論文内容の要旨

申請者氏名　GAO PEI

The emergence of antibacterial resistance necessitates the development of novel antibiotics. Natural products (NPs) have demonstrated potential as candidates for antibiotic development. However, in silico approaches are required to efficiently identify NPs with antibiotic potential. This study focuses on identifying NPs with antibacterial efficacy, guided by modern medicine and traditional herbal medical systems, specifically Traditional Chinese Medicine (TCM) (China) and Ayurveda (India). To address the absence of an equivalent concept for antibiotics in TCM, we constructed a knowledge-based network integrating NPs, herbs, TCM concepts, and modern treatment protocols for infectious diseases. 5,164 NP candidates were screened, forming a dataset of 2,256 TCM formulae. Ayurveda benefits from a well-curated dataset comprising 285 Ayurvedic formulae and 293 herbal candidates, adapted for this research. For both datasets, machine learning-based feature selection methods were employed to refine the NP sample space and identify potential candidates for new antibiotic design. The TCM dataset achieved a classification accuracy of 0.942 using a Multi-Layer Perceptron model, while the Ayurveda dataset achieved an accuracy of 0.824 using a Random Forest model with Synthetic Minority Over-sampling Technique (SMOTE) preprocessing. These results confirm the reliability of the constructed datasets for classification tasks. NP candidates with antibacterial potential were extracted from the high-performance models. Validation involved extensive peer-reviewed literature analysis to verify the antibacterial properties of these plants and NPs. This process identified 70 NP candidates with antibacterial potential from the TCM dataset and 17 candidates from the Ayurveda dataset. This research highlights the potential of herbal medicine as a valuable resource for discovering new therapeutics.

論文審査結果の要旨

申請者氏名　　GAO PEI

中医学（TCM）ならびにアユルベーダの配合生薬をもとに、新規抗生物質を特に天然物化合物から探索する方法を、ケモインフォマティクスならびにバイオインフォマティクスにより検討した。まずはじめに、感染症に関する5164種の天然物を対象としたデータセットを構築した。さらに、TCM データセットをもとに、深層学習法、ならびにランダムフォーレスト法による良好な分類が可能なった。具体的には、感染症に関して82.4％を正しく認識できた。また、有効化合物のスクリーニングにより70種の天然物化合物を選択することができた。