

# Supplementary Materials: Estimating Missing Data in Temporal Data Streams Using Multi-directional Recurrent Neural Networks

Jinsung Yoon\*, William R. Zame, and Mihaela van der Schaar, *Fellow, IEEE*

## STANDARD DEVIATIONS FOR THE MAIN RESULTS

TABLE I  
STANDARD DEVIATIONS OF TABLE II

Category	Algorithm	Mean RMSE (% Gain of M-RNN (Multiple Imputations))				
		MIMIC-III	Deterioration	UNOS-Heart	UNOS-Lung	Biobank
M-RNN	M-RNN (MI)	0.0021	0.0019	0.0023	0.0038	0.0044
	M-RNN (SI)	0.0022	0.0022	0.0025	0.0041	0.0043
RNN-based	[1]	0.0030	0.0027	0.0034	0.0045	0.0049
	[2]	0.0024	0.0023	0.0025	0.0040	0.0046
	[3]	0.0027	0.0027	0.0031	0.0046	0.0049
Interpolation	Spline	0.0023	0.0029	0.0025	0.0039	0.0041
	Cubic	0.0028	0.0029	0.0031	0.0045	0.0048
Imputation	MICE	0.0025	0.0023	0.0026	0.0042	0.0043
	MissForest	0.0024	0.0023	0.0027	0.0043	0.0054
	EM	0.0028	0.0039	-	-	0.0067
Others	Matrix Completion	0.0022	0.0023	0.0025	0.0030	0.0034
	Auto-encoder	0.0020	0.0027	0.0023	0.0037	0.0044
	MCMC	0.0024	0.0023	0.0024	0.0041	0.0054

TABLE II  
STANDARD DEVIATIONS OF TABLE III

Algorithm	Mean RMSE (% Gain from Imputation Algorithm)				
	MIMIC-III	Deterioration	UNOS-Heart	UNOS-Lung	Biobank
Spline + MICE	0.0025	0.0023	0.0037	0.0043	0.0039
Spline + MissForest	0.0032	0.0039	0.0041	0.0045	0.0047
Cubic + MICE	0.0021	0.0024	0.0041	0.0039	0.0043
Cubic + MissForest	0.0034	0.0031	0.0045	0.0036	0.0050

Manuscript received November 23, 2017; revised February 28, 2018. This work was supported by the Office of Naval Research (ONR) and the NSF (Grant number: ECCS1462245, ECCS1533983, and ECCS1407712).

\*J. Yoon is with the Department of Electrical and Computer Engineering, University of California, Los Angeles, CA, 90095 USA; (correspondence e-mail: jsoon0823@ucla.edu).

W. R. Zame is with the Departments of Economics and Mathematics, University of California, Los Angeles, CA, 90095 USA.

M. van der Schaar is with the Department of Engineering Science, University of Oxford, UK and the Alan Turing Institute, London, UK.

TABLE III  
STANDARD DEVIATIONS OF TABLE IV

Datasets	M-RNN (Std of RMSE)		
	Only Interp	Only Impute	Interp + Impute
MIMIC-III	0.0015	0.0027	0.0021
Deterioration	0.0017	0.0032	0.0019
UNOS-Heart	0.0021	0.0033	0.0023
UNOS-Lung	0.0030	0.0041	0.0038
Biobank	0.0032	0.0037	0.0044

TABLE IV  
STANDARD DEVIATIONS OF TABLE V

Category	Algorithm	AUROC				
		MIMIC-III	Deterioration	UNOS-Heart	UNOS-Lung	Biobank
M-RNN	M-RNN (MI)	0.0058	0.0064	0.0080	0.0112	0.0133
	M-RNN (SI)	0.0047	0.0067	0.0085	0.0125	0.0129
RNN-based	[1]	0.0079	0.0059	0.0092	0.0133	0.0143
	[2]	0.0069	0.0089	0.0096	0.0129	0.0149
	[3]	0.0083	0.0053	0.0076	0.0099	0.0131
Interpolation	Spline	0.0081	0.0072	0.0061	0.0108	0.0107
	Cubic	0.0091	0.0104	0.0093	0.0149	0.0146
Imputation	MICE	0.0078	0.0075	0.0095	0.0115	0.0128
	MissForest	0.008	0.0074	0.0088	0.0133	0.0143
	EM	0.0088	0.0097	-	-	0.0154
Others	Matrix Completion	0.0047	0.0077	0.0058	0.0126	0.0146
	Auto-encoder	0.0060	0.0065	0.0057	0.0092	0.0117
	MCMC	0.0059	0.0081	0.0072	0.0126	0.0137

TABLE V  
STANDARD DEVIATIONS OF TABLE VII

Algorithm	Std of RMSE	
	MCAR	MAR
M-RNN (MI)	0.0021	0.0028
[3]	0.0027	0.0031
Cubic Interpolation	0.0028	0.0032
MissForest	0.0024	0.0034
Matrix Completion	0.0022	0.029

## REFERENCES

- [1] E. Choi et al, "Doctor ai: Predicting clinical events via recurrent neural networks," *arXiv preprint arXiv:1511.05942*, 2015.
- [2] Z. C. Lipton et al, "Directly modeling missing data in sequences with rnns: Improved classification of clinical time series," *arXiv preprint arXiv:1606.04130*, 2016.
- [3] Z. Che et al, "Recurrent neural networks for multivariate time series with missing values," *arXiv preprint arXiv:1606.01865*, 2016.