INVESTIGATION OF THE RELATIONSHIP BETWEEN BREATH AND BLOOD GLUCOSE MEASUREMENT AND DEVELOPMENT OF BREATH BLOOD GLUCOSE METER PROTOTYPE WITH MOBILE APPLICATION FOR DIABETIC PATIENTS

2023
MASTER THESIS
BIOMEDICAL ENGINEERING

Taleh BİNNATOV

Thesis Advisors
Prof. Dr. Tamila ANUTGAN
Assist. Prof. Dr. Hakan YILMAZ

INVESTIGATION OF THE RELATIONSHIP BETWEEN BREATH AND BLOOD GLUCOSE MEASUREMENT AND DEVELOPMENT OF BREATH BLOOD GLUCOSE METER PROTOTYPE WITH MOBILE APPLICATION FOR DIABETIC PATIENTS

Taleh BİNNATOV

Thesis Advisors
Prof. Dr. Tamila ANUTGAN
Assist. Prof. Dr. Hakan YILMAZ

T.C.
Karabuk University
Institute of Graduate Programs
Department of Biomedical Engineering
Prepared as
Master Thesis

KARABUK
June 2023

I certify that in my opinion the thesis submitted by Taleh BİNNATOV titled "INVESTIGATION OF THE RELATIONSHIP BETWEEN BREATH AND BLOOD GLUCOSE MEASUREMENT AND DEVELOPMENT OF BREATH BLOOD GLUCOSE METER PROTOTYPE WITH MOBILE APPLICATION FOR DIABETIC PATIENTS" is fully adequate in scope and in quality as a thesis for the degree of Master of Science.

Prof. Dr. Tamila ANUTGAN .....
Thesis Advisor, Department of Medical Engineering

Assist. Prof. Dr. Hakan YILMAZ .....
2nd Advisor, Department of Medical Engineering

This thesis is accepted by the examining committee with a unanimous vote in the Department of Biomedical Engineering as a Master of Science thesis. 23/06/2023

Examining Committee Members (Institutions)	Signature
Chairman : Prof. Dr. Kadir GÖKŞEN (DÜ)	.....
Member : Prof. Dr. Tamila ANUTGAN (KBÜ)	.....
Member : Assist. Prof. Dr. Hakan YILMAZ (KBÜ)	.....
Member : Assoc. Prof. Dr. Hacı Mehmet KAYILI (KBÜ)	.....
Member : Assoc. Prof. Dr. Daver ALİ (KBÜ)	.....

The degree of Master of Science by the thesis submitted is approved by the Administrative Board of the Institute of Graduate Programs, Karabuk University.

Prof. Dr. Müslüm KUZU .....
Director of the Institute of Graduate Programs

"I declare that all the information within this thesis has been gathered and presented in accordance with academic regulations and ethical principles and I have according to the requirements of these regulations and principles cited all those which do not originate in this work as well."

Taleh BİNNATOV

ABSTRACT

M. Sc. Thesis

INVESTIGATION OF THE RELATIONSHIP BETWEEN BREATH AND BLOOD GLUCOSE MEASUREMENT AND DEVELOPMENT OF BREATH BLOOD GLUCOSE METER PROTOTYPE WITH MOBILE APPLICATION FOR DIABETIC PATIENTS

Taleh BİNNATOV

Karabuk University
Institute of Graduate Programs
The Department of Biomedical Engineering

Thesis Advisor:
Prof. Dr. Tamila ANUTGAN
2nd Advisor
Assist. Prof. Dr. Hakan YILMAZ
June 2023, 62 pages

Diabetes Mellitus (DM), characterized by high blood glucose (BG) levels, can be controlled if monitored continuously. BG is commonly followed via a home-based glucometer, but due to its invasiveness patients reduce the frequency of the BG measurements. Non-invasive BG measurement methods proposed in the literature usually have insufficient data, include complex systems and are not comprehensively discussed because some parameters related to patients are not considered. In this thesis, a cheap, reliable, reproducible, accurate breath blood glucose meter (BBGM) for non-invasive BG measurement was developed.

Firstly, 5 commercial metal oxide semiconductor (MOS) gas sensors were tested to determine the most suitable to be used in BBGM.

Then in the preliminary study, BA data via BBGM and BG data via home-based glucometer were collected from non-diabetic (ND) and DM volunteers. After training the collected 141 data with 3 different machine learning (ML) methods, the most successful method gave 0.89 R^2 score.

Afterwards, in the study conducted in the clinical environment, BA data were collected by BBGM, while BG data were extracted from venous blood. As a result of the training of the collected 415 data with 4 different ML methods, the most successful method resulted in 0.986 R^2 score. The effect of the biological parameters of the volunteers on the training was evaluated.

It was shown that the high training scores can be obtained for both preliminary and clinical environment studies by only measuring BA values via BBGM and using some biological parameters to estimate BG values with acceptable error.

Key Words : Blood glucose, Diabetes Mellitus, Breath acetone, MOS gas sensor, Machine learning, Breath blood glucose meter

Science Code : 92505

ÖZET

Yüksek Lisans Tezi

NEFES İLE KAN ŞEKERİ ÖLÇÜMÜ ARASINDAKİ İLİŞKİNİN İNCELENMESİ VE DİYABET HASTALARI İÇİN MOBİL UYGULAMALI NEFESTEN KAN ŞEKERİ ÖLÇÜM CİHAZI PROTOTİPİNİN GELİŞTİRİLMESİ

Taleh BİNNATOV

Karabük Üniversitesi
Lisansüstü Eğitim Enstitüsü
Biyomedikal Mühendisliği Anabilim Dalı

Tez Danışmanı
Prof. Dr. Tamila ANUTGAN
2. Danışman
Dr. Öğr. Üyesi Hakan YILMAZ
Haziran 2023, 62 sayfa

Yüksek kan glukozu (BG) seviyeleri ile karakterize olan Diabetes Mellitus (DM) sürekli izlendiği takdirde kontrol altına alınabilir. BG genellikle ev-tipi glukometre ile takip edilmektedir, fakat yöntemin invaziv olması sebebiyle hastalar BG ölçüm sıklığını azaltmaktadır. Literatürde önerilen non-invaziv BG ölçüm yöntemleri genellikle yetersiz sayıda veriye sahip, karmaşık sistemler içerir ve hastalara ait bazı parametreler dikkate alınmadığından kapsamlı bir şekilde ele alınmamıştır. Bu tezde ise non-invaziv BG ölçümü için ucuz, güvenilir, tekrarlanabilirliği ve doğruluğu yüksek bir nefesten kan glukozu ölçüm cihazı (BBGM) geliştirilmiştir.

İlk olarak, BBGM cihazında kullanıma en uygun sensör için 5 adet ticari metal oksit yarıiletken (MOS) gaz sensörünün testleri yapılmıştır.

Ardından, yapılan ön çalışmada diyabetik olmayan (ND) ve DM gönüllülerinden BBGM cihazı ile BA verileri ve ev-tipi glukometre ile BG verileri toplanmıştır. Toplanan 141 verinin 3 farklı makine öğrenmesi (ML) yöntemi ile eğitimi sonrası en başarılı yöntemle 0,89 R^2 skoru elde edilmiştir.

Daha sonra klinik ortamda yapılan çalışmada BA verileri BBGM ile BG değerleri ise venöz kandan toplanmıştır. Toplanan 415 verinin 4 farklı ML yöntemi ile eğitimi sonucunda en başarılı yöntemle 0,986 R^2 skoru elde edilmiştir. Gönüllülere ait biyolojik parametrelerin eğitime olan etkisi değerlendirilmiştir.

Hem ön çalışma hem de klinik çalışma için elde edilen yüksek eğitim skorlarının sadece BBGM ile ölçülen BA değerleri ve hastaların bazı parametreleri ile BG düzeylerinin kabul edilebilir hata içinde tahmin edilebileceği gösterilmiştir.

Anahtar Kelimeler : Kan glukozu, Diabetes Mellitus, Nefesteki aseton, MOS gaz sensörü, Makine öğrenmesi, Nefesten kan glukozu ölçüm cihazı

Bilim Kodu : 92505

ACKNOWLEDGMENT

I would like to express my endless thanks to my advisor, respected academician Prof. Dr. Tamila ANUTGAN, whose knowledge and experience I benefited from at every stage of this thesis and who did not hesitate to support me in every subject. Also, I would like to greatly thank my second advisor, respected academician Assist. Prof. Dr. Hakan YILMAZ, for all his support and information provided within the scope of the thesis.

This thesis work was supported by Karabuk University Scientific Research Project Coordination Unit under the Project no: KBÜBAP-21-YL-098.

I also would like to thank TÜBİTAK for providing financial support for the clinical study of this thesis under the Project no: 122E006.

I would like to thank Kastamonu University Clinical Research Ethics Committee for their support in obtaining the ethics committee document used within the scope of clinical study of this thesis.

I would like to thank Assoc. Prof. Dr. Nurhayat ÖZKAN SEVENCAN and Fatime ŞAHİN for their support during clinical data collection.

Finally, I would like to thank my beloved family for their unwavering support throughout this process.

CONTENTS

Page

Page

Page

LIST OF FIGURES

Page

Page

LIST OF TABLES

Page

INDEX OF SYMBOLS AND ABBREVITIONS

SYMBOLS

R^2 : coefficient of determination

β : beta

C_3H_6O : acetone

$C_4H_6O_3$: acetoacetate

R : correlation factor

R_s : sensor resistance

R_l : load resistance

V_{Rl} : voltage of load resistance

V_c : circuit voltage

R_0 : sensor's resistance in fresh air

ABBREVIATIONS

DM : Diabetes Mellitus
BG : Blood Glucose
BBGM : Breath Blood Glucose Meter
MOS : Metal Oxide Semiconductor
ND : Nondiabetic People
ML : Machine Learning
DKA : Diabetic Ketoacidosis (DKA)
T1D : Type 1 Diabetes
T2D : Type 2 Diabetes
SMBG : Self-Monitoring of Blood Glucose
VOC : Volatile Organic Compounds
3HB : 3-B-Hydroxybutyrate
BA : Breath Acetone
PPM : Parts Per Million
GC : Gas Chromatography
GC-MS : Gas Chromatography Mass Spectrometry
OGTT : Oral Glucose Tolerance Test
MLR : Multiple Linear Regression
HS-SPME : Headspace Solid-Phase Microextraction
CRDS : Cavity Ring-Down Spectroscopy
MEMS : Micro-Electromechanical Systems
QCM : Quartz Crystal Microbalance
NMR : Nuclear Magnetic Resonance
ANN : Artificial Neural Network
HbA1c : Hemoglobin A1C
SVOR : Support Vector Ordinal Regression
SVM : Support Vector Machine
SRC : Sparse Representation-Based Classification
KNN : K-Nearest Neighbor
CNN : Convolutional Neural Network
LTCC : Low Temperature Co-Fired Ceramics

GND : Grounding

VCC : Power Supply

AQS : Air Quality Sensor V1.3 Module

LoA : Limits of Agreement

BMI : Body Mass Index

KS : Kolmogorov-Smirnov

ECDF : Empirical Cumulative Distribution Function

CDF : Cumulative Distribution Function

D : Maximum Difference

GB : Gradient Boosting

MAE : Mean Absolute Error

MAPE : Mean Absolute Percentage Error

CHAPTER 1

INTRODUCTION

It is well known that when insulin, which enables the glucose in the blood to enter the cell and turn it into the energy, is not produced enough and/or the cells do not respond to insulin well, the blood glucose (BG) level rises. Such high BG characterizes a rapidly spreading chronic metabolic disease, Diabetes Mellitus (DM) [1]. According to the International Diabetes Federation Atlas 2019 data, the number of people with DM is predicted to reach 578 million by 2030 [2]. The most important metabolic complications of DM caused by a severe hyperglycemic state are diabetic ketoacidosis (DKA) and/or hyperglycemic hyperosmolar syndromes. In addition, in the long term, DM may cause damage to the blood vessels in the retina (diabetic retinopathy), to the kidney at the microvascular level (nephropathy), DM may lead to the diabetic foot ulcer, the peripheral neuropathy (amputation and Charcot joint disease) and the autonomic neuropathy (damage in the cardiovascular, gastrointestinal, genitourinary and neurovascular systems). Also, people with DM exhibit more abnormalities of lipoprotein metabolism and hypertension [1].

As proposed by the American Diabetes Association, it is still accepted to identify type 1 diabetes (T1D) and type 2 diabetes (T2D) as common classes of DM. While T1D occurs due to the destruction of β cells that secrete insulin in the pancreas, T2D occurs mainly due to the insulin resistance mostly caused by obesity [3]. In either case, glycemic control is important for DM management. Therefore, people with DM should monitor their BG levels several times (2-5) per day by self-monitoring of BG (SMBG) technique via home-based glucometer measurements [4]. The recommended number of measurements for T2D patients is at least 2 times a day [5], whereas for T1D is 6-10 times a day [6]. Since a blood sample is taken from the finger during each BG level measurement, DM patients do not reach this recommended number of measurements

in practice. Indeed, there are situations when T2D patients do not check their BG level even for months because of the invasiveness of this SMBG procedure.

To overcome the invasiveness problem of the home-based glucometer, different non-invasive BG level measurement techniques were proposed up to now and the promising one is to estimate BG level from the breath because the BG values are found to be related to the volatile organic compounds (VOC) in the exhaled breath, specifically to the breath ethanol and acetone concentration [7]. Since the breath ethanol concentration is much lower than the acetone one [7], BG is usually correlated solely with the acetone concentration in the breath. This correlation is based on the fact that the rise of BG levels in DM patients starts when their body can not use glucose as energy and as a response to this energy deficiency their liver burns fat as an alternative fuel source. During this fat-burning, ketone bodies are formed: acetone (C_3H_6O), acetoacetate ($C_4H_6O_3$) and 3-β-hydroxybutyrate (3HB). Since acetone is more volatile than other ketones, it is present in the exhaled breath [8]. Acetone smell in the breath is observed also for nondiabetic people (ND) who are hunger and/or exercising for a long time so that they experience fat burning process [9]. There are already commercially available non-invasive portable small-sized ketone breath analyzers replacing urine strips or blood ketone invasive devices [10, 11]. These ketone breath analyzers provide the users of a ketogenic diet with the level of their ketosis state, but it is a rough classification of the measured acetone concentration in the exhaled breath. On the other hand, for DM patients the research is based on the delicate correlation of the breath acetone (BA) concentration with the BG values, which is a much more complex problem. In addition, DM patients (mostly T1D) sometimes may experience very fast fat burning, which rises their BG level and at the same time ketones, so that the blood becomes acidic, i.e. DKA process [12]. As a result, during DKA, a strong acetone smell is developed in a breath with an acetone concentration even much higher than for the ND individuals undergoing starvation ketosis [11]. On the other hand, T2D patients during hunger, exercising, and/or ketogenic diet may experience fat burning just like an ND individual [13], so the two different reasons behind their fat-burning can overlap leading to a more complex relationship between their exhaled acetone concentration and BG level. Therefore, in order to comprehensively examine the relationship between BA and BG, it is necessary to consider the patient's satiety status, DM class and biological parameters as well as

BA. With this motivation in this thesis, it is aimed to estimate the amount of BG according to the BA concentration by taking into account the patient's parameters.

In the second part of this thesis, the literature review for studies examining the relationship between BA and BG is summarized. In the third chapter, experimental device construction, breath tests and comparison of sensors are explained. In the fourth chapter, the establishment of breath blood glucose meter (BBGM) is described. In the fifth chapter of the thesis, the preliminary study with the prototype device is explained. In the sixth chapter, the experimental work performed with the prototype device in the clinical envitonment is presented. In the seventh chapter of the thesis, the results of the preliminary and clinical studies are summarized.

CHAPTER 2

LITERATURE REVIEW

Since the BA is of the order of parts per million (ppm), it is not easy to detect it accurately. Trotter et al. presented in 1971 the first quantitative detection of acetone from the exhaled breath using the gas chromatography (GC) technique and found a positive linear relationship between GC-measured BA and plasma acetone (standard chemical method) for 9 patients with DKA [14]. As a result, BA was shown to reffect the Blood acetone which is turn affected by BG.

In the other study, where GC Mass Spectrometry (GC-MS) technique was used, the breath measurements were taken within 120 minutes from the volunteers subjected to Oral Glucose Tolerance Test (OGTT). By multiple linear regression (MLR) method it was found that the BA concentration was associated with BG with the correlation factor (R) 0.95 [7]. In another work, the amount of BA in 30 T2D fasted volunteers and 28 ND fasted volunteers was analyzed by Headspace Solid-Phase Microextraction (HS-SPME) and GC-MS [15]. As a result, these GC methods are concluded to be sensitive but are of high cost, have low portability due to heavy equipment, can not be used for real-time analysis and their usage is difficult.

In addition to the GC methods, various other techniques were used to measure BA and determine its relationship to BG: cavity ring-down spectroscopy (CRDS) based breath analysis system [16-17], CMOS Micro-electromechanical system-based (MEMS) system [18], an electronic nose based on a Quartz Crystal Microbalance (QCM) sensor system [19], Nuclear Magnetic Resonance (NMR) technique [20].

To examine the relationship between BA and BG, Sun et al. developed a CRDS-based BA analyzer [16]. 20 T1D patients and 5 ND volunteers who fasted for 8 hours blew into approximately 100 mL breathing gas bags for 30 days. A total 600 measurements

were taken and analyzed with the LaserBreath-001 device based on the CRDS technique. In addition, daily continuous monitoring of subjects' BG and blood ketones was performed. For 20 outpatient T1D patients, the range of BA concentrations ranged from 0.5 to 49.5 ppm with an average of 3.1 ppm, while in ND subjects BA ranged from 0.2 to 6.2 ppm with an average of 1.9 ppm. As a result it was reported that no correlation was observed between BA and BG [16].

