-series inputs. (2) It performs *cross-modal imputation* which exploits correlations across heterogeneous data to impute the missing values of both static and time-series variables. (3) It jointly optimizes the classification and imputation objectives in an end-to-end deep model, such that the imputation objective can help supervise the classification objective and vice versa. (4) It has simpler structure than recent state-of-the-art models (e.g., BRITS). The computational efficiency makes our model easier to be deployed in real-time EWS at large scales. The overall architecture of CrossNet is shown as Figure 2. As we have introduced multi-modal fusion in the last section, we now focus on the cross-modal imputation components of CrossNet in the rest of this section.

We first introduce the representation of the data matrices used for encoding the missing value information. To encode missing value information in static variables $\mathbf{s} \in \mathbb{R}^{D_s}$ of sample i, we use an additional mask vector $\mathbf{m_s} \in \{0, 1\}^{D_s}$, which has the same dimension as static variable vector $\mathbf{s}$. On the other hand, to encode missing value information in time-series variables, we choose to use both the mask matrix $\mathbf{m}$ and the time interval matrix δ defined in the

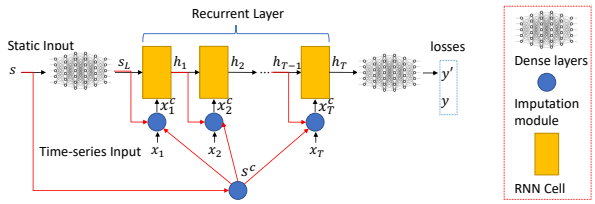

Figure 2: CrossNet architecture

same way as previous approaches, e.g., GRU-D [5] and BRITS [4]. We denote the multivariate time series of the i-th sample as $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_T]^{\mathrm{T}}$ for total T timestamps. We denote the observation at t-th timestamp as $\mathbf{x}_t \in \mathbb{R}^{D_t}$. The corresponding mask vector at t is $\mathbf{m}_t \in \{0, 1\}^{D_t}$. $\mathbf{m_s}$ and $\mathbf{m}_t$ are formally defined as:

$$\mathbf{m_s}^d = \begin{cases} 0 & \text{if } \mathbf{s}^d \text{ missing} \\ 1 & \text{otherwise} \end{cases} \quad \mathbf{m}_t^d = \begin{cases} 0 & \text{if } \mathbf{x}_t^d \text{ missing} \\ 1 & \text{otherwise} \end{cases} \tag{6}$$

where d denotes the d-th variable in either static or time-series variables. The time interval vector $\delta_t \in \mathbb{R}^{D_t}$ captures how long a particular variable has been missing continuously. Therefore, each entry of the time interval vector is defined as:

$$\delta_t^d = \begin{cases} ts_t - ts_{t-1} + \delta_{t-1}^d & \text{if } t > 1, \mathbf{m}_{t-1}^d = 0 \\ ts_t - ts_{t-1} & \text{if } t > 1, \mathbf{m}_{t-1}^d = 1 \\ 0 & \text{if } t = 1 \end{cases} \tag{7}$$

where ts_t represents the normalized time at the timestamp t. Figure 3 shows an example on how the masks and time interval matrix are calculated from the raw static variables and time-series variables. In order to impute the missing values in static variables using

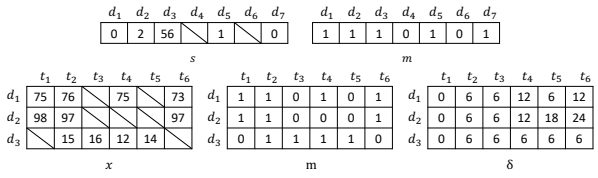

Figure 3: The upper one is an illustration of generating the mask vector for static input, where d_1 to d_7 represent the static variables. The lower one is an illustration of generating mask matrix and time interval matrix for time-series input. d_1, d_2 and d_3 represent the time-series variables; $t_1 = 0$, $t_2 = 6$, $t_3 = 12$, $t_4 = 18$, $t_5 = 24$, $t_6 = 30$ represent the timestamps, where they are offset by 6 hours.

their correlations with other static variables, we apply a linear transformation on the original static variables $\mathbf{s}$ and then use the mask vector $\mathbf{m_s}$ of the static variables to determine whether to use the original values or imputed values as the input. Thus, the imputed static variables $\mathbf{s}^c$ can be derived as:

$$\hat{\mathbf{s}} = \mathbf{W}_s \mathbf{s} + \mathbf{b}_s \tag{8}$$

$$\mathbf{s}^c = \mathbf{m_s} \odot \mathbf{s} + (1 - \mathbf{m_s}) \odot \hat{\mathbf{s}} \tag{9}$$

In Eq. (8), $\mathbf{W}_s$ and $\mathbf{b}_s$ are the parameters of the linear transformation. Different from the original LSTM model, we introduce the temporal

decay γ_t, which addresses the effect of previous hidden state on the calculations at current time step:

$$\gamma_t = \exp\{-\max(0, \mathbf{W}_\gamma \delta_t + \mathbf{b}_\gamma)\} \tag{10}$$

The procedure of imputing the missing values in time-series variables is more complicated. The missing values are imputed based on three types of data, which are their own historical values, the contemporary values of other time-series variables, and the static variables. The imputed time-series variables $\mathbf{x}^c$ are calculated by:

$$\tilde{\mathbf{x}}_t = \mathbf{W}_x \mathbf{x}_t + \mathbf{b}_x \tag{11}$$

$$\hat{\mathbf{x}}_t = \mathbf{W}_h(\gamma_t \odot \mathbf{h}_{t-1}) + \mathbf{W}_{xs}\mathbf{s}^c + \tilde{\mathbf{x}}_t + \mathbf{b}_h \tag{12}$$

$$\mathbf{x}_t^c = \mathbf{m}_t \odot \mathbf{x}_t + (1 - \mathbf{m}_t) \odot \hat{\mathbf{x}}_t \tag{13}$$

where $\mathbf{W}_h$, $\mathbf{W}_{xs}$, $\mathbf{W}_x$, $\mathbf{b}_x$ and $\mathbf{b}_h$ are the parameters of the linear transformation, and $\mathbf{h}_{t-1}$ is the previous hidden state of LSTM. At each recurrent step, $\mathbf{x}_t^c$, $\mathbf{m}_t$, γ_t, $\mathbf{h}_{t-1}$ are the inputs to the LSTM cell:

