Breath acetone classification using XGBoost algorithm for diabetes detection

Anna Paleczek, Dominik Grochala, Artur Rydosz

Institute of Electronics, Faculty of Computer Science, Electronics and Telecommunications, AGH University of Science and Technology, al. A. Mickiewicza 30, 30-059 Krakow, Poland

1. INTRODUCTION

- According to WHO, there are currently around 500 million diabetics in the world.
- There are two main types of diabetes mellitus type 1 (T1DM) and type 2 (T2DM).
- Exhaled air consists of over 3500 Volatile Organic Compounds (VOCs), but a single breath consists of around 500 various VOCs.
- Acetone is a known biomarker of diabetes in the exhaled air.

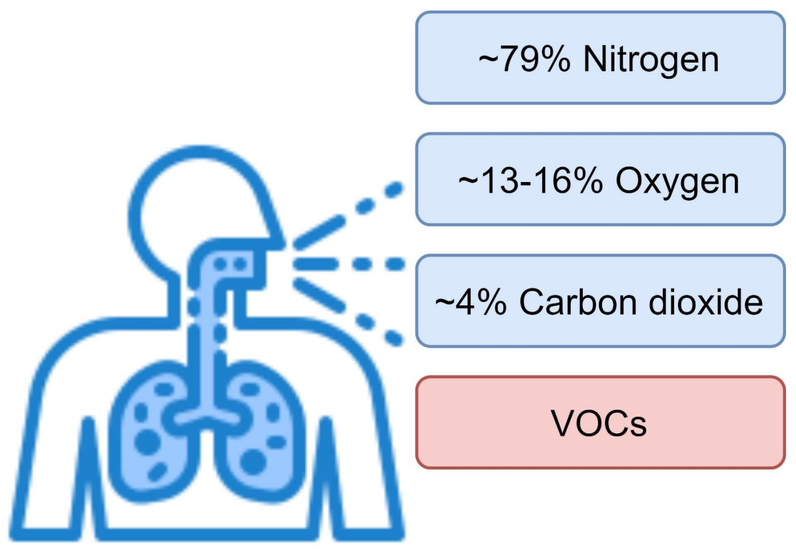

Fig 1. Composition of humans' exhaled air

2. METHODS

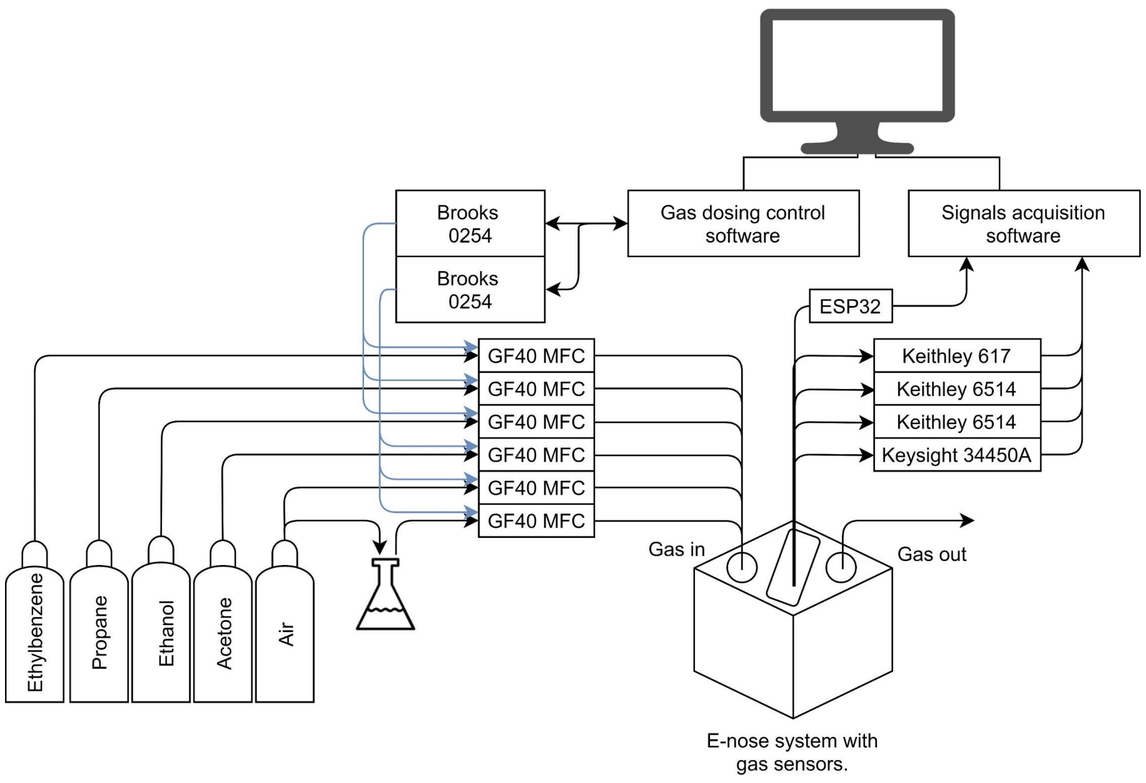

Fig 2. Measurement system

- Sensors used in the measurement system: TGS1820, TGS2620, TGS8100, MICS5524, MQ3, SGP30
- Simulated relative humidity range: 0 - 70%
- „Healthy" and „diabetes" artificial breath classification based on acetone concentration (≥1.5 ppm for diabetes patient).
- Data preprocessing consisted of baseline substraction and calculation of sensor response (S).

$$S = \frac{R_{sensor\ exposed\ to\ target\ gas}}{R_{sensor\ exposed\ to\ pure\ synthetic\ air}}$$

- Classification performed using XGBoost.

