TE-ESN: Time Encoding Echo State Network for Prediction Based on Irregularly Sampled Time Series Data

Chenxi Sun
sun_chenxi@pku.edu.cn
Peking University
Beijing, China

Hongyan Li
leehy@pku.edu.cn
Peking University
Beijing, China

ABSTRACT

Prediction based on Irregularly Sampled Time Series (ISTS) is of wide concern in real-world applications. Different from the ordinary time series, ISTS is characterized by irregular time intervals of intra-series and different sampling rates of inter-series. However, existing methods have suboptimal predictions due to artificially introducing new dependencies in a time series and biasedly learning relations among time series when modeling these two characteristics. In this work, we propose a novel Time Encoding (TE) mechanism. TE can embed the time information as time vectors in the complex domain. It has the properties of absolute distance and relative distance under different sampling rates, which helps to represent two irregularities. Meanwhile, we create a new model named Time Encoding Echo State Network (TE-ESN). It is the first ESNs-based model that can process ISTS data. Experiments on one chaos system and three real-world datasets show that TE-ESN performs better than all baselines and has better reservoir property.

CCS CONCEPTS

• **Computing methodologies** → **Neural networks**.

KEYWORDS

irregularly sampled time series, echo state networks, time encoding.

ACM Reference Format:
Chenxi Sun and Hongyan Li. 2021. TE-ESN: Time Encoding Echo State Network for Prediction Based on Irregularly Sampled Time Series Data. In *MileTS '21: 7th KDD Workshop on Mining and Learning from Time Series, August 14th, 2021, Singapore.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

Time series prediction widely exists in many scenarios, such as healthcare and meteorology [20, 21]. However, in real-world applications, it's usually Irregularly Sampled Time Series (ISTS), having two irregularities under the aspects of intra-series and inter-series:

- Intra-series irregularity is the irregular time intervals between observations within a time series. In Figure 1, the time between a COVID-19 patient's blood sample could be 1 hour or even 7 days. Uneven intervals changes the dependency between observations and large intervals add time sparsity factor [11].
- Inter-series irregularity is the different sampling rates among time series. For a COVID-19 patient in Figure 1, heart rate is measured in seconds, while blood sample is collected in days. The difference of sampling rates is not conducive to data preprocessing and model design [12].

However, grasping both two irregularities is challenging. In real-world applications, a model usually has multiple time series as input. If seeing the input as a multivariate time series, data alignment with up/down-sampling and imputation occur. But it will artificially introduce some new dependencies while omit some original dependencies, causing suboptimal prediction [16]; If seeing the input as multiple separated time series and modeling them separately, it will encounter the problem of bias, embedding stronger dependency in high sampled time series due to smaller time intervals. This is not necessarily the case, for example, although the detection of blood pressure is not frequent than heart rate in clinical practice, its values have a strong diurnal correlation [18].

In order to get rid of the above dilemmas, modeling all irregularities without introducing new dependency is feasible [8]. The premise is that ISTS can't be interpolated, which makes the alignment impossible, leading to batch gradient descent for multivariate time series hard to implement, aggravating the non-converging and instability of error Back Propagation Recurrent Neural Networks (BPRNNs) [5], the basis of existing methods [16]. Echo State Networks (ESNs) is a simple type of RNNs and can avoid non-converging and computationally expensive by applying least square problem as the alternative training method [9]. But ESNs can only process uniform TS by assuming time intervals are equally distributed, with no mechanism to model ISTS. For solving all the difficulties mentioned above, we design a new structure to enable ESNs to handle ISTS data, where a novel mechanism makes up for the disadvantage of no learning of irregularity.

- We introduce a novel mechanism named Time Encoding (TE). TE represents time points as dense vectors in complex domain. It injects the absolute and relative distance properties to model both intra-series irregularity and inter-series irregularity of ISTS.
- We design a model named Time Encoding Encoding Echo State Network (TE-ESN). In addition to the ability of modeling both two ISTS irregularities, TE-ESN can learn the long short-term memories longitudinally and fuses the relations horizontally.
- We evaluate TE-ESN for early prediction and one-step-ahead forecasting on four datasets. TE-ESN outperforms state-of-the-art models and has better reservoir property.