On the other hand, Rabih et al. proposed MEMS sensor device for non-invasive screening of DM [18]. The sensing mechanism of the sensor depends on the change in mass of the moving plate after the adsorption/desorption of acetone on the surface of the sensing layer. After the frequency change, the signal is digitized by the measuring unit. Synthetic acetone vapor was used experimentally in that study. It was reported that the sensor had a linear relationship between 0.5-5 ppm for acetone measurement and R of 0.93 was calculated. In addition, cross-sensitivity testing of the sensor was performed for 5 ppm 2-propanol and methanol. However, the sensor was 24% and 33% less responsive to 2-propanol and methanol, respectively, compared to acetone. Measurement with this method requires a multi-component system, extra power supply, and multiprocessing. In addition, the response time of the sensor (7 minutes) and reset time (1 minute) last longer than other methods. The shift of the baseline of the output voltage of the system, which working principle depends on the frequency, affects the measurement accuracy [18].

In another study, it was aimed to determine human BG and HbA1c levels from the exhaled breath as a non-invasive method with the help of an electronic nose system based on QCM sensors [19]. Composed of a chemical sensing mechanism such as a set of electronic sensors and a pattern recognition mechanism such as a neural network, electronic noses can identify many different physical and chemical compounds and odors by converting them into quantities like frequency and voltage. Samples taken by 30 patients blowing into a breathing balloon were examined with 9 QCM sensors. QCM sensors naturally generate a curve for each breath sample. The frequency differences between the peak and trough in this curve indicate the density of gas molecules. Each sensor's response to the sample is different. All sensor responses were analyzed for each patient, and the sensors with the most appropriate response were

used to train the neural network. In that work, radial basis function neural networks were preferred because they give better educational results in clustered data sets and perform better in function convergence problems. Thanks to the radial basis function, in practice, by changing the mesh weights and continuously spreading the coefficient factor, the most suitable constants are found, so that the most suitable mesh for the data group is created. In the program written in Matlab, the data was used with the desired clustering method with the K-fold cross-validation test and the radial function diffusion range was changed randomly and continuously. The number of neurons in the hidden layers was limited to 50 due to a lack of data. By subtracting the obtained artificial neural network (ANN) results from the normal, BG and hemoglobin A1C (HbA1c) estimation values were found. Accordingly, training results showed an accuracy rate of 74.76 % for the glucose parameter and an accuracy rate of 83.03 % for HbA1c [19]. However the system included costly multi-process complex systems.

Luaibi et al. reported in their study that NMR is a promising technique for non-invasive BG measurement [20]. In that study, the results obtained from 2 human subjects with an error rate of 3 % to 14 % showed applicability of the technique. However, the number of the tests should be increased for the system accuracy. In addition, the system has disadvantages such as being expensive and requiring different calibrations for each sample [20].

In summary, some of the above described methods and systems are costly, difficult to use (GC), some require a calibration procedure (CRDS), some are easily perishable and have limited sensitivity (MEMS), and some have costly multi-process complex systems (QCM) [21]. In addition, some of these studies reported no or weak association between BG and BA [16-17].

On the other hand, it is more appropriate to use MOS based gas sensors with high sensitivity, low cost and simple measurement for the same purpose [22-27].

In one of the BG and BA classification studies, a sensor system consisting of 12 MOS sensors (Figaro) was used for the collection and analysis of breath samples from DM patients. Respiratory data were obtained using breathing balloons from a total of 192

fasting DM patients, 123 T2D patients and the rest with T1D and other types of DM. Afterward, the fasting BG levels of the subjects were measured. BA according to the hunger BG levels, the data were classified into 4 categories as “well-controlled”, “slightly controlled”, “poorly controlled” and “uncontrolled” using the support vector ordinal regression (SVOR) method. In addition to SVOR, support vector machine (SVM) and sparse representation-based classification (SRC) were also used as classifiers. As a result, it was found that SVOR (68.66 %) outperformed SRC and SVM in solving sequential problems from a medical point of view [22]. Unfortunately, the result showed that the classification accuracy for the clinical application of that system was too low [22].

In another classification study, a new breath analysis system based on the e-nose principle was proposed for the diagnosis of DM. Yan et al. used the k-nearest neighbor (KNN) classification in 2 categories with more data (294 ND+294 DM) [23]. An array of 7 chemical MOS sensors was used to acquire breath data. Data were collected from DM volunteers and ND individuals. Half of the collected data (equal number of data for both categories) was used for classification training and the other half for testing. The results showed that the system can distinguish ND and DM samples with simple feature extraction and classification algorithms. It was reported that the uncontrolled DM samples were different from ND human samples and that the controlled DM obesity did not affect respiratory values [23].

Lekha et al. proposed the use of a 1-dimensional (1D) modified convolutional neural network (CNN) algorithm, which combined feature extraction and classification techniques of breath signals for DM classification [24]. A total of 25 signals were collected from ND, T1D, and T2D subjects with MQ3 and MQ5 MOS sensors. 15 examples were used for training and 10 examples for testing the classifier. It was reported that the proposed approach significantly reduced the limitations associated with using these techniques separately, thereby further improving the performance of the classifier [24]. It should be noted that in these classification studies, BG values were not obtained after training, they only reported the ranges of BG levels. It is known that training should be given in the classification study to achieve high accuracy.

In another study in order to detect exhaled acetone concentrations below the detection limit of gas sensors, a microsystem in low temperature co-fired ceramics (LTCC) technology was introduced [25]. An array of 6 MOS gas sensors (Figaro) was used as the detector unit for acetone detection. First BA and BG measurements of 8 ND volunteers and 10 T1D volunteers included in the study were taken during 8 hours of fasting. Afterwards, BA and BG values of the subjects were measured again with 1-hour intervals for 7 hours after the meal. As a result, it was observed that exhaled acetone levels were higher in DM volunteers than in ND patients, so the fasting period affected acetone levels in DM volunteers and a linear relationship was found only for some patients [25].

Thati et al. used an acetone-sensitive tin oxide sensor (Figaro-TGS822) to detect the acetone concentration in the exhaled breath [26]. The sensing principle of the sensor is based on changes in electrical conductivity due to reactions between the oxygen on the sensor surface and the measured gas. The resistance of the sensor varies depending on the amount of acetone present and can be detected with a potential divider circuit. In that study 30 volunteers blew into a 215 cm^3 chamber containing a temperature and humidity sensor as well as a gas sensor. Then, the volunteer's BG value was measured with a home-based glucometer (Accucheck). Data from sensors and glucometer were trained using an ANN. Using the proposed system after training, BG was estimated to be within a margin of error of ±7.5 mg/dL [26].

In a recent promising study, 100 breath data from 100 DM subjects were collected by a linear regression classifier with only one feature was used to train these data, resulting in a correlation score of 0.92 [27].

As a result of the literature review it was concluded that the use of the commercial MOS-based sensors for BA mrasurements is pratical and non-expensive. In addition, it was concluded that the DM class, satiety state & other biological parameters should be added in the training of the data to obtain a healthy correlation between BA and BG values.

CHAPTER 3

EXAMINATION OF BREATH ACETONE BY MASS SPECTROSCOPY, SETTING UP A DEVICE COMPRISING DIFFERENT AIR QUALITY SENSORS, CONDUCTING BREATH TESTS AND ANALYZING EXPERIMENTAL DATA

3.1. MASS SPECTROSCOPY EXPERIMENTS

Before conducting experiments of BA detection by the MOS-based sensors the detectability of acetone in exhaled breath was examined by mass spectroscopy device (HIDEN Ins., Figure 3.1).

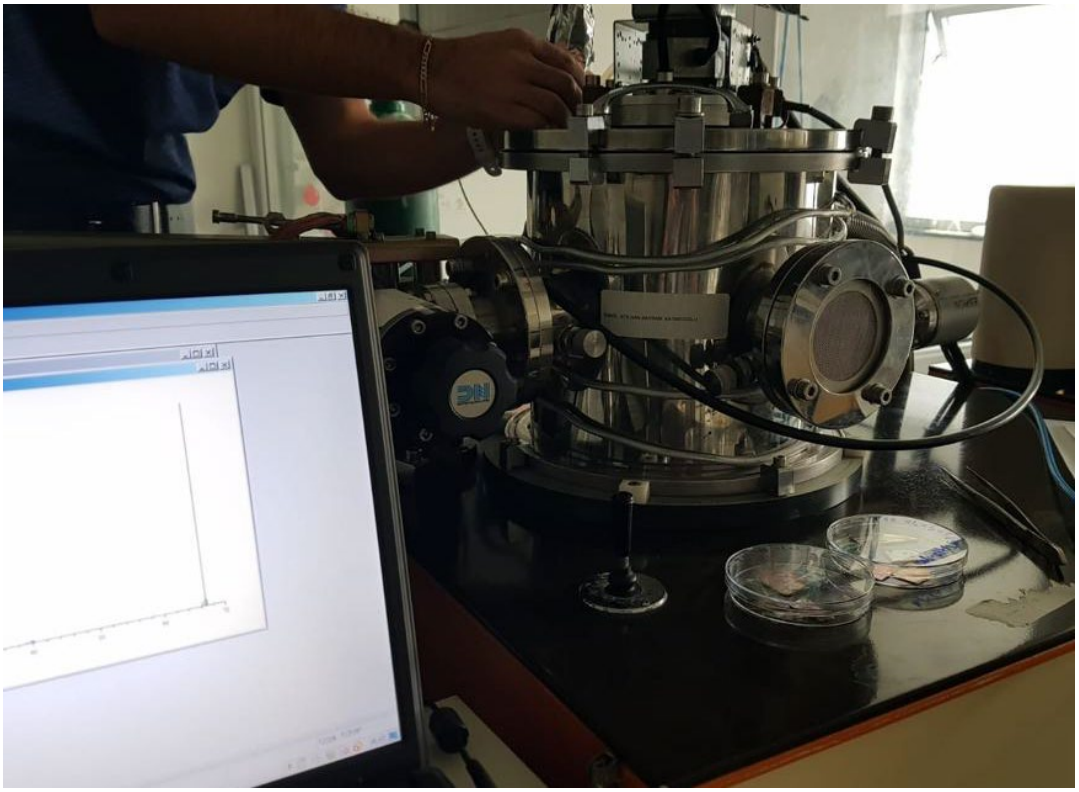


Figure 3.1. An experiment of the detection of acetone from breath by mass spectroscopy device.

Firstly, background spectrum of the vacuum chamber before blowing in it was measured Figure 3.2 (a). Then the spectrum was again obtained Figure 3.2 (b) after blowing into the vacuum chamber. As can be seen from these graphs, acetone base peak (amu=43) [28] and carbon dioxide base peak (amu=44) [29] appeared after blowing together with the background spectrum.

This indicates that the acetone is present in an exhaled breath with the detectability level close to the carbon dioxide. Thanks to this relatively high amount of the BA most of the MOS-based sensors are able to measure it after blowing on them.

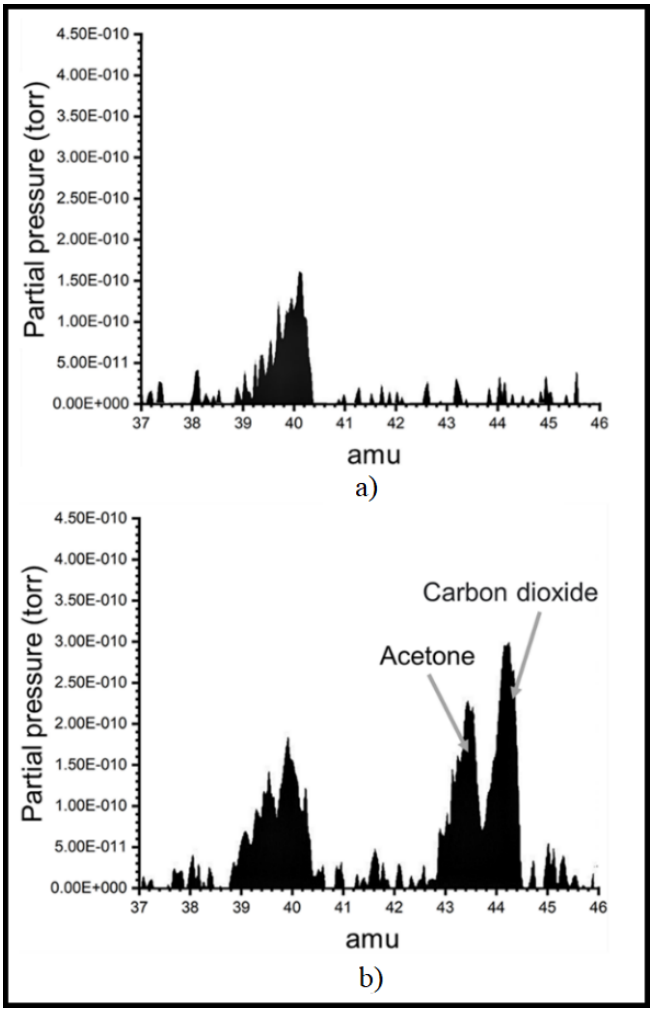


Figure 3.2. a) Background value before blowing, b) Obtaining acetone and carbon dioxide peaks after blowing

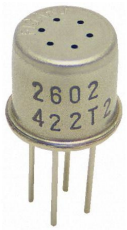
3.2. FEATURES OF USED MOS GAS SENSORS

5 MOS gas sensors sensitive to volatile organic compounds such as acetone were selected to find the most suitable sensor to be used in the determination of acetone levels in the breath. MOS is used to detect various gases in the environment and to measure gas concentration. The concept of MOS was first discovered in the 1960s by researchers who observed changes in the electrical conductivity of metal oxide materials when exposed to different gases [30]. The detection mechanism of MOS gas sensors is due to the reactions occurring between the target gas and the sensor surface. The sensor contains the heater and the MOS material on the ceramic substrate. The contact of gases with the surface of MOS gas sensors, which have chemical resistance features, increases the conductivity of the sensor, so the change in conductivity values is converted to the concentration of the target gas. Contact with the surface of the gases allows the MOS to remove more electrons in the conduction band. Therefore, the conductivity of MOS depends on the concentration of the target gas [31-33].

Grove Air Quality Sensor v1.3 module (AQS) with Mp503 sensor, MQ3, MQ135, TGS822, TGS2602 sensors were selected to be used in the test device. The specifications, applications and sensitivity features of these sensors are given in Table 3.1.

Table 3.1. Specifications, applications and sensitivity features of the AQS, MQ3, MQ135, TGS822 and TGS2602 gas sensors used in the test device.

Sensors	Specifications	Application	Sensitivity Features
AQS [32,33] 	• High sensitivity • Low power consumption • Long lifetime • Good stability • Quick response • Automatic preheat function • Warm-up between 10-180 seconds	• Air cleaner system • Clean air system • Smart integrated ceiling • Air quality detector • Fan • Air conditioning	Detection of 1-50 ppm acetone, toluene, formaldehyde, methyl, alcohol, breath (within the AQS module)

MQ3 [34] 	• High sensitivity to alcohol • Low sensitivity to gasoline • Quick answer • High sensitivity • Stable and long lasting • Simple driver circuit	• Alcohol control devices • Alcohol meters	Detection of alcohol, gasoline, methyl, hexane, LPG, carbon monoxide between 1-100 ppm
MQ135 [35] 	• Quick answer • High sensitivity • Stable and long lasting • Simple driver circuit	• Air quality control systems	Detection of carbon monoxide, carbon dioxide, alcohol, methane, toluene, acetone between 10-200 ppm
TGS822 [36] 	• High sensitivity to VOCs • High stability and reliability over a long period of time • Long lifetime • Uses a simple electrical circuit • Wide detection scope	• Breath alcohol detectors • Gas leak detectors/alarms • Solvent detectors for semiconductors	Detection of 50-5000 ppm methane, carbon monoxide, isobutane, n-hexane, benzene, ethanol, acetone
TGS2602 [37] 	• High sensitivity to VOCs, gaseous air pollutants, odorous gases • low power consumption • Long lifetime • Uses simple electrical circuit • Small-sized	• Air purifiers • Ventilation control • Air quality monitors • VOC monitors • Odor monitors	Detection of hydrogen, ammonia, ethanol, toluene between 10-30 ppm, Detection of hydrogen sulfide between 0.1-3 ppm

Among the selected gas sensors MP503 is a gas sensor with high sensitivity against acetone, toluene, formaldehyde, methane, and alcohol combinations between 1-50 ppm. In addition, the availability of this sensor for the breath detection within the AQS

module has been reported [32]. The most prominent features of the MP503 sensor are fast warming up, fast resetting, good stability, quick response, low power consumption and long service life. The sensor is reported to have been calibrated prior to shipment. Each time the sensor is connected to power, preheating is performed between 10-180 seconds with the automatic preheating function. Preheating is important as it affects the correct measurement of the sensor. MP503 is used in air cleaner and fresh air systems, smartly integrated ceilings, air quality detectors, ventilators and air conditioners [33]. As it concluded later, in this thesis BBGM device uses AQS with an MP503 sensor.

