

Abstract

**I. Introduction**

**II. Repeated elastic net technique for feature selection**

A. Ensemble training and selection criteria  
B. Hyperparameter selection  
C. Training runtime complexity of RENT

**III. Experiments**

A. Experimental setup and datasets  
B. Evaluation metrics  
C. RENT hyperparameter selection  
D. Validation study of features selected with RENT  
E. Comparison of RENT with established feature selectors

**IV. Exploratory post-hoc analysis**

A. Analysis of training objects  
B. Principal component analysis (PCA) on selected features

**V. Discussion**

**VI. Conclusion**

Acknowledgment

References

Abstract

**1 Background**

1.1 Deep learning  
1.2 Preprocessing  
1.3 Prediction models

**2 Datasets**

2.1 The FTIR spectra  
2.2 The AVIRIS remote sensing data

**3 Methods**

3.1 Validation

**4 Results**

4.1 The preprocessing hyperparameters  
4.2 Pipeline comparison  
4.3 Time usage

**5 Discussion**

5.1 DL-based preprocessing with ANN prediction models  
5.2 DL-based preprocessing with PLS prediction models  
5.3 ANN as a feature extractor for linear models  
5.4 Interpretability  
5.5 Neural networks and small datasets  
5.6 Time usage  
5.7 Efforts required

**6 Conclusion**

Conflict of interest

Peer review  
ORCID  
References  
Supporting information

Abstract

**1. Introduction**

**2. Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (FT-ICR MS)**

**3. Text mining**

3.1. Search strategy

**4. Results**

4.1. Text analysis study  
4.2. Classification vs. Regression  
4.3. Ordinary Least Squares (OLS)  
4.4. Latent variable-based methods  
4.5. Regularisation-based methods  
4.6. Data used in literature

**5. Conclusions and future perspectives**

Declaration of competing interest  
Data availability  
Acknowledgements  
Appendix A. Supplementary data  
References