ACM ISBN 978-1-4503-9999-9/18/06…$15.00
https://doi.org/10.1145/1122445.1122456

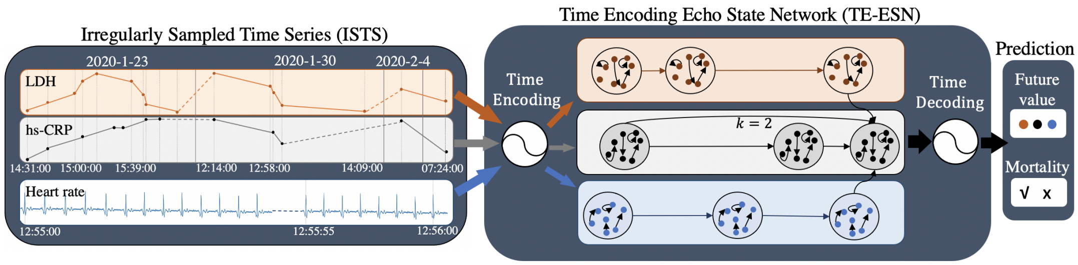

Figure 1: An ISTS example of a COVID-19 patient and the structure of TE-ESN

2 RELATED WORK

Existing methods can be divided into two categories:

Missing data perspective. It discretizes the time axis into non-overlapping intervals, points without data are considered as missing data. M-RNN [11] handled missing data by operating time series forward and backward. GRU-D [3] used decay rate to weigh the correlation between missing data and other data. But data imputation may artificially introduce new dependency beyond original relations and totally ignore ISTS irregularities.

Raw data perspective. It constructs models which can directly receive ISTS as input. T-LSTM [1] used the elapsed time function for modeling irregular time intervals. IPN [15] used three time perspectives for modeling different sampling rates. However, they just performed well in the univariate time series, for multiple time series, they had to apply alignment first, causing the data missing in some time points, back to the defects of the first category.

3 TIME ENCODING ECHO STATE NETWORK

3.1 Definitions

Definition 1 (Irregularly Sampled Time Series ISTS). *A time series u with sampling rate $r_s(d), d \in= \{1, ..., D\}$ has several observations distributed with time $t, t \in= \{1, ..., T\}$. u_t^d represents an observation of a time series with sampling rate $r_s(d)$ in time t.*

ISTS has two irregularities: (1) Irregular time intervals of intra-series: $t_i - t_{i-1} \neq t_j - t_{j-1}$. (2) Different sampling rate of inter-series: $r_s(d_i) \neq r_s(d_j)$. For prediction tasks, one-step-ahead forecasting is using the observed data $u_{1:t}$ to predict the value of u_{t+1}, and continues over time; Early prediction is using the observed data $u_{1:t}$ $(t < t_{pre})$ to predict the classes or values in time t_{pre}.

Definition 2 (Time Encoding TE). *Time encoding mechanism aims to design methods to embed and represent every time point information of a time line.*

TE mechanism extends the idea of Positional Encoding (PE) [7, 17] in natural language processing, shown in Equation 1. Where pos indicates the position of a word, d_{model} is the embedding dimension.

$$\begin{cases} PE(pos, 2i) = \sin\left(\frac{pos}{10000^{\frac{2i}{d_{model}}}}\right) \\ PE(pos, 2i+1) = \cos\left(\frac{pos}{10000^{\frac{2i}{d_{model}}}}\right) \end{cases} \tag{1}$$

3.2 Time Encoding Mechanism

3.2.1 Time Vector with Fixed Sampling Rate. Only consider one time series, whose irregularity is just reflected in the irregular time intervals, we apply Time Vector (TV) to note the time as vector. Each time vector has d_{TV} embedding dimensions. Each dimension corresponds to a sinusoidal wave, forming a geometric progression from 2π to $MT\pi$. MT is the maximum number of input time points.

$$\begin{aligned} TV(t) &= [..., \sin(c_i t), \cos(c_i t), ...] \\ c_i &= MT^{-\frac{2i}{d_{TV}}}, i = 0, ..., \frac{d_{TV}}{2} - 1 \end{aligned} \tag{2}$$

TV can simulate the time intervals between two observations by its properties of absolute distance and relative distance.

