UA-CRNN: Uncertainty-Aware Convolutional Recurrent Neural Network for Mortality Risk Prediction

Qingxiong Tan[1], Andy Jinhua Ma[2], Mang Ye[1], Baoyao Yang[1], Huiqi Deng[1, 2], Vincent Wai-Sun Wong[3], Yee-Kit Tse[3], Terry Cheuk-Fung Yip[3], Grace Lai-Hung Wong[3], Jessica Yuet-Ling Ching[3], Francis Ka-Leung Chan[3], Pong C Yuen[1]*

[1]Department of Computer Science, Hong Kong Baptist University
[2]School of Data and Computer Science, Sun Yat-Sen University
[3]Department of Medicine and Therapeutics, The Chinese University of Hong Kong
[1]{csqxtan, mangye, byyang, pcyuen}@comp.hkbu.edu.hk [2]andyjinhuama@gmail.com [1,2]hqdeng@comp.hkbu.edu.hk
[3]{wongv, yktse, wonglaihung, jessicaching, fklchan}@cuhk.edu.hk [3]terryfungyip@gmail.com

ABSTRACT

Accurate prediction of mortality risk is important for evaluating early treatments, detecting high-risk patients and improving healthcare outcomes. Predicting mortality risk from the irregular clinical time series data is challenging due to the varying time intervals in the consecutive records. Existing methods usually solve this issue by generating regular time series data from the original irregular data without considering the uncertainty in the generated data, caused by varying time intervals. In this paper, we propose a novel Uncertainty-Aware Convolutional Recurrent Neural Network (UA-CRNN), which incorporates the uncertainty information in the generated data to improve the mortality risk prediction performance. To handle the complex clinical time series data with sub-series of different frequencies, we propose to incorporate the uncertainty information into the sub-series level rather than the whole time series data. Specifically, we design a novel hierarchical uncertainty-aware decomposition layer (UADL) to adaptively decompose time series into different sub-series and assign them proper weights according to their reliabilities. Experimental results on two real−world clinical datasets demonstrate that the proposed UA-CRNN method significantly outperforms state-of-the-art methods in both short-term and long-term mortality risk predictions.

CCS CONCEPTS

• Analytics and machine learning → data mining, machine learning and AI • Special data processing and mining → sequential data • Applications → digital health, biomedical informatics

KEYWORDS

Uncertainty-Aware Prediction, Mortality Risk Prediction, Time Series Decomposition, Convolutional Recurrent Neural Network, Machine Learning

CIKM '19, November 3–7, 2019, Beijing, China.

ACM ISBN 978-1-4503-6976-3/19/11...$15.00.
https://doi.org/10.1145/3357384.3357884

ACM Reference format:

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 *The 28th ACM International Conference on Information and Knowledge Management (CIKM '19), November 03–07, 2019, Beijing, China.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3357384.3357884

1 Introduction

Mortality risk prediction, which detects high-risk patients and evaluates early treatment effects, is important for improving clinical outcomes. The widely adopted electronic health records (EHRs) produce large volumes of clinical data, such as time series (e.g., laboratory parameters), static data (e.g., gender and birthdate) and text data (e.g., diagnosis codes) [2, 4, 5, 6]. Among them, clinical time series data, which record values of important health indicators at different time stamps, are very important because they contain critical medical information on dynamic changes in a patient's health status. EHRs are influenced by a substantial number of factors, such as health status and personal arrangement of patients, and the schedule of attending doctors and medical centers [4]. These factors often induce severe irregularity and complexity in clinical time series data [8, 35].

Time intervals between consecutive records vary a lot, ranging from hours to years [1]. The irregularity makes the trajectory of clinical time series hard to capture. Since most machine learning models are designed for regular data, they may suffer from suboptimal performance if directly used for irregular data. Hence, many clinical outcome prediction approaches choose to generate regular data from the original irregular data first [4, 7, 46]. Based on the assumption that the generated data are correct, these methods ignore the uncertainty in the generated data, which refers to the difference between the generated data and the actual data, as shown in Figure 1(a). Due to the varying time intervals in the original time series data, it will result in different uncertainty levels in the generated data. Therefore, it is reasonable to estimate and incorporate the uncertainty information in the generated data

to learn an accurate mortality risk prediction model. In recent years, some uncertainty mechanism based methods have been successfully used for various tasks [28-33, 36], such as image segmentation [31], power system prediction [29] and traffic flow prediction [28]. However, these methods deal with uncertainty caused by other sources, namely model parameters [36], unseen data [28, 32, 33], or measurement errors in the low-quality data [29-31]. Furthermore, all these methods are designed for regular data, which cannot handle the irregular time series data in our mortality risk prediction problem. Therefore, it is important to develop a new diagnosis prediction model that can tackle the uncertainty in the process of generating regular time series data from the original irregular data.

Since every influencing factor has its unique pattern of change, sub-components induced by different factors have different frequencies and reliabilities [9]. Due to this, clinical time series is complex and contains sub-series of different frequencies [8]. This motivates us to incorporate the uncertainty information at the more precise sub-series level. Rather than incorporating the uncertainty at the whole time series data, the sub-series level incorporation provides a much more accurate and robust solution. There are some signal processing techniques (wavelet decompositions (WD) [3], empirical mode decomposition (EMD) [10] and ensemble EMD (EEMD) [9]) for time series decomposition. However, they are not globally optimized with the widely adopted machine learning neural networks. As a result, it is urgent to develop a decomposition network that can adaptively modify the decomposition process with the neural network in the learning process.

In this paper, we propose a novel method called Uncertainty-Aware Convolutional Recurrent Neural Network (UA-CRNN) to address the above challenges. Specifically, Gaussian Process (GP) [11] is first utilized to generate a regular time series for each original time series data and meanwhile estimating corresponding uncertainty information. After that, the proposed end-to-end trainable UA-CRNN method jointly models the generated data and corresponding uncertainty information. In the first layer, a novel uncertainty-aware decomposition layer (UADL) is designed to transform uncertainty information into weight by using a decay function to ensure large uncertainty corresponds to a small weight. Furthermore, to introduce the uncertainty-aware mechanism into the more precise sub-series level, UADL utilizes a hierarchical decomposition structure to decompose time series into several sub-series and assigns them different weights according to their reliabilities. Convolutional layers [20] are then used to learn a feature sequence from each time stamp of uncertainty-aware sub-series features. Finally, recurrent layers [13] are utilized to predict mortality risk by extracting temporal dependencies information from these feature sequences. The major contributions of this paper include:

- We propose a novel UA-CRNN method to incorporate both the generated time series data and corresponding uncertainty information into deep learning architecture to ensure reliable data points receive more attention.
- We design a novel hierarchical uncertainty-aware decomposition layer (UADL) to adaptively decompose time series into different sub-series and assign them proper weights in accordance with their reliabilities.
- We introduce an end-to-end trainable architecture to jointly optimize each component in our proposed UA-CRNN, achieving global optimal solution.
- We empirically show that the proposed UA-CRNN method significantly outperforms existing methods in both short-term and long-term mortality risk predictions on two real-world datasets.