$$\mathbf{h}_t = \text{LSTMCell}(\gamma_t \odot \mathbf{h}_{t-1}, [\mathbf{x}_t^c; \mathbf{m}_t]) \tag{14}$$

where $[\mathbf{x}_t^c; \mathbf{m}_t]$ denotes the concatenation of $\mathbf{x}_t^c$ and $\mathbf{m}_t$. The model's final probabilistic output can be computed from the hidden states $\mathbf{h}$ of the LSTM:

$$y' = \sigma(\text{MLP}(\mathbf{h}_T)) \tag{15}$$

where $\sigma(\cdot)$ is the sigmoid function and $\mathbf{h}_T$ is the last hidden state, MLP stands for multilayer perceptron. There are multiple ways of computing the probabilistic output, such as directly using the last hidden state, mean pooling of hidden states from all recurrent time steps and deriving via attention mechanism. In our implementation, we choose to obtain the output directly from the last hidden state, since it is computational efficient and yields the best results. The objective function consists of classification loss, imputation loss and regularization terms, which jointly optimizes the model for achieving the goal of imputation and classification simultaneously. For classification loss, we choose the binary cross entropy loss, which is defined as:

$$\mathcal{L}_C = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log y_i' + (1 - y_i)\log(1 - y_i')] \tag{16}$$

where y_i' is the probabilistic output defined by Eq. (15) as a function of static variables $\mathbf{s}$ and time-series variables $\mathbf{X}$. The imputation loss is defined as a Hadamard product of element-wise error and mask:

$$\mathcal{L}_I(\mathbf{x}_i', \mathbf{x}_i) = \frac{1}{N}\sum_{i=1}^{N} \|\mathbf{m}_i \odot (\mathbf{x}_i' - \mathbf{x}_i)\|_1 \tag{17}$$

where $\mathbf{x}_i'$ represents any imputed values, such as $\mathbf{s}^c$ or $\mathbf{X}^c$, and $\mathbf{x}_i$ represents any original values, such as $\mathbf{s}$ or $\mathbf{X}$. The intuition of $\mathcal{L}_I$ is to only compute the mean absolute errors of variables that originally exist in EHR. In addition, we devise regularization terms to help the imputation modules generalize better:

$$\mathcal{R} = \sum_{\alpha \in A} \|\mathbf{W}_\alpha\|_1 + \|\mathbf{w}_x'\|_1 + \|\mathbf{w}_s'\|_1 \tag{18}$$

where the first term $\sum_{\alpha \in A} \|\mathbf{W}_\alpha\|_1$ enforces learning sparse weight matrices $\mathbf{W}_{\alpha \in A} = \{\mathbf{W}_s, \mathbf{W}_h, \mathbf{W}_{xs}, \mathbf{W}_x\}$ for all the transformations in the imputation modules; the terms $\|\mathbf{w}_x'\|_1$, $\|\mathbf{w}_s'\|_1$ are L1-norms of matrix diagonal $\mathbf{w}_x' = [w_x^{d,d}]_{d=1}^{D_t}$ and $\mathbf{w}_s' = [w_s^{d,d}]_{d=1}^{D_s}$, which enforce that the cross-modal imputation computes the imputed values based on other variables. The overall loss function is a weighted sum of all the classification loss, imputation losses and regularization terms:

$$\mathcal{L} = \mathcal{L}_C + \lambda_1 \mathcal{L}_I(\mathbf{s}^c, \mathbf{s}) + \lambda_2 \mathcal{L}_I(\mathbf{X}^c, \mathbf{X}) + \lambda_3 \mathcal{R} \tag{19}$$

where $\lambda_1, \lambda_2, \lambda_3$ represent the loss weights which can be determined via hyper-parameter tuning. By introducing the static variables into the CrossNet model and its loss function together with classification objective, we exploit the cross-modal imputation to guide the model learning a good representation for the classification task.

We note the significant differences between CrossNet and earlier RNN-based imputation models. Different from BRITS and GRU-D that only perform imputation for time-series input, in CrossNet the intermediate imputed value $\hat{\mathbf{x}}_t$ is also inferred by the cross-modal information from imputed static variables $\mathbf{s}^c$, as defined in Eq. (12). Moreover, our model also differs substantially from BRITS in that we use a linear combination of transformed data from different sources, such as time-series variables, static variables and previous hidden states, as shown by Eq. (11), (12), (13). We rely on L1 regularization to supervise the learning of parameters related to cross-modal imputation. In contrast, BRITS uses nested matrix multiplications to perform imputation, which incurs higher computation cost. We discover that some of the nested matrix multiplications without non-linearities added in between are redundant. This induces extra computational burden on the forward pass and back-propagation for the recurrent layer without improving the predictive performance. Hence, CrossNet improves the predictive performance by exploiting cross-modal correlations while enhancing computational efficiency. The experimental results on predictive performance and computational efficiency are presented in Section 4.4 and Section 4.7, respectively.

4 EXPERIMENTAL EVALUATION

We evaluate the proposed approaches on a real-world dataset of oncology inpatients. To assess the generality of our approaches, we also present experimental results on the MIMIC-III ICU dataset [17]. We first compare the predictive performance of our approaches against state-of-the-art models for clinical prediction. We then evaluate the imputation accuracy under cross-modal imputation. To assess the impact of CrossNet on clinical practice we present a case study on configuring CrossNet for realistic clinical early warning scenarios. Finally, we report the computation efficiency of CrossNet.

4.1 Datasets and Preprocessing

The oncology dataset was extracted from an EHR of adult oncology patients in the general hospital wards of a major research hospital in the United States. The EHR contained static variables (e.g., patient demographics and comorbidities) and time-series variables (e.g., vital signs, laboratory results, and medications) of 20,700 adult hospitalizations from 2014 to 2017 for cancer or stem cell transplant. 9% of the patients encountered clinical deterioration defined as a composite outcome of ICU transfer or mortality during their hospital stays. Some of the time-series variables were collected manually

by nurses at a frequency of several readings per day, while others were more frequent measurements. The irregularity of the data collection frequency made the overall data space sparse. To make it easier for the recurrent models to learn useful representation, the original EHR data were down-sampled into 6-hour segments. For time-series variables with multiple values within a 6-hour segment, we used the last observed values in that 6-hr segment. If there was no value observed in the 6-hr segment, we treated this as a missing case for the corresponding variable. The number of extracted static variables is 128 and the number of extracted time-series variables is 41.

