Supplementary Material: Learning Representations for Incomplete Time Series Clustering

Qianli Ma[1,2*], Chuxin Chen [1*], Sen Li [1*], Garrison W. Cottrell [3]

[1] South China University of Technology, Guangzhou, China
[2] Key Laboratory of Big Data and Intelligent Robot (South China University of Technology), Ministry of Education,
[3] University of California, San Diego, CA, USA
qianlima@scut.edu.cn, {sherlockem,awslee}@foxmail.com, gary@ucsd.edu

Section A: Dataset Introduction

In Section **Experiments**, we report the experimental results of CRLI on 8 real-world datasets. Here, we show the statistics of these datasets.

- Ali-v1∼v3 (Alizadeh et al. 2000), Chen (Chen et al. 2002) and Liang (Liang et al. 2005): We collected 5 datasets in microarray domain (Alizadeh et al. 2000; Chen et al. 2002; Liang et al. 2005). Each dataset contains gene expression information of different subtypes of different cancer. The use of clustering methods for the discovery of cancer subtypes has drawn a great deal of attention in the scientific community (de Souto et al. 2008). These datasets are preprocessed as the approach proposed in (de Souto et al. 2008).
- BloodSample (Bianchi, Mikalsen, and Jenssen 2017): Each patient in the dataset is represented by a multivariate time serie of blood samples extracted within 20 days after surgery. The multivariate time series contain measurements of 10 variables, which are alanine aminotransferase, albumin and alkaline phosphatase, creatinine, CRP, hemoglobine, leukocytes, potassium, sodium and thrombocytes. Class label indicates whether a patient is with and without surgical site infections.
- Congressional Voting Records (Dua and Graff 2017): This dataset includes votes for the U.S. House of Representatives Congressmen on the 16 key votes. This dataset is treated as a time-series dataset since the votes are sequential.
- Physionet 2012 (Silva et al. 2012): This dataset is released by the Physionet 2012 Challenge. Each sample contains a patient's measurements records in the first 48 hours after the patient's admission to ICU. Class label indicates the patient's survival. The Challenge contains 3 sub-sets and up to 12000 samples. We conduct experiments on set-a (with 4000 samples) following GRUI (Che et al. 2018) and BRITS (Cao et al. 2018). Up to 42 variables were recorded at least once, but some variables are *general descriptors* (collected on admission, not time-series). We select 35 time-series variables as what was did in BRITS, and preprocess the data (nomarlization & deletion of 3 empty samples) using code provided by BRITS.

Dataset	Clusters	Length	Dim	Missing ratio(%)	#Train	#Test	Domain
Ali-v1	2	932	1	1.84	29	13	microarray
Ali-v2	3	1030	1	2.59	43	19	microarray
Ali-v3	4	1030	1	2.59	43	19	microarray
BloodSample	2	20	10	85.72	707	176	medical
Chen	2	2328	1	2.31	125	54	microarray
Vote	2	16	1	5.41	303	131	media
Liang	3	2505	1	0.79	25	12	microarray
Physionet	2	48	35	80.52	2798	1199	medical

Table 1: Statistics of the used time series data sets

Section B: Details of Baseline Methods

We compare CRLI with both one-stage method (state-of-the-art incomplete time series deep clustering, VaDER (de Jong et al. 2019)) and two-stage methods. For two-stage methods, we first impute missing values with two SOTA imputation methods (BRITS (Cao et al. 2018), GAIN (Yoon, Jordon, and Der Schaar 2018)) and one common used ZERO imputation method, and then apply the existing SOTA clustering methods (KS (Paparrizos and Gravano 2015), DEC (Xie, Girshick, and Farhadi 2016), IDEC (Guo et al. 2017), DTC (Madiraju et al. 2018), DTCR (Ma et al. 2019)). The details are as follows:

- VaDER (de Jong et al. 2019): A generative clustering model that instantiates the variational auto-encoder clustering framework (VaDE (Jiang et al. 2017)) into RNN and integrates imputation into training process to directly cluster incomplete time series data.
- BRITS (Cao et al. 2018): A time series imputation model that jointly optimizes imputation and classification . The category information is introduced to obtain better imputed value.
- GAIN (Yoon, Jordon, and Der Schaar 2018): Generative imputation method that proposes a hint mechanism to ensure to learn the correct generator.
- ZERO: Simply replace all the missing value with zero.
- K-shape (KS) (Paparrizos and Gravano 2015): A time series clustering algorithm that adopts a scalable iterative

