Improving identification of animal secretory proteins

Jiratchaya Nuanpirom [1] and Unitsa Sangket [2,*]

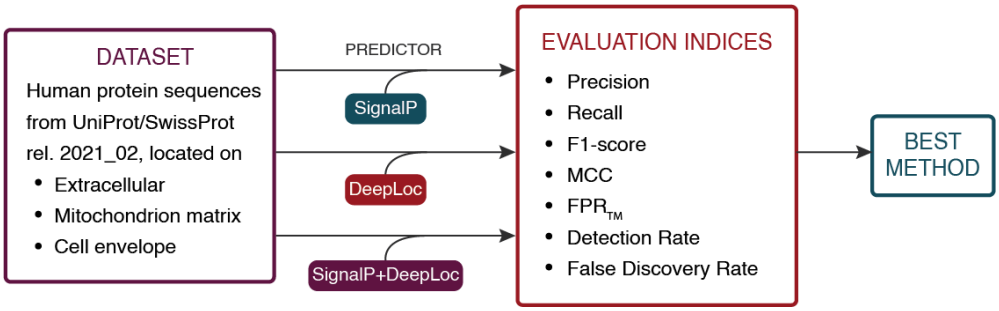
[1] Division of Biological Science, Faculty of Science, Prince of Songkla University, Hat Yai, Thailand. 90110; jirath.nuan@gmail.com

[2] Center for Genomics and Bioinformatics Research, Division of Biological Science, Faculty of Science, Prince of Songkla University, Hat Yai, Thailand. 90110; unitsa.s@psu.ac.th

* Correspondence: unitsa.s@psu.ac.th

Abstract: Nascent protein translated inside cells is controlled by a signal peptide to various cellular compartments, including the secretory protein. A preproprotein sequence has a signal peptide flanking it at the N-terminus. To distinguish the signal peptide from proteins that migrate to other compartments, several state-of-the-art tools have been developed. Because the signal peptide and the transmembrane protein are too close to each other, problems have arisen. In addition, several proteins can be found in multiple locations. Therefore, we proposed to use the integrated approach to bootstrap the performance of the traditional prediction method. The combination of SignalP and DeepLoc can provide a better result than any single predictor alone. This study was conducted using the protein sequences from the recently reviewed database and applied the scoring indices derived from the confusion matrix. In terms of recall and F1 score, the results show that the integrated method outperforms the individual predictors. Some indices are slightly different from those of the single predictor. Moreover, the integrated method increases the detection rates while decreasing the false discovery rates. It can be shown that the combination of multiple predictor algorithms outperforms the conventional predictor method.

Graphical abstract:



Keywords: SignalP; DeepLoc; signal peptide; extracellular

Funding: This research received no external funding

Data Availability Statement: Dataset, Commands, R script, and additional files are deposited in GitHub repository https://github.com/JirathNuan/BMB2021.

Acknowledgments: We would like to thank the Center for Genomics and Bioinformatics Research, Faculty of Science, Prince of Songkla University, Thailand for supporting the computer server in this research. Although the pandemic is still underway, we thank the BMB2021 committee for organizing this international conference.

Conflicts of Interest: The authors declare no conflict of interest.