Property 1 (Absolute Distance Property). *For two time points with distance p, the time vector in time point $t + p$ is the linear combination of the time vector in time point t.*

$$\begin{aligned} TV(t+p) &= (a,b) \cdot TV(t) \\ a = TV(p, 2i+1)&, b = TV(p, 2i) \end{aligned} \tag{3}$$

Property 2 (Relative Distance Property). *The product of time vectors of two time points t and $t + p$ is negatively correlated with their distance p.*

$$TV(t) \cdot TV(t+p) = \sum_{i=0}^{\frac{d_{TV}}{2}-1} \cos(c_i p) \tag{4}$$

3.2.2 Time Encoding with Different Sampling Rates. When the input is multi-series, the irregularity of different sampling rates occurs. Then, TV will encounter the problem of bias, embedding more associations between observations with high sampling rate according to the Property 2. Thus, we design an advanced version of time vector, noted Time Encoding (TE), to encode time within multiple ISTS. TE extends TV to complex-valued domain. For a time point t in the d-th ISTS with $r_s(d)$ sampling rate, the time code is in Equation 5, where ω is the frequency.

$$TE(d, t) = e^{i(\omega t)}, \omega = \omega_d \cdot r_s^{-1}(d) \tag{5}$$

TE not only keeps the property 1 and 2 but also incorporates the influence of frequency ω, making time codes consistent at different sampling rates. ω reflects the sensitivity of observation to time, where a large ω leads to more frequent changes of time codes and more difference between the adjacent time points.

Property 3 (Relative Distance Property with ω). *The product of time encoding of two time points t and $t+p$ is positive correlated with the frequency ω.*

$$TE(t) \cdot TE(t+p) = e^{i\omega t} \cdot e^{i\omega(t+p)} = e^{i\omega(2t+p)} \tag{6}$$

In TE, we set $\omega = \omega_d \cdot r_s^{-1}(d)$. ω_d is the frequency parameter of d-th sampling rate. TE fuses the sampling rate term $r_s^{-1}(d)$ to avoid the bias of time vector causing by only considering the distance p. Each time point can be embeded into d_{TE} dimensions with more options of frequencies by setting different $\omega_{j,k}$ in Equation 7. $\omega_{j,k}$ means the time vector in dimension j has k frequencies.

$$\begin{aligned} &TE(d,t) = e^{i(\omega t)}, \omega = \omega_{j,k} \cdot r_s^{-1}(d) \\ &j = 0, ..., d_{TV} - 1, k = 0, ..., K - 1 \end{aligned} \tag{7}$$

3.2.3 The Relations between Different Mechanisms.

- TV is a special case of TE. If we set $\omega_{k,j} = c_i$, then $TE(d,t) = TV(t, 2i+1) + iTV(t, 2i)$.
- TE is a special case of $r \cdot e^{i \cdot (\omega x + \theta)}$. $\theta = 0$ as we focus more on the relation between different time points than the value of the first point; r is the representation of observations and leave it to learn by computing models. Besides, TE inherits the properties of position-free offset transformation and boundedness [19].

3.3 Injecting Time Encoding Mechanism into Echo State Network

ESN is a fast and efficient RNN. A typical ESN consists of an input layer $W_{in} \in R^{N \times D}$, a reservoir layer $W_{res} \in R^{N \times N}$ and an output layer $W_{out} \in R^{M \times N}$. $u(t) \in R^D$, $x(t) \in R^N$ and $y(t) \in R^M$ denote the input value, reservoir state and output value at time t. The existing ESNs-based methods cannot model the irregularities of ISTS. Thus, we make up for this by proposing Time Encoding Echo State Network (TE-ESN) as shown in Figure 1.

