International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 05 | May - 2022 Impact Factor: 7.185 ISSN: 2582-3930

IOT based Air Pollution Monitoring System with Machine Learning Techniques

* Dr. P.Punitha Ponmalar [1], Dr. C.R. Vijayalakshmi [2]
[1] Dept. of Computer Science , Sri Meenakshi Govt Arts College for Women, *M*adurai-2,TN, India
[1*] p.punithaponmalar@gmail.com

[2]Dept. of Computer Science
Govt. Arts & Science College , Aundipatti, Theni, TN, India
vijinsc@yahoo.in

Abstract. Air quality may be a universal challenge for governments and citizens. Many governments are capitalizing more to measure the levels, causes of pollution and to convalesce the air quality. They are empowering cities to tackle air pollution locally. In the earlier days air quality monitoring is performed with the help of pricy scientific instruments installed in small number of fixed locations. Nowadays the analysis and detection of the harmful gases present in the atmosphere which will be easily identified by the newest technology of 'IoT-Internet of Things'. Analytics and machine learning, can then be applied to the present data to know the causes, prediction and fluctuations in pollution.

Keywords: IOT, air pollution monitoring, Data aggregation, machine learning, data analysis .

1 Introduction

Air pollution has developed a foremost environmental risk to the extent that public health is concerned. Air pollution is excessive quantities of substances including gases, particulates and biological molecules are introduced into Earth's atmosphere.In many cities pollution levels exceed allowed and World Health Organization (WHO) limits. Reduced air quality is causing a public health problem, since breathing polluted air increases the risk of debilitating and lethal diseases such as lung cancer, stroke, heart disease and chronic bronchitis. Air pollution causes one in ten deaths.

The reduction in levels of air pollution may result in lessening of the global burden of disease. Air pollution monitoring and management has been administered for an elongated time. Management of air pollution is fronting challenges due to lack of obtainability of suitable tools and techniques.

Taking advantage of advances in communications and sensor technologies a new generation of low-cost sensor devices are available recently in the market. The IoT allows physical objects and people to be connected and monitored through the internet. In the IOT, billions of objects [14] can be found of various types such as GPRS, temperature, pressure and humidity sensors which are connected to the internet and transforms information from the physical world in to the digital world shown in Fig. 1.

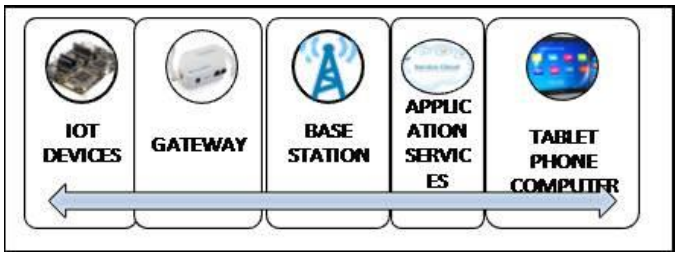

Fig. 1. IOT Physical Design

Air Pollution Monitoring with these IoT connected devices sense the environment several times a minute. In the IoT, sensing devices [18], [6] are normally powered by limited batteries and energy supplies. Therefore, it is necessary to improve the lifetime of long-term applications such as weather forecasting. The data aggregation strategy is to aggregate and collect the data packets in an effective manner in order to improve the energy consumption and network lifetime. Then, the aggregated value to a connected analytics solution. Analytical solution that delivers dynamic, local information to stakeholders, early indications of pollution hotspots, giving citizens the opportunity to avoid those areas.

Machine learning can help public to accurately predict hourly pollutant concentrations. Recently, various algorithms are used to predict pollutant levels, including the traditional machine learning method such as decision tree (DT), logistic regression and so on. In this work SVM and KNN is used to predict the pollutant concentrations. **K-Nearest Neighbour (KNN) Classifier:** KNN is a supervised non-parametric learner which classifies the data to a given category based on training sample. **Support Vector Machine (SVM):** SVM is a supervised machine learning algorithm used for both classification and regression.

The rest of the paper is systematized as follows. Section II discusses the related work, section III, and describes the research problem definitions. Section IV focuses on proposed work and its architecture and Section V discusses on Results and discussion.

2 Related Work

Yi[20] discussed various sensing technologies and MEMEs used in air pollution monitoring. Kelly [13] deliberated temperature monitoring, humidity and light intensity, based on a distributed sensing unit, data aggregation and context awareness. Yang [2] proposed a monitoring system which gives the concentration of Carbon-di-oxide of remote area. The system also reports temperature humidity and light intensity of the outdoor monitoring area. Kim [7] proposed a real time indoor air quality monitoring system to monitor seven different gases.

Sulayman K. Sowe [12] discussed typical air automatic observation system. IOT, will cut back the hardware value as lower as before.Thangarajet.al[9] have focused on a review of data aggregation techniques in WSN and various types of data aggregation. Rahman et al[10] have discussed a few methods, which are used for data aggregation.

