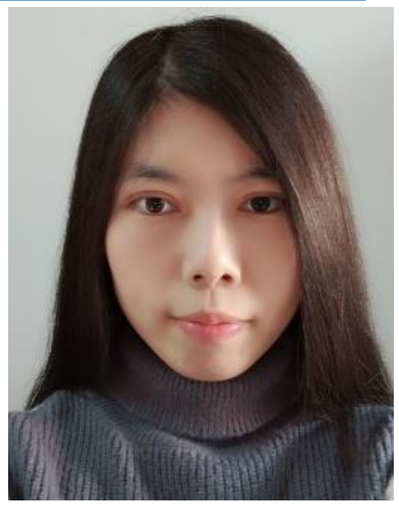
P8-6

Discrimination of periodontal pathogens using Raman spectroscopy combined with machine learning algorithms

Yiping Liu[1,#], Hongxiao Li[2,#], Shisheng Cao[1], Xin Li[1], Jing Cao[1], Baoxin Fan[1], Xingtong Mu[1], Ying Li[1,], Xiaoxi Dong[2,*], Juan Zhang[1,*]*

[1] *School of Dentistry, Tianjin Medical University, Tianjin 300070, China.*

[2] *Chinese Academy of Medical Science & Peking Union Medical College, Institute of Biomedical Engineering, Tianjin 300192, China.*



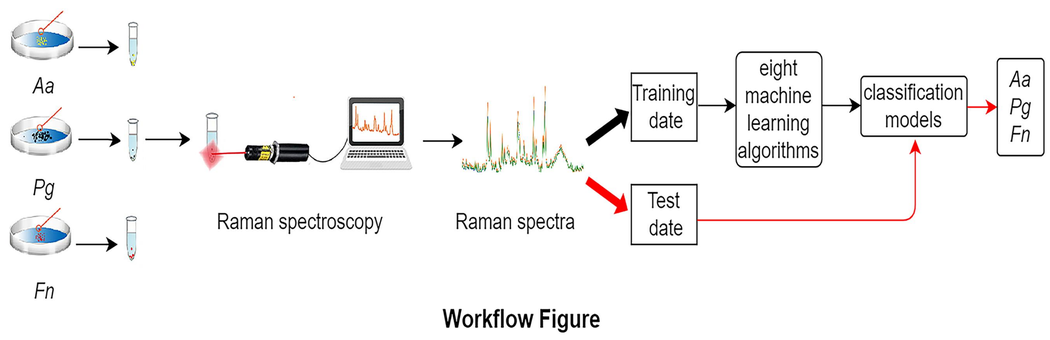
Introduction

- Periodontitis is closely related to many systemic diseases linked by different periodontal pathogens. To unravel the relationship between periodontitis and systemic diseases, it is very important to correctly discriminate major periodontal pathogens.
- Many studies have been conducted on the identification of bacteria in food and medical area using Raman spectroscopy combined with machine learning algorithms.
- Few studies reported its usage in the dental, including periodontal field.

Aims

To realize convenient, efficient, and high-accuracy bacterial species classification, the authors use Raman spectroscopy combined with machine learning algorithms to distinguish three major periodontal pathogens *Porphyromonas gingivalis* (*Pg*), *Fusobacterium nucleatum* (*Fn*), and *Aggregatibacter actinomycetemcomitans* (*Aa*).

Materials and Methods



Workflow Figure

- **Bacterial culture:** at a concentration of $3x10^9$ CFU/ml
- **Sample preparation:** 45 *Pg*, 53 *Fn*, and 54 *Aa* samples, each sample had ten spectra
- **Raman measurements:** Raman spectrometer (SLSR-ProTT, Enwave Optronics, USA), 785 nm, 450mW
- **Data preprocessing:** the spectra baselines were removed with the SNIP algorithm
- **Machine learning:** the data were randomly divided into training (75%) and testing (25%) data

Results

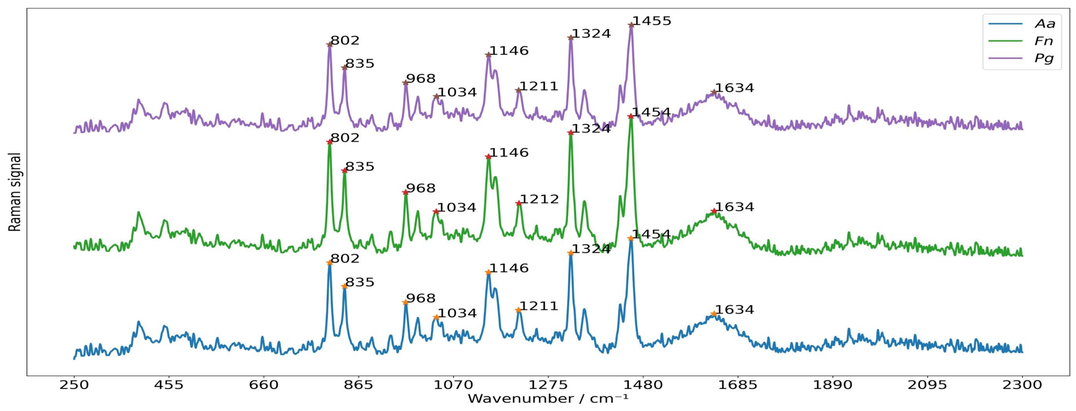


Fig 1 The average spectra of the three bacteria after baseline smoothed.