3.3.1 Time Encoding Phase. TE-ESN has D reservoirs, assigning each input time series an independent reservoir by:

$$\begin{aligned} \tilde{x}_t^d &= \gamma_f x_t^{d'} + (1-\gamma_f) x^{D \backslash d} && \textit{Reservoir} \\ x_t^{d'} &= \gamma_l x_t^d + (1-\gamma_l)(x_{t-1}^d + x_{t-k}^d) && \textit{Long short} \\ x_t^d &= \tanh(TE(d,t) + W_{in}^d u_t^d + W_{res}^d \tilde{x}_{t-1}^d) && \textit{Time encoding} \\ x^{D \backslash d} &= \frac{1}{D-1} \sum_{i \in D \backslash d} \tilde{x}^i && \textit{Neighbor} \end{aligned} \tag{8}$$

- Time encoding mechanism (TE). TE-ESN integrates time by changing reservoir state through TE term to *Time encoding state*.
- Long short-term memory mechanism (LS). TE-ESN incorporates long short term memories from former k time to *Long short state*.
- Series fusion (SF). TE-ESN considers the horizontal time series to change *Reservoir state* by the *Neighbor state* in other time series.

3.3.2 Time Decoding Phase. For final value prediction, TE-ESN decodes the time information and get the real estimated value at time t_{pre} by Equation 9. Further, by changing the time t_{pre}, we can get different prediction results in different time points.

$$y(t_{pre}) = W_{out}(\tilde{x}(t) - TE(t_{pre})) \tag{9}$$

$$\begin{aligned} &\min_{W_{out}} ||Y_{pre} - Y||_2^2 + \lambda ||W_{out}||_2^2 \\ &W_{out} = Y(\tilde{X} - TE)^T((\tilde{X} - TE)(\tilde{X} - TE)^T + \lambda I)^{-1} \end{aligned} \tag{10}$$

Algorithm 1 shows how TE-ESN predicts. Assuming reservoir size is N, maximum time is T, input has D data, the complexity is:

$$C = O(\alpha T N^2 + TND) \tag{11}$$

Algorithm 1 TE-ESN

Input: training input $X_{train} = \{x_t^d, t\}$, teacher signal $Y_{train} = \{y_t^d, t\}$,
test input $X_{test} = \{x_t^d, t\}$, maximum time MT,
leaky rate γ_l, fusion rate γ_f, time span k, regularization λ
input scale w_{in}, spectral radius $\rho(W_{res})$, sparsity α,
Output: Y_{pre}: prediction result.

Randomly initialized W_{in} in $[-w^{in}, w^{in}]$;
Randomly initialized W_{res} with α and $\rho(W_{res})$.
for $i = 1$ to $|U_{train}|$ **do**
  for $t = 1$ to MT **do**
    Compute $TE(d,t)$ by Equation 7
    Compute $\tilde{x}(t)$ by Equation 8
  end for
end for
$\tilde{X} = \{\tilde{x}(t)\}$
$TE = \{TE(t)\}$
Compute W_{out} by Equation 10
for $t = 1$ to T_{test} **do**
  Compute $TE_{test}(d,t)$ by Equation 7
  Compute $\tilde{x}_{test}(t)$ by Equation 8
end for
$\tilde{X}_{test} = \{\tilde{x}_{test}(t)\}$
$TE_{test} = \{TE_{test}(t)\}$
$Y_{pre} = W_{out}(\tilde{X}_{test} - TE_{test})$

4 EXPERIMENTS

4.1 Experiment Setting

Datasets

- *MG* [13] chaotic system. $y(t+1) = y(t) + \delta(a\frac{y(t-\frac{\tau}{\delta})}{1+y(t-\frac{\tau}{\delta})^n} - by(t))$. $\delta, a, b, n, \tau, y(0) = 0.1, 0.2, -0.1, 10, 17, 1.2$. t random increases.
- *SILSO* [2] provides open-source monthly sunspot series from 1749 to 2020. It has irregular time intervals, from 1 to 6 month.
- *USHCN* [14] has U.S. daily climate data from 1887 to 2014. Time intervals from 1 to 7 days. Sampling rates from 0.33 to 1 per day.
- *COVID-19* [23] contains blood samples with 80 features and 6877 records, from Tongji Hospital, Wuhan, China. Time intervals are from 1 minus to 12 days. Sampling rates are from 0 to 6 per day.

Baselines

- *BPRNNs-based*: There are 3 methods designed for ISTS data with BP training - M-RNN [11], T-LSTM [1] and GRU-D [3].
- *ESNs-based*: There are 4 methods designed based on ESNs - ESN [9], Leaky-ESN [10], DeepESN [6] and LS-ESN [24].