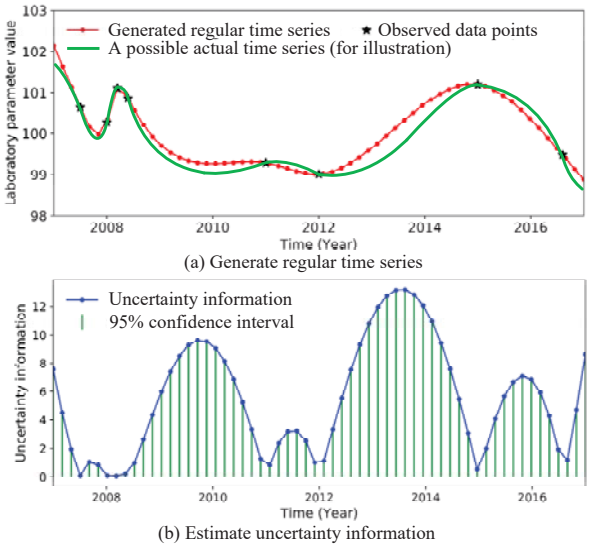


Figure 1. An example of using Gaussian Process (GP) to generate regular time series from the original irregular data. (a) The generated regular time series data fit well to observed data points, but they may deviate from actual data in periods without observations; (b) The uncertainty information evaluates the reliability of the generated data by quantifying the degree in which they may deviate from the actual data.

2 Related Work

In this section, we review the research work on using machine learning methods to conduct diagnosis prediction, especially for mortality risk prediction. After that, we introduce some research work on uncertainty-aware models.

2.1 Machine Learning for Diagnosis Prediction

In recent years, machine learning methods, especially deep learning methods, are widely used for different tasks [38-45], such as predicting the clinical outcome of patients [4, 7, 34]. According to the characteristic of clinical time series data, these methods can be categorized into two categories, i.e., regular data-based and irregular data-based methods. The first category often models data with an equal time interval, e.g., 1 hour [14, 15] and 12 hours [16]. In [14], based on hourly sampled clinical data, the incorporation of missing data patterns into deep learning architecture is proved useful. In [15], the authors show the effectiveness of using the switching-state autoregressive model

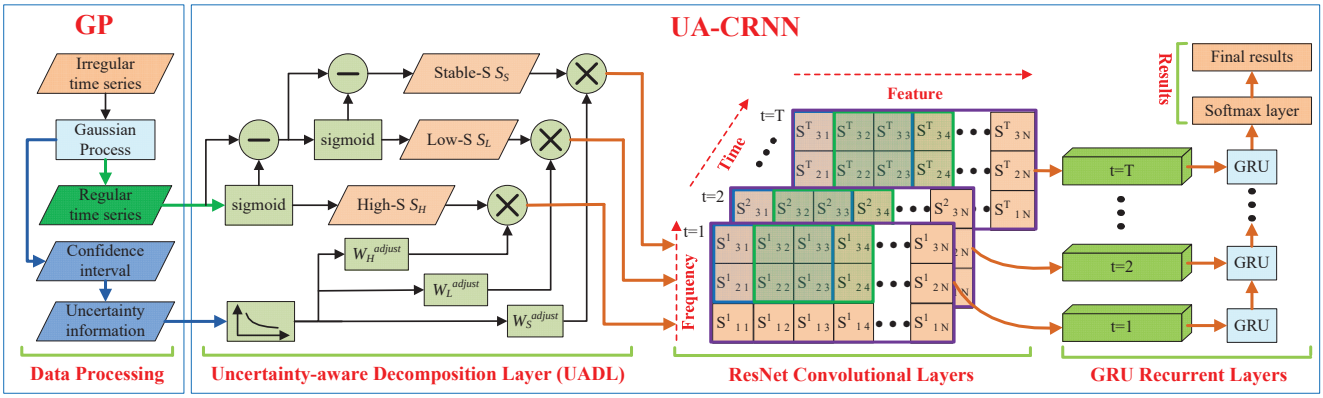


Figure 2. An overview of the proposed UA-CRNN prediction method.

for hourly sampled EHRs data. However, influenced by a substantial number of factors, most clinical datasets are sampled irregularly. Thus, these methods often select a small regularly sampled subset from the whole dataset [17]. As a result, these methods have limited application scenarios.

The irregular data-based methods often use signal processing techniques to process original data first and then use standard machine learning methods to model the generated regular data [18, 19]. In [18], the last occurrence carries forward strategy is used to impute missing values. In [26], a moving window with a time interval of 12 hours is adopted to get regular features. Authors in [19] use population medians and modes to impute missing numeric values and categorical points respectively to obtain regular data. After obtaining regular data, most methods utilize standard machine learning models to predict the clinical outcome by only considering the generated data. However, we can only see a few observed data points from the EHRs data. The way to estimate the data processing quality is checking whether the generated time series fits well with these recorded data points. However, this cannot ensure the generated data totally represent the ground truth of the trajectory because data points in many periods are often missing. Moreover, clinical data have strong irregularity and complexity, which further increases the uncertainty of the generated data. To solve this problem, we propose to incorporate the uncertainty information into deep learning architecture to ensure reliable data points in the generated data receive larger weights and play more important roles.

2.2 Uncertainty-Aware Models

In recent years, some uncertainty-aware methods have been developed to solve different tasks [28-33, 36, 37, 47], such as traffic flow prediction [28], power equipment lifetime prediction [29], image segmentation [31] and autonomous driving [33]. To estimate the uncertainty caused by model parameters, a Bayesian ensembling method is proposed to regularize model parameters around values taken from a prior distribution [36]. To handle the uncertainty that occurs when faced with unseen data, an ensemble strategy together with dropout and bootstrapping is utilized to estimate the predictive uncertainty when encountering dynamic obstacles in new scenarios [33]. The authors in [28] propose to pay particular attention to the non-deterministic property of traffic scenes and better estimate its uncertainty.

The noise in low measuring quality data could also cause uncertainty. A Bayesian inference method is used to analyze the influence of measurement errors and quantify the caused uncertainty to obtain more accurate power equipment lifetime prediction results [29]. The authors in [30] propose to generate attention for every feature with different degrees of noise according to the uncertainty caused by the low measuring quality problem. A new evaluation framework is designed to adjust the segmentation model to consider the uncertainty caused by nighttime images under adverse lighting conditions [31].

These methods are proposed to handle the uncertainty caused by model parameters, unseen data, or measurement errors in the low-quality data. However, all these methods are designed for regular data, which cannot handle the irregular time series data in the mortality risk prediction problem. Therefore, we design a novel uncertainty-aware method to deal with the uncertainty caused by the processing of irregular time series data.

3 Proposed Method

The overview of the proposed UA-CRNN method is given in Figure 2. This architecture contains four parts, namely Gaussian Process (GP) for data processing, uncertainty-aware decomposition layer (UADL), Residual Network (ResNet) convolutional layers and Gated Recurrent Unit (GRU) recurrent layers. First, GP is used to generate regular time series from the original irregular data and meanwhile estimating corresponding uncertainty information. Then, UA-CRNN takes both the generated times series and the uncertainty information as inputs. In the first layer, UADL is designed to transform uncertainty information into weight by using a decay function. Furthermore, to incorporate uncertainty information into the sub-series level, UADL utilizes a hierarchical decomposition structure to decompose time series into different sub-series and assign them different weights according to their reliabilities. Then ResNet is utilized to learn a feature sequence from each time stamp of uncertainty-aware sub-series features. Finally, GRU is adopted to predict mortality risk by extracting temporal dependencies from these feature sequences.

It is notable that UA-CRNN is designed with an end-to-end trainable architecture to enable different parts to optimize their parameters simultaneously. Specifically, in the training process, prediction errors of mortality risk labels are back propagated to the recurrent layers, thus guiding GRU Recurrent Layers to adaptively refine parameters, capture temporal dependencies information and evolve into a sounder network. Furthermore, prediction errors are further back propagated to convolutional layers and finally back to UADL, thus further help direct the feature sequences extraction and the time series decomposition. The end-to-end trainable mechanism enables different parts of UA-CRNN to train their parameters simultaneously to achieve a global optima solution.