3. RESULTS

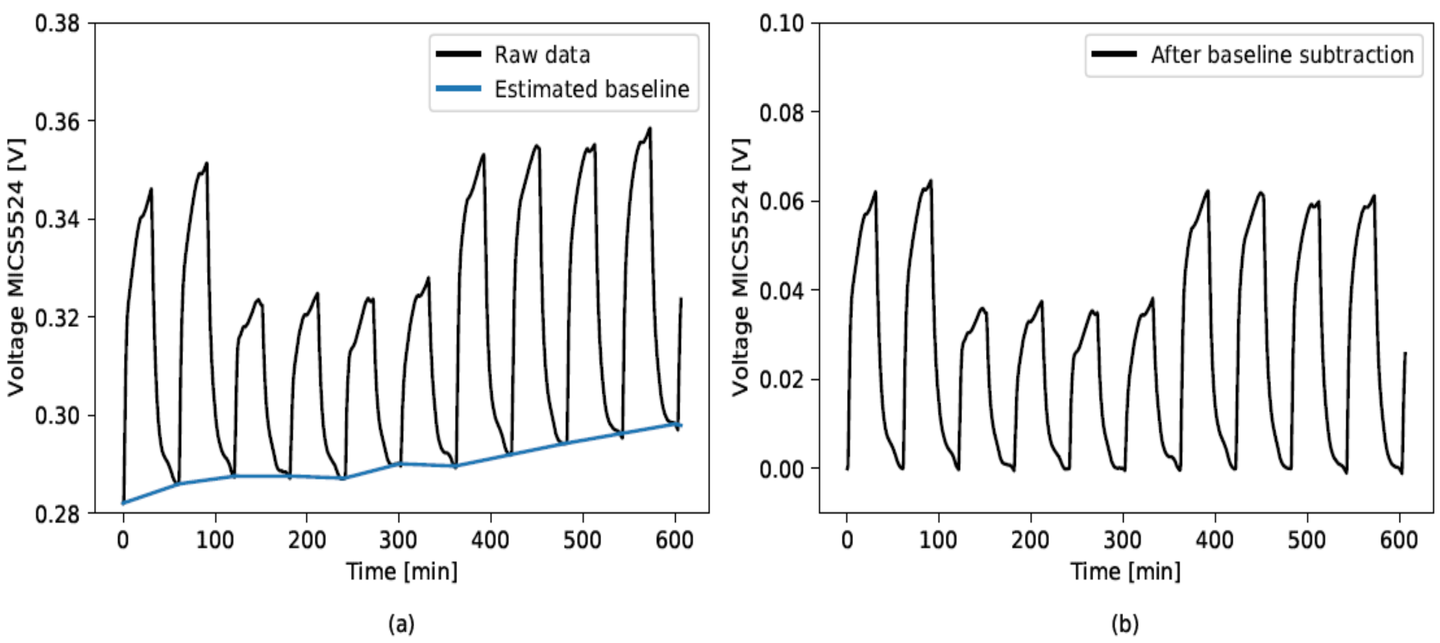

Fig 3. Baseline subtraction. (a) Sensor raw response with fitted baseline; (b) result of the baseline subtraction.

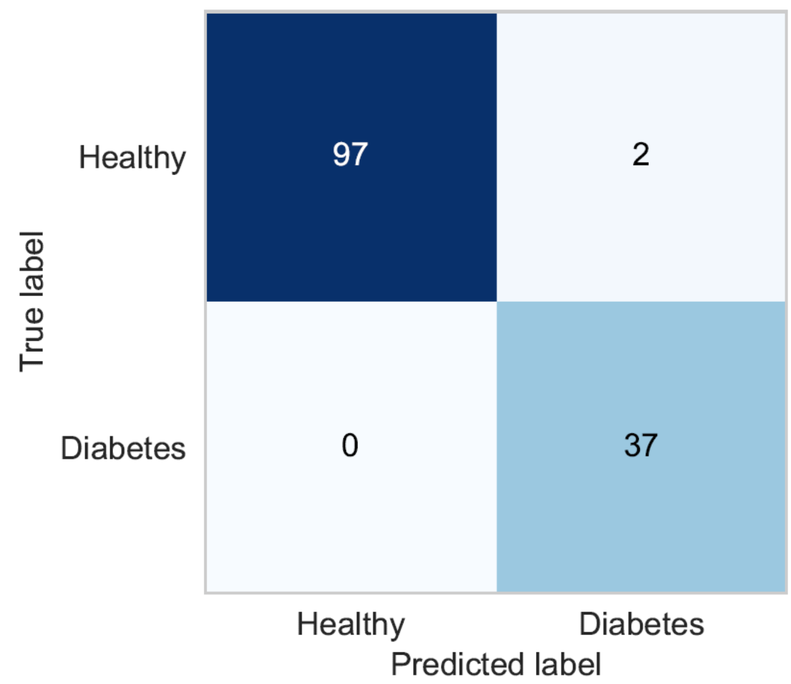

Fig 4. Confusion matrix

4. CONCLUSIONS

Studies have shown that the prepared set of sensors is highly selective for acetone, which is a known biomarker of diabetes, also in various humidities. With the use of the XGBoost algorithm for classification, very good results were obtained: accuracy 99%, recall 100%, F1-score 97.4%. The prepared set of sensors and algorithms can be used in the future for further research on human breathing, and later for the development of a device for non-invasive blood sugar monitoring.