Metrics Use genetic algorithm [25] to optimize hyper-parameters in Table 5. Evaluate prediction by AUC-ROC and MSE. Evaluate network property by Memory Capability (MC) [4]: $MC = \sum_{k=0}^{\infty} r^2(u(t-k), y(t))$

4.2 Results

Prediction Results. TE-ESN outperforms all baselines on four datasets shown in Table 1 and Figure 2. TE-ESN is better than TV-ESN in multivariable time series (COVID-19, USHCN) shows the effect of Property 3; TE-ESN is better than TV-ESN in univariable time series (SILSO, MG) shows the advantage of multiple frequencies options of TE.

Time Encoding Mechanism Analysis. Dot product between two sinusoidal positional encoding decreases with increment of absolute value of distance [22]. Figure 4 and Table 3 show that using multiple frequencies will enhance monotonous of negative correlation and can improve the accuracy.

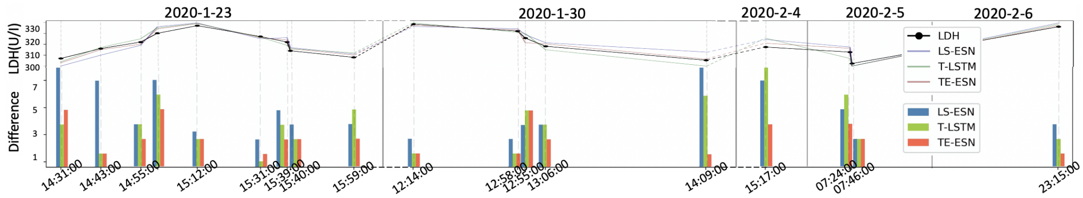

Figure 2: Lactic dehydrogenase (LDH) forecasting for a 70-year-old female COVID-19 patient

	BPRNNs-based			ESNs-based			Ours		
	M-RNN	T-LSTM	GRU-D	ESN	leaky-ESN	DeepESN	LS-ESN	TV-ESN	TE-ESN
MG	0.232±0.005	0.216±0.003	0.223±0.005	0.229±0.001	0.213±0.001	0.197±0.000	0.198±0.000	0.204±0.001	**0.195±0.001**
SILSO	2.950±0.740	2.930±0.810	2.990±0.690	3.070±0.630	2.950±0.590	2.800±0.730	2.540±0.690	2.540±0.790	**2.390±0.780**
USHCN	0.752±0.320	0.746±0.330	0.747±0.250	0.868±0.290	0.857±0.200	0.643±0.120	0.663±0.150	0.647±0.150	**0.640±0.190**
COVID-19	0.098±0.005	0.096±0.007	0.100±0.005	0.136±0.006	0.135±0.007	0.129±0.006	0.120±0.007	0.115±0.005	**0.093±0.005**
	0.959±0.004	0.963±0.003	0.963±0.004	0.941±0.003	0.942±0.003	0.948±0.003	0.949±0.003	0.958±0.002	**0.965±0.002**

Table 1: Prediction results of nine methods on four datasets (COVID-19 mortality in AUC-ROC; Others in MSE)

Parameters	Value range	Parameters	Value range
w^{in}, α, ρ	$(0, 1]$	γ_l, γ_f	$[0, 1]$
k	$\{2, 4, 6, 8, 10, 12\}$	λ	$\{10^{-4}, 10^{-2}, 1\}$

Table 2: Search settings of hyper-parameters

	$c_i, 32$	$c_i, 64$	$\omega_{d,i}, 32$	$\omega_{d,i}, 64$
MG	0.226±0.001	0.204±0.001	0.210±0.001	**0.193±0.001**
SILSO	2.690±0.600	2.540±0.79	2.550±0.750	**2.390±0.780**
USHCN	0.681±0.180	0.670±0.200	0.673±0.170	**0.640±0.190**
COVID-19	0.105±0.006	0.099±0.005	0.101±0.005	**0.093±0.005**
	0.949±0.002	0.952±0.003	0.950±0.002	**0.965±0.002**

