Breath Biopsy Conference 2022

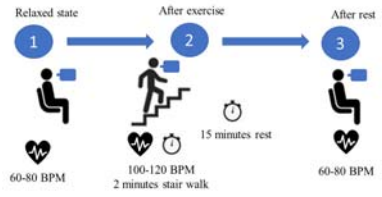
Machine Learning Based Clinical Study of Target Breath Biomarkers Concentration Profile for Diagnosis of Liver Disease

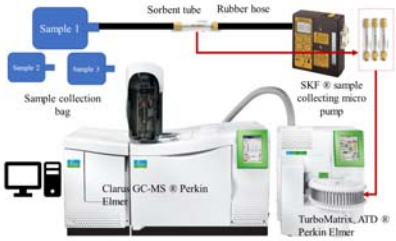
Authors： Rakesh Kumar Patnaik[1], Yu-Chen Lin[1], Ashish Agarwal[1], Ming-Chih Ho[2] and , J. Andrew Yeh[1]
Presenting author： Rakesh Kumar Patnaik
INEMS, National Tsing Hua University, Taiwan[1]
Department of Surgery, NTU Hospital, Taiwan[2]

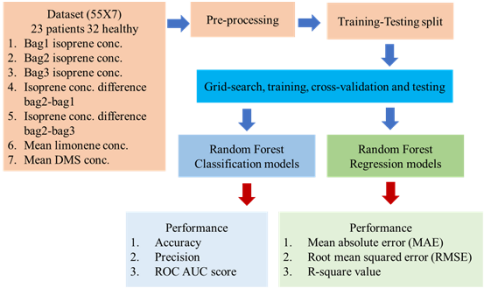
Abstract: Diagnosis at an early stage may reduce the risk of disease progression and improves the mortality rate. Liver function impairment dysregulates the volatile organic compounds (VOCs) that are found in the exhaled breath. The purpose of this work is to distinguish and categorize liver patients (based on Child-Pugh's score, APRI score and MELD score) and healthy controls using isoprene, limonene and dimethyl sulfide as a panel of biomarkers and a machine learning (ML) model.

Introduction: Isoprene is a by-product of the glycolysis cycle and cholesterol synthesis. Liver disease destroys the active mass and downregulates the concentration of isoprene. Limonene is an exogenous compound results from the indigestion of citrus contain diet due to alteration of P450 liver enzymes. Studies show that hepatic dysfunction disturbs the process by affecting the enzymes responsible for H_2S absorption so part of the H_2S is converted to DMS. In the proposed work we not only observe biomarkers concentration but also try to predict scores using breath data and compare it with the real value of scores which are derived from blood test data. A Random Forest classification and regression algorithms are employed to observe the predictions on a designed dataset that comprises the said VOCs concentration.

Method: This study is conducted at National Taiwan University Hospital, Taipei. So far, a total of 55 study subjects (32 controls and 23 patients) were recruited. The collected breath samples were quantified using a GC-MS setup. Then the dataset is used to train and predict using various ML models for both classification and regression.

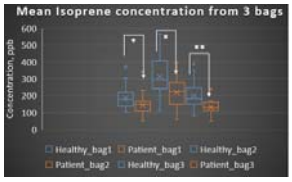


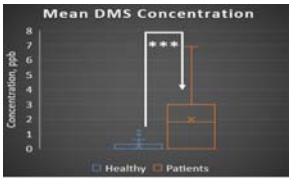




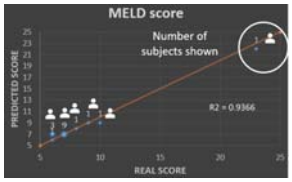
Result

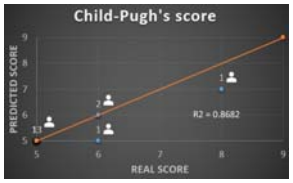
The mean isoprene concentration of patient group is less than controls. The mean concentration of limonene and DMS is more in patients compare to controls.





The comparison of actual and predicted Child-Pugh's, APRI and MELD score (using RF Regression) are shown, and it yields a acceptable R-square value for each target variables.

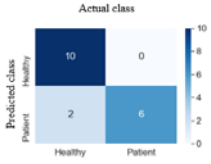




Regression Test results

	Child Pugh's score	APRI score	MELD score
Mean	5.5	0.5	7.5
MAE	0.117	0.088	0.294
RMSE	0.342	0.228	0.542

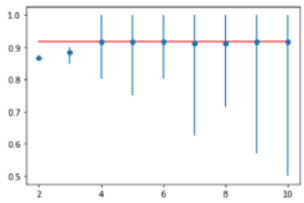
Confusion matrix



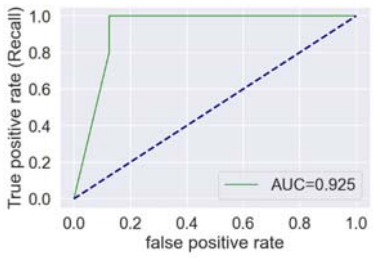
Classification report

Accuracy	88.89%			
	Precision	**Recall**	**F1-score**	**Support**
Patient	1.00	0.75	0.86	8
Healthy	0.83	1.00	0.91	10

Cross-validation result of RF Classification model



Receiver Operating Curve (ROC) of RF Classification model



Conclusion: The results showed a high classification accuracy, AUC and precision in the classification of liver patients and healthy control. The regression result can estimate the clinical scores which imply the concentration of biomarkers varies according to the liver condition. Combining the breath biomarker and machine learning approach increases the opportunities of utilizing the data to facilitate and improvise the existing diagnosis procedure.