[*] These authors contributed equally to this work

Imputation	ZERO					GAIN					BRITS					VaDER	CRLI
Cluster	KS	DEC	IDEC	DTC	DTCR	KS	DEC	IDEC	DTC	DTCR	KS	DEC	IDEC	DTC	DTCR		
Ali-v1	0.50(0.06)	**0.85(0.00)**	0.72(0.00)	0.47(0.01)	**0.85(0.00)**	0.63(0.24)	0.46(0.00)	0.62(0.00)	0.49(0.04)	0.56(0.16)	0.50(0.06)	**0.85(0.00)**	0.82(0.06)	0.47(0.01)	0.82(0.06)	0.47(0.04)	**0.85(0.00)**
Ali-v2	0.69(0.09)	0.77(0.00)	0.79(0.06)	0.74(0.05)	0.47(0.03)	0.70(0.16)	0.51(0.00)	0.60(0.00)	0.54(0.10)	0.78(0.07)	0.60(0.02)	0.77(0.00)	0.83(0.08)	0.70(0.02)	0.48(0.01)	0.49(0.08)	**0.93(0.04)**
Ali-v3	0.75(0.03)	0.77(0.00)	0.77(0.02)	0.73(0.03)	0.63(0.06)	0.75(0.03)	0.51(0.00)	0.64(0.00)	0.48(0.18)	0.72(0.01)	0.72(0.03)	0.81(0.00)	0.74(0.03)	0.70(0.01)	0.69(0.01)	0.29(0.06)	**0.92(0.05)**
BloodSample	0.59(0.07)	0.67(0.00)	0.67(0.00)	0.72(0.06)	0.50(0.00)	0.62(0.14)	0.67(0.00)	0.65(0.00)	0.55(0.04)	0.51(0.00)	0.69(0.00)	0.62(0.00)	0.64(0.00)	0.75(0.04)	0.59(0.05)	0.67(0.06)	**0.85(0.02)**
Chen	0.59(0.09)	0.51(0.00)	0.50(0.00)	0.56(0.06)	0.51(0.01)	0.58(0.12)	**0.77(0.00)**	0.49(0.00)	0.50(0.01)	0.51(0.02)	0.54(0.06)	0.65(0.00)	0.61(0.00)	0.54(0.01)	0.50(0.00)	0.50(0.01)	0.63(0.05)
Vote	0.57(0.02)	0.83(0.00)	0.84(0.00)	0.57(0.09)	0.50(0.00)	0.69(0.05)	0.81(0.00)	0.76(0.00)	0.63(0.15)	0.50(0.01)	0.62(0.07)	0.81(0.00)	0.87(0.00)	0.69(0.11)	0.76(0.01)	0.58(0.03)	**0.91(0.05)**
Liang	0.55(0.15)	0.64(0.00)	0.70(0.00)	0.64(0.00)	**0.82(0.17)**	0.64(0.10)	0.42(0.00)	0.42(0.00)	0.56(0.00)	0.75(0.14)	0.58(0.12)	0.64(0.00)	0.70(0.00)	0.64(0.00)	0.68(0.13)	0.70(0.13)	0.67(0.03)
Physionet	0.50(0.00)	0.52(0.00)	0.52(0.00)	0.50(0.00)	0.64(0.00)	0.50(0.00)	0.51(0.00)	0.51(0.00)	0.66(0.04)	0.73(0.00)	0.50(0.00)	0.51(0.00)	0.50(0.00)	0.50(0.00)	0.52(0.01)	0.70(0.10)	**0.76(0.00)**
Best	0	1	0	0	2	0	1	0	0	0	0	1	0	0	0	0	**6**
AVG RANK	11.25	5.875	5.625	9.375	10.25	9.625	10.625	11.625	12.625	8.75	10.5	5.5	5.875	9.75	10.625	10.75	**2**
AVG RI	0.5936	0.6938	0.6890	0.6180	0.6134	0.6387	0.5835	0.5866	0.5520	0.6332	0.5948	0.7068	0.7134	0.6227	0.6299	0.5503	**0.8140**
***p*-value**	5.41E-04	3.48E-03	2.86E-03	2.65E-03	3.25E-02	6.75E-04	1.02E-02	2.30E-05	8.59E-04	1.54E-02	4.82E-04	1.84E-02	2.77E-02	1.82E-03	8.79E-03	1.18E-02	-