Table 3: Prediction results of TE-ESN with different ω, d_{TE}

	w/o TE	w/o LS	w/o SF	TE-ESN
MG	0.210±0.001	0.213±0.001	0.193±0.001	**0.193±0.001**
SILSO	2.790±0.630	2.930±0.690	2.390±0.780	**2.390±0.780**
USHCN	0.713±0.120	0.757±0.210	0.693±0.160	**0.640±0.190**
COVID-19	0.135±0.006	0.130±0.006	0.125±0.007	**0.093±0.005**
	0.943±0.003	0.949±0.003	0.956±0.003	**0.965±0.002**

Table 4: Prediction results with different mechanisms

	w^{in}	α	ρ	γ_l	k	γ_f	λ
MG	1	0.1	0.7	0.8	6	1.0	10^{-2}
SILSO	1	0.1	0.6	0.8	10	1.0	10^{-2}
USHCN	1	0.1	0.7	0.8	12	0.8	10^{-2}
COVID-19	1	0.2	0.8	0.8	2	0.8	10^{-2}
	1	0.3	0.9	0.7	4	0.9	10^{-2}

Table 5: Best settings of hyper-parameters of TE-ESN

	ESN	leaky-ESN	DeepESN	LS-ESN	w/o TE	TE-ESN
MC	35.05	39.65	42.98	46.05	40.46	**47.83**

Table 6: Memory capacity results of ESNs-based methods

TE-ESN Ablation Study of TE, LS and SF. Table 4 show that all theses three mechanisms contribute to the final prediction tasks.

Hyper-Parameters Analysis of TE-ESN. Figure 3 shows that setting different hyper-parameters for each reservoir has little effect on the prediction results. Table 5 is the best settings.

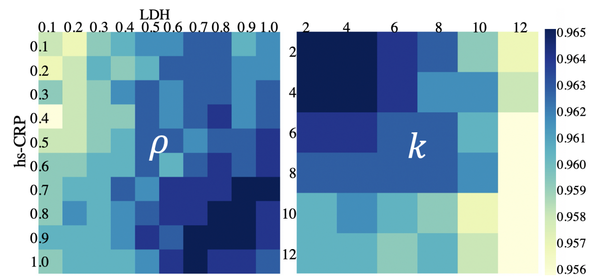

Figure 3: Prediction of TE-ESN with different ρ and k

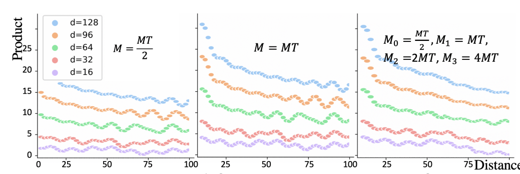

Figure 4: Dimension and frequency setting of time encoding

Memory Capability Analysis. Table 6 shows that TE-ESN obtains the best MC value and TE can increase the memory capability.

5 CONCLUSIONS

In this paper, we propose a novel Time Encoding (TE) mechanism in complex domain to model the time information of ISTS. It can represent the irregularities of intra-series and inter-series. We create a novel Time Encoding Echo State Network (TE-ESN), which is the first method to enable ESNs to handle ISTS. We evaluate the method and give several model related analysis in two prediction tasks on four datasets. The results show that TE-ESN outperforms the existing state-of-the-art and has good properties. Future works will focus on the dynamic reservoir properties and hyper-parameters optimization of TE-ESN, and will incorporate deep structures to TE-ESN for better prediction accuracy.

ACKNOWLEDGMENTS

This work was supported by the National Key Research and Development Program of China under Grant 2021YFE0205300.

REFERENCES

[1] Inci M. Baytas, Cao Xiao, Xi Zhang, Fei Wang, and Jiayu Zhou. 2017. Patient Subtyping via Time-Aware LSTM Networks. In *KDD 2017*. 65–74.

[2] SILSO World Data Center. 2016. The international sunspot number, Int. Sunspot Number Monthly Bull. Online Catalogue (1749-2016). http://www.sidc.be/silso/.

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

[4] Igor Farkas, Radomír Bosák, and Peter Gergel. 2016. Computational analysis of memory capacity in echo state networks. *Neural Networks* 83 (2016), 109–120.

[5] Hassan Ismail Fawaz, Germain Forestier, Jonathan Weber, Lhassane Idoumghar, and Pierre-Alain Muller. 2019. Deep learning for time series classification: a review. *Data Min. Knowl. Discov.* 33, 4 (2019), 917–963.

