Bigram-PGK: phosphoglycerylation prediction using the technique of bigram probabilities of position specific scoring matrix

Abel Chandra[5,*], Alok Sharma[1,2,3,5,7,*], Abdollah Dehzangi[4], Daichi Shigemizu[2,3,6,7], and Tatsuhiko Tsunoda[2,3,7]

[1] Institute for Integrated and Intelligent Systems, Griffith University, Brisbane, QLD-4111, Australia
[2] Department of Medical Science Mathematics, Medical Research Institute, Tokyo Medical and Dental University (TMDU), Tokyo, 113-8510, Japan
[3] Laboratory for Medical Science Mathematics, RIKEN Center for Integrative Medical Sciences, Yokohama, 230-0045, Kanagawa, Japan
[4] Department of Computer Science, Morgan State University, Baltimore, Maryland, USA
[5] School of Engineering and Physics, Faculty of Science Technology and Environment, University of the South Pacific, Suva, Fiji
[6] Medical Genome Center, National Center for Geriatrics and Gerontology, Obu, Aichi 474-8511, Japan
[7] CREST, JST, Tokyo, 102-8666, Japan
[*] Corresponding authors
abelavit@gmail.com, alok.sharma@griffith.edu.au

Table 1. Comparison of Bigram-PGK predictor and Bigram-PGK predictor without the profile bigrams using 10-fold cross-validation scheme and **polynomial SVM kernel**. Highest values of the metrics are highlighted in bold.

Predictor	Sensitivity	Specificity	Precision	Accuracy	MCC	AUC
Bigram-PGK (without bigram)	0.1517	**0.9562**	NaN	0.6890	NaN	0.5543
Bigram-PGK	**0.9642**	0.8973	**0.8253**	**0.9193**	**0.8330**	**0.9306**

Table 2. Comparison of Bigram-PGK predictor and Bigram-PGK predictor without the profile bigrams using 10-fold cross-validation scheme and **linear SVM kernel**. Highest values of the metrics are highlighted in bold.

Predictor	Sensitivity	Specificity	Precision	Accuracy	MCC	AUC
Bigram-PGK (without bigram)	0.5914	0.8316	0.6381	0.7520	0.4346	0.7125
Bigram-PGK	**0.8427**	**0.9239**	**0.8454**	**0.8984**	**0.7689**	**0.8855**