Towards Self-Tracking Personal Pollution Exposure using Wearables

BY

Nina Sakhnini
B.S., Jordan University of Science and Technology, June 2016

THESIS

Submitted as partial fulfillment of the requirements
for the degree of Master of Science in Computer Science
in the Graduate College of the
University of Illinois at Chicago, 2019

Chicago, Illinois

Defense Committee:

Debaleena Chattopadhyay, Chair and Advisor
Ugo Buy
Tanya Berger-Wolf
Rachael-Mary Jones, School of Public Health
Oliver Amft, Friedrich Alexander University Erlangen-Nürnberg

To my parents Samar Nsour and Muhammad Sakhnini whom with their endless love and support I could prosper and accomplish ..

To Saleh for his infinite patience ..

To my friends who went through hell with me and yet always managed to lift me up with doses of happiness hormones :D

Nina

ACKNOWLEDGEMENTS

First of all, I want to acknowledge my advisor Professor Debaleena Chattopadhyay of the UIC Computer Science department. Professor Chattopadhyay always answered questions and saved the day when I ran into trouble in my research. She consistently allowed this thesis to be my work but steered me in the right direction when I drifted off track.

I would love to thank Professor Rachael M. Jones of the UIC School of Public Health for her time and support and for making parts of this thesis work come true by giving me access to instruments I needed. I would also like to thank the experts who were involved in this research by providing generous guidance and advising: Professor Oliver Amft and Asta Roseway. I also would love to thank the rest of my committee: Professor Ugo Buy and Professor Tanya Berger-Wolf.

Moreover, I want to thank professor Isabel Cruz for helping me make connections between my thesis project and semantics. Also, I would love to thank my colleague Ja Eun Yu as my work and her work met at some points and we collaborated for aceiving some parts. Finally, I want to thank my lab mates at the UIC HCI lab for bearing with my noisy experiments.

NS

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

CHAPTER **PAGE**

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AQI Air Quality Index. vii

AWS Amazon Web Services. vii

dB Decibel. vii

dBA A-weighted Decibel. vii

DC Direct Current. vii

FFT Fast Fourier Transform. vii

HCI Human Computer Interaction. vii

IoT Internet of Things. vii

MCI Mild Cognitive Impairment. vii

NIST National Institute of Standards and Technology. vii

NRDC Natural Resources Defense Council. vii

PM Particulate Matter. vii

PM2.5 Particulate Matter 2.5 um/g$^3.vii$

SoC System on Chip. vii

SPL Sound Pressure Level. vii

SVM Support Vector Machine. vii

SWE Sensor Web Enablement. vii

SUMMARY

Recent epidemiological studies have shown that long-term exposure to air pollution is positively associated with mild cognitive impairment (MCI). Although interest in pollution monitoring is proliferating, self-tracking personal pollution exposure is little explored. In this thesis, I adopt a human-centered computing approach to explore the design space of personal pollution tracking wearables. This work makes three contributions to human-computer interaction: 1) design guidelines for rapid-prototyping low-cost, sub-optimal personal pollution tracking wearables and a physical prototype that measures $PM_{2.5}$ and ambient noise which are the pollutants that epidemiological studies have demonstrated their association to MCI, 2) exploration of different calibration techniques to improve the accuracy of low-cost $PM_{2.5}$ sensors, and 3) a characterization of how human interference, our day-to-day activities, significantly affect the operation of personal pollution tracking wearables. In sum, this thesis informs design guidelines about how to physically prototype personal pollution tracking wearables and where to wear them—beyond citizen-science efforts of data collection—rather toward monitoring personal long-term pollution exposures to mitigate the environmental risk factors for many illnesses such as early dementia.

CHAPTER 1

INTRODUCTION

Recent epidemiological studies have shown that prolonged exposure to some air pollutants is positively associated with mild cognitive impairment (MCI). In spite of the fact that Ubiquitous computing is a common-place now and in spite of the increase in the interest in pollution monitoring, personal pollution exposure is little explored. This thesis resembles an effort towards a more user-friendly ubiquitous personal pollution exposure using wearable technology.

In this work, we created a low-cost pollution exposure measurement wearable for older adults in particular, and for the public in general. The wearable measures the twp pollutants that are proven to be related to mild cognitive impairment in older adults: $PM_{2.5}$ and noise. This study informs a basic calibration algorithm to improve the accuracy of low-cost $PM_{2.5}$ sensors. Using the wearable, we also studied the effects of different human interference factors on $PM_{2.5}$ due to day-to-day activities and the proximity to the human body. The results inform design guidelines for where to wear the devices and open the door for creating personal pollution monitoring wearables that make an account for the humanly generated particulate matter as opposed to the industrial particulate matter. The wearable allows users to monitor their personal environmental exposures, be aware of potential pollution exposures when planning day-to-day activities, and take day-to-day actions to combat the environmental risk factors for the early onset of dementia.

1 .1 Thesis Overview

In Chapter 1 we introduce the thesis work. We also talk about the motivation behind this work and we give a short background study on the aspects of the thesis work. Finally, we discuss the theoretical framework of this thesis work. Chapter 2 is describing and discussing the personal pollution exposure monitoring prototype we built. The prototype technical components, as well as human-centered decisions, were discussed extensively in the chapter. Chapter 3 explores the calibration of the prototype using a high-cost high-precision $PM_{2.5}$. In Chapter 4, human interference with low-cost personal pollution exposure monitoring tools is explored. Chapter 5 concludes the thesis.

1 .2 Motivation

Recent epidemiological studies marked air pollution and noise as the two pollutants associated with Mild Cognitive Impairment (MCI) in older adults.[4–6] MCI is an intermediate stage between normal age-related cognitive decline and early dementia, e.g., amnestic type of MCI (aMCI) is an early onset of Alzheimer disease, and non-amnestic MCI (naMCI) is the early onset of vascular and other forms of dementia.[7,8] The cognitive decline is adversely affected by the prolonged exposure to $PM_{2.5}$ and noise. The biological reasons for these associations are not fully known yet. By 2020, it is expected that 42.7-48.1 million people will be suffering from Dementia.[9] With the increase of the number of people having Dementia, the urgency for early diagnosis of mild cognitive impairment (MCI) and for monitoring the environmental risk factors (ERF) for MCI in real-time is expanding. Early diagnosis of MCI and for monitoring the ERF for MCI in real-time play a critical role in controlling and delaying the onset of Dementia.

Epidemiology studies investigate the health effect of pollution at a population level, using data from few fixed monitoring stations such as EPA monitoring stations[10] and the Array of Things in Chicago[11] and assuming static population distributions.[12]

There are several challenges when it comes to personal pollution monitoring exposure. First, low-cost wearable environmental sensors are suboptimal and pose high sensitivity towards the human body, clothing, and, everyday activity.[13] Second, pollution concentrations measurements from fixed pollution monitoring sites lack spatial and temporal resolution in spite of being equipped with high-cost gold standard sensors. Finally, although outdoor pollution monitoring is becoming more of a commonplace, and smart home bundles includes pollution monitoring devices. Indoor air quality monitoring is becoming prevalent, ways to manage long-term environmental exposures and adopt healthy lifestyle changes are mainly limited to large-size expensive equipment and wearables intended for the use of scientists and quantified-selfers. These tools are not designed for the use of the intended at-risk population, who need and benefit from these tools the most.

Monitoring exposure to $PM_{2.5}$ and noise pollution and using that data to combat the risk of pollution exposure is promising for the at-risk population and their caregivers, but, it forms challenges.[14–16] With that in mind, the goal of this thesis work is to study the effect of human interference on low-cost $PM_{2.5}$ sensors in order to inform the design and development a personal pollution monitoring wearable that can estimate $PM_{2.5}$ concentrations and ambient noise to help people, particularly the elderly, to monitor personal pollution exposures ($PM_{2.5}$ and noise). This will give the elderly the opportunity to benefit from the latest ubiquitous computing technologies to monitor their exposures to pollutants in order to combat MCI risk factors. This wearable will help the elderly to adjust their day-to-day schedule and activities in order to avoid the risks of long-term exposure to these pollutants and therefore keep their cognitive functions healthy by lowering the chance of suffering of MCI.

1 .3 Background

Particulate Matter (PM) is the body of particles, solid and liquid, suspended in the atmosphere. These particles differ in shape, size, and source. PM is classified into three size classes: coarse with a diameter ranging from $10.0\,\mu m$ to $2.5\,\mu m$, fine with a diameter ranging from $2.5\,\mu m$ to $0.1\,\mu m$, and, ultra-fine which is smaller than $0.1\,\mu m$. Based on sizes, PM is defined as one of three types of PM; PM_{10}, $PM_{2.5}$, and, $PM_{0.1}$. Different size of particulate matter are produced from different sources and have varying chemical components. PM_{10} is mainly coal dust and fly ash. $PM_{2.5}$ are produced from smoking tobacco and some metal fumes. Diesel engines produce $PM_{1.0}$. Other manufacturing and industrial procedures produce PM of different sizes as well. Figure 1 shows the classifications of particulate matter.

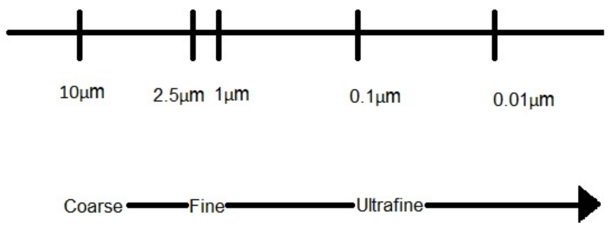

Figure 1: Classification of Particulate Matter based on size and source. (From[3])

1 .3.1 Humans and Particulate Matter

Environmental studies have identified sources of PM in different microenvironments, such as indoor and outdoor. The outdoor PM has two general sources. First is natural sources such as natural chemical interactions, for example, decomposition of organic compounds, and volcanic eruptions. The other contributor to outdoor PM is anthropogenic emissions such as traffic emissions, industrial emissions, and emissions from power plants. Anthropogenic emissions are responsible for a majority of the harmful outdoor PM as the natural sources contribute to the natural environmental cycle to maintain it at equilibrium.[17,18]

Indoor PM is composed, generally, of two main sources. First, indoor generated PM such as PM generated by activities such as cleaning[19,20], frying and cooking fumes[21,22], fireplaces[23], tobacco smoke[24,25], candles[26], spraying products[26], results of chemical interactions of organic components indoors due to heat or other triggers, and, -one important source- PM generated from human body and clothing, bioaerosol emissions of breath[27], and, physical activity.

The PM emitted from the human body and clothing contributes to the human's personal cloud.[28] The personal cloud is defined as the difference between the PM measurement of a personal pollution exposure monitoring tool, that is near the human body or attached to it, and the PM value read by a standard area-level or population-level pollution monitoring station at a certain geo-spatial point at a certain point or period of time.[29]

A study conducted in a sealed chamber to asses the emissions of the clothed human body on the personal cloud has found that human activity increases the PM emission rates. Also, the study found that among different textiles, cotton has the highest PM emission rate.[30] The second source of indoor PM is the outdoor particulate matter that reaches the indoor microenvironment by several means, such as opening windows and doors and penetrating through building cracks, or, outdoor PM travels indoor by human activities. The indoor-generated PM is the most significant contributor to indoor PM.[29,31] People spend a vast majority of their time indoors, thus, they are mostly exposed to indoor PM.[32]

Environmental scientist are exerting efforts to identify and classify sources of PM based on toxicity and the adverse health risk of exposure.[33] The chemical composition of PM increases toxicity by reacting with the atoms of the human body. The presence of PM in certain size and shape pose a threat of toxicity as well. Different sources of PM produce mixtures of particles with different chemical composition as well as different physical features. It is believed that the whole mixture plays a role in the exposure's health effect.[33] Studies have shown that PM from traffic emissions is more harmful than coal-fired power plants emissions and other secondary carbon sources.[34,35]

1 .3.2 Personal Pollution Exposure

Personal pollution exposure, from a quantitative point of view, is defined as the total number of pollution particles measured by a personal pollution monitoring tool within a person's breathing area. Personal pollution exposure should not be impacted by the PM in the person's exhaled breath.[29] In other words, personal pollution exposure is the amount of PM inhaled at a given instant or a period of time in a certain location or a set of locations trailing the person's movement. Personal pollution exposure assessment is a process in which magnitude, frequency, and duration of exposure to a pollutant or a set of pollutants is measured.[36]

Assessing personal pollution exposure is becoming more and more important since it is a crucial health indicator, especially in urban and industrial areas where pollution is a serious concern. Facing complications from human interaction with the surrounding environment is inevitable when assessing personal pollution exposure, therefore, it is challenging. Wearable, portable, environmental sensors are a compromise between cost and optimality.[13] These sensors are usually sensitive to interference by everyday activities of the user, such as human skin emissions or textile emissions.[13]

Ubiquitous computing has radically changed the way people track their activities and behaviors.[37] People are gaining access to data about their own activities, behaviors, feelings, and health indicators in orders of magnitude richer than any previously available. This is because of the leap in sensing technology, device miniaturization, and data analysis approaches. Market analysis forecasts a dramatic increase in the smart wearable market share; it is expected to double from 2018 to 2022.[38]

In spite of the fact that epidemiology gives vigorous pieces of evidence that long-term pollution exposure has adverse health, ubiquitous computing has limited progresses towards robust personal pollution exposure tracking systems that can be used openly by the public. Complex interactions between the human body, clothing, and, belongings and the surrounding environment makes assessing personal pollution exposure perplexing.[12,15] That is a reason behind the gap between self-monitoring tools for pollution exposure and these tools for monitoring other health indicators.[39]

Indoor air quality (IAQ) monitoring is one of the technologies available for the public for monitoring pollution exposure in indoor microenvironments. IAQs are increasingly becoming prevalent (e.g., Dylos DC1100, Foobot, uHoo,[40]). These tools provide household-level pollutant data, such as volatile organic compounds (e.g., CO, NO2, O3) and $PM_{2.5}$. Despite that IAQ monitors are portable, they are designed to sit at a fixed indoor location (e.g., placed in the living room[40]) and thus, do not offer, at a the fundamental level, monitoring of personal pollution exposure across different situations of human exposure. No personal pollution exposure monitoring wearable that monitors exposures in real-time and everywhere, is currently available for the public on the market.

Other current solutions for personal pollution exposure monitoring are designed for the use of scientists and quantified-selfers.[39] Moreover, tracking personal pollution exposure in real-time and keeping traces of time-activity-exposure patterns offers a considerable promise for people, particularly, the at-risk community such as older adults, people with asthma, and people living or working in areas marked as polluted.[14–16] Thus, this thesis is an effort towards setting design guidelines for self-tracking personal pollution exposure, in order to reduce the gap separating personal pollution exposure monitoring and other personal activity and behavioural monitoring and to give the at-risk community a tool to monitor their personal pollution exposure as a step to combat environmental risk factors for health problems.

Traditional approaches to track pollution exposure that build upon pollution concentration values measured by fixed pollution monitoring stations lack spatial and temporal resolution. For instance, there are four fixed governmental pollution monitoring sites in Chicago, with a 1-in-few days pollution concentrations reading schedule.[12] Currently, with the availability of affordable pollution monitoring sensors and low-cost computing, tracking personal pollution exposure with higher spatial and temporal resolution is becoming more prevalent. Table I shows a comparison between recent works on personal pollution exposure monitoring tools and how do they compare to my thesis project.

Pollution monitoring system	Pollutant	Form factor	Settings	Calibration study	Managing Human Interference	Personal exposure monitoring
Budde et al.[41]	$PM_{2.5}$, PM10	Portable	O			
AirSense[42]	$PM_{2.5}$	Portable	I, O			✓
inAir[40]	$PM_{2.5}$	Portable	I			
MyPart[43]	PM10	Wearable	I, O	✓		✓
MAQS[44]	CO2	Portable	I			✓
CitiSense[45]	CO, NO2, O3	Portable	I, O			
Common Sense[46]	CO, NOx, O3	Handheld	O			
Oletic et al.[47]	CO, NO2, SO2	Handheld	O	✓		
Piedrahita et al.[48]	CO, CO2, NO2, O3	Portable	I, O	✓		✓
W-Air[13]	CO2, O3	Wearable	I, O	✓	✓(partially)	✓
Ambiciti[49]	Noise	Handheld	I, O			✓
proposed project	$PM_{2.5}$, Noise	Handheld	I, O	✓	✓	✓

TABLE I: Comparison of portable air and/or noise pollution sensing devices. Personal exposure monitoring refers to individuals, not a room or enclosed space (I = indoor, O = outdoor).

In this work, we measure two types of pollutants: $PM_{2.5}$ and noise. But, we focus our studies on $PM_{2.5}$. Noise is a commonplace for scientists in computing as well as other disciplines related to acoustics. Noise and microphones have been extensively studied; microphone calibration have been done for stand alone microphones as well as smartphone microphones.[50,51] Microphones are used for human activity recognition[52,53], incorporated in smart home applications[54], development of acoustic localization technology that is used to locate the source of a sound in 3 dimensions both in air and underwater[55–58], and in passive Human-Robot interaction[59]. Thus, we focus on exploring human interaction with low-cost $PM_{2.5}$ sensors.

CHAPTER 2

PERSONAL POLLUTION EXPOSURE MONITORING PROTOTYPE

Ubiquitous computing is becoming more and more prevalent and wearable technologies are becoming a part of people's everyday life. People use wearables to monitor their daily activities and feelings. Monitoring activities and habits helps in understanding these behaviours and, as a result, adjust them for a healthier life style.

2 .1 Motivation

Recent epidemiological studies have shown that mild cognitive impairment (MCI) is associated positively with the long-term exposure of older adults to $PM_{2.5}$ and noise. MCI is a precursor to a disease that has no cure yet: Dementia.

Available solutions to monitor environmental pollution as a health risk factor do not target the at-risk. Instead, these solutions are either designed for the use of scientists and quantified-selfers, or, they are designed to stay at a fixed location. Thus, comes the decision to design a wearable for personal pollution exposure.

We created a wearable prototype to measure personal pollution exposure. The prototype measures pollutants associated with mild cognitive impairment; $PM_{2.5}$ and noise. The prototype is a wearable worn most comfortably as a handheld and is used throughout this study to inform design decisions for iterating towards a human-centered wearable device. The wearable prototype works side by side with a smartphone running Android. The prototype is shown in figure 2.