To simulate an early warning system in a retrospective study, a sliding window of 84 hours and a stride of 6 hours were applied to extract samples and labels from the raw EHR data. Within the window, we used the data in the first 72 hours (containing 12 6-hour segments) as the input, and the onset of deterioration (ICU transfer or death) from the 78-th to the 84-th hour as the label to ensure a minimum of 6-hour lead time. We choose 6 hours as the prediction horizon because it gives clinicians sufficient lead time to apply interventions, while allowing models to predict events in the near future with reasonable accuracy. We excluded the encounters that last shorter than 84 hours from analysis. For static variables, we duplicated their values while sliding the window and applied one-hot encoding for the categorical static variables. After preprocessing, we got a total of 449,672 samples, where only 0.31 % of them were the cases with deterioration events. The low positive rate was because only a small portion (9%) of the patients had deterioration and the deterioration events most likely happened only once during the entire hospital stay. To assess the generalizability of our approaches, we repeated the evaluation on the MIMIC-III dataset, a public dataset with over 60,000 intensive care unit (ICU) admissions [17]. The prediction task is to generate early warning of patient death in the ICU. We extracted 25 time-series variables (e.g., vital signs, laboratory results) and 32 static variables (e.g., demographics, diagnosis) from the MIMIC-III dataset. We used a similar sliding window approach to create samples but with more granular segments in consistency with the ICU setting. For each time-series variable, we extracted first 9 hours' data (containing 18 30-minute segments) from the Chartevents table as the time-series input. If there were multiple values within a 30-minute segment, we used the last observed value to represent the value in that segment. If there was no value within a 30-minute segment, we treated the corresponding variable as missing in this segment. The demographics information from the Admissions table and the diagnosis from Diagnoses_icd table were selected as the static variables. The label was defined as the onset of death from the 10-th to the 11-th hour, which yielded a minimum of one-hour lead time for preparing intervention. Patients whose stays were shorter than 11 hours were excluded from the samples. After preprocessing, we obtained 99,622 samples with a positive rate of 2.3%. The samples obtained from the MIMIC dataset are generally sparser, where all the time-series variables have data missing rate larger than 0.5. Thus, the evaluation results on the MIMIC dataset reflected how the models performed on sparser input data.

4.2 Evaluation Setting

To evaluate a model's performance on unseen patients, we randomly split the dataset by patients so that 75% of the patients belong to the training set and the rest 25% of the patients belong to the testing set. Among the patients in the training set, we further split 10% of them as a validation set for hyperparameter tuning. For the oncology dataset, the positive ratios in the training set (0.30%), validation set (0.28%), and testing set (0.34%) remain close to the positive ratio of the whole dataset (0.31%). After partitioning the dataset, we perform the aforementioned preprocessing to obtain samples for model training, validation and testing. Since the oncology dataset is extremely unbalanced, we apply random over-sampling to the training set, where the resulting positive ratio is balanced at 50%. All the data are normalized before being fed to the models. The results are obtained from 10 repeated runs with both means and standard deviations reported. To ensure a fair comparison, all the models used in the evaluation have the same number of hidden units in RNN ($h = 108$), network depth, size of dense layers, and batch size of 64. The models and experiments are implemented via PyTorch [32] and the code is available to the public [1].

Table 1: Predictive performance of RNN models with time-series variables and RNN models integrating static and time-series variables with different methods ($mean(\sigma)$). (All the AUPRC results from oncology dataset are relatively low due to the fact that the positive ratio is only 0.31%)

	Oncology		MIMIC	
Model	AUROC	AUPRC	AUROC	AUPRC
LSTM	0.8132(0.0012)	0.0258(0.0014)	0.7414(0.0044)	0.1379(0.0113)
m-RNN	0.7983(0.0021)	0.0260(0.0012)	0.7629(0.0021)	0.1456(0.0062)
GRU-D	0.8042(0.0023)	0.0252(0.0008)	0.7838(0.0065)	0.1694(0.0157)
BRITS	0.8174(0.0017)	0.0280(0.0016)	0.7810(0.0043)	0.1676(0.0107)
LGnet	0.8130(0.0013)	0.0268(0.0014)	0.7786(0.0028)	0.1664(0.0088)
C. LSTM	0.8108(0.0066)	0.0259(0.0017)	0.8383(0.0003)	0.1687(0.0027)
S. LSTM	0.7996(0.0077)	0.0234(0.0011)	0.8342(0.0019)	0.1692(0.0089)
M. LSTM	**0.8237(0.0031)**	**0.0288(0.0012)**	**0.8409(0.0006)**	**0.1729(0.0054)**
C. m-RNN	0.7940(0.0075)	0.0236(0.0010)	0.8489(0.0030)	0.1906(0.0195)
S. m-RNN	0.7626(0.0063)	0.0174(0.0021)	0.8414(0.0007)	0.1931(0.0112)
M. m-RNN	**0.8063(0.0012)**	**0.0268(0.0011)**	**0.8561(0.0082)**	**0.1958(0.0235)**
C. GRU-D	0.8081(0.0074)	0.0249(0.0010)	0.8536(0.0019)	0.2245(0.0111)
S. GRU-D	0.7899(0.0015)	0.0232(0.0029)	0.8540(0.0090)	0.2190(0.0184)
M. GRU-D	**0.8213(0.0021)**	**0.0279(0.0004)**	**0.8557(0.0051)**	**0.2439(0.0213)**
C. BRITS	0.8172(0.0015)	0.0275(0.0012)	0.8498(0.0007)	0.2056(0.0003)
S. BRITS	0.8226(0.0014)	0.0270(0.0009)	0.8521(0.0033)	0.2103(0.0070)
M. BRITS	**0.8339(0.0022)**	**0.0291(0.0012)**	**0.8553(0.0054)**	**0.2186(0.0023)**
C. LGnet	0.8120(0.0043)	0.0278(0.0011)	0.8490(0.0069)	0.1915(0.0165)
S. LGnet	0.8095(0.0043)	0.0262(0.0026)	0.8493(0.0028)	0.1932(0.0507)
M. LGnet	**0.8240(0.0036)**	**0.0286(0.0008)**	**0.8520(0.0043)**	**0.2074(0.0502)**

4.3 Impact of Multi-Modal Fusion

The first set of experiments is designed to evaluate the impact of multi-modal fusion on predictive performance. We use the area under the ROC curve (AUROC) and the area under the precision-recall curve (AUPRC) as performance metrics to evaluate predictive performance.

