

MICROPROCESSOR BASED EXHALED GAS COMPOSITION ANALYSIS FOR THE DETECTION OF RESPIRATORY DISORDERS

Submitted in partial fulfillment of the requirements for the award of
Bachelor of technology in Biomedical Engineering

by

J. MEENA MANJURI (Reg no:39240025)

P.LOCHANA (Reg no:39240023)



**DEPARTMENT OF BIOMEDICAL ENGINEERING SCHOOL OF BIO
AND CHEMICAL ENGINEERING**

SATHYABAMA

**INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)**

Accredited with Grade "A" by NAAC

JEPPIAAR NAGAR, RAJIV GANDHI SALAI, CHENNAI - 600 119

APRIL-2023

MICROPROCESSOR BASED EXHALED GAS COMPOSITION ANALYSIS FOR THE DETECTION OF RESPIRATORY DISORDERS

Submitted in partial fulfillment of the requirements for the award of
Bachelor of technology in Biomedical Engineering

by

J. MEENA MANJURI (Reg no:39240025)

P.LOCHANA (Reg no:39240023)



**DEPARTMENT OF BIOMEDICAL ENGINEERING SCHOOL OF BIO
AND CHEMICAL ENGINEERING**

SATHYABAMA

**INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)**

Accredited with Grade "A" by NAAC

JEPPIAAR NAGAR, RAJIV GANDHI SALAI, CHENNAI - 600 119

APRIL-2023



SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY)

Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE

www.sathyabama.ac.in

DEPARTMENT OF BIOMEDICAL ENGINEERING

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the Bonafide work of **J.Meena Manjuri (39240025), P.Lochana (39240023)**, who carried out the project work entitled **"Microprocessor based exhaled gas composition analysis for the detection of respiratory disorders"** under my supervision from **October 2022 to April 2023**.

Internal Guide

(Dr.P.GRACE KANMANI PRINCE)

Head of the department

(DR.T. SUDHAKAR)

HEAD OF THE DEPARTMENT
DEPARTMENT OF BIOMEDICAL ENGINEERING
SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Kudathur Nagar Rajiv Gandhi Salai
Chennai - 600 079

Submitted for Viva-voce Examination held on ____

Internal examiner

External examiner

DECLARATION

We, , J. Meena Manjuri, P. Lochana hereby declare that the Project Report entitled "**Microprocessor based exhaled gas composition analysis for the detection of respiratory disorders**" done by us under the guidance of **Dr. P.Grace Kanmani Prince**, at Sathyabama Institute of Science and Technology is submitted in partial fulfillment of the requirements for the award of Bachelor of Technology degree in Biomedical engineering.

DATE:

PLACE: CHENNAI

J. Meena Manjuri

1. J.MEENA MANJURI

Lochana.P

2. LOCHANA.P

SIGNATURE OF THE CANDIDATE

ACKNOWLEDGEMENT

We are pleased to acknowledge my sincere thanks to the **Board of Management of SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. We are grateful to them.

We convey our thanks to **Dr.T.Sudhakar M.Sc.,Ph.D., Head of the Department, Dept. of Biomedical Engineering** for providing us necessary support and details at the right time during the progressive reviews.

We would like to express my sincere and deep sense of gratitude to our **Project Guide Dr.Grace Kanmani Prince.P M.E.,Ph.D.**, for her valuable guidance, suggestions and constant encouragement paved the way for the successful completion of our project work.

We wish to express our thanks to all **Teaching and Non-teaching** staff members of the **Department of Biomedical Engineering** who were helpful in many ways for the completion of the project.

ABSTRACT

Today more than 1 billion people are suffering from respiratory disorders like asthma, tuberculosis, pneumonia etc. So, detecting them at an early stage is very essential. In our project, the human exhaled breath is analyzed using gas sensors (like CO_2 , TVOC), microphone, humidity and temperature sensors. The microphone is used to estimate the breathing pattern. The gas sensors are used to analyze the CO_2 , TVOC levels. The temperature and humidity of the breath is estimated by the humidity and temperature sensors. Analysis of Total Volatile Organic Compounds (TVOCs) emanating from human exhaled breath can provide deep insight into the status of various biochemical processes in the human body. TVOCs can serve as potential biomarkers of physiological and pathological conditions related to several diseases like liver and lung diseases, different types of cancers and infections. Thus, the methods we imply are non-invasive, inexpensive and less time consuming. This ensures the early detection of respiratory disorders. Thus, the methods we imply are noninvasive, inexpensive and less time consuming. This ensures the early detection of respiratory disorders.

KEYWORDS : Exhaled breath, Analysis, respiratory disorders, sensors, noninvasive.

CONTENTS

CHAPTER NO	TITLE	PAGE NO
	ABSTRACT	i
	LIST OF ABBREVIATIONS	ii
	LIST OF FIGURES	iii
1	INTRODUCTION	1
	1.1 BASIC INTRODUCTION	1
	1.2 RESPIRATORY DISEASES: CAUSES, SYMPTOM	2
	1.3 EXHALED BREATH ANALYSIS	3
2	LITERATURE SURVEY	4
3	AIM AND SCOPE	14
	3.1 AIM	
	3.2 SCOPE	
4	MATERIALS AND METHODS	15
	4.1 BLOCK DIAGRAM WITH EXPLANATION	15
	4.2 METHODOLOGY	16
	4.2.1. COLLECTION AND PROCESSING OF THE SAMPLES	16
	4.2.2. CLASSIFICATION OF DATA USING THE NAIVE BAYES CLASSIFIER	17
	4.2.3. CLASSIFICATION OF DATA USING THE RANDOM FOREST CLASSIFIER	18

	4.3	HARDWARE USED	19
	4.3.1	ARDUINO UNO	19
	4.3.2	MICROPHONE MAX – 9814	20
	4.3.3	GAS SENSOR CJMCU-811	21
	4.3.4	LCD DISPLAY	22
	4.3.5	THE TEMPERATURE AND HUMIDITY SENSOR (BME680)	23
	4.3.6	THE ESF NON-CONTACT HUMAN BODY TEMPERATURE SENSOR (MLX90614)	24
	4.4	SOFTWARE USED	25
	4.4.1	DOCKER DESKTOP SOFTWARE	25
5		RESULTS AND DISCUSSION	26
6		SUMMARY AND CONCLUSION	30
	6.1	SUMMARY	
	6.2	CONCLUSION	
		REFERENCES	32
		APPENDIX	34

LIST OF ABBREVIATIONS

S. NO	ABBREVIATION	FULL FORM
1.	AC	Alternating Current
2.	ADC	Analogue Digital Converter
3.	AGC	Automatic Gain Control
4.	ASIC	Application Specific Integrated Circuit
5.	BioVOC	Bio-volatile Organic Compound
6.	BMI	Body Mass Index
7.	BMP	Basic Metabolic Panel
8.	CO	Carbon Monoxide
9.	CO ₂	Carbon dioxide
10.	COPD	Chronic Obstructive Pulmonary Disease
11.	dB	Decibel
12.	DC	Direct Current
13.	DLCO	Diffusing Capacity of the lungs for Carbon monoxide
14.	DSP	Digital Signal Processors
15.	EBC	Electromagnetic Battery Charger
16.	EBT	Electronic Benefits Transfer
17.	eNO	Exhaled Nitric Oxide
18.	ESF	Electro sinter forging
19.	FEV	Forced expiratory volume
20.	FeNO	Fractional Exhaled Nitric Oxide
21.	GND	Ground
22.	HPA	High Power Amplifier
23.	ICSP	In-circuit serial programming
24.	I ² C	Inter-Integrated Circuit
25.	LCD	Liquid Crystal Display
26.	MEMS	Micro Electro Mechanical System
27.	ms	Microsecond
28.	PH	Pulmonary Hypertension

29.	PAH	Pulmonary arterial hypertension
30.	PPM	Parts per million
31.	PWM	Pulse-width modulation
32.	RH	Relative Humidity
33.	SSc	Systemic Sclerosis
34.	SCL	Clock line
35.	SPI	Serial Peripheral Interface
36.	TVOC	Total Volatile Organic Compounds
37.	USB	Universal Serial Bus
38.	VOC	Volatile Organic Compounds
39.	VGA	Video Graphics Array
40.	VCC	Voltage Common Collector
41.	WAK	Wireless Access Kit
42.	WHO	World Health Organization
43.	XH2	X-Linked Helicase 2
44.	°C	Celsius
45.	μF	Micro Farad

LIST OF FIGURES

FIGURE NO	DESCRIPTION	PAGE NO
4.1	Block diagram	15
4.2	Prototype model	16
4.3	Naive Bayes classifier	17
4.4	Random forest classifier	18
4.5	Arduino UNO	19
4.6	Microphone MAX-9814	20
4.7	Gas Sensor(CJMCU 811)	21
4.8	LCD Display	22
4.9	The temperature and humidity sensor(BME680)	23
4.10	The ESF Non-Contact Human Body Temperature	24
4.11	Docker Desktop Software	25
5.1	Breath analyzer - Output	26
5.2	Graph-Normal level-Healthy person	27
5.3	Graph- Abnormal level- Un Healthy person	27
5.4	Output using Naive Bayes and Random Forest classifier - TVOC	28
5.5	Output using Naive Bayes and Random Forest Classifier - CO2	28
5.6	Output using Naive Bayes and Random Forest classifier - Microphone	29