Table 2: Rand Index (RI) comparisons on 8 real incomplete time series datasets (the values in parentheses present standard deviations)

Imputation	ZERO					GAIN					BRITS					VaDER	CRLI
Cluster / Dataset	KS	DEC	IDEC	DTC	DTCR	KS	DEC	IDEC	DTC	DTCR	KS	DEC	IDEC	DTC	DTCR		
Ali-v1	0.06(0.09)	**0.68(0.00)**	0.38(0.00)	0.02(0.02)	0.68(0.00)	0.31(0.44)	0.50(0.00)	0.23(0.00)	0.20(0.19)	0.25(0.24)	0.13(0.15)	**0.68(0.00)**	0.64(0.08)	0.01(0.01)	0.65(0.08)	0.41(0.00)	**0.68(0.00)**
Ali-v2	0.54(0.10)	0.68(0.00)	0.70(0.06)	0.64(0.05)	0.04(0.02)	0.51(0.15)	0.30(0.00)	0.37(0.00)	0.31(0.20)	0.59(0.05)	0.43(0.04)	0.68(0.00)	0.74(0.08)	0.62(0.07)	0.17(0.05)	0.22(0.00)	**0.81(0.07)**
Ali-v3	0.63(0.07)	0.68(0.00)	0.66(0.03)	0.56(0.09)	0.29(0.08)	0.59(0.08)	0.23(0.00)	0.46(0.00)	0.09(0.63)	0.48(0.04)	0.55(0.08)	0.72(0.00)	0.55(0.05)	0.51(0.08)	0.45(0.04)	0.23(0.00)	**0.79(0.09)**
BloodSample	0.08(0.07)	0.21(0.00)	0.25(0.00)	0.29(0.08)	0.04(0.00)	0.21(0.17)	0.20(0.00)	0.20(0.00)	0.00(0.00)	0.03(0.00)	0.24(0.00)	0.27(0.00)	0.29(0.00)	0.39(0.05)	0.06(0.03)	0.20(0.00)	**0.52(0.04)**
Chen	0.17(0.16)	0.05(0.00)	0.04(0.00)	0.15(0.06)	0.02(0.01)	0.14(0.21)	**0.44(0.00)**	0.06(0.00)	0.01(0.02)	0.06(0.05)	0.10(0.13)	0.25(0.00)	0.19(0.00)	0.16(0.05)	0.01(0.01)	0.01(0.00)	0.22(0.11)
Vote	0.14(0.03)	0.64(0.00)	0.66(0.00)	0.12(0.16)	0.00(0.00)	0.32(0.08)	0.60(0.00)	0.54(0.00)	0.00(0.00)	0.01(0.01)	0.23(0.09)	0.60(0.00)	0.66(0.00)	0.35(0.21)	0.42(0.02)	0.14(0.00)	**0.78(0.09)**
Liang	0.43(0.24)	0.58(0.00)	0.65(0.00)	0.58(0.00)	**0.79(0.19)**	0.56(0.16)	0.13(0.00)	0.13(0.00)	0.00(0.00)	0.71(0.17)	0.46(0.20)	0.58(0.00)	0.65(0.00)	0.58(0.00)	0.59(0.18)	0.51(0.00)	0.61(0.03)
Physionet	0.00(0.00)	0.01(0.00)	0.01(0.00)	0.00(0.00)	0.00(0.00)	0.00(0.00)	0.00(0.00)	0.00(0.00)	0.00(0.00)	0.01(0.00)	0.00(0.00)	0.02(0.00)	**0.03(0.00)**	0.01(0.00)	0.00(0.00)	0.03(0.00)	0.01(0.01)
Best	0	1	0	0	1	0	1	0	0	0	0	1	1	0	0	0	**5**
AVG RANK	11.25	5.875	5.625	9.125	12.25	9.125	10.5	11.125	15.375	10	10.5	3.625	3.75	7.875	11.375	11.5	**2.5**
AVG NMI	0.2558	0.4412	0.4193	0.2963	0.2310	0.3315	0.3012	0.2503	0.0764	0.2663	0.2677	0.4746	0.4700	0.3288	0.2936	0.2184	**0.5530**
***p*-value**	1.09E-02	1.89E-02	1.48E-02	2.74E-02	4.09E-02	6.66E-03	3.32E-02	1.19E-03	1.09E-03	2.21E-02	4.70E-03	6.30E-02	5.99E-02	2.73E-02	1.59E-02	5.48E-03	-