In addition to the MP503 sensor 2 gas sensors of the MQ series were used in the test device: MQ3 and MQ135. MQ3 is a fast-responding, stable and long-lasting gas sensor with a simple driver circuit that can detect 1-100 ppm alcohol, gasoline, methane, hexane, LPG and carbon monoxide compounds [34]. The sensitivity of this sensor is high to alcohol and low to gasoline. MQ3 is widely used in alcohol control devices and breathalyzers [34]. MQ135 sensor is sensitive to combinations of carbon monoxide, carbon dioxide, alcohol, methane, tolen and acetone in the range of 10-200 ppm [35]. Quick response, stable and long lifetime of this sensor with a simple driver circuit are important features. MQ135 is used in air quality control systems for buildings and offices.

The last 2 sensors used in the test device study are TGS822 and TGS2602 from the Figaro TGS series. TGS822 can detect and measure methane, carbon monoxide, isobutane, n-hexane, benzene, ethanol and acetone in the range of 50-5000 ppm [36]. The sensor's high sensitivity to VOCs and wide detection scope are its most important features. It has been reported that the TGS822 sensor, which has a simple electrical circuit, has a long service life, measures high stability and high reliability throughout the period. This sensor is used as breath alcohol detectors, gas leak detectors/alarms and solvent detectors for semiconductors [36]. TGS2602 sensor detects hydrogen between 10-30 ppm, ammonia, ethanol, toluene and hydrogen sulfide between 0.1-3 ppm [37]. TGS2602, which has a simple electrical circuit, shows high sensitivity to VOCs, gaseous air pollutants and odorous gases. Low power consumption, long lifetime and small dimensions are other important features of this sensor. The

TGS2602 sensor has areas of use such as air purification systems, ventilation control, air quality monitors, VOC monitors and odor monitors [37].

3.3. CIRCUIT DESIGN FOR THE TEST DEVICE

In the test device Arduino Nano integrated development board was used to obtain the 5 gas sensor values and display them on the computer. Connections of the 5 sensors with Arduino were made. Because the AQS, MQ3 and MQ135 sensors used are in the form of modules, grounding (GND), power supply (VCC-5V) and analog inputs were connected. Since these sensors have internal resistance (AQS-10K Ohm, MQ3-200K Ohm, MQ135-20K Ohm), no resistance was added to the circuit [32, 34, 35]. A 10K Ohm resistor was added in the circuit of TGS822 and TGS2602 gas sensors. The design of the test device circuit design is shown in Figure 3.3 and its constructed circuit is provided in Figure 3.4.

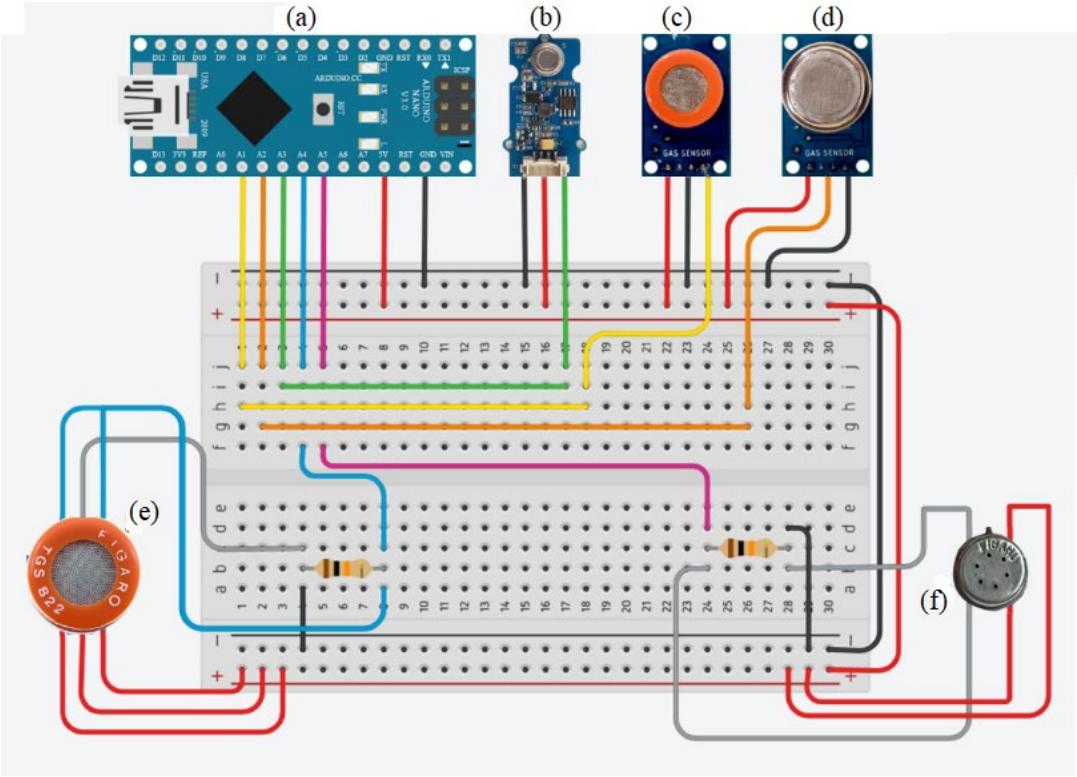


Figure 3.3. The design of the test device circuit: a) Arduino Nano, b) AQS, c) MQ3, d) MQ135, e) TGS, f) TGS2602, Resistor-10k Ohm.

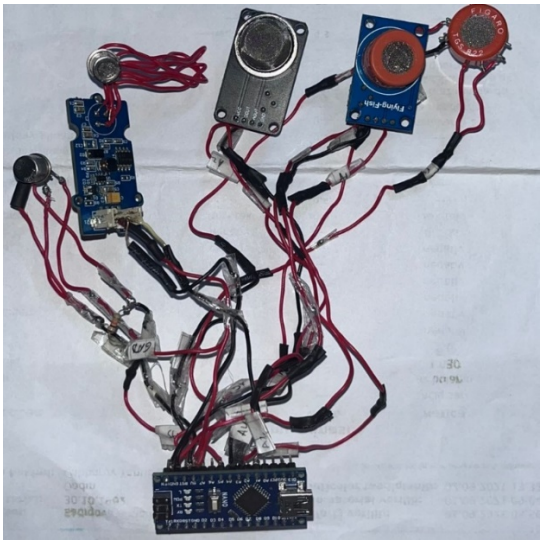


Figure 3.4. The constructed circuit of the test device.

3.4. ARDUINO CODING OF THE TEST DEVICE

After constructing the circuit, the Arduino coding of the test device was done. All necessary libraries were defined. To read the analog signal analog pin, mq3, mq135, mp503, tgs822 and tgs2602 were defined as variables. The values obtained from the A1, A2, A3, A4, and A5 pins were transferred to the mq3, mq135, mp503, tgs822 and tgs2602 variables respectively and the values were displayed on the serial port screen. The created Arduino code is provided in Figure 3.5, while the serial port screen image is given in Figure 3.6.

```
int analogPin = 0;
int mq3 = 0;
int mp503 = 0;
int mq135 = 0;
int tgs2602 = 0;
int tgs822 = 0;

void setup() {
  Serial.begin(9600);
}

void loop() {
  mq3 = analogRead(A1);
  mq135= analogRead(A2);
  mp503 = analogRead(A3);
  tgs822 = analogRead(A4);
  tgs2602 = analogRead(A5);

  Serial.print(" mq3 : ");
  Serial.print(mq3);
  Serial.print("  ||   mq135 : ");
  Serial.print(mq135);
  Serial.print("  ||   mp503 : ");
  Serial.print(mp503);
  Serial.print("  ||   tgs822 : ");
  Serial.print(tgs822);
  Serial.print("  ||   tgs2602 : ");
  Serial.println(tgs2602);
  delay(3000);}
```

Figure 3.5. The created Arduino code of the test device.

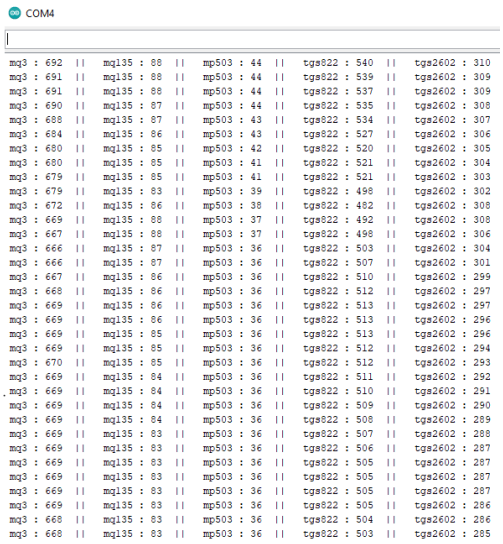


Figure 3.6. Serial port screen of the test device.

3.5. THE DESIGN OF THE TEST DEVICE

The design of the test device was made with the Fusion 360 program. The technical drawings of 5 sensors were added to Fusion 360 to determine the test device dimensions and sensor locations. The 5 sensors were positioned to fit the circular region and to ensure that the sensing part is at the same height (Figure 3.7). The sensor chamber was designed in a way to ensure that the gases are measured in a fast and accurate manner by the sensors without spreading. 4 holes were added to the chamber for a quick reset of the sensors after the measurement is done. The body part of the test device was designed by taking into account the Arduino Nano and all connection dimensions. A cable entry for Arduino Nano was added to the body. Finally, the blowing cover was designed. The whole test device design is provided in Figure 3.8.

Figure 3.7. The sensor chamber and body of the test device with added 5 sensors.

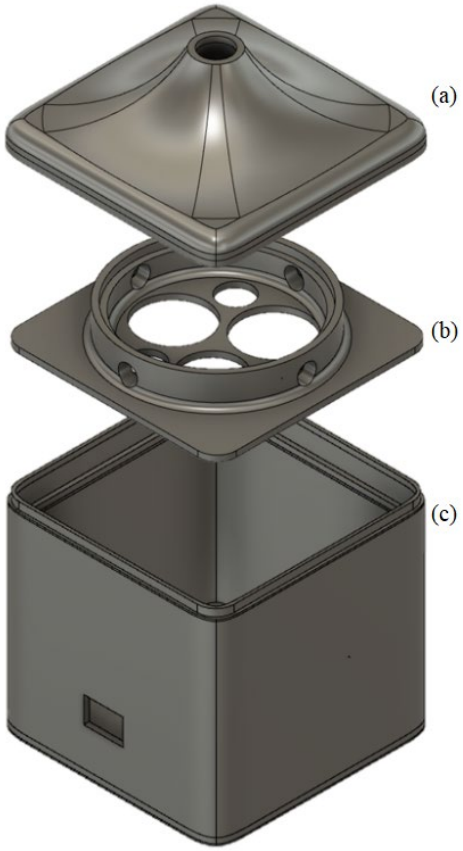


Figure 3.8. The design of the test device: a) blowing cover, b) sensor chamber, c) body.

The whole test device design was printed with a 3D printer (Figure 3.9) and the circuit elements were assembled. The final version of the assembled test device is given in Figure 3.10.

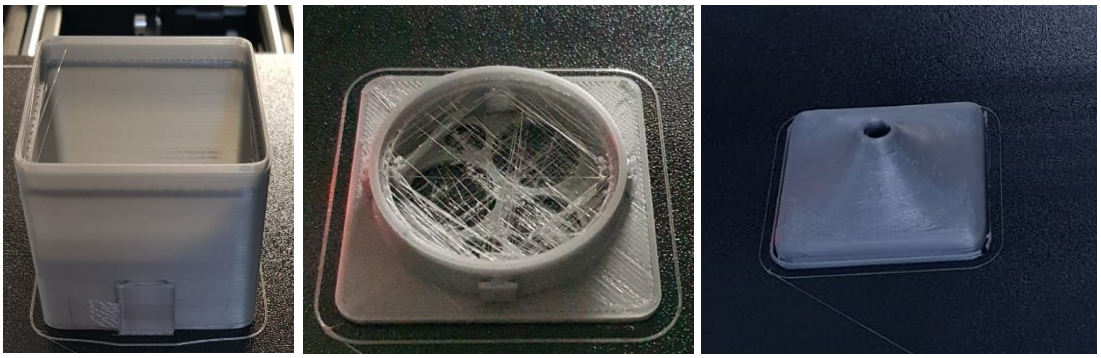


Figure 3.9. The test device parts printed with a 3D printer.

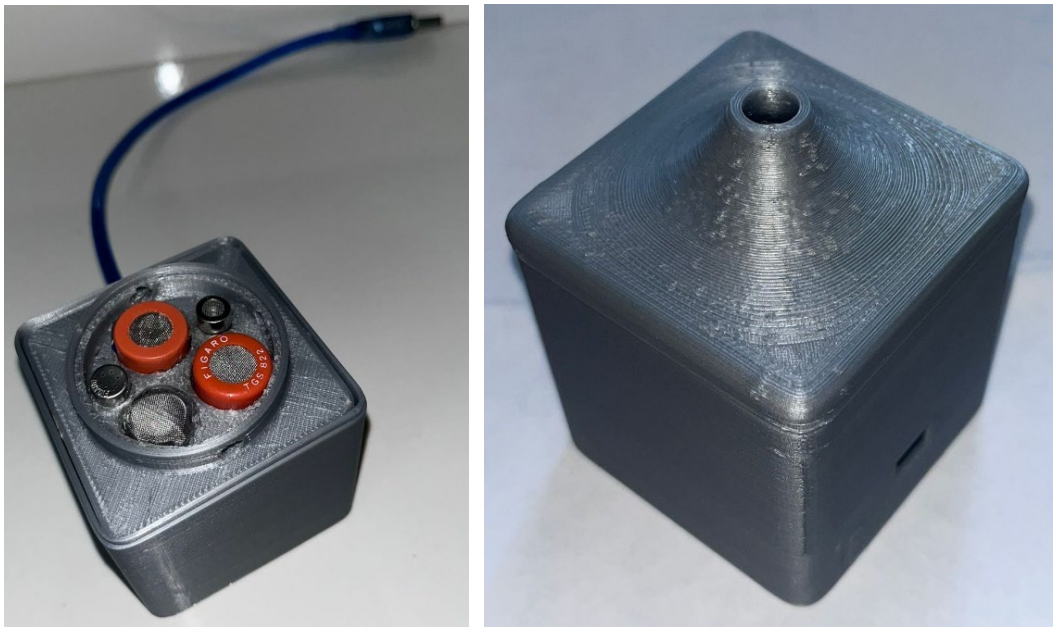


Figure 3.10. The final version of constructed test device.

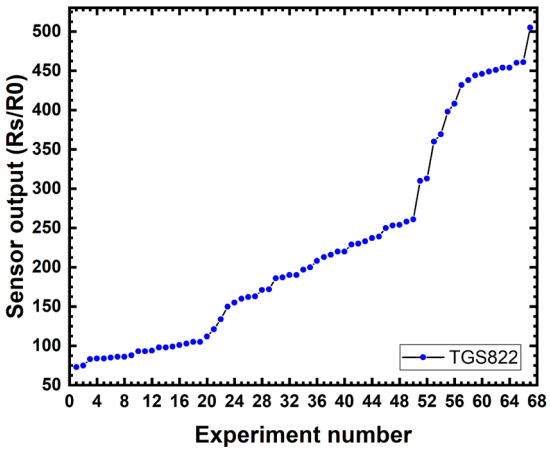
3.6. BREATH TESTS CONDUCTED VIA CONSTRUCTED TEST DEVICE

The preheating process was carried out in order to ensure stable operation and accurate measurement of the gas sensors in the test device. For this, the test device was heated by connecting it to the power supply for 7 days. After the ambient values were stabilized, breath tests were performed with the test device.

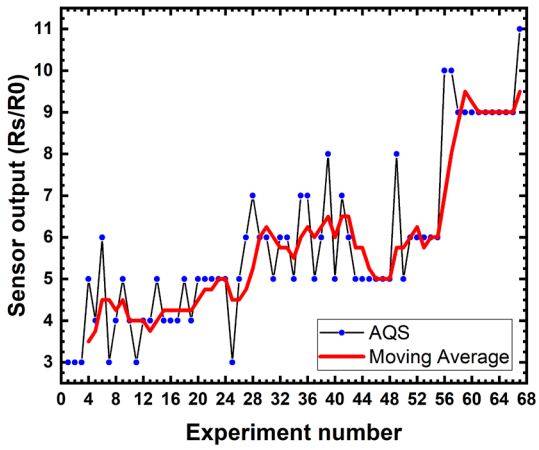
The test device was connected to the computer and preheated before each measurement. When the ambient sensor values were stable and suitable for the measurement, the device was blown from the blowing cover with a disposable mouthpiece. After blowing, the change in the sensor value was observed and the maximum values of each sensor were noted on the Arduino serial port screen. These data were transferred to the data set. A total of 67 breath data were collected from 5 people. These subjects did not use alcohol at least 24 hours before blowing and did not smoke at least half an hour before blowing. In addition, these subjects did not have any disease affecting their breathing. The subjects did not eat food or snacks that would cause odor, such as garlic and did not drink odor-inducing beverages at least 2 hours before blowing.

