SOMI CHA, Ph.D.

Postdoctoral Researcher in Mathematics and Statistics

Department of Mathematics and Statistics, Chonnam National University, Republic of Korea

RESEARCH & COLLABORATION PROFILE |

KEY WORDS:, high-dimensional clustering; variable selection

Research Snapshot. Her research focuses on high-dimensional statistical analysis, particularly variable selection and clustering of mixed-type data. Her work combines spectral methods, statistical hypothesis testing, and multiple-testing procedures. Her research also includes deep learning applications in medical image analysis and natural language processing.

RESEARCH EXPERTISE

- **High-dimensional clustering:** Development of clustering methods for high-dimensional data containing continuous and categorical variables using spectral methods.
- **Variable selection and statistical inference:** Identification of informative variables through statistical hypothesis testing and multiple-testing procedures to improve clustering performance.
- **Deep learning and medical image analysis:** Attention-based transfer learning for pneumonia detection using chest X-ray images.
- **Natural language processing:** Keyphrase extraction from traffic accident reports using pre-trained language models and BiLSTM-CRF.

RECENT RESEARCH HIGHLIGHTS

- **High-dimensional clustering (Ph.D. Dissertation, 2026):** Developed a variable-selection framework for clustering high-dimensional continuous and categorical data using spectral methods and statistical hypothesis testing.

- **Natural language processing (IEEE Access, 2023):** Developed a keyphrase extraction approach using pre-trained language models and BiLSTM-CRF for traffic accident reports.

- **Medical image analysis (Applied Sciences, 2021):** Applied attention-based transfer learning to pneumonia detection using chest X-ray images.

POTENTIAL COLLABORATION RESEARCH

Best-fit collaborator areas: high-dimensional statistics & clustering | variable selection & statistical inference | medical image analysis | natural language processing

- **1. High-dimensional clustering and variable selection:** Joint development of statistical methods for clustering high-dimensional mixed-type data through spectral methods, hypothesis testing, and multiple-testing procedures.
- **2. Medical image analysis:** Joint development and evaluation of deep learning methods for medical imaging, including attention-based transfer learning.
- **3. Natural language processing:** Collaborative research on keyphrase extraction and information extraction from unstructured text using pre-trained language models.

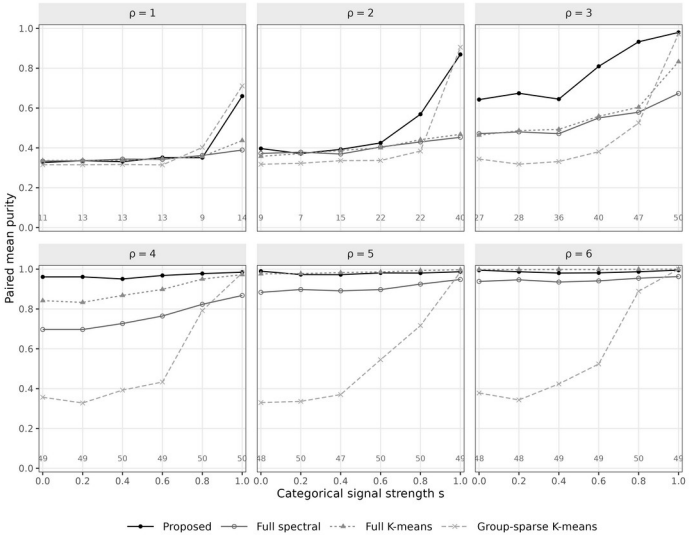
SELECTED RECENT PUBLICATIONS

- **IEEE Access 2023, 11, 139521–139533—** Extracting Fallen Objects on the Road From Accident Reports Using a Natural Language Processing Model-Based Approach.
- **Appl. Sci. 2021, 11, 1242 —** Attention-Based Transfer Learning for Efficient Pneumonia Detection in Chest X-ray Images.

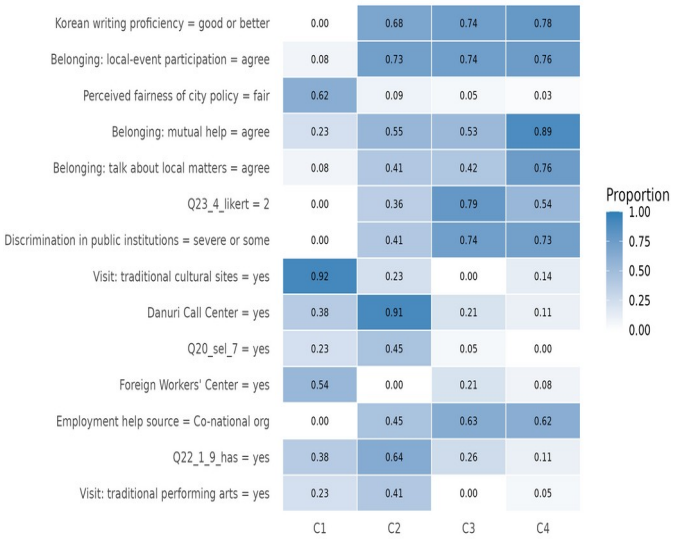
CONTACT

ssomchaa@jnu.ac.kr

REPRESENTATIVE RESEARCH FIGURES / SCHEMES



Comparison of clustering purity across methods and signal strengths.



Cluster profiles based on selected variables from the 2024 Seoul Survey on Foreign Residents.