CHAPTER -1

INTRODUCTION

1.1 BASIC INTRODUCTION

Respiration is a metabolic process that occurs in all organisms. It is a biochemical process that occurs within the cells of organisms. Respiration involves two processes Inhalation and Exhalation. Inhalation is a process of taking in air and other substances into our lungs. The air, along with other gases, enters our body through inhalation. Exhalation is defined as the process wherein the air moves out from the lungs through the airways in the body. The respiratory system disorders or respiratory diseases are the medical terms used to study the various types of infections, allergies and other diseases related to the different organs, tissues and specialized cells of the human respiratory system. According to WHO, in India, deaths due to respiratory diseases account for 11% of total deaths in the country. In every one lakh people, an estimated amount of 142 people died from lung disease. Asthma, chronic bronchitis, bronchiectasis and chronic obstructive pulmonary disease (COPD) are all obstructive lung diseases characterised by airway obstruction. This limits the amount of air that is able to enter alveoli because of constriction of the bronchial tree, due to inflammation. Obstructive lung diseases are often identified because of symptoms and diagnosed with pulmonary function tests such as spirometry. Many obstructive lung diseases are managed by avoiding triggers (such as dust mites or smoking), with symptom control such as bronchodilators, and with suppression of inflammation (such as through corticosteroids) in severe cases. One common cause of COPD including emphysema, and chronic bronchitis, is tobacco smoking, and common causes of bronchiectasis include severe infections and cystic fibrosis. The respiratory disease causes an immense health burden. It is estimated that worldwide 235 million people suffer from asthma, more than 200 million people have chronic obstructive pulmonary disease (COPD), 65 million endure moderate-to-severe COPD, 1–6% of the adult population (more than 100 million people)

experience sleep-disordered breathing, 9.6 million people develop tuberculosis (TB) annually, millions live with Pulmonary Hypertension and more than 50 million people struggle with occupational lung diseases, totalling more than 1 billion people suffering from chronic respiratory conditions. At least 2 billion people are exposed to the toxic effects of biomass fuel consumption, 1 billion are exposed to outdoor air pollution and 1 billion are exposed to tobacco smoke. Each year, 4 million people die prematurely from chronic respiratory disease.

1.2 RESPIRATORY DISEASES: CAUSES, SYMPTOMS

The main cause of COPD in developed countries is tobacco smoking. In the developing world, COPD often occurs in people exposed to fumes from burning fuel for cooking and heating in poorly ventilated homes. Only some chronic smokers develop clinically apparent COPD, although many smokers with long smoking histories may develop reduced lung function. Some smokers develop less common lung conditions. They may be misdiagnosed as having COPD until a more thorough evaluation is performed. The symptoms of respiratory diseases; cough, shortness of breath, chest pain and hemoptysis (blood in the sputum) are the predominant symptoms. The cough may present for a long time along with expectoration and sometimes hemoptysis. The presence of blood in the sputum is an ominous finding and the patient should be investigated properly. Asthma patients usually present with cough with expectoration along with easy fatigability may have a seasonal variation. COPD patients are present with cough with expectoration and easy fatigability which are gradually progressive unless properly treated. Tuberculosis is a very common disease and patients usually present with cough, expectoration, hemoptysis, weight loss and myriad other symptoms. Lung cancer is currently the second most common cancer in both men and women. Cigarette smoking is the predominant cause of lung cancer. In non-smokers, it may develop due to radon exposure, air pollution, and asbestos exposure, etc. Lung cancer patients are usually present with cough, hemoptysis, weight loss, accumulation of fluid in the pleural cavity, etc. The prognosis is usually grave if the patients are in the late stage of their disease.

1.3 EXHALED BREATH ANALYSIS

Human exhaled breath mostly contains, nitrogen (78.04%), oxygen [16%], carbon dioxide [4%-5%], hydrogen [5%], inert gases [0.9%] and water vapor. Exhaled breath analysis is an important area of research work which has gained tremendous interest recently owing to the advances in analytical techniques and nanotechnology. The analysis of exhaled breath to detect VOCs could indicate the diseased or healthy state of an individual. Researchers have documented the evidence of the presence of detectable VOCs in the breath related to breast and lung cancer. This is a non-invasive method for disease detection, therapeutic monitoring, and metabolic status monitoring by analysing the volatile organic compounds (VOCs) present in the exhaled breath. Along with point of care detection breath analysis has the advantage of non-invasive, cost-effective, real time, qualitative/quantitative disease diagnosis and hence it is striking. On-line analysis of exhaled breath offers insight into a person's metabolism without the need for sample preparation or sample collection. Due to its noninvasive nature and the possibility to sample continuously, the analysis of breath has great clinical potential. The breath samples hold an 86% diagnostic accuracy rate for disease detection.

CHAPTER-2

LITERATURE SURVEY

Abd Alghani Khamis , Aida Idris ,etal, Development and Performance Evaluation of an lo T Integrated Breath Analyzer, International Journal of Environmental Research and Public Health, 2023, Although alcohol consumption may produce effects that can be beneficial or harmful, alcohol consumption prevails among communities around the globe . Additionally, alcohol consumption patterns may be associated with several factors among communities and individuals. Numerous technologies and methods are implemented to enhance the detection and tracking of alcohol consumption, such as vehicle-integrated and wearable devices. In this paper, we present a cellular-based Internet of Things (IoT) implementation in a breath analyzer to enable data collection from multiple users via a single device. Cellular technology using hypertext transfer protocol (HTTP) was implemented as an IoT gateway. IoT integration enabled the direct retrieval of information from a database relative to the device and direct upload of data from the device onto the database. A manually developed threshold algorithm was implemented to quantify alcohol concentrations within a range from 0 to 200 mcg/100 mL breath alcohol content using electrochemical reactions in a fuel-cell sensor. Two data collections were performed: one was used for the development of the model and was split into two sets for model development and on-machine validation, and another was used as an experimental verification test. An overall accuracy of 98.16% was achieved, and relative standard deviations within the range from 1.41% to 2.69% were achieved, indicating the reliable repeatability of the results. The implication of this paper is that the developed device (an lo T-integrated breath analyzer) may provide practical assistance for healthcare representatives and researchers when conducting studies involving the detection and data collection of alcohol consumption patterns.

Anna Paleczek, Artur Rydosz, Review of the algorithms used in exhaled breath analysis for the detection of diabetes, *Journal of Breath Research*, 2022, Currently, intensive work is underway on the development of truly noninvasive medical diagnostic systems, including respiratory analysers based on the detection of biomarkers of several diseases including diabetes. In terms of diabetes, acetone is considered as one of the potential biomarkers, although it is not the single one. Therefore, the selective detection is crucial. Most often, the analysers of exhaled breath are based on the utilization of several commercially available gas sensors or on specially designed and manufactured gas sensors to obtain the highest selectivity and sensitivity to diabetes biomarkers present in the exhaled air. An important part of each system are the algorithms that are trained to detect diabetes based on data obtained from sensor matrices. The prepared review of the literature showed that there are many limitations in the development of the versatile breath analyser, such as high metabolic variability between patients, but the results obtained by researchers using the algorithms described in this paper are very promising and most of them achieve over 90% accuracy in the detection of diabetes in exhaled air. This paper summarizes the results using various measurement systems, feature extraction and feature selection methods as well as algorithms such as support vector machines, k-nearest neighbors and various variations of neural networks for the detection of diabetes in patient samples and simulated artificial breath samples.

Bartolomé R. Celli, Jadwiga A. Wedzicha, Update on Clinical Aspects of Chronic Obstructive Pulmonary Disease, *the new England journal of medicine*, 2019. Chronic obstructive pulmonary disease (COPD) is the third leading cause of death worldwide; COPD led to 3.2 million deaths in 2017, a toll expected to reach 4.4 million yearly by 2040.^{1,2} With a worldwide prevalence of 10.1%, COPD afflicts many people in low-income, middle-income, and wealthy countries (Fig. 1), and years of life lost prematurely increased 13.2% between 2007 and 2017.¹ Although COPD has traditionally been considered a disease that affects men, in some countries, the prevalence and associated mortality are higher among women than among men. In this review, we update the clinical face of COPD, concentrating on

the pulmonary aspects of the disease, which also affects many other organ systems. The pathogenesis of COPD is discussed in a companion article by Agustí and Hogg in this issue of the Journal, 3 and the review of muco - obstructive lung diseases in a recent issue of the Journal⁴ complements this article.