Figure 2: The wearable

2 .2 The System

Several studies have shown a positive association between household income and pollution exposure. A recent study of the Natural Resources Defense Council (NDRC) has shown a map of Chicago that associates income with pollution levels. The map is built using data from EPA. The pollution levels were shown to be higher in the south and west sides of Chicago, the areas where there is a lower socioeconomic status.[60] With that in mind, the decision to create a low-cost personal pollution exposure tool came in order to give the population with lower socioeconomic status to have at hand a personal pollution exposure tool. Moreover, market statistics show that the primary reason people turn away from a wearable gadget is the cost.[61] With having access to such a tool, health risk factors, in general, and the MCI risk factor, particularly, will be manageable. Having cost efficiency in mind as one of the main goals, we built the prototype using off-the-shelf low-cost components. As an estimation, the cost for building the prototype is 80 USD.

2 .2.1 Mechanism

The wearable reads current local $PM_{2.5}$ values using a low-cost $PM_{2.5}$ sensor. If WiFi is available, the wearable sends the timestampped data over WiFi to a cloud-based database dedicated for this project hosted on Amazon Web Services (AWS)[62] using the DynamoDB engine[63] which is a NoSQL online database service.

If WiFi is not available, the wearable connects to a smartphone using Bluetooth. The smartphone, running a dedicated application, sends the data to the AWS database using GSM Network or any available internet connection. If neither WiFi nor Bluetooth is available, the wearable saves the data locally until a connection is available. Later, when a connection is available, the online database is synchronized with the local database.

The smartphone application collects the current noise level using its built-in microphone. Also, the smartphone application captures contextual data in order to provide an understanding of the context of which the pollution exposure instant happened. Next, the data is processed on the phone to display an estimation of the user's personal pollution exposure.

2 .2.2 Hardware

Since cost efficiency is one of the goals of this project, as an effort to enable people with lower socioeconomic status to monitor pollution since it is a health concern, we built the wearable using off-the-shelf low-cost components.

System on Chip. The computing power of the wearable comes from a Raspberry Pi 3b+ shown in figure 3 connected to the sensor near a smartphone.

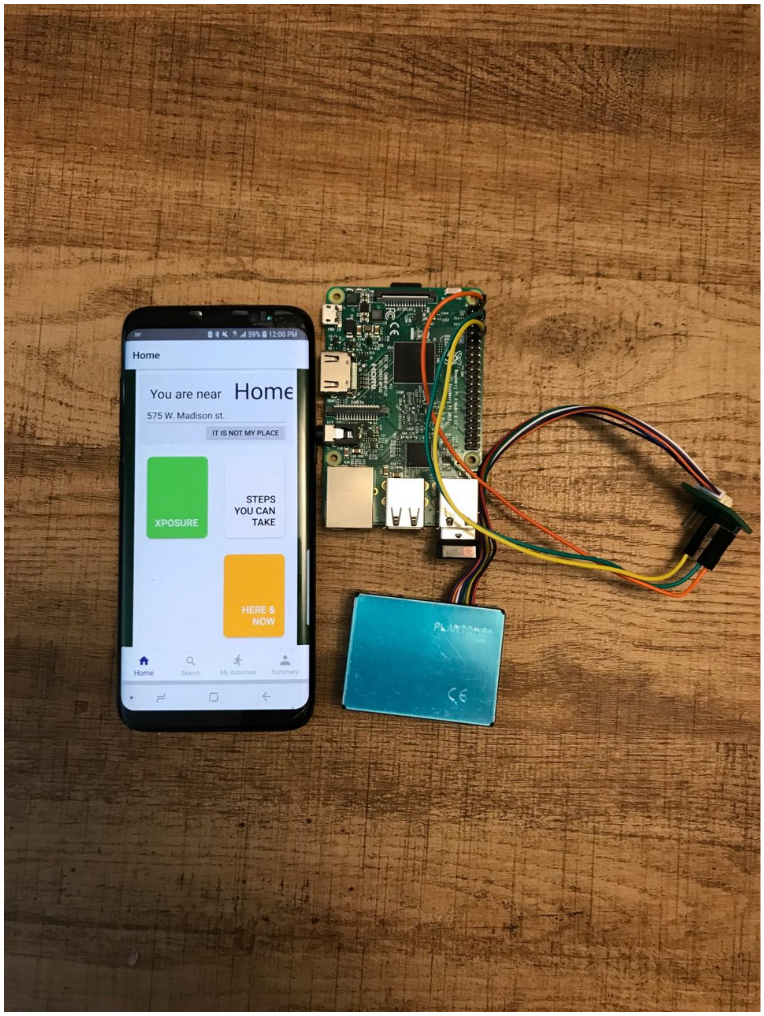

Figure 3: Raspberry Pi 3b+

Raspberry Pi is a low-cost small-sized hackable System on Chip (SoC) with adequate computing power. It was originated as an education-friendly computer board in 2012. Raspberry Pi is widely used in the Internet of Things (IoT) applications.[64] Raspberry Pi runs an operating system of choice by installing it on a microSD card. The microSD card operates as the hard disk of the system. The Raspberry Pi 3b+ uses the microSD external memory to run and save the data. There are two famous operating systems modified for Raspberry Pi. Windows 10 Internet of Things (IoT) is a lightweight version of Windows 10 developed for board computers.[65] The other operating system is Rasbian. Rasbian is a light-weight version of Linux Debian created for Raspberry Pi. In this project, Raspberry Pi ran Rasbian installed on a 16 GB microSD card.

Raspberry Pi then acts like a normal computer and requires input devices to be connected using USB or any of the available ports e.g. keyboard. The board of a Raspberry Pi is similar to a card-sized motherboard with fundamental system components such as computing and graphics processing units and memory. In addition to that, the board has a set of components to communicate with different types of peripherals for input and output. Raspberry Pi 3b+ is powered by a Broadcom BCM2837B0, Cortex-A53 64-bit SoC @ 1.4 GHz processor. The processor can be overclocked for more computing power.[66] But, within the scope of my application, there is no need for overclocking.

The Raspberry Pi 3b+ was chosen because it offers built communication facilities. The Pi allows for WiFi and Bluetooth communications with built-in WiFi and Bluetooth modules. This eliminates the need to attach a WiFi or a Bluetooth module to the board. As a result, the system has a minimal thickness. This is important because minimizing thickness is one important aspect of wearables design because it adds safety and comfort both physically and perceptually.[67]

The sensor. The Pi is interfaced using its serial UART to a low-cost $PM_{2.5}$ sensor Plantower PMS7003.[68] At an early stage, we interfaced the Pi to a low-cost $PM_{2.5}$ sensor Plantower PMS5003. In a later stage, we upgraded the sensor to Plantower PMS7003. The latter improves upon the last by that PMS7003 is approximately half of the size of the PMS5003 which makes the sensor more portable and more adequate to be a part of a wearable as it does not contribute to the thickness of the system.

The sensor reads 3 size categories of particulate matter (PM). These are the particulate matter with a diameter less than or equal to $1.0\,\mu\text{m}$. The smallest diameter for particulate matter which the sensor is capable of measuring is the particulate matter with diameter $0.3\,\mu\text{m}$ as per the manufacturer's evaluations.[68] The second category of which the sensor measures is the particulate matter with a diameter larger than $1.0\,\mu\text{m}$ and less than or equal to $2.5\,\mu\text{m}$. The last category is the particulate matter with a diameter up to $10.0\,\mu\text{m}$.

The sensor uses a laser scattering principle to measure PM. Laser scattering is a method for characterizing micro-particles. A laser beam is released across a chamber where atmospheric air can enter. Particles then radiate the laser beam causing it to scatter to a certain degree based on its size. The degrees of scattering and the number of scattering events happening at a point of time is then transmitted to the Pi as numbers of $PM_{1.0}$, $PM_{2.5}$, and, PM_{10}.

Powering. The wearable is powered using a 4000 mAh lithium-ion battery as a rechargeable power bank.[69] The power bank is approximately the size of the Raspberry Pi board. The power bank takes approximately 4 hours to charge completely and runs the Pi for an approximation of 6 hours per charge. The device has boost converters that provide 5 V(DC) up to 1A via a USB A port. The 1A output is key for powering Raspberry Pi to avoid under-powering. Under-power is when the device is getting a current smaller than what its default. Long intervals of under-power might cause damage to the device's power supply. When a Raspberry Pi is underpowered it shows a small lightning strike shape on the top-right corner of the screen.

2 .2.3 Software

The Raspberry Pi used to build the wearable runs Rasbian. The latest version of Rasbian was installed on the microSD card and all the local data and codes are stored there. The codes running the wearable were written in Python. The codes use threading because it gives it robustness. When the Pi is connected to the power source, an interfacing script runs initiating two threads. The first thread handles communications as follows:

- The thread checks if WiFi is available
- If WiFi is available, the thread reads the earliest entry in the database that was not written to the database and it writes it to the database. Then, it marks it as read and it checks again for WiFi availability and loops as long as WiFi is available.
- If WiFi is not available, the thread tries connecting to an available smartphone running the application. If the connection is successful, the thread reads the earliest entry in the database that was not written to the database and it writes it to the database. Then, it marks it as read and it checks again for WiFi availability, then Bluetooth connection availability. The thread loops on that.

- If neither of the connections is available, the threads keeps checking for connections every 5 seconds.

The second thread handles reading data from the sensor and writing it to the local database. Having this in a separate thread ensures that no data is missed and reading data is not affected by the availability of a connection.

The computing capabilities of the Pi handles the two threads without any interruptions or slowing down. Also, it does not heat up the Processor of the Pi. As a result, the pi does not require a heat sink and runs without excessive or disturbing heating in room temperature as well as hot temperature as we have tested by running pilots in summer weather as well as winter weather. This holds as long as the wearable is not wrapped in an isolating and non-breathable packaging. This protects the device and gives it longevity. Also, increases the wearability by increasing comfort as it remains at an adequate temperature.

2 .3 Smartphone Application

As a part of the system, we created a smartphone application for Android devices. The application works hand-in-hand with the wearable to obtain an inclusive personal pollution exposure estimation. The other pollutant that is associated with MCI that the wearable system measures is the noise. Noise is obtained by sampling sound levels periodically. The noise is measured by the smartphone's application using the smartphone's microphone. The application is developed in Java using Android Studio.

The application has a simple interface showing current $PM_{2.5}$ values (If connected to wearable) and noise levels. Part of the application was built in collaboration with my colleague Ja Eun Yu as a part of the myCityMeter project.[70]

2 .3.1 Contextual Data Collection

Contextual resolution is an important factor when trying to understand personal pollution exposure. Time and location, for example, are critical in order to make sense of the pollution data for a better understanding of an exposure event. Weather also affects the pollution levels, for example, wind changes the concentration of PM. Thus, the collection of contextual data alongside pollution data was essential to create a full understanding of personal pollution exposure. Contextual data was collected using the smartphone's built-in sensors and data from weather stations.

The smartphone application timestamps the data and geotags it using fine GPS data. The smartphone's WiFi and GSM signal levels and the screen's brightness are collected from the smartphone's system. The built-in light sensor collects the luminosity value. Using Google and Yahoo! weather APIs, weather information such as outdoor temperature, humidity, dew point, weather condition, and wind speed. We also collected an estimation of ambient temperature using battery temperature. Studies could model a formula for estimating ambient temperature based on the smartphone's battery temperature.[71] The formula is as follows:

$$t_{ambient} = 2.085 + 0.874t_{battery} - 0.0004v$$

Where $t_{ambient}$ is the estimated ambient temperature in degree Celsius at a certain time instant, $t_{battery}$ is the smartphone's battery temperature in degree Celsius at the time instant, and v is the battery's voltage at that same time instant. It is worth mentioning that the ambient temperature collected using this method is not always the actual instant temperature since might be affected by a lot of factors that might isolate the device from the surroundings or affect the temperature of the device.

Finally, we needed to manually tag the microenvironment as one of four, outdoor, semi-outdoor, indoor, deep indoor, and the setting, for example, cooking, or in the library, of which the current exposure is happening in for calibration purposes. To do that, we modified the application to manually type a setting and choose a microenvironment. The edited application interface is shown in the figure 4.

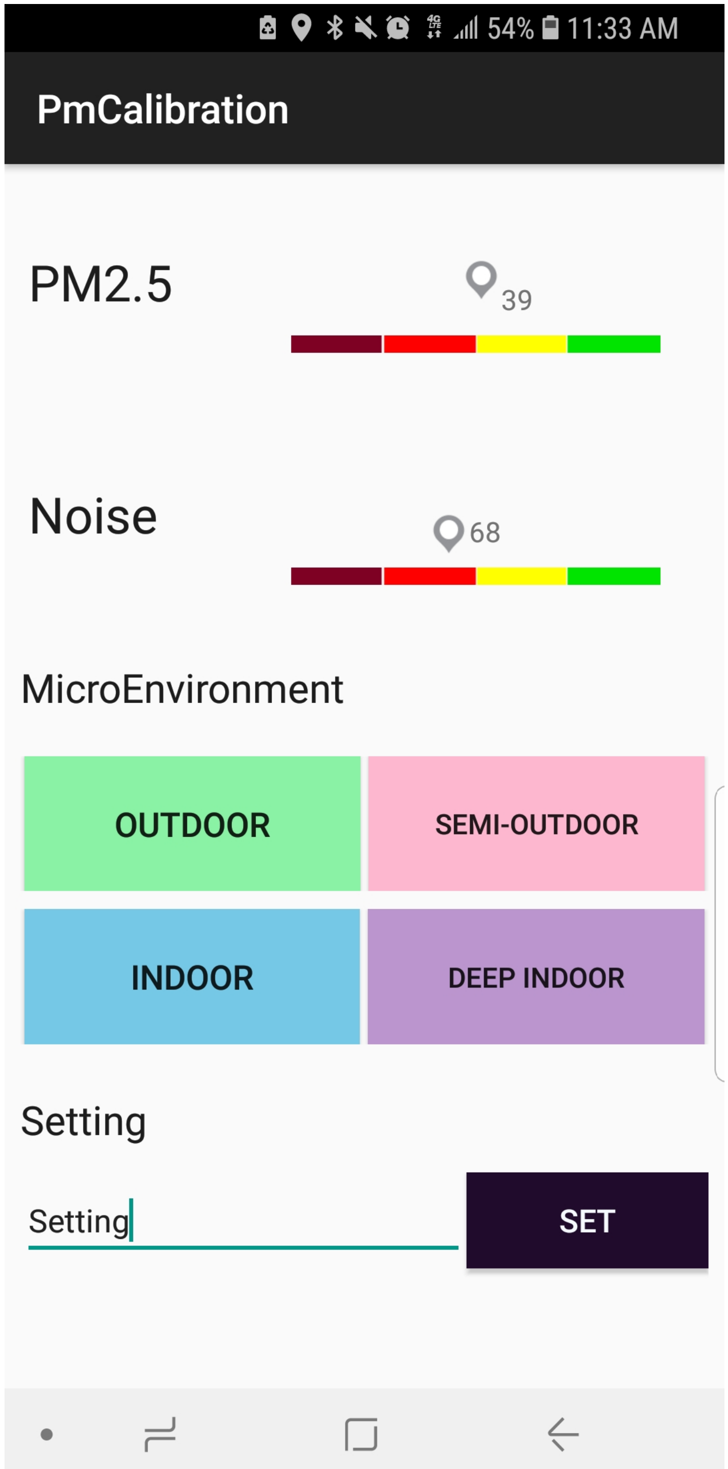

Figure 4: Application’s Interface for Contextual Data

2 .4 Ontology

Personal pollution monitoring systems are built with a set of sensors that provide sensor data streams for pollution as well as context-related data. This data is and will be produced in high volumes given that the data streams are being produced by each user for a set of sensors over lengthy time periods. This torrent of data will be problematic if it was not managed systematically and formally. Gigabytes of raw sensor data are meaningless if this data was not semantically organized. The organization will allow the data to be leveraged well. It will ease the access and operation on this data as well as communicating and visualizing the data as needed. Personal pollution data convey valuable knowledge for users, thus, it must be handled carefully.

Given that an ontology is a formal way to describe the concepts and relationships that can exist for an agent or a community of agents[72], an ontology is an approach to formally and semantically represent the personal pollution exposure of a person. Using semantic web allows for the formality and systematizing needed for the previously mentioned data. The work towards pollution in the Semantic Web is oriented towards sensors and sensor readings.

Sensor Web Enablement (SWE)[73] is a suite of specifications related to sensors by Open Geospatial Consortium. Sensor Web Enablement enables sensors to be accessible and controllable by the Web. Raw sensors data is by nature opaque, so, metadata is critical in managing sensor data. According to SWE, a semantically rich sensor network provides three types of metadata: Spatial metadata, Temporal metadata, and, Thematic metadata.

Every sensing system has unique domain-specific information. The Sensor Web Enablement framework supports simple spatial and temporal concepts.[73]

Semantic Sensor Web (SSW)[74] is a framework that works towards giving situational awareness for sensor observations by improving upon standard sensor languages of the SWE. SSW gives sensor data more contextual meaning to help the user understand the situation of which the sensor was operating in. SSW connects SWE XML-based metadata standards to the Semantic Web.

Several ongoing initiatives to build relevant ontologies

- Sensor Standards Harmonization by the National Institute of Standards and Technology (NIST)
- W3C Geospatial Incubator Group Ontology
- OGC's Geographic Markup Language Ontology
- OWLTime for temporal-based Ontologies[75]
- W3C Semantic Sensor Network[76]
- Domain-specific ontologies provide semantic descriptions of thematic entities: Indoor Environment Quality Ontology[77]

Semantic Sensor Network[76] is an OWL 2 ontology to describe sensors and observations. Semantic Sensor Network Describes capabilities, measurement processes, observations and deployments of sensors. The ontology offers an inclusive view of a sensor. The ontology can describe: Sensors, Accuracy of sensors, Observations, Deployment, Methods of sensing, Concepts for operating, and Ranges. Semantic Sensor Network introduces Stimulus-Sensor-Observation pattern. The Semantic Sensor Network ontology has four perspectives: Observation perspective, Sensor perspective, System perspective, and, Feature and property perspective.

The available semantic works extensively describe sensors and sensor event. But, these works have minimal consideration for sensor context, which is a core component to measuring personal pollution exposure. Current ontologies focus on the sensor rather than the exposure. The ontologies focus on sensors, per single sensor. These ontologies extensively describe sensors, but not the exposure. These ontologies lack accurate contextual resolution. Here comes the need for formalizing the data to provide ease of access by communicating pollution data with a high contextual resolution. To achieve that, we created an ontology to leverage the ability to systematically access user data to improve the application's algorithms. The ontology aims to facilitate blending data from different users and sources about the same location to give more accurate pollution exposure estimation at a certain location or time.

