Geolocation based air pollution mobile monitoring system

Aya Mazin Talib¹, Mahdi Nsaif Jasim²
¹Computers Sciences Department, Information institute for Postgraduate Studies, Baghdad, Iraq
²University of Information Technology and Communications, Baghdad, Iraq

ABSTRACT

Air pollution is conducted to harmful substances like solid particles, gases or liquid droplets. More pollutants CO, SO₂, NOₓ, CO₂. This research is proposed the design and implementation of mobile, low cost and accurate air pollution monitoring system using Arduino microcontroller and gas sensor like MQ2, MQ131, MQ135, MQ136, DHT22, measuring materials mentioned above, smoke, Acetone, Alcohol, LPG, Toluene, temperature, humidity and GPS sensor “NEO-6M” that track the location of air pollution data, and display the analysis result on ESRI maps. The system also save the results on SQL server DB. The data is classified using data mining algorithms, presenting the result on a map helps governmental organizations, nature guards, and ecologists to analyze data in real time to simplify the decision making process. The proposed system uses J48 pruning tree classifier generated using cross validation of fold (10) with highest accuracy 100%, while IBK ≈99.67, Naïve bays ≈90.89, and SVM ≈81.4. It’s found that the common air quality for Baghdad (study area) is between (“Good”, “Satisfactory”, and “Moderately”) for 1835 records of air samples during (January and February 2021) time period.

This is an open access article under the CC BY-SA license.

1. INTRODUCTION

Air pollution is a collection of gases and solid particles in the air. Chemicals from factories, dust, car emissions, mold spores and pollen may be suspended as particles [1]. Some air pollutants are poisonous. Inhaling them can increase the chance that you’ll have health problems [2]. People with lung or heart disease, children and older adults are at greater risk from air pollution [3]. Air pollution isn’t just outside-the air inside building can also be polluted and effect our health [4].

Human bodies don’t care whether they’re exposed to air pollution while we’re taking a walk outside or indoors [2]. They all has a health impact. Most people find that air pollution more exposed in indoors than outdoors [5]. Also, according to the World Health Organization, indoor smoke presents a serious health problem almost half the entire world population, for about 3 billion people who cook and heat their homes with biomass, kerosene fuels and coal [6]. In addition to indoors activities that cause air pollution (like candle burning, cooking, and using certain types of cleaning products) [7]-[11]. There are several types of air pollutant but the most commonly monitored are: Ozone, Sulfur Dioxide, Carbon Monoxide, Nitrogen Dioxide, Carbon Dioxide, LPG and Smoke.
1.1. Literature review

T. Landolsi A., et al., [12] and their study of “Pollution Monitoring System using Position-Aware Drones with 802.11 Ad-Hoc Networks ”they present the design of a pollution monitoring system consisting of an ad-hoc Wi-Fi network of unmanned aerial vehicles or drones, equipped with a single-board computer, several gas sensors such as carbon, ozone, monoxide, and methane and a video camera [12]. S. Dhingra, et al., in their study of “internet of things mobile-air pollution monitoring system (IoT-Mobair)” they propose a three-phase air pollution monitoring system. An IoT kit was developed using gas sensors, Arduino IDE (integrated development environment), and a Wi-Fi module. This kit can be physically placed in various cities to monitoring air pollution [13]. In 2020, V. T. Pyagay, et al., in their study of “Analysis and processing of environmental monitoring system” they present an approach to the construction using the LoRaWAN (long range wide area network) technology. They have identified certain patterns in collected data, such as dependence on the time of year, weather, the location of certain industrial facilities near the observed zone [4]. Jing Ma, et al., [14], in their study “Assessing Mobility-Based Real-Time Air Pollution Exposure in Space and Time Using Smart Sensors and GPS Trajectories in Beijing “by using real-time data from sensors and smartphone GPS trajectories collected in Beijing, China [14]. In 2020, Y. Narasimha Rao, et al., they presents study on real-time air pollution monitoring using wireless sensors on public transport vehicles [11]. In 2020, L. Bertrand, et al., in their study “How to choose healthier urban biking routes: CO as a proxy of traffic pollution ”they present study demonstrates the usefulness of a compact CO sensor (alpha sense CO-B4) mounted on a bicycle to evaluate atmospheric levels of CO [6].