[6] Claudio Gallicchio, Alessio Micheli, and Luca Pedrelli. 2017. Deep Reservoir Computing: A Critical Experimental Analysis. *Neurocomputing* 268, dec.11 (2017), 87–99.

[7] Jonas Gehring, Michael Auli, David Grangier, Denis Yarats, and Yann N. Dauphin. 2017. Convolutional Sequence to Sequence Learning. In *ICML 2017*. 1243–1252.

[8] Yifan Hao and Huiping Cao. 2020. A New Attention Mechanism to Classify Multivariate Time Series. In *IJCAI 2020*. ijcai.org, 1999–2005.

[9] Herbert Jaeger. 2002. Adaptive Nonlinear System Identification with Echo State Networks. In *NIPS 2002*. 593–600.

[10] Herbert Jaeger, Mantas Lukosevicius, Dan Popovici, and Udo Siewert. 2007. Optimization and applications of echo state networks with leaky- integrator neurons. *Neural Networks* 20, 3 (2007), 335–352. https://doi.org/10.1016/j.neunet.2007.04.016

[11] Yoon Jinsung, Zame William R., and Mihaelaå Van Der Schaar. 2017. Estimating Missing Data in Temporal Data Streams Using Multi-Directional Recurrent Neural Networks. *IEEE Trans. Biomed. Eng.* PP (2017), 1–1.

[12] Fazle Karim, Somshubra Majumdar, and Houshang Darabi. 2019. Insights Into LSTM Fully Convolutional Networks for Time Series Classification. *IEEE Access* 7 (2019), 67718–67725.

[13] M. Mackey and L Glass. 1977. Oscillation and chaos in physiological control systems. *Science* 197, 4300 (1977), 287–289.

[14] Williams C. Menne, M. and Vose R. 2010. Long-term daily and monthly climate records from stations across the contiguous united states. https://cdiac.ess-dive.lbl.gov/epubs/ndp/ushcn/ushcn.html.

[15] Satya Narayan Shukla and Benjamin M. Marlin. 2019. Interpolation-Prediction Networks for Irregularly Sampled Time Series. In *ICLR*.

[16] Chenxi Sun, Shenda Hong, Moxian Song, and Hongyan Li. 2020. A Review of Deep Learning Methods for Irregularly Sampled Medical Time Series Data. *CoRR* abs/2010.12493 (2020).

[17] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is All you Need. In *NIPS 2017*. 5998–6008.

[18] Imran S. Virk. 2006. Diurnal Blood Pressure Pattern and Risk of Congestive Heart Failure. *Congestive Heart Failure* 12, 6 (2006), 350–351.

[19] Benyou Wang, Donghao Zhao, Qiuchi Li, Peng Zhang, and Jakob Grue Simonsen. 2020. Encoding word order in complex embeddings. In *ICLR*.

[20] Yishu Wang, Ye Yuan, Yuliang Ma, and Guoren Wang. 2019. Time-Dependent Graphs: Definitions, Applications, and Algorithms. *Data Sci. Eng.* 4, 4 (2019), 352–366.

[21] Zhengzheng Xing, Jian Pei, and Eamonn J. Keogh. 2010. A brief survey on sequence classification. *SIGKDD Explorations* 12, 1 (2010), 40–48.

[22] Hang Yan, Bocao Deng, Xiaonan Li, and Xipeng Qiu. 2019. TENER: Adapting Transformer Encoder for Named Entity Recognition. *CoRR* abs/1911.04474 (2019).

[23] Goncalves J et al. Yan L, Zhang H T. 2020. An interpretable mortality prediction model for COVID-19 patients. *Nature, Machine intelligence* 2 (2020).

[24] Kaihong Zheng, Bin Qian, Sen Li, Yong Xiao, Wanqing Zhuang, and Qianli Ma. 2020. Long-Short Term Echo State Network for Time Series Prediction. *IEEE Access* 8 (2020), 91961–91974.

[25] Shisheng Zhong, Xiaolong Xie, Lin Lin, and Fang Wang. 2017. Genetic algorithm optimized double-reservoir echo state network for multi-regime time series prediction. *Neurocomputing* 238 (2017), 191–204.