3 Research Problem Definition

Air pollution monitoring system includes large number of devices connected and interacted to monitor the environment. So, an enormous volume of data is exchanged, the data might also be formatted or unformatted. The types of data determine the backend storage. The formatted and unformatted data needs to be accumulated, assessed with efficient data analytic methods. Few researchers are focused on the data gaining and aggregation. Some are focused on the application framework. Few researchers focused attention to the analytics of data with machine learning techniques to provide a prediction on air pollution. But the data aggregation with machine learning techniques revolves about the efficient use of data from IoT devices and to provide an exact prediction on air pollution.

4 Proposed Architecture

Smart Air pollution monitoring system continuously monitors solid and liquid particles and certain gases that are suspended in the air. The air pollution data are collected from the environment through the sensors in the IoT architectures. The information regarding the pollution levels are collected through the sensors placed on the environment. The sensors are battery driven and it cannot be supplanted frequently. To deterioration the energy utilization of the battery and the upswing in the amount of data diffusion are handled with data agglomeration and the data get stored in the database. Finally, Machine learning techniques such as classification and clustering are activated on the gathered data to deliver more precise prediction results and to provide public awareness.

3.1 Architecture

The system architecture of the Smart air pollution monitoring system shown in Fig. 2, which includes four layers.: Perception Layer ,Network Layer, Middleware Layer, and Application Layer.

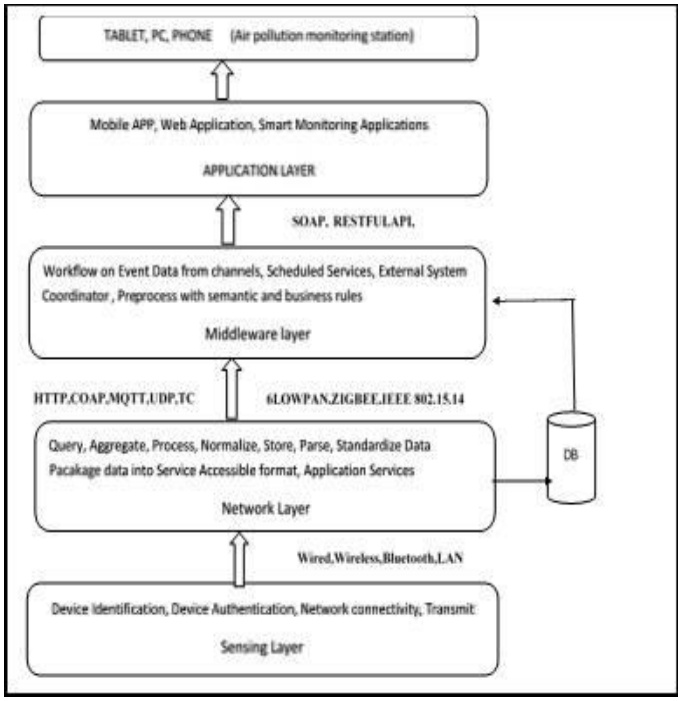

Fig 2: Smart Air Pollution system architecture

A. Perception Layer:

Perception Layer is used in data gathering from IoT devices like sensors and actuators. This Layer collects air pollution data over air pollution monitoring equipment. It measures various gas levels found in the environment and to give accurate readings of O3 or NO2 in ambient air.

The smart gadgets gather information and deliver out in the format of the opcode. Sensor uses low energy consumptionnetwork protocol called IPv6 over Low-Power Wireless Personal Area Networks (6LoWPAN) and the sensor gateway relays sensor data in Sensor Markup Language (SenML) format via WiFi connection.

B. Network layer:

This Layerperforms functions such as Data accumulation, normalization and validation. Sensors collects the data and transmits it

repeatedly that data are often discerned as base data. That also create data redundancy. Data aggregation evades redundant transfer of packets and it lessens the energy depletion, increases the packet delivery ratio and elongates the lifespan of the IoT.

The application reliant aggregated data could also be considered as initial data. So the data arecleaned, validated, pervaded and then assessed with the rules and conditions constituted. Data normalization eliminates redundancy and standardize the information. The data exchange format of this layer might be the XML or JSON.Datastorage options for owing be TinyDB, SINA and etc. The transmissiondeal options ranges from IEEE 802.15.4e ,2G-3G, LTE and CoAP.

C. Middleware Layer:

The middleware services could be event triggered and service oriented. The middleware layer furnishes the services are Air pollution device observation services, Web proclamation services, Sensor observant services. Application and Business Services, WorkFlow on Event Data from Channels, Scheduled Services, Report Engine, Alert Engine, External System Coordinator, Business Rule Management Services, User Profile and Previlege setting.