3.1 Gaussian Processes for Data Processing

Let us consider a collection of independent observed (training) data $D = \{(t_i, y_i) \mid i = 1, \cdots, N\}$, where the input is time t and the target is value y. Time elapses between consecutive data points are usually irregular, namely, the time interval set $\{t_{i+1} - t_i \mid i = 1, \cdots, N\}$ has more than one value. The target of this step is to utilize knowledge learned from D to predict target Y_* for equally spaced time points.

The basic assumption of GP [11] is that observed value y in a regression problem can be expressed as the superposition of a GP over input space X, output of model $f(t)$ and Gaussian noise, namely:

$$y = f(t) + \varepsilon \tag{1}$$

where $f(t) \sim GP(m(t), k(t,t))$; $m(t)$ and $k(t,t)$ are the mean function and the covariance function of the process respectively; $p(\varepsilon) \sim N(\varepsilon \mid 0, \sigma^2)$.

Since the sum of independent Gaussian variables is also Gaussian, the joint distribution of observed target values $Y = \{y_i\}_{i=1}^N$ and an unknown target Y_*, which can be approximated by values f_* of the postulated GP evaluated for t_*, is represented as the following expression:

$$N\left(\begin{bmatrix} Y \\ f_* \end{bmatrix} \mid 0, \begin{bmatrix} K(T,T) + \sigma^2 I_N & k(t_*) \\ k(t_*)^T & k(t_*, t_*) \end{bmatrix}\right) \tag{2}$$

where $k(t_*) \sim [k(t_1, t_*), \cdots, k(t_N, t_*)]^T$, $T = \{t_i \mid i = 1, \cdots, N\}$; I_N is $N \times N$ unit matrix; K demonstrates the matrix of the covariance between observed data.

Therefore, if we have some observed data points D, we can derive the following predictive distribution, and corresponding mean and covariance. For new input t_*:

$$p(f_* \mid t_*, D) = N(f_* \mid \mu_*, \sigma_*^2) \tag{3}$$

$$\mu_* = k(t_*)^T (K(T,T) + \sigma^2 I_N)^{-1} Y \tag{4}$$

$$\sigma_*^2 = k(t_*, t_*) - k(t_*)^T (K(T,T) + \sigma^2 I_N)^{-1} k(t_*) \tag{5}$$

In this research, we make use of Eq. (4) to generate a regular time series for each original time series data. Correspondingly, we calculate the 95% confidence interval (CI) as uncertainty information by using the following expression to process covariance given in Eq. (5):

$$uncertainty_* = 1.96 \times \sqrt{\sigma_*^2} \tag{6}$$

Figure 1 introduces an example of using GP to generate regular time series data from the original irregular data. The generated regular time series seems reliable because it is close to the observed data points. However, most periods have no observations, which means that actual trajectories in these regions can fluctuate in a wide range. Uncertainty information provides a natural way to evaluate the reliability of the generated data by quantifying the degree in which the generated data may deviate from the actual data. We can see data points generated at periods with intensive observations have narrow CIs and small uncertainty scores. Conversely, data points generated at periods with sparse observations have large uncertainty scores. This is reasonable because when more data points are recorded, more ground truth information about the disease trajectory is kept, and the generated data is more reliable. Therefore, the estimated uncertainty information can well quantitatively evaluate the reliability of the generated data.

3.2 Uncertainty-Aware Decomposition Layer

Uncertainty-aware decomposition layer (UADL) takes both generated time series and corresponding uncertainty information as inputs, as shown in Figure 2. To incorporate uncertainty information into the sub-series level, namely assigning proper weights to different sub-series in accordance with their reliabilities, there are three tasks in UADL. The first task is transforming uncertainty information into appropriate weight. The second task is decomposing time series into different sub-series. Finally, the obtained weights are assigned to these sub-series according to their reliabilities.

Firstly, transform uncertainty into weight. For every time series $X = [x_1, x_2, \cdots, x_T]$ generated by GP, there is a sequence of uncertainty scores $U = [u_1, u_2, \cdots, u_T]$ which quantitatively describes the reliability of the generated data. To achieve more reliable and accurate results, larger attention should be assigned to data with small uncertainty to ensure these reliable data play a more active role. A decay function with the following equation is used to transform uncertainty into weight:

$$w_u = \frac{\max([u_1, u_2, \cdots, u_T]) - u}{\max([u_1, u_2, \cdots, u_T]) - \min([u_1, u_2, \cdots, u_T])} \tag{7}$$

where T is the length of the time series.

This function has a similar structure with the widely used min-max normalization method. Therefore, it produces proper weight, ensuring a large weight is assigned to high-quality data. Besides, it increases the compatibility of the obtained weights by unifying their dimension.

Secondly, decompose time series into several sub-series. Different sub-series often have different stabilities and cause different degrees of uncertainty. To achieve better results, it is more feasible to incorporate uncertainty information into a more accurate sub-series level rather than the whole time series level.

Furthermore, time series of different health indicators often have correlations because changes of one indicator can influence the future trajectory of other indicators. E.g., if the density of white blood cell increases rapidly, it is possible that the patient is suffering from an acute infection and other indicators will change accordingly. To consider the correlation relationship, UADL decomposes time series of different indicators jointly rather than individually. Use the matrix X to represent the time series of different indicators:

$$X = \begin{bmatrix} x_{1\,1} & x_{1\,2} & \cdots & x_{1\,N} \\ x_{2\,1} & x_{2\,2} & \cdots & x_{2\,N} \\ \vdots & \vdots & \vdots & \vdots \\ x_{T\,1} & x_{T\,2} & \cdots & x_{T\,N} \end{bmatrix} \tag{8}$$

where T is the length of time series; N is the number of health indicators.

UADL utilizes a hierarchical structure to decompose the matrix X into three different sub-series, namely high-frequency sub-series (High-S) S_H, low-frequency sub-series (Low-S) S_L and stable sub-series (Stable-S) S_S, which respectively refers to sub-series that fluctuates frequently with strong instability, moderately with medium instability and slowly with weak instability. High-S S_H is first extracted by using a network with the following expression:

$$S_H = \sigma(W_H X + b_H) \tag{9}$$

where $\sigma(\cdot)$ is a sigmoid function; $W_H \in \Re^{T\times T}$ and $b_H \in \Re^{T\times N}$ are trainable weight vectors and bias vectors respectively.

High-S S_H is then subtracted from X. Thus, the residual component of High-S is obtained by using $S^{res}{}_H = X - S_H$. After this, another network is utilized to decompose the residual component of $S^{res}{}_H$ further so that Low-S S_L and Stable-S S_S can be separated. The expression of the network used to extract Low-S S_L from $S^{res}{}_H$ is as the following:

$$S_L = \sigma(W_L S^{res}{}_H + b_L) \tag{10}$$

where W_L and b_L are trainable weight vectors and bias vectors having the same shape with W_H and b_H respectively.