2 .4.1 Personal Pollution Exposure Ontology

An ontology was designed in OWL using Protege 5.5.0 build beta-3. The ontology describes personal pollution exposure events. Core requirement for the ontology is that an exposure instance is expressed in an extensive form. The knowledge for the design of the ontology was acquired by collecting information from already existing sensing ontologies such as Semantic Sensor Network.[76] Also, prior knowledge gained by working on other related projects was employed in the ontology design.

The core class is Exposure_Object has two subclasses Pollutants and Exposure_Context. Pollutants class currently looks at two pollutants: $PM_{2.5}$ measured by μ/m^3 and Noise measured by dB(A) which is a transformation of dB to suit the human ear. Exposure_Context describes the location where an exposure happened. Also, it describes the microenvironment. A microenvironment is the local environment of an event and can be classified into four microenvironments: outdoor, semi-outdoor, indoor, deep-indoor. Also, the context describes the timestamp at which the exposure instant happened as well as the weather at that point since weather plays a critical role in the pollution levels at a certain point.

External ontologies were used. First, the SSN was used since it extensively describes sensors. Second, OWL Time was used to represent temporal events in an exposure object's context. Third, an RDF Vocabulary: WGS84 Geo Positioning was used to describe the exposure context's location.[78] Finally, a weather ontology by the Smart Energy-Aware Systems was used to describe the weather as a part of an exposure context.[79]

2 .4.2 API

In order to make use of the ontology, an API to communicate and query the data is necessary. Thus, we designed and partially prototyped an API for the data collected in this project using Java. The API gets the raw sensor data in a CSV format from AWS's database. The API, then, process the data and produces OWL individuals using OWLAPI[80] and attaches these individuals to the Ontology.

2 .5 Conclusion

In this thesis, we present an effort towards creating low-cost personal pollution exposure wearables. We created a wearable prototype to measure $PM_{2.5}$ and noise as the two pollutants positively associated with early-stage dementia as demonstrated in recent epidemiology studies.

In this chapter we extensively discuss how we created the wearable using off-the-shelf system-on-chip Raspberry Pi 3b+ and a low-cost $PM_{2.5}$ sensor. Also, we explore an Android application we created to inform pollution levels and to measure noise and collect contextual data. Finally, we discuss the creation of an ontology to describe personal pollution exposure in an extensive way supported by contextual variables. The prototype was used throughout the thesis work to collect data and conduct interaction studies with the human body.

CHAPTER 3

LOW-COST $PM_{2.5}$ SENSORS CALIBRATION

Scientists and engineers are developing new sensing and measurement technologies every day. New breakthroughs in sensing technology are breaking size and cost limits for sensing in a way that makes access to these tools within the reach of a vast majority of people; from students to amateurs to researchers and even business professionals. Thus, these sensors are being used in an ample set of applications.

But, these sensors and measurement tools vary in accuracy due to a set of reasons. Some low-cost sensors are built using low-cost materials and technology which leads to the sensor's sub-optimality. Another factor that affects measurement accuracy is that the accuracy is reduced with use in time. Sensor parts may differ in shape or functionality due to being exposed to sunlight, heat, rain, and other natural conditions that might affect the sensor's performance.

Also, some unexpected accidents might affect a sensor's measurement accuracy such as falling to the ground and bumping into objects. The sensor measures faulty data when such a condition applies. Thus, to make use of the sensor, a correction to the data is needed. Calibration is a measurement rectification technology. Calibration is defined as the correction of a measurement of a certain tool by evaluating the tool's measurements against ground truth data gathered by a gold standard measurement tool.

3 .1 Motivation

In this thesis, we are using a low-cost $PM_{2.5}$ sensor as a part of the creation of a low-cost personal pollution monitoring wearable. Since the sensor we are using is a low-cost $PM_{2.5}$ sensor with sub-optimality as the main concern, conducting a calibration study is an important step toward the creation of the wearable. In this chapter, we discuss our calibration study and our results.

3 .2 Tools

For the calibration study, we used the SidePak™AM510 Personal Aerosol Monitor as a reference for ground truth $PM_{2.5}$ concentrations. The calibrated sensor represented by the prototype we developed and discussed in chapter 2 was the sensor to be calibrated.

3 .2.1 SidePak™

The SidePak™AM510 Personal Aerosol Monitor is a laser photometer that measure personal pollution exposure in real-time. SidePak is used as a pollution monitoring tool for measuring pollution concentration in different types of settings and microenvironments.[81–86]

SidePak is a portable device that can be worn on a belt or attached to clothing using a clip fixed on the back of the device. It has a tube that is attached to a sampling pump that facilitates the monitoring of pollution concentrations in the personal cloud. Although it is considered quiet when compared to other pollution monitoring tool, the pump noise is considerable, our noise meter shows a noise average of 70 dBA (That is the noise of a vacuum cleaner at an approximately 10 ft distance[87]) comparing to an average of approximately 40 dBA (Quiet living room noise[87]) when it is turned off.

SidePak requires approximately 3.5 hours to recharge and runs for approximately 7 hours. It has internal memory with an approximate size < 100 MB (The manual[88] and the specifications sheet[89] don't mention the memory size, the estimation is based on our experiments). SidePak is with a moderate weight of approximately a pound and it has dimensions of 4.2 x 3.7 x 2.8 inches (10.6 x 9.2 x 7.0 cm). It is worth mentioning that the SidePak costs a few thousands of US dollars.

In a procedure of laser scattering, when SidePak is on, the pump regularly conveys air samples from the surrounding atmosphere to an optical chamber inside the device passing through a filter that removes all particles with diameters larger than 2.5 μm. There are multiple filters with multiple sizes for measuring the concentration of different sizes of PM. The particles with the intended size or less make their way to the optical chamber where a ray of laser is shot and a set of lenses captures the scattered rays in a process to count the number of particles per size. SidePak has a frequency up to 1 reading per second and can function in pollution concentrations reaching a limit of 20,000 μg/m^3.

The SidePak we used in our experiments is calibrated by the manufacturer (TSI). The calibration report mentions that in the calibration process, the device is adjusted in accord with standard ISO 12103-1, A1 test dust (Arizona Test Dust). This indicates that SidePak when measuring $PM_{2.5}$ concentrations will assume that the physical properties of the measured particulate matter are similar to that of the Arizona Test Dust. SidePak is used as the ground truth for our low-cost $PM_{2.5}$ sensor and it is shown in figure 5.

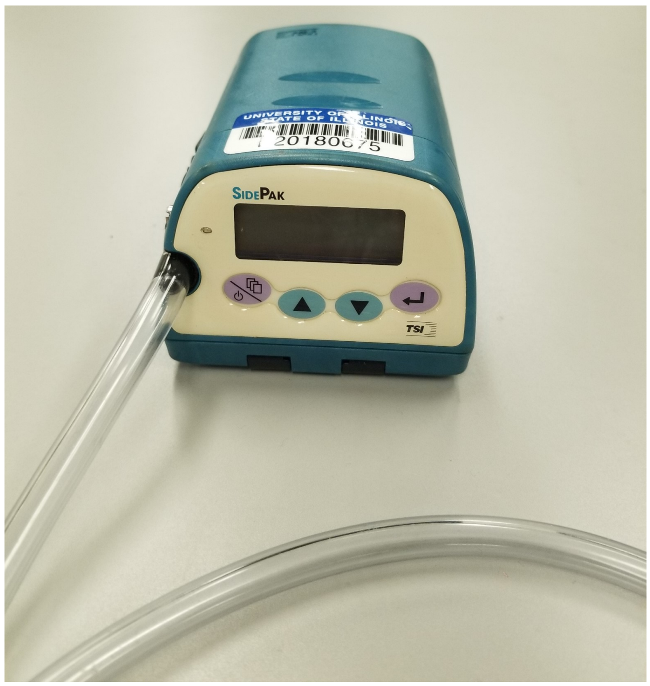

Figure 5: SidePak™AM510 Personal Aerosol Monitor

3 .2.2 Low-cost $PM_{2.5}$ Sensor

The low-cost $PM_{2.5}$ sensor that was calibrated is the Plantower PMS7003. It is a low-cost off-the-shelf $PM_{2.5}$ sensor. The sensor is deployed in the prototype developed as a part of this project. The sensor measures $PM_{2.5}$ concentrations using the laser scattering method. Please refer to 2.1.2 (Chapter 2) for further descriptions and details.

3 .3 Conceptualization

Independent Variables:

Pollutants. Our pollutants are PM_1, $PM_{2.5}$, PM_{10}, and noise measured in dBA.

Type of sensing device. We used two types of sensing devices: SidePak as a high-cost high-precision pollution monitoring device and Plantower PMS7003 as low-cost low-precision $PM_{2.5}$ sensor.

Contextual Features. Contextual features are the sets of data collected by the smartphone application. These are timestamps, geotags represented by longitude and latitude, WiFi and GSM signal levels, screen's brightness and luminosity value, weather information such as outdoor temperature, humidity, dew point, weather condition, and wind speed, ambient temperature inferred from the battery temperature, microenvironment, and, setting.

Dependent Variable: Value (distribution).

3 .4 Data Collection

SidePak and the prototype were used to collect data for this study. The tube opening of the SidePak where the pump draws air samples from was attached to the prototype. Figure 6 shows a picture during one of the data collection trips.

Figure 6: Data Collection for Calibration

For this study, we collected data for an approximation of 20 hours with a read rate of 1 read/second for both SidePak and our prototype. The data collection was done over a period of 4 weeks in different locations near Chicago and UIC campus as well as in Chicago suburbs in different times of the day. We collected data in settings of high particulate matter concentrations such as a kitchen when cooking, and low concentration such as in a computing lab. Also, we collected data in different weather conditions with varying humidity and temperatures in stationary and during movement (walking and in a car).

The data collection was done in four different microenvironments: deep indoor which resembles a room with no windows or doors that lead to straight to outdoor, indoor which is a room or an indoor space that has windows and/or doors leading outdoors, semi-outdoor which is a microenvironment that is outdoor but has significant walls or roofs or buildings, and, outdoor which is an open outdoor area. Figure 7 shows a side of data collection in an outdoor microenvironment.

Figure 7: Outdoor Data collection

The smartphone collected contextual data while SidePak and the wearable collected particulate matter data. We collected contextual data to examine the effect of different factors that might affect the $PM_{2.5}$ concentration. Also, we collected PM_{10} and $PM_{1.0}$ using our sensor as it collects all three values of $PM_{2.5}$ at the same time instance. Studies have shown a strong correlation between each two consecutive sizes.[90] Thus, we believe that this data can improve our calibration model.

We created python scripts to aggregate the data from SidePak and from our prototype and smartphone application data based on timestamp matching. The data was then saved to a CSV file and it was analyzed using R. The final data set had 70558 observations of 22 variables. Table II shows a summary of the collected $PM_{2.5}$ concentrations collected using SidePak.

	Min.	Max.	Median	Mean
Value in μg/m^3	0.00	794.00	10.00	15.92

TABLE II: SidePak $PM_{2.5}$ values overview

Table III shows an overview of the $PM_{2.5}$ concentrations collected using our wearable. Comparing to table II, difference in means of measurements stresses the need for calibration.

	Min.	Max.	Median	Mean
Value in μg/m^3	0.000	100.000	5.000	9.417

TABLE III: Wearable $PM_{2.5}$ values overview

Although in this work we are not focusing on calibrating microphones for noise data, we collected noise data as a contextual variable in order to test if noise level can be a feature for predicting actual $PM_{2.5}$ concentrations. A summary of noise data is shown in table IV.

	Min.	Max.	Median	Mean
Value in dBA	-25.86	92.13	60.51	62.23

TABLE IV: Noise values overview

Finally, we show in table V the distribution of our observations against microenvironments. A majority of the data was collected in indoor and deep indoor microenvironments. We report testing the effect of microenvironment on calibration.

Experiment	Deep indoor	Indoor	Semi-outdoor	Outdoor
Number of observations	42437	18309	4142	5670

TABLE V: Data by Microenvironment

3 .5 Data Analysis

In this quantitative study for exploring low-cost $PM_{2.5}$ sensors and adapting them for creating low-cost personal pollution exposure monitoring wearables, we were interested in calibrating suboptimal low-cost $PM_{2.5}$ sensors. We collected a data set using SidePak as a ground truth reference and our prototype as the area of study for calibrating the sensor. The data collected from the both SidePak and the low-cost $PM_{2.5}$ sensor did not follow a normal distribution. Statistical analysis using Wilcoxon rank sum test did show a significant difference between the means of $PM_{2.5}$ values from the SidePak and the means of $PM_{2.5}$ values from the low-cost $PM_{2.5}$ sensor with a p-value <0.0001.

Figure 8 shows the $PM_{2.5}$ data in the data set plotted as a time series. Figure 8 shows trends in the data affecting all measuring devices. PM25 is $_{2.5}$ value measured by our wearable, PM10 is PM_1 measured by our wearable as well, also, PM100 is the concentration of PM_{10} as measured by our wearable. PM25_sp is $PM_{2.5}$ concentration measured by SidePak.

Trends in concentrations of $PM_{2.5}$ are perceived in a noticeably similar fashion by both SidePak and our wearable. But, differences still need to be tackled in order to get optimal sensing.

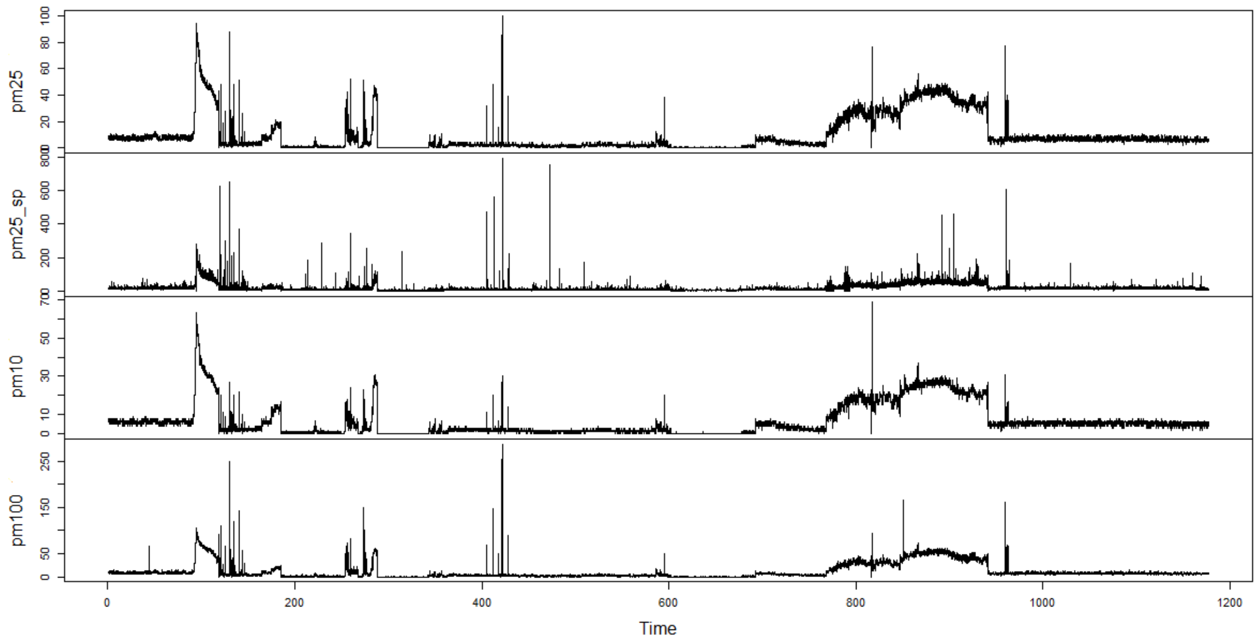

Figure 8: Calibration data plot

In this analysis, we are exploring different paths to create a model to predict $PM_{2.5}$ concentrations from the low-cost $PM_{2.5}$ sensor's readings supported by contextual data. We use the data we collected using SidePak to infer the model. We kickoff this analysis by creating and testing a linear regression model. Next, we explore a set of machine learning algorithms: nïve bayes, random forest, and, support vector machine (SVM).

In order to reduce the noise in the data, we classified the $PM_{2.5}$ values into 6 categories based on the Air Quality Index (AQI) reference provided by the EPA. EPA has published a reference for air quality index and the associated $PM_{2.5}$ values. This reference categorizes pollution concentrations based on the anticipated health concern of exposure over a 24-hour period to pollution with the given concentrations.[1,2] Table VI shows the $PM_{2.5}$ values and their corresponding pollution levels and the anticipated health effect.

$PM_{2.5}$ concentration range in $\mu g/m^3$	Air Pollution Level	Health Implications
0.0 - 12.0	Good	Pollution has no or little risk
12.1 - 35.4	Moderate	Acceptable pollution with a moderate health concern
35.5 - 55.4	Unhealthy for Sensitive Groups	Children, older adults, and people with health concerns such as respiratory health concern are at risk.
55.5 - 150.4	Unhealthy	Might affect the health of the general public. Poses a serious health issues for sensitive groups.
150.5 - 250.4	Very Unhealthy	Health Warning and the general public is more likely to be affected.
250.5 and up	Hazardous	Health Alert. Serious heath threat for everyone.

TABLE VI: $PM_{2.5}$ Index: health effects of $PM_{2.5}$ concentrations according to EPA Air Quality Index, adapted from[1,2]

The classified SidePak $PM_{2.5}$ observations are distributed as shown in table VII. We notice that a majority of the observations fall in group 1, 2, and 3. This indicates that a majority of our data is with low $PM_{2.5}$ concentrations. We report the effect of $PM_{2.5}$ concentrations on our calibration models below.

Group	1	2	3	4	5	6
Number of observations	39742	22575	5053	3029	136	23

TABLE VII: Data distribution per EPA's health-effect-based grouping of SidePak readings

3 .6 Calibration Model

Low-cost $PM_{2.5}$ sensors have shown to be sub-optimal by deviation in readings from the high-cost high-precision pollution monitoring tool we used, the SidePak. Thus, we collected data in order to create a model that infers actual $PM_{2.5}$ concentration levels from patterns and knowledge held in the data.

3 .6.1 Linear Regression

We start our analysis with creating a linear regression model to predict the concentrations. As a beginning, we calculated 5-minute time series means in order to reduce the noise and improve the regression. We calculated Kendall rank correlation between the means of the low-cost sensor's $PM_{2.5}$ and $PM_{2.5}$ measured using SidePak and it was equal 0.864 which indicates the existence of such relationship. Figure 9 is a plot of the relationship between the two variables at hand: Low-cost sensor $PM_{2.5}$ means and SidePak $PM_{2.5}$ means.

