IoT Environmental Analyzer using Sensors and Machine Learning for Migraine Occurrence Prevention

Rosemarie J. Day
Data Science Program
Worcester Polytechnic Institute
Worcester, MA, U.S.A
rday@wpi.edu

Hassan S. Salehi
Department of Electrical and Computer Engineering
California State University, Chico
Chico, CA, U.S.A
hsalehi@csuchico.edu

Mahsa H. Javadi
Division of Genetic Research
University of Connecticut Health Center
Farmington, CT, U.S.A
mjavadi@uconn.edu

***Abstract*— Everyday migraines are affecting more than one billion people worldwide. This headache disorder is classified as the sixth most disabling disease in the world. Migraines are just one chronic illness affected by environmental triggers due to changes that occur inside the home. Migraines share this characteristic with sinus headaches and thus are often misdiagnosed. In this research work, an iOS-based environmental analyzer was designed, implemented and evaluated for migraine sufferers with the use of sensors. After the data collection and cleaning, five machine learning model were used to estimate prediction accuracy of migraines in terms of the environment. The data was evaluated against the models using K-Fold cross validation. The algorithm accuracy comparison showed that Linear Discriminant Analysis (LDA) produced highest accuracy for the testing data at a mean of 0.938. Preliminary results demonstrate the feasibility of using machine learning algorithms to perform the automated recognition of migraine trigger areas in the environment.**

***Index Terms*— Data Acquisition, Data Analysis, Machine Learning, Internet of Things (IoT), Mobile Applications, Smart Devices, Smart Homes**

I. INTRODUCTION

MIGRAINES are often misdiagnosed as sinus headaches due to their similarities in being affected by environmental triggers. Often thought of as a neurological disorder affected by the chemicals in the brain and nerve pathways, migraines are commonly affected by environmental changes inside and outside the home, which include pressure, temperature, humidity, noise and light [1-3]. While researching migraines, it was also noted that there is a distinction between an individual who suffers from chronic migraines versus one who suffers from chronic headaches that is not dependent on pain level [4]. It has been stated that "weather changes can cause biological changes in the body's chemical balance and thus precipitate a migraine headache in some sensitive people. Weather conditions also can increase the severity of a headache induced by other factors." [5]
Knowing this, an application can be developed to track and analyze changes in migraines in response to the home environment changes. Many migraine sufferers relate changes in their environment to migraine triggers such as weather changes, altitude and more [6]. With this, it can be said that indoor environments contribute to migraine attacks, which result in individuals requesting time off from work or special accommodations. One researcher, Friedman, looked at two methods of reporting migraines affected by environmental changes, (1) self-reporting by the individual, and (2) using diary information from the individual. Of the 38 patients with migraines and 17 with tension-type headaches, weather was the most frequent factor used to distinguish migraines from tension headaches. 71% of migraine sufferers reported being affected by the changes in weather conditions while only 35% of tension headache sufferers noted this. The conclusion of this research showed that many environmental factors are often reported to trigger migraines and other headaches but those who suffer from migraines are reported to be more sensitive to environmental changes than others [6].

Often, migraines are caused by other triggers and then worsened by environmental changes that can be avoidable in the home. Two common environmental conditions that are associated with migraines are dry / dusty environments and extremely cold conditions. With this, migraines can be further provoked when an individual enters a home whose environmental conditions are not ideal, causing this individual to adjust and adapt to the surrounding conditions [7].

Another common occurrence with migraines is to see an aura before or after an attack. Aura's are visual, sensory, motor or verbal disturbances to a person [7]. Typically occurring before an attack, auras are one sign that an individual is about to suffer from a migraine.

Despite many migraine sufferers reporting issues with environmental triggers being linked to migraine attacks, very few controlled studies have been performed to explain this link. Studies that have been done have shown a conflicting conclusion on the link to these triggers [6]. Each person reacts differently to environmental triggers, having their own threshold for the stimuli, such as amount of light or humidity changes.

This work has been partially supported by iDevices LLC.