1.2. The problem statement

Solid state fixed place gas monitor unit is high costly, present data for average 24 hours and off line with no advanced result like using data mining to analysis data.

1.3. The proposed solution

Design and implement a mobile air pollution monitoring system with low cost, has high accuracy, high sensitivity, effective, and presenting the results in real time. Providing advance results indexed by their GPS location, presented on city maps and easy to use with high speed. The data is classified using data mining techniques and the results overlaid on study area maps with different colors representing different pollution levels. The research aims to measure the pollution levels on the study area and present the results indexed by GPS coordinates and overlaid on digital maps classified with the aid of data mining techniques to simplify decision making process. Integrating Visual studio, ESRI maps, and data mining techniques to create mobile air pollution monitoring system with aforementioned features.

1.4. The contribution

1. Solid state gas monitor costing up to $150,000 per station while the proposed unit is mobile cost 3000$.
2. Solid state working off line collecting data while the proposed unit is online.
3. The precision of our mobile unit is more adhere the real environment then fixed place solid station.
4. The proposed system measure more than 10 different gases while the solid fixed station measure 4 gases only.
5. Old station has no mining process to discover data behavior while the proposed new unit has the ability to use data mining techniques for more result analysis.

2. RESEARCH METHOD

To build air pollution monitoring system hardware and software are needed. The hardware consists of an Arduino Uno,"NEO-6M" GPS sensor, “MQ2", “MQ131", “MQ135", and “MQ136” sensors to display air pollution concentration. Buzzer sensor to alarm detected gas threshold violation,”DHT22” sensor to measure temperature and humidity value, light emitted diode (LED) green/red to indicate safe or polluted area. Figure 1 display comprehensive view of air pollution system.

2.1. Arduino UNO

The Arduino Uno is a microcontroller. It has 6 analog inputs, 14 digital input/output pins, a USB connection and a reset button; easily connect it to a computer with a USB cable or battery or power it with AC-to-DC adapter to get started. Arduino can be powered with a regulated 5V or 3.3V supply. It is programmed with the Arduino IDE software integrated to C# of Visual studio 2019.
2.2. Sensors

MQ-2 gas sensor has good sensitivity CO gas in wide range; High sensitivity to LPG; Propane, Hydrogen, and CO; long life and low cost and concentration 300-10000 ppm [15]. MQ135 sensor wide detecting scope fast response and high sensitivity; detect/measure NH4, LPG, NOx, alcohol, Benzene, smoke, CO2; Sensing Resistance 30KΩ-200KΩ [16]; MQ136 gas sensor has high sensitivity to SO2, also could be used to detect other vapor which contains Sulfur [17]. MQ131 gas sensor is suitable for detecting NOx.; also can use to detect O3 concentration scope 10ppb-2ppm [18]. All MQ sensors has Analog output voltage: 0V to 5V; digital output voltage: 0V or 5V Preheat duration 20 seconds; Can be used as analog or digital sensor. NEO-6M global positioning system (GPS) module, a very popular, cost-effective, high-performance. GPS module has Horizontal position accuracy GPS 2.5 m, Maximum Navigation update rate 5Hz [19].

DHT22 humidity and temperature measurement sensor has high precision; full range temperature compensated; calibrated digital signal; outstanding long-term stability; extra components not needed; long transmission distance, up to 100 meters; low power consumption; 4 pins packaged and fully interchangeable; power supply 3.3-5.5V DC; operating range humidity 0-100%RH; temperature -40-80Celsius [20]-[22].

2.3. Circuit diagram

Component of circuit diagram in air pollution system are display in Table 1. Each VCC, GND in sensors are connected to VCC and GND in Arduino. When connection is done, the system starts to sense the gases, and gets the pollution level in parts per million (PPM). If sensors have moderate values the blue led is turn on, but when detected pollution levels come out of range the buzzer will start beeping and the red led turns on. See Figure 2.

