

A comprehensive evaluation of a network feature construction algorithm FeCO3 based on inter-feature correlation coefficients

Shenggeng Lin^{1,2,*}, Yuqi Lin^{1,*}, Kexin Wu^{1,*}, Yueying Wang³, Zixuan Feng¹, Meiyu Duan¹, Shuai Liu¹, Jingxuan Huang⁴, Zhao Wang⁴, Yusi Fan⁴, Lan Huang¹, Fengfeng Zhou^{1,#}.

¹ College of Computer Science and Technology, and Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun, Jilin 130012, China.

² State Key Laboratory of Microbial Metabolism, and School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China.

³ Department of Epidemiology and Biostatistics, School of Public Health, Jilin University, Changchun, Jilin Province, China.

⁴ College of Software, and Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun, Jilin 130012, China.

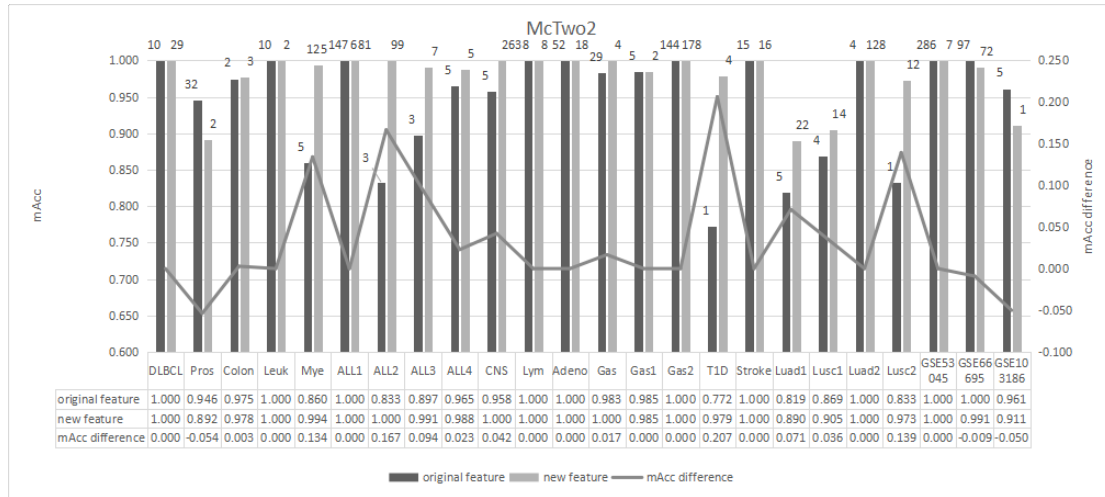
* Contributed equally to this study.

Correspondence may be addressed to Fengfeng Zhou: FengfengZhou@gmail.com or ffzhou@jlu.edu.cn . Lab web site: <http://www.healthinformatics.org/> . Phone: +86-431-8516-6024. Fax: +86-431-8516-6024.