978-1-7281-4550-1/19/$31.00 ©2019 IEEE
DOI 10.1109/ICMLA.2019.00239

The research objective of this study is to examine triggers for migraines affected by environmental changes to develop an environmental analyzer that can take in a user's migraine data and examine it in terms of the environment. Data is collected and fed back into an app that will track changes in a user's migraines as well as the home environment. Other apps for migraines have been designed to provide on-demand access to instructions to guide users into how to manage migraines and track them, monitoring key migraine and medication variables [8].

The application developed during this research gives users the ability to understand trigger areas in their home environment by analyzing how frequently migraines occur, the severity of migraines, and patterns that begin to emerge from their migraine attacks or in their environment. By implementing an application in this manner, users can then develop a migraine profile in which will help explain the environmental triggers link to their migraines. This application, currently developed for migraines, can later be expanded to include other environmental related chronic illnesses such as joint pain and asthma.

This paper is an extended version of work published in [1] and [2]. The work was extended from previous work with the addition of the analytics process and a brief study of migraines and the environment through descriptive statistics and machine learning. The rest of the paper is organized as follows: Section II provides the approach for realization into the environmental analysis of the home with a focus on how data moves from the hardware through the software. Section III overviews a brief study done on the application and the results that were obtained for two locations, as well as, introduces and discusses machine learning techniques used on the data. This section provides examples of the type of data that can be obtained from this application.

II. Home Environmental Analytics Process

This section discusses the design of the IoT devices, data collection, and application development.

A. Data Collection and Sensor Selection

The design of the environmental analyzer links the Particle Photon firmware to the iOS application using Amazon Web Services to clean and store data. Particle is an open-source platform for scalable, reliable and secure Internet of Things (IoT) devices. They have repositories available through GitHub for both hardware and software.

AWS is utilized in the backend of the system to link the application programming interface (API) gateway to a Lambda function that fills the home environmental data into the DynamoDB database. The Lambda function is used in the backend to read a JSON payload from the Particle Photon cloud that is delivered by the API gateway. This data is then stored into the DynamoDB database by deviceID which identifies which device the data belongs to. The Lambda function from AWS is used as a mobile backend to retrieve and transfer data to the DynamoDB database.

Utilizing this structure, the Particle Photon's firmware was designed to read data from sensors connected to the board. This data is then published to the Particle cloud and can be accessed by AWS through a web hook. The web hook allows for data to be sent from one device or service to another. Once the data has been collected, the iOS application can begin analyzing migraines and the home environment.

During this research, the main sensor utilized was the BME280 sensor (Bosch-Sensortec [9-11]). This integrated sensor can read temperature, humidity, pressure and altitude inside the home environment. With a high linearity and high sensitivity, the BME280 is designed for low current consumption. The pressure sensor in the BME280 has a high accuracy and very low noise while the humidity sensor has a fast response [10]. This sensor operates within a temperature range of –40 to 85 degrees Celsius with a ±1.0°C accuracy. With that, there is a ±3% accuracy for the humidity sensor, a ±1 hPa absolute accuracy for the pressure sensor, and a ±1-meter accuracy in measuring the current altitude.

The BME280 is a sensor that can be utilized using either serial peripheral interface (SPI) or inter-integrated circuit (I2C) buses on a microcontroller. For this research, the sensor utilizes the I2C bus, which requires four connections [11]. The first two connections provide power and ground to the sensor with a voltage supply on the Particle Photon of 3.3V. Utilized with two 4.7K pull up resistors, this voltage input works well with the BME280 sensor. With that, the BME280 also requires connection for SDA and SCL. SDA is the sensors connection bi-directional serial data to and from the microcontroller while the SCL connection is the sensors connection to the serial clock on the microcontroller.

Utilizing the I2C bus, the firmware was programmed with functions to read the desired sensor and obtain characteristics the function is looking for. After the sensor data has been grabbed, the values are stored into a variable that can then be used to send data up to the Particle cloud in string format. This is where the web hook, mentioned previously, connects the Particle cloud to the AWS API. Through the firmware, the data is sent to the Particle cloud at a specified rate through a delay in the main loop. The original delay was to update the Particle cloud data every 5 seconds but this was later changed to update the cloud every hour with new home environmental data. Now that this data has been sent to the Particle cloud it can be retrieved and stored in an AWS database.