[1] https://github.com/dingwenli/crossnet

We report the performance of four groups of models:

- *RNN models for time-series data*: Bidirectional LSTM without imputation, and state-of-the-art recurrent models with imputation components including m-RNN [29], GRU-D [5], BRITS [4] and LGnet [37]. While the RNN models use only time-series data, we include their results as baselines to assess the potential contributions of static variables to clinical predictions.
- *Concatenation* version of the RNN models (denoted as **C. Model**): We apply the commonly used representation concatenation approach to the RNN models, which concatenates the learned representations of static and time-series variables at latter layers.
- *Static-repeat* version of the RNN models (denoted as **S. Model**): The representations learned from static variables are repeatedly fed together with time-series variables to RNN at each recurrent step, which is an integration approach adopted by state-of-the-art clinical prediction models [20, 34].
- *Multi-modal* version of the RNN models (denoted as **M. Model**): We apply the proposed multi-modal fusion to the RNN models.

The RNN models use time-series input only. The concatenation, static-repeat and multi-modal versions of the RNN models use both static and time-series variables as inputs.

We have the following observations on the results in Table 1. (1) The multi-modal version of each RNN model outperforms both the concatenation version and the static-repeat version of the same model in terms of AUROC (p<0.05) and AUPRC (p<0.05). The results show the multi-modal fusion is a more effective approach to integrate static and time-series variables than concatenation and static-repeat. (2) The consistent performance improvement of multi-modal fusion on different RNN models demonstrates its generality as a approach to integrate static variables with RNN models. (3) While the multi-modal versions of the RNN models deliver superior performance over the original RNN models with only time-series input (all with p<0.05), the concatenation version and static-repeat version do not always outperform the original models for the oncology dataset. These results show that, while there are significant potential benefits to integrate static variables into clinical predictive models, it is important to employ an effective approach to build the integrated model, which is the contribution of the proposed multi-modal fusion approach.

4.4 Impact of Cross-Modal Imputation

The second set of experiments focuses on evaluating the impact of cross-modal imputation on predictive performance. To compare cross-modal imputation to various types of imputation approaches, we select the following three groups of models:

- *Multi-modal bidirectional LSTM trained on imputed data:* The first three baselines follow a traditional two-step process including imputation followed by model training based on the imputed dataset. We select three imputation methods, KNN imputation [3], MICE [1] as standard imputation methods and E^{2}GAN [25] as a state-of-the-art imputation method published recently. For KNN imputation and MICE, we apply two KNN imputation or MICE procedures to static data and time-series data separately. For E^{2}GAN, since it is designed for accepting time-series input, we use it to impute the time-series data and rely on MICE for the static variables. Then, we train multi-modal bidirectional LSTM on the imputed data as the classification model. These models are denoted KNN*, MICE* and E^{2}GAN*.
- *Multi-modal version of RNN-based imputation models:* The second set of baselines are the multi-modal fusion versions of RNN models that integrate imputation of time-series data and predictions, including m-RNN [29], GRU-D [5], BRITS [4] and LGnet [37]. As the original RNN models can only impute time-series data, we perform MICE or KNN imputation for the static variables of these models, denoted with superscripts $^{\mathrm{m}}$ and $^{\mathrm{k}}$, respectively. Thus, these multi-modal RNN models have their own imputation components for the time-series variables and MICE or KNN imputation for the static variables, but they do not exploit cross-modal correlation.
- *CrossNet:* We build CrossNet with bidirectional LSTM cell as the RNN cell. CrossNet integrates prediction and cross-modal imputation of static and time-series data.

Table 2: Predictive performance of different imputation models on two datasets ($mean(\sigma)$).

	Oncology		MIMIC	
Model	AUROC	AUPRC	AUROC	AUPRC
KNN*	0.8201(0.0027)	0.0272(0.0019)	0.8434(0.063)	0.1792(0.0103)
MICE*	0.8217(0.0022)	0.0283(0.0022)	0.8422(0.0058)	0.1695(0.0118)
E^{2}GAN*	0.8236(0.0039)	0.0276(0.0018)	0.8467(0.0047)	0.1750(0.0157)
M. m-RNN$^{\mathrm{k}}$	0.8089(0.0014)	0.0268(0.0008)	0.8488(0.0075)	0.1932(0.0236)
M. m-RNN$^{\mathrm{m}}$	0.8078(0.0012)	0.0261(0.0006)	0.8524(0.0087)	0.1945(0.0149)
M. GRU-D$^{\mathrm{k}}$	0.8197(0.0015)	0.0276(0.0011)	0.8528(0.0069)	0.2334(0.0152)
M. GRU-D$^{\mathrm{m}}$	0.8221(0.0016)	0.0280(0.0012)	0.8541(0.0078)	0.2393(0.0114)
M. BRITS$^{\mathrm{k}}$	0.8328(0.0014)	0.0268(0.0013)	0.8530(0.0123)	0.2188(0.0195)
M. BRITS$^{\mathrm{m}}$	0.8343(0.0011)	0.0265(0.0009)	0.8554(0.0101)	0.2169(0.0206)
M. LGnet$^{\mathrm{k}}$	0.8258(0.0029)	0.0278(0.0017)	0.8531(0.0090)	0.2133(0.0168)
M. LGnet$^{\mathrm{m}}$	0.8254(0.0018)	0.0288(0.0012)	0.8526(0.0083)	0.2118(0.0174)
CrossNet	**0.8425(0.0009)**	**0.0312(0.0020)**	**0.8607(0.0091)**	**0.2617(0.0285)**

The results reported in Table 2 show that with cross-modal imputation CrossNet can effectively exploit correlations across static and time-series variables to improve predictive performance. We further analyze the performance of CrossNet on samples with different missing rates. The models are evaluated on the subsets of the oncology dataset with different missing rates (all the variables extracted from the MIMIC dataset have missing rates larger than 0.5). As shown in Table 3, CrossNet consistently outperforms all the baseline models for all three subsets with different missing rates. The performance improvement is more significant for higher missing rates. For the samples with missing rate above 0.5, the AUPRC is improved by 125.9% over that of multi-modal BRITS with MICE for static variables. For those samples with high missing rates, we observe that most of these missing data are caused by losing large segments of time-series data. In such cases, cross-modal imputation is particular beneficial by exploiting the correlation between static variables and time-series variables. We observe that MICE [1] and

Table 3: Predictive performance for different missing rates ($mean(\sigma)$).

