Development of a Low-Cost Portable Electronic Nose for the Detection of Colorectal Cancer By Using Convolutional Neural Networks

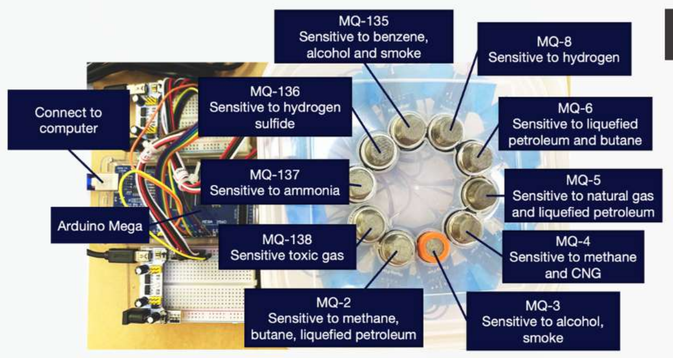
YUELONG WANG, JIAHAO LIANG
Macau Kao Yip Middle School

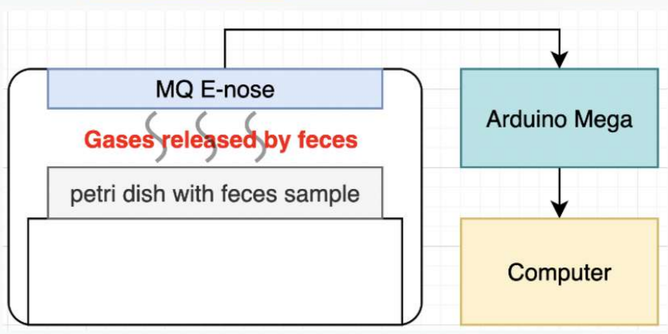
Background

- Colorectal cancer(CRC) ranking third in the global incidence of cancer (10.2%) and second in mortality (9.2%).
- In China, around 1.2~1.5% of people in the age between 50 to 64 had a preliminary screening for colorectal cancer.
- In China, current screening method is immune fecal occult blood test (iFOBT) with a sensitivity around 80%.
- The type and characteristics of the intestinal flora can be considered as diagnostic markers of colorectal cancer.
- These intestinal flora produce different metabolized gasses like volatile organic compounds (VOCs) which are emitted through breathing.
- These gases are considered biomarkers of colon cancer (CRC), such as 1-iodinonane (1-iodononane) and benzene compounds (benzene).

Hence, we want to develop a simple, portable, and low-cost electronic nose, with accurate computational methods for automated CRC non-invasive preliminary screening analysis using deep learning.

Materials and Methods





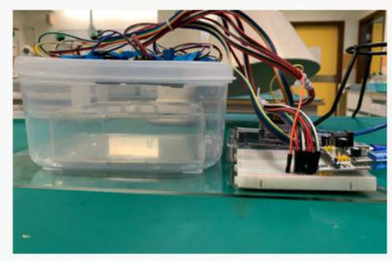
Portable, and low-cost E-nose(~$50):
Contained 10 commonly MQ gas modules which are arranged in a circle shape.
Place a petri dish with sample in the box and release VOCs to the E-nose directly

Step 1

Collecting feces from the University of Macau
12 Healthy rats
6 rats with CRC

Step 2

Culturing fecal samples
Prepared and autoclaved LB broth solution with a concentration of 2.5%.
Mixed the samples from healthy rats in different combinations in a clean centrifuge tube. Added 20 ml of LB broth in each treatment group with label.
Ditto but using CRC samples.
Cultured all samples in an incubator



Step 3

Sampling gases released by the feces via E-nose
Sampling gases released by each feces with E-nose (sampling rate: 0.5Hz)
Sampling each feces around 1 to 3 mistunes
Data cleaning and finally got 70 csv files with total 12,325 entries x 11 columns dataset.

Result and Conclusion

Initial Approach

Feature Extraction

- apply sliding window(k=20) calculate reaction rate & standard deviation
- Normalizing (MinMax scale)

Modeling

- Random Forest (RF) model (max depth = 6 & number of trees = 250)
- Neural Network (NN) (3 hidden layers, 32, 64, 128 neurons respectively, 64 batch size and 15 training epochs)

	Training Set	Validation Set	Testing Set
RF	99.96%	99.82%	80.89%
NN	99.89%	99.96%	78.98%

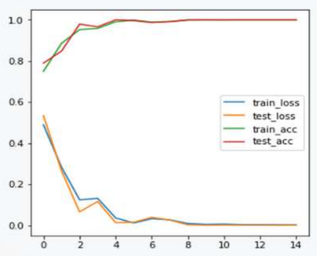
Accuracy of RF/NN on training set, validation set and testing set

We can find that the accuracy of two models of testing set is better than the first approach, with around 80% accuracy, it is not good enough,

Final Approach

Since some gases released by CRC feces are considered as biomarkers and gases react with the sensitive material on the MQ sensor, therefore the reacting rates of different gases are different.

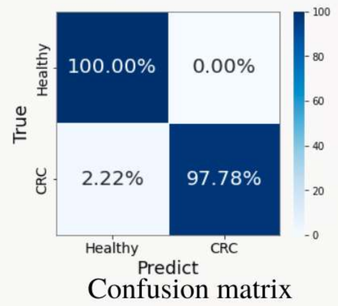
- Appling two convolution layers for considering the co-reaction rate of different MQ sensors



The loss curve of CNN

	CNN
Training	100%
Validating	99.95%
Testing	98.63%

Accuracy of CNN on training set, validation set and testing set



Confusion matrix

The CNN model processed blind-tested using a new testing dataset sampling from another 14 feces, The results showed that accuracy was 98.63%, sensitivity was 97.78%, and specificity was 100%.
Our E-nose shows the possible implications of gases pattern recognition for the diagnosis of CRC. And our e-nose is an advancement over current work, for the following reasons:

- Portable and low cost
- Rapid (around 1~3 mins sampling and screening)
- high accuracy, sensitivity, and specificity