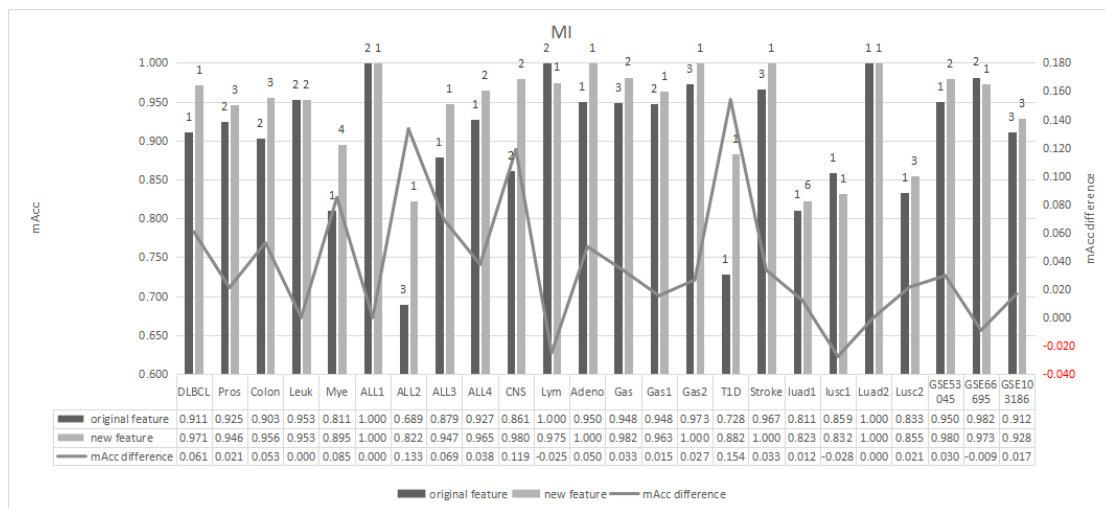
Supplementary Figure S1

Comparison of the classification metric mAcc and the number of features between the original features and the FeCO3 features using the feature selection algorithm McTwo2. The series “original” gives the data of the original features, and the series “FeCO3” illustrates the status of the FeCO3 features. The series “Improvement” is the mAcc value of the FeCO3 features minus the mAcc value of the original features. The classification accuracy is calculated

using the 5-fold cross validation strategy. The data was calculated using the feature selection algorithms

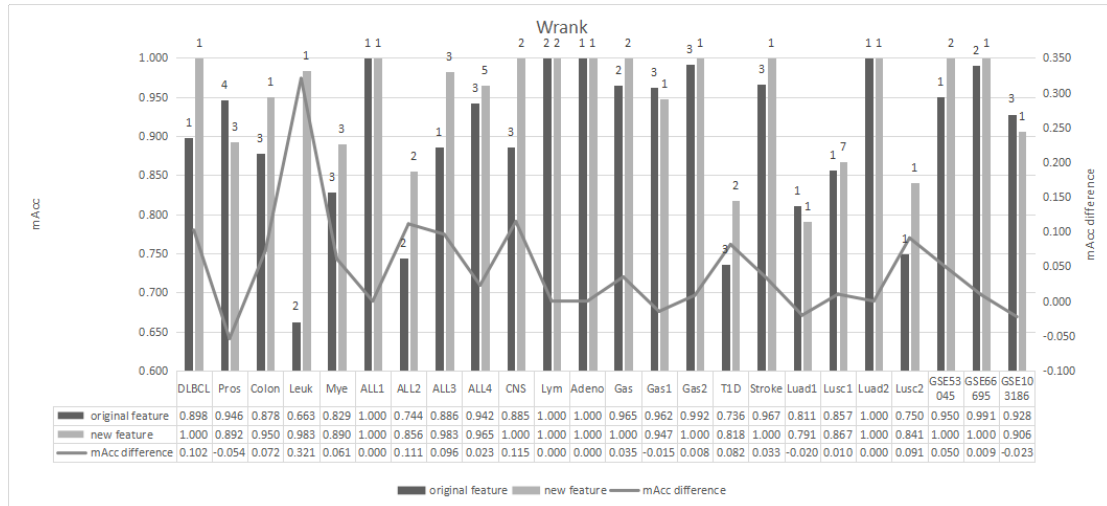


Comparison of the classification metric mAcc and the number of features between the original features and the FeCO3 features using the feature selection algorithm MI. The series “original” gives the data of the original features, and the series “FeCO3” illustrates the status of the FeCO3 features. The series “Improvement” is the mAcc value of the FeCO3 features minus the mAcc value of the original features. The classification accuracy is calculated using the 5-fold cross validation strategy. The data was calculated using the feature selection algorithms