	Missing rate < 0.2		Missing rate $[0.2, 0.5)$		Missing rate ≥ 0.5	
Model	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC
KNN*	0.8215(0.0092)	0.0276(0.0024)	0.6520(0.0185)	0.0145(0.0032)	0.6056(0.0366)	0.0078(0.0014)
MICE*	0.8232(0.0104)	0.0282(0.0020)	0.6522(0.0226)	0.0141(0.0023)	0.5852(0.0431)	0.0034(0.0008)
E^{2}GAN*	0.8254(0.0047)	0.0279(0.0039)	0.6810(0.0165)	0.0130(0.0014)	0.6025(0.0372)	0.0098(0.0019)
M. m-RNN^k	0.8103(0.0022)	0.0271(0.0018)	0.6530(0.0171)	0.0081(0.0019)	0.6015(0.0352)	0.0045(0.0010)
M. m-RNN^m	0.8082(0.0026)	0.0265(0.0016)	0.6527(0.0162)	0.0068(0.0016)	0.5966(0.0323)	0.0031(0.0005)
M. GRU-D^k	0.8210(0.0034)	0.0278(0.0013)	0.7019(0.0206)	0.0118(0.0028)	0.6157(0.0298)	0.0033(0.0004)
M. GRU-D^m	0.8233(0.0042)	0.0283(0.0009)	0.7026(0.0193)	0.0117(0.0031)	0.6203(0.0339)	0.0027(0.0004)
M. BRITS^k	0.8338(0.0036)	0.0293(0.0024)	0.7227(0.0196)	0.0168(0.0022)	0.6467(0.0326)	0.0326(0.0043)
M. BRITS^m	0.8349(0.0039)	0.0290(0.0027)	0.7232(0.0188)	0.0173(0.0011)	0.6479(0.0349)	0.0317(0.0016)
M. LGnet^k	0.8270(0.0032)	0.0286(0.0022)	0.7194(0.0201)	0.0177(0.0024)	0.6360(0.0318)	0.0306(0.0016)
M. LGnet^m	0.8265(0.0057)	0.0281(0.0017)	0.7209(0.0189)	0.0180(0.0027)	0.6366(0.0351)	0.0295(0.0009)
CrossNet	**0.8432(0.0028)**	**0.0330(0.0026)**	**0.7638(0.0188)**	**0.0286(0.0179)**	**0.6930(0.0416)**	**0.0716(0.0094)**

E^{2}GAN [25] do not perform well in predictive performance, which suggests the importance of considering the classification objective in the imputation process to exploit informative missingness of clinical data for prediction.

4.5 Imputation Accuracy

While our primary objective is to improve predictive performance, we also report the imputation accuracy of cross-modal imputation in comparison to both pure imputation methods and RNN models integrating imputation. Since state-of-the-art RNN-based imputation models can only impute time-series variables, we focus on the comparison on the time-series data. The baseline models are categorized into two groups:

- *Imputation-only models:* These models include traditional imputation approaches such as Mean imputation, KNN imputation [3] and MICE [1]. We also select a recent GAN-based model, E^{2}GAN [25] as a baseline.
- *RNN-based imputation models:* These models are the recent RNN-based imputation models that perform imputation and classification in an end-to-end manner, including m-RNN [29], GRU-D [5], BRITS [4] and LGnet [37].[2]

In the evaluation, we compare the imputation accuracy in terms of mean absolute error (MAE) and mean relative error (MRE), defined as:

$$\text{MAE} = \frac{\sum_{i=1}^{N} |\mathbf{x}'_i - \mathbf{x}_i|}{N} \quad \text{MRE} = \frac{\sum_{i=1}^{N} |\mathbf{x}'_i - \mathbf{x}_i|}{\sum_{i=1}^{N} |\mathbf{x}_i|} \tag{20}$$

where $\mathbf{x}'_i$ is the imputed data and $\mathbf{x}_i$ is the original data. In the experiment, we randomly hold out 10% of non-missing values in the time-series data as the test set. Similarly, we first evaluate the imputation accuracy of different methods on both the oncology and the MIMIC datasets. The results are shown in Table 4. Then, we analyze the imputation accuracy of the models when applied on data with various missing rates, as shown in Table 5. Again, the evaluation is performed on the oncology dataset because all the variables we extract from the MIMIC dataset have missing rates higher than 0.5.

[2]In fairness to the baseline RNN imputation models, we report the imputation accuracy of the original models without multi-modal fusion. Multi-modal fusion has negligible impact on imputation accuracy of the time-series data in our experiments.

Table 4: Imputation accuracy on two datasets ($mean(\sigma)$).

	Oncology		MIMIC	
Model	MAE	MRE(%)	MAE	MRE(%)
Mean	0.2188(0.0010)	76.11%(0.67%)	0.0041(0.0003)	153.38%(11.21%)
KNN	0.0104(0.0004)	3.61%(0.14%)	0.0015(0.0003)	56.11%(11.10%)
MICE	0.0072(0.0006)	2.52%(0.24%)	$7.76e^{-4}(8.59e^{-5})$	29.11%(3.21%)
E^{2}GAN	0.0062(0.0007)	2.20%(0.23%)	$2.04e^{-4}(1.83e^{-5})$	7.61%(0.68%)
m-RNN	0.0102(0.0017)	3.59%(0.61%)	$8.82e^{-4}(9.18e^{-5})$	32.89%(3.42%)
GRU-D	0.0100(0.0041)	3.50%(0.71%)	0.0024(0.0003)	90.78%(1.85%)
BRITS	0.0059(0.0008)	2.09%(0.27%)	$1.92e^{-4}(1.51e^{-5})$	7.16%(0.56%)
LGnet	0.0060(0.0009)	2.11%(0.30%)	$2.30e^{-4}(1.88e^{-5})$	8.58%(0.70%)
CrossNet	**0.0052(0.0008)**	**1.85%(0.26%)**	$\mathbf{1.66e^{-4}(1.35e^{-5})}$	**6.21%(0.51%)**

We observe that CrossNet outperforms the baseline models for all subsets of samples with different missing rates. The improvement of imputation accuracy is more statistically significant for the samples with larger missing rates (CrossNet, 0.0081(0,0012), versus BRITS, 0.0109(0.0017), p=$1.9e^{-4}$ for missing rate larger than or equal to 0.5), which reflects the benefit of exploiting the cross-modal correlation to impute the missing values when there is a significant amount of data missing in the time-series variables.