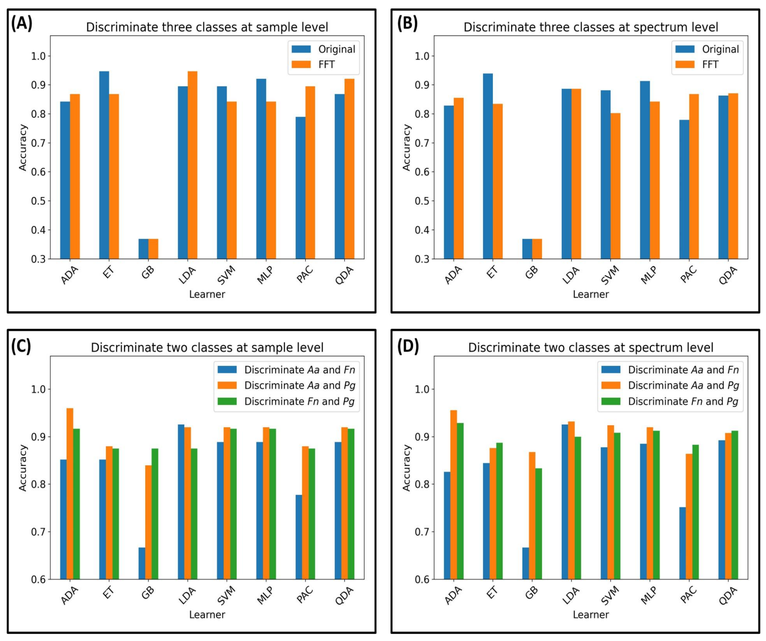


Fig 2 Accuracy of the classifier discriminating three and two classes at sample or spectrum level. ADA: Adaboost, ET: Extra trees, GB: Gradient boosting, LDA: Linear discriminant analysis, SVM: Support vector machine, MLP: Multi-layer perceptron, PAC: Passive-aggressive classifier, QDA: Quadratic discriminant analysis.

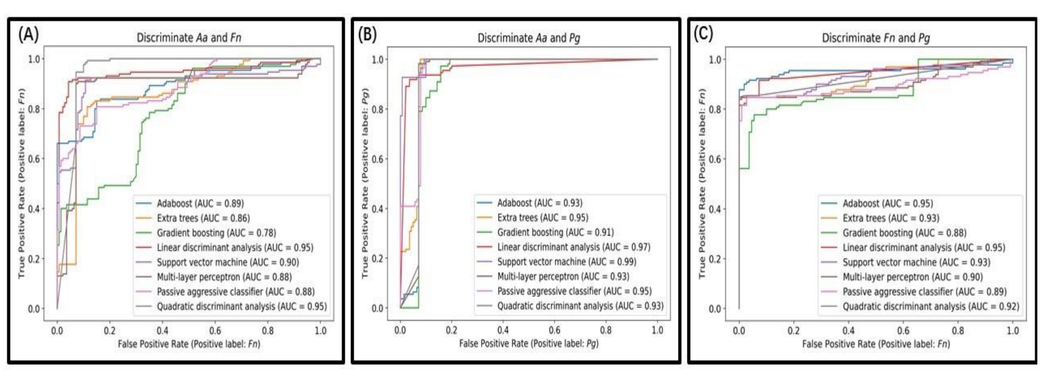


Fig 3 ROC curves of the best models from eight machine learning algorithms for the three groups of binary-class data: (A) *Aa* and *Fn*, (B) *Aa* and *Pg*, (C) *Fn* and *Pg*. These results were generated at the spectrum level.

The classification accuracies for the three categories of the original date were 94.7% at the sample level and 93.9% at the spectrum level by the machine learning algorithm extra trees. The linear discriminant analysis trained on the Fourier transformed data produced the optimal classifier at both sample (accuracy of 94.7%) and spectrum (accuracy of 88.7%) levels.

At the sample level (Figure 2 C), the best models for the three groups of binary-classes data are from linear discriminant analysis algorithm (*Aa* VS *Fn*, accuracy of 92.6%), AdaBoost algorithm (*Aa* VS *Pg*, accuracy of 96%), and four algorithms (*Fn* VS *Pg*, accuracy of 91.7%), respectively. At the spectrum level (Figure 2 D), the best classifiers for the three groups of binary-classes data are from linear discriminant analysis algorithm (*Aa* VS *Fn*, accuracy of 92.6%), AdaBoost algorithm (*Aa* VS *Pg*, accuracy of 95.6%), and AdaBoost algorithms (*Fn* VS *Pg*, accuracy of 92.9%), respectively.

Conclusion

Raman spectroscopy combined with machine learning algorithms may be an easy, fast, sensitive, and accurate method for the identification and classification of major periodontal pathogens *Pg*, *Fn*, and *Aa*. The results shed light on a novel method to discriminate periodontal pathogens. This provides a new strategy to differentiate periodontal pathogens for future discovering the underlying pathogenic mechanism of systemic diseases related to periodontitis.

Acknowledgments

This research was funded by the Major Program of Social Science Foundation of Tianjin Municipal Education Commission (2019JWZD53).