<table>
<thead>
<tr>
<th>Component</th>
<th>Sensor pins</th>
<th>Arduino UNO</th>
</tr>
</thead>
<tbody>
<tr>
<td>MQ2 sensor</td>
<td>Analog pin</td>
<td>A0</td>
</tr>
<tr>
<td>MQ131 sensor</td>
<td>Analog pin</td>
<td>A1</td>
</tr>
<tr>
<td>MQ135 sensor</td>
<td>Analog pin</td>
<td>A2</td>
</tr>
<tr>
<td>MQ136 sensor</td>
<td>Analog pin</td>
<td>A4</td>
</tr>
<tr>
<td>DHT22 sensor</td>
<td>Digital pin</td>
<td>D11</td>
</tr>
<tr>
<td>Buzzer</td>
<td>Digital pin</td>
<td>D8</td>
</tr>
<tr>
<td>Neo-6M</td>
<td>RX</td>
<td>D3</td>
</tr>
<tr>
<td></td>
<td>TX</td>
<td>D4</td>
</tr>
<tr>
<td>Blue led</td>
<td>Digital pin</td>
<td>D10</td>
</tr>
<tr>
<td>Red led</td>
<td>Digital pin</td>
<td>D12</td>
</tr>
</tbody>
</table>

Figure 1. Comprehensive view of air pollution system

Table 1 Circuit diagram of air pollution system
2.4. The proposed system software

The software system consists of three parts, the first is the Arduino IDE program to deal with the sensors measurement. The second is written in Visual Studio/C# language that is responsible for controlling the hardware and analysis air quality using data mining algorithm to classify and display the results on the study area map. See Figure 3 the third part is data presentation using ESRI ARC-GIS 10.6 shape file to be overlaid on study area map depending on GPS sensor latitude and longitude readings.

![Figure 2. Air pollution unit circuit diagram](image)

![Figure 3. Sensor reading and classification](image)

2.5. The proposed system database

SQL Server database is created to store the sensor reading data in database table with a unique GPS reference location of each area point scanned, see Table 2.

<table>
<thead>
<tr>
<th>Lati.</th>
<th>Lon.</th>
<th>Hu</th>
<th>Tem</th>
<th>NOx</th>
<th>CO</th>
<th>SO2</th>
<th>CO2</th>
<th>NH4</th>
<th>Sma.</th>
<th>Aceto.</th>
<th>Alco</th>
<th>LPG</th>
<th>Tolu.</th>
<th>Date/time</th>
</tr>
</thead>
<tbody>
<tr>
<td>33.30</td>
<td>928</td>
<td>44.3</td>
<td>24.6</td>
<td>0</td>
<td>4.95</td>
<td>25</td>
<td>3.35</td>
<td>4.99</td>
<td>19</td>
<td>0.57</td>
<td>1.59</td>
<td>4</td>
<td>0.67</td>
<td>2/1/2021-2:47 PM</td>
</tr>
<tr>
<td>33.30</td>
<td>938</td>
<td>44.3</td>
<td>49.1</td>
<td>0</td>
<td>4.65</td>
<td>25</td>
<td>3.21</td>
<td>4.8</td>
<td>19</td>
<td>0.54</td>
<td>1.51</td>
<td>4</td>
<td>0.63</td>
<td>2/1/2021-2:57 PM</td>
</tr>
<tr>
<td>33.30</td>
<td>504</td>
<td>33.2</td>
<td>24.1</td>
<td>0</td>
<td>38.62</td>
<td>24</td>
<td>14.94</td>
<td>18.15</td>
<td>38</td>
<td>3.29</td>
<td>8.37</td>
<td>8</td>
<td>4.05</td>
<td>2/1/2021-3:17 PM</td>
</tr>
</tbody>
</table>

Geolocation based air pollution mobile monitoring system (Aya Mazin Talib)
2.6. Mining process

Data mining is an iterative process which involves various steps [23]. SQL server data table has columns as indicated in Table 2.

2.6.1. Preprocessing

The database table is passed to be preprocessed by MS Excel in order to normalize the sensors data, remove outlier values. The total records number is 1835 and the system focused on dealing with (Latitude, longitude, NOx, CO, SO2, CO2, AVG and class) fields as in Table 3. These Data field are passed to J48 classifier to classify the map areas according to their pollution levels to help decision makers easily deals with pollution sources. The filed class represents the average classification results from data mining process.