4.6 Case Study for Clinical Early Warning

To complement the general evaluation of the models in term of predictive accuracy, we now assess the practical impact of our model on clinical care based on the oncology dataset. In hospital wards, it is critical to avoid alarm fatigue caused by too many alarms that overwhelm clinicians. It is hence crucial to configure an EWS for *alarm rate control*, i.e., to bound the alarm rate generated by an EWS to avoid alarm fatigue. We tune a threshold on the training set such that the model generates no more than 48 alarms on any day for all patients in the training set combined. The average alarm rate of one alarm per 30 minutes is chosen based on the general resource capacity in the oncology wards of our hospital. We compare the models by the sensitivity and specificity at the one alarm per 30 minutes cutoff. The resulting sensitivity and specificity of the integrated model are 0.4218(0.0130) and 0.9486(0.0093) respectively, which significantly outperforms the Modified Early Warning Scores (MEWS) [35] used by many hospitals (p<0.05), as well as

Table 5: Imputation accuracy for different missing rates ($mean(\sigma)$).

	Missing rate < 0.2		Missing rate $[0.2, 0.5)$		Missing rate ≥ 0.5	
Model	MAE	MRE(%)	MAE	MRE(%)	MAE	MRE(%)
Mean	0.2187(0.0009)	76.08%(0.67%)	0.2190(0.0009)	67.55%(0.28%)	0.2193(0.0012)	50.08%(0.27%)
KNN	0.0101(0.0004)	3.50%(0.16%)	0.0133(0.0009)	4.10%(0.26%)	0.0142(0.0005)	3.24%(0.12%)
MICE	0.0071(0.0006)	2.49%(0.26%)	0.0089(0.0011)	2.75%(0.34%)	0.0158(0.0014)	3.61%(0.32%)
E^2GAN	0.0059(0.0006)	2.09%(0.27%)	0.0068(0.0007)	2.10%(0.22%)	0.0112(0.0011)	2.56%(0.25%)
m-RNN	0.0093(0.0024)	3.32%(0.85%)	0.0129(0.0022)	3.97%(0.68%)	0.0907(0.0056)	20.81%(1.28%)
GRU-D	0.0092(0.0038)	3.27%(1.35%)	0.0184(0.0043)	5.69%(1.31%)	0.0331(0.0049)	7.60%(1.12%)
BRITS	0.0058(0.0012)	2.07%(0.41%)	0.0065(0.0010)	2.01%(0.31%)	0.0109(0.0017)	2.49%(0.39%)
LGnet	0.0059(0.0013)	2.09%(0.45%)	0.0071(0.0009)	2.19%(0.28%)	0.0127(0.0021)	2.90%(0.47%)
CrossNet	**0.0050(0.0008)**	**1.77%(0.26%)**	**0.0058(0.0008)**	**1.79%(0.25%)**	**0.0081(0.0012)**	**1.90%(0.27%)**

the concatenation version (p<0.05) and the static-repeat version (p<0.05) of BRITS (state-of-the-art models), as shown in the Table 6. In the second experiment, we simulate the implementation of a more proactive EWS with *false alarm control*, where the alarm rate can be high as long as the number of false alarms is limited to a certain level. In this case, we choose a threshold to set the specificity around 0.95, which helps reduce false alarms. The models are compared based on the sensitivity (also known as detection rate in this scenario). Our model again outperforms the baseline models by significant margins in terms of sensitivity (p<0.05). The results indicate our model can maintain a high detection rate (sensitivity 0.3952(0.0118)) while generating few false alarms (specificity 0.9558(0.0067)).

Table 6: Configuring on EWS with controlled alarm rate ($mean(\sigma)$)

Model	Alarm rate control		False alarm control
	Sensitivity	Specificity	Sensitivity
MEWS	0.3358(0.0115)	0.8257(0.0142)	0.0392(0.0062)
C. BRITS	0.3899(0.0134)	0.9394(0.0097)	0.3581(0.0092)
S. BRITS	0.3891(0.0122)	0.9396(0.0105)	0.3586(0.0131)
CrossNet	**0.4218(0.0130)**	**0.9486(0.0093)**	**0.3952(0.0118)**

4.7 Speed-up of Imputation

In CrossNet, we simplify the computation by avoiding nested matrix multiplications and replace them with sums of the linear transformations of time-series variables, static variables and previous hidden states. The baseline implementation uses nested matrix multiplications to achieve imputation, which is denoted as CrossNet$^+$. The evaluation is to compare our design of cross-modal imputation (CrossNet) against CrossNet$^+$ and BRITS [4]. From previous performance evaluations, we know that BRITS yields the best result among the state-of-the-art models and it is mainly built upon nested matrix multiplications for imputation.

As a result, the number of model parameters for performing cross-modal imputation with our proposed simplified transformation (CrossNet) is 13.6% less than the one implemented by the nested matrix multiplications (CrossNet$^+$). To fairly compare the runtime of CrossNet and CrossNet$^+$, we used the same batch size and number of hidden unit of RNN and tuned other hyperparameters so that the two variants achieved similar optimal predictive performance. We empirically evaluated the runtime of the imputation component of each model trained on the oncology dataset. We executed all the experiments on the same machine with Nvidia GTX1080Ti and Intel Xeon Gold 5118 2.3GHz. CrossNet took 26.2% less time on the forward pass than CrossNet$^+$, as shown in Table 7.

Table 7: Actual runtime of forward pass with different imputation methods ($mean(\sigma)$)

Model	Parameters	Runtime (milliseconds)
BRITS	15,852	1.4672(0.2521)
CrossNet$^+$	37,653	1.5241(0.0856)
CrossNet	32,528	**1.1253(0.2609)**

Interestingly, even though BRITS does not have the static variables imputation and the cross-modal imputation, it is still slower than CrossNet in forward pass, due to the massive amount of nested matrix multiplication. Since forward pass is performed at both learning and inference stage, our simplified computation for cross-modal imputation can expedite the learning process as well as inference without sacrificing the predictive performance. Faster inference is important to clinical early warning systems, which makes it easier for the model to run in real-time for numerous patients.