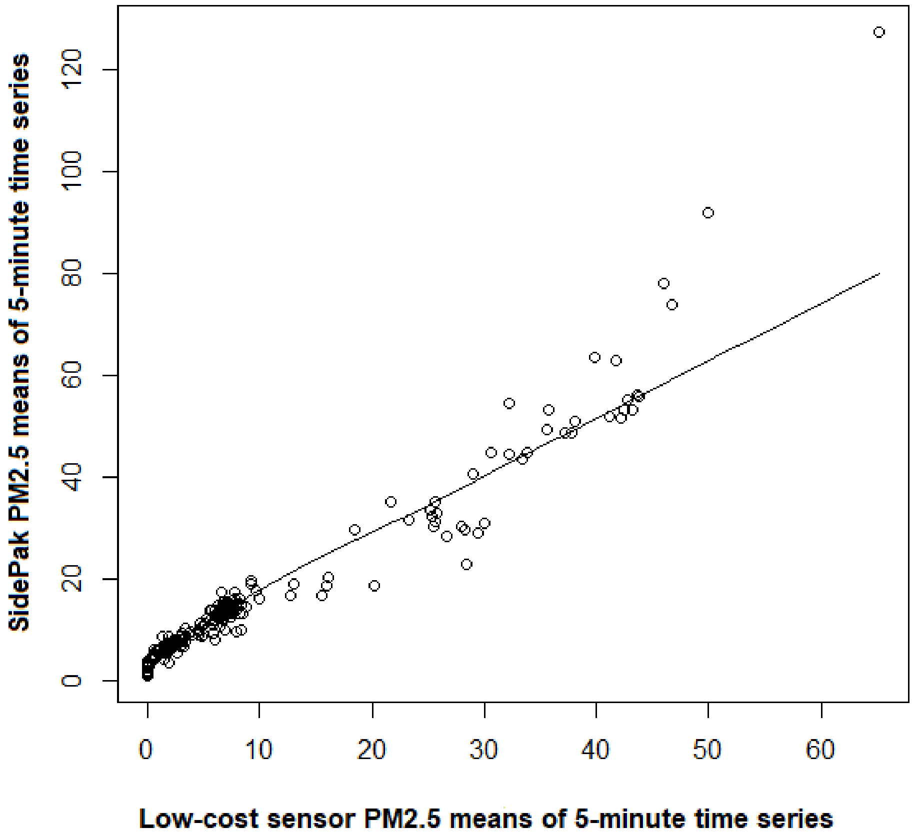

Figure 9: Low-cost sensor and SidePak $PM_{2.5}$ Means Relationship

Next we divided the means data set into training (160 observations) and testing(69 observation) sets. The training data was used to create a linear regression model to predict actual $PM_{2.5}$ concentrations. The model was then tested using the test data set. The model shows a weak correlation between the predicted and the actual values with a correlation coefficient of -0.019. This indicates that the relationship between them is weak. The root mean square error is 25.820 which is high value indicating that the model does not perform very well.

The raw sensor readings, as is, had a lot of noise, so, we labeled the data according to EPA's grouping per health effect shown previously in table VI. We created a set of classification models by training machine learning algorithms on the labeled data using R. We tried creating models based our data with a variety of classifiers and using different subsets of our contextual variables, we chose 3 classifiers to compare, Naïve Bayesian classification, Support Vector Machine (SVM) with a radial basis function (RBF) kernel, and, Random Forest Classification.

Table VIII shows a summary of our classifiers when trained and tested on all of our observation (or a random subset for a Naïve Bayes classifier). We report in the table the train and test data sets sizes and the set of features we used to train each model, and, finally, the accuracy of each model. As the table implies, our best classifier is a Naïve Bayes with an accuracy of 71.1% (fairly good) and $PM_{2.5}$, $PM_{1.0}$, $PM_{10.0}$ as our feature set. We examined all of our features, and we chose the best model for each classification method shown.

In order to better understand the nature of our data, we created a set of classifiers based on subsets of data that were divided based on two criteria: $PM_{2.5}$ concentrations and the microenvironment of which the observation was captured in. We decided on those

Classifier	Train Size	Test Size	Features	Accuracy
Naïve Bayes	28386	42172	$PM_{2.5}$, PM_1, PM_{10}	71.1%
Naïve Bayes (on a sample of the data)	6420	9580	$PM_{2.5}$, PM_1, PM_{10} luminosity, noise level, microenvironment, and, GSM signal strength	67.8%
Random Forest	28431	42127	$PM_{2.5}$	61.8%
Support Vector Machine	28431	28431	$PM_{2.5}$, PM_1, PM_{10} luminosity, noise level, microenvironment, and, GSM signal strength	42.0%

TABLE VIII: Comparison of Classification models built using the complete calibration data set

two criteria because there was less effect of other features on the classification. Table IX shows our $PM_{2.5}$ concentration based classification.

As table IX shows, we divided our data to two sets. First, set of data where concentration of $PM_{2.5}$ measured by SidePak was less than or equal to 55.4 µg/m^3 (data that belonged to group 1, 2, and, 3 in the health-effect-based classification). The other set was the complement of the first set with observation with SidePak reading greater than 55.4 µg/m^3 (belonging to group 4, 5, and, 6).

We trained 3 models on each set using the naïve Bayes, Random Forest, and Support Vector Machine classifiers. We report in the table train and test data sizes, feature set for each model, and, accuracy of the model. Models trained on larger $PM_{2.5}$ concentrations outperformed with higher accuracy. This indicates that calibration is more accurate at situations where concentrations of $PM_{2.5}$ are higher. It is worth mentioning that Random

Forest classifier performed the best and incorporating the feature set improved the model as opposed to the model built using the full data set.

Data	$PM_{2.5}$ <= 55.4 µg/m^3				$PM_{2.5}$ > 55.4 µg/m^3			
Classifier	Train Size	Test Size	Features	Accuracy	Train Size	Test Size	Features	Accuracy
Naïve Bayes	27091	40279	$PM_{2.5}$, PM_1, PM_{10}	58.7%	1249	1939	$PM_{2.5}$, PM_1, PM_{10}	91.7%
Random Forest	27091	40279	$PM_{2.5}$, PM_1, PM_{10} luminosity, noise level, microenvironment, and, GSM signal strength	55.8%	1292	1896	$PM_{2.5}$, PM_1, PM_{10} luminosity, noise level, microenvironment, and, GSM signal strength	92.6%
Support Vector Machine	27091	27091	$PM_{2.5}$, PM_1, PM_{10} luminosity, noise level, microenvironment, and, GSM signal strength	45.0%	1292	1292	$PM_{2.5}$, PM_1, PM_{10} luminosity, noise level, microenvironment, and, GSM signal strength	91.6%

TABLE IX: Comparison of Classification models built by training by $PM_{2.5}$ concentrations

Our other intra-group models were across microenvironments. We divided our 4 microenvironments (deep indoor, indoor, semi-outdoor, and, outdoor) to 2 general groups (Out and In) based on high-level similarity. We report our division in the table X.

Group	Microenvironments	Similarity
Out	Outdoor, Semi-outdoor	Not inside a closed area with 4 walls and a roof
In	Indoor, Deep indoor	In an area surrounded by walls and a roof

TABLE X: Microenvironment Groups

After dividing the data into two sets using the subsetting concept shown before, we trained and tested classifiers using each sub-group. We report a summary of the results in table XII. For each sub-group, the table shows the sizes of the train and test data for each classifier, as well as, features we used to train the classifier and the accuracy of that classifier. Our results show that calibration of indoor and deep indoor data outperformed calibration in outdoor and semi-outdoor setting. The model built using naïve Bayes classification for indoor and deep indoor data outperformed all the other models significantly.

Next, we discuss each of our models and show confusion matrices for each classifier.

Data	Out				In			
Classifier	Train Size	Test Size	Features	Accuracy	Train Size	Test Size	Features	Accuracy
Naïve Bayes	3966	5846	$PM_{2.5}$, PM_1, PM_{10}	65.5%	24429	36317	$PM_{2.5}$, PM_1, PM_{10}	82.9%
Random Forest	3957	5855	$PM_{2.5}$, PM_1, PM_{10} luminosity, noise level, microenvironment, and, GSM signal strength	36.9%	24429	36317	$PM_{2.5}$, PM_1, PM_{10} luminosity, noise level, microenvironment, and, GSM signal strength	53.2%
Support Vector Machine	3957	3957	$PM_{2.5}$, PM_1, PM_{10} luminosity, noise level, microenvironment, and, GSM signal strength	35.6%	24429	24429	$PM_{2.5}$, PM_1, PM_{10} luminosity, noise level, microenvironment, and, GSM signal strength	50.4%

TABLE XI: Comparison of Classification models built by training by microenvironment

3 .6.2 Naïve Bayes

Naïve Bayes classification is a probabilistic classification that adapts the Bayesian algorithm for conditional probabilities assuming independence. We created multiple models using the naïve Bayes classification method on the whole data set as well as for different subsets. For each model, we divided our data set randomly to train data and test data at a 2:3 ratio. We trained our model using the data for training with the following variables: SidePak $PM_{2.5}$ labels as the classifications, $PM_{2.5}$, PM_1, and PM_{10} values from our prototype. For one model, we incorporated, in addition to the previous, noise values in dBA, microenvironments, GSM signal strength, and, luminosity values. Models trained with PM data only outperformed models using feature set or feature subsets, thus, we chose to report the models with PM data as features. We used the model to classify test data. Train data consists of 28386 observations and test data consists of 42172 observations. The model had an accuracy of 71.1%. We show the confusion matrix in figure 10.

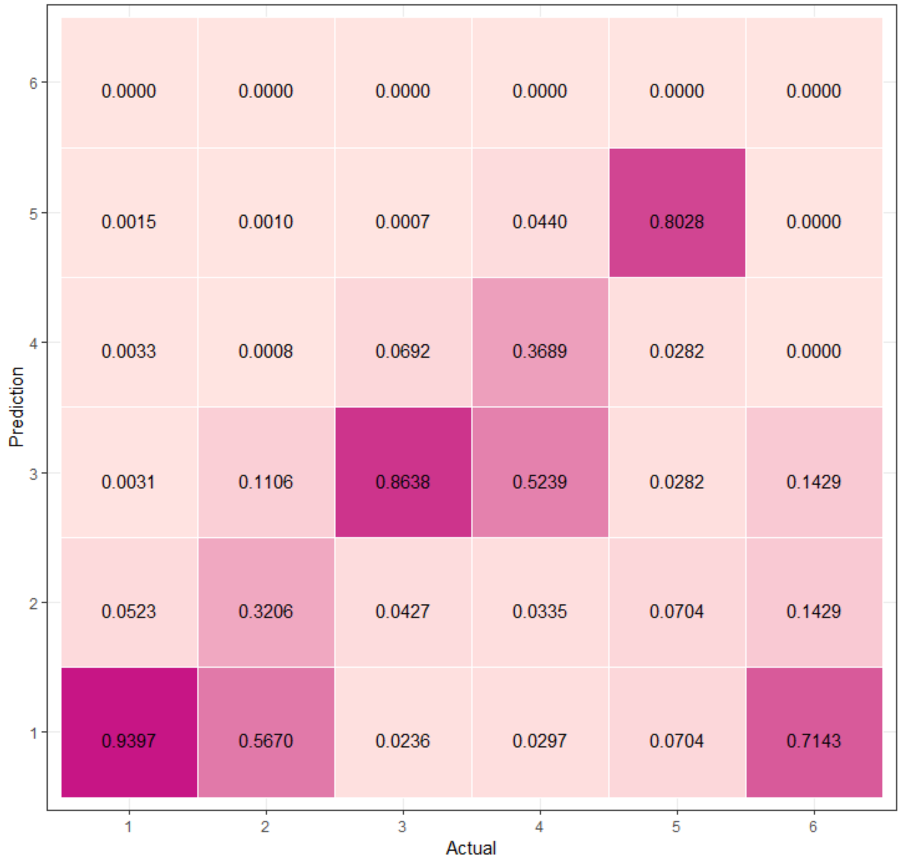

Figure 10: Naïve Bayes confusion matrix for calibration (complete data set)

In figure 10, the X-axis represents the Actual values of measured $PM_{2.5}$ for the test data represented by the health-based group it belongs to. The Y-axis shows the predicted group. In other words, what the model classified a certain data point. The color intensity indicates the percentage of the actual x group that was predicted as y group. So, the columns sum to 1 as for 100% of the actual x group. This idea applies to all confusion matrices shown in the thesis.

That said, what we are looking for in a confusion matrix is high color intensity when x = y. In this model, the accuracy was 71.1%. The model performed best at classifying group 1, 2, and, 5. The model classifies group 2 as group 1 and group 4 as group 3, this can be due to the proximity between these two sets of groups. The model is offset with classifying group 6 as it mostly classifies it as group 1. This could be because of the low number of observations in group 6 as per table VIII.

As to better understand the effect of feature set on naïve Bayes classification. We show the best model we created that classified the data with PM concentrations as well as noise values in dBA, microenvironments, GSM signal strength, and, luminosity values as features. The model was trained on a data set with 6420 observations and tested on a test data set of size 9580 observations. The model has an accuracy of 67.8%. Figure 11 shows the confusion matrix for this model.

Figure 11: Confusion matrix of the naïve Bayes classifier built using a set of features on a random subset of the data

As the confusion matrix shows, the model is best at predicting group 1, 3, and, 5. For group 3, the model predicted a majority of the data as group 1 and 2 in a similar fashion to that of the previous model. Similarly, the model predicts majority of group 4 as group 4 and 3. The noticeable difference here is that this model predicts the majority of group 6 as group 3 which is closer in value to the actual group. Thus, adding features improved the prediction of the high-value groups and reduced the accuracy of predicting the low-value groups.

We divided our data set to two groups; in and out as explained in table XI. We created two naïve Bayes models for "in" and "out" data to observe the effect of general microenvironments on classification. We created general microenvironment groups because classifying per microenvironment did show the exact pattern we observed when classified per group, so, we grouped as an effort to generalize the results. The "in" model, our train data had 24429 observations and the test data had 36317 observations. We used the PM data as feature set: $PM_{2.5}$, PM_1, and, PM_{10} concentrations. The accuracy of our model when used to predict test data was 82.9%. Figure 12 shows the confusion matrix for "in" calibration model.

Figure 12: Confusion matrix for "in" calibration model using naïve Bayes

The confusion matrix shows that the model correctly predicted a majority of group 1, 2, 4, and, 5. In a pattern similar to the model of the complete data set, this model classified most of the group 6 data points as group 1. This model classifies a majority group 3 as group 4 as opposing to classifying as the lower group in the complete data set model.

Next, we created a model using the "out" data set. We divided the set to a train set of size 3966 observations and a test set of size 5846 observations. The features used to train the model are PM concentrations: $PM_{2.5}$, PM_1, and, PM_{10}. The model had an accuracy of 65.5% and with a performance lagging behind the model built to classify indoor data. Figure 13 shows the confusion matrix for the "out" model.

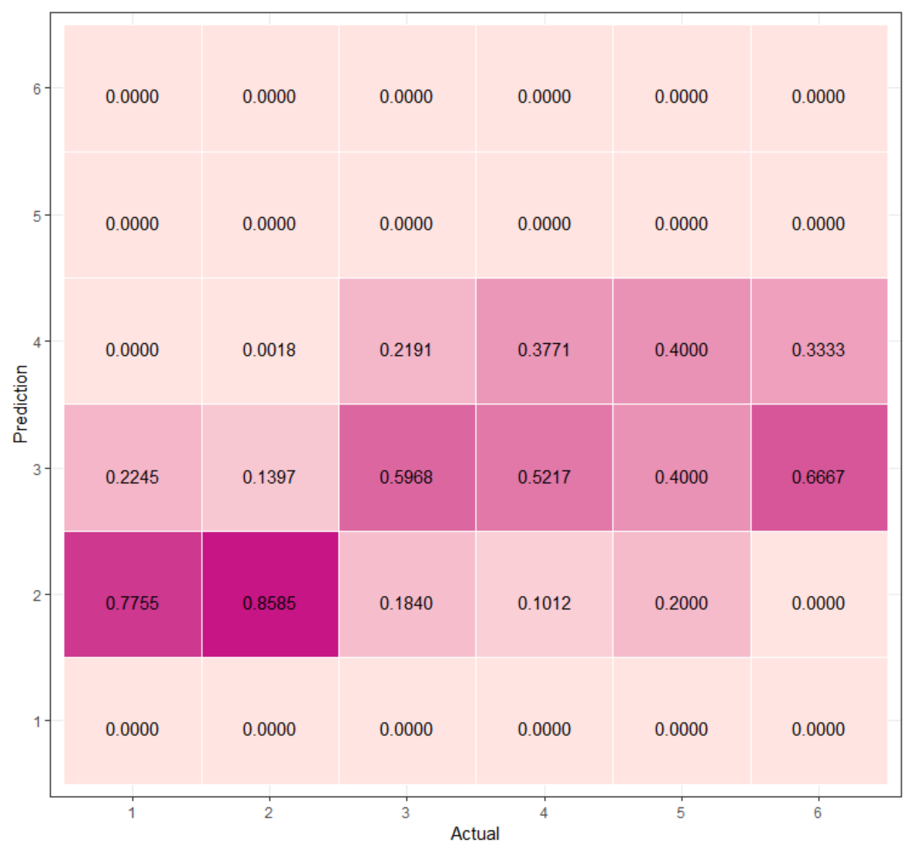

Figure 13: Confusion matrix for naïve Bayes "out" model

The confusion matrix in figure 13 shows clearly the low accuracy of the model. The model does not classify any data point to group 1, group 5, or, group 6 correctly. The model is best as classifying group 2 and group 3. This can be understood as due to the generally larger concentrations of PM in the two outdoor microenvironments. Similar to the model built using the large feature set, the model classifies group 6 as group 3. The model classifies groups 4 and 5 as groups 3 and 4.

Other than subsetting the data per general microenvironment groups, we subsetted the data per $PM_{2.5}$ concentrations. We divided the data to two sets: with $PM_{2.5}$ concentration $<= 55.4$ µg/m^3 and with $PM_{2.5}$ concentration > 55.4 µg/m^3. We created a model for each set.

We first created a model with $PM_{2.5}$ concentration $<= 55.4$ µg/m^3. The model was trained using a train set of size 27091 observation and we tested it on a test set of size 40279. The test yielded and accuracy of 58.7%. Which indicates that the naïve Bayes classification is worse when classifying smaller concentrations of $PM_{2.5}$. Figure 14 shows the confusion matrix for this model. The confusion matrix shows that the model classified all the data as group 1.