An air quality index (AQI) is used by government to measure air pollution. Public health risks increase as the AQI rises. Different countries have their own air quality indices by average air quality over time [4], [24], [25], see Table 4.

Table 3. Air pollution chosen module

<table>
<thead>
<tr>
<th>latitude</th>
<th>longitude</th>
<th>NOx</th>
<th>CO</th>
<th>SO2</th>
<th>CO2</th>
<th>AVG</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>33.30928</td>
<td>44.33004</td>
<td>0</td>
<td>4.95</td>
<td>25</td>
<td>3.35</td>
<td>8.325</td>
<td>moderately polluted</td>
</tr>
<tr>
<td>33.30938</td>
<td>44.33007</td>
<td>0</td>
<td>4.65</td>
<td>25</td>
<td>3.21</td>
<td>8.215</td>
<td>moderately polluted</td>
</tr>
<tr>
<td>33.30894</td>
<td>44.33157</td>
<td>0</td>
<td>38.62</td>
<td>24</td>
<td>14.94</td>
<td>19.39</td>
<td>severe</td>
</tr>
<tr>
<td>33.30881</td>
<td>44.33393</td>
<td>0</td>
<td>29</td>
<td>24</td>
<td>12.13</td>
<td>16.2825</td>
<td>very poor</td>
</tr>
</tbody>
</table>

Table 4. AQI concentration labeling

<table>
<thead>
<tr>
<th>AQI</th>
<th>NOx</th>
<th>CO</th>
<th>SO2</th>
<th>CO2</th>
<th>AQI Value</th>
<th>Color</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good</td>
<td>0-40</td>
<td>0-1.0</td>
<td>0-40</td>
<td>0-1.0</td>
<td>0-30</td>
<td>Green</td>
</tr>
<tr>
<td>Satisfactory</td>
<td>41-80</td>
<td>1.1-2.1</td>
<td>41-80</td>
<td>1.1-2.1</td>
<td>51-100</td>
<td>Yellow</td>
</tr>
<tr>
<td>Moderately polluted</td>
<td>81-180</td>
<td>2.1-10</td>
<td>81-380</td>
<td>2.1-10</td>
<td>101-150</td>
<td>Orange</td>
</tr>
<tr>
<td>Poor</td>
<td>181-280</td>
<td>10-17</td>
<td>381-800</td>
<td>10-17</td>
<td>151-200</td>
<td>Red</td>
</tr>
<tr>
<td>Very poor</td>
<td>281-400</td>
<td>17-34</td>
<td>801-1600</td>
<td>17-34</td>
<td>201-300</td>
<td>Purple</td>
</tr>
<tr>
<td>Severe</td>
<td>400+</td>
<td>34+</td>
<td>1600+</td>
<td>34+</td>
<td>301-500</td>
<td>Maroon</td>
</tr>
</tbody>
</table>

2.6.2. Classification

When we consider the distribution of the “CO, NOx, SO2, CO2” we find that, NOx (1835 “good”), SO2 (1824 ”good”,11” Satisfactory”), air quality in CO (709 ”moderately polluted”, 63 ”Poor”, 6 “very Poor”, 546 “Satisfactory”, 475 ”good” and 36 “severe”), and CO2 (567 ”moderately polluted”, 807 ”Satisfactory”, 422 ”good”, 18 ”Poor”, 21 “Very Poor”) in CO2. We find that CO is between (0.32-65.52), CO2 is (0.46-21.95), SO2 is (15-56), and NOx is 0 all the time as in Figure 4.

Table 5. Distribution of NOx, CO, SO2, and CO2

<table>
<thead>
<tr>
<th>Gas</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>StdDev</th>
</tr>
</thead>
<tbody>
<tr>
<td>CO</td>
<td>0.32</td>
<td>65.92</td>
<td>3.46</td>
<td>7.517</td>
</tr>
<tr>
<td>CO2</td>
<td>0.46</td>
<td>21.95</td>
<td>2.213</td>
<td>2.689</td>
</tr>
<tr>
<td>SO2</td>
<td>15</td>
<td>56</td>
<td>24.43</td>
<td>2.8</td>
</tr>
<tr>
<td>NOx</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Figure 4. Distribution of NOx, CO, SO2, and CO2
2.6.3. Decision tree classification

By using Weka 3.8.4, a J48 Pruning tree is generated using cross validation of fold (10) with highest accuracy result 100%, while IBK ≈99.67, Naïve bays ≈90.89, and SVM ≈81.4. The J48 Tree diagram shown in Figure 5 illustrates the classification path that the data suggests.