5 CONCLUSION

This paper presents an integrated clinical predictive model to exploit static and time-series clinical variables in a holistic framework. The evaluation on a large oncology inpatient dataset and the public MIMIC-III dataset demonstrate the superior performance of the proposed approaches in predicting clinical deterioration as well as the generalizability for different RNN models. A case study under realistic clinical settings demonstrates our model can support effective EWS by providing accurate early warnings at moderate alarm rates.

ACKNOWLEDGEMENT

Research reported in this publication was supported by the Washington University Institute of Clinical and Translational Sciences grant UL1TR002345 from the National Center for Advancing Translational Sciences (NCATS) of the National Institutes of Health (NIH), by the Foundation for Barnes-Jewish Hospital and by the Fullgraf Foundation.

REFERENCES

[1] Melissa J. Azur, Elizabeth A. Stuart, Constantine Frangakis, and Philip J. Leaf. 2011. Multiple imputation by chained equations: what is it and how does it work? *International journal of methods in psychiatric research* 20, 1 (Mar 2011), 40–49.

[2] A. D. Bedoya, M. E. Clement, M. Phelan, R. C. Steorts, C. O'Brien, and B. A. Goldstein. 2019. Minimal Impact of Implemented Early Warning Score and Best Practice Alert for Patient Deterioration. *Critical Care Medicine* 47, 1 (Jan 2019), 49–55.

[3] Lorenzo Beretta and Alessandro Santaniello. 2016. Nearest neighbor imputation algorithms: a critical evaluation. *BMC medical informatics and decision making* 16 Suppl 3, Suppl 3 (25 Jul 2016), 74–74.

[4] Wei Cao, Dong Wang, Jian Li, Hao Zhou, Lei Li, and Yitan Li. 2018. BRITS: Bidirectional Recurrent Imputation for Time Series. In *Advances in Neural Information Processing Systems 31*. 6775–6785.

[5] Zhengping Che, Sanjay Purushotham, Kyunghyun Cho, David Sontag, and Yan Liu. 2018. Recurrent Neural Networks for Multivariate Time Series with Missing Values. *Scientific Reports* 8, 1 (17 Apr 2018), 6085.

[6] Zhengping Che, Sanjay Purushotham, Robinder Khemani, and Yan Liu. 2017. Interpretable Deep Models for ICU Outcome Prediction. *AMIA ... Annual Symposium proceedings. AMIA Symposium* 2016 (10 Feb 2017), 371–380. 28269832[pmid].

[7] Edward Choi, Mohammad Taha Bahadori, Joshua A. Kulas, Andy Schuetz, Walter F. Stewart, and Jimeng Sun. 2016. RETAIN: An Interpretable Predictive Model for Healthcare Using Reverse Time Attention Mechanism. In *Proceedings of the 30th International Conference on Neural Information Processing Systems* (Barcelona, Spain) *(NIPS'16)*. Red Hook, NY, USA, 3512–3520.

[8] Edward Choi, Mohammad Taha Bahadori, Andy Schuetz, Walter F. Stewart, and Jimeng Sun. 2016. Doctor AI: Predicting Clinical Events via Recurrent Neural Networks. *JMLR workshop and conference proceedings* 56 (Aug 2016), 301–318.

[9] Edward Choi, Mohammad Taha Bahadori, Elizabeth Searles, Catherine Coffey, Michael Thompson, James Bost, Javier Tejedor-Sojo, and Jimeng Sun. 2016. Multi-Layer Representation Learning for Medical Concepts. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*. 1495–1504.

[10] Edward Choi, Mohammad Taha Bahadori, Le Song, Walter F. Stewart, and Jimeng Sun. 2017. GRAM: Graph-Based Attention Model for Healthcare Representation Learning. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Halifax, NS, Canada) *(KDD '17)*. 787–795.

[11] Edward Choi, Cao Xiao, Walter F. Stewart, and Jimeng Sun. 2018. MiME: Multilevel Medical Embedding of Electronic Health Records for Predictive Healthcare. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems* (Montréal, Canada) *(NIPS'18)*. 4552–4562.

[12] C. Esteban, D. Schmidt, D. Krompaß, and V. Tresp. 2015. Predicting Sequences of Clinical Events by Using a Personalized Temporal Latent Embedding Model. In *2015 International Conference on Healthcare Informatics*. 130–139.

[13] Cristóbal Esteban, Oliver Staeck, Yinchong Yang, and Volker Tresp. 2016. Predicting Clinical Events by Combining Static and Dynamic Information Using Recurrent Neural Networks. *CoRR* abs/1602.02685 (2016). arXiv:1602.02685

[14] J. Ferlay, M. Colombet, I. Soerjomataram, C. Mathers, D.M. Parkin, M. Piñeros, A. Znaor, and F. Bray. 2019. Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *International Journal of Cancer* 144, 8 (2019), 1941–1953.

[15] Jingyue Gao, Xiting Wang, Yasha Wang, Zhao Yang, Junyi Gao, Jiangtao Wang, Wen Tang, and Xing Xie. 2019. Camp: Co-attention memory networks for diagnosis prediction in healthcare. In *2019 IEEE International Conference on Data Mining (ICDM)*. IEEE, 1036–1041.

[16] V. Huddar, B. K. Desiraju, V. Rajan, S. Bhattacharya, S. Roy, and C. K. Reddy. 2016. Predicting Complications in Critical Care Using Heterogeneous Clinical Data. *IEEE Access* 4 (2016), 7988–8001.

[17] Alistair E.W. Johnson, Tom J. Pollard, Lu Shen, Li-wei H. Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G. Mark. 2016. MIMIC-III, a freely accessible critical care database. *Scientific Data* 3, 1 (24 May 2016), 160035.

[18] Andrej Karpathy and Fei-Fei Li. 2015. Deep Visual-Semantic Alignments for Generating Image Descriptions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

[19] Dingwen Li, Patrick G. Lyons, Chenyang Lu, and Marin Kollef. 2020. DeepAlerts: Deep Learning Based Multi-Horizon Alerts for Clinical Deterioration on Oncology Hospital Wards. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020*. 743–750.