Bhagaban Behera, Rathin Joshi , Anil Vishnu GK, Sanjay Bhalerao , Hardik J Pandya, Electronic nose: a non-invasive technology for breath analysis of diabetes and lung cancer patients, Journal of Breath Research, 2019 . In human exhaled breath, more than 3000 volatile organic compounds (VOCs) are found, which are directly or indirectly related to internal biochemical processes in the body. Electronic noses (E-noses) could play a potential role in screening/analyzing various respiratory and systemic diseases by studying breath signatures. An E-nose integrates a sensor array and an artificial neural network that responds to specific patterns of VOCs, and thus can act as a non-invasive technology for disease monitoring. The gold standard blood glucose monitoring test for diabetes diagnostics is invasive and highly uncomfortable. This contributes to the massive need for technologies which are non-invasive and can be used as an alternative to blood measurements for glucose detection. While lung cancer is one of the deadliest cancers with the highest death rate and an extremely high yearly global burden, the conventional diagnosis means, such as sputum cytology, chest radiography, or computed tomography, do not support wide-range population screening. A few standard non-invasive techniques, such as mass spectrometry and gas chromatography, are expensive, non-portable, and require skilled personnel for operation and are again not suitable for large-scale screening. Breath contains markers for both diabetes and lung cancer along with markers for several diseases and thus, a non-invasive technique such as the E-nose would greatly improve analysis procedures over existing invasive methods. This review shows the state-of-the-art technologies for VOC detection and machine learning approaches for two clinical models: diabetes and lung cancer detection.

Catharina C. Moor, Judith C. Oppenheimer, et al , Exhaled breath analysis by use of eNose technology: a novel diagnostic tool for interstitial lung disease, *European Respiratory Journal*, 2021, Early and accurate diagnosis of interstitial lung diseases (ILDs) remains a major challenge. Better noninvasive diagnostic tools are much needed. We aimed to assess the accuracy of exhaled breath analysis using eNose technology to discriminate between ILD patients and healthy controls, and to distinguish ILD subgroups. In this cross-sectional study, exhaled breath of consecutive ILD patients and healthy controls was analysed using eNose technology (Spiro Nose). Statistical analyses were done using partial least square discriminant analysis and receiver operating characteristic analysis. Independent training and validation sets (2:1) were used in larger subgroups. A total of 322 ILD patients and 48 healthy controls were included: sarcoidosis (n=141), idiopathic pulmonary fibrosis (IPF) (n=85), connective tissue disease-associated ILD (n=33), chronic hypersensitivity pneumonitis (n=25), idiopathic nonspecific interstitial pneumonia (n=10), interstitial pneumonia with autoimmune features (n=11) and other ILDs (n=17). eNose sensors discriminated between ILD and healthy controls, with an area under the curve (AUC) of 1.00 in the training and validation sets. Comparison of patients with IPF and patients with other ILDs yielded an AUC of 0.91 (95% CI 0.85–0.96) in the training set and an AUC of 0.87 (95% CI 0.77–0.96) in the validation set. The eNose reliably distinguished between individual diseases, with AUC values ranging from 0.85 to 0.99. eNose technology can completely distinguish ILD patients from healthy controls and can accurately discriminate between different ILD subgroups. Hence, exhaled breath analysis using eNose technology could be a novel biomarker in ILD, enabling timely diagnosis in the future

Jan Krajsa, Lubomír Straka, et al , Confectionary containing alcohol and their effect on breath analyzer results: a preliminary study, *Journal of Slovak republic*, 2022 . A variety of food and soft drinks were tested and found to contain low concentration of alcohol. The potential for these products to generate false positive results on an evidential breath-alcohol analyzers (1). It is well known that alcohol is retained in the mouth for a short period following consumption of alcohol-

containing breath sprays, mouthwashes, or alcohol-fortified food (1,2). Other examples of potential confounders are E-cigarette smoking (as the liquid may contain ethanol), use of alcohol-based hand rubs, disinfectants, or cleaners and, oddly, alcohol spill in the car (3,4). Breath-alcohol tests given immediately following the use of these products can cause inaccurate results, and a 15-min alcohol deprivation period is there an appropriate part of any evidential breath-alcohol test (1,2,5). The kinetics of mouth-alcohol elimination can be affected by dental adhesives or even oral jewelry in a pierced tongue (1). In the case of positive breath test result, it is necessary to exclude all possibilities of a distorted result. Recently in expert activities, questions regarding if, how and especially how long, food (most often confectionery for children) can affect the results of alcohol-breath tests, appear more and more often. Previous experiments, developed by authors, described cases in which driver breath out alcohol vapor even after 13 minutes post washing the windscreen by alcohol-contained liquid (6). This fact led to attempt an experiment with over-the-counter confectionery, available also at petrol stations, to measure the alcohol concentration after consumption and their possibility to react with breath analyzers.

Jyoti Kanwar Shekhawat, Mithu Banerjee, Role of Breath Biopsy in COVID-19, The Journal of applied Laboratory Medicine, 2022. COVID-19 is a highly contagious respiratory disease that can be transmitted through human exhaled breath. It has caused immense loss and has challenged the healthcare sector. It has affected the economy of countries and thereby affected numerous sectors. Analysis of human breath samples is an attractive strategy for rapid diagnosis of COVID-19 by monitoring breath biomarkers. Breath collection is a noninvasive process. Various technologies are employed for detection of breath biomarkers like mass spectrometry, biosensors, artificial learning, and machine learning. These tools have low turnaround time, robustness, and provide onsite results. Also, MS-based approaches are promising tools with high speed, specificity, sensitivity, reproducibility, and broader coverage, as well as its coupling with various chromatographic separation techniques providing better clinical and biochemical understanding of COVID-19 using breath samples. Herein, we have tried to review

the MS-based approaches as well as other techniques used for the analysis of breath samples for COVID-19 diagnosis. We have also highlighted the different breath analyzers being developed for COVID-19 detection.

Ileana Andreea Ratiu, Tomasz Ligor, Volatile Organic Compounds in Exhaled Breath as Fingerprints of Lung Cancer, Asthma and COPD, *Journal of Clinical Medicine*, 2020. Lung cancer, chronic obstructive pulmonary disease (COPD) and asthma are inflammatory diseases that have risen worldwide, posing a major public health issue, encompassing not only physical and psychological morbidity and mortality, but also incurring significant societal costs. The leading cause of death worldwide by cancer is that of the lung, which, in large part, is a result of the disease often not being detected until a late stage. Although COPD and asthma are conditions with considerably lower mortality, they are extremely distressful to people and involve high healthcare overheads. Moreover, for these diseases, diagnostic methods are not only costly but are also invasive, thereby adding to people's stress. It has been appreciated for many decades that the analysis of trace volatile organic compounds (VOCs) in exhaled breath could potentially provide cheaper, rapid, and non-invasive screening procedures to diagnose and monitor the above diseases of the lung. However, after decades of research associated with breath biomarker discovery, no breath VOC tests are clinically available. Reasons for this include the little consensus as to which breath volatiles (or pattern of volatiles) can be used to discriminate people with lung diseases, and our limited understanding of the biological origin of the identified VOCs. Lung disease diagnosis using breath VOCs is challenging. Nevertheless, the numerous studies of breath volatiles and lung disease provide guidance as to what volatiles need further investigation for use in differential diagnosis, highlight the urgent need for non-invasive clinical breath tests, illustrate the way forward for future studies, and provide significant guidance to achieve the goal of developing non-invasive diagnostic tests for lung disease. This review provides an overview of these issues from evaluating key studies that have been undertaken in the years 2010–2019, in order to present objective and comprehensive updated information that presents the progress that has been made in this field. The potential of this approach is

highlighted, while strengths, weaknesses, opportunities, and threats are discussed. This review will be of interest to chemists, biologists, medical doctors and researchers involved in the development of analytical instruments for breath diagnosis.

Patrik Španěl, David Smith, Quantification of volatile metabolites in exhaled breath by selected ion flow tube mass spectrometry, SIFT-MS, Volume 16, 2020. Selected ion flow tube mass spectrometry, SIFT-MS, is a non-separative method for direct quantitative analyses of volatile compounds, VOCs, in air and humid breath based on chemical ionization. Selected reagent ions, either H_3O^+ , NO^+ or O_2^+ (non-reactive with major components of air), ionize analyte molecules during a defined time in a flow tube by ion-molecule reactions thus producing analyte ions that are characteristic of the neutral analyte VOCs. Concentrations can be calculated in real-time from the ion count rates. Direct on-line analysis of single or multiple breath exhalations or off-line analysis of breath samples collected into bags can be performed. Several volatile breath metabolites have been quantified by SIFT-MS, including ammonia, acetone, hydrogen cyanide, alcohols, pentane, acetic acid, methane, and sulphur compounds. Their potential as biomarkers is discussed.

Peter J. Mazzone, Analysis of Volatile Organic Compounds in the Exhaled Breath for the Diagnosis of Lung Cancer, Journal of Thoracic Oncology, 2015. Volatile organic compounds are able to be detected in the exhaled breath by a variety of sensing techniques. These volatiles may be produced by cellular metabolic processes, or inhaled/absorbed from exogenous sources. Lung cancer cells may produce and process these compounds differently than normal cells. The differences may be detectable in the breath. The following manuscript will review the evidence supporting the premise that a unique chemical signature can be detected in the breath of patients with lung cancer, discuss the results of studies using mass spectrometry and nonspecific chemical sensing techniques to detect the unique lung cancer signature, and speculate on the advancements that must occur to develop a breath test accurate enough to be clinically useful.