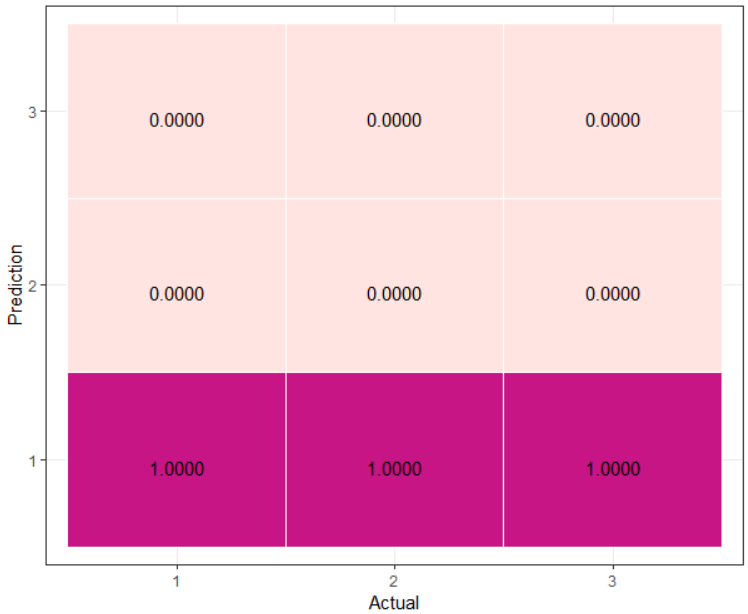

Figure 14: Naïve Bayes classification's confusion matrix for lower $PM_{2.5}$ concentrations model

Finally, we classify the larger concentrations data set. We divided the set to train and test at a 2:3 ratio with the train set of size 1249 and a test set of size 1939. We used $PM_{2.5}$, PM_1, and, PM_{10} as features. When testing the model using the test data set, it resulted and accuracy of 91.7%. Thus, naïve Bayes classification performs best at higher $PM_{2.5}$ concentrations. Figure 15 shows the confusion matrix of the model.

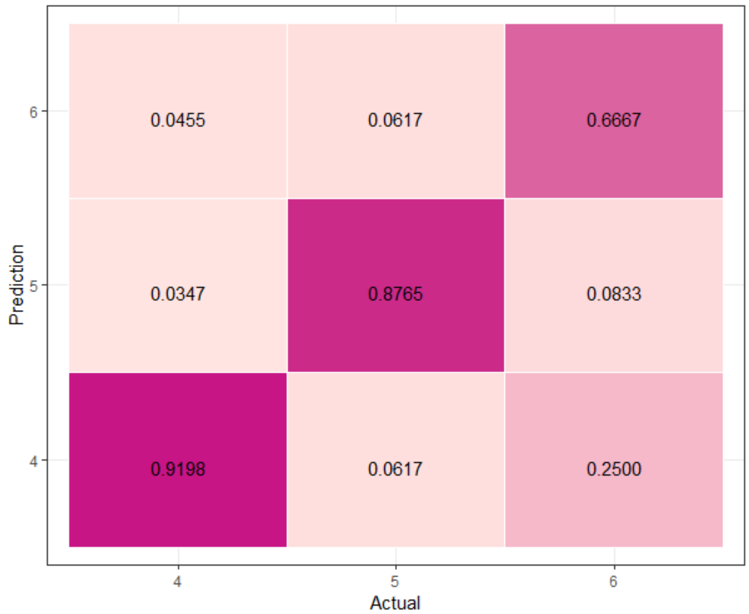

Figure 15: Confusion matrix of naïve Bayes classifier using data of higher $PM_{2.5}$ concentrations

The confusion matrix shows that the model classifies most of the group 4 and group 5 data correctly. The model classifies quarter of group 6 as group 4. This pattern of classifying group 6 to lower groups was seen on all models, but, this model is the best at classifying group 6.

Naïve Bayes classifiers show clearly the effect of changing contextual features on the classification. The classifiers perform fairly good for our calibration data.

3 .6.3 Random Forest Classification

Our second model is a random forest classifier. A random forest classifier is built using a set of decision trees, which are trees used to make decisions, where nodes are classes and branches are decision results. Random Forest classification follows the method of ensemble learning where it combines a set of sub-classifiers (Decision trees) for improved classification.

We used the calibration data we collected to train and test a random forest classifier. We obtained the train and test set by randomly dividing the data set to train and test at a 2:3 ratio (28431 observations for training and 42127 observation for testing). The features which we used for training the model are the labeled SidePak $PM_{2.5}$ data as the classification and the $PM_{2.5}$ data collected by our sensing model. We tested the random forest classifier model on the test data and it yielded an accuracy of 61.8%. Figure 16 shows the confusion matrix for our classifier.

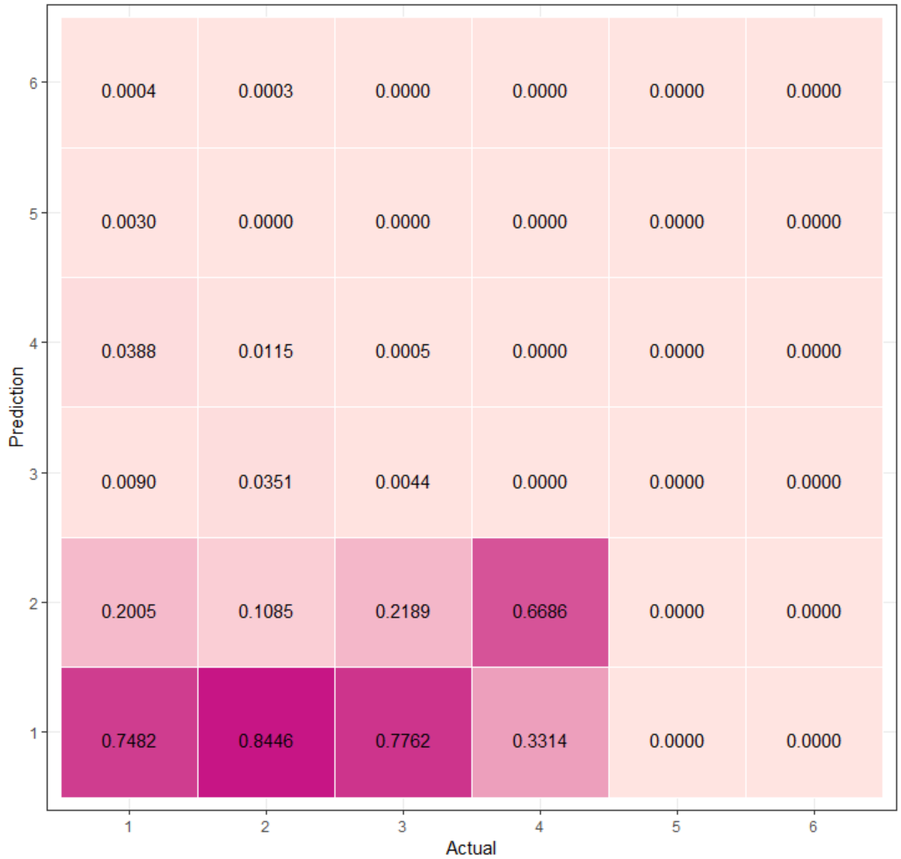

Figure 16: Random Forest classification confusion matrix for calibration

From the confusion matrix shown in figure 16, the classifier is classifying most of the data group 1 and group 2. The test data appear to not have any group 6 data points. The classifier is not reliable.

We used Random Forest classification to classify "in" and "out" data subsets. We use PM concentrations, noise values in dBA, microenvironments, GSM signal strength, and, luminosity values as features.

For the "in" model, we divided the data to train and test sets. Train set has 24429 observations and test has 36317 observations. The model has an accuracy of 53.2% with the confusion matrix shown in figure 17.

Figure 17: Random Forest classification of "in" data

From the confusion matrix, the model classifies most of group 1 and 2 as group 1 and most of group 4 and 5 as group 2. Also, the model has a pattern of classifying most group 3 as groups 4 and 5 and most of group 6 as group 4. The confusion matrix shows that the model is faulty.

Next, we create a model to classify "out" data using random forest. The "out" model uses the same feature set as the "in" model. Also, the "out" data set was divided into a train set of 3957 observations and 5855 observations for the test set. The model has an accuracy of 36.9%. Figure 18 shows the confusion matrix for "out" data.

Figure 18: Random Forest Classifier for "out" data

As clear in the confusion matrix, the "out" classifier has a tendency to classify data to groups that represent higher $PM_{2.5}$ concentrations. This can be interpreted as the effect of having higher values of $PM_{2.5}$ concentrations.

We divided our data based on $PM_{2.5}$ concentrations to two groups with the value 55.4 μg/m^3 as the critical value inclusive of the lower concentration group. Then we used

PM concentrations, noise values in dBA, microenvironments, GSM signal strength, and, luminosity values as features to train two models. The first model classifies lower $PM_{2.5}$ concentration. This model was trained using a set of 27091 observations and tested using a set of 40279 observations. The model has an accuracy of 55.8%. Figure 19 shows the confusion matrix for the model.

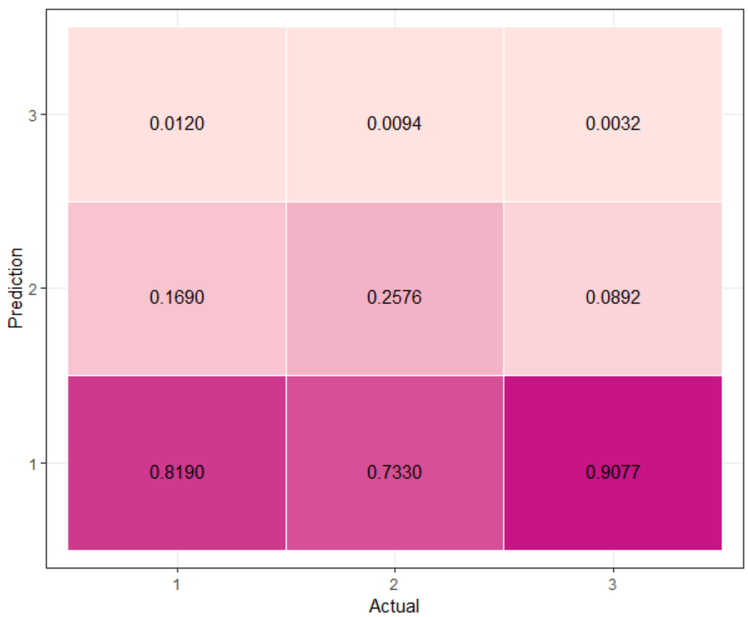

Figure 19: Confusion matrix for Random Forest classifier of lower $PM_{2.5}$ concentration data

The confusion matrix shows that the model is biased towards group 1. Thus, random forest is not a good classifier for lower $PM_{2.5}$ concentration.

The second model is trained using 1292 observations at high $PM_{2.5}$ concentrations using the same feature set used for the previous model. This model was tested on 1896 observations and yielded an accuracy of 92.6%. Figure 20 shows the confusion matrix of the model's prediction of test data.

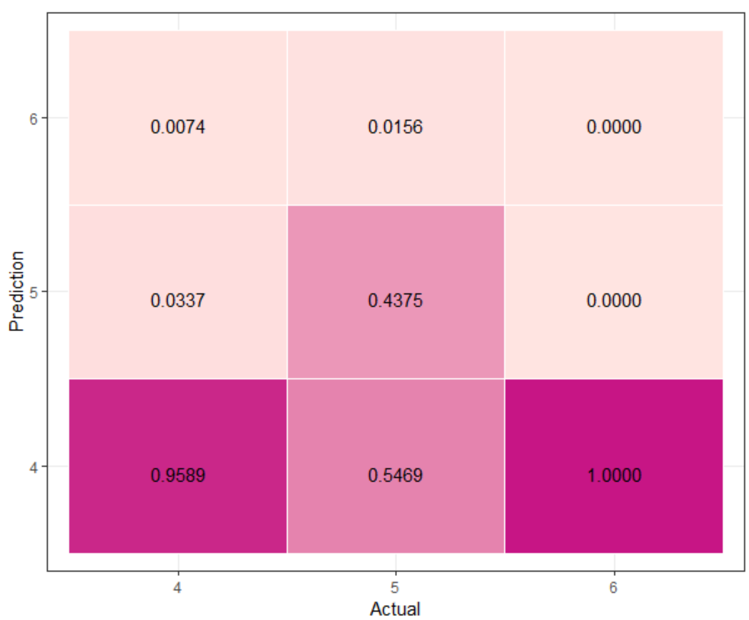

Figure 20: Random Forest classifier's confusion matrix for high $PM_{2.5}$ concentration data

Similar to the prediction of the low concentration model, the model is biased towards group 4. According to table VIII, the high accuracy of the model is due to the majority of the data being in group 4.

We notice that Random Forest classifiers are less affected by contextual features and more affected by the number of observations per group in the training data.

3 .6.4 Support Vector Machine (SVM) with RBF kernel

Our last model is a support vector machine classifier with a radial function as a kernel. SVM functions by creating hyperplanes that separate data classes and the prediction of new data depends on where does the data point fall in the space containing these hyperplanes. Since the data is complicated and could not be separated linearly, we used the kernel method to transform the space using a radial function. The RBF kernel function is:

$$K(x, y) = exp(-\gamma \sum_{j=1}^{p}(x_{ij} - y_{ij})^2)$$

Our SVM model was trained using the same train set that we used for the other models which is obtained by randomly dividing the data set to train and test at a 1:1 ratio since the model performed the best at this ratio. The features which we used for training the model are PM_1, $PM_{2.5}$, PM_{10} data that was collected using the sensor in our prototype, in addition to noise values, luminosity values, microenvironments, and GSM signal levels. These properties apply to all the models we built using SVM classification.

Our first model is built using a random sample of the data set divided into train set and test set of size 28431. We tested the SVM model on the test data and it yielded an accuracy of 42%. We show the confusion matrix in figure 21.

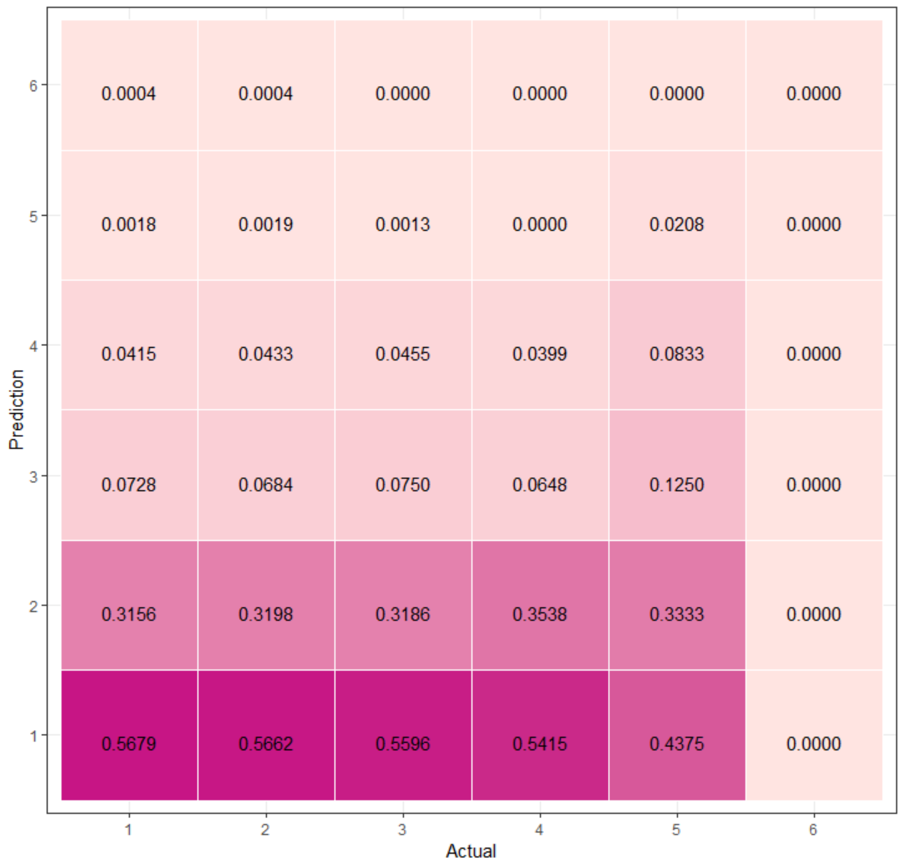

Figure 21: SVM with RBF kernel confusion matrix for calibration

Similar to random forest classification, SVM has predicts most of the data as group 1 and 2. The cause of this is the distribution of the data.

We used SVM to classify the microenvironment subsetted data. First, we created a model using the "in" data set. The train and test data sets had 24429 observations and we used the same feature set as the previous SVM models. The accuracy of this model was 50.4%. Figure 22 shows the confusion matrix for this model.

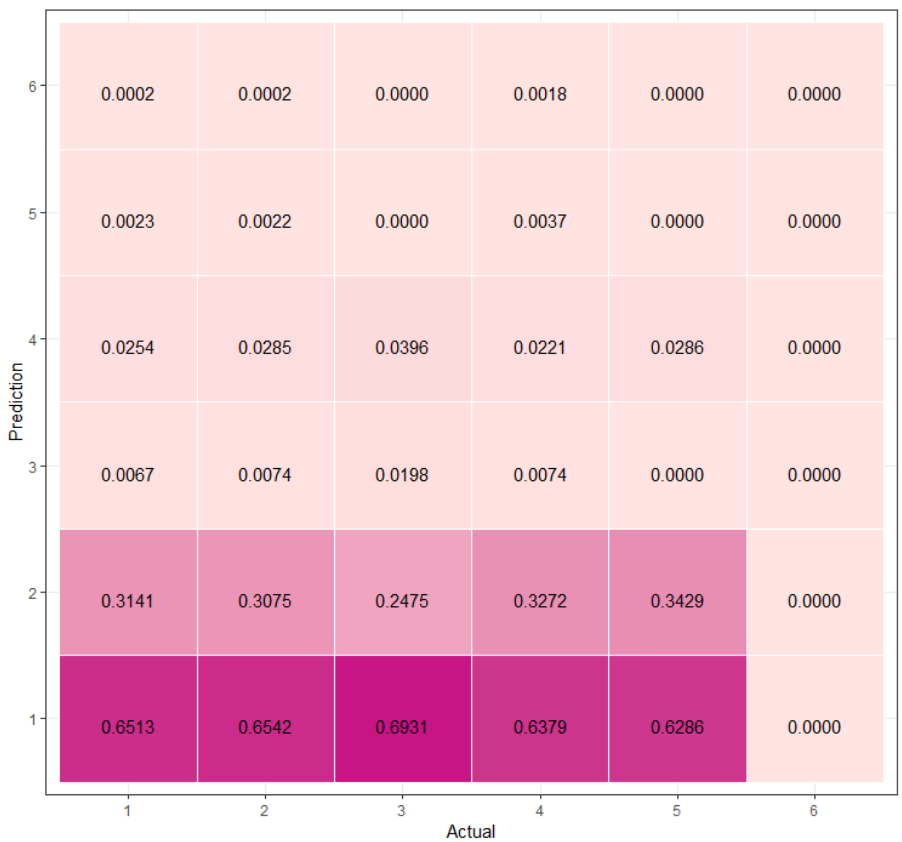

Figure 22: Confusion matrix for SVM with RBF kernel "in" model

The confusion matrix is similar to the confusion matrix of the previous model. Thus, subsetting did not affect the classification.