Table 3: Normalized Mutual Information (NMI) comparisons on 8 real incomplete time series datasets (the values in parentheses present standard deviations)

Imputation	ZERO					GAIN					BRITS					VaDER	CRLI
Cluster / Dataset	KS	DEC	IDEC	DTC	DTCR	KS	DEC	IDEC	DTC	DTCR	KS	DEC	IDEC	DTC	DTCR		
Ali-v1	0.62(0.09)	**0.92(0.00)**	0.85(0.00)	0.57(0.04)	0.92(0.00)	0.71(0.21)	0.54(0.00)	0.77(0.00)	0.60(0.08)	0.68(0.14)	0.62(0.09)	**0.92(0.00)**	0.91(0.03)	0.55(0.03)	0.91(0.03)	0.55(0.08)	**0.92(0.00)**
Ali-v2	0.80(0.04)	0.84(0.00)	0.87(0.03)	0.83(0.02)	0.68(0.00)	0.79(0.07)	0.68(0.00)	0.74(0.00)	0.76(0.07)	0.80(0.02)	0.75(0.02)	0.84(0.00)	0.88(0.06)	0.83(0.02)	0.68(0.00)	0.68(0.02)	**0.94(0.02)**
Ali-v3	0.67(0.04)	0.74(0.00)	0.74(0.04)	0.64(0.09)	0.56(0.07)	0.67(0.02)	0.47(0.00)	0.58(0.00)	0.47(0.11)	0.65(0.03)	0.64(0.09)	0.79(0.00)	0.71(0.03)	0.56(0.03)	0.65(0.03)	0.38(0.04)	**0.93(0.03)**
BloodSample	0.76(0.02)	0.79(0.00)	0.80(0.00)	0.83(0.04)	0.74(0.00)	0.79(0.08)	0.79(0.00)	0.78(0.00)	0.74(0.00)	0.74(0.00)	0.81(0.00)	0.75(0.00)	0.77(0.00)	0.85(0.03)	0.75(0.02)	0.80(0.05)	**0.92(0.01)**
Chen	0.69(0.12)	0.59(0.00)	0.57(0.00)	0.68(0.07)	0.60(0.02)	0.67(0.13)	**0.87(0.00)**	0.57(0.00)	0.57(0.00)	0.60(0.04)	0.63(0.09)	0.78(0.00)	0.74(0.00)	0.65(0.01)	0.57(0.00)	0.57(0.02)	0.73(0.06)
Vote	0.69(0.02)	0.91(0.00)	0.92(0.00)	0.68(0.10)	0.61(0.00)	0.81(0.05)	0.89(0.00)	0.86(0.00)	0.72(0.15)	0.61(0.00)	0.74(0.08)	0.89(0.00)	0.93(0.00)	0.80(0.11)	0.86(0.00)	0.68(0.05)	**0.95(0.03)**
Liang	0.85(0.09)	0.92(0.00)	0.92(0.00)	0.92(0.00)	**0.95(0.05)**	0.88(0.07)	0.75(0.00)	0.75(0.00)	0.75(0.00)	0.93(0.04)	0.85(0.09)	0.92(0.00)	0.92(0.00)	0.92(0.00)	0.88(0.07)	0.83(0.07)	0.92(0.00)
Physionet	0.86(0.00)	0.86(0.00)	0.86(0.00)	0.86(0.00)	0.86(0.00)	0.86(0.00)	0.86(0.00)	0.86(0.00)	0.86(0.00)	0.86(0.00)	0.86(0.00)	0.86(0.00)	0.86(0.00)	0.86(0.00)	0.86(0.00)	0.86(0.00)	**0.86(0.00)**
Best	0	1	0	0	1	0	1	0	0	0	0	1	0	0	0	0	**6**
AVG RANK	9.125	5.125	5.5	7.75	9.75	7.875	10	10.5	12.75	9	9.25	4.375	4.5	7.875	9.75	12.375	**1.625**
AVG PUR	0.7432	0.8214	0.8151	0.7516	0.7403	0.7740	0.7326	0.7390	0.6842	0.7347	0.7378	0.8442	0.8393	0.7531	0.7711	0.6705	**0.8954**
***p*-value**	6.70E-03	2.47E-02	1.41E-02	2.20E-02	2.74E-02	5.85E-03	5.22E-02	2.80E-03	2.79E-03	8.87E-03	4.52E-03	9.79E-02	1.01E-01	3.03E-02	1.27E-02	8.44E-03	-