3.7. ANALYSIS OF THE BREATH DATA COLLECTED FROM THE TEST DEVICE

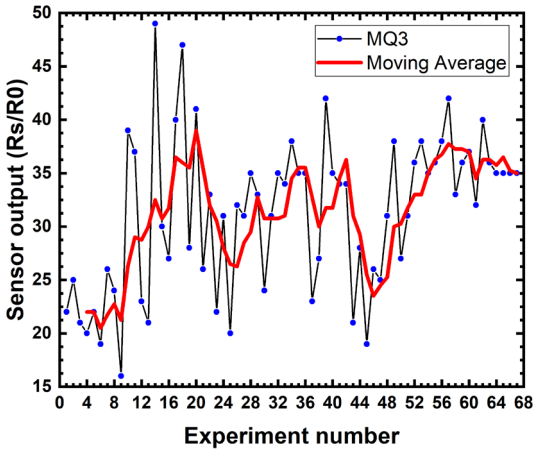
Since the TGS822 sensor was used in some studies for BA measurement in the literature and reasonable results were obtained and the sensor evaluation of the test device at hand was based on this sensor [26]. 67 data of TGS822 were listed in ascending order according to which graph of TGS822, AQS, MQ3, MQ135, TGS2602 sensor values were drawn with the Origin program (Figure 3.11). In addition, a 4-period moving average was added to the graphs to follow the curves and examine them more thoroughly.



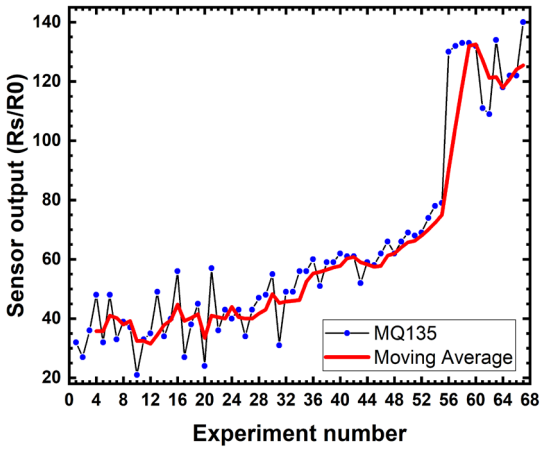
(a)



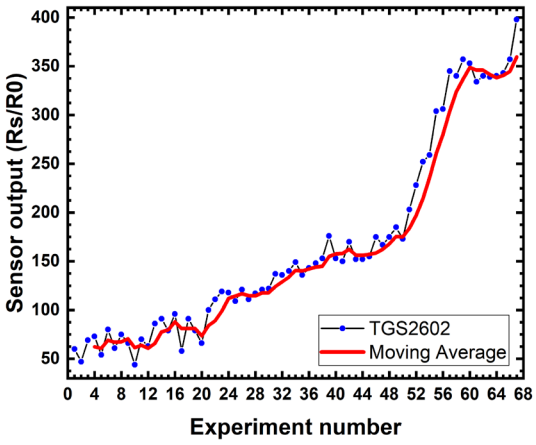
(b)



(c)



(d)



(e)

Figure 3.11. BA measurements collected by the test device with different gas sensors: a) TGS822, b) AQS, c) MQ3, d) MQ135, e) TGS2602.

When the moving averages of the AQS, TGS2602 and MQ135 curves are examined, it is seen that they are similar to the TGS822 sensor curve, whereas the MQ3 sensor responds quite differently to the same blows. Therefore it was concluded that the MQ3 sensor usage for BBGM device in unreasonable.

3.8. COMPARISON OF THE AIR QUALITY SENSORS USED IN THE TEST DEVICE

In addition to the BA response characteristics, the performances of the sensors were compared according to their warm-up, response and reset times, sensitivity, selectivity, stability, ease of use, size, price and accessibility (Table 3.2) [38].

Table 3.2. Comparison of the air quality sensors used in the test device.

Sensor	Preheat time	Response time	Reset time	Price (TL)	Accessibility
AQS	3 dk	5-15 s	15-35 s	275	+
MQ3	14 dk	15-25 s	90-120 s	45	+
MQ135	12 dk	15-25 s	90-150 s	50	+
TGS822	10 dk	20-30 s	45-180 s	800	-
TGS2602	6 dk	20-30 s	55-200 s	400	-

The preheating times of the sensors were measured since all sensors had to be preheated before each measurement in order to obtain stable data. It takes 3 minutes for the AQS sensor to be ready for stable measurement, while for the TGS2602, TGS822 and MQ135 sensors, this time is 6, 10 and 12 minutes, respectively. The sensor requiring the longest preheat time was determined to be MQ3 one. As can be seen from the response times of the sensors reported in the Table 3.2, all sensors detect the BA value within a certain period of time with maximum 30 seconds. It was observed that MQ and TGS series sensors have similar response times among themselves. As a result of the blowing tests, it was found that the MQ3 and MQ135 sensors detect the maximum breath value in the range of 15-25 seconds, while the TGS822 and TGS2602 sensors in the range of 20-30 seconds. However, the AQS sensor responds to the breath more quickly, in 5-15 seconds. Response time, as well

as reset time, are important for fast measurements. It was observed that the TGS2602 and TGS822 sensors take a minimum of 45 seconds and a maximum of 200 seconds to return to the ambient value before the next blowing can be done. This time varies between 90-150 seconds for MQ3 and MQ135 sensors. It was observed that the AQS sensor is reseted within 15-35 seconds. When evaluated in terms of price and accessibility, MQ series sensors are available in Turkey in the range of 45 – 50 TL, while TGS series sensors are much more expensive: TGS822 – 800 TL, TGS2602 – 400 TL. In addition, the TGS series sensors are not available in Turkey. AQS sensor module with a price of 275 TL is available in Turkey.

On other hand AQS, MQ3 and MQ135 sensors were used as modules, while TGS822 and TGS2602 were used as sensors. The use of the sensor is practical in that it is small in size, but it requires an additional resistor connection. However, the modules are self-resistive in accordance with the sensitivity of the sensor.

As a result of the comparison of the 5 sensors used in the test device, the AQS sensor was chosen to be used in the BBGM device based on the shortness of its preheat, response and reset times, its sensitivity to the breath value in the sensor guide, stability, ease of use and accessibility.

CHAPTER 4

SETTING UP AND EVALUATING A DEVICE WHICH MEASURES BLOOD GLUCOSE FROM THE BREATH ACETONE

4.1. CIRCUIT DESIGN AND ARDUINO CODING OF THE DEVICE

After selecting the appropriate gas sensor the circuit was designed to create a BBGM. In addition to the selected AQS sensor, the DHT22 temperature and humidity sensor was used to measure humidity and thus detect whether there is blowing into the device or not. A rechargeable battery to provide a power supply to the device, an OLED screen to display the BA, ambient values and other information, a switch to turn the device on/off, push buttons to reset the device and fix the values on the screen were added to the device circuit. All selected circuit elements were connected to Arduino Nano (Fiqure 4.1) and the Arduino code of the device was written and run. After all necessary libraries and sensor variables were defined, the BA measured analog values were transferred to these variables. Analog values were converted to ppm by the appropriate procedure (described in detail in section 4.3) and displayed on the OLED screen.

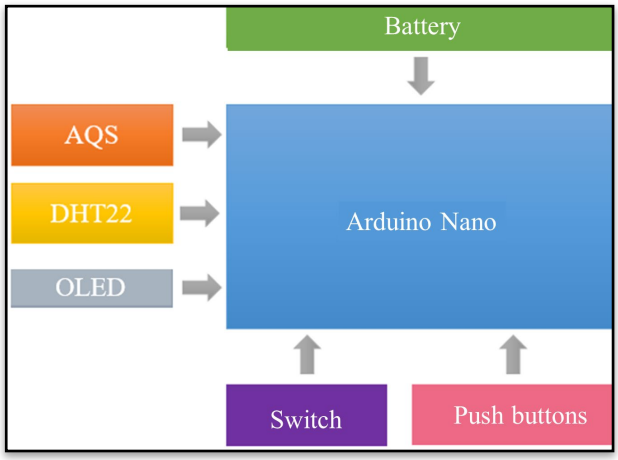


Figure 4.1. The circuit design of BBGM device.

4.2. 3D DESIGN, PRINTING AND ASSEMBLY OF THE BBGM DEVICE PROTOTYPE

The BBGM device prototype was designed with the Fusion 360 program according to the selected circuit elements, connection and locations. In the program environment the circuit elements were assembled and the device was designed according to this assembly. The device design consisted of three parts: the top cover, the sliding cover and the body (Figure 4.2). The OLED screen was located on the top cover of the device for friendly use. A sliding cover was designed for faster resetting of the sensors. The blowing hole was designed on the sliding cover to blow into the device with a disposable mouthpiece. Another design to reduce the reset time of sensitive sensors after measurement was to add 3 holes to the top cover. The fix and reset buttons for the gas sensor values were placed on the front body of the device, while the hole for charging, on/off key switch was placed on the side part of the device's body.

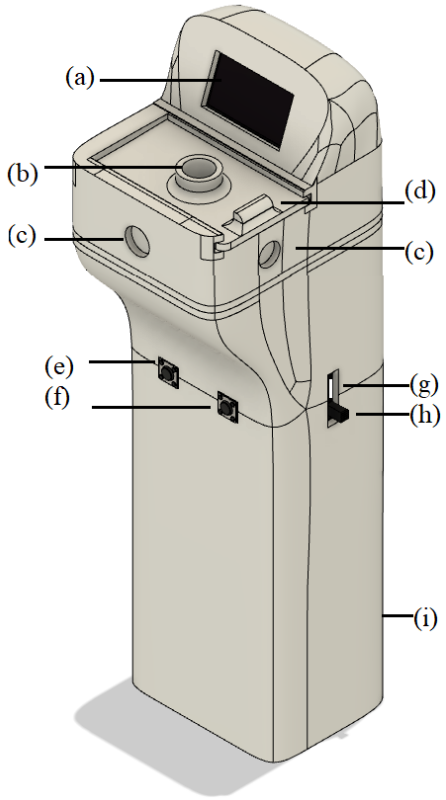


Figure 4.2 General view of the designed BBGM device prototype: a) OLED screen location, b) mouthpiece hole – blow hole, c) reset holes, d) sliding cover, e) fix button of the gas sensor value, f) reset button of the gas sensor value, g) charging port, h) on/off key switch, i) back cover of the device's body.

A sensor chamber was designed in the body of the device to reduce the effect of the external environment on the sensors and to accurately measure the trapped breath. In the body part, a cable entry (hole) was added to the design for the passage of cable connections. A total of 6 reset holes were drilled in the chamber walls to facilitate sensor's reset (Figure 4.3).

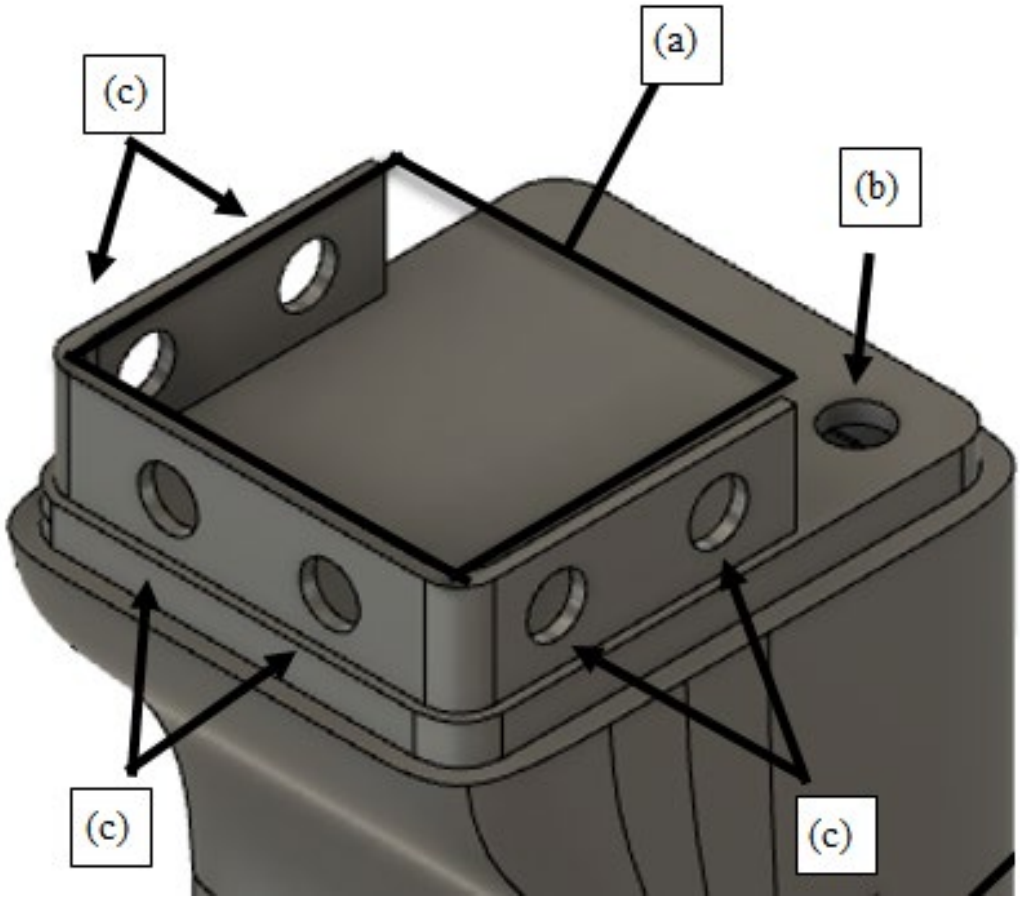


Figure 4.3. The design of the sensor chamber of BBGM device prototype: a) sensor chamber space, b) cable entry, c) reset holes.

The device design was printed with a 3D printer and the circuit elements were assembled. The assembled BBGM device prototype is given in Figure 4.4. The patent application related to this original design of BBGM device was submitted to Karabuk University Technology Transfer Office.

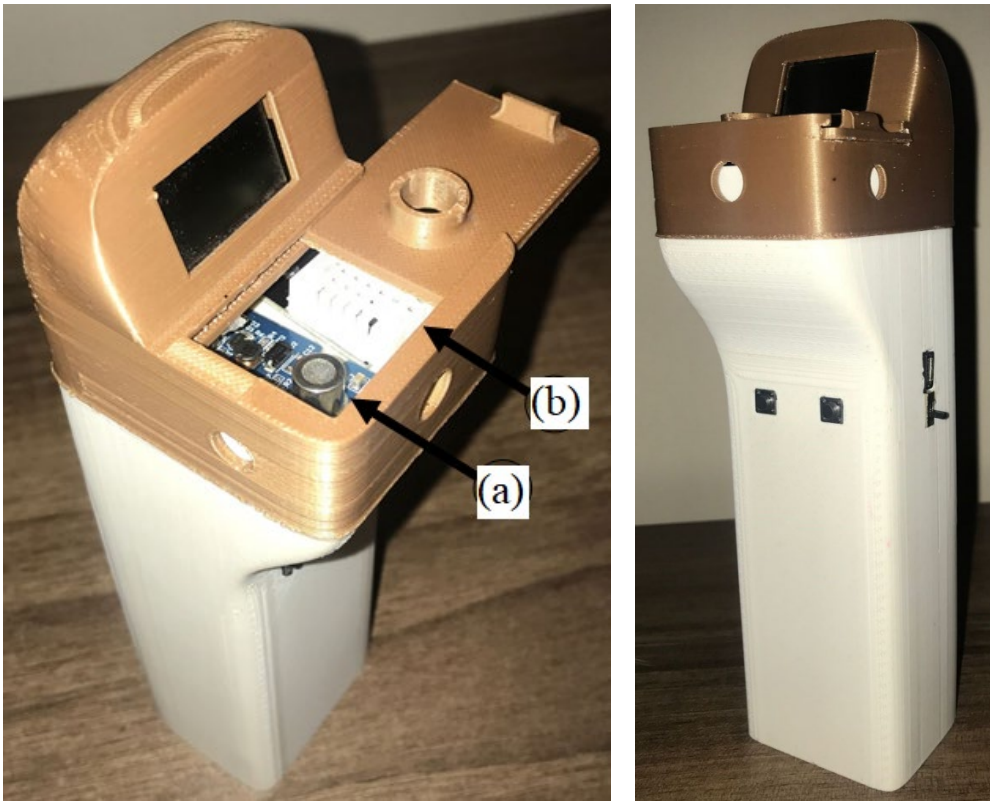


Figure 4.4 The assembled BBGM device prototype: a) AQS, b) DHT22 sensors.

4.3 CONVERTING ANALOG VALUES OF THE BBGM DEVICE TO PPM

Since the values are read in analog form from the AQS sensor in the BBGM device, their correlation to the ppm values was performed. Looking at the sensitivity curve given in the technical file of the MP503 sensor, it is seen that there is a relationship between the R_s/R_0 ratio and the acetone concentration in ppm, where R_s is the sensor resistance at various gas concentrations, R_0 is the sensor resistance in fresh air [33].

To make a convertion of analog values to ppm, first the sensitivity curve of the sensor was digitized. For this process, the sensitivity curve photograph of the AQS sensor was loaded into the digitizer as a 2D (X-Y) plot. The start and end points for the X and Y axes and the appropriate values in the graph image were entered. Horizontal and vertical intersection points were marked on the acetone curve. X and Y values of the curve points entered with the digitizer were obtained. The curve was drawn according to these values transferred to Excel program and the formula of the curve was obtained

(Figure 4.5, Equation 4.1). Since Y= R_s/R_0 and X=ppm, X was found from Equation 4.1 and it was written as in Equation 4.2.