The second model classified the "out" data in a manner similar to that of the "in" data classifier. The model was created and tested with train and test data sets of size 3957 observations. The model yielded an accuracy of 35.6%. Figure 23 shows the confusion matrix for this model.

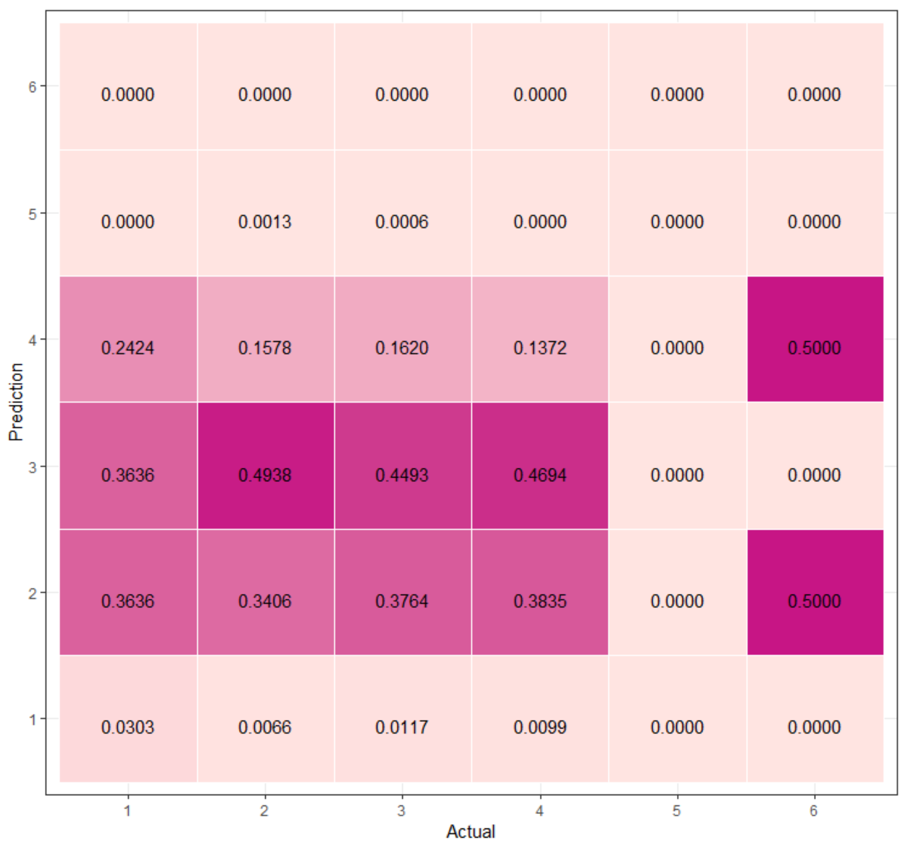

Figure 23: SVM with RBF kernel "out" classifier’s confusion matrix

In this confusion matrix, the classifier classifies mainly to group 2, 3, and, 4. This is a result of having higher values of $PM_{2.5}$ in "out" microenvironments, thus, the bias shifted to a higher concentration.

Finally, we divided our data based on $PM_{2.5}$ values to two groups: values less than or equal to 55.4 μg/m^3 and larger than 55.4 μg/m^3. Using the same feature set, we created two classification models using SVM with RBF kernel.

The first model of lower $PM_{2.5}$ concentrations was trained and tested using 27091 observations for each. We used the same feature set used for previous SVM models. The model yielded an accuracy of 45%. Figure 24 shows the confusion matrix of the model's predictions.

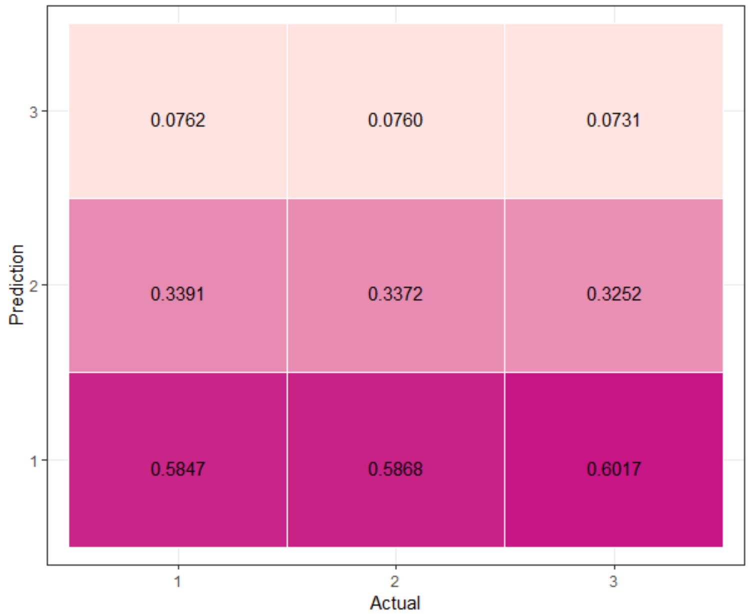

Figure 24: Confusion matrix for lower $PM_{2.5}$ concentration classification using SVM with RBF kernel

The confusion matrix, similar to previous, is biased towards group 1, then group 2.

Lastly, we created a model using data that has $PM_{2.5}$ concentrations higher than 55.4 µg/m^3. The model was built using train and test data set each with a size of 1292 observations. The model has an accuracy of 91.6%. Figure 25 shows the confusion matrix of the model.

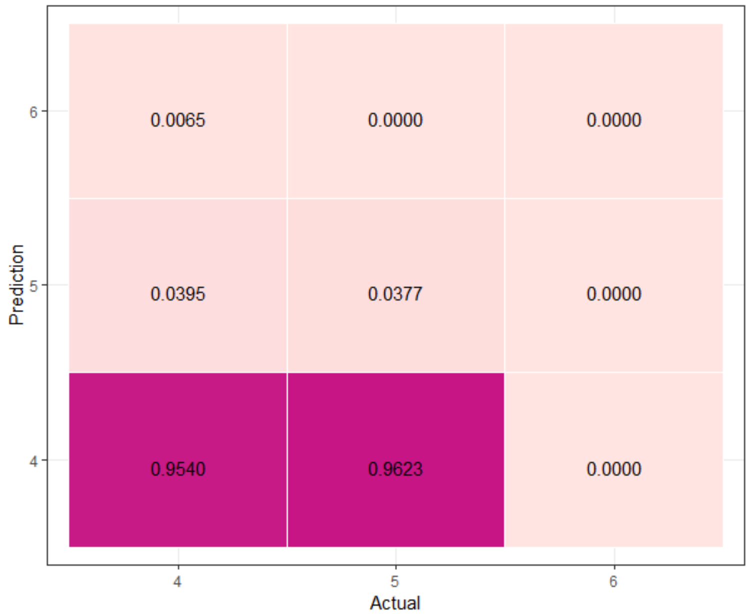

Figure 25: Confusion matrix for SVM with RBF kernel classifier for data with higher $PM_{2.5}$ concentration

In spite of the high accuracy of the model, it is faulty and predicts mostly as group 4. We notice that different features do not affect the SVM model. It is affected mostly by the number of observations per group.

3 .7 <u>Findings</u>

This thesis is an effort towards creating a low-cost wearable for personal pollution exposure monitoring for the public in general and for the at-risk communities particularly. We firstly created a prototype using low-cost $PM_{2.5}$ sensor and powering it using a low-cost computing chip. Next, we explored the performance of the low-cost $PM_{2.5}$ sensor as opposing to high-cost high-accuracy pollution monitoring tool: SidePak.

Our studies have shown significance between the readings of the low-cost sensors and SidePak when running side-by-side measuring the same actual $PM_{2.5}$ concentrations. Based on that, we studied creating a model to calibrate low-cost $PM_{2.5}$ sensors.

We examined different algorithms with different sets of features by subsetting our variables. We started the analysis by creating a linear regression model for the means of 5-minute time series of the data at hand. The model performed poorly. Thus, relying on trade-offs between efficiency and accuracy, we chose 3 classification models to compare and discuss.

Starting with the SVM model, with the accuracy of the model we think that it is not functional to calibrate the sensor's readings. Also, our SVM models show that it is mostly affected by the number of observations per group. The models perform best using a large set of features, yet, it does not show effect when trained on data subsetted using these features. We think that collecting more data with an equal distribution amongst groups can improve the model as future work.

Next, the random forest model performed the best with no features, only the low-cost sensor's $PM_{2.5}$ readings and the labeled SidePak $PM_{2.5}$ data when trained on the full data set. The model performs poorly despite that recent works in calibration uses random forest.[91] Our confusion matrices have shown that, similarly to our SVM classifiers, the classification using random forest is unreliable and its prediction biased towards the group with the larger number of observations. Moreover, prediction using our random forest model takes significantly more time comparing to the other models.

Finally, our naïve Bayes model performed at a fairly good accuracy of 71.1% when it was trained on the full data set. The classifier performance drops when incorporating set of features. But, when training the classifier of data sets subsetted based on features, the classifier showed higher sensitivity to these features. This allowed us to understand that classification is better at higher $PM_{2.5}$ concentrations and it is better at indoor and deep indoor environments. With this model performing fairly good but not satisfactory, we plan on building on this model to improve our system. Prolonged data collection across different ranges of contextual data, for example, different seasons and locations, might improve this model for future work.

CHAPTER 4

HUMAN INTERFERENCE WITH LOW-COST $PM_{2.5}$ SENSORS

When assessing personal pollution monitoring, physical characteristics of the $PM_{2.5}$ measured in a human's personal cloud vary per source. Some studies have shown variation in toxicity of $PM_{2.5}$ based on its source.[33–35] Measuring $PM_{2.5}$ near the human body is prone to $PM_{2.5}$ resulting from the body's emission, such as skin and hair emissions[13] as well as bio-aerosol emissions of breath.[27] Also, clothing and textile contribute to the $PM_{2.5}$ in the personal cloud. Studies have shown that human body emissions increase with the increase of physical activity.[30]

4 .1 Motivation

Studying human interference with low-cost $PM_{2.5}$ sensors informs design guidelines for a personal pollution exposure monitoring wearable, particularly, where to wear the sensor so that the measured $PM_{2.5}$ is least interfered with. Moreover, studying the patterns of different human interference situations can help in predicting the situation or the activity the user is currently taking in order to infer more accurate personal pollution exposure monitoring.

We studied different situations of human interference with low-cost $PM_{2.5}$ sensors using the prototype we created and supported with ground truth measurements using Side-Pak.[88] We explored human interference using the two tools in an exploratory pilot study of which results, in addition to wearability and comfort, we made our decisions for our human interference experiments.

4 .2 Conceptualization

Independent Variables:

Pollutants. Our pollutants are PM_1, $PM_{2.5}$, PM_{10}, and noise measured in dBA

Type of the sensing device. We used two types of sensing devices: SidePak as a high-cost high-precision pollution monitoring device and Plantower PMS7003, deployed in our wearable, as low-cost low-precision $PM_{2.5}$ sensor.

Distances. We measured human-interference-induced disturbance using two sensors. One of the sensors is a reference sensor that was set at a fixed distance from the experiment. We chose 2 fixed distances for reference in our experiments, first is 30 cm (11.8 inches) and the second is 100 cm (39.3 inches). In our pilot study, we noticed that for some human interference situations, the effect of that situation on the $PM_{2.5}$ concentration differs within the range of 30 cm and the range of 100 cm. Thus, we chose 30 cm, and, 100 cm as our distances.

Human Interference. We measure changes in $PM_{2.5}$ at a set of human-interference conditions inferred by the set of preliminary experiments in our pilot study we conducted in the lab. These are:

- Control: Which is an experiment where no interference happens as the sensors are on a clean flat surface. We did the control experiments using SidePak as well as our prototype. Control experiment serves as a baseline with no human interference in our analysis. So, whenever we talk about the *baseline* we indicate that it is this control experiment.

- Skin: We examined a set of skin experiments: first, we measured $PM_{2.5}$ near bare skin with no touching or interference, second, we measured the effect of touching skin, next, we measured scratching skin, also, we measured near skin wearing bracelets, and finally, we measured sweaty skin with no interference. In the skin experiments, the sensor was worn like a wristwatch with no interference with the outfit. The experiment was conducted while typing on a laptop or reading a book as a lightweight activity. We measured skin experiments using SidePak as well as our prototype.

- Breath: We inspected the effect of breath on $PM_{2.5}$ values in the personal cloud. We conducted a set of experiments to explore 4 breath and breath-like situations: normal breathing, yawning, coughing, and, laughing. We measured breath experiments using SidePak as well as our prototype. Breath experiments were conducted with the sensor clipped on a collar and while doing lightweight activities such as typing on a laptop.

- Hair: We explored hair emissions and their effect on $PM_{2.5}$ concentrations. We studied two situations: touching and playing with hair and heat styling. These experiments were done with medium length hair. Hair experiments were conducted with the sensor clipped on a collar.

- Toiletries: As studies have shown that spraying products affect the concentration of PM[26], we studied the use of toiletries as a part of people's day-to-day life to examine the effect on personal pollution exposure. We conducted experiments where the experimenter was spraying perfume, facial spray, hair spray, and using loose-powder makeup. For this set of experiments, the experimenter wore sensor by clipping it to the collar. SidePak was used to measure $PM_{2.5}$ as well as our prototype.

- Textile: Textile emissions also affect the amount of $PM_{2.5}$ in the personal cloud. We examined 6 types of textile: Cotton, Leather, Silk, Synthetic fabrics, Wool, and, Fur. For this set of experiments, the experimenter wore sensor by clipping it on shirt pocket (or attaching it to the approximate location when no pocket existed). The experimenter was doing lightweight activity during these experiments. Also, we tested the existence of static electricity in the clothing items during the experiments using an off-the-shelf multi-meter by grounding it using a metal surface.

Contextual Features. Contextual features are the sets of data collected by the smartphone application. these are timestamps, geotags represented by longitude and latitude, WiFi and GSM signal levels, screen's brightness and luminosity value, weather information such as outdoor temperature, humidity, dew point, weather condition, and wind speed, ambient temperature inferred from the battery temperature, microenvironment, setting. Dependent Variable: Value (distribution)

4 .3 Data Collection

We conducted the experiments explained in the previous section in 3 settings: Lab, Home, and Library. Each experiment was repeated 3 times at each distance. Most of the experiments lasted for 15 minutes except for the experiments that will not hold for 15 minutes. These experiments are skin scratching, skin touching, cough, laugh, yawn, both of the hair experiments, and all of the toiletries experiment. SidePak was used with some of the experiments as discussed previously.

For the human interference experiments, two replicas of the prototype were needed. One prototype will serve as the experimental prototype and it is worn or attached as previously described. The other sensor is used as a reference sensor where it was measuring the ground condition by being set on a clean surface. The reference sensor was measuring at two distances; 30 cm and 100 cm, as discussed earlier. Figure 26 shows our two prototype replicas.

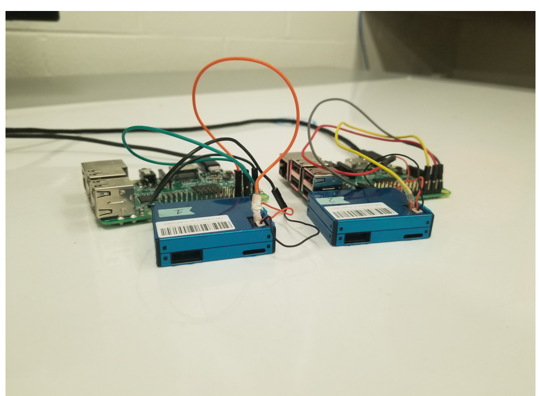

Figure 26: Prototype Replicas for Human interference

Data was collected over a period of 6 weeks. Data collected from both sensors and contextual data from the smartphone app was matched based on timestamp using a python script we wrote for this mission. After matching the dates and removing data lines that contained null values, we had 65468 data lines. Each line represents the values for 1 second with a total of around 18 hours of data. Experiments were manually labeled. Table XII below shows the number of observations for each experiment in our data set.

bracelet	breath	control	cotton	skin no interference	fur
5260	5111	5904	6085	6907	6200
hair spray	hair touch	laugh	leather	powder makeup	silk
684	3443	818	6138	753	1776
skin sweat	skin touch	synthetic	wool	spray perfume	yawn
5016	1464	2799	4230	651	336
face spray	hair heat style	skin scratch	cough		
557	596	534	206		

TABLE XII: Experiments data overview

4 .4 Data Analysis

$PM_{2.5}$ data collected by both $PM_{2.5}$ sensors as well as SidePak does not follow a normal distribution, thus, we analyzed the data using non-parametric statistical tests. Also, since the data was collected as separate time series (experiments), we compared means in parts of our analysis.

4 .4.1 Low-cost $PM_{2.5}$ sensors in Control Experiment

We measured significance between the two low-cost $PM_{2.5}$ sensors in the two replicas of our prototype using Wilcoxon rank sum test. The raw data we compared was the data collected in the baseline experiment. The result yielded a p-value = 0.058 and we calculated the effect size r = -0.045. This indicates that no significance worth considering between the two copies of sensors used for our experiments. Variance for the reference sensor data = 399.042, and variance for the experiment sensor data = 398.506. Median of the reference sensor data = 5, and that for the experiment sensor = 6. Finally, the mean for reference sensor is = 12.561 and for the experiment sensor is = 13.082.

Also, we labeled sensor's data using the EPA health-based $PM_{2.5}$ groups. We measured significance between the two groups of labels using Wilcoxon rank sum test. The test resulted a p-value = 0.142 with an effect size r = -0.019. The result confirms that the two copies of sensors perform similarly.

4 .4.2 Data Analysis per Experiment

We analyzed the human interference data experiment-wise for both low-cost sensors and SidePak. We explored the differences between baseline data that was collected as a control experiment by having the sensor sitting on a surface with no interference. Experiment data that was collected by originating a human interference situation by the experimenter. Our analysis is demonstrated below.

