

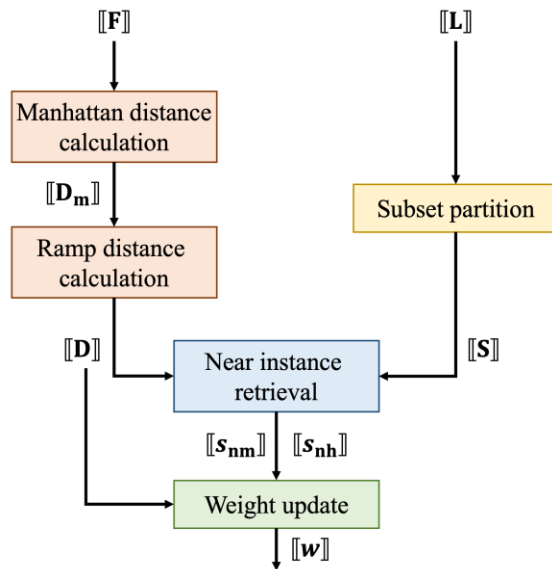
Enhancing Data Quality with Effective Feature Selection and Privacy Protection

**Luyao WANG, Zhusen LIU, Qi FENG, Weibin WU,
Lu ZHOU**

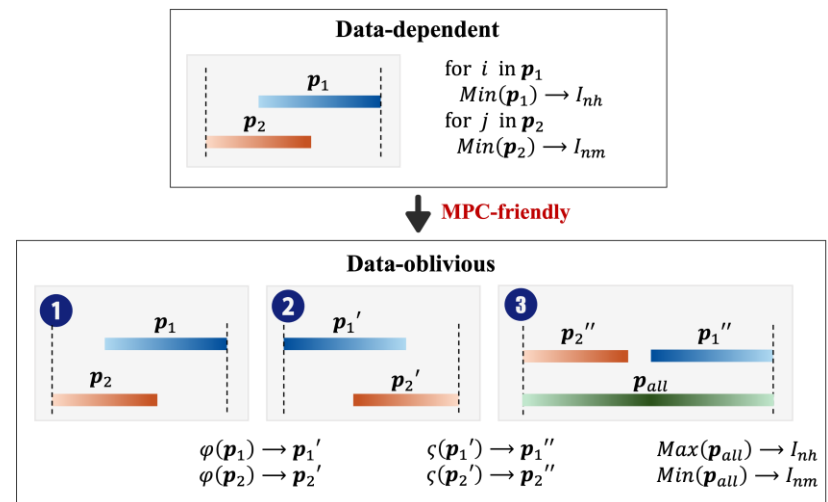
Frontiers of Computer Science, DOI: [10.1007/s11704-025-41074-0](https://doi.org/10.1007/s11704-025-41074-0)

Problems & Ideas

- Problems of privacy-preserving feature selection schemes:
 - Most MPC-based feature selection schemes overlook the correlation between features and perform poorly for model training when handling datasets containing both numerical and categorical attributes.
 - Implementing the feature selection methods using a concatenation-based approach with MPC is ineffective and may lead to a privacy breach.
- Ideas: Converting all steps into data-oblivious and constructing a bidirectional vector to estimate the near instances.



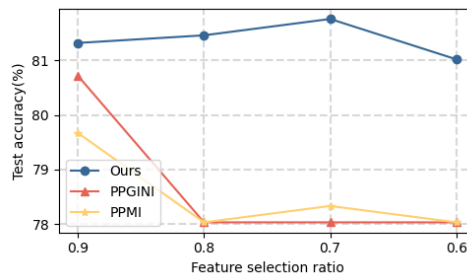
The essential steps of MPC-Relief



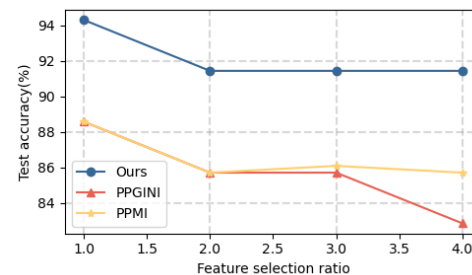
MPC-friendly bidirectional vector construction method

Main Contributions

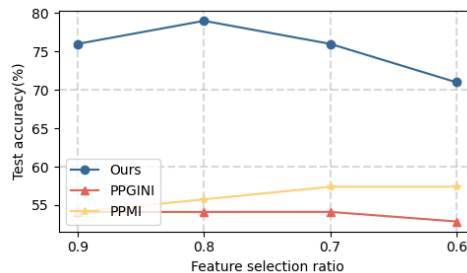
- Contributions:
 - Privacy-preserving distance calculation approach with mix-type features that maintains the same utility.
 - Privacy-preserving feature selection scheme with Relief called MPC-Relief under the setting with multiparty computation. ;
 - A bidirectional vector construction method to estimate the near instances, which avoids conditional judgments.



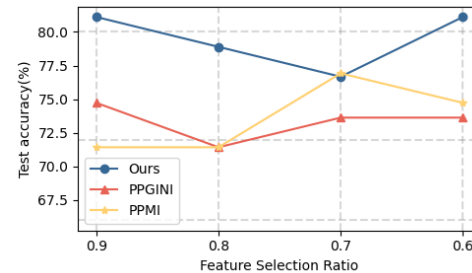
Mixed dataset Adult



Mixed dataset Darwin



Mixed dataset Heart disease



Mixed dataset Arrhythmias

The proposed scheme can effectively enhance the accuracy of model training and exhibits superior performance on datasets with mixed features.