Backward Feature Elimination for Accurate Pathogen Recognition Using Portable Electronic Nose

Mukunthan Tharmakulasingam[1], Cihan Topal[1], Anil Fernando[1] and Roberto La Ragione[2]

[1]Centre for Vision Speech and Signal Processing, University of Surrey, United Kingdom
[2]Department of Pathology and Infectious Diseases, School of Veterinary Medicine, University of Surrey, United Kingdom
Email: {m.tharmakulasingam, c.topal, w.fernando, r.laragione}@surrey.ac.uk

Abstract—This paper presents the application of the backward feature elimination technique on an electronic nose (E-nose) to aid the rapid detection of pathogens using Volatile Organic Compounds (VOCs). The timely identification of pathogens is vital to facilitate control of diseases. E-noses are widely used for the identification of VOCs as a non-invasive tool. However, the identification of VOC signatures associated with microbial pathogens using E-nose is currently inefficient for the timely identification of pathogens. Therefore, we proposed an E-nose system integrating the backward feature elimination. Comprehensive experiments of backward feature elimination showed that they improve the classification accuracy.

Index Terms—Electronic nose, Pathogen detection, Backward feature elimination, Machine learning

I. Introduction

Consumer electronics (CE) has permeated every aspect of the health care domain. An Electronic nose (E-nose) is an intelligent sensing device with a gas sensor array which is used to identify odours and gas in biomedical, agriculture, environmental, food and various scientific research fields. E-nose analyses the odour measurement and provides a fast response for clinical diagnosis. An electronic nose consists of a sensor array which acts as detecting system, microprocessors and connected computers to act as computing system and delivery system to send the volatile organic compounds into the detection system of the electronic nose. Appropriate data analysis methods and pattern recognition algorithms are developed in computers and connected to e-nose thorough software and microprocessors to accurately predict the gases [1].

Accurate pathogen identification is essential in the health care domain to facilitate disease diagnosis and treatment, and to identify Antimicrobial Resistance (AMR) which is a growing public health threat associated with 700,000 deaths per year [2]. Therefore, the timely identification of the presence of pathogens in humans and animals is a key step to prevent AMR. Furthermore, it may enable more targeted prescribing of antibiotics and thus reducing the emergence of AMR. As significant AMR genes vary for different pathogens [3], the identification of pathogens will be the first step for identifying AMR using Machine learning approaches.

Traditional methodologies for the identification of bacteria depend on the culture of clinical specimens which can be time consuming, invasive and expensive. Volatile Organic Compounds (VOCs) which are identified in e-nose can be used as biological markers to confirm the presence of pathogens as humans do not produce the same array of VOCs. Thus, combinations of VOCs can be used as a rapid, and economic option for the diagnosis of infectious diseases [4]. Even though e-noses have been used to identify gases, they are not widely used to recognize pathogens directly as a non-invasive and fast response system.

E-noses uses surface acoustic waves (SAW), a quartz crystal micro-balance (QCM), metal-oxide semiconductors (MOS), conductive polymers (CP) and optical sensors in sensor array to sense the gases [5]. Responses of the sensor will be sent to a microprocessor to convert analogue responses to digital values. Connection to an external computer is used to control the data recording and to run a pattern-recognition system to discriminate odours and gases bases on trained data. Graphical analysis, multivariate data analysis and network analysis such as artificial neural network (ANN), radial basis function (RBF) etc. are used to identify VOCs from sensor responses [6]. This conventional e-nose setup requires external computers to train data to identify gases and to identify pathogens amidst the high complexity and time consuming methods. Other drawbacks of e-nose are low sensitivity and specificity in comparison with microbiological and molecular methods in identifying VOCs and pathogens using VOCs. As numerous VOC are produced by bacteria, training model with all VOCs is not feasible to identify pathogens [7]. To reduce the complexity and time consuming, VOCs contributing to the identification of pathogens should be identified. However, there is no fixed set of VOCs to identify pathogens and VOCs are changing over time and place. For these reasons, machine learning can be applied to identify relevant sets of VOCs to accurately predict pathogens.

In this study, we propose a portable e-nose design to run Machine learning algorithm in an optimized manner. This will facilitate numerous applications for the e-nose in the medical