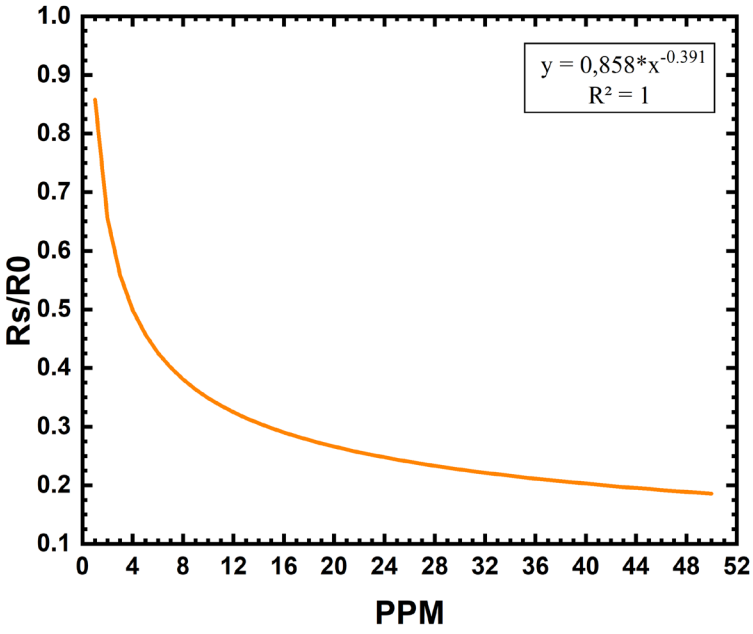


Figure 4.5. Acetone sensitivity curve of MP503 drawn using digitizer.

$$Y=0.858*X^{-0.391} \quad (4.1)$$

$$PPM=(1.165*R_s/R_0)^{-2.557} \quad (4.2)$$

With this formula of Equation 4.2, it is necessary to calculate the R_0 and R_s values to determine the ppm value. R_s and R_0 values are reported to be equal in fresh air. By definition, R_0 is the sensor's resistance in fresh air, so in this case only R_s changes for each gas [32]. It is reported in the AQS technical file that the gas sensor at hand has a voltage divider circuit with R_s and R_l (load resistance). The R_l resistance for AQS is given as 10K Ohms in the sensor technical file [32]. The voltage divider circuit formula was adapted to the values and R_s was found from the formula given in Equation 4.3, where V_c is the circuit voltage (5V), R_l is the load resistance (10K Ohm) and V_{Rl} is the voltage of load resistance.

$$R_s = [(V_c \times R_l) / V_{Rl}] - R_l \quad (4.3)$$

It was found that the maximum value obtained when the sensor is connected to a 5V power supply is equal to 800. Using this relationship, the AQS analog value was converted to V_{Rl} via the Equation 4.4, where a is the AQS analog value, b is the maximum analog value seen on the sensor when 5V is connected (b=800) and c is the power supply of the sensor (c=5V).

$$V_{rl} = a/b*c \tag{4.4}$$

In order to calculate the R_s value, firstly the analog value of the sensor (a) was measured in the 'fresh air'. The ambient value of the AQS sensor was determined as 10 in the ventilated environment. According to this value, V_{Rl} was calculated using Equation 4.4 in Arduino code and R_0 value was calculated using Equation 4.3. After defining the R_0 value in the Arduino code, the R_s value was calculated in the same way. By adding Equation 4.2 to the Arduino code, the ppm value was calculated according to the R_s/R_0 ratio and printed on the device screen. The device image with collected ppm value is given in Figure 4.6.

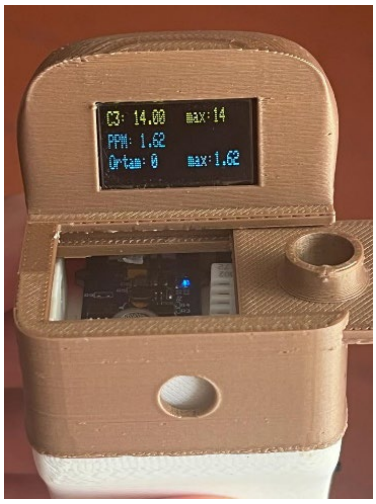


Figure 4.6. The BBGM device measuring BA value in ppm.

After writing algoritm for breath data collection in Arduino code and making device tests, the BBGM device was ready for the data collection.

CHAPTER 5

CONDUCTING PRELIMINARY WORK WITH THE CONSTRUCTED BBGM DEVICE

In order to examine the relationship between BA and BG for the first time BA data were collected from the volunteers via constructed BBGM device prototype and their BG data were collected with a home-based glucometer. Thirty-seven volunteers with T1D, T2D and ND (control group) health status, who live in Karabuk/Turkey and have a wide range of age, height and weight characteristics were included in this preliminary study. T2D subjects were taking only 1 or 2 oral antidiabetic drugs per day, but not any injections. On the other hand, T1D subjects received at least 2 insulin injections per day.

5.1. BA DATA ACQUISITION OF THE PRELIMINARY WORK

All subjects blew into the BBGM device prototype with a disposable mouthpiece for an average of 1.5 seconds. After reseting the BBGM device the blows were repeated. 3 blows were collected from each subject during each BA measurement. The average of these 3 values was recorded in the data set as the BA data. The following conditions were met during the collection of BA data:

- All subjects were in normal respiratory rhythm;
- All subjects did not have lung, liver, kidney or similar diseases that could affect the amount of exhaled BA;
- All subjects did not eat fruit or other foods (eg garlic) that would cause a strong odor for at least 2 hours before blowing into the BBGM device;
- All subjects did not drink an odorous beverage (eg coffee) at least half an hour before blowing into the BBGM device;

- All subjects did not eat snacks such as chewing gum, menthol candies, and did not do oral care (eg, mouthwash, brushing teeth) at least 1 hour before blowing into the BBGM device;
- All subjects did not drink alcohol for at least 24 hours before blowing into the BBGM device;
- Volunteers who smoke blew into the BBGM device at least half an hour after smoking;
- If all subjects did not eat at least 3 hours before blowing into the BBGM device, they were labeled as "preprandial" (marked as "0" in the data set) and "postprandial" (marked "1" in the data set) for all subjects who received food 2 hours ago.

5.2. BG DATA ACQUISITION OF THE PRELIMINARY WORK

After collection of each BA data the BG level was self-measured by volunteers using the same invasive commercial glucometer device Bayer Contour TS BG monitor with the disposable lancets and test strips. All volunteers pricked their middle finger and after wiping away the first two blood drops they extracted the third drop on the Bayer Contour TS test strip. BG values were self-measured by volunteers immediately after their BA data were collected by the BBGM device.

5.3. INVESTIGATION ON THE CORRELATION BETWEEN THE COLLECTED BG AND BA VALUES, ANALYSIS OF THE DATA WITH MACHINE LEARNING

The detailed characteristics of 37 people included in this preliminary study are provided in Table 5.1: 26 men, 11 women; 23-68 age range; 11 Normal Weight (NW), 15 Pre-Obese (PO), 8 Obese Class 1 (OC1), 2 Obese Class 2 (OC2), 1 Obese Class 3 (OC3); 26 ND, 7 T2D, 4 T1D. From 26 ND individuals 96 data: 35 preprandial, 61 postprandial; from 7 T2D subjects 33 data: 7 preprandial, 26 postprandial; from 4 T1D subjects 12 data: 5 preprandial, 7 postprandial were collected. As a result, the total number of collected data was 141. On the other hand, a wide range of age and BMI was used in this preliminary study with the statistics summarized in Table 5.2. Also,

Figure 5.1 suggests that there is a weak second-degree correlation between the collected age and BMI data. Although the number of volunteers in the preliminary study was small (37), BMI and age ranges are statistically sufficient, since the correlation curve in Figure 5.1 is similar to the statistical correlation curve found before for very large populations [39, 40, 41]. The number of subjects versus BMI category for ND, T2D and T1D volunteers is provided in Figure 5.2. Here the ND subjects are found to have low BMI values, while T2D subjects possess high BMI values (PO and Obese) and T1D subjects' BMI is in average related to PO. Even with such a low number of subjects at hand, the results of Figure 5.2 are consistent with the previous extensive studies: 50% of individuals with T2D are obese and approximately 90% are PO [42], while 186 T1D patients were found to have higher BMI values than 15,771 ND individuals [43]. In total, the volunteers of this preliminary study exhibit a wide range of age, BMI and DM classes that can be used for preliminary statistical examination.

Table 5.1. The characteristics of the volunteers who participated in the preliminary study. (According to the World Health Organization (WHO), NW: 18.5−24.9 kg/m^2; PO: 25.0−29.9 kg/m^2; OC1: 30.0−34.9 kg/m^2; OC2: 35.0−39.9 kg/m^2; OC3: ≥40.0 kg/m^2 [39].)

Subject	Gender	Age (years)	BMI (kg/m^2)	BMI Category	# of data	Total # of data
ND-1	M	23	21.9	NW	31	
ND-2	M	27	22	NW	2	
ND-3	F	51	31.6	OC1	1	
ND-4	F	36	21.5	NW	17	
ND-5	M	37	23.5	NW	14	
ND-6	M	34	33.7	OC1	5	
ND-7	M	30	28.4	PO	3	
ND-8	F	33	21.1	NW	1	
ND-9	M	35	26.5	PO	2	
ND-10	M	50	32.1	OC1	1	
ND-11	F	27	27.3	PO	2	
ND-12	F	31	37.7	OC2	2	
ND-13	M	30	19.9	NW	1	96
ND-14	M	34	32.9	OC1	1	
ND-15	M	41	27.1	PO	1	
ND-16	M	35	24.2	NW	1	
ND-17	M	52	27.8	PO	1	

ND-18	F	30	22.5	NW	1	
ND-19	M	60	26.1	PO	2	
ND-20	M	35	23.9	NW	1	
ND-21	F	38	27.1	PO	1	
ND-22	M	38	26.6	PO	1	
ND-23	M	36	31.5	OC1	1	
ND-24	M	50	27.7	PO	1	
ND-25	M	42	25.8	PO	1	
ND-26	M	25	23.4	NW	1	
T2D-1	F	41	36.6	OC2	10	
T2D-2	F	67	33.8	OC1	13	
T2D-3	M	43	42.9	OC3	3	
T2D-4	F	48	29.4	PO	2	33
T2D-5	M	41	28.4	PO	2	
T2D-6	F	67	30.5	OC1	1	
T2D-7	M	41	28.7	PO	2	
TID-1	M	68	24.6	NW	9	
TID-2	M	56	30.4	OC1	1	12
T1D-3	M	49	28.1	PO	1	
TID-4	M	36	28.4	PO	1	

Table 5.2. Descriptive statistics of volunteers' age and BMI values used in the preliminary study.

N:37 (F:11 M:26)	Max	Min	Mean	Median	SD
Age	68.00	23.00	41.00	38.00	11.60
BMI	42.90	19.90	27.99	27.70	4.92

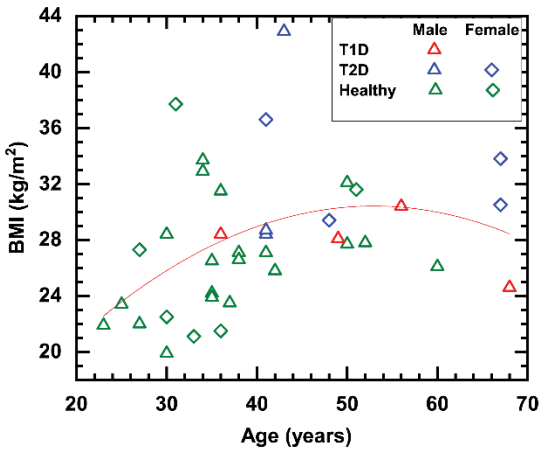


Figure 5.1. Correlation found between BMI and age in the preliminary work.

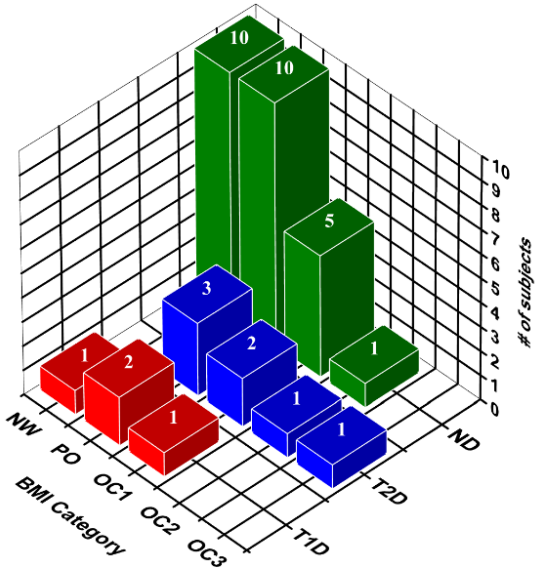


Figure 5.2. The number of the subjects versus BMI category for ND, T2D and T1D volunteers in the preliminary study.

5.3.1. Correlation Between BG And BMI Values Of The Preliminary Work.

Before going into the details of the correlation between BG and BA, which is the main point of the present study, the dependence of the volunteers' glucometer-measured BG on their BMI data was examined (Figure 5.3). The green symbols in Figure 5.3 show that while the BMI range of ND people is low (between ~20 and ~38 kg/m^2) their BG values are limited by 130 mg/dL. This maximum BG value is expected for the people without DM as their upper limit is accepted to be 140 mg/dL [44]. The weak positive linear correlation between BG and BMI values is observed for the preprandial state of the ND subjects, whereas no dependence if found for their postprandial state. On the other hand, T2D patients (blue symbols in Figure 5.3) have in average both higher BMI (~28 – ~43 kg/m^2) and BG (83 – 222 mg/dL) values than ND subjects. There is also a weak positive linear relationship between BMI and BG values for both preprandial and postprandial status in T2D patients. This is consistent with the previous study performed on 83 T2D patients, where a significant linear correlation between BMI and BG level was determined [45]. Finally, for a small number of the collected T1D patient data (red symbols in Fig. 5.3), BG values range from 119 to 262 mg/dL, which is in average much higher than BG values for ND and T2D patients. Overall, the ranges of the measured BG values of 3 different DM classes are found to

be within the expected limits according to the previous studies done with a higher amount of data.

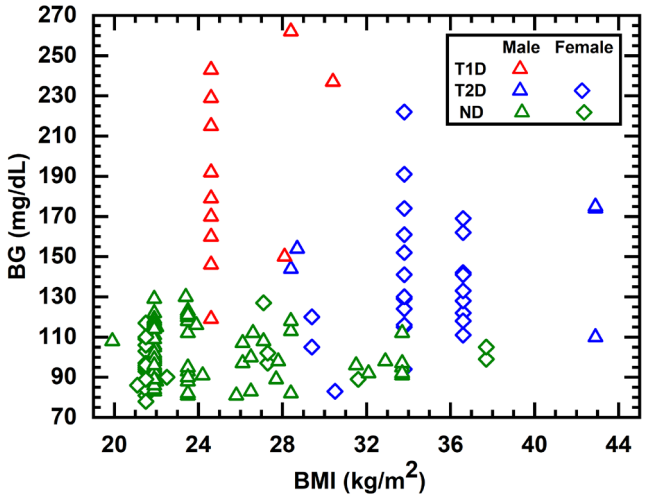


Figure 5.3. Correlation found between glucometer-measured BG and BMI.

5.3.2. Correlation Between BG And BA Values

The 141 BG and BA data collected for all 3 DM classes at both preprandial and postprandial status is presented in Figure 5.4. The maximum acetone concentration is found to be 6.74 ppm for T2D patients and 5.64 ppm for T1D patients, while this value was only 4.12 ppm for ND subjects. In the previous studies, it has been reported that the exhaled acetone level is generally 0.2–1.8 ppm in ND people, 1.25–2.5 ppm in DM patients, and may also increase up to 21 ppm [46] and 25 ppm [47] in some T1D patients. In another study, it was reported that the BA value ranged between 0.22 ppm and 0.80 ppm in ND individuals and between 1.76 ppm and 3.73 ppm in DM patients [48]. It should be noted that these ppm values are measured by different detectors/methods, while in the present study ppm level is recalculated from the commercial gas sensor output resistance values: the discrepancy between the absolute ppm values is possible, so the relative ppm level is rather considered in this study. Fig. 5.4 also suggests that while the BG values are in average much higher for T2D and T1D patients compared with ND people, the BA concentration of ND people does not much differ from that of diabetics: diabetics have higher BA than ND people for only 5 data out of 141. Similar result was found by Turner et al., where BA levels in T2D

and especially T1D patients were not generally much higher than in ND subjects [46]. Since a moderate positive correlation between BG and BA with a coefficient around 0.55 is found (not shown in Figure 5.4), it seems that a very strong BG-BA relationship can not be established.

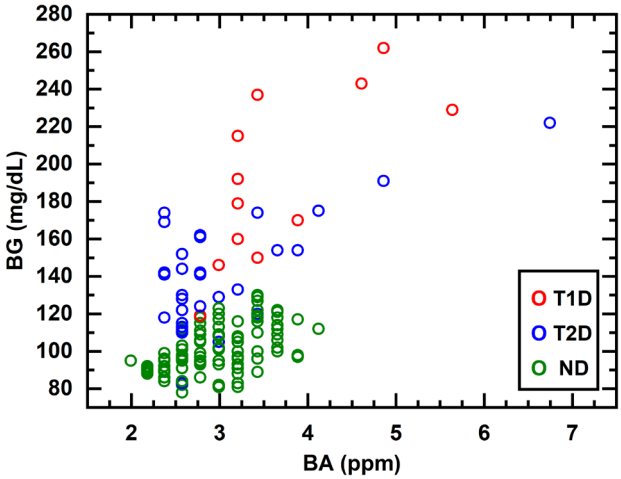


Figure 5.4. Correlation found between glucometer-measured BG and BA values for ND, T2D and T1D subjects.