Through the connection of the Particle Photon cloud and the AWS API gateway, the finalized system architecture for the environmental analyzer can be broken up into frontend and backend components. The front end of the application is made up of the Particle Photon device connected to sensors and the iOS application programmed to analyze the collected data. The backend of the system is comprised of the Particle cloud and the AWS cloud and API.

B. Cleaning and Normalizing the Data

Taking the sensor data gathered from the BME280, the data collected in the DynamoDB and iOS application can be cleaned and normalized for analysis. When first looking at the acquired data, as seen in Table 1-A, the data only contains the sensor data collected in the DynamoDB database. This data collected contains unneeded or duplicate fields from the collection process that can be removed. For analysis, the DeviceID column is not needed. This DeviceID is used in the code to separate users from one another. This table in DynamoDB only refers to one user. After removing this column, light can also be removed as it was not used in this study. The light column did not present useful data when collected, as can be seen Thein Table 1-A, the column reads 0. Lastly, publishedAt can be removed as it refers to the epoch time. This column is used to place items in order and give them a unique ID in the table but is not needed for analysis as time is already changed into a human readable format when the Lambda code transfers the data from the Particle cloud to the AWS database.

$$z_i = \frac{x_i - \min(x)}{max(x) - \min(x)} \tag{1}$$

where $x = (x_1, \ldots, x_n)$ and z_i is now your i^{th} normalized data

Cleaning this data and combining it with the migraine data collected in the iOS application will yield the result seen in Table 1-B. This table shows the remaining fields and how each day relates to a user having a migraine (1) or not having a migraine (0). After cleaning and combining the data into one data set for the user, the data can be normalized to one using equation 1, and will resemble the data seen in Table 1-C.

C. Numerical and Graphical Representation of the Data

After collection, cleaning and normalizing the data it was analyzed and displayed in an iOS application, designed to present data in both numerical and graphical form for the user. The user is given three different screen options that provide data for migraines and different environmental variables. For users, one screen in the application is labeled migraines and is where the user could see a list of all migraines that occurred along with data logged for each migraine.

A graphical representation is provided to users to allow them to see how drastically or suddenly their home environment is changing over time. These changes may not be apparent to the user or noticeable beforehand. This allows users a quick glance into the home environment without needed to focus on reading and interpreting numbers. With this, data is also presented in numerical form to the users, showing the mean, variance, standard deviation, and coefficient of variance for the home environment.

III. A BRIEF STUDY ON MIGRAINES AND THE ENVIRONMENT

The finalized application with the hardware implementation was tested over a nine-day span. Utilizing one individual as a test subject, room data was collected in three different homes, two located in Torrington, CT and one located in West Hartford, CT. Utilizing this room data, five migraines were recorded in this span of time. The purpose of using three different homes was to show the changes in altitude with migraines.

With this, the study is broken down into three main sections: descriptive statistics, Matlab visualizations, and machine learning methods.

A. Descriptive Statistics

After completing this nine-day trial, the results were analyzed and plotted to show how room variables changed as migraines occurred. First looking at the data for pressure and altitude, the altitude varied greatly near the last few migraines of the nine-day trial while pressure remained very close to the same value until the end. Looking at temperature and humidity compared to migraines, the temperature had slight variations over the week but the humidity varied drastically during the days of migraines occurring.

