Journal of Data Science 24 (2), 293–295
April 2026
DOI: 10.6339/26-JDS242EDI
Editorial

Editorial: Bridging Disciplines: Advancing AI, Statistics, and Data Science Together

Sy Han Chiou[1,*], Mingzhao Hu[2], and Jing Cao[1]

[1]*Department of Statistics and Data Science, Southern Methodist University, Dallas, TX, USA*
[2]*Department of Quantitative Health Sciences, Mayo Clinic College of Medicine and Science, Rochester, MN, USA*

In this special issue of the *Journal of Data Science*, we present nine peer-reviewed papers from the 2025 Symposium on Data Science and Statistics (SDSS 2025), themed "Bridging Disciplines: Advancing AI, Statistics, and Data Science Together." As data-driven problems grow in complexity and scale, progress increasingly depends on integrating statistical theory, machine learning, and domain-specific knowledge. This theme emphasizes not only methodological innovation but also the synthesis of ideas across traditionally distinct areas. The contributions in this special issue reflect this direction through work on dependence modeling, extreme value analysis, interpretable machine learning, AI-enabled research workflows, visualization, and software tools for applied data analysis.

Methodological contributions in the *Statistical Data Science* section span dependence modeling, sentiment analysis, visualization, and reliability assessment. Pareek and Ghosh (2025) develop a semiparametric dynamic copula framework for rolling-window portfolio optimization. The approach combines nonparametrically estimated empirical beta copulas with parametric marginals from the skewed generalized t-family to capture asymmetry, heavy tails, and time-varying dependencies. The model is applied to financial returns from the United States, India, and Hong Kong to study the influence of dynamic market conditions. Yang et al. (2025) propose a fine-grained attention-based multiple instance classification model for word-level sentiment analysis. The model uses only document-level sentiment labels but operates at the word level to improve interpretability, producing outputs such as contextual weighting and negation cues that show how context shapes sentiment. Berry et al. (2025) develop a Bayesian hierarchical probit framework for comparing intraclass correlation coefficients across pre- and post-intervention settings with ordinal ratings. The model accounts for rater- and item-level correlations, and simulation studies show that modeling the pre-post correlation improves estimation accuracy. An application to a multicenter study on conjunctival inflammation illustrates the approach.

The *Data Science in Action* section features applied work in vocational rehabilitation, environmental statistics, and education. Le et al. (2025) investigate how interactive feature selection affects graphical perception by extending a previous lineup study with toggleable aesthetic features such as cluster coloring, ellipses, and trendlines. A generalized linear mixed model identifies mixture proportion as the strongest predictor of target selection, contributing to understanding how users engage with interactive tools in exploratory data analysis. Taylor et al. (2025) develop an `R` package for cleaning and analyzing the Rehabilitation Services Administration Case Service Report (RSA-911) data, a standardized dataset collected quarterly on vocational rehabilitation participants. The package also handles the newly introduced Transition Readiness Toolkit scores. The authors provide the first analysis linking these scores to RSA-911 demographic data and

*Corresponding author. Email: schiou@smu.edu or hu.mingzhao@mayo.edu or jcao@smu.edu.

deliver an online dashboard built with the Shiny framework. Russell et al. (2025) study extreme precipitation in the Northern US Rocky Mountains using tools from multivariate extreme value theory. Specifically, they use asymptotic dependence measures to assess how well satellite precipitation estimates from the Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks–Climate Data Record (PERSIANN-CDR) agree with ground observations from the Snowpack Telemetry (SNOTEL) network during extreme events. They present two estimators of asymptotic dependence and find that both indicate positive spatial dependence. A spatial correlation analysis shows that elevation is negatively associated with asymptotic dependence while average summer temperature is positively associated. Chickering et al. (2025) describe how artificial intelligence and automation support qualitative research workflows in an educational setting. They present three interconnected pipelines developed for the Illinois Needs Assessment project that automate transcription, report generation, and summary assembly, replacing ad hoc manual workflows with reproducible systems.

The *Computing in Data Science* section includes two `R` packages for applied data analysis. Burger and Bean (2025) introduce `cdlsim` for quantifying the sensitivity of land use and land cover metrics derived from the USDA Cropland Data Layer. The package simulates data at the patch level using accuracy statistics published by the National Agricultural Statistics Service. A case study on a mixed agricultural and grassland landscape in South Dakota illustrates how the package enables researchers to assess the impact of classification error on downstream metrics. Linse et al. (2025) present `modeldiagramR` for creating model structure diagrams for hierarchical linear mixed models fitted with `lme4` or `nlme`. The package generates a visualization directly from the data and the fitted model object, helping practitioners understand the relationships among random effect variables in models with nested random effects.

Together, these nine papers illustrate the breadth of the SDSS 2025 symposium, from foundational statistical methodology to domain-specific applications and accessible software tools. We thank the anonymous referees who contributed their time and expertise to the peer-review process for this special issue.

References

Berry A, Cao J, Zhang S (2025). A Bayesian approach to pre-post comparison of inter-rater agreement in ordinal ratings. *Journal of Data Science*, 24(2): 338–351. Accepted paper for the SDSS 2025 special issue. https://doi.org/10.6339/25-JDS1213

Burger H, Bean B (2025). Quantifying the sensitivity of land use land cover metrics through simulation techniques. *Journal of Data Science*, 24(2): 436–454. Accepted paper for the SDSS 2025 special issue. https://doi.org/10.6339/25-JDS1215

Chickering G, Jones C, Blaushild N (2025). Leveraging artificial intelligence and automation for enhancing school improvement efforts. *Journal of Data Science*, 24(2): 411–435. Accepted paper for the SDSS 2025 special issue. https://doi.org/10.6339/26-JDS1216.

Le D, Rogers R, Robinson E (2025). Clusters, trends, and choices: Feature selection in interactive statistical graphics. *Journal of Data Science*, 24(2): 352–372. Accepted paper for the SDSS 2025 special issue. https://doi.org/10.6339/26-JDS1221.

Linse G, Greenwood M, June R (2025). Data-driven model structure diagrams for hierarchical linear mixed models. *Journal of Data Science*, 24(2): 455–475. Accepted paper for the SDSS 2025 special issue. https://doi.org/10.6339/26-JDS1222.

Pareek S, Ghosh SK (2025). Semiparametric dynamic copula models using rolling-window portfolio optimization. *Journal of Data Science*, 24(2): 296–318. Accepted paper for the SDSS 2025 special issue. https://doi.org/10.6339/26-JDS1229.

Russell BT, Momo Nizegha K, Malshika D, Huang WK (2025). Investigating spatial dependence in the degree of asymptotic dependence between a satellite precipitation product and station data in the Northern US Rocky Mountains. *Journal of Data Science*, 24(2): 394–410. Accepted paper for the SDSS 2025 special issue. https://doi.org/10.6339/26-JDS1220.

Taylor R, Bean B, Phillips B, Fleming A (2025). Designing accessible and dependable tools for vocational rehabilitation data analysis. *Journal of Data Science*, 24(2): 373–393. Accepted paper for the SDSS 2025 special issue. https://doi.org/10.6339/26-JDS1228.

Yang C, Larson E, Cao J (2025). Interpretable word-level context-based sentiment analysis. *Journal of Data Science*, 24(2): 319–337. Accepted paper for the SDSS 2025 special issue. https://doi.org/10.6339/26-JDS1225.