Stable-S S_S is finally obtained by subtracting S_L from the residual component of $S^{res}{}_H$ via the expression:

$$S_S = S^{res}{}_H - S_L \tag{11}$$

Finally, assign weights obtained at task one to sub-series obtained at task two according to their reliabilities. Different sub-series have different stabilities and reliabilities. Stable-S S_S mainly contains stable sub-component induced by slow-changing factor and causes small uncertainty, while High-S S_H consists of fast-changing sub-components and causes severe uncertainty. Based on the principle that large attention and weight should be given to stable and reliable sub-series to help them play a more important and active role, weights assigned to these sub-series are further adjusted by using different weight adjusting parameters with the following expressions:

$$S^{adjust}{}_H = S_H \times W_u \times W_H^{adjust} \tag{12}$$

$$S^{adjust}{}_L = S_L \times W_u \times W_L^{adjust} \tag{13}$$

$$S^{adjust}{}_S = S_S \times W_u \times W_S^{adjust} \tag{14}$$

where $W_H^{adjust}, W_L^{adjust}, W_S^{adjust}$ $(W_S^{adjust} > W_L^{adjust} > W_H^{adjust})$ are weight adjusting parameters; W_u is calculated from the uncertainty score by using Eq. (7).

3.3 Residual Network Convolutional Layers

ResNet was proposed for image recognition [12] and has been used for EHRs data [20, 27]. UA-CRNN constructs convolutional layers by taking residual learning blocks, which consist of several convolutional units and a shortcut connection, from ResNet. The shortcut connection enables ResNet to avoid the vanishing gradient problem, thus gaining strong modeling capacity. The output of a residual learning block for input X is:

$$Y = F(X) + X \tag{15}$$

where $F(\cdot)$ consists of convolution units, batch normalization units and rectified linear units.

Some methods have been proposed to analyze the correlation between multiple clinical time series [21, 22]. However, few methods consider the possibility that such a correlation relationship could come from the sub-series level. E.g., if stable sub-series of a health indicator rises slowly, such as blood glucose level, other health indicators will probably experience dramatic changes in their unstable sub-series. ResNet can jointly analyze different variables by using convolution units and increase the depth of the network without vanishing gradient problem by using shortcut connections, thus effectively capturing correlations between different variables [27]. Therefore, we use ResNet to extract correlation information from the sub-series of different features to learn informative feature sequences.

3.4 Gated Recurrent Unit Recurrent Layers

GRU [13] is a variation of RNN with a simpler structure and can achieve state-of-the-art performance for time series modeling. For a time series $X_t \in \Re^M$, $t = 1,\cdots,T$, GRU has a reset gate r_t^j and an update gate z_t^j for every hidden state h_t^j. The update function is as follows:

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z) \tag{16}$$

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r) \tag{17}$$

$$\tilde{h}_t = \tanh(W x_t + U(r_t \odot h_{t-1}) + b) \tag{18}$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \tag{19}$$

where $\sigma(\cdot)$ is sigmoid function; W_z, U_z, b_z, W_r, U_r and U are trainable matrix; b_z, b_r and b are trainable vectors.

Clinical data contain important temporal dependencies information about dynamic changes in the patient's health status,

which is crucial for improving mortality risk prediction accuracy. Therefore, GRU is utilized to capture temporal dependencies information from representative feature sequences outputted by ResNet convolutional layers. The output of GRU is finally fed to a dense layer to predict the mortality risk of patient.

4 Experiments

In this section, to evaluate the performance of the proposed UA-CRNN method for both short-term and long-term mortality risk predictions, we conduct experiments on two real-world EHRs datasets, i.e., an intensive care unit (ICU) dataset and a chronic disease clinical dataset. Experimental results demonstrate that UA-CRNN outperforms existing methods in various levels of mortality risk predictions and prove its utility for different prediction sceneries.

4.1 Dataset Description

The two real-world datasets used in this experiments are the Peptic Ulcer Bleeding (PUB) dataset and the MIMIC-III [25] dataset.

PUB Dataset

The PUB dataset is a specialist clinical dataset with a cohort of 6,367 Peptic Ulcer Bleeding (PUB) patients, who were under the care of the Endoscopy Center, Prince of Wales Hospital, Hong Kong from January 2007 to December 2016. Patients in the PUB dataset have long visit records. Seven types of laboratory parameters, which are closely related to the PUB disease, are irregularly recorded.

MIMIC-III Dataset

The MIMIC-III dataset is collected from the intensive care units (ICU) Beth Israel Deaconess Medical Center from 2001 to 2012. Different from the PUB dataset, the MIMIC-III dataset contains 753 kinds of laboratory parameters. Since patients normally only stay in ICU for only a few days, patients in this dataset have short visits. For the MIMIC-III dataset, we select all the patients aged 18 years or above and get a cohort of 38,549 adult patients.

4.2 Experimental Settings

In this subsection, we first introduce the arrangement of the observation window and the prediction window. Then we discuss the mortality risk prediction levels for the PUB dataset and the MIMIC-III dataset. After that, comparing methods are introduced. Finally, we describe the implementation details.

Observation Window and Prediction Window

Figure 3 demonstrates the relationship between the observation window, the prediction window, and clinical outcome diagnosis date. We use features in the observation window as input to build models and predict the clinical outcome at the diagnosis date ahead length of the prediction window. After getting a patient's mortality risk level, we can feed this important information back to doctors to help evaluate early treatment effects. For patients with a high probability of an unfavorable outcome, a more aggressive and effective treatment plan could be taken to improve the final healthcare outcome.

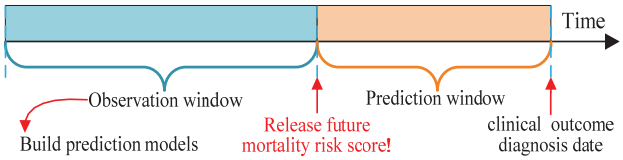
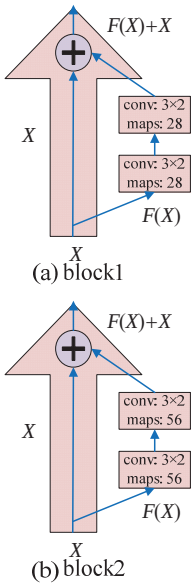


Figure 3. The relation between the observation window, the prediction window, and clinical outcome diagnosis date.



Type	Configurations
Input	Temporal shape: $T\times N$ Uncertainty shape: $T\times N$
UADL	output shape: $T\times N\times 3$
Conv	#maps: 28, k: 3×2 s: 1, p: same
Max pooling	k: 3×2, s: 2, p: same
block 1	#maps: 28
block 2	#maps: 56
Average Pooling	#maps: 56, s: 1 output shape: $T\times 1\times 1$
Flatten	output shape: $T\times 56$
GRU	#hidden units: 56
GRU	#hidden units: 28
GRU	#hidden units: 14
Dense	#units: 2, softmax

(c) Network configuration summary

Figure 4. The network configuration of UA-CRNN.

Prediction Levels

Patients in the PUB dataset have years of records, thus making it possible for models to learn long-term patterns of change of the disease and predict the mortality risk of patients ahead a long time. Therefore, for the PUB dataset, we perform long-term mortality risk prediction. Specifically, we predict whether a patient dies within 10 years after admission ahead six different time scales, namely, from 0.5 year to 3 years with an interval of 0.5 year. We use all the seven types of laboratory parameters mentioned above as input data.

However, for the MIMIC-III dataset, the records are short and thus can only be used to learn the changes in a patient's health status over a short time. Therefore, we conduct in-hospital mortality risk prediction, i.e., predicting whether a patient dies during the treatment in the hospital, and short-term mortality risk prediction, i.e., predicting whether a patient dies a few days after the observation window, i.e., 10 days, 20 days, 30 days, 60 days and 90 days. We select the 20 most frequent laboratory parameters as input and set the observation window as 10 days.

Comparing Methods

To evaluate the prediction performance of the proposed UA-CRNN method, we compare it with the following methods:

Table 1. The AUC scores (95% confidence interval) of long-term mortality risk prediction results for the PUB dataset.