When the diabetic classes (ND, T2D and T1D) and satiety status are taken into account and reexamined a much more meaningful relationships are obtained as it is shown in Figure 5.5 via redrawn 6 different graphs. Some of these graphs were fitted with quadratic polynomial (Figure 5.5 a, b, c, d), while others were fitted with linear lines (Figure 5.5 e, f) as guides for eye. As a result of this fitting, the positive relationship between BG and BA was obtained for postprandial cases (Figure 5.5 a, c, e, f), whereas for preprandial ND [49] and T2D cases the negative correlation is found to be present for low BA values overlapping with the positive one for high BA values (Figure 5.5 b, d). Such negative correlation at hunger state was previously found for volunteers experiencing ketogenesis [50]. Therefore, for ND and T2D subjects in this work at their preprandial state two different processes, i.e. ketosis and normal hungry state, probably overlap to give this complex relationship between BG and BA (Figure 5.5 b, d). Since both linear and non-linear correlations are found for all 6 cases, a more complicated treatment of the 141 data together with 6 different parameters seems to be inevitable. Therefore, the training of this data with different regression classifiers is done to choose the best method to predict the BG values from the measured BA ones.

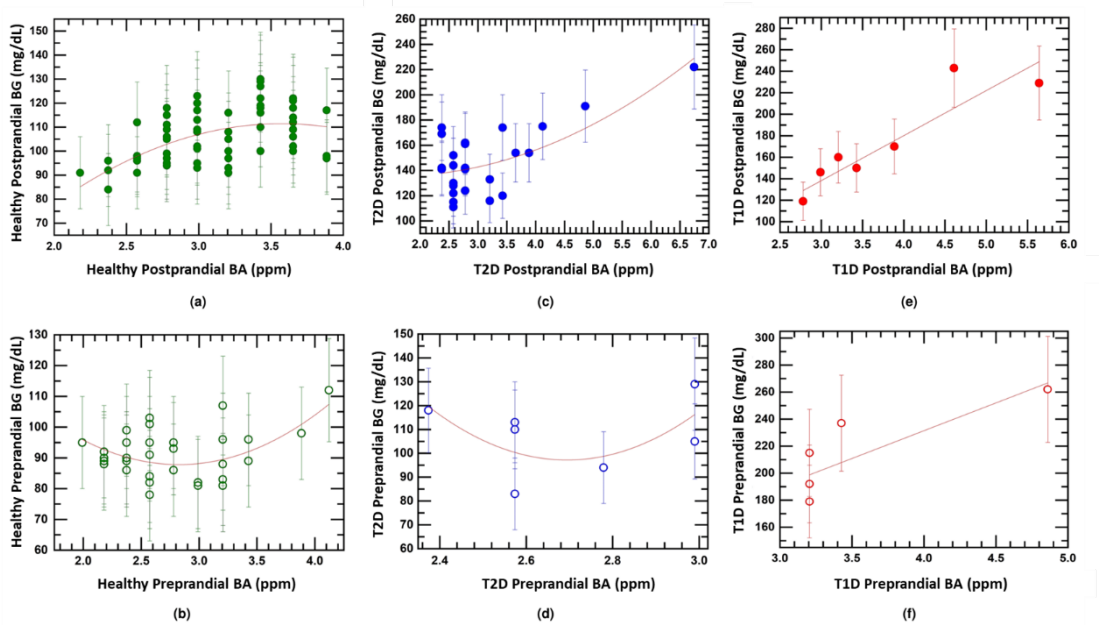


Figure 5.5. Correlation of the glucometer-measured BG and BA values redrawn separately for: a) ND postprandial, b) ND preprandial, c) T2D postprandial, d) T2D preprandial, e) T1D postprandial and f) T1D preprandial subjects.

5.3.3. ML Methods Used in the Preliminary Study

As it is well known, ML is a branch of science that deals with extracting information from existing data, it is the intersection of statistics and computer science and it is also a subfield of artificial intelligence. ML is also referred to predictive analytics or statistical learning. ML methods have become common in almost every field including daily life in recent years [51] and are frequently used in the health science. ML methods are classified according to the type of supervision they receive during training. The type of training in which causes and results are given to the system is called supervised learning [52]. There are two major types of supervised learning: classification and regression. The purpose of classification is to select a class from predefined probabilities. Regression, on the other hand, tries to estimate a continuous number [51]. In this preliminary study, three methods belonging to the regression group of supervised learning types were used to estimate the amount of the glucose in the blood from the breath data. These methods are polynomial linear regression (PLR), multi-layer perceptron (MLP) and support vector regressor (SVR).

5.3.3.1. PLR

Linear regression is a widely used method for finding the relationship between linear data and making predictions according to this relationship. This method is advantageous as it is fast and simple. However, it gives poor performance when the data is nonlinear. One of the ways to overcome this problem is to add the powers of each feature as new features and then train a linear model on this extended set of features. This technique is called PLR. Especially when multiple features are used for regression, the relationships can be found by adding combinations between features. This is a property that the linear regression model can not do. For example, if an input is two-dimensional and has the form [a, b], the degree-2 polynomial features are [1, a, b, a^2, ab, b^2] [52]. As for this preliminary study, degree-3 polynomial features were used.

5.3.3.2. MLP

Perceptron is one of the simplest neural network architectures in which a weighted z value is obtained using an addition function of inputs and weights and gives output as a result of applying an activation function to this z value [52]. But this method has some limitations. It can not produce solutions to nonlinear problems such as XOR problems. This can be overcome with MLP. MLP is a neural network model that produces output as a result of connecting multiple perceptrons as hidden layers [51].

5.3.3.3 SVR

Support vector machines (SVM) are popular methods for classification and regression. SVM allows to divide the z space consisting of input vectors from the optimal location through hyperplanes. The goal here is to have the margins at the maximum value when they are drawn for creating hyperplanes [52]. This principle gives quite good results for classification, but it is applied with some small differences for the regression because it is difficult to predict numbers with infinite probability. An approximate tolerance margin (epsilon) is added to eliminate this difficulty. In regression, as in classification, it is aimed to find hyperregulations that will maximise the margin. This

is an optimisation problem and solving it is simpler with its Lagrange dual formulation. The approximate function that occurs after this approach is given in Equation 5.1 [54]. This equation is used to estimate new values. $(-a_i + a_i^*)$ stores the difference between two Lagrange multipliers of support vectors, x_i stores support vectors and b stores bias.

$$\sum_{i=1}^{l}(-a_i + a_i^*)K(x_i, x) + b. \tag{5.1}$$

5.3.4. Training the Data of the Preliminary Study with PLR, MLP and SVR

The collected 141 data were trained with PLR, MLP and SVR (Table 5.3). The equations used for the determination of R^2 score, Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) are provided in Equation 5.2, Equation 5.3 and Equation 5.4, respectively. As seen from the compared coefficient of determination (R^2 score), MAE and MAPE values, the best result is obtained by PLR method: R^2=0.89, MAE=9.18, MAPE=7.85%. The R^2 score provides information about the goodness of a model's fit: it is a statistical measure of how well the predicted BG values approximate the true BG values [55]. The R^2 score is calculated to be 0.89 in this preliminary study, which is very close to its maximum value 1 [55]. In addition, the MAE and MAPE values being equal to 9.18 mg/dL and 7.85%, respectively, indicate that the estimation accuracy of the proposed method can compete with that of the commercial glucometer: the MAE and MAPE values for the glucometer at hand are 15 mg/dL and 15%, respectively [56]. In this context, it is clear that the proposed method can be preferred for painless and high accuracy alternative BG measurement.

Table 5.3. BG training results of 141 data obtained via PLR, MLP and SVR methods.

Metric	**PLR**	**MLP**	**SVR**
R^2 score	0.89	0.76	0.09
MAE (mg/dL)	9.18	12.81	20.23
MAPE (%)	7.85	10.76	14.74

$$R^2(y,\hat{y}) = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \quad (5.2)$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|e_t| \quad (5.3)$$

$$MAPE = \frac{100\%}{n}\sum_{t=1}^{n}\left|\frac{e_t}{y_t}\right| \quad (5.4)$$

5.3.5. Evaluation of the BG Values Measured by the Glucometer and Predicted by PLR Method

The comparison between the BG values measured by the glucometer and predicted via BA by PLR method is summarized in Table 5.4. The mean and SD of the glucometer-measured and predicted BG values are found to be very close to each other. The relationship between the predicted and measured BG values is examined further by plotting a scatter plot together with the line of equality, which is acting as a reference line [57] (Figure 5.6). This plot shows how far are the predicted BG values from the measured BG ones. If all the points lie along the line of equality (the black line in Fig. 6), it indicates a perfect agreement [58]. It is also observed that the BG values are mostly within the glucometer error band, which is shown by the gray region within the red lines in Figure 5.7. This error range is drawn according to the standarts imposed on the glucometers: at least 95% of BG values should be within the range of ±15 mg/dL at BG concentrations <100 mg/dL and ±15% at ≥100 mg/dL [56].

Table 5.4. Descriptive statistics of BG prediction by PLR method in the preliminary work.

N: l41	**Max**	**Min**	**Mean**	**Median**	**SD**
Measured BG (mg/dL)	262.00	78.00	117.13	108.00	35.09
Predicted BG (mg/dL)	266.00	87.00	117.12	106.00	33.20
Absolute Error (mg/dL)	33.00	0.00	9.18	8.00	6.75
Absolute Perc. Error (%)	25.30	0.00	7.85	7.37	5.27

Indeed, out of 141 data at hand only 10 data is found to be outside this error band, i.e. ~93% of the predicted data is acceptable. Although the scatterplot of Figure 5.7 is

useful to compare the measured and predicted BG values, it is difficult to evaluate the differences between them: usually the data points cluster around the line [58]. Therefore, the Bland-Altman plot is also utilized to compare the predicted and measured BG values (Figure 5.7) [58]. Here it is important to compute the Mean value of the difference, its SD, the confidence intervals for 95% limits of agreement (LoA) [58] (the red dashed lines in Figure 5.7).

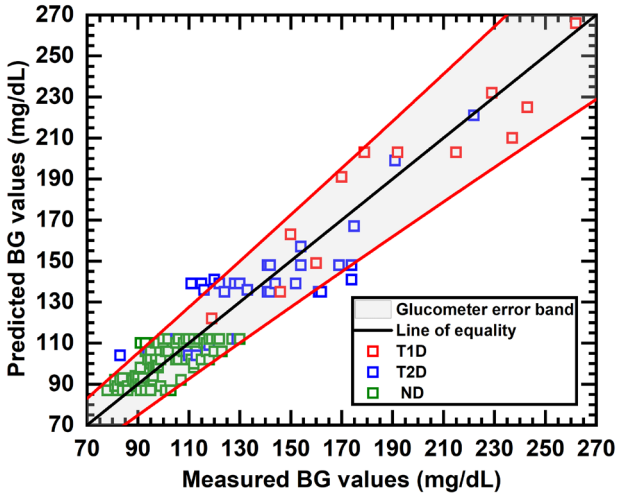


Figure 5.6. Correlation of the PLR predicted and glucometer-measured BG values for ND, T2D and T1D subjects.

The difference between the glucometer-measured and PLR predicted BG values for 141 data at hand resulted in Mean value of 0.007 mg/dL, its SD of 11.43 mg/dL, lower LoA of -22.39 mg/dL, upper LoA of 22.40 mg/dL. Since the difference between the two methods resulted in the very low Mean value (almost close to zero), the predicted and measured BG values are found to be in a good agreement.

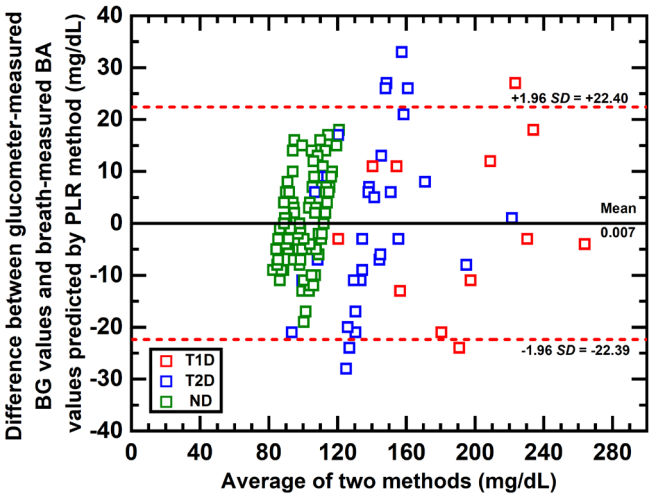


Figure 5.7. Bland-Altman plot for the glucometer-measured and predicted by PLR method BG values.

CHAPTER 6

CLINICAL STUDY WITH THE BBGM DEVICE

After completing the preliminary study where BG data were collected by home-based glucometer, the clinical study was conducted with the same BBGM device but BG data were collected from patients' venous blood. In order to reveal the relationship between BA and BG in a more comprehensive way, BG and BA measurements were made again separately in the satiety and hunger states of T1D, T2D and ND individuals in the clinical environment. In this clinical study, besides the BA value, waist circumference, body mass index (BMI), height, weight, age, hunger-satiety status, hunger-satiety duration, gender and health status (T1D, T2D or ND) values collected from volunteers were included as parameters in the machine learning methods. For BG - BA correlation, the data were trained with different ML methods and the training results were compared. In addition, the effect of other collected parameters on the training score was investigated.

6.1. DATA ACQUISITION IN THE CLINICAL ENVIRONMENT

In order to examine the relationship between BA and BG, 426 data were collected from 403 volunteers who applied to Karabuk University Training and Research Hospital Internal Diseases Polyclinic within the scope of Kastamonu University Clinical Research Ethics Committee approval. The volunteers involved in this clinical study live in the Turkish province of Karabuk and have a wide range of age, height and weight characteristics. The volunteers are between 18-83 years old, male or female, with T1D, T2D or ND, without lung disease, liver cirrhosis, chronic kidney failure, etc., who have not consumed alcohol in the last 24 hours. T2D patients who applied to the polyclinic for the routine control of BG values every 3 months and ND volunteers who came to have BG analysis for general control purposes were informed about this study. By obtaining BGOF approval from these volunteers the BMI, height, weight,

age, hunger-satiety status, hunger-satiety duration, waist circumference, gender and health status (T1D, T2D or ND) values were recorded. Since the smell of alcohol used during BG measurement would affect the BBGM device with a gas sensor, the BA measurement was made before the BG measurement.

6.1.1. BA Data Acquisition In The Clinical Environment

In this clinical study, the BA data were collected with the same BBGM device used in the preliminary work (Chapter 5). The conditions determined in the preliminary study with the device during the collection of BA data were also met in the clinical study. The subjects with normal respiratory rhythm and without lung, liver, kidney, or similar diseases that could affect the amount of administered BA concentration were included in this study. Before blowing into the BBGM device, the subjects did not eat foods that would cause a strong odor at least 2 hours before, did not smoke or drink odorous drinks at least half an hour before, did not eat snacks such as chewing gum, menthol candy for at least 1 hour and they didn't do any oral care (e.g. mouthwash, tooth brushing). Also the subjects did not consume alcohol for at least 24 hours before blowing. The subjects who had not eaten for at least 3 hours before the blowing were labeled as “preprandial” (marked “0” in the dataset), subjects who received food 2 hours before the blowing were labeled as “postprandial” (marked “1” in the dataset). To make further differentiation between thre subjects, the fasting-fulness time of the subjects was also questioned and included in the data set in different categories: 0-4 hours – “0”, 4-8 hours – “1”, 8-12 hours – “2”. Snacks and fruits were not accepted as a meal, only a meal with certain nutritional values was considered as “postprandial” status.

All subjects blew into the BBGM device with a disposable mouthpiece for an average of 1.5 seconds. During each BA measurement, 3 consecutive blows were collected from the subject and the time between each blow was approximately 15 seconds. The average of these 3 values was recorded as BA measurement data.

6.1.2 BG Data Acquisition In The Clinical Environment

After each BA measurement, BG measurement from venous blood was done. Preprandial and postprandial BG values were obtained for T1D and T2D patients who applied to the polyclinic for the routine control of BG values once in every 3 months. In addition, preprandial and postprandial BG values of ND volunteers who came to have BG analysis for general control purposes were recorded. The control group was formed from ND volunteers. No blood collection was performed specifically for this study.

6.2 INVESTIGATION ON THE CORRELATION BETWEEN BG AND BA ACQUIRED IN THE CLINICAL ENVIRONMENT, DATA ANALYSIS VIA ML

In this study, 426 data were collected from a total of 403 volunteers and added to the data set.

6.2.1. Preliminary Clinical Data Analysis

After examining the data, T1D data were excluded from the study, as the number of T1D data was not sufficient for evaluation (4 data from 4 volunteers). Afterwards, outlier data analysis was performed in terms of BA and BG.

6.2.1.1. Outlier Data Analysis

In general outliers may indicate variability in measurement, experimental errors or an innovation [55]. As a result of the outlier analysis for BG data, it was seen that the BG values of the 7 ND volunteers were above the normal range. Since the normal BG values for ND individuals have been reported to be less than 120 mg/dL for preprandial and less than 140 mg/dL for postprandial [59], the health category of these 5 ND volunteers with high BG values was rechecked by the internist who concluded that after BG results were received these volunteers were diagnosed with T2D. Therefore, these 5 volunteers' health status in the dataset was changed from ND to T2D. 2 T2D

data with BG values above 400 mg/dL were excluded from the data set because they were determined as outliers. As a result of the analysis for BA, it was seen that 5 T2D data were outliers. Some of these subjects received dental canal treatment, while others were in inconsistent health. These 5 data were not included in the ML analysis of this study. Thus, the data set of the clinic study was reduced to 415 data.