4 .4.2.1 Measuring Human Interference Using SidePak

As a part of our experiments, we were interested in seeing how SidePak will perceive different human interference situations as opposed to our low-cost $PM_{2.5}$ sensors. Thus, we used SidePak to measure $PM_{2.5}$ for some of our human interference experiments. We kick off this analysis with table XIII showing the significance of the SidePak baseline experiment over the other experiments of which SidePak was used in. We compared means for each experiment (means for an experiment's 1-second time series measurements). The means of SidePak $PM_{2.5}$ data does not follow a normal distribution. Thus, the comparison was done using the Kruskal-Wallis test and Wilxocon rank sum test with Benjamini-Hochberg p-value adjustment for pairwise comparisons as a post hoc. We also report effect size (r) for each comparison.

The significance test shows a significance between the baseline experiment and all other experiments except for the skin with no interference, skin scratching, and, skin touching. For these experiments, the experimenter had the SidePak tube was attached near the low-cost sensor as a wrist band. The reason that the test does not show significance for these experiment in our opinion is that the skin experiments produced a short-time spike in the $PM_{2.5}$ concentration near the sensor. Calculating the mean, the short-term spike effect will be with a small effect. Thus, the test did not show a signif-

Experiment	P-value	Effect Size (r)
bracelet	0.026	-1.055
breath	0.026	-0.906
cough	0.018	-0.963
face spray	0.018	-0.963
hair spray	0.018	-0.892
laugh	0.026	-0.993
powder makeup	0.018	-0.629
skin no interference	ns	-0.690
skin scratch	ns	-0.906
skin sweat	0.043	-0.993
skin touch	ns	-1.321
spray perfume	0.018	-0.892
yawn	0.018	-0.892

TABLE XIII: Significance between Control experiment and the rest of the experiments using SidePak

icance of means comparing to other experiments such as spray experiments where the effect of the interference lasts for a longer while. Other than that, SidePak could identify the human interference in the different experiments.

4 .4.2.2 Measuring Human Interference Using low-cost $PM_{2.5}$ sensors

Similar to the analysis of SidePak data, we conducted an analysis comparing the baseline data measured by the low-cost $PM_{2.5}$ sensors and the experiment's data measured by the same sensors at different times. The analysis was using the Kruskal-Wallis test and Wilxocon rank sum test with Benjamini-Hochberg p-value adjustment for pairwise comparisons as a post hoc. We also report effect size (r) for each comparison. SidePak was used for a partial set of experiments, our prototype was used to collect data for all of the experiments. Table XIV below shows a comparison between the experiment means of our prototype in the baseline experiment and the rest of the experiments.

As the analysis results suggest, data from experiments differ to data from baseline experiment data except for 3 experiments. First, similar to SidePak, the wearable measurements have no significance between baseline and skin touching experiment. Also, the test shows no significance between baseline and skin scratching experiment. The reason behind that, similarly to SidePak data analysis, is an effect of the way $PM_{2.5}$ concentration level responds to the skin experiments in short spikes.

Also, the wearable did not detect any difference for the silk experiment. The silk experiment is a part of the textile experiment group and was not tested using SidePak. In this experiment, the experimenter is wearing a silk garment with the sensor attached to the shirt pocket or clipped to its approximate location and doing only lightweight activities such as reading a book. The reason to this is that the silk, in its nature, has an emission rate that is negligible when using the low-cost $PM_{2.5}$ sensor. Overall, the wearable did detect human interference.

Experiment	P-value	Effect Size (r)
bracelet	0.010	-1.159
breath	0.008	-1.077
cotton	0.008	-1.077
cough	0.008	-1.077
face spray	0.008	-1.077
fur	0.008	-1.077
hair heat style	0.0367	-1.206
hair spray	0.008	-0.997
hair touch	0.0485	-0.746
laugh	0.010	-1.159
leather	0.0167	-1.197
powder makeup	0.008	-1.077
silk	ns	-1.218
skin no interference	0.010	-1.159
skin scratch	0.010	-1.159
skin sweat	ns	-0.276
skin touch	ns	-1.218
spray perfume	0.008	-0.997
wool	0.010	-1.159
yawn	0.008	-0.997

TABLE XIV: Low-cost $PM_{2.5}$ sensor significance of experiments over control experiment

4 .4.2.3 Measuring the effect of Human interference at different distances

As a part of our experiments, we used to replicas of our prototype. One prototype copy served as an experiment sensor that was measuring $PM_{2.5}$ at the created human interference situation. The other copy was used as a reference measuring the $PM_{2.5}$ concentrations away from the interference. For each of the experiments, we repeated the experiment when the distance between the reference and the experimental prototype was 30 cm and when the distance between them was 100 cm. In this section, we present the significance observed between the data read by the experiment sensor and the data read by the reference sensor at each of the two distances for all the experiments. The significance was tested for each distance and for each experiment using Wilxocon rank sum test. Also, we calculated effect size r. We report our the distance-based analysis results in table XV.

In this analysis, we compared the two replicas of our prototype together at different distances. Since one of our sensors is deployed as the experiment measurement tool and the other is used as a reference to obtain a sub-optimal -but with the same level of sub-optimality to the experiment sensor.

The analysis results suggest that for bracelet, skin scratch, and skin sweat both or the sensor readings did not differ for both distances. For breath, skin with no interference, and, yawn experiments there was a difference between reference senor and experiment sensor at 30 cm and no difference when the distance was 100 cm. For the wool, silk, fur experiment, there was no difference at 30 cm, but, there was a significant difference at

Distance	@30 cm		@100 cm	
Experiment	P-value	Effect Size (r)	P-value	Effect Size (r)
bracelet	ns	r = -0.009	ns	r = -0.026
breath	0.020	r = -0.032	ns	r = -0.012
cotton	<0.0001	r = -0.100	<0.0001	r = -0.160
cough	ns	r = -0.058	0.001	r = -0.224
face spray	< 0.0001	r = -0.501	<0.0001	r = -0.397
fur	ns	r = -0.024	<0.0001	r = -0.254
hair heat style	<0.0001	r = -0.532	<0.0001	r = -0.654
hair spray	<0.0001	r = -0.303	<0.0001	r = -0.386
hair touch	<0.0001	r = -0.512	<0.0001	r = -0.076
laugh	0.039	r = -0.072	<0.0001	r = -0.272
leather	<0.0001	r = -0.164	<0.0001	r = -0.433
powder makeup	ns	r = -0.028	<0.0001	r = -0.237
silk	ns	r = -0.030	<0.0001	r = -0.457
skin no interference	<0.0001	r = -0.047	ns	r = -0.009
skin scratch	ns	r = -0.055	ns	r = -0.027
skin sweat	ns	r = -0.023	ns	r = -0.013
skin touch	<0.0001	r = -0.578	<0.0001	r = -0.547
spray perfume	<0.0001	r = -0.344	<0.0001	r = -0.259
synthetic	< 0.0001	r = -0.370	< 0.0001	r = -0.094
wool	ns	r = -0.007	<0.0001	r = -0.187
yawn	<0.0001	r = -0.295	ns	r= -0.097

TABLE XV: Human Interference Effects on Different Distances

100 cm. These all are textile experiments which means that the emissions of these textile can be sensed from a radius of 30 cm. Also, cough and powder experiments can be sensed from a distance of 30 cm. The rest of the experiments were significantly different from reference at 30 cm and at 100 cm.

4 .4.3 Data Analysis per Experiment group

Our experiments fall into more general experiment categories based on experiment similarity. We sorted our experiments to groups as shown in table XVI. We labeled the data by experiment group. Some analysis was done on the group level and experiment-group labeling can be used to improve classification by human interference type.

Group	Experiments				
Control	Control				
Breath	Breath	Cough	Laugh	Yawn	
Textile	Cotton	Fur	Leather	Wool	Synthetic Silk
Skin	Skin no interference	Skin scratch	Skin sweat	Skin touch	Bracelet
Hair	Hair heat style	Hair touch			
Toiletries	Face spray	Hair spray	Powder makeup	Spray perfume	

TABLE XVI: Classifying Experiments per Experiment Group

4 .4.3.1 Significance of Experiment Groups in SidePak

As a part of our analysis for the effect of human interference on the concentrations of $PM_{2.5}$ in the personal cloud, we analyzed the relationship between baseline experiment and the experiment groups of which SidePak was used in. The groups measured using SidePak are breath, skin, and, toiletries. The analysis was done using the group comparison testing of experiment means via Kruskal-Wallis test and Wilxocon rank sum test with Benjamini-Hochberg p-value adjustment for pairwise comparisons as a post hoc. Also, we report the effect size (r). Table XVII shows the results of the group-wise comparison with the baseline experiment in SidePak .

Experiment	P-value	Effect size (r)
Breath	0.008	-0.545
Skin	0.021	-0.493
Toiletries	0.006	-0.541

TABLE XVII: Comparison of Experiment Groups over Control in SidePak

The group comparison with baseline for the data collected by SidePak shows a difference between the data for all of the experiment groups. However, the difference between each group and the baseline varies. This is interpreted as that SidePak is able to detect human interference at the general level.

4 .4.3.2 Significance of Experiment Groups in Low-cost $PM_{2.5}$ Sensors

Similar to the group-based analysis for SidePak data, we analyzed the relationship between baseline experiment and all of the experiment groups' data that was collected using the replicas of our prototype. The analysis was using Kruskal-Wallis test and Wilxocon rank sum test with Benjamini-Hochberg p-value adjustment for pairwise comparisons as a post hoc. Also, we report the effect size (r). The comparison in this analysis is between the baseline control group and all the other groups. The analysis results are shown in table XVIII.

Experiment	P-value	Effect Size (r)
Breath	0.006	-0.567
Hair	ns	-0.396
Skin	0.008	-0.568
Textile	0.005	-0.565
Toiletries	0.004	-0.564

TABLE XVIII: Significance of Experiment Groups over Control in Low-cost $PM_{2.5}$ Sensors

The group-wise comparison of wearable's measurements of the experiments was significant to the baseline for all of the groups except the hair group. Hair experiments were not performed using SidePak. Similarly to skin experiments, hair experiments did affect the concentrations of $PM_{2.5}$ in the surrounding region in a short spike fashion. Thus, comparing the means did not show a difference. For the other experiments as whole, the low-cost $PM_{2.5}$ sensor could detect the existence of the interference.

4 .4.4 Findings

In the process of designing a low-cost personal pollution monitoring tool, we incorporate the use of a $PM_{2.5}$ sensor. But, the readings of the $PM_{2.5}$ sensor are affected by emissions from the human body and clothing. Emissions from the human body differ to that from the exhaust of a diesel engine, thus, recognizing and pointing out $PM_{2.5}$ resulting from human interference is needed to design a wearable with low interference and to further study the difference of the health effect between the different sources of $PM_{2.5}$. With that, we conducted a study to explore different human interference situations and their effect on low-cost $PM_{2.5}$ sensors. In addition to that, we studied some of the human interference situations using high-cost high-accuracy $PM_{2.5}$ monitoring tool: SidePak.

We conducted a set of baseline control experiments to explore human interference. We quantify our human interference experiments into 6 main groups. First, is a control experiment without any human interference as the baseline experiment. Second, breath experiments studying normal breathing, coughing, yawning and laughing. Third, we conducted experiments to study skin emissions by measuring $PM_{2.5}$ concentration near skin with no interference, touching, scratching, and near sweaty skin. Next, we explored hair emissions by measuring $PM_{2.5}$ concentration while touching hair and while heat styling.

Moreover, we studied the change in $PM_{2.5}$ concentrations during the use of toiletries; face spray, hair spray, perfume, and loose-powder makeup. Finally, we studied the emissions of different textile while doing lightweight exercise in normal settings. The textile we explored were cotton, fur, leather, wool, synthetic fabrics, and, silk. In the textile experiments, we were interested in seeing if there is any effect of static electricity on the concentrations of $PM_{2.5}$ near the textiles. We measured static electricity using a multi-meter in fabrics during the textile experiments and we found no static electricity during our experiments.

Our data analysis shows, using both SidePak and the low-cost $PM_{2.5}$ sensor that there is a difference between baseline experiments and a large set of our human interference experiments. Next, results have shown the variability of significance probabilities and their corresponding effect sizes.

For each of the human interference situations, we deployed two copies of the low-cost $PM_{2.5}$ sensor. One copy measured the human interference while the other was used as a reference to measure the change that happened due to the human interference situation. We repeated experiments with the reference sensor sitting at two different distances from the human interference. The distances were 30 cm and 1 m. Our analysis has shown that low-cost $PM_{2.5}$ sensor could capture changes of $PM_{2.5}$ concentrations due to different human interference situations. Finally, the analysis has shown different effects (and no effect for some cases) of human interference on the reference sensor at different distances.

Using our data as a piece of basic evidence for design decisions, the data shows that the toiletries emissions have the highest effect on the wearable and SidePak. Thus, the designer should minimize the possibility of having the sensor interact with toiletries. Next, breath emissions then skin emissions where the highest, thus, the designer should be careful when designing for the head area and for when being around exposed skin. Also, textile had an effect on the $PM_{2.5}$ concentrations and it is, for most cases, reaching a 30 cm distance. A good suggestion is to wear the sensor on the sleeve.

Further explorations of other human interference situations such as belongings or shoes might be needed for a more evident inferring of a where to wear design decision. Moreover, for our future work, we are planning on collecting more data to build a model to detect and classify human interference. This model can help in understanding the health effects of different types of $PM_{2.5}$ generated by different sources in a more exploratory approach.

CHAPTER 5

CONCLUSION

Epidemiological studies suggest that prolonged exposure to pollutants have adversarial health threats. Some studies have shown a positive association between exposure to $PM_{2.5}$ and noise and mild cognitive impairment as a precursor to dementia.

Recent works in ubiquitous computing for personal monitoring focus more on tracking eating, exercise, and sleep habits and feelings. Little work is being done towards monitoring personal pollution exposure, particularly for the at-risk communities. Most of the available pollution monitoring tools are designed for scientists and quantified-selfers or they are designed to be at a fixed location, e.g home air quality meters to be fixed in the living room. In this thesis, we presented our effort towards a more user-oriented personal pollution exposure monitoring, especially for the at-risk communities.

We created a low-cost $PM_{2.5}$ and noise monitoring prototype using low-cost Plantower PMS7003 $PM_{2.5}$ sensor and facilitating the at-hand microphone of the smartphone to serve as a noise monitoring tool at hand. We studied the low-cost $PM_{2.5}$ sensor and we tackled the low-accuracy problem by creating a fair calibration model using naïve Bayes classification.

The emissions from the human body and clothing affect the concentrations of $PM_{2.5}$ in the personal cloud of the person at hand. We explored different situations of human interference and their effects on low-cost $PM_{2.5}$ sensors as well as high-cost high-precision sensors. Our studies have shown significant changes due to some of the human interference situations we studied. With our current results we recommend, for designing a sensor with little human interference, to keep a distance from bare skin areas and the human head.

Contributions to HCI. This thesis contributes to Human-Computer Interaction (HCI) by exploring the effect of human physical interaction with low-cost $PM_{2.5}$ sensors. We identify a set of somatic human interference situations and textile emissions. We studied how these situations can affect the sensors as design guidelines for personal pollution monitoring wearables. This knowledge lead the design decisions for personal pollution exposure monitoring wearables. Our framework can be used to study different types of $PM_{2.5}$ sensors (for example sensors that don't use laser scattering). Also, our framework can guide studies to explore sensors that measure other types of pollutants.

Future Work. For future work, we are planning to collect more data in wider set of contexts to improve the sensor calibration model. Also, we will fabricate a smaller circuit board to replace the Raspberry Pi board in order to reduce the size and the weight of the wearable towards better wearablitiy and improved use experience. The wearable fabrication will be using human-in-the-loop design approach. For human interference, we are planning to improve our design guidelines be experimenting for more where-to-wear candidates as an effort towards offering a variety of choices for designing low-cost personal pollution monitoring wearables. Finally, we want to create a human interference detection and classifying model to distinguish different human interference situation and make an account for them in order to help understand the effect of different types of particulate matter on the human health, particularly, the health of the at-risk.

CITED LITERATURE

1. Air quality index (aqi) basics. https://airnow.gov/index.cfm?action=aqibasics.aqi.
2. Aqi breakpoints. https://aqs.epa.gov/aqsweb/documents/codetables/aqi_breakpoints.html.
3. Muhlfeld, C., Rothen-Rutishauser, B., Blank, F., Vanhecke, D., Ochs, M., and Gehr, P.: Interactions of nanoparticles with pulmonary structures and cellular responses. American Journal of Physiology-Lung Cellular and Molecular Physiology, 2008.
4. Power, M. C., Adar, S. D., Yanosky, J. D., and Weuve, J.: Exposure to air pollution as a potential contributor to cognitive function, cognitive decline, brain imaging, and dementia: a systematic review of epidemiologic research. Neurotoxicology, 56:235–253, 2016.
5. Tzivian, L., Dlugaj, M., Winkler, A., Weinmayr, G., Hennig, F., and Fuks, K. B. ... & J"ockel, K. H.: Long-term air pollution and traffic noise exposures and mild cognitive impairment in older adults: a cross-sectional analysis of the Heinz Nixdorf recall study. Environmental health perspectives, 2016.
6. Weuve, J.: Invited commentary: how exposure to air pollution may shape dementia risk, and what epidemiology can say about it. American journal of epidemiology, 180(4):367–371, 2014.
7. Petersen, R. C., Smith, G. E., Waring, S. C., Ivnik, R. J., Tangalos, E. G., and Kokmen, E.: Mild cognitive impairment: clinical characterization and outcome. Archives of neurology, 56(3):303–308, 1999.
8. Peterson, R. C.: Mild cognitive impairment as a diagnostic entity. J Intern Med, 256(3):183–94, 2004.
9. Prince, M., Bryce, R., Albanese, E., Wimo, A., Ribeiro, W., and Ferri, C. P.: The global prevalence of dementia: a systematic review and metaanalysis. Alzheimer's & Dementia, 9(1):63–75, 2013.
10. Air data: Air quality data collected at outdoor monitors across the us. https://www.epa.gov/outdoor-air-quality-data, Oct 2018.
11. Catlett, C. E., Beckman, P. H., Sankaran, R., and Galvin, K. K.: Array of things: a scientific research instrument in the public way: platform design and early lessons learned. In Proceedings of the 2nd International Workshop on Science of Smart City Operations and Platforms Engineering, pages 26–33. ACM, 2017.
12. Özkaynak, H., Baxter, L. K., Dionisio, K. L., and Burke, J.: Air pollution exposure prediction approaches used in air pollution epidemiology studies. Journal of Exposure Science and Environmental Epidemiology, 23(6):566, 2013.