RosaAlba , Sola Martínez, et al , Data preprocessing work flow of exhaled breath analysis by GC/ MS using open sources, Scientific Reports, 2020. The noninvasive diagnosis and monitoring of high prevalence diseases such as cardiovascular diseases, cancers and chronic respiratory diseases are currently priority objectives in the area of health. In this regard, the analysis of volatile organic compounds (VOCs) has been identified as a potential noninvasive tool for the diagnosis and surveillance of several diseases. Despite the advantages of this strategy, it is not yet a routine clinical tool. The lack of reproducible protocols for each step of the biomarker discovery phase is an obstacle of the current state. Specifically, this issue is present at the data preprocessing step. Thus, an open source workflow for preprocessing the data obtained by the analysis of exhaled breath samples using gas chromatography coupled with single quadrupole mass spectrometry (GC/MS) is presented in this paper. This workflow is based on the connection of two approaches to transform raw data into a useful matrix for statistical analysis. Moreover, this workflow includes matching compounds from breath samples with a spectral library. Three free packages (xcms, cliqueMS and eRah) written in the language R are used for this purpose. Furthermore, this paper presents a suitable protocol for exhaled breath sample collection from infants under 2 years of age for GC/MS.

Sagnik Das and Mrinal Pal ,Non-Invasive Monitoring of Human Health by Exhaled Breath Analysis: A Comprehensive Review,Journal of The Electrochemical Society, 2020. Exhaled human breath analysis is a very promising field of research work having great potential for diagnosis of diseases in non-invasive way. Breath analysis has attracted huge attention in the field of medical diagnosis and disease monitoring in the last two decades. VOCs/gases (Volatile Organic Compounds) in exhaled breath bear the finger-prints of metabolic and biophysical processes going on in human body. It's a non-invasive, fast, non-hazardous, cost effective, and point of care process for disease state monitoring and environmental exposure assessment in human beings. Some VOCs/gases in exhaled breath are biomarkers of different diseases and their presence in excess amount is indicative of

un-healthiness. Breath analysis has the potential for early detection of diseases. However, it is still underused and commercial device is yet not available owing to multifarious challenges. This review is intended to provide an overview of major biomarkers (VOCs/gases) present in exhaled breath, importance of their analysis towards disease monitoring, analytical techniques involved, promising materials for breath analysis etc. Finally, related challenges and limitations along with future scope will be touched upon.

Sean P.Duffy , Gerard J.Criner, Chronic Obstructive Pulmonary disease Evaluation and Management, Medical Clinics, 2019. Chronic obstructive pulmonary disease (COPD) is a leading cause of death nationally and worldwide. Cigarette smoking is the most common risk factor in the development of COPD. Disease course is variable with some patients having a high degree of obstruction and minimal symptoms, whereas others with better lung function have a greater symptoms burden. The goal of pharmacologic therapy is to minimize symptoms, improve exercise tolerance, and reduce exacerbation risk. No pharmacologic therapy has been shown to improve survival in COPD. Pulmonology referral is recommended for patients with COPD with symptoms despite first-line inhaled therapy, frequent exacerbations, any hospitalizations, or moderate-to-severe disease.

Ting Chen , Tiannan Liu , Ting Li, et al , Exhaled breath analysis in disease detection, Elsevier, 2021 . Investigating the use of exhaled breath analysis to diagnose and monitor different diseases has attracted much interest in recent years. This review introduces conventionally used methods and some emerging technologies aimed at breath analysis and their relevance to lung disease, airway inflammation, gastrointestinal disorders, metabolic disorders and kidney diseases. One section correlate breath components and specific diseases, whereas the other discusses some unique ideas, strategies, and devices to analyze exhaled breath for the diagnosis of some common diseases. This review aims to briefly introduce the potential application of exhaled breath analysis for the diagnosis and

screening of various diseases, thereby providing a new avenue for the detection of non-invasive diseases.

Wadah Ibrahim, Rebecca L. Cordell, Michael J. Wilde, Diagnosis of COVID-19 by exhaled breath analysis using gas chromatography–mass spectrometry, European Respiratory Society, 2021. The ongoing corona virus disease 2019 (COVID-19) pandemic has claimed over two and a half million lives worldwide so far. Severe acute respiratory syndrome corona virus 2 (SARS-CoV-2) infection is perceived to be seasonally recurrent, and a rapid noninvasive biomarker to accurately diagnose patients early on in their disease course will be necessary to meet the operational demands for COVID-19 control in the coming years. The aim of this study was to evaluate the role of exhaled breath volatile biomarkers in identifying patients with suspected or confirmed COVID-19 infection, based on their underlying PCR status and clinical probability. The prospective, real-world, observational study was carried out, recruiting adult patients with suspected or confirmed COVID-19 infection. Breath samples were collected using a standard breath collection bag, modified with appropriate filters to comply with local infection control recommendations, and samples were analysed using gas chromatography–mass spectrometry (TD-GC-MS). 81 patients were recruited between April 29 and July 10, 2020, of whom 52 out of 81 (64%) tested positive for COVID-19 by reverse transcription–polymerase chain reaction (RT-PCR). A regression analysis identified a set of seven exhaled breath features (benzaldehyde, 1-propanol, 3,6-methylundecane, camphene, beta-cubebene, iodobenzene and an unidentified compound) that separated PCR- positive patients with an area under the curve (AUC): 0.836, sensitivity: 68%, specificity: 85%. GC-MS- detected exhaled breath biomarkers were able to identify PCR- positive COVID-19 patients. External replication of these compounds is warranted to validate these results.

CHAPTER-3

AIM AND SCOPE

3.1 AIM

- Using Arduino incorporated with mic (MAX – 9814) and gas sensor (CJMCU – 811) we are diagnosing the respiratory disorder.
- The mic (MAX – 9814) is a sensitive sensor which is used to analyze the breathing patterns of an individual.
- The gas sensor (CJMCU – 811) is used to estimate the amount of CO₂ and TVOC (Total Volatile Organic Compounds) in the exhaled gas.
- The temperature and humidity sensor (BME680) are used to estimate the temperature and humidity of the exhaled breath.
- The ESF non-contact human body temperature sensor (MLX90614) is used to measure the temperature of the human body.
- And all the values and Datas will be displayed in an LCD display.

3.2 SCOPE

- It is essential to monitor Breathing Pattern, CO₂ and TVOC, temperature and humidity levels for the early diagnosis of respiratory disorder.
- The scope of this project is to epitomize a new risk free and potentially inexpensive diagnostics of respiratory disorders.

CHAPTER - 4

MATERIALS AND METHODS

The Human exhaled breath is analyzed using gas sensors (like Co₂, TVOC), microphone, humidity and temperature sensors. The microphone is used to estimate the breathing pattern. The gas sensors are used to analyze the co₂, TVOC levels. The temperature and humidity of the breath is estimated by the humidity and temperature sensors. The methods that are implied are Non-Invasive.

4.1 BLOCK DIAGRAM WITH EXPLANATION

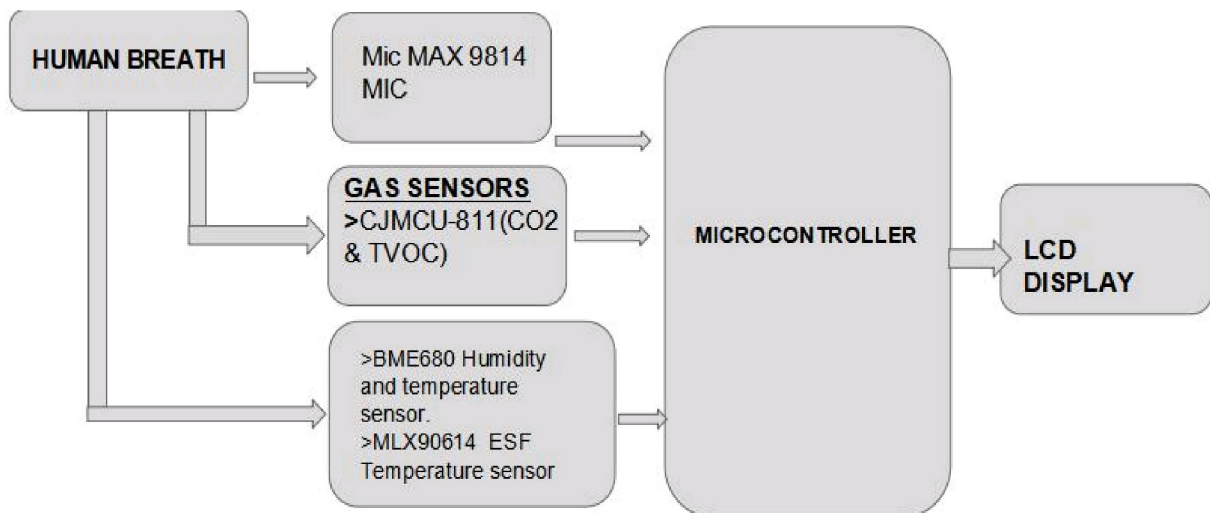


Fig 4.1 Block diagram

As in (fig4.1), when we exhale, the exhaled breath of an human being is analyzed by highly sensitive sensors Mic MAX 9814, gas sensor CJMCU-811, the temperature and humidity sensor (BME680), the ESF non-contact human body temperature sensor (MLX90614). These sensors help to record the breathing pattern, CO₂ level, TVOC level along with the temperature and humidity of the exhaled breath and then these data are delivered to the Arduino. The Arduino

does the necessary processing. And then the digital output is finally displayed on an LCD display.