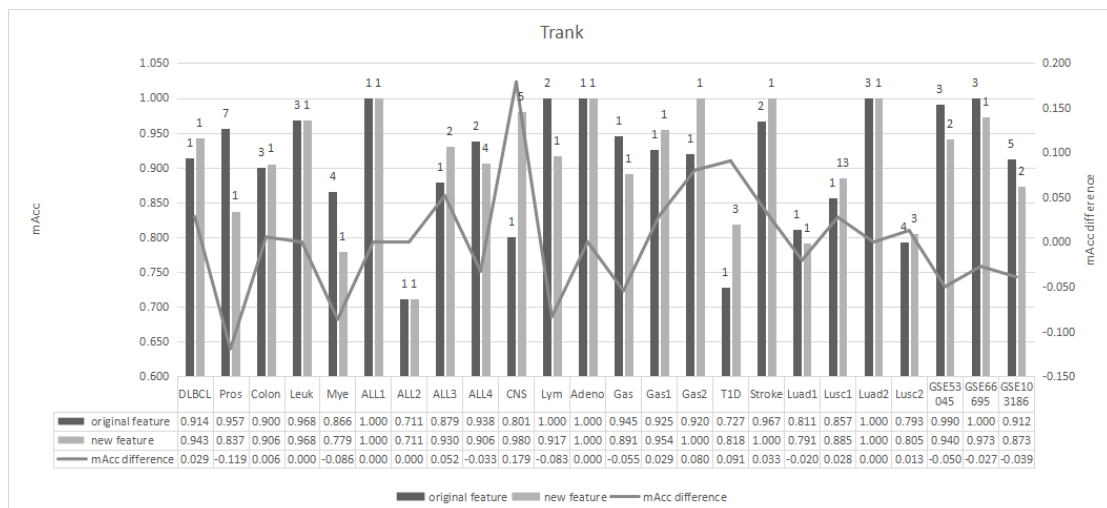


Comparison of the classification metric mAcc and the number of features between the original features and the FeCO3 features using the feature selection algorithm Wrank. The series “original” gives the data of the original features, and the series “FeCO3” illustrates

the status of the FeCO3 features. The series “Improvement” is the mAcc value of the FeCO3 features minus the mAcc value of the original features. The classification accuracy is calculated using the 5-fold cross validation strategy. The data was calculated using the feature selection algorithms

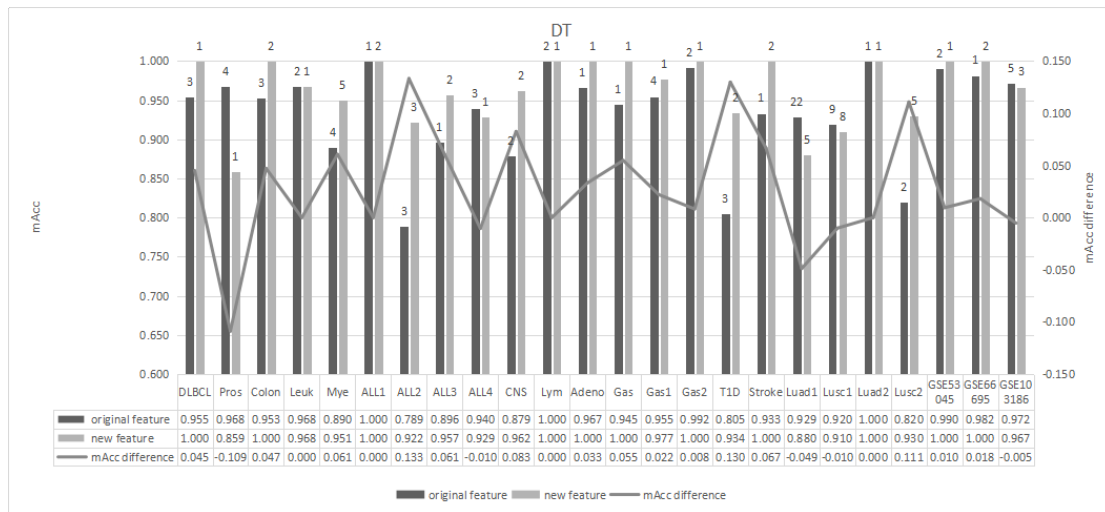


Comparison of the classification metric mAcc and the number of features between the original features and the FeCO3 features using the feature selection algorithm Trank. The series “original” gives the data of the original features, and the series “FeCO3” illustrates the status of the FeCO3 features. The series “Improvement” is the mAcc value of the FeCO3 features minus the mAcc value of the original features. The classification accuracy is calculated using the 5-fold cross validation strategy. The data was calculated using the feature selection algorithms