Table 4: Cluster Purity (PUR) comparisons on 8 real incomplete time series datasets (the values in parentheses present standard deviations)

Imputation	ZERO					GAIN					BRITS					VaDER	CRLI
Cluster / Dataset	KS	DEC	IDEC	DTC	DTCR	KS	DEC	IDEC	DTC	DTCR	KS	DEC	IDEC	DTC	DTCR		
Ali-v1	0.62(0.09)	**0.92(0.00)**	0.85(0.00)	0.57(0.04)	0.92(0.00)	0.71(0.21)	0.54(0.00)	0.77(0.00)	0.60(0.08)	0.68(0.14)	0.62(0.09)	**0.92(0.00)**	0.91(0.03)	0.55(0.03)	0.91(0.03)	0.55(0.08)	**0.92(0.00)**
Ali-v2	0.64(0.13)	0.84(0.00)	0.85(0.06)	0.62(0.11)	0.45(0.07)	0.66(0.19)	0.53(0.00)	0.68(0.00)	0.57(0.09)	0.77(0.05)	0.54(0.02)	0.84(0.00)	0.88(0.06)	0.56(0.03)	0.45(0.05)	0.68(0.04)	**0.94(0.02)**
Ali-v3	0.62(0.06)	0.74(0.00)	0.74(0.04)	0.58(0.15)	0.51(0.07)	0.63(0.05)	0.37(0.00)	0.42(0.00)	0.47(0.11)	0.58(0.05)	0.62(0.09)	0.68(0.00)	0.65(0.03)	0.48(0.04)	0.53(0.04)	0.38(0.04)	**0.93(0.03)**
BloodSample	0.70(0.10)	0.79(0.00)	0.80(0.00)	0.83(0.04)	0.54(0.00)	0.71(0.15)	0.79(0.00)	0.78(0.00)	0.65(0.09)	0.59(0.01)	0.81(0.00)	0.75(0.00)	0.77(0.00)	0.85(0.03)	0.70(0.07)	0.79(0.05)	**0.92(0.01)**
Chen	0.69(0.12)	0.59(0.00)	0.57(0.00)	0.68(0.07)	0.60(0.02)	0.66(0.15)	**0.87(0.00)**	0.54(0.00)	0.56(0.03)	0.57(0.07)	0.63(0.09)	0.78(0.00)	0.74(0.00)	0.65(0.01)	0.55(0.02)	0.56(0.02)	0.73(0.06)
Vote	0.69(0.02)	0.91(0.00)	0.92(0.00)	0.66(0.12)	0.51(0.01)	0.81(0.05)	0.89(0.00)	0.86(0.00)	0.72(0.15)	0.55(0.04)	0.74(0.08)	0.89(0.00)	0.90(0.00)	0.78(0.15)	0.86(0.00)	0.68(0.05)	**0.95(0.03)**
Liang	0.57(0.14)	0.67(0.00)	0.67(0.00)	0.67(0.00)	0.80(0.18)	0.63(0.07)	0.50(0.00)	0.50(0.00)	0.75(0.00)	0.73(0.15)	0.62(0.11)	0.67(0.00)	0.67(0.00)	0.67(0.00)	0.68(0.09)	**0.80(0.07)**	0.68(0.04)
Physionet	0.52(0.01)	0.61(0.00)	0.60(0.00)	0.53(0.01)	0.76(0.00)	0.52(0.00)	0.57(0.00)	0.56(0.00)	0.78(0.03)	0.84(0.00)	0.53(0.00)	0.57(0.00)	0.55(0.00)	0.52(0.01)	0.58(0.04)	0.82(0.11)	**0.86(0.00)**
Best	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	1	**6**
AVG RANK	11.5	5.25	5.625	9.75	10.375	10.125	11	11.25	11	9.625	10.625	5.375	6	10.375	10.5	9.5	**1.875**
AVG ACC	0.6303	0.7586	0.7481	0.6428	0.6361	0.6670	0.6318	0.6386	0.6373	0.6638	0.6368	0.7632	0.7586	0.6336	0.6576	0.6583	**0.8661**
***p*-value**	3.84E-04	9.89E-03	4.88E-03	3.36E-03	2.53E-02	1.07E-03	2.07E-02	1.70E-03	5.94E-03	8.96E-03	1.31E-03	4.67E-02	4.44E-02	5.81E-03	1.17E-02	2.32E-02	-