4.2. METHODOLOGY

4.2.1. COLLECTION AND PROCESSING OF THE SAMPLES



FIG. 4.2. Prototype Model

Using the prototype model as in fig (4.2) , collect the exhaled breath sample using the sensors incorporated with the micro controller . The Arduino uno is connected to the LCD crystal display which shows the temperature, humidity, Co2 and TVOC levels along with breath rate here the breath rate is initially zero and, in the program, it is multiplied by 6 times the breath rate is taken for about 10 seconds so that according to the program it is multiplied by 6 which gives a value for 1 minute. and the Arduino is connected to four sensitive sensors CJMCU 811, BME680, MLX90614 and MAX 9814 CJMCU-811 which is used to calculate and show the Co2 and TVOC levels, temperature and humidity levels in LCD display. When exhaled the co2 and TOC levels increase and when inhaled the co2 and TVOC levels decrease in LCD display.

4.2.2. CLASSIFICATION OF DATA USING THE NAIVE BAYES CLASSIFIER

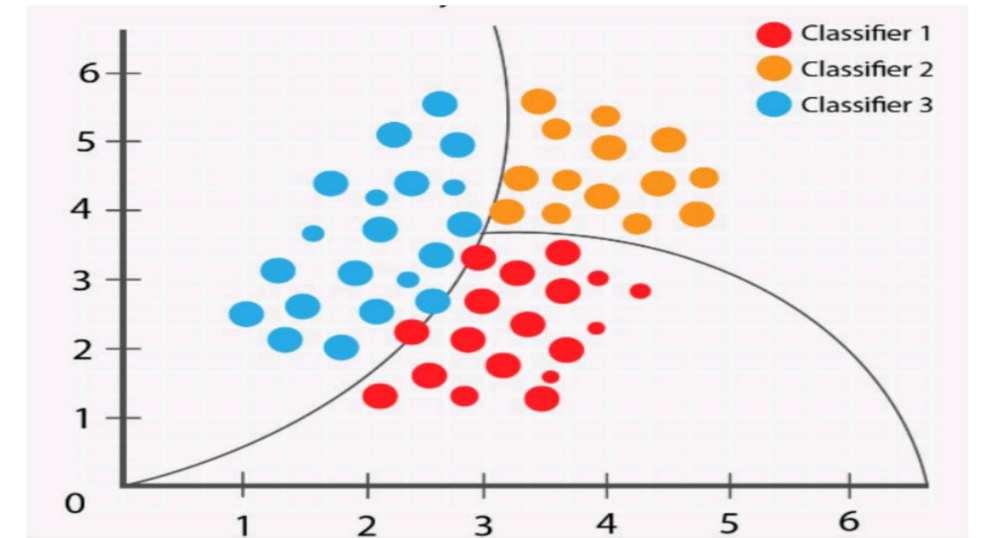


FIG 4.3. Naive Bayes Classifier

Naive Bayes is a simple technique for constructing classifiers: models that assign class labels to problem instances, represented as vectors of feature values, where the class labels are drawn from some finite set. There is not a single algorithm for training such classifiers, but a family of algorithms based on a common principle: all naive Bayes classifiers assume that the value of a particular feature is independent of the value of any other feature, given the class variable. For example, a fruit may be considered to be an apple if it is red, round, and about 10 cm in diameter. A naive Bayes classifier considers each of these features to contribute independently to the probability that this fruit is an apple, regardless of any possible correlations between the color, roundness, and diameter features.

In many practical applications, parameter estimation for naive Bayes models uses the method of maximum likelihood; in other words, one can work with the naive Bayes model without accepting Bayesian probability or using any Bayesian methods. Using this technique the collected samples are classified as normal and abnormal values. The accuracy of this Naive Bayes technique is 83.33%

4.2.3. CLASSIFICATION USING RANDOM FOREST CLASSIFIER

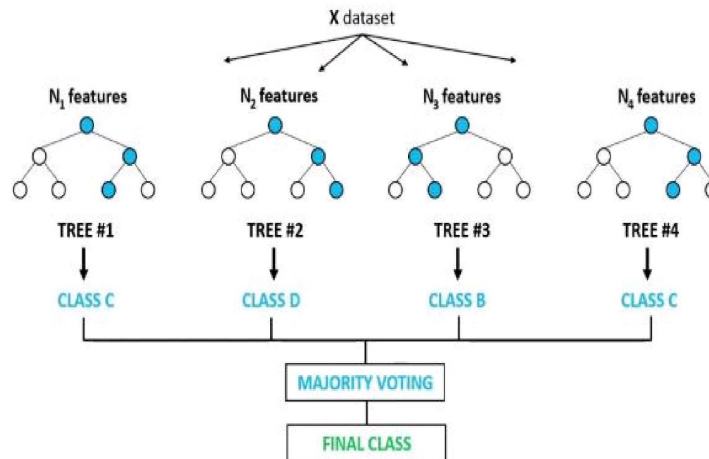


FIG 4.4 Random Forest Classifier

Random Forest (fig 4.4) is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model. As the name suggests, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given data set and takes the average to improve the predictive accuracy of that data set." Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output. The greater number of trees in the forest leads to higher accuracy and prevents the problem of over fitting. Using this technique the collected samples are classified as normal and abnormal values. The accuracy of this Random Forest Classifier technique is 100%.

4.3. HARDWARE USED

4.3.1. ARDUINO UNO



Fig 4.5 Arduino UNO

Arduino Uno(fig 4.2) is a micro controller board based on the ATmega328P. It has 14 digital input/output pins (of which 6 can be used as PWM outputs), 6 analog inputs, a 16 MHz quartz crystal, a USB connection, a power jack, an ICSP header and a reset button. It contains everything needed to support the micro controller; we can simply connect it to a computer with a USB cable or power it with an AC-to-DC adapter or battery to get started. "Uno" means one in Italian. The Uno board is the first in a series of USB Arduino boards, and the reference model for the Arduino platform; for an extensive list of current, past or outdated boards see the Arduino index of boards. Arduino UNO is a low-cost, flexible, and easy-to-use programmable open-source micro controller board that can be integrated into a variety of electronic projects. This board can be interfaced with other Arduino boards, Arduino shields, Raspberry Pi boards and can control relays, LED s, servos, and motors as an output.

4.3.2. MICROPHONE MAX-9814



Fig 4.6 Microphone MAX-9814

The MAX-9814(fig4.3) is a low-cost, high-performance microphone amplifier with automatic gain control (AGC) and low noise microphone bias. The device features a low-noise front-end amplifier, a variable gain amplifier (VGA), an output amplifier, a microphone bias voltage generator, and an AGC control circuit. The low-noise pre amplifier has a fixed 12dB gain, while the VGA gain automatically adjusts from 20dB to 0dB, depending on the output voltage and the AGC threshold. The output amplifier offers select able gains of 8dB, 18dB, and 28dB. With no compression, the cascade of the amplifiers results in an overall gain of 40dB, 50dB, or 60dB. A tri-level digital input programs the output amplifier gain. An external resistive divider controls the AGC threshold and a single capacitor programs the attack/release times. A tri-level digital input programs the ratio of attack-to-release time. The hold time of the AGC is fixed at 30ms. The low-noise microphone-bias-voltage generator can bias most electric microphones. We can also change the Attack/Release ratio, from the default 1:4000 to 1:2000 or 1:500. The output from the amp is about 2Vpp max on a 1.25V DC bias, so it can be easily used with any Analog/Digital converter that is up to 3.3V input. If you want to pipe it into a Line Input, just use a 1-100uF blocking capacitor in series (100uF sounds best).

4.3.3. GAS SENSOR CJMCU-811

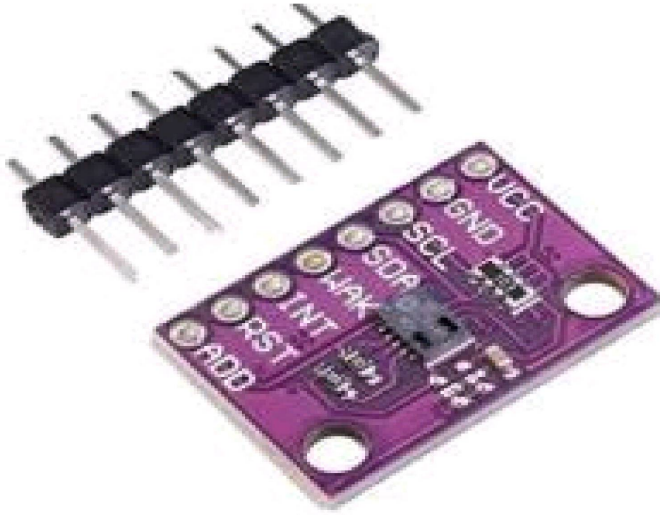


Fig 4.7 Gas Sensor CJMCU-811

The CJMCU-811(fig4.4) is a digital sensor for measuring changes in tVOCs(total volatile organic compounds), in particular Carbon Dioxide (CO₂). Simply to supply power, connect WAK to Ground, and connect the SDA and SCL pins to the I2C pins on your Arduino, ESP8266, Raspberry Pi, etc. Operating voltage of this sensor ranges from 3~5VDC. 2C address 0x5A.It is easy to use on a breadboard with the included 8-pin header. CCS811 Carbon Monoxide (CO) Breakout Board for Arduino Raspberry Pi ESP8266 I2C 3~5VDC CJMCU-811.

