Beyond Random Missingness: Clinically Rethinking for Healthcare Time Series Imputation

- This study emphasizes the critical role of masking strategies (RMEO vs. RMOD) and normalization timing (before/after masking) in clinical time-series imputation, evaluated on the PhysioNet 2012 dataset. Results demonstrated that self-attention-based models (SAITS) excel under dynamic overlay masking (RMOD), while bidirectional RNNs (BRITS) remain robust across strategies, with both outperforming traditional methods. Their work also highlights that clinically realistic missingness simulations and normalization protocols significantly influence imputation accuracy and downstream predictive performance in healthcare AI.

Top 3 references:

1. Cao, W., Wang, D., Li, J., Zhou, H., Li, L., Li, Y.: Brits: Bidirectional recurrent imputation for time series. Advances in neural information processing systems 31 (2018)
2. Du, W., Côté, D., Liu, Y.: Saits: Self-attention-based imputation for time series. Expert Systems with Applications 219, 119619 (2023)
3. Tashiro, Y., Song, J., Song, Y., Ermon, S.: Csdi: Conditional score-based diffusion models for probabilistic time series imputation. Advances in Neural Information Processing Systems 34, 24804–24816 (2021)

Accounting for Missing Data in Public Health Research Using a Synthesis of Statistical and Mathematical Models

- The paper introduces a synthesized framework combining statistical and mathematical models to address positivity violations in missing data using the NHANES data by estimating SBP (Systolic Blood Pressure) in children aged 2-17. They partitioned the problem into a positive region (ages 8–17), modeled via saturated g-computation, and a nonpositive region (ages 2–7), imputed using external SBP distributions. As a result, the synthesized estimator produced better SBP estimates than the conventional approaches.

Addressing Missing Data Challenges in Geriatric Health Monitoring: A Study of Statistical and Machine Learning Imputation Methods

- This paper evaluated statistical (EM, matrix completion), machine learning (KNN, SVM), and deep learning (GAIN, VAE, GRU-D) approaches for imputing the missing sensor data in geriatric health monitoring system, where missing data were categorized as MCAR, MAR and MNAR. As a result, the KNN and SVM outperformed other techniques across all missingness patterns with the lowest normalized MSE/MAE. GRU-D, a deep learning approach was comparatively better in MNAR scenarios.

Name: Rabin BK
Matriculation Number: 23272000
FAU mail: rabin.bk@fau.de