Prediction level	Logistic Regression	Random Forests	Phased LSTM	Independent RNN	GRU Temporal	GRU Uncertainty	ResNet Temporal	ResNet Uncertainty	UA-CRNN
0.5 year	0.817 (0.798-0.837)	0.840 (0.823-0.856)	0.858 (0.841-0.875)	0.851 (0.831-0.869)	0.857 (0.841-0.874)	0.907 (0.892-0.921)	0.891 (0.876-0.905)	0.917 (0.903-0.930)	**0.944 (0.933-0.954)**
1 year	0.810 (0.790-0.829)	0.829 (0.811-0.848)	0.845 (0.827-0.862)	0.850 (0.832-0.868)	0.852 (0.835-0.870)	0.891 (0.875-0.907)	0.877 (0.861-0.892)	0.894 (0.880-0.908)	**0.935 (0.924-0.947)**
1.5 years	0.801 (0.780-0.819)	0.817 (0.799-0.836)	0.830 (0.811-0.848)	0.830 (0.811-0.848)	0.842 (0.824-0.860)	0.872 (0.856-0.889)	0.862 (0.847-0.878)	0.881 (0.866-0.896)	**0.923 (0.911-0.935)**
2 years	0.794 (0.772-0.815)	0.816 (0.799-0.835)	0.827 (0.808-0.844)	0.826 (0.807-0.845)	0.835 (0.816-0.853)	0.854 (0.837-0.871)	0.849 (0.832-0.867)	0.831 (0.812-0.849)	**0.898 (0.884-0.912)**
2.5 years	0.793 (0.772-0.813)	0.810 (0.792-0.830)	0.831 (0.813-0.849)	0.819 (0.799-0.838)	0.833 (0.813-0.852)	0.831 (0.814-0.850)	0.831 (0.812-0.851)	0.803 (0.783-0.824)	**0.879 (0.863-0.894)**
3 years	0.781 (0.760-0.802)	0.787 (0.766-0.807)	0.814 (0.795-0.832)	0.804 (0.785-0.824)	0.821 (0.803-0.840)	0.820 (0.802-0.839)	0.812 (0.791-0.832)	0.793 (0.773-0.814)	**0.865 (0.848-0.881)**

As baselines, Logistic Regression (LR) and Random Forests (RF) are used to predict mortality risk by using temporal information (temporal data is flattened for LR and RF). We also used LR and RF to build models for mean values of time series data but found that the results are not as good as that of using temporal data. Therefore, we use them to model temporal information.

To our best knowledge, this paper is the first attempt to integrate uncertainty caused by the processing of irregular data into deep learning architecture to assign attention to generated data according to their reliability. Existing methods cannot jointly model uncertainty and temporal information. Therefore, we compare it with two state-of-the-art sequential data modeling methods, i.e., Phased LSTM (PLSTM) [23] and Independent Recurrent Neural Network (IndRNN) [24]. By adding a time gate controlled via a parametrized oscillation, PLSTM convergences faster than regular LSTMs for long sequences [23]. Different from traditional RNNs, neurons in the same layer of IndRNN are independent of each other and connected across different layers, thus avoiding the gradient vanishing problem. They both show a strong modeling capacity for many sequential data modeling tasks. Therefore, we use them to model temporal information.

Finally, we utilize ResNet and GRU as another two comparing methods, which have the same structure and parameter setting with the convolutional layers and the recurrent layers of UA-CRNN respectively and are followed by a dense layer. Both ResNet and GRU are utilized to model temporal information and uncertainty information respectively.

Implementation Details

For both datasets, we randomly choose 70% of patients as the training set and use rest patients as the test set. Experimental results are evaluated by using receiver operator characteristic (ROC) curves and the area under ROC curves (AUC) score.

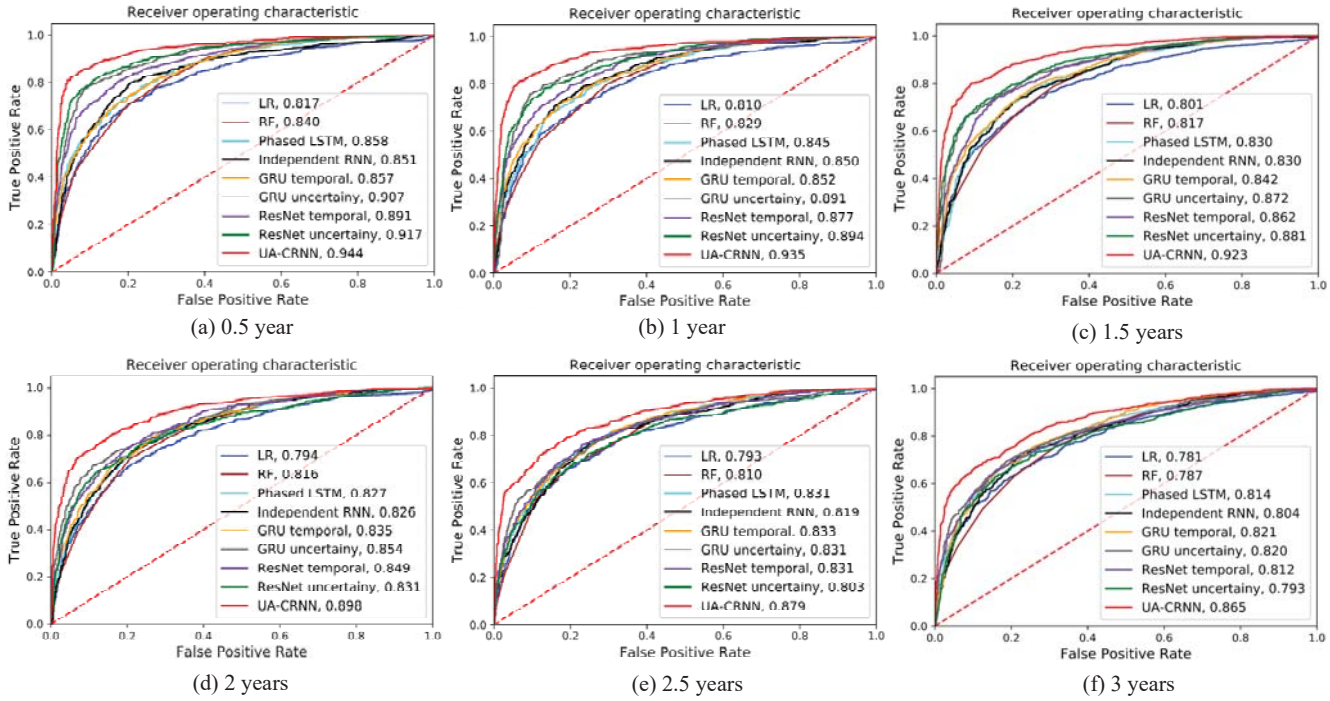
Figure 4 gives the network configuration of UA-CRNN. Convolutional layers are based on ResNet with 3×2 convolution units. The number of maps for residual learning block 1 and block 2 is set as 28 and 56 respectively. Recurrent layers consist of three-layer GRU with hidden units of 56, 28 and 14.

We implement the proposed UA-CRNN method, GRU, ResNet, PLSTM and IndRNN in Keras platform with Tensorflow backend. All these methods are trained for 100 epochs with a batch size of 512 and by using Adam as the optimizer. The baselines LR and RF are implemented with Scikit-learn.