Table 5: Cluster Accuracy (ACC) comparisons on 8 real incomplete time series datasets (the values in parentheses present standard deviations)

refinement procedure to explore the shapes of time series that have a normalized cross-correlation measure

- DEC (Xie, Girshick, and Farhadi 2016): Learns a mapping from the data space to a lower-dimensional feature

Metric	RI				NMI				PUR				ACC			
Dataset \ Imputation	ZERO	GAIN	BRITS	CRLI	ZERO	GAIN	BRITS	CRLI	ZERO	GAIN	BRITS	CRLI	ZERO	GAIN	BRITS	CRLI
Ali-v1	0.50(0.02)	0.49(0.00)	0.49(0.03)	**0.85(0.00)**	0.17(0.05)	0.15(0.00)	0.11(0.06)	**0.68(0.00)**	0.63(0.03)	0.62(0.00)	0.62(0.05)	**0.92(0.00)**	0.63(0.03)	0.62(0.00)	0.62(0.05)	**0.92(0.00)**
Ali-v2	0.76(0.14)	0.65(0.00)	0.85(0.12)	**0.93(0.04)**	0.69(0.19)	0.35(0.00)	0.79(0.14)	**0.81(0.07)**	0.86(0.08)	0.79(0.00)	0.92(0.07)	**0.94(0.02)**	0.65(0.20)	0.79(0.00)	0.82(0.20)	**0.94(0.02)**
Ali-v3	0.78(0.01)	0.44(0.01)	0.76(0.02)	**0.92(0.05)**	0.75(0.03)	0.30(0.01)	0.73(0.05)	**0.79(0.09)**	0.76(0.03)	0.53(0.00)	0.74(0.04)	**0.93(0.03)**	0.76(0.03)	0.52(0.02)	0.74(0.04)	**0.93(0.03)**
BloodSample	0.61(0.00)	0.61(0.00)	0.81(0.00)	**0.85(0.02)**	0.01(0.00)	0.01(0.00)	0.45(0.00)	**0.52(0.04)**	0.74(0.00)	0.74(0.00)	0.89(0.00)	**0.92(0.01)**	0.73(0.00)	0.73(0.00)	0.89(0.00)	**0.92(0.01)**
Chen	0.49(0.00)	0.50(0.00)	0.49(0.00)	**0.63(0.05)**	0.10(0.00)	0.03(0.02)	0.10(0.00)	**0.22(0.11)**	0.57(0.00)	0.57(0.00)	0.57(0.00)	**0.73(0.06)**	0.50(0.00)	0.56(0.00)	0.50(0.00)	**0.73(0.06)**
Vote	0.81(0.00)	0.50(0.00)	0.78(0.01)	**0.91(0.05)**	0.60(0.00)	0.02(0.00)	0.56(0.02)	**0.78(0.09)**	0.89(0.00)	0.61(0.00)	0.88(0.00)	**0.95(0.03)**	0.89(0.00)	0.50(0.00)	0.88(0.00)	**0.95(0.03)**
Liang	0.64(0.00)	0.64(0.00)	0.64(0.00)	**0.67(0.03)**	0.58(0.00)	0.48(0.00)	0.58(0.00)	**0.61(0.03)**	**0.92(0.00)**	0.83(0.00)	**0.92(0.00)**	**0.92(0.00)**	0.67(0.00)	**0.75(0.00)**	0.67(0.00)	0.68(0.04)
Physionet	0.52(0.00)	**0.76(0.00)**	0.50(0.00)	**0.76(0.00)**	0.00(0.00)	0.00(0.00)	0.01(0.00)	**0.01(0.01)**	0.85(0.00)	0.85(0.00)	0.85(0.00)	**0.86(0.00)**	0.59(0.00)	0.85(0.00)	0.54(0.00)	**0.86(0.00)**
***p*-value**	6.45E-04	2.87E-03	3.43E-03	-	1.76E-02	1.61E-03	3.98E-02	-	6.99E-03	2.08E-03	2.18E-02	-	6.67E-04	8.43E-03	3.52E-03	-