[20] C. Lin, Y. Zhang, J. Ivy, M. Capan, R. Arnold, J. M. Huddleston, and M. Chi. 2018. Early Diagnosis and Prediction of Sepsis Shock by Combining Static and Dynamic Information Using Convolutional-LSTM. In *2018 IEEE International Conference on Healthcare Informatics (ICHI)*. 219–228.

[21] Zachary C. Lipton, David C. Kale, Charles Elkan, and Randall Wetzel. 2017. Learning to Diagnose with LSTM Recurrent Neural Networks. arXiv:1511.03677 [cs.LG]

[22] Zitao Liu and Milos Hauskrecht. 2017. A Personalized Predictive Framework for Multivariate Clinical Time Series via Adaptive Model Selection. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management (CIKM '17)*. 1169–1177.

[23] Scott M. Lundberg, Bala Nair, Monica S. Vavilala, Mayumi Horibe, Michael J. Eisses, Trevor Adams, David E. Liston, Daniel King-Wai Low, Shu-Fang Newman, Jerry Kim, and Su-In Lee. 2018. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nature Biomedical Engineering* 2, 10 (01 Oct 2018), 749–760.

[24] Junyu Luo, Muchao Ye, Cao Xiao, and Fenglong Ma. 2020. HiTANet: Hierarchical Time-Aware Attention Networks for Risk Prediction on Electronic Health Records. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '20)*. 647–656.

[25] Yonghong Luo, Ying Zhang, Xiangrui Cai, and Xiaojie Yuan. 2019. E^{2}GAN: End-to-End Generative Adversarial Network for Multivariate Time Series Imputation. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*. 3094–3100.

[26] Patrick G. Lyons, Jeff Klaus, Colleen A. McEvoy, Peter Westervelt, Brian F. Gage, and Marin H. Kollef. 2019. Factors Associated With Clinical Deterioration Among Patients Hospitalized on the Wards at a Tertiary Cancer Hospital. *Journal of oncology practice* 15, 8 (Aug 2019), e652–e665.

[27] Fenglong Ma, Radha Chitta, Jing Zhou, Quanzeng You, Tong Sun, and Jing Gao. 2017. Dipole: Diagnosis Prediction in Healthcare via Attention-Based Bidirectional Recurrent Neural Networks. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Halifax, NS, Canada) *(KDD '17)*. 1903–1911.

[28] Fenglong Ma, Quanzeng You, Houping Xiao, Radha Chitta, Jing Zhou, and Jing Gao. 2018. KAME: Knowledge-Based Attention Model for Diagnosis Prediction in Healthcare. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management* (Torino, Italy) *(CIKM '18)*. New York, NY, USA, 743–752.

[29] Junhua Mao, Wei Xu, Yi Yang, Jiang Wang, Zhiheng Huang, and Alan Yuille. 2015. Deep Captioning with Multimodal Recurrent Neural Networks (m-RNN). *ICLR* (2015).

[30] Yi Mao, Wenlin Chen, Yixin Chen, Chenyang Lu, Marin Kollef, and Thomas Bailey. 2012. An Integrated Data Mining Approach to Real-Time Clinical Monitoring and Deterioration Warning. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 1140–1148.

[31] Paolo Palatini. 1999. Need for a Revision of the Normal Limits of Resting Heart Rate. *Hypertension* 33, 2 (1999), 622–625.

[32] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. 2017. Automatic differentiation in PyTorch. In *NIPS-W*.

[33] Zhi Qiao, Zhen Zhang, Xian Wu, Shen Ge, and Wei Fan. 2020. *MHM: Multi-Modal Clinical Data Based Hierarchical Multi-Label Diagnosis Prediction.* 1841–1844.

[34] Molla Hafizur Rahman, Shuhan Yuan, Charles Xie, and Zhenghui Sha. 2020. Predicting human design decisions with deep recurrent neural network combining static and dynamic data. *Design Science* 6 (2020), e15.

[35] C.P. Subbe, M. Kruger, P. Rutherford, and L. Gemmel. 2001. Validation of a modified Early Warning Score in medical admissions. *QJM: An International Journal of Medicine* 94, 10 (10 2001), 521–526.

[36] Qingxiong Tan, Andy Jinhua Ma, Mang Ye, Baoyao Yang, Huiqi Deng, Vincent Wai-Sun Wong, Yee-Kit Tse, Terry Cheuk-Fung Yip, Grace Lai-Hung Wong, Jessica Yuet-Ling Ching, Francis Ka-Leung Chan, and Pong C. Yuen. 2019. UA-CRNN: Uncertainty-Aware Convolutional Recurrent Neural Network for Mortality Risk Prediction. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management (CIKM '19)*. 109–118.

[37] Xianfeng Tang, Huaxiu Yao, Yiwei Sun, Charu C. Aggarwal, Prasenjit Mitra, and Suhang Wang. 2020. Joint Modeling of Local and Global Temporal Dynamics for Multivariate Time Series Forecasting with Missing Values. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020*. 5956–5963.

[38] O. Vinyals, A. Toshev, S. Bengio, and D. Erhan. 2015. Show and tell: A neural image caption generator. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 3156–3164.

[39] Yanbo Xu, Siddharth Biswal, Shriprasad R. Deshpande, Kevin O. Maher, and Jimeng Sun. 2018. RAIM: Recurrent Attentive and Intensive Model of Multimodal Patient Monitoring Data. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '18)*. 2565–2573.

[40] Jinsung Yoon, Ahmed Alaa, Scott Hu, and Mihaela Schaar. 2016. ForecastICU: A Prognostic Decision Support System for Timely Prediction of Intensive Care Unit Admission. In *Proceedings of Machine Learning Research*, Vol. 48. 1680–1689.

[41] Jinsung Yoon, James Jordon, and Mihaela van der Schaar. 2018. GAIN: Missing Data Imputation using Generative Adversarial Nets. In *Proceedings of the 35th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 80)*. Stockholmsmässan, Stockholm Sweden, 5689–5698.

[42] J. Yoon, W. R. Zame, and M. van der Schaar. 2019. Estimating Missing Data in Temporal Data Streams Using Multi-Directional Recurrent Neural Networks. *IEEE Transactions on Biomedical Engineering* 66, 5 (2019), 1477–1490.