4.3 Results

Table 1 and Table 2 demonstrate AUC scores (95% confidence interval) of long-term and short-term mortality risk prediction results for patients in the PUB dataset and the MIMIC-III dataset respectively. Corresponding ROC curves for both datasets are given in Figure 5 and Figure 6. From these experimental results, we can draw the conclusion that the proposed UA-CRNN method achieves the best results for both datasets. Specifically, UA-CRNN obtains the largest AUC scores at all the experiments. E.g., for the PUB dataset, at the 0.5 year prediction scale, the AUC score of UA-CRNN is 0.944, which is much larger than the AUC scores of other methods, such as 0.851 achieved by IndRNN and 0.858 achieved by PLSTM. For the MIMIC-III dataset, UA-CRNN achieves the AUC score of 0.898 for the in-hospital mortality risk prediction, which is much larger than the best result of comparing methods, namely 0.845 achieved by ResNet (Temporal). Correspondingly, we can see that the ROC curves of the proposed UA-CRNN method lie completely above ROC curves of other methods, which further demonstrates that UA-CRNN has the strongest prediction capacity. There are several possible reasons that enable UA-CRNN to outperform other methods. Firstly, UA-CRNN incorporates uncertainty information into deep learning architecture to ensure reliable data in the generated time series receive larger attention. Furthermore, this uncertainty-aware mechanism is further introduced into the sub-series level by designing a novel UADL to decompose time series into several sub-series and adjust weights assigned to them to ensure reliable sub-series receive large weight. Furthermore, parameters in UADL are optimized jointly with rest parts of UA-CRNN in an end-to-end trainable way, ensuring that obtained sub-series collaborate well with other layers. Besides this, several other conclusions can be drawn.

Firstly, there is an interesting phenomenon that modeling uncertainty information sometimes can even result in better results than modeling temporal information. As shown in Table 1, for the PUB dataset, GRU achieves more accurate results when using uncertainty information to predict the mortality risk of patients from the level of 0.5 year ahead all the way to the level of 2 years ahead than using temporal information. Similarly, ResNet also obtains better results when modeling uncertainty information from the 0.5 year level all the way to the 1.5 years level for the PUB dataset. This is probably because uncertainty



(a) 0.5 year (b) 1 year (c) 1.5 years

(d) 2 years (e) 2.5 years (f) 3 years

Figure 5. The ROC curves of long-term mortality risk prediction results for the PUB dataset.

information can extract valuable information about the health status of patients. Specifically, uncertainty scores are large at periods with sparse observations and are small at periods with intensive observations, as shown in Figure 1. Furthermore, the density of observation points can reflect the doctor's judgment on the health condition of the patient. For example, for a patient with a good status, it is probable that the doctor will arrange the patient to undertake fewer examinations to reduce the patient's pain and expense. However, for a patient in severe health conditions, intensive examinations are necessary so that doctors can make effective and timely treatments according to changes in the patient's conditions. As a result, uncertainty scores at different periods form a time series, which records dynamic changes in the patient's health status and sometimes can even be more informative than the temporal information.

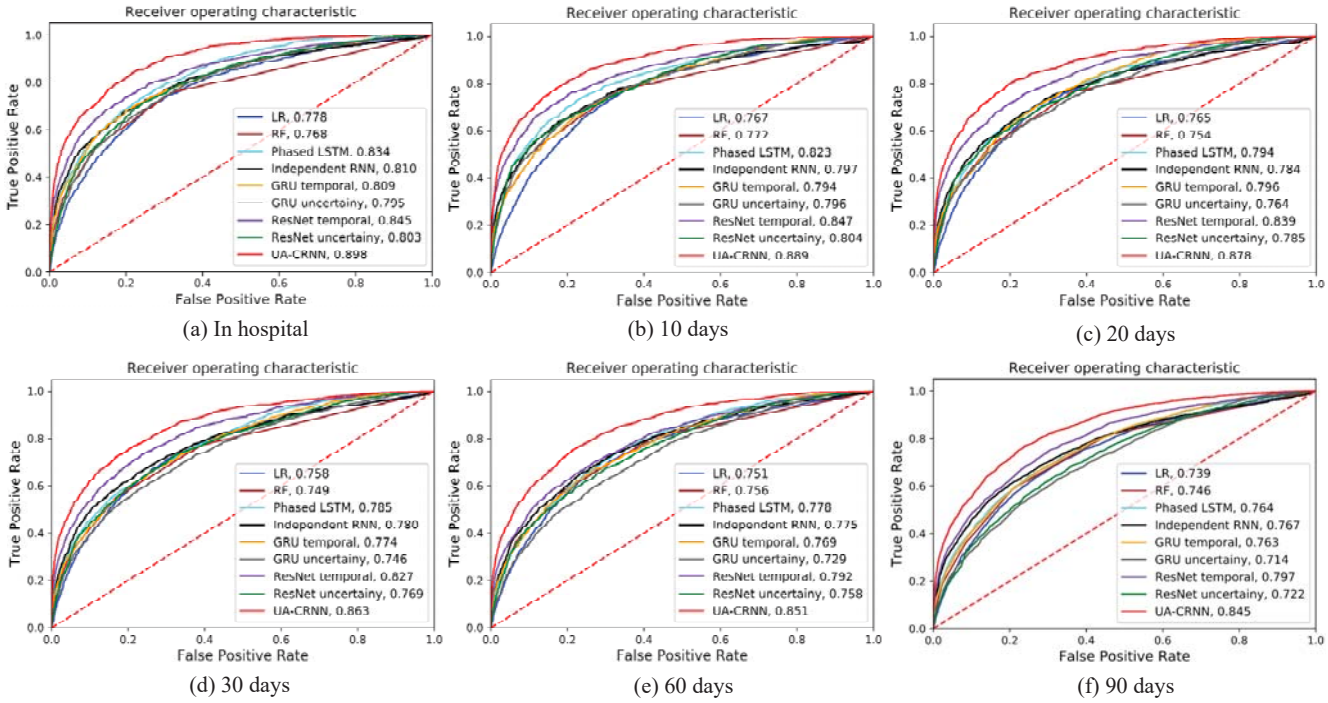
Secondly, ResNet obtains competitive or even better prediction results compared with GRU. E.g., for the PUB dataset, it can be seen from Table 1 that when modeling temporal information data, ResNet obtains more accurate prediction results than GRU from the 0.5 year ahead prediction level all the way to the 2 years prediction level. Similarly, for the MIMIC-III dataset, as shown in Table 2, when using temporal information to predict the mortality risk of patients, ResNet consistently outperforms GRU regardless of the prediction levels. These results demonstrate that when using multiple time series data to predict the clinical outcome of patients, we should not only use sequential data analysis methods, such as GRU, to capture temporal dependencies information, but also should use correlation analysis models, such as ResNet, to extract correlation information between different time series features.

Thirdly, variations of RNN obtain similar results, but all outperform LR and RF. E.g., for the MIMIC-III dataset, as shown in Table 2, at the in-hospital and the 10 days ahead prediction levels, PLSTM obtains slightly larger AUC scores than IndRNN and GRU. Then, GRU becomes slightly better than IndRNN and PLSTM at the 20 days prediction level. After this, PLSTM becomes the best among these three methods with a slight advantage at the 30 days and 60 days ahead prediction levels. Finally, for the 90 days ahead prediction level, IndRNN outperforms PLSTM and GRU slightly. In addition, the ROC curves of these three methods, as shown in Figure 5 and Figure 6, are close to each other, which further supports this conclusion. This is probably because temporal information about dynamic changes in the patient's health condition contained in the generated regular time series is straightforward and easy to capture. As a result, though these variations of RNN have different advantages, they all can capture temporal dependencies information from the generated time series data. However, it can be seen that the prediction accuracy of these methods is not high. This is probably because the amount of useful information in the generated time series is limited. Furthermore, these methods give equal weight to different data points, failing to highlight the importance of reliable data points. Even so, it can be seen that these methods still outperform LR and RF on a large scale, which proves deep learning models' advantages over traditional methods.