6.2.1.2 Normality Test

The distribution of 415 data was examined with the Kolmogorov-Smirnov (KS) test and it was seen that the data were normally distributed. The KS test is a statistical hypothesis test that measures the goodness of fit between two probability distributions [60, 61]. This test compares the empirical cumulative distribution function (ECDF) of the sample with the theoretical cumulative distribution function (CDF) of the specified distribution and calculates the maximum difference (D) between them. The significance level of the KS test is typically chosen as 0.05. If the P value is less than 0.05, the null hypothesis is rejected and it is concluded that the sample comes from a different distribution than the reference distribution. If the P value is greater than or equal to 0.05, the null hypothesis is accepted [62, 63]. The KS test is widely used in various fields such as physics, engineering, social sciences and finance [64, 65]. KS values of clinic study were calculated separately for ND, T2D and all data according to BG and BA parameters (Table 6.1). The fact that all obtained KS values are greater than 0.05 indicate that the clinical data are normally distributed.

Table 6.1. KS test of ND, T2D and all data according to BA and BG parameters

	Data	Number	KS value
BG	ND	201	0.079
	T2D	214	0.143
	All data	415	0.189
BA	ND	201	0.101
	T2D	214	0.112
	All data	415	0.105

The relationship between BG and BA values for 415 data obtained in the clinical environment is given in Figure 6.1. Similar to the preliminary study, the average BG value of the T2D clinical data shown in blue is greater than ND data shown in green. The lowest value for BA was obtained as 1.51 ppm and the highest value was measured as 5.29 ppm. For BG, the lowest value was 66 mg/dL and the highest value was 413 mg/dL. BG values were found to be between 66-144 mg/dL for ND volunteers and between 72-413 mg/dL for T2D patients. BA values were found to be between 1.51-5.29 ppm for ND volunteers and between 1.64-4.54 ppm for T2D patients.

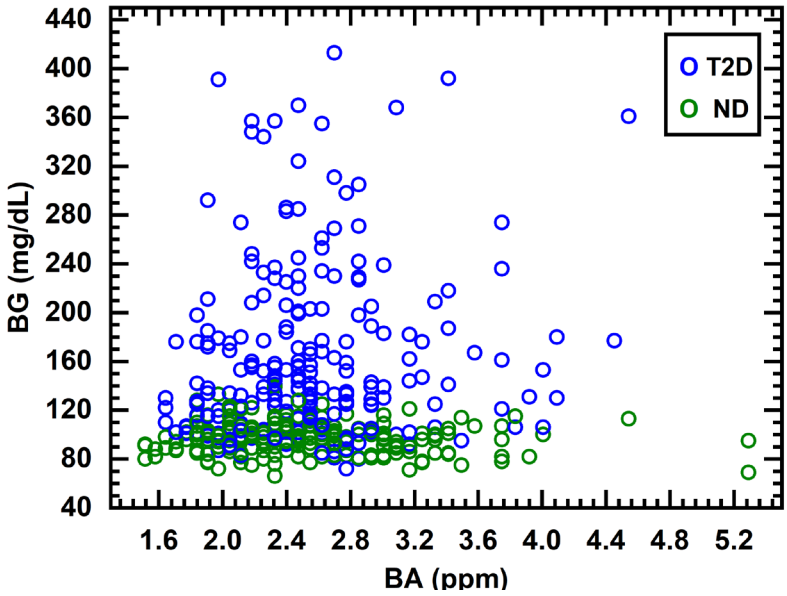


Figure 6.1. The relationship between the BG and BA values measured in the clinical environment.

Out of 415 data, 201 were ND and 214 were T2D. Data were collected from 261 female and 154 male subjects. The mean, median, standard deviation, minimum and maximum values of 415 data were calculated for age, weight, height, BMI, waist circumference (Table 6.2). 3 ND volunteers with BA values of 5.29 ppm, 5.29 ppm and 4.54 ppm were in a 12-hour hunger state. The high BA level was thought to be related to the hunger ketoacidosis state in the volunteers' bodies due to long-term hunger [66]. The mean, median, standard deviation, minimum and maximum values for BA and BG parameters were also examined separately, taking into account the ND and T2D categories (Table 6.3).

Table 6.2. Evaluation of the parameters for all data of the clinical study.

N=415	Mean	Median	Standard Deviation	Min	Max
Age	52.40	53.00	14.30	18.00	83.00
Weight	81.63	80.00	16.61	36.00	146.00
Height	1.60	1.59	0.10	1.37	1.91
BMI	31.94	31.50	6.20	17.36	56.96
Waist circumference	97.72	99.00	14.82	63.00	190.00

Table 6.3. Evaluation of BA and BG parameters of ND and T2D data of the clinical study.

	Data	Number	Mean	Median	Standard Deviation	Min	Max
BA	ND	201	2.52	2.39	0.62	1.51	5.29
	T2D	214	2.56	2.47	0.53	1.64	4.54
BG	ND	201	97.08	96.00	13.05	66.00	144.00
	T2D	214	167.91	145.50	70.91	72.00	413.00

6.2.2. Application of ML Methods

Various ML methods were used to examine the correlation between BA and BG clinical data and also to determine the effect of the parameters on the training score. In this clinical study, 4 methods belonging to the regression group of supervised learning types were used to examine the relationship between BA and BG data. These methods are MLP, PLR, SVM and gradient boosting (GB).

6.2.2.1 GB ML Method

GB is a popular ML algorithm used for regression and classification tasks. The algorithm creates a collection of weak prediction models, such as decision trees, by iteratively adding new models that focus on reducing the errors of previous models [67]. New models try to reduce the faults of previous models. This approach minimizes the error rate of the ensemble model, resulting in a more robust and accurate prediction

model [68]. The algorithm continues until a predetermined number of iterations or until the improvement in the loss function falls below a threshold. GB is widely used in various fields such as finance, marketing, health due to its efficiency, accuracy and interpretability and performs well in large datasets [67, 68].

6.2.3 Training Data with PLR, MLP, SVM, GB

The relationship between BA and BG was examined by the PLR, MLP, SVM, GB methods by adding the parameters of waist circumference, BMI, height, weight, age, hunger-satiety status, hunger-satiety duration, gender and health status (T2D or ND). 415 data were trained with MLP, PLR, SVM and GB methods and the training was tested at a train-test rate of 0.1. As a result of the training, the highest metric was found with the GB method as R^2 score of 0.986 and MAPE value as 4.41. The R^2 score obtained shows that the model can explain the variance in the data at the rate of 98.6% for 415 data. In the training tests with other methods, the R^2 score was 0.794, the MAPE value was 14.97 with the PLR method, the R^2 score was 0.577, the MAPE value was 4.93 with the SVM method, the R^2 score was 0.386 and the MAPE value was 23.37 with the MLP method (Table 4.6).

Table 6.4. Training results of 415 data by PLR, MLP, SVM, GB methods.

Metric	**PLR**	**MLP**	**SVM**	**GB**
R^2 score	0.794	0.386	0.577	0.986
Average Difference	19.0	32.62	13.31	5.49
MAPE (%)	14.97	23.37	4.93	4.41

6.2.4 The Effect of Different Physiological Parameters on the Correlation between BA and BG

After determining that the best ML method is GB, the effect of different physiological parameters on the score was investigated. All parameter combination were used with GB method as it is shown in Table 6.5. For 415 data, BA, waist circumference, BMI, height, weight, age, hunger-satiety status, hunger-satiety duration, gender parameters

were added to the training in 37 different possibilities and the training results were obtained. The lowest values of MAE and MAPE (MAE=3.748, MAPE=3.9) were found when all the parameters were added, while the highest values (MAE=9.456, MAPE=9.9) were observed when BA was used as a single parameter. Thus, it has been shown that inclusion of all variables has a positive effect on the education score by reducing the error.

Table 6.5. MAE and MAPE values measured according to the different combinations of BGOF variables with GB method.

No	MAE	MAPE	Variables
1	3.748	3.9	BA, waist circumference, BMI, height, weight, age, hunger-satiety status, hunger-satiety duration, gender
2	4.034	4.2	BA, waist circumference, BMI, height, weight, age, hunger-satiety status, gender
3	4.145	4.3	BA, waist circumference, BMI, height, weight, hunger-satiety status, hunger-satiety duration, gender
4	4.16	4.3	BA, waist circumference, BMI, height, weight, age, gender
5	4.16	4.3	BA, waist circumference, BMI, height, weight, age
6	4.229	4.4	BA, waist circumference, BMI, height, weight, gender
7	4.229	4.4	BA, waist circumference, BMI, height, weight
8	4.288	4.4	BA, waist circumference, BMI, height, weight, hunger-satiety status, gender
9	4.352	4.5	BA, BMI, height, weight, age
10	4.466	4.6	BA, waist circumference, BMI, height, hunger-satiety status, hunger-satiety duration, gender
11	4.572	4.7	BA, BMI, height, weight
12	4.639	4.8	BA, height, weight, age
13	4.704	4.9	BA, waist circumference, BMI, hunger-satiety status, hunger-satiety duration, gender
14	4.847	5	BA, waist circumference, BMI
15	4.867	5	BA, waist circumference, BMI, height, gender
16	4.867	5	BA, waist circumference, BMI, height
17	4.921	5.1	BA, waist circumference, BMI, height, hunger-satiety status, gender
18	4.921	5.1	BA, waist circumference, BMI, gender
19	4.955	5.1	BA, weight, age
20	5.047	5.2	BA, height, weight

21	5.077	5.2	BA, waist circumference, BMI, hunger-satiety status, gender
22	5.184	5.4	BA, BMI, height
23	5.385	5.6	BA, weight
24	6.11	6.3	BA, BMI
25	6.198	6.4	BP, waist circumference, hunger-satiety status, hunger-satiety duration, gender
26	6.521	6.7	BW, waist circumference, hunger-satiety status, gender
27	6.547	6.8	BA, waist circumference, gender
28	6.594	6.9	BA, age
29	6.633	6.8	BA, waist circumference
30	7.142	7.5	BA, height
31	7.861	8.3	BA, hunger-satiety status, hunger-satiety duration, gender
32	8.093	8.5	BA, hunger-satiety status, hunger-satiety duration
33	8.196	8.6	BA, hunger-satiety duration
34	8.744	9,1	BA, hunger-satiety status, gender
35	8.95	9.4	BA, hunger-satiety status,
36	9.166	9.6	BA, gender
37	9.456	9.9	BA

6.2.5 Displaying the Values of the BBGM Device on the Mobile Device

A Bluetooth module has been added to the BBGM device prototype to transfer data to the Android device. The module was coded and the values were displayed on the Android device (Figure 6.2). The device values that rise when pure acetone is sniffed are given in Figure 6.3 and the device circuit is given in Figure 6.4. It turned out that the thesis time should be longer for the development of a special mobile application for the BBGM device.

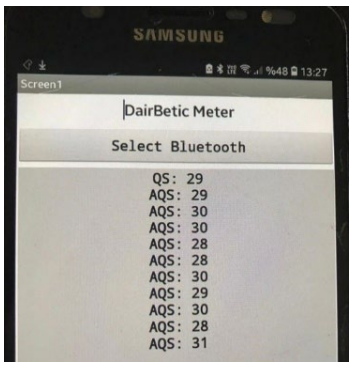


Figure 6.2. BA values of BBGM device prototype displayed on Android device after blowing.

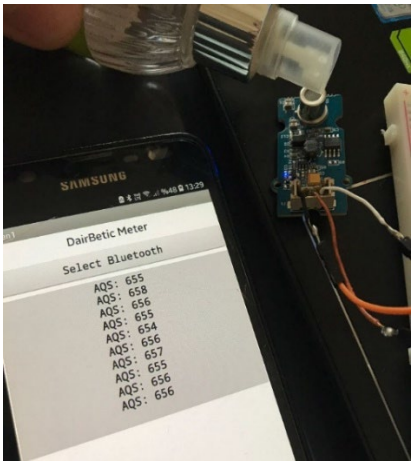


Figure 6.3. BA values of BBGM device prototype displayed on Android device when pure acetone is sniffed.

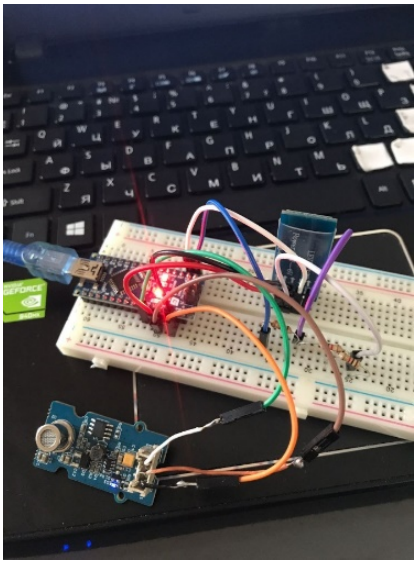


Figure 6.4. BBGM device circuit obtained by the addition of the Bluetooth module.

6.2.6. Evaluation of the BG Values Measured by Venous Blood and Predicted by GB Method

The relationship between the predicted and measured BG values is examined further by plotting a scatter plot together with the line of equality, which is acting as a reference line [57] (Figure 6.5). Since all points lie along the line of equality (the black line in Figure 6.5) the estimated BG values are close to the measured BG values. Also, the results are expected to improve further as the number of data with high BG values increases. The closness of the estimated BG values to the measured ones seems to be better for hifh BG values.

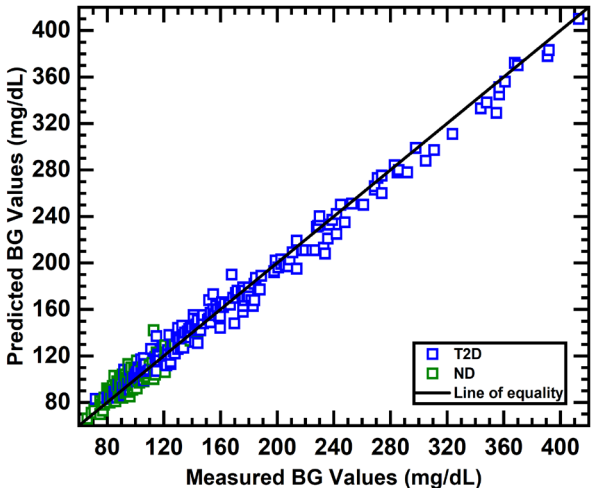


Figure 6.5. Correlation of the GB predicted and venous BG values for ND and T2D subjects.

Since most of the data are clustered around the line, the Bland-Altman plot is also utilized to compare the predicted and measured BG values (Figure 6.6). The difference between the venous and GB predicted BG values for 415 data at hand resulted in Mean value of 0.016 mg/dL, its SD of 7.36 mg/dL, lower LoA of -14.41 mg/dL, upper LoA of 14.45 mg/dL. Since the difference between the two methods resulted in the very low Mean value (almost close to zero), the predicted and measured BG values are found to be in a good agreement.

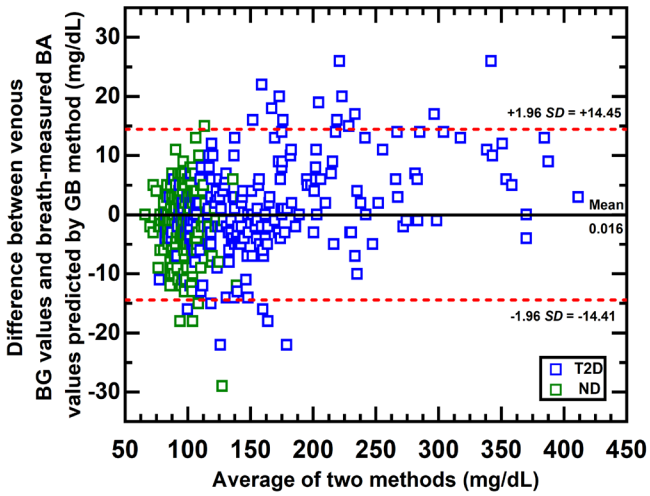


Figure 6.6. Bland-Altman plot for the venous and predicted by GB method BG values.

The BG values estimated by the GB method were examined whether the glucometers were in the error band range: it was observed that only 5 data were excluded. Error

range for glucometers: at least 95% of BG values at concentrations <100 mg/dL should be within ±15 mg/dL and within ±15% at ≥100 mg/dL [56]. The error range obtained with the GB method compared to glucometers is quite low: 4.41% for all values.

CHAPTER 7

CONCLUSION

In this thesis, the BBGM device prototype based on the cheap MOS sensor was developed for accurate estimation of BG from the human breath.

Firstly, 67 breath data and performances of 5 commercial gas sensors (AQS, MQ3, MQ135, TGS822, TGS2602) were compared and the AQS sensor was selected as the most suitable sensor to be used in the BBGM device. AQS has high sensitivity to the respiratory value in the sensor guide, high stability, ease of use, accessibility, and short preheating, response and reset times.

The BBGM prototype was constructed using AQS, DHT22, Arduino Nano, OLED screen, buttons and rechargeable battery. It measures BA value in ppm, has original software and design. The patent application of the original design of the BBGM device was submitted to Karabuk University Technology Transfer Office.

In the preliminary study 141 BA data were collected with a BBGM device prototype and BG data with a home-based glucometer. As a result of the training of these 141 data with PLR, MLP and SVR methods, the best training score was obtained by the PLR method with R^2=0.89, MAE=9.18, MAPE=7.85% scores.

