2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Prediction of TCM Effective against Bacterial Pneumonia and Identification of Antibacterial Natural Product

Pei Gao[1], Zheng Chen[1], Ming Huang[1], Naoaki Ono[1,2], Shigehiko Kanaya[1,2] and MD Altaf-Ul-Amin[1,*]

***Abstract*—Based on the empirical data from the formulae of Traditional Chinese Medicine, this study screened out the natural products with antibacterial potential. The efficacy of natural products was evaluated by regression and classification models as candidates for new antibiotics.**

I. Introduction

Finding new antibiotics is a valuable effort to tackle the issue of antibiotic resistance in bacterial pneumonia. NP is intuitively associated with Traditional Chinese Medicine (TCM), hence it is reasonable to pay attention again to natural products (NP) that are the largest source of antibiotics. In practice, the TCM formula can directly guide the combination medication of NP compositions while it is effective in the antibacterial application of NP [1]. Here, a five-level complementary network provides candidates of the antibiotics for involved bacteria.

II. Experiments and Results

To eliminate the terminological differences and the indeterminate correspondence, we firstly developed a complementary network (as shown in Figure 1) from TCM formulae to bacterial pneumonia based on our previous work [2]. By building a bridge between TCM syndromes and bacterial pneumonia, we collected 2258 formulae effective for respiratory system disease composed of total 316 NP. Out of them, 1208 formulae were labeled as the positive samples that were considered to be potentially effective for bacterial pneumonia, others were negative samples. With the presence status of NP, a data matrix of NP composition and efficacy relations was constructed. Then three machine learning methods (Lasso regression, Random Forest, and XGboost) were used to evaluate the performance of efficacy prediction. The objective of machine learning is not classification, but to reveal the key features/NP by ranking their contributions to the models.

The performance of the three models is shown in Table 1. For the tree-based models, the contribution of representation highly relies on the accuracy of classification. Here, metrics (precision, recall, and specificity) in RF and XGboost perform a promising overall result for the classification decision. That is, the feature ranking can be considered as reliable. To mutually evaluate the importance of each candidate NP, we also introduced LASSO to remove the useless features. As shown in Figure 2 (b)-(d), the thresholds of cumulative importance were 0.50, 0.75, and 0.95, while the output of the models was cataloged into different colors. Moreover, the overall distribution of the salient NP is shown in Figure 2 (a). Noteworthy, with the least 0.50 threshold, four NP make the greater contribution to both classification and regression. Furthermore, we search the literature to support the reliability of the result. For instance, Morus alba against staphylococcus pneumonia is mentioned in [3]. In summary, the candidates of antibiotics for treating bacterial pneumonia identified by the proposed method can be supported by strong evidence.

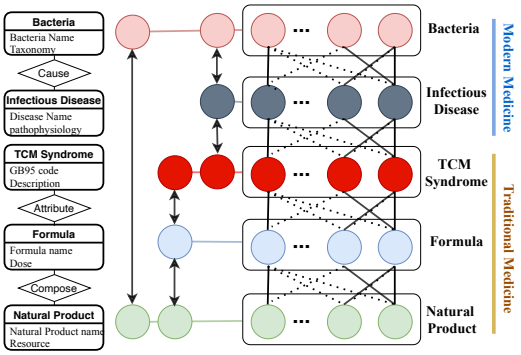

Fig.1 The five-level complementary network to bridge modern medicine and TCM. The entity relationship diagram shows the details of data structure in each level.

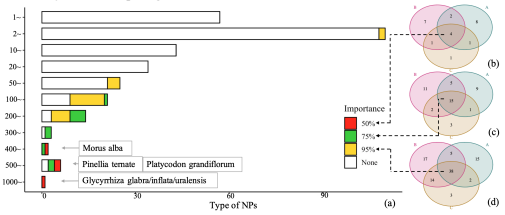

Fig.2 (a) shows the results of feature importance ranking and statistical analysis. (b) to (d) show the output of different models (pink: RF, green: LASSO, yellow: XGB).

Method		LASSO		RF	XGB
	AUC	0.9486	Precision	0.8786	0.8991
Metric(%)	MSE	0.01383	Recall	0.8984	0.9131
			Specificity	0.8586	0.9028

Tabel. 1

This work was supported by the Next Generation Interdisciplinary Research Project of NAIST and Ministry of Education, Culture, Sports, Science, and Technology of Japan (20K12043 and 16K07223)

The authors are with Graduate School of Science and Technology, Nara Insitute of Science and Technology, Takayamacho 8916-5, Ikoma, Japan. (e-mail: {gao.pei.gi3, amin-m }@is.naist.jp)

Naoaki Ono and Shigehiko Kanaya are with Data Science Center, Nara Insitute of Science and Technology, Takayamacho 8916-5, Ikoma, Japan.

References

[1] A. L. Harvey, R. Edrada-Ebel, and R. J. Quinn, "The re-emergence of natural products for drug discovery in the genomics era," *Nature reviews drug discovery*, vol. 14, no. 2, pp. 111–129, 2015.

[2] P. Gao, Z. Chen, M. Huang, N. Ono, S. Kanaya, and M. Altaf-UI-Amin, "An approach to construct and validate tcm dataset effective against bacterial pneumonia," pp. 102–103, 2021.

[3] S.-C. Wu, F. Han, M.-R. Song, S. Chen, Q. Li, Q. Zhang, K. Zhu, and J.-Z. Shen, "Natural flavones from morus alba against methicillin-resistant staphylococcus aureus via targeting the proton motive force and membrane permeability," *Journal of agricultural and food chemistry*, vol. 67, no. 36, pp. 10 222–10 234, 2019.