

# A Comprehensive Survey on Imbalanced Data Learning

**Xinyi GAO, Dongting XIE, Yihang ZHANG, Zhengren WANG,  
Chong CHEN, Conghui HE, Hongzhi YIN, Wentao ZHANG**

Frontiers of Computer Science, DOI: [10.1007/s11704-025-50274-7](https://doi.org/10.1007/s11704-025-50274-7)

# Problems

- Imbalanced data distributions are prevalent in various types of raw data and severely hinder the performance of ML by biasing the decision-making processes.
- To deepen the understanding of imbalanced data and facilitate the related research and applications, this survey systematically analyzes various real-world data formats and concludes existing researches for different data formats into diverse distinct categories.

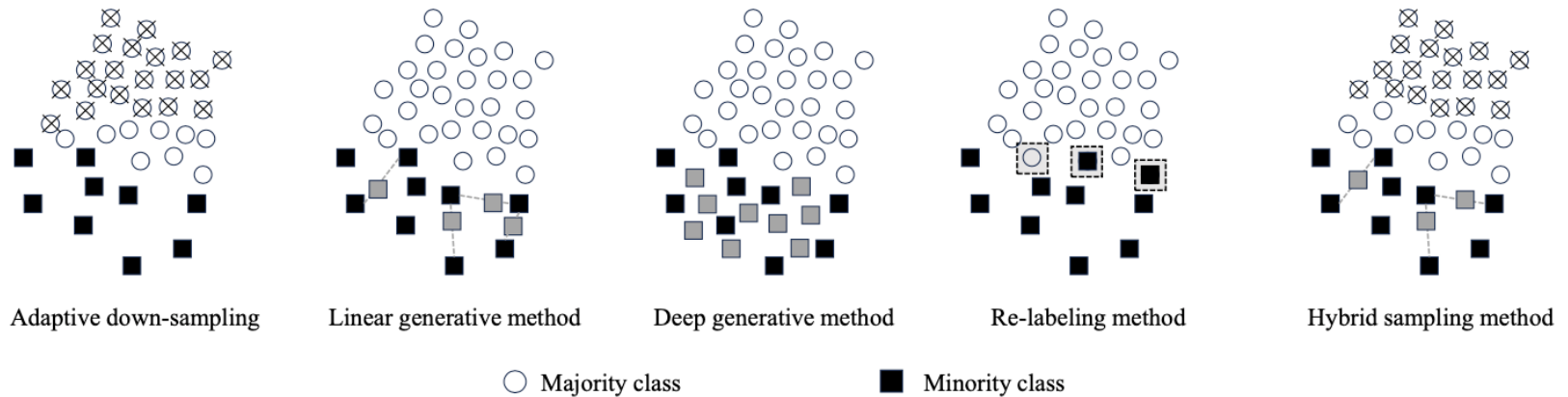


Fig. 1 The data imbalance and re-balancing methods.

# Main Contributions

- Contributions:
  - We offer a structured overview of methods that are rooted in the foundational processes of ML.
  - We conduct a thorough analysis across various data formats, including images, text, and graphs.
  - We highlight resources available for tackling imbalanced data and explore current challenges and future directions.