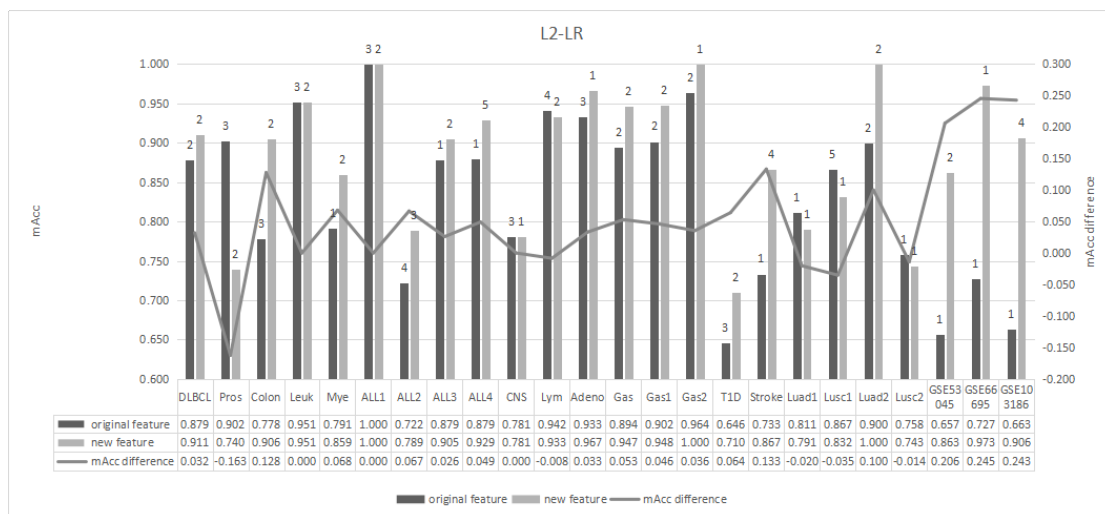


Comparison of the classification metric mAcc and the number of features between the original features and the FeCO3 features using the feature selection algorithm DT. The series “original” gives the data of the original features, and the series “FeCO3” illustrates the status of the FeCO3 features. The series “Improvement” is the mAcc value of the FeCO3 features minus the mAcc value of the original features. The classification accuracy is calculated using the 5-fold cross validation strategy. The data was calculated using the feature selection

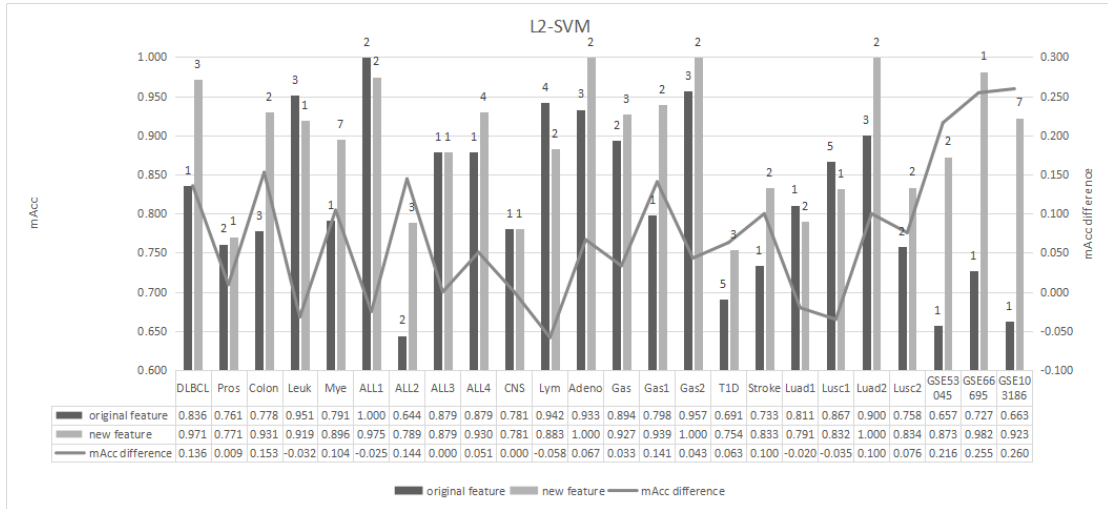
algorithms



Comparison of the classification metric mAcc and the number of features between the original features and the FeCO3 features using the feature selection algorithm L2-LR. The series “original” gives the data of the original features, and the series “FeCO3” illustrates the status of the FeCO3 features. The series “Improvement” is the mAcc value of the FeCO3 features minus the mAcc value of the original features. The classification accuracy is calculated using the 5-fold cross validation strategy. The data was calculated using the feature selection algorithms



Comparison of the classification metric mAcc and the number of features between the original features and the FeCO3 features using the feature selection algorithm L2-SVM. The series “original” gives the data of the original features, and the series “FeCO3” illustrates the status of the FeCO3 features. The series “Improvement” is the mAcc value of the FeCO3 features minus the mAcc value of the original features. The classification accuracy is calculated using the 5-fold cross validation strategy. The data was calculated using the feature selection algorithms



Comparison of the classification metric mAcc and the number of features between the original features and the FeCO3 features using the feature selection algorithm XGB. The series “original” gives the data of the original features, and the series “FeCO3” illustrates the status of the FeCO3 features. The series “Improvement” is the mAcc value of the FeCO3 features minus the mAcc value of the original features. The classification accuracy is calculated using the 5-fold cross validation strategy. The data was calculated using the feature selection algorithms