Finally, the performance of different models tends to decrease with the increase of the length of the prediction window. For the PUB dataset, as shown in Table 1, when the prediction level increases from 0.5 year to 3 years with the time interval of 0.5 year, the AUC scores of all these models drop steadily. Similarly, when

Table 2. The AUC scores (95% confidence interval) of short-term mortality risk prediction results for the MIMIC-III dataset.

Prediction level	Logistic Regression	Random Forests	Phased LSTM	Independent RNN	GRU Temporal	GRU Uncertainty	ResNet Temporal	ResNet Uncertainty	UA-CRNN
In hospital	0.778 (0.766-0.791)	0.768 (0.754-0.784)	0.834 (0.823-0.845)	0.810 (0.797-0.824)	0.809 (0.796-0.823)	0.795 (0.782-0.807)	0.845 (0.832-0.857)	0.803 (0.791-0.816)	**0.898 (0.889-0.906)**
10 days	0.767 (0.752-0.780)	0.772 (0.757-0.785)	0.823 (0.809-0.836)	0.797 (0.781-0.811)	0.794 (0.780-0.808)	0.796 (0.783-0.810)	0.847 (0.835-0.860)	0.804 (0.790-0.816)	**0.889 (0.880-0.899)**
20 days	0.765 (0.751-0.778)	0.754 (0.741-0.769)	0.794 (0.781-0.807)	0.784 (0.771-0.798)	0.796 (0.783-0.808)	0.764 (0.751-0.777)	0.839 (0.827-0.851)	0.785 (0.772-0.798)	**0.878 (0.868-0.888)**
30 days	0.758 (0.745-0.771)	0.749 (0.736-0.762)	0.785 (0.774-0.797)	0.780 (0.766-0.793)	0.774 (0.762-0.787)	0.746 (0.733-0.759)	0.827 (0.816-0.838)	0.769 (0.757-0.781)	**0.863 (0.853-0.872)**
60 days	0.751 (0.738-0.763)	0.756 (0.743-0.767)	0.778 (0.767-0.789)	0.775 (0.764-0.787)	0.769 (0.758-0.781)	0.729 (0.717-0.741)	0.792 (0.780-0.803)	0.758 (0.746-0.770)	**0.851 (0.842-0.860)**
90 days	0.739 (0.727-0.750)	0.746 (0.733-0.758)	0.764 (0.754-0.777)	0.767 (0.754-0.779)	0.763 (0.752-0.775)	0.714 (0.702-0.726)	0.797 (0.786-0.808)	0.722 (0.709-0.733)	**0.845 (0.836-0.854)**



(a) In hospital (b) 10 days (c) 20 days

(d) 30 days (e) 60 days (f) 90 days

Figure 6. The ROC curves of short-term mortality risk prediction results for the MIMIC-III dataset.

conducting short-term prediction for the MIMIC-III dataset, most models achieve better results for the more recent prediction levels. These results demonstrate that the difficulty in performing precise prediction is in coincidence with how long into future the prediction is made. One possible reason is that many unknown factors could occur in the long future and make the trajectory of the patient's health state full of uncertainty. Nevertheless, based on the observed visit records, our proposed UA-CRNN method can effectively consider the uncertainty caused by the processing of irregular time series data, which is a necessary process for diagnosis prediction due to the irregular sampling problem, and improve final prediction results.

5 Conclusion

In this paper, we propose a novel UA-CRNN model for mortality risk prediction. The proposed method analyzes the uncertainty of generated regular data and assigns more attention to reliable data by incorporating uncertainty information into a deep architecture. Furthermore, since the degree of uncertainty caused by sub-series of different frequencies of clinical time series data is different, we further introduce this uncertain-aware mechanism into the more precise sub-series level by designing a novel uncertainty-aware decomposition layer (UADL) to decompose time series and assign the obtained sub-series with appropriate weights according to their reliabilities. To evaluate the prediction performance of the proposed UA-CRNN method for different prediction levels, we conduct experiments on two real-world EHRs datasets, i.e., the MIMIC-III dataset with short visit records and the PUB dataset with long visit records. Extensive experiment results demonstrate that the proposed UA-CRNN method significantly outperforms existing methods in both short-term and long-term mortality risk prediction tasks.

REFERENCES

[1] Baytas, I. M., Xiao, C., Zhang, X., Wang, F., Jain, A. K., and Zhou, J. Patient subtyping via time-aware LSTM networks. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, 65-74, 2017.

[2] Liu, Z., and Hauskrecht, M. A Personalized Predictive Framework for Multivariate Clinical Time Series via Adaptive Model Selection. In *Proceedings of the 26th ACM on Conference on Information and Knowledge Management*, 1169-1177, 2017.

[3] Elisha O and Dekel S. Wavelet decompositions of Random Forests: smoothness analysis, sparse approximation and applications. *The Journal of Machine Learning Research*, 17(1): 6952-6989, 2016.

[4] Shickel, B., Tighe, P. J., Bihorac, A., and Rashidi, P. Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. *IEEE journal of biomedical and health informatics*, 22(5), 1589-1604, 2018.

[5] Cai, X., Gao, J., Ngiam, K. Y., Ooi, B. C., Zhang, Y., and Yuan, X. Medical concept embedding with time-aware attention. *arXiv preprint arXiv*:1806.02873, 2018.

[6] Yip, T. F., Ma, A. J., Wong, V. S., Tse, Y. K., Chan, H. Y., Yuen, P. C., and Wong, G. H. Laboratory parameter-based machine learning model for excluding non-alcoholic fatty liver disease (NAFLD) in the general population. *Alimentary pharmacology & therapeutics*, 46(4), 447-456, 2017.

[7] Yadav, P., Steinbach, M., Kumar, V., and Simon, G. Mining electronic health records (EHRs): a survey. *ACM Computing Surveys (CSUR)*, 50(6), 85, 2018.

[8] Cao, Z., and Lin, C. T. Inherent fuzzy entropy for the improvement of EEG complexity evaluation. *IEEE Transactions on Fuzzy Systems*, 26(2), 1032-1035, 2018.

[9] Naik, G. R., Selvan, S. E., and Nguyen, H. T. Single-channel EMG classification with ensemble-empirical-mode-decomposition-based ICA for diagnosing neuromuscular disorders. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 24(7), 734-743, 2016.

[10] Ren, P., Tang, S., Fang, F., Luo, L., Xu, L., Bringas-Vega, M. L., Yao, D., Kendrick, K. M. and Valdes-Sosa, P. A. Gait rhythm fluctuation analysis for neurodegenerative diseases by empirical mode decomposition. *IEEE Transactions on Biomedical Engineering*, 64(1), 52-60, 2017.

[11] Wu, A., Roy, N. G., Keeley, S., and Pillow, J. W. Gaussian process based nonlinear latent structure discovery in multivariate spike train data. In *Proceedings of Advances in neural information processing systems*, 3496-3505, 2017.

[12] He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770-778, 2016.

[13] Cho, K., Van Merriënboer, B., Bahdanau, D., and Bengio, Y. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv*:1409.1259., 2014.

[14] Che, Z., Purushotham, S., Cho, K., Sontag, D., and Liu, Y. Recurrent neural networks for multivariate time series with missing values. *Scientific reports*, 8(1), 6085, 2018.