With the approval of the Clinical Research Ethics Committee received from Kastamonu University, 415 data were collected in the clinical environment from ND and DM volunteers. In this clinical study BA data were collected again by the BBGM device and BG data were collected with the clinical method from the venous blood. As a result of training of 415 data with MLP, PLR, SVM and GB methods the best training score was obtained by the GB method with R^2=0.986, mean difference=5.49, MAPE=4.41%.

In order to evaluate the effect of the patients' physiological parameters on the education score in the clinical study 37 different parameter combinations were trained together with BA. The highest error values (MAE=9.456, MAPE=9.9) were obtained when BA was used as an only parameter, while the lowest error values (MAE=3,748, MAPE=3.9) were found when all parameters were used.

The high ML training scores obtained in both preliminary and clinical studies conducted within the scope of this thesis have shown that the BG levels can be predicted within an acceptable error from the BA values measured via the devepoled BBGM device prototype.

Within the scope of this thesis, the obtained publications are listed:

- Conference oral presentation: Binnatov T., Anutgan T., Yılmaz H. (2022) "Blood Glucose Prediction from Nondiabetic and Diabetic Exhaled Breath via Machine Learning", 4th International Conference on Applied Engineering and Natural Sciences, Konya, Türkiye (Kasım 2022).
- Book Chapter: Binnatov T., Anutgan T., Yılmaz H. (2022) "Blood Glucose Prediction from Nondiabetic and Diabetic Exhaled Breath via Machine Learning", International Research in Engineering, İzmir/Türkiye: Serüven Publishing
- Article: Binnatov T., Anutgan T., Yılmaz H., Sevencan N. Ö., Şahin F., (2023) "Venous Blood Glucose Prediction from Non-Diabetic and Diabetic Exhaled Breath via Machine Learning" in progress.

REFERENCES

1. American Diabetes Association, “Diagnosis and Classification of Diabetes Mellitus", ***Diabetes Care***, 37 (Supplement_1): S81–S90 (2014).

2. Saeedi, P., Petersohn, I., Salpea, P., Malanda, B., Karuranga, S., Unwin, N., Colagiuri, S., Guariguata, L., Motala, A. A., Ogurtsova, K., Shaw, J. E., Bright, D., and Williams, R., “Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9th edition”, ***Diabetes Research And Clinical Practice***, 157: 107843-107853 (2019).

3. Kharroubi, A. T., “Diabetes mellitus: The epidemic of the century”, ***World Journal Of Diabetes***, 6 (6): 850-867 (2015).

4. Kovatchev, B. P., “Diabetes Technology: Markers, Monitoring, Assessment, and Control of Blood Glucose Fluctuations in Diabetes”, ***Scientifica***, 2012: 1–14 (2012).

5. Internet: Türkiye Diyabet Vakfı, “Diyabet Hakkında Herşey - Tip 2 Diyabet”, **https://www.turkdiab.org/diyabet-hakkinda-hersey.asp?lang=TR&id=48** (2023).

6. American Diabetes Association, "7. Diabetes Technology: *Standards of Medical Care in Diabetes—2020*", ***Diabetes Care***, 43 (Supplement 1): S77–S88 (2020).

7. Galassetti, P. R., Novak, B., Nemet, D., Rose-Gottron, C., Cooper, D. M., Meinardi, S., Newcomb, R., Zaldivar, F., and Blake, D. R., “Breath Ethanol and Acetone as Indicators of Serum Glucose Levels: An Initial Report”, ***Diabetes Technology & Therapeutics***, 7 (1): 115–123 (2005).

8. Turner, C., “Potential of breath and skin analysis for monitoring blood glucose concentration in diabetes”, ***Expert Review Of Molecular Diagnostics***, 11 (5): 497–503 (2011).

9. Bovey, F., Cros, J., Tuzson, B., Seyssel, K., Schneiter, P., Emmenegger, L., and Tappy, L., “Breath acetone as a marker of energy balance: an exploratory study in healthy humans”, ***Nutrition & Diabetes***, 8 (1): 50-54 (2018).

10. Internet: Ketonix, “Products”, **https://www.ketonix.com/webshop/products** (2023).

11. Internet: Ketosischeck, “Keto Breath Monitor -KT003W”, **https://ketosischeck.com/products/keto-breath-monitor-kt003w** (2021).

12. Dhatariya, K. K., Glaser, N. S., Codner, E., and Umpierrez, G. E., "Diabetic ketoacidosis", ***Nature Reviews Disease Primers***, 6 (1): 40-101 (2020).

13. Yuan, X., Wang, J., Yang, S., Gao, M., Cao, L., Li, X., Hong, D., Tian, S., and Sun, C., "Effect of the ketogenic diet on glycemic control, insulin resistance, and lipid metabolism in patients with T2DM: a systematic review and meta-analysis", ***Nutrition & Diabetes***, 10 (1): 38-46 (2020).

14. Trotter, M. D., Sulway, M. J., and Trotter, E., "The rapid determination of acetone in breath and plasma", ***Clinica Chimica Acta***, 35 (1): 137–143 (1971).

15. Saasa, V., Beukes, M., Lemmer, Y., and Mwakikunga, B., "Blood Ketone Bodies and Breath Acetone Analysis and Their Correlations in Type 2 Diabetes Mellitus", ***Diagnostics***, 9 (4): 224-234 (2019).

16. Sun, M., Wang, Z., Yuan, Y., Chen, Z., Zhao, X., Li, Y., and Wang, C., "Continuous Monitoring of Breath Acetone, Blood Glucose and Blood Ketone in 20 Type 1 Diabetic Outpatients Over 30 Days", ***Journal Of Analytical & Bioanalytical Techniques***, 8 (5): 386-394 (2017).

17. Gong, Z., Y., Sun, M., X., Jian, C., Y., Wang, Z., N., Kang, M., L., Li, Y., X., and Wang C., "A Ringdown Breath Acetone Analyzer: Performance and Validation Using Gas Chromatography-Mass Spectrometry", ***Journal Of Analytical & Bioanalytical Techniques***, S7 (013): 1-8 (2014).

18. Rabih, A. A. S., Dennis, J. O., Ahmed, A. Y., Khir, M. H. M., Ahmed, M. G. A., Idris, A., and Mian, M. U., "MEMS-Based Acetone Vapor Sensor for Non-Invasive Screening of Diabetes", ***IEEE Sensors Journal***, 18 (23): 9486–9500 (2018).

19. Saraoglu, H. M., Selvi, A. O., Ebeoglu, M. A., and Tasaltin, C., "Electronic Nose System Based on Quartz Crystal Microbalance Sensor for Blood Glucose and HbA1c Levels From Exhaled Breath Odor", ***IEEE Sensors Journal***, 13 (11): 4229–4235 (2013).

20. Luaibi, A. Y., Ghusain, A. J. A., Rahman, A., Sayah, M. H. A., and Nashash, H. A. A., "Noninvasive blood glucose level measurement using nuclear magnetic resonance", ***2015 IEEE 8th GCC Conference and Exhibition (GCCCE)***, Muscat, 1-4 (2015).

21. Lekha, S. and Suchetha M., "Recent Advancements and Future Prospects on E-Nose Sensors Technology and Machine Learning Approaches for Non-Invasive Diabetes Diagnosis: A Review", ***IEEE Reviews In Biomedical Engineering***, 14: 127–138 (2021).

22. Guo, D., Zhang, D., Zhang, L., and Lu, G., "Non-invasive blood glucose monitoring for diabetics by means of breath signal analysis", ***Sensors And Actuators B: Chemical***, 173: 106–113 (2012).

23. Yan, K. and Zhang, D., "A novel breath analysis system for diabetes diagnosis", ***2012 International Conference on Computerized Healthcare (ICCH)***, Hong Kong, 166-170 (2012).

24. Lekha, S. and Suchetha M., "Real-Time Non-Invasive Detection and Classification of Diabetes Using Modified Convolution Neural Network", ***IEEE Journal Of Biomedical And Health Informatics***, 22 (5): 1630–1636 (2018).

25. Rydosz, A., Wincza, K., and Gruszczynski, S., "Microsystem in LTCC technology for the detection of acetone in healthy and diabetes breath", ***2016 IEEE ANDESCON***, Arequipa, 1-4 (2016).

26. Thati, A., Biswas, A., Chowdhury, R. S., Sau, K. T., "BREATH ACETONE-BASED NON-INVASIVE DETECTION OF BLOOD GLUCOSE LEVELS", ***International Journal On Smart Sensing And Intelligent Systems***, 8 (2): 1244–1260 (2015).

27. Hassan, T., Rehman, T., Qasim A. A., Ahmad, S., "Blood Glucose Level Measurement From Breath Analysis", ***International Journal of Medical, Medicine and Health Sciences***, 12 (9): 379-382 (2018)

28. Internet: The National Institute of Standards and Technology (NIST), "Acetone", **https://webbook.nist.gov/cgi/cbook.cgi?ID=C67641&Mask=200** (2023).

29. Internet: The National Institute of Standards and Technology (NIST), "Carbon dioxide", **https://webbook.nist.gov/cgi/cbook.cgi?ID=C124389&Units=SI&Mask=200#Mass-Spec** (2023).

30. Seiyama, T., Kato, A., Fujiishi, K., and Nagatani, M., "A New Detector for Gaseous Components Using Semiconductive Thin Films.", ***Analytical Chemistry***, 34 (11): 1502–1503 (1962).

31. Patil, S. J., Patil, A. V., Dighavkar, C. G., Thakare, K. S., Borase, R. Y., Nandre, S. J., Deshpande, N. G., and Ahire, R. R., "Semiconductor metal oxide compounds based gas sensors: A literature review", ***Frontiers Of Materials Science***, 9 (1): 14–37 (2015).

32. Internet: Seeedstudio, "Grove - Air Quality Sensor v1.3 - Arduino Compatible", **https://www.seeedstudio.com/Grove-Air-Quality-Sensor-v1-3-Arduino-Compatible.html** (2023).

33. Internet: Winsen, "ZP07-MP503 Air-Quality Detection Module—Winsen", **https://www.winsen-sensor.com/sensors/air-quality-sensor/zp07-mp503.html** (2023).

34. Internet: HANWEI ELETRONICS, "MQ-3 GAS SENSOR", **https://www.sparkfun.com/datasheets/Sensors/MQ-3.pdf (2023).**

35. Internet: Waveshare, “MQ-135 Gas Sensor - Waveshare”, **https://www.waveshare.com/wiki/MQ-135_Gas_Sensor** (2023).

36. Internet: Figaro, “TGS 822 | FIGARO”, **https://www.soselectronic.com/en/products/figaro/tgs-822-7719** (2023).

37. Internet: Figaro, “TGS2602 : Gas Sensors & Modules - Products - Figaro Engineering Inc.”, **https://www.figarosensor.com/product/entry/tgs2602.html#ti** (2023).

38. Goel, N., Kunal, K., Kushwaha, A., and Kumar, M., “Metal oxide semiconductors for gas sensing”, ***Engineering Reports***, e12604:1-22 (2022).

39. World Health Organization, “Obesity: preventing and managing the global epidemic.”, ***World Health Organization***, ***WHO/NUT/NCD/98.1***, ***Geneva***, 1-276 (2000).

40. Maillard, G., Charles, M., Thibult, N., Forhan, A., Sermet, C., Basdevant, A., and Eschwège, E., “Trends in the prevalence of obesity in the French adult population between 1980 and 1991”, ***International Journal Of Obesity***, 23 (4): 389–394 (1999).

41. Jayne, J. M., Blake, C. E., Frongillo, E. A., Liese, A. D., Cai, B., Nelson, D. A., Kurina, L. M., and Funderburk, L., “Trajectories of body mass index among active-duty U.S. Army soldiers, 2011–2014”, ***Preventive Medicine Reports***, 14: 100818-100827 (2019).

42. Garber, A. J., “Obesity and type 2 diabetes: which patients are at risk?”, ***Diabetes, Obesity And Metabolism***, 14 (5): 399–408 (2012).

43. Fellinger, P., Fuchs, D., Wolf, P., Heinze, G., Luger, A., Krebs, M., and Winhofer, Y., “Overweight and obesity in type 1 diabetes equal those of the general population”, ***Wiener Klinische Wochenschrift***, 131 (3–4): 55–60 (2019).

44. American Diabetes Association, "Postprandial Blood Glucose", ***Diabetes Care***, 24 (4): 775–778 (2001).

45. Amelia, R., “The Correlation Between Body Mass Index and Self-Efficacy with Blood Glucose Level in Type 2 Diabetes Mellitus”, ***Advanced Science Letters***, 23 (4): 3606–3609 (2017).

46. Turner, C., Walton, C., Hoashi, S., and Evans, M., “Breath acetone concentration decreases with blood glucose concentration in type I diabetes mellitus patients during hypoglycaemic clamps”, ***Journal Of Breath Research***, 3 (4): 1-9 (2009).

47. Masikini, M., Chowdhury, M., and Nemraoui, O., “Review—Metal Oxides: Application in Exhaled Breath Acetone Chemiresistive Sensors”, ***Journal Of The Electrochemical Society***, 167 (3): 1-11 (2020).

48. Deng, C., Zhang, J., Yu, X., Zhang, W., and Zhang, X., "Determination of acetone in human breath by gas chromatography–mass spectrometry and solid-phase microextraction with on-fiber derivatization", ***Journal Of Chromatography B***, 810 (2): 269–275 (2004).

49. Rydosz, A., "A Negative Correlation Between Blood Glucose and Acetone Measured in Healthy and Type 1 Diabetes Mellitus Patient Breath", ***Journal Of Diabetes Science And Technology***, 9 (4): 881–884 (2015).

50. Güntner, A. T., Kompalla, J. F., Landis, H., Theodore, S., Geidl, B., Sievi, N., Kohler, M., Pratsinis, S., and Gerber, P., "Guiding Ketogenic Diet with Breath Acetone Sensors", ***Sensors***, 18 (11): 3655-3667 (2018).

51. Müller, A. C. and Guido, S., "Introduction to Machine Learning with Python: A Guide for Data Scientists", First edition. Ed., ***O'Reilly Media, Inc***, Sebastopol, CA, 1-376 (2016).

52. Géron, A., "Hands-on Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems", First edition. Ed., ***O'Reilly Media***, Beijing, Boston, 1-551 (2017).

53. Vapnik, V. N., "The Nature of Statistical Learning Theory", ***Springer New York : Imprint : Springer***, New York, NY, (2000).

54. Chang, C. C. and Lin, C.-J., "LIBSVM: A library for support vector machines", ***ACM Transactions On Intelligent Systems And Technology***, 2 (3): 1–27 (2011).

55. Tabachnick, B. G., and Fidell, L. S. "Using Multivariate Statistics", Sixth edition, Ed., ***Pearson***, Boston, MA, 1-1018 (2013).

56. Internet: Bayer, "Bayer HealthCare Contour TS User Manual | Manualzz", **https://manualzz.com/doc/55223320/bayerhealthcare-contour-ts-user-manual** (2023).

57. Kotz, S., Read, C. B., Banks, D. L. "Encyclopedia of Statistical Sciences. Update, Volume 1", ***Wiley***, New York, 1-455 (1997).

58. Bland J. M. and Altman, D. G. "STATISTICAL METHODS FOR ASSESSING AGREEMENT BETWEEN TWO METHODS OF CLINICAL MEASUREMENT", ***The Lancet,*** 327 (8476): 307–310 (1986).

59. Internet: Türkiye Diyabet Vakfı, "Diyabet Hakkında Herşey - Diyabet Nedir?", **https://www.turkdiab.org/diyabet-hakkinda-hersey.asp?lang=TR&id=46** (2023).

60. Kolmogorov, A. N., "Sulla Determinizione Empirica di una Legge di Distribuzione", ***Giornale dell Istituto Italiano degli Attuari***, 4: 83-91 (1933).

61. Smirnov, N., “Table for Estimating the Goodness of Fit of Empirical Distributions”, **Annals of Mathematical Statistics**, 19: 279-281 (1948).

62. Massey, F. J., “The Kolmogorov-Smirnov Test for Goodness of Fit”, ***Journal Of The American Statistical Association***, 46 (253): 68–78 (1951).

63. Conover, W. J., "Practical Nonparametric Statistics 2nd ed”. Ed., ***Wiley***, New York, 1-511 (1999).

64. Stephens, M. A., “Use of the Kolmogorov-Smirnov, Cramér-Von Mises and Related Statistics Without Extensive Tables”, ***Journal Of The Royal Statistical Society: Series B (Methodological)***, 32 (1): 115–122 (1970).

65. Kim, Y., White, H. 2004. “On more robust estimation of skewness and kurtosis: Simulation and application to the S&P500 index”, ***Journal of Empirical Finance***, 11(4): 585-603 (2004).

66. Ghimire, P. and Dhamoon, A. S., “Ketoacidosis”, StatPearls, ***StatPearls Publishing***, Treasure Island (FL) (2021).

67. Friedman, J. H., “Greedy function approximation: A gradient boosting machine.”, ***The Annals Of Statistics***, 29 (5): 1189-1232 (2001).

68. Hastie, T., Tibshirani, R., and Friedman, J. H., “The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed.”, Ed., ***Springer***, New York, NY, 1-745 (2009)

RESUME

Taleh BİNNATOV received his primary, middle and high school education in Baku. After that, he joined undergraduate program in Karabuk University Medical Engineering Department in 2015. In 2020, he started his graduate education at Karabuk University Biomedical Engineering. He is currently working as a medical device engineer in a private company that sells and repairs medical devices.