Table 6: Comparison on 8 real incomplete time series with DTW+Spectral clustering with different imputation methods. (the values in parentheses present standard deviations)

space in which it iteratively optimizes a clustering objective.

- IDEC (Guo et al. 2017): Manipulates feature space to scatter data by optimizing a KL divergence-based clustering loss and maintains the local structure carefully.
- DTC (Madiraju et al. 2018): Takes the KL divergence between predicted and target distribution as guidance to learn non-linear features in a deep framework
- DTCR (Ma et al. 2019): A representative time series clustering model that utilizes a fake sample mechanism to enhance ability of the temporal encoder and guides representation learning by a novel k-means objective.

Section C: Comparison with SOTA Methods

The metricsNormalized Mutual Information (NMI), Cluster Purity (PUR) and Cluster Accuracy (ACC) of each approach on 8 real-world datasets are presented in Table 3, Table 4 and Table 5 respectively. CRLI achieves the lowest average rank and the highest average performance on all three metrics. To compare with traditional clustering methods, we also conduct experiment on DTW+Spectral clustering with different imputation methods. Due to space limit, results of DTW+Spctral clustering are presented in a new table, Table 6.

Section D: Ablation Analysis

The detail results of Ablation Analysis is presented in Figure 7, which indicates that the full CRLI is significantly superior to all of its ablations at $p < 0.05$ level, demonstrating the effectiveness of all its components.

Dataset	w/o featured	w/o fc	w/o adv	w/o jointly	CRLI
Ali-v1	0.50(0.03)	0.54(0.07)	0.85(0.19)	0.82(0.06)	**0.85(0.00)**
Ali-v2	0.51(0.12)	0.60(0.03)	0.70(0.21)	0.51(0.00)	**0.93(0.04)**
Ali-v3	0.67(0.12)	0.48(0.09)	0.67(0.00)	0.51(0.17)	**0.92(0.05)**
BloodSample	0.69(0.00)	0.55(0.09)	0.64(0.15)	0.64(0.09)	**0.85(0.02)**
Chen	0.61(0.16)	0.51(0.00)	0.51(0.17)	0.61(0.06)	**0.63(0.05)**
Vote	0.87(0.14)	0.50(0.09)	0.69(0.16)	0.57(0.03)	**0.91(0.05)**
Liang	0.58(0.10)	0.42(0.07)	0.67(0.44)	0.42(0.05)	**0.67(0.03)**
Physionet	0.69(0.24)	0.51(0.15)	0.51(0.08)	0.50(0.01)	**0.76(0.00)**
Best	0	0	0	0	**8**
AVG RANK	2.75	4.25	2.625	3.625	**1**
AVG RI	0.6413	0.5139	0.6535	0.5750	**0.8140**
***p*-value**	1.32E-02	6.60E-05	3.77E-03	3.04E-03	-

Table 7: Rand Index(RI) ablation study of CRLI

References

Alizadeh, A. A.; Eisen, M. B.; Davis, R. E.; Ma, C.; Lossos, I. S.; Rosenwald, A.; Boldrick, J. C.; Sabet, H.; Tran, T.; Yu, X.; et al. 2000. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 403(6769): 503–511.