With this, let's examine the data collected and analyzed by the application through mean, standard deviation, and coefficient of variance. Standard deviation (2) is used to quantify the amount of variation on a set of data points which can be interpreted as such, a low standard deviation has data points that tend to be close to the mean while a higher standard deviation indicates a wider spread of points. Using the standard deviation and mean, the coefficient of variance can be calculated as the ratio of the standard deviation to the mean. Through this knowledge, the environmental changes of the

TABLE 1-A
COLLECTED SENSOR DATA

DeviceID	publishedAt	Altitude (km)	Humidity (%)	Light (%)	Pressure (hPA)	publishtedAtString	Temperature (°C)
220049	1491274623067	2.309227	28.995117	0	1012.972656	Monday, April 03, 2017 02:57:03	28.830000
220049	1491529916864	541.190796	80.121094	0	949.913452	Friday, April 07, 2017 01:51:56	24.420000

TABLE 1-B
COLLECTED DATA CLEANED AND COMBINED WITH MIGRAINE DATA

publishtedAtString	Altitude (km)	Humidity (%)	Pressure (hPA)	Temperature (°C)	Migraine
Monday, April 03, 2017 02:57:03	2.309227	28.995117	1012.972656	28.830000	1
Friday, April 07, 2017 01:51:56	541.190796	80.121094	949.913452	24.420000	0

TABLE 1-C
NORMALIZED COLLECTED DATA

publishtedAtString	Altitude (km)	Humidity (%)	Pressure (hPA)	Temperature (°C)	Migraine
Monday, April 03, 2017 02:57:03	0.979880499	0	0.017957117	0.665784862	1
Friday, April 07, 2017 01:51:56	0.737158607	1	0.240731591	0.276895904	0

three houses analyzed can be looked at in terms of their mean, standard deviation, and coefficient of variance.

As seen in Table 2, the data has been split up and analyzed for two locations, West Hartford and Torrington. When analyzing the variable changes between the two locations, West Hartford had a much larger change in environmental variables over time than Torrington. For example, humidity in West Hartford had a 18.754% increase in relative standard deviation from Torrington. Seeing as all relative standard deviations are higher in West Hartford than in Torrington for the data collected, it can be said that during this time, weather changes in West Hartford varied more than those in Torrington.

$$\mu = \sqrt{\frac{\sum(x - \tilde{x})^2}{(n-1)}} \tag{2}$$

where $\tilde{x}$ is the average value of x for the sample population and n is the population size.

Taking this into account, the overall trend between the two locations shows that humidity and altitude had the highest relative standard deviation when getting migraines, therefore seemed to have the highest impact of migraines. These variables are where the user should look to make changes in their environment.

B. *Data Visualizations*

After normalizing the data from the database, it can be analyzed in graphical form. Analysis was done using a self-organizing map (SOM) and piecewise cubic interpolation graphs in Matlab.

The SOM is used to analyze correlations of variables to one another [12]. Using Matlab's SOM Weight Planes, a weight plane is calculated for each element in the input vector resulting in a visualization of the weights where lighter and darker colors represent larger and smaller weights, respectively, as seen in Fig. 1. Using these weight planes, if connection patterns are very similar, it can be assumed they are highly correlated.

Interpolation is a method of estimating values between already known data points. Therefore, when graphical data contains a gap in data but there is data on either side of the gap or in a few specific points, interpolations allows us to estimate the values within the gap.

Utilizing this knowledge, a self-organizing map (SOM) was implemented to analyze variable weights. As seen in Fig. 1, each weight plane created for each element of the input vector is displayed using 100 neurons. These visualizations of the weights show how each input is connected to the weights.

Using this, similarities can be seen between the weights from Input 1 (Humidity) and Input 5 (Pressure), as well as the weights from Input 2 (Temperature) and Input 3 (Migraines). Looking at these similarities, it can be said that if the connection patterns of two inputs are very similar in nature, the inputs are highly correlated. Thus, it can be said that Input 4 (Altitude) does not have a high correlation with any variable in the data as it does not share a similar pattern to any other variable.

Therefore, three variables were extracted from this analysis, temperature, humidity and migraines. A piecewise cubic interpolation graph can be shown to illustrate how temperature and humidity react to migraines over changes in each variable. In the data, migraines are given a value of 1 or 0 to indicate a migraine occurred or did not occur. With that, the temperature range for the nine-day trial was 21.28° to 36.62° Celsius, while humidity ranged from 29.00% to 80.12%. Like stated earlier, these values were normalized using (2).