Features:

- I2C bus Interface
- Ultra-low power consumption can be used in battery powered equipment
- high sensitivity
- fast heating
- Integrated MCU
- Onboard Processing
- Five Operating Modes

4.3.4. LCD DISPLAY



Fig 4.8 LCD Display

This LCD Screen(fig 4.5) is using an I2C communication interface. It means it only needs 4 pins for the LCD display: VCC, GND, SDA, SCL. It will save at least 4 digital/analog pins on Arduino. All connectors are standard XH2. An electronic device that is used to display data and the message is known as LCD 16×2. As the name suggests, it includes 16 Columns & 2 Rows so it can display 32 characters ($16 \times 2 = 32$) in total & every character will be made with 5×8 (40) Pixel Dots. The I2C protocol involves using two lines to send and receive data: a serial clock pin (SCL) that the Arduino Master board pulses at a regular interval, and a serial data pin (SDA) over which data is sent between the two devices.

4.3.5. THE TEMPERATURE AND HUMIDITY SENSOR (BME680)

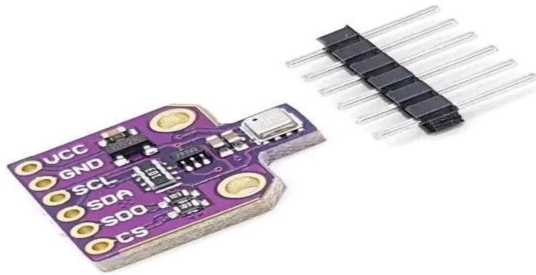


Fig 4.9 Temperature and humidity sensor (BME680)

The BME680(fig 4.6) Environmental Sensor is a breakout that combines a gas sensor with temperature, humidity and barometric pressure sensing for a complete environmental sensor in a single package. The gas sensor on the BME680 can detect a wide variety of volatile organic compounds (or VOC for short) to monitor indoor air quality. Combine that with precise temperature, humidity, and barometric pressure and the BME680 can work as a completely standalone environmental sensor all in a 1in. x 1in. breakout. It is especially developed for mobile applications and wearables where size and low power consumption are critical requirements. The sensor communicates via either I²C or SPI. We've broken out both the I²C and SPI pins to standard 0.1in.-spaced pins. Like the BME280 & BMP280, this precision sensor from Bosch can measure humidity with $\pm 3\%$ accuracy, barometric pressure with ± 1 hPa absolute accuracy, and temperature with $\pm 1.0^{\circ}\text{C}$ accuracy. Because pressure changes with altitude, and the pressure measurements are so good, you can also use it as an altimeter with ± 1 meter or better accuracy.

4.3.6. THE ESF NON CONTACT HUMAN BODY TEMPERATURE SENSOR (MLX90614)

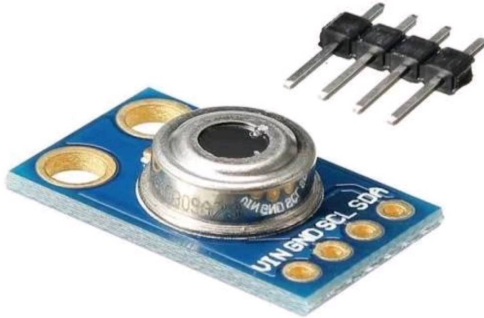


Fig 4.10 The ESF Non-Contact Human Body Temperature Sensor (MLX90614)

The MLX-90614 ESF(fig4.7) is an Infra-Red thermometer for non-contact temperature measurements. Both the IR-sensitive thermopile detector chip and the signal conditioning ASIC are integrated into the same TO-39 can. The Integrated MLX90614 GY-906 is a low noise amplifier, 17-bit ADC, and powerful DSP unit thus achieving high accuracy and resolution of the thermometer. The user can configure the digital output to be PWM. As a standard, the 10-bit PWM is configured to continuously transmit the measured temperature in the range of -20 to 120 °C, with an output resolution of 0.14 °C.

Features: Works from 3.3V to 5V input, Module has power regulator IC built-in. Standard I2C interface with built 2x pull up resistors. When measuring the temperature, please maintain a measuring distance of 1 cm.

4.4. SOFTWARE USED

4.4.1 DOCKER DESKTOP SOFTWARE

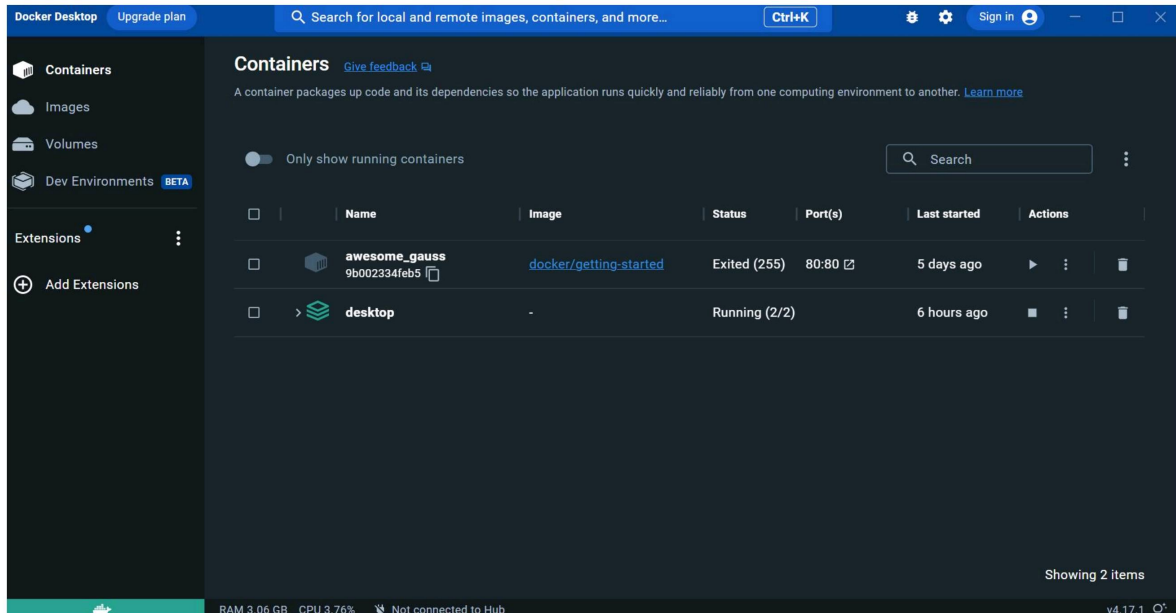


Fig 4.11 Docker Desktop Software

Docker Desktop(fig 4.8) is a one-click-install application for your Mac, Linux, or Windows environment that enables you to build and share containerized applications and micro-services.Docker is an open platform for developing, shipping, and running applications. Docker enables you to separate your applications from your infrastructure so you can deliver software quickly. With Docker, you can manage your infrastructure in the same ways you manage your applications. Docker Desktop does not run processes directly on your Mac or Windows machine, instead Docker Desktop provides a virtual machine that runs Linux, and when you ask Docker to run a command it will run that command on this virtual machine. This is useful technology and makes Docker available to most working programmers.

CHAPTER - 5

RESULTS AND DISCUSSION

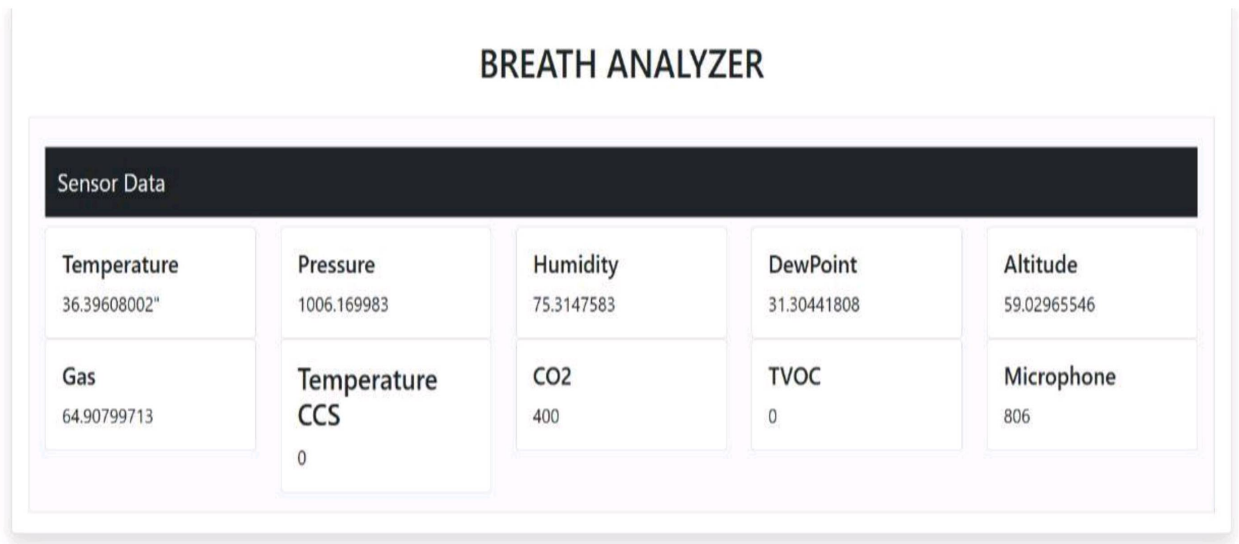


FIG 5.1 Breath Analyzer - Output

Hence, the breath analyzer (fig 5.1) shows the CO₂ and TVOC levels, temperature and humidity levels of the exhaled breath have been obtained by using the sensors and the breathing pattern is also obtained using mic.