13. Maag, B., Zhou, Z., and Thiele, L.: W-Air: Enabling Personal Air Pollution Monitoring on Wearables. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2(1):24, 2018.

14. Setton, E., Marshall, J. D., Brauer, M., Lundquist, K. R., Hystad, P., Keller, P., and Cloutier-Fisher, D.: The impact of daily mobility on exposure to traffic-related air pollution and health effect estimates. Journal of Exposure Science and Environmental Epidemiology, 21(1):42, 2011.

15. Steinle, S., Reis, S., and Sabel, C. E.: Quantifying human exposure to air pollutionâĂŤMoving from static monitoring to spatio-temporally resolved personal exposure assessment. Science of the Total Environment, 443:184–193, 2013.

16. Steinle, S., Reis, S., Sabel, C. E., Semple, S., Twigg, M. M., Braban, C. F., Leeson, S. R., Heal, M. R., Harrison, D., Lin, C., and others: Personal exposure monitoring of PM2. 5 in indoor and outdoor microenvironments. Science of the Total Environment, 508:383–394, 2015.

17. Popescu, F. and Ionel, I.: Anthropogenic air pollution sources. INTECH Open Access Publisher, 2010.

18. Mazzei, F., D'alessandro, A., Lucarelli, F., Nava, S., Prati, P., Valli, G., and Vecchi, R.: Characterization of particulate matter sources in an urban environment. Science of the Total Environment, 401(1-3):81–89, 2008.

19. He, C., Morawska, L., Hitchins, J., and Gilbert, D.: Contribution from indoor sources to particle number and mass concentrations in residential houses. Atmospheric environment, 38(21):3405–3415, 2004.

20. Hussein, T., Glytsos, T., Ondráček, J., Dohányosová, P., Ždímal, V., Hämeri, K., Lazaridis, M., Smolík, J., and Kulmala, M.: Particle size characterization and emission rates during indoor activities in a house. Atmospheric Environment, 40(23):4285–4307, 2006.

21. Hussein, T., Korhonen, H., Herrmann, E., Hämeri, K., Lehtinen, K. E., and Kulmala, M.: Emission rates due to indoor activities: indoor aerosol model development, evaluation, and applications. Aerosol Science and Technology, 39(11):1111–1127, 2005.

22. Huboyo, H. S., Tohno, S., and Cao, R.: Indoor PM2. 5 characteristics and CO concentration related to water-based and oil-based cooking emissions using a gas stove. Aerosol and Air Quality Research, 11(4):401–411, 2011.

23. Fan, C.-W. and Zhang, J. J.: Characterization of emissions from portable household combustion devices: particle size distributions, emission rates and factors, and potential exposures. Atmospheric environment, 35(7):1281–1290, 2001.

24. Spengler, J., Dockery, D., Turner, W., Wolfson, J., and Ferris Jr, B.: Long-term measurements of respirable sulfates and particles inside and outside homes. Atmospheric Environment (1967), 15(1):23–30, 1981.

25. Wang, B., Ho, S. S. H., Ho, K. F., Huang, Y., Chan, C. S., Feng, N. S. Y., Ip, S. H. S., and others: An environmental chamber study of the characteristics of air pollutants released from environmental tobacco smoke. Aerosol Air Qual. Res, 12(6):1269–1281, 2012.

26. Afshari, A., Matson, U., and Ekberg, L.: Characterization of indoor sources of fine and ultrafine particles: a study conducted in a full-scale chamber Indoor Air 15: 141-150. Find this article online, 2005.

27. Xu, C., Wu, C.-Y., and Yao, M.: Fluorescent bioaerosol particles resulting from human occupancy with and without respirators. Aerosol Air Qual. Res, 17:198–208, 2017.

28. Wallace, L. A., Emmerich, S. J., and Howard-Reed, C.: Effect of central fans and in-duct filters on deposition rates of ultrafine and fine particles in an occupied townhouse. Atmospheric Environment, 38(3):405–413, 2004.

29. Wilson, W. E., Mage, D. T., and Grant, L. D.: Estimating separately personal exposure to ambient and nonambient particulate matter for epidemiology and risk assessment: why and how. Journal of the Air & Waste Management Association, 50(7):1167–1183, 2000.

30. You, R., Cui, W., Chen, C., and Zhao, B.: Measuring the short-term emission rates of particles in the "personal cloud" with ifferent clothes and activity intensities in a sealed chamber. Aerosol and Air Quality Research, 13(3):911–921, 2013.

31. Chen, C. and Zhao, B.: Review of relationship between indoor and outdoor particles: I/O ratio, infiltration factor and penetration factor. Atmospheric Environment, 45(2):275–288, 2011.

32. Klepeis, N. E., Nelson, W. C., Ott, W. R., Robinson, J. P., Tsang, A. M., Switzer, P., Behar, J. V., Hern, S. C., and Engelmann, W. H.: The national human activity pattern survey (nhaps): a resource for assessing exposure to environmental pollutants. Journal of Exposure Science and Environmental Epidemiology, 11(3):231, 2001.

33. Kelly, F. J. and Fussell, J. C.: Size, source and chemical composition as determinants of toxicity attributable to ambient particulate matter. Atmospheric environment, 60:504–526, 2012.

34. Grahame, T. J. and Schlesinger, R. B.: Health effects of airborne particulate matter: do we know enough to consider regulating specific particle types or sources? Inhalation toxicology, 19(6-7):457–481, 2007.

35. De Kok, T. M., Driece, H. A., Hogervorst, J. G., and Briedé, J. J.: Toxicological assessment of ambient and traffic-related particulate matter: a review of recent studies. Mutation Research/Reviews in Mutation Research, 613(2-3):103–122, 2006.

36. McAlexander, T. P., Gershon, R. R., and Neitzel, R. L.: Street-level noise in an urban setting: assessment and contribution to personal exposure. Environmental Health, 14(1):18, 2015.

37. De Nazelle, A., Seto, E., Donaire-Gonzalez, D., Mendez, M., Matamala, J., Nieuwenhuijsen, M. J., and Jerrett, M.: Improving estimates of air pollution exposure through ubiquitous sensing technologies. Environmental Pollution, 176:92–99, 2013.

38. Lamkin, P.: Smart wearables market to double by 2022: $27 billion industry forecast. https://www.forbes.com/sites/paullamkin/2018/10/23/smart-wearables-market-to-double-by-2022-27-billion-industry-forecast/158addd12656, Oct 2018.

39. Choe, E. K., Lee, N. B., Lee, B., Pratt, W., and Kientz, J. A.: Understanding quantified-selfers' practices in collecting and exploring personal data. In Proceedings of the 32nd annual ACM conference on Human factors in computing systems, pages 1143–1152. ACM, 2014.

40. Kim, S. and Paulos, E.: InAir: sharing indoor air quality measurements and visualizations. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pages 1861–1870. ACM, 2010.

41. Budde, M., El Masri, R., Riedel, T., and Beigl, M.: Enabling low-cost particulate matter measurement for participatory sensing scenarios. In Proceedings of the 12th international conference on mobile and ubiquitous multimedia, page 19. ACM, 2013.

42. Zhuang, Y., Lin, F., Yoo, E.-H., and Xu, W.: Airsense: A portable context-sensing device for personal air quality monitoring. In Proceedings of the 2015 Workshop on Pervasive Wireless Healthcare, pages 17–22. ACM, 2015.

43. Tian, R., Dierk, C., Myers, C., and Paulos, E.: Mypart: Personal, portable, accurate, airborne particle counting. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, pages 1338–1348. ACM, 2016.

44. Jiang, Y., Li, K., Tian, L., Piedrahita, R., Yun, X., Mansata, O., Lv, Q., Dick, R. P., Hannigan, M., and Shang, L.: Maqs: a personalized mobile sensing system for indoor air quality monitoring. In Proceedings of the 13th international conference on Ubiquitous computing, pages 271–280. ACM, 2011.

45. Nikzad, N., Verma, N., Ziftci, C., Bales, E., Quick, N., Zappi, P., Patrick, K., Dasgupta, S., Krueger, I., Rosing, T. Š., et al.: Citisense: improving geospatial environmental assessment of air quality using a wireless personal exposure monitoring system. In Proceedings of the conference on Wireless Health, page 11. ACM, 2012.

46. Dutta, P., Aoki, P. M., Kumar, N., Mainwaring, A., Myers, C., Willett, W., and Woodruff, A.: Common sense: participatory urban sensing using a network of handheld air quality monitors. In Proceedings of the 7th ACM conference on embedded networked sensor systems, pages 349–350. ACM, 2009.

47. Oletic, D. and Bilas, V.: Design of sensor node for air quality crowdsensing. In 2015 IEEE Sensors Applications Symposium (SAS), pages 1–5. IEEE, 2015.

48. Piedrahita, R., Xiang, Y., Masson, N., Ortega, J., Collier, A., Jiang, Y., Li, K., Dick, R. P., Lv, Q., Hannigan, M., et al.: The next generation of low-cost personal air quality

sensors for quantitative exposure monitoring. Atmospheric Measurement Techniques, 7(10):3325–3336, 2014.

49. Ambiciti - environmental awareness for all. http://ambiciti.io/.

50. Plinge, A., Jacob, F., Haeb-Umbach, R., and Fink, G. A.: Acoustic microphone geometry calibration: An overview and experimental evaluation of state-of-the-art algorithms. IEEE Signal Processing Magazine, 33(4):14–29, 2016.

51. Zhu, Y., Li, J., Liu, L., and Tham, C.-K.: ical: Intervention-free calibration for measuring noise with smartphones. In 2015 IEEE 21st International Conference on Parallel and Distributed Systems (ICPADS), pages 85–91. IEEE, 2015.

52. Nemirovski, G. G.: Sensor pair for detecting changes within a human ear and producing a signal corresponding to thought, movement, biological function and/or speech, November 11 2003. US Patent 6,647,368.

53. Stork, J. A., Spinello, L., Silva, J., and Arras, K. O.: Audio-based human activity recognition using non-markovian ensemble voting. In 2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication, pages 509–514. IEEE, 2012.

54. Brdiczka, O., Langet, M., Maisonnasse, J., and Crowley, J. L.: Detecting human behavior models from multimodal observation in a smart home. IEEE Transactions on automation science and engineering, 6(4):588–597, 2009.

55. Birchfield, S. T. and Gillmor, D. K.: Fast bayesian acoustic localization. In 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing, volume 2, pages II–1793. IEEE, 2002.

56. Birchfield, S. T. and Gangishetty, R.: Acoustic localization by interaural level difference. In Proceedings.(ICASSP'05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005., volume 4, pages iv–1109. IEEE, 2005.

57. Erol-Kantarci, M., Mouftah, H. T., and Oktug, S.: A survey of architectures and localization techniques for underwater acoustic sensor networks. IEEE Communications Surveys & Tutorials, 13(3):487–502, 2011.

58. Haddad, K., Hald, J.-r., et al.: 3d localization of acoustic sources with a spherical array. Journal of the Acoustical Society of America, 123(5):3311, 2008.

59. Sekmen, A. S., Wilkes, M., and Kawamura, K.: An application of passive human-robot interaction: human tracking based on attention distraction. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 32(2):248–259, 2002.

60. Geertsma, M.: New map shows chicago needs environmental justice reforms. https://www.nrdc.org/experts/meleah-geertsma/new-map-shows-chicago-needs-environmental-justice-reforms, Oct 2018.

61. Page, T.: A forecast of the adoption of wearable technology. International Journal of Technology Diffusion (IJTD), 6(2):12–29, 2015.

62. Amazon web services (aws) - cloud computing services. https://aws.amazon.com/.

63. Sivasubramanian, S.: Amazon dynamodb: a seamlessly scalable non-relational database service. In Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data, pages 729–730. ACM, 2012.

64. Maksimović, M., Vujović, V., Davidović, N., Milošević, V., and Perišić, B.: Raspberry pi as internet of things hardware: performances and constraints. design issues, 3(8), 2014.

65. Windows 10 internet of things. https://developer.microsoft.com/en-us/windows/iot.

66. Raspberry pi 3 model b. https://www.raspberrypi.org/products/raspberry-pi-3-model-b-plus/.

67. Gemperle, F., Kasabach, C., Stivoric, J., Bauer, M., and Martin, R.: Design for wearability. In digest of papers. Second international symposium on wearable computers (cat. No. 98EX215), pages 116–122. IEEE, 1998.

68. plantower pms7003 datasheet. https://download.kamami.com/p564008-p564008-PMS7003 series data manua_English_V2.5.pdf.

69. Industries, A.: Usb battery pack for raspberry pi - 4000mah - 5v @ 1a. https://www.adafruit.com/product/1565.

70. Sakhnini, N., Yu, J. E., and Chattopadhyay, D.: myCityMeter: Helping Older Adults Manage the Environmental Risk Factors for Cognitive Impairment. In Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers, pages 235–238. ACM, 2018.

71. Chau, N. H.: A new approach to estimate urban air temperature using smartphones. In Asian Conference on Intelligent Information and Database Systems, pages 633–641. Springer, 2018.

72. Gruber, T.: What is an Ontology. WWW Site http://www-ksl. stanford. edu/kst/whatis-an-ontology. html (accessed on 07-09-2004), 1993.

73. Botts, M., Percivall, G., Reed, C., and Davidson, J.: OGC® sensor web enablement: Overview and high level architecture. In GeoSensor networks, pages 175–190. Springer, 2008.

74. Sheth, A., Henson, C., and Sahoo, S. S.: Semantic sensor web. IEEE Internet computing, 12(4), 2008.

75. Hobbs, J. R. and Pan, F.: Time ontology in OWL. W3C working draft, 27:133, 2006.

76. Compton, M., Barnaghi, P., Bermudez, L., GarcíA-Castro, R., Corcho, O., Cox, S., Graybeal, J., Hauswirth, M., Henson, C., Herzog, A., and others: The SSN ontology of the W3c semantic sensor network incubator group. Web semantics: science, services and agents on the World Wide Web, 17:25–32, 2012.

77. Adeleke, J. A. and Moodley, D.: An ontology for proactive indoor environmental quality monitoring and control. In Proceedings of the 2015 Annual Research Conference on South African Institute of Computer Scientists and Information Technologists, page 2. ACM, 2015.

78. W3C: WGS84 Geo Positioning: an RDF vocabulary. https://www.w3.org/2003/01/geo/wgs84_pos, 2009.

79. SEAS: SEAS-WeatherOntology ontology. https://ci.mines-stetienne.fr/seas/WeatherOntology, 2017.

80. OWLCS: OWL API main repository, 2018. Published: https://github.com/owlcs/owlapi/wiki/Documentation.

81. Semple, S., Apsley, A., Galea, K. S., MacCalman, L., Friel, B., and Snelgrove, V.: Secondhand smoke in cars: assessing children's potential exposure during typical journey conditions. Tobacco control, 21(6):578–583, 2012.

82. Rees, V. W. and Connolly, G. N.: Measuring air quality to protect children from secondhand smoke in cars. American journal of preventive medicine, 31(5):363–368, 2006.

83. Soule, E. K., Maloney, S. F., Spindle, T. R., Rudy, A. K., Hiler, M. M., and Cobb, C. O.: Electronic cigarette use and indoor air quality in a natural setting. Tobacco control, 26(1):109–112, 2017.

84. Cheng, K.-C., Zheng, D., Tetteh, A. O., Park, H.-K., Nadeau, K. C., and Hildemann, L. M.: Personal exposure to airborne particulate matter due to residential dryer lint cleaning. Building and Environment, 98:145–149, 2016.

85. Semple, S., Ibrahim, A. E., Apsley, A., Steiner, M., and Turner, S.: Using a new, low-cost air quality sensor to quantify second-hand smoke (shs) levels in homes. Tobacco Control, 24(2):153–158, 2015.

86. Fiala, S. C., Morris, D. S., and Pawlak, R. L.: Measuring indoor air quality of hookah lounges. American journal of public health, 102(11):2043–2045, 2012.

87. of Environmental Quality, O.: Impact of noise on people. 1977.

88. TSI: Sidepak personal aerosol monitor model am510. https://www.tsi.com/getmedia/51f3ccb6-780e-4386-b8fb-60d688d37a18/SidePak_AIM510_US_19804 web?ext=.pdf.

89. TSI: Sidepak personal aerosol monitor model am510. https://www.tsi.com/getmedia/03b9b3e3-c1e0-4217-ad6f-4188940cda40/SidePakAM510_2980194-USA-web?ext=.pdf.

90. Keywood, M., Ayers, G., Gras, J., Gillett, R., and Cohen, D.: Relationships between size segregated mass concentration data and ultrafine particle number concentrations in urban areas. Atmospheric Environment, 33(18):2907–2913, 1999.

91. Zimmerman, N., Presto, A. A., Kumar, S. P., Gu, J., Hauryliuk, A., Robinson, E. S., Robinson, A. L., and Subramanian, R.: A machine learning calibration model using random forests to improve sensor performance for lower-cost air quality monitoring. Atmospheric Measurement Techniques, 11(1), 2018.

92. Ott, W. R.: Concepts of human exposure to air pollution. Environment International, 7(3):179–196, 1982.

93. Kioumourtzoglou, M.-A., Schwartz, J. D., Weisskopf, M. G., Melly, S. J., Wang, Y., Dominici, F., and Zanobetti, A.: Long-term pm2. 5 exposure and neurological hospital admissions in the northeastern united states. Environmental health perspectives, 124(1):23–29, 2015.

94. Hosford-Dunn, H., Roeser, R. J., and Valente, M.: Audiology: diagnosis. Thieme, 2007.

95. A-weighting. https://www.nti-audio.com/en/support/know-how/frequency-weightings-for-sound-level-measurements.

96. Martin, W.: Decibel-the name for the transmission unit. Journal of the AIEE, 48(3):223–223, 1929.

97. Martin, W.: The transmission unit and telephone transmission reference systems. Transactions of the American Institute of Electrical Engineers, 43:797–801, 1924.

VITA

Nina Sakhnini

Education

M.S. Candidate, Computer Science 2019
Computer Science Department, College of Engineering
University of Illinois at Chicago, Chicago, Illinois, USA
Major GPA: 3.50 / 4.00

B.S., Computer Engineering 2016
Department of Computer Engineering, College of Computer and Information Technology
Jordan University of Science & Technology, Irbid, Jordan
GPA: 80.5%

Experience

Teaching Assistant Spring, 2019
Department of Computer Science, College of Engineering, University of Illinois at Chicago, Chicago, IL , USA

Teaching Assistant Fall, 2017
Department of Computer Science, College of Engineering, University of Illinois at Chicago, Chicago, IL , USA

Research Intern Summer, 2015
Texas Tech University 2015 NSF Research Experiences for Undergraduates Site Program