Excluding altitude, which had the highest variations but no correlation, the room variables analyzed in terms of migraines in this study were humidity, temperature, and pressure. Looking at one individual's data, humidity varied the most over the nine-day trial and temperature had the highest correlation to

TABLE 2
STATISTICAL CALCULATION OF ROOM VARIABLES FOR MIGRAINES BY LOCATION

Room Variable	Mean		Variation		Standard Deviation		Relative Standard Deviation		
WEST HARTFORD, CT									
Humidity	39.165	%	208.441	%	13.988	%	±	35.714	%
Temperature	26.650	°C	13.921	°C	3.492	°C	±	13.102	%
Pressure	989.576	hPA	7974.218	hPA	78.576	hPA	±	7.940	%
Altitude	49.944	km	1085.594	km	35.215	km	±	70.510	%
TORRINGTON, CT									
Humidity	61.631	%	109.257	%	10.453	%	±	16.960	%
Temperature	26.515	°C	1.440	°C	1.120	°C	±	4.525	%
Pressure	961.197	hPA	244.641	hPA	15.641	hPA	±	1.672	%
Altitude	443.462	km	30.551	km	135.480	km	±	30.551	%

migraines. Using the interpolation graph in Fig. 2, the estimated values of migraines in terms of temperature and humidity can be filled in. With this, it can be said migraines typically occur in two key areas: dry with varying temperatures and high humidity and mid-range temperatures.

With this, the room variables analyzed in this trial did produce expected results for this individual. This same analysis can then be used on other individuals to see how their migraines react with the environment in which they live in. This application does have room for expansion to include other chronic illnesses affected by environmental changes to encompass a broader range of issues.

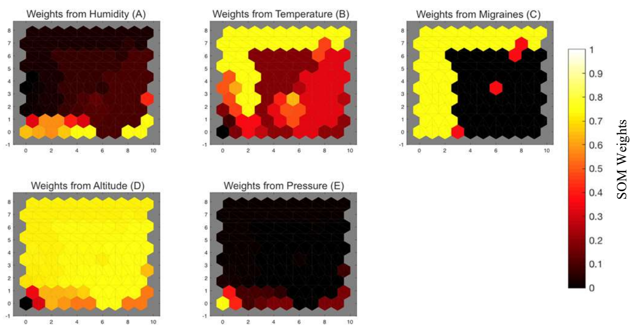

Fig. 1. SOM example data of weights from each input of the data.

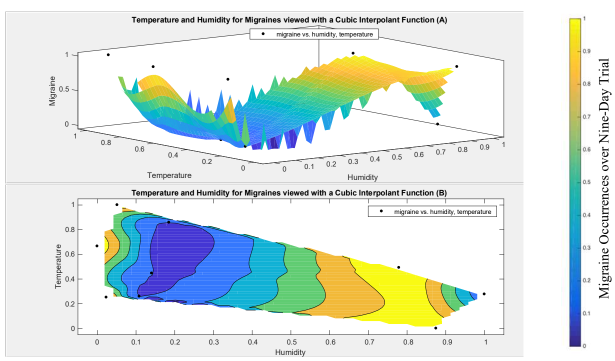

Fig. 2. Cubic interpolation graph of temperature and humidity data for migraines over the nine-day trial period.

C. *Machine Learning Methods*

With this, further analysis was done using machine learning techniques to look at prediction accuracy of migraines in terms of the environment. Using a custom Docker container set up with the Intel® Distribution for Python that contained optimized Scikit-learn, the data was loaded into Jupyter notebook.

Two sets of data were analyzed during this portion of the study. The first dataset was the original data from above and the second dataset included the original data collected above along with historical weather data from the area which included added columns for temperature, dew point, humidity, pressure, visibility, wind speed and conditions. The historical weather data was collected from Weather Underground. Using these two datasets, the data was first cleaned to remove unnecessary columns, as before, and then the date column was converted into a Timestamp value. Next, columns that contained category or object types were converted into an encoding in the dataframe. This allowed for all numeric columns in the data. Once the data was cleaned, it was examined to look at the variable being predicted, migraines. The data showed almost 50% split in the migraine column for (1) migraine and (0) no migraine. This means, half of the observations include days in which the individual suffered a migraine. Lastly, the data was normalized using the same equation (1) as before.