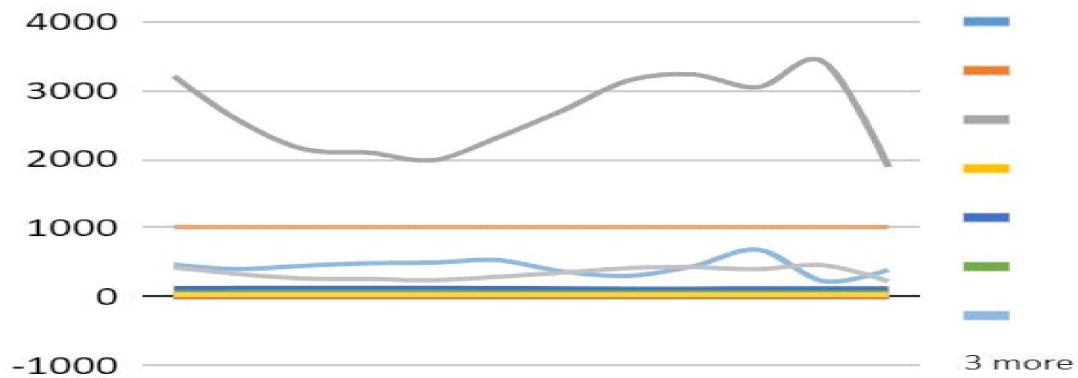


FIG 5.2 Graph - Normal Level - Healthy Person

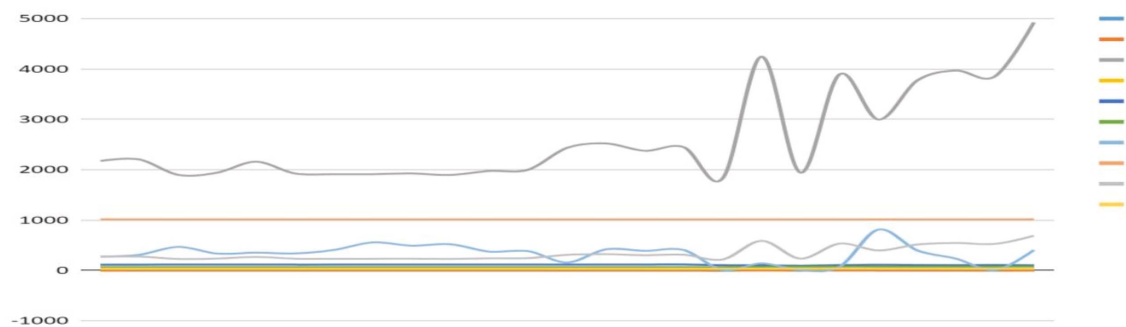


FIG 5.3 Graph - Abnormal Level - Unhealthy Person (With Respiratory Disease)

As in the fig 5.2 the graph represents the individual with normal levels of CO₂,TVOC, humidity and temperature along with normal breath rate .

The fig 5.3 , the graph represents the individual with abnormal levels of CO₂,TVOC, humidity and temperature along with abnormal breath rate.

So the individual has respiratory disease.

The fig5.4 represents the output derived using Naive Bayes and Random forest classifier .The normal and abnormal levels of TVOC is found out.

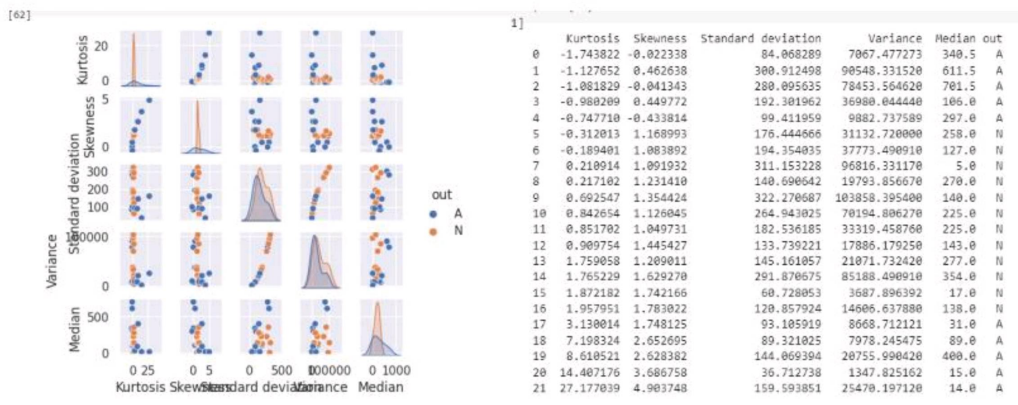


FIG 5.4 Output Using Naive Bayes and Random Forest classifier -TVOC

The fig5.5 represents the output derived using Naive Bayes and Random forest classifier .The normal and abnormal levels of CO2 is found out.

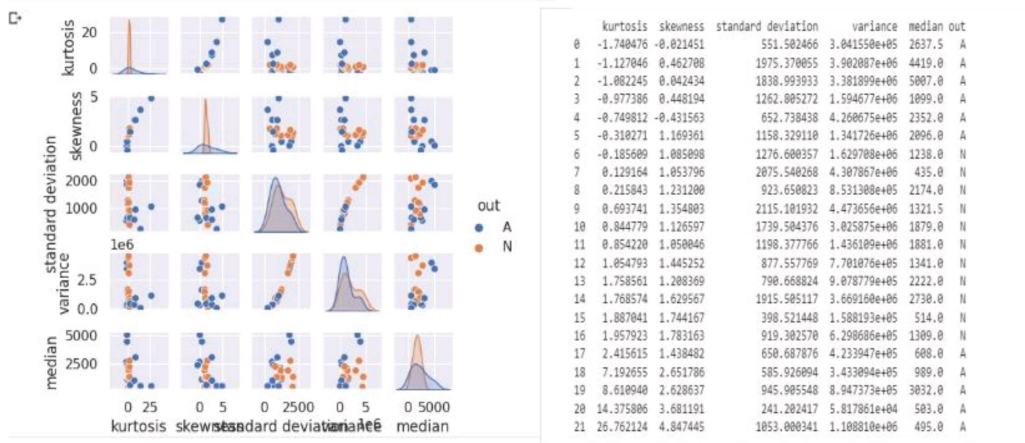


FIG 5.5 Output Using Naive Bayes and Random Forest classifier - CO2

The fig5.6 represents the output derived using Naive Bayes and Random forest classifier .The normal and abnormal levels of Microphone is found out.

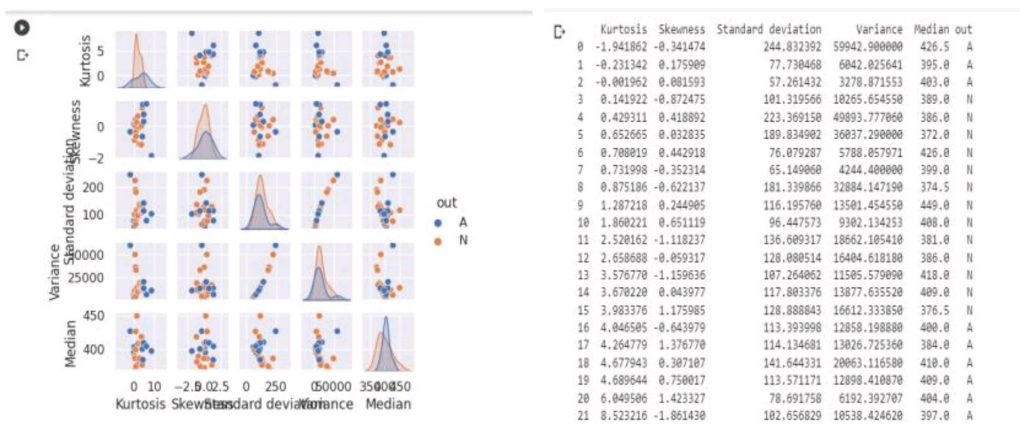


FIG 5.6 Output Using Naive Bayes and Random Forest Classifier-Microphone

Using Naive Bayes and Random Forest classifier ,the normal and abnormal levels of TVOC, CO₂ and Microphone values have been determined .The normal level of TVOC ranges from 0.2 - 1.95 ppb and the values lesser than 0.2 ppb and the values greater than 1.95 are abnormal levels of TVOC in exhaled breath.