[15] Wu, M., Ghassemi, M., Feng, M., Celi, L. A., Szolovits, P., and Doshi-Velez, F. Understanding vasopressor intervention and weaning: Risk prediction in a public heterogeneous clinical time series database. *Journal of the American Medical Informatics Association*, 24(3), 488-495, 2017.

[16] Sherman, E., Gurm, H., Balis, U., Owens, S., and Wiens, J. Leveraging Clinical Time-Series Data for Prediction: A Cautionary Tale. In *Proceedings of American Medical Informatics Association Annual Symposium*, 1571-1580, 2017.

[17] Liu, L., Shen, J., Zhang, M., Wang, Z., and Tang, J. Learning the Joint Representation of Heterogeneous Temporal Events for Clinical Endpoint Prediction. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[18] Che, C., Xiao, C., Liang, J., Jin, B., Zho, J., and Wang, F. An RNN Architecture with Dynamic Temporal Matching for Personalized Predictions of Parkinson's Disease. In *Proceedings of the 2017 Society for Industrial and Applied Mathematics International Conference on Data Mining*, 198-206, 2017.

[19] Harrison, E., Chang, M., Hao, Y., & Flower, A. Using machine learning to predict near-term mortality in cirrhosis patients hospitalized at the University of Virginia health system. In *IEEE Systems and Information Engineering Design Symposium*, 2018.

[20] Zhang, X., Henao, R., Gan, Z., Li, Y., and Carin, L. Multi-Label Learning from Medical Plain Text with Convolutional Residual Models. *arXiv preprint arXiv*:1801.05062., 2018.

[21] Mei, J., Liu, M., Wang, Y. F., and Gao, H. Learning a mahalanobis distance-based dynamic time warping measure for multivariate time series classification. *IEEE transactions on Cybernetics*, 46(6), 1363-1374, 2016.

[22] Wang, Z., Song, W., Liu, L., Zhang, F., Xue, J., Ye, Y. and Xu, M. Representation Learning with Deconvolution for Multivariate Time Series Classification and Visualization. *arXiv preprint arXiv*:1610.07258, 2016.

[23] Neil, D., Pfeiffer, M., and Liu, S. C. Phased lstm: Accelerating recurrent network training for long or event-based sequences. In *Proceedings of Advances in neural information processing systems*, 3882-3890, 2016.

[24] Li, S., Li, W., Cook, C., Zhu, C., and Gao, Y. Independently recurrent neural network (indrnn): Building A longer and deeper RNN. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 5457-5466, 2018.

[25] Johnson, A. E., Pollard, T. J., Shen, L., Li-wei, H. L., Feng, M., Ghassemi, M., Moody B., and Mark, R. G. MIMIC-III, a freely accessible critical care database. *Scientific data*, 3, 160035, 2016.

[26] Xu, Y., Biswal, S., Deshpande, S. R., Maher, K. O., and Sun, J. RAIM: Recurrent Attentive and Intensive Model of Multimodal Patient Monitoring Data. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, 2565-2573, 2018.

[27] Tan, Q., Ma, A. J., Deng, H., Wong, V. W. S., Tse, Y. K., Yip, T. C. F., Wong, G. L., Ching, J. Y., Chan, F. K and Yuen, P. C. A Hybrid Residual Network and Long Short-Term Memory Method for Peptic Ulcer Bleeding Mortality Prediction. In *Proceedings of American Medical Informatics Association Annual Symposium*, 998-1007, 2018.

[28] Bhattacharyya, A., Fritz, M., and Schiele, B. Long-term on-board prediction of people in traffic scenes under uncertainty. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4194-4202, 2018.

[29] Aizpurua, J. I., McArthur, S. D., Stewart, B. G., Lambert, B., Cross, J. G., and Catterson, V. M. Adaptive Power Transformer Lifetime Predictions Through Machine Learning and Uncertainty Modeling in Nuclear Power Plants. *IEEE Transactions on Industrial Electronics*, 66(6): 4726-4737, 2019.

[30] Heo, J., Lee, H. B., Kim, S., Lee, J., Kim, K. J., Yang, E., and Hwang, S. J. Uncertainty-aware attention for reliable interpretation and prediction. In *Proceedings of Advances in neural information processing systems*, 909-918, 2018.

[31] Sakaridis C, Dai D, Van Gool L. Semantic Nighttime Image Segmentation with Synthetic Stylized Data, Gradual Adaptation and Uncertainty-Aware Evaluation. *arXiv preprint arXiv*:1901.05946, 2019.

[32] Sensoy, M., Kaplan, L., and Kandemir, M. Evidential deep learning to quantify classification uncertainty. In *Proceedings of Advances in neural information processing systems*, 3179-3189, 2018.

[33] Lütjens, B, Everett, M, How, J. P. Safe reinforcement learning with model uncertainty estimates. *arXiv preprint arXiv*:1810.08700, 2018.

[34] Ma, F., You, Q., Xiao, H., Chitta, R., Zhou, J., and Gao, J. Kame: Knowledge-based attention model for diagnosis prediction in healthcare. In *Proceedings of the 27th ACM on Conference on Information and Knowledge Management*, 743-752, 2018.

[35] Hosseini, A., Chen, T., Wu, W., Sun, Y., and Sarrafzadeh, M. HeteroMed: Heterogeneous Information Network for Medical Diagnosis. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, 763-772, 2018.

[36] Pearce, T., Zaki, M., Brintrup, A., and Neel, A. Uncertainty in neural networks: Bayesian ensembling. *arXiv preprint arXiv*:1810.05546, 2018.

[37] Kendall, A., and Gal, Y. What uncertainties do we need in bayesian deep learning for computer vision? In *Advances in neural information processing systems*, 5574-5584. 2017.

[38] Ye, M., Lan, X., Wang, Z., and Yuen, P. C. Bi-directional Center-Constrained Top-Ranking for Visible Thermal Person Re-Identification. *IEEE Transactions on Information Forensics and Security*, 2019.

[39] Ye, M., Zhang, X., Yuen, P. C., and Chang, S. F.. Unsupervised Embedding Learning via Invariant and Spreading Instance Feature. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 6210-6219, 2019.

[40] Yang, B., Ma, A. J., and Yuen, P. C. Learning domain-shared group-sparse representation for unsupervised domain adaptation. *Pattern Recognition*, 81, 615-632, 2018.

[41] Yang, B., and Yuen, P. C. Cross-Domain Visual Representations via Unsupervised Graph Alignment. In *Thirty-Three AAAI Conference on Artificial Intelligence*, 5613-5620, 2019.

[42] Ye, M., Li, J., Ma, A. J., Zheng, L., and Yuen, P. C. Dynamic graph co-matching for unsupervised video-based person re-identification. *IEEE Transactions on Image Processing*, 28(6), 2976-2990, 2019.

[43] Yang, B., Ma, A. J., and Yuen, P. C. Domain-shared group-sparse dictionary learning for unsupervised domain adaptation. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 7453-7460, 2018.

[44] Ye, M., Lan, X., and Yuen, P. C. Robust anchor embedding for unsupervised video person re-identification in the wild. In *Proceedings of the European Conference on Computer Vision*, 170-186, 2018.

[45] Pang, M., Cheung, Y. M., Liu, R., Lou, J., and Lin, C. Toward efficient image representation: Sparse concept discriminant matrix factorization. *IEEE Transactions on Circuits and Systems for Video Technology*, 2018.

[46] Luo, Y., Cai, X., Zhang, Y., and Xu, J. Multivariate time series imputation with generative adversarial networks. In *Advances in Neural Information Processing Systems*, 1596-1607, 2018.

[47] Gal, Y., and Ghahramani, Z. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, 1050-1059, 2016.