Bianchi, F. M.; Mikalsen, K. Ø.; and Jenssen, R. 2017. Learning compressed representations of blood samples time series with missing data. In *European Symposium on Artificial Neural Networks*.

Cao, W.; Wang, D.; Li, J.; Zhou, H.; Li, L.; and Li, Y. 2018. Brits: Bidirectional recurrent imputation for time series. In *Advances in Neural Information Processing Systems*, 6775–6785.

Che, Z.; Purushotham, S.; Cho, K.; Sontag, D.; and Liu, Y. 2018. Recurrent neural networks for multivariate time series with missing values. *Scientific reports* 8(1): 1–12.

Chen, X.; Cheung, S. T.; So, S.; Fan, S. T.; Barry, C.; Higgins, J.; Lai, K.-M.; Ji, J.; Dudoit, S.; Ng, I. O.; et al. 2002. Gene expression patterns in human liver cancers. *Molecular Biology of the Cell* 13(6): 1929–1939.

de Jong, J.; Emon, M. A.; Wu, P.; Karki, R.; Sood, M.; Godard, P.; Ahmad, A.; Vrooman, H.; Hofmann-Apitius, M.; and Fröhlich, H. 2019. Deep learning for clustering of multivariate clinical patient trajectories with missing values. *GigaScience* 8(11): giz134.

de Souto, M. C.; Costa, I. G.; de Araujo, D. S.; Ludermir, T. B.; and Schliep, A. 2008. Clustering cancer gene expression data: a comparative study. *BMC Bioinformatics* 9(1): 497.

Dua, D.; and Graff, C. 2017. UCI Machine Learning Repository. http://archive.ics.uci.edu/ml/, last accessible on 2021/3/15.

Guo, X.; Gao, L.; Liu, X.; and Yin, J. 2017. Improved deep embedded clustering with local structure preservation. In *International Joint Conference on Artificial Intelligence*, 1753–1759.

Jiang, Z.; Zheng, Y.; Tan, H.; Tang, B.; and Zhou, H. 2017.

Variational deep embedding: An unsupervised and generative approach to clustering. In *International Joint Conference on Artificial Intelligence*, 965–1972.

Liang, Y.; Diehn, M.; Watson, N.; Bollen, A. W.; Aldape, K. D.; Nicholas, M. K.; Lamborn, K. R.; Berger, M. S.; Botstein, D.; Brown, P. O.; et al. 2005. Gene expression profiling reveals molecularly and clinically distinct subtypes of glioblastoma multiforme. *Proceedings of the National Academy of Sciences* 102(16): 5814–5819.

Ma, Q.; Zheng, J.; Li, S.; and Cottrell, G. W. 2019. Learning Representations for Time Series Clustering. In *Advances in Neural Information Processing Systems*, 3781–3791.

Madiraju, N. S.; Sadat, S. M.; Fisher, D.; and Karimabadi, H. 2018. Deep Temporal Clustering: Fully unsupervised learning of time-domain features. *ArXiv Preprint ArXiv:1802.01059* .

Paparrizos, J.; and Gravano, L. 2015. K-shape: Efficient and accurate clustering of time series. In *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, 1855–1870.

Silva, I.; Moody, G.; Scott, D. J.; Celi, L. A.; and Mark, R. G. 2012. Predicting in-hospital mortality of icu patients: The physionet/computing in cardiology challenge 2012. In *2012 Computing in Cardiology*, 245–248. IEEE.

Xie, J.; Girshick, R.; and Farhadi, A. 2016. Unsupervised deep embedding for clustering analysis. In *International Conference on Machine Learning*, 478–487.

Yoon, J.; Jordon, J.; and Der Schaar, M. V. 2018. GAIN: Missing Data Imputation using Generative Adversarial Nets. In *International Conference on Machine Learning*, 5675–5684.