Using the same classification technique mentioned above, the normal level of CO₂ ranges from -0.18 - 1.9 ppm and all the values lesser than -0.18 ppm and the values greater than 1.9 ppm are abnormal levels of CO₂ in exhaled breath.

Similarly, the normal level of breath rate measured using microphone ranges from 0.141 - 3.98 and all the values lesser 0.141 and the values greater than 3.98 are abnormal levels of breath rate measured using microphone.

The normal levels of TVOC, CO₂ and Microphone indicates a individual with healthy respiratory system and the abnormal levels of TVOC, CO₂ and Microphone indicates a individual with unhealthy respiratory system .The algorithms that have been used to classify the samples have high accuracy rate.The Naive Bayes classifier have 83.33% accuracy rate. And the Random Forest Classifier have 100% accuracy rate.

CHAPTER - 6

SUMMARY AND CONCLUSION

6.1. SUMMARY

In our project, the human exhaled breath is analyzed using gas sensors (like Co₂, TVOC), microphone, humidity and temperature sensors. The microphone is used to estimate the breathing pattern. The gas sensors are used to analyze the co₂, TVOC levels. The temperature and humidity of the breath is estimated by the humidity and temperature sensors. Finding the respiration rate is essential, identifying abnormal breathing patterns in real-time could be used to gain further insights into respiratory disorders and refine diagnostic procedures. And the samples collected using the prototype has been classified using two algorithms : Naive Bayes classifier and Random Forest classifier. This classification helps to find out the normal and abnormal values of TVOC, CO₂ and Microphone. The composition of exhaled breath may contain valuable information for patients presenting with asthma, renal and liver diseases, lung cancer, chronic obstructive pulmonary disease, inflammatory lung disease, or metabolic disorders. Analysis of Total Volatile Organic Compounds (tVOCs) emanating from human exhaled breath can provide deep insight into the status of various biochemical processes into the human body. tVOCs can serve as potential biomarkers of physiological and patho physiological conditions related to several diseases like liver and lung diseases, different types of cancers and infections. The algorithms that have been used to classify the samples have high accuracy rate. The Naive Bayes classifier have 83.33% accuracy rate. And the Random Forest Classifier have 100% accuracy rate. Thus, the methods we imply are non-invasive, inexpensive and less time consuming. This ensures the early detection of respiratory disorders.

6.2. CONCLUSION

Examination of exhaled breath has the potential to change the existing routine approaches in human medicine. The rapidly developing new analytical and computer technologies along with novel, unorthodox ideas are prerequisites for future advances in this field. Measurement of exhaled breath is safe, rapid, simple to perform, and effort independent. Given that human breath contains upwards of 250 chemicals, the potential for developing new applications is high.

Exhaled breath constitutes more than 3,500 components, the bulk of which are volatile organic compounds in miniature quantities. Many of these characterize the functioning of the organism as a whole (systemic biomarkers), but some are related to processes taking place in the respiratory system and the airways in particular (lung biomarkers). Exhaled breath analysis has enormous potential as an easy, non-invasive means of monitoring inflammation and oxidative stress in the airways, particularly in non-diseased subjects. Exhaled air can be collected without the need of unpleasant stimulation of the airways as occurs in sputum induction or lavage sampling. Thus, the CO₂ and TVOC levels, temperature and humidity levels and breathing pattern has been obtained using the sensors and this method is non-invasive user friendly, cost effective and it doesn't need any kind of professional help for usage.

REFERENCES

- [1] Abd Alghani Khamis , Aida Idris, etal , Development and Performance Evaluation of an IoT Integrated Breath Analyzer, International Journal of Environmental Research and Public Health, Volume 20, 2023
- [2] Anna Paleczek, Artur Rydosz, Review of the algorithms used in exhaled breath analysis for the detection of diabetes, Journal of Breath Research, Volume 16 number 2, 2022.
- [3] Bartolomé R. Celli, Jadwiga A. Wedzicha, Update on Clinical Aspects of Chronic Obstructive Pulmonary Disease, the new England journal of medicine, 2019.
- [4] Bhagaban Behera, Rathin Joshi , Anil Vishnu G K , Sanjay Bhalerao , Hardik J Pandya, Electronic nose: a non-invasive technology for breath analysis of diabetes and lung cancer patients, Journal of Breath Research, Volume 13 Number 2 , 2019 .
- [5] Catharina C. Moor, Judith C. Oppenheimer, etal , Exhaled breath analysis by use of eNose technology: a novel diagnostic tool for interstitial lung disease, European Respiratory Journal, 2021.
- [6] Jan Krajsa, Ľubomír Straka, etal , Confectionary containing alcohol and their effect on breath analyzer results: a preliminary study, Journal of Slovak republic, 2022.
- [7] Jyoti Kanwar Shekhawat, Mithu Banerjee, Role of Breath Biopsy in COVID-19, The Journal of applied Laboratory Medicine, Volume 7 issues 5, 2022.

[8] Ileana Andreea Ratiu, Tomasz Ligor, Volatile Organic Compounds in Exhaled Breath as Fingerprints of Lung Cancer, Asthma and COPD, Journal of Clinical Medicine, Volume 10, 2020.

[9] Patrik Španěl, David Smith, Quantification of volatile metabolites in exhaled breath by selected ion flow tube mass spectrometry, SIFT-MS, Volume 16, 2020, Page number 18-24.

[10] Peter J. Mazzone, Analysis of Volatile Organic Compounds in the Exhaled Breath for the Diagnosis of Lung Cancer, Journal of Thoracic Oncology, Volume 3 Issue 7, 2015, Pages 774-780

[11] Rosa Alba, Sola Martínez, et al, Data preprocessing workflow for exhaled breath analysis by GC/MS using open sources, Scientific Reports, 2020.

[12] Sagnik Das and Mrinal Pal, Non-Invasive Monitoring of Human Health by Exhaled Breath Analysis: A Comprehensive Review, Journal of The Electrochemical Society, Volume 167 Number 3, 2020.

[13] Sean P. Duffy, Gerard J. Criner, Chronic Obstructive Pulmonary disease Evaluation and Management, Medical Clinics, Volume 103 issue 3, 2019.

[14] Ting Chen, Tiannan Liu, Ting Li, et al, Exhaled breath analysis in disease detection, Elsevier, Volume 515, 2021, Pages 61-72.

[15] Wadah Ibrahim, Rebecca L. Cordell, Michael J. Wilde, Diagnosis of COVID-19 by exhaled breath analysis using gas chromatography–mass spectrometry, European Respiratory Society, Volume 7 Number 3, 2021.

APPENDIX

```
from google.colab import files

uploaded = files.upload()

for fn in uploaded.keys():
    print('User uploaded file "{name}" with length {length} bytes'.format(
        name=fn, length=len(uploaded[fn])))

import seaborn as sns
import pandas as pd

Kurtos = pd.read_csv('/content/Microphone.csv')
df = pd.DataFrame(Kurtos)

K1=(df[['Kurtosis','Skewness','Standard deviation','Variance','Median','out']])
print(K1)

%matplotlib inline
import seaborn as sns; sns.set()
sns.pairplot(K1,hue='out', height=1);

x=K1.drop(columns=['out'])
y=K1['out']

import sklearn

from sklearn.model_selection import cross_val_score

#Import Module
from sklearn.model_selection import train_test_split
```



```

train_X, test_X, train_y, test_y = train_test_split(x, y,
                                                    train_size=0.7,
                                                    test_size=0.3,
                                                    random_state=1)

```

```

print("Labels for training and testing data")
print(train_y)
print(test_y)

```

USING NAIVE BAYES CLASSIFIER

```

from sklearn.naive_bayes import GaussianNB # 1. choose model class
model = GaussianNB()                      # 2. instantiate model
model.fit(train_X, train_y)                # 3. fit model to data
y_model = model.predict(test_X)

```

```

from sklearn.metrics import accuracy_score
accuracy_score(test_y, y_model)

```

```

custom_input = pd.DataFrame({'Kurtosis': [-1.7], 'Skewness': [-0.02], 'Standard deviation': [84], 'Variance': [1300],
                             'Median': [5] })
predicted_output = model.predict(custom_input)

```

```

predicted_output

```

```

from sklearn.metrics import confusion_matrix
cm_nb = confusion_matrix(test_y, y_model)
cm_nb

```

Using Random Forest Classifier

```

from sklearn.ensemble import RandomForestClassifier

```

```

model = RandomForestClassifier(n_estimators=19000)

# fit the model to the training data
model.fit(train_X, train_y)

# make predictions on the test data
y_predict = model.predict(test_X)

accuracy_score(test_y, y_predict)

custom_input = pd.DataFrame({'Kurtosis': [-1.7], 'Skewness': [-0.02], 'Standard deviation': [84], 'Variance': [1300],
                             'Median': [5] })
predicted_output = model.predict(custom_input)

predicted_output

```