The five models used for evaluation with both datasets were Logistic Regression, Linear Discriminant Analysis (LDA), K-Nearest Neighbors (KNN), Decision Tree Classifier and Gaussian Naïve Bayes Classifier.

Moving forward, the data was split into feature and response data in which the migraines column was the response and all others were features. This split allowed for the dataset to be broken down into four datasets for training and validation. Two sets were kept for training and two for validation of the model.

Once split, the data was evaluated against the five machine learning models using K-Fold Cross Validation. This validation was run eight times against the models which returns the mean accuracy of all the models. Looking just as the original dataset, as seen in Fig. 3, without added features, the algorithm accuracy comparison showed that the LDA produced highest accuracy for the testing data at a mean of 0.906. Using this model to move forward, further validation was done to make predictions, and the data was tested again with K-Fold Cross Validation, with ten folds, to produce a mean accuracy of 0.917. After which, the final validation was completed on the dataset to produce a resulting validation mean accuracy of 0.800 using the LDA model.

Seeing the accuracy for the original dataset being between 80 to 90%, the five machine learning models were reevaluated using the second dataset which contained external weather data for the area, as seen in Fig. 4. Following the same process of using K-Fold Cross Validation, these models were run eight times against and retuned the mean accuracy of all the models. The algorithm accuracy comparison showed that the LDA produced highest accuracy for the testing data at a mean of 0.938. Using this model to move forward, further validation was done to make predictions, and the data was tested again with K-Fold Cross Validation, with ten folds, to produce a mean accuracy of 0.925. After which, the final validation was completed on the dataset to produce a resulting validation mean accuracy of 0.900 using the LDA model.

Final validation between the two datasets showed that the addition of external weather data aided in the accuracy of the predictions for migraines using the LDA method. This additional information increased the accuracy of the final validation model by 0.100 and shows how external weather conditions can also have effects on migraine sufferers.

IV. CONCLUSION

The environmental analyzer for migraines was designed to take in environmental data from a user's surroundings, i.e. temperature, humidity, pressure and altitude, while also taking in user data about migraines. These two datasets are then combined and cleaned to define a user's common environment and migraine occurrences. After the data has been cleaned, an iOS app analyzes the migraines in terms of the changes going on in the home environment, looking for common trigger points and areas of concern.

The environmental analyzer developed during this research can read room data variables while collecting a user's migraine data to examine migraines in terms of the environment in which the application then provided users with both a visual and numerical representation of the data after analysis.

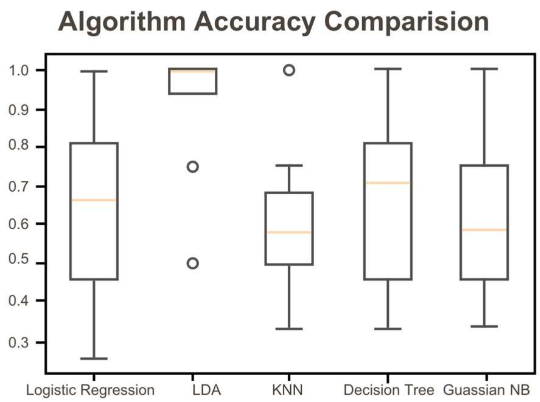

Fig. 3. Algorithm Accuracy Comparison for the five machine learning methods utilized in this study with the original data.

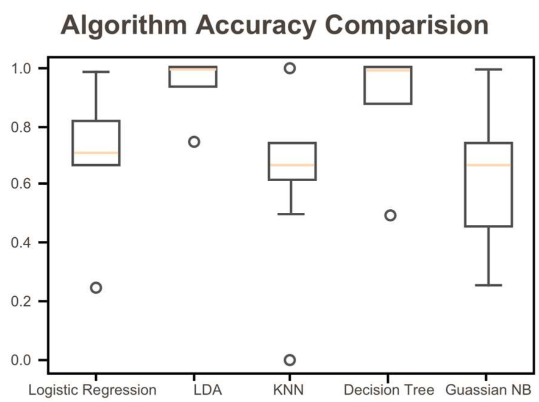

Fig. 4. Algorithm Accuracy Comparison for the five machine learning methods utilized in this study with added external weather data.