![Decision Tree Diagram](image)

**Figure 5.** Decision tree classification

2.7. Data presentation

The proposed system consists of two main methods to solve the problem:

1) Draws each pollution concentrations level independently on ArcGIS map at zone/WGS_1984_UTM_Zone_38N, see Figure 6.

2) Calculate the AQI as (1) and draw air pollution concentrations by IDW in ArcGIS zone/WGS_1984_UTM_Zone_38N at [13], see Figure 7.

\[
AQI = \frac{I_{lo} - I_{lo}}{BP_{lo} - BP_{lo}} + (C_p - BP_{lo}) + I_{lo}
\]

- AQI = the AQI for the pollutant
- BPHi = the breakpoint that is equal to or greater than Cp
- Ilo = the AQI value corresponding to BPLo
- BPLo = the breakpoint that is equal to or less than Cp
- IHi = the AQI value corresponding to Bphi
- Cp = the rounded concentration of the pollutant

![Air Pollution Map](image)

**Figure 6.** Air pollution result on map
2.8. Data analysis

According to data sampling throw (January and February 2021) time period we find that the common air quality of Baghdad (study area) is between (“good”, “Satisfactory”, and “Moderately polluted”) Orange color (“Moderately polluted”). Green color (“good”) in Jazirat ‘um Alkhanażir near Tigris River, Baghdad University and gardens which has a lots of green trees. It is Maroon color (“Severe”) in “Msfa Aldawruh” which it is oil refinery. And in the east of “Msfa Aldawruh” it is red color (“Poor”) and in Southwest of “Msfa Aldawruh” it is Purple (“very poor”). Yellow color (“Satisfactory”) in Baghdad University and near rewers with less population.

3. RESULTS AND ANALYSIS

A mobile, low cost air pollution monitoring system is high sensitive and very effective to measure air pollution in real time with ability to track pollution location using GPS sensor. Air pollution affected by the level of humidity & temperature in the air, in Figure 8 we display the level of humidity and temperature for 1835 collected records for weather during (January and February 2021) time period. According to collected data sampling we find that the common air quality for Baghdad (study area) is between (“good”, “Satisfactory”, and “Moderately polluted”) because of cars emition, crowds, public power generators, burning things, and sewers.

Among the difficulties encountered in the system, the accuracy of system results are affected by the number of scanned data. Some areas are not reachable because of security or technical restrictions. The lack of high precision and utmost updated maps.

![Figure 7. Air pollution result in IDW](image)

![Figure 8. Humidity and temperature reading system](image)

4. CONCLUSION

The implemented mobile air pollution monitoring system reduces the cost of air pollution monitoring system by replacing the high cost fixed monitoring stations. The system usage increases the monitoring accuracy by using high sensitive sensors. These sensors can detect more gases for each, so we can use it to detect more available concentration in future. In addition to the data of the old fixed units system are referenced to only five locations, while in the new system the collected data are referenced to its actual GPS locations so the new system is more accurate in data acquisition and presentation. Ease of use and simple to
understand results using maps. GPS referenced data to simplify data analysis. This system can be used to detect 10 different gases concentration and their position can be regulated according to scanning speed to save time, disc space, and eliminate errors. The system can be mounted on drowns to give it more scanning coverage and speed.

REFERENCES


**BIOGRAPHIES OF AUTHORS**

Assit. Prof. Mahdi Nsaif Jasim (PhD), University Of Information Technology And Communications, College of Business Informatics Dept. of Management Information Systems, Born in Babylon/Iraq, lives in Baghdad. Interest: information systems, data and information security, mining in vector data, GIS, database systems. The researcher has interest in SDN data acquisition and data processing. He also Supervised a number of PhD and MSc. Students in different Iraqi universities.