This device's target audience is those who suffer from migraine attacks that they believe has some form of environmental factor contributing to the migraines. Migraines have many similarities to sinus headaches and can often be misdiagnosed as such. Some common triggers that can affect migraine sufferers include pressure, temperature, humidity, noise, and light.

Through the nine-day study on one individual, the data collected confirmed that there were five migraines captured. During this trial, data was analyzed to show that temperature changes had the highest correlation to migraine occurrences while humidity had the highest variation in value in both locations that were tested. Together, humidity and temperature showed the two areas in which migraines most occur for this individual: dry with varying temperatures and high humidity and mid-range temperatures. Using this information, five machine learning techniques were used to examine data with and without external weather information provided. This produced two models with high accuracy for predicting migraine attacks, with added external weather data producing a higher accuracy than the data without it.

Utilizing information collected from this analysis, the individual can begin to change their home environment to aide in reducing future migraine attacks or worsening already occurring migraines.

ACKNOWLEDGMENT

The authors are grateful to the engineers of iDevices, Jonathan Day, Chris and Kaitlyn Turner, and Michael Tetreault, for their invaluable support and assistance during the realization of the "Environmental Analyzer" project.

REFERENCES

[1] R. Day and H. S. Salehi, "Development of Smart Interior Environmental Mobile App for Health Monitoring," in IEEE Green Tech 2018, Austin, TX, 2018, pp. 115-118.

[2] R. Day and H. S. Salehi, "Design and Implementation of an Interior Automated Environmental Analyzer with App," in 2017 IEEE MIT Undergraduate Research Technology Conference, Cambridge, MA, 2017.

[3] "Migraine Research Foundation," Migraine Research Foundation, 2016. [Online]. Available: https://migraineresearchfoundation.org/about-migraine/migraine-facts/. [Accessed 2018].

[4] F. Mongini, A. Deregibus, F. Raviola and T. Mongini, "Confirmation of the Distinction Between Chronic Migraine and Chronic Tension-type Headache by the McGill Pain Questionnaire," *Headache: The Journal of Head and Face Pain*, vol. 43, no. 8, pp. 867-77, Sep 2003.

[5] D. Friedman, "Migraine and the Environment," *Headache: The Journal of Head and Face Pain*, vol. 49, no. 6, pp. 941-952, 2009.

[6] "Environmental & Physical Factors," National Headache Foundation, 25 Oct 2007. [Online]. Available: https://headaches.org/2007/10/25/environmental-physical-factors/. [Accessed 2016].

[7] M. Dahlem, "A computational perspective on migraine aura," *Progress in Neurobiology*, vol. 74, no. 6, pp. 351-361, Dec 2004.

[8] C. Liu, K. A. Holroyd, Q. Zhu, K. Shen and W. Zhou, "Design and implementation of a behavioral migraine management iPhone app for adolescents with migraine," in IEEE International Symposium on "A World of Wireless, Mobile and Multimedia Networks" (WoWMoM), Montreal, QC, 2010, pp. 1-6.

[9] *"BME260 Datasheet," Bosch Sensortec. [Online]. Available: https://www.bosch-sensortec.com/bst/products/all_products/bme680. [Accessed 2017].*

[10] *"BME280 Datasheet," Bosch Sensortec. [Online]. Available: https://www.bosch-sensortec.com/bst/products/all_products/bme280. [Accessed 2017].*

[11] *"BME280 Datasheet," Bosch Sensortec, 7 May 2015. [Online]. Available: https://cdn-shop.adafruit.com/datasheets/BST-BME280_DS001-10.pdf. [Accessed 2017].*

[12] Kaski, S., "Data exploration using self-organizing maps". *Acta Polytechnica Scandinavica, Mathematics, Computing and Management in Engineering Series* No. 82, Espoo 1997, 57 pp. Published by the Finnish Academy of Technology.