D. Application Layer:

This layer helps the people to know about the information on air pollution. It helps to handle an crunchy position and to escalate the efficacy and performance of the analysis. It supports the people to preclude from getting the ailment in the initial stage of air pollution.

3. 2 Workflow model

Model, predict, and monitor air quality is becoming more and more important due to the pragmatic critical effects of air pollution for populaces and the environment Smart Air pollution system continuously monitors the environment and collects concentrations of gases such as CO, CO_2, SO_2 and NO_2 using semiconductor sensors. Collected data aggregated to reduce the redundancy and transmitted to end devices via various wireless technologies and stored in the database. Machine learning techniques such as classification and clustering are applied to provide more accurate prediction results about the air pollution level.

The Process flow of the smart Air pollution system is shown in Fig.3 and Fig.4 shows the algorithm for the system.

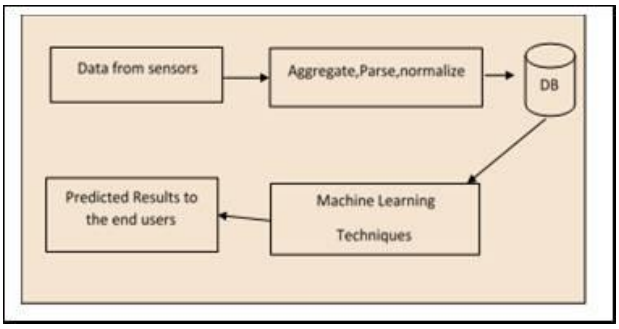

Fig. 3. Process Flow of Smart Airpollution System

3. 3 Machine Learning Algorithms

Machine learning algorithms are used to organize, classify and to make automated decisions. Using the machine learning algorithms, the dataset for a air pollution monitoring system is trained and analysis is done based on the test set.

A. Support Vector Machine

The famous supervised machine learning algorithm used for both classification and regression is support vector machine. At first SVM was applied for classification problems [16] and later on it was used for solving regression, image processing etc. [15]. SVM

generates hyperplanes which distinguishes different classes. The best possible distribution is attained only when the hyperplane has the maximum functional margin [17].

Algorithm 1: Disintegration and Data corroboration from IOT Air Pollution Monitoring Devices

Input:IOT_AP: Dev={ IOT_AP_Reading, pollution level},i=1..n
For Each event of the Device (Dev) € IOT_AP **Do**
Begin Authenticate Device , Create Database connection (con)
Insert SmartDataObject(Dev) , ExecuteQuery(con)
If DevTriggers Transmission **Then** CreateRootNode(Dev)
Refer Device Definition, Business Rule from Database(con)
Add SmartDataObjectTag in EXI XML() , Get Attribute() and AttributeValue()
Create EXI XML ChildNode for all attributes() , Create APRecord in EXI XML format()
Normalize the APRecord() as per standard(con)
Validate Data for Device(Min, Max Range) from DB(con)
Check for Datatype, Dataformat, Correct timestamp from Device Standards Described in DB(cc
. **If** Check or Validation fails **Then**
Follow the Fallback Action as per workflow configuration **EndIf**
Process the APRecord (Message Composition, Content Categorization, Message Filtering, Parsi
)Aggregate the APRecord
Analyze the APRecord using Machine Learning Techniques for Prediction Model
On the Fly Data Transformation(Source, Target) using Business rules in DB(con) , Repackage
APRecord into Target Data Model()
Identify Network connection
Transmit APRecord in the Protocol transmission format **End**

Fig. 4. Algorithm for smart air pollution monitoring

B. K-Nearest Neighbor (KNN) Classifier

KNN [19] , [11] is a supervised classifier which stores all available cases and categorizes new one with the help of distance measures. It is a non-parametric algorithm. KNN is also called as lazy learner because first it stores the training data and then waits for a testing data. When the test data arrives, it performs the classification based on the most similar data using distance measures. Here Euclidean distance measure is used for finding the closest neighbors.

5 Implementation

The Smart Air pollution monitoring system was implemented using WEKA[3] tool. The SmartSVM is implemented by LibSVM library [1] and SmartKNN is designed by 1BK algorithm. The potentiality of the smart systems are compared against normal SVM and KNN. The evaluation of experiments was made by Intel Core i5 2.67 GHz with 4 GB RAM**,** running Windows 7. The

emphirical experiments were conducted on the Air quality dataset [4] with the help of following measures. In this study, 10 fold cross validation method is used .

- Accuracy: The proportion of correctly classified tuples
- Runtime: The time taken to generate and test the classifier in seconds
- F-measure: Harmonic mean of the precision and recall of the test.
- Specificity**:** Defined as the ratio of true negative divided by the sum of a true negative and false positive.
- Sensitivity: Defined as the ratio of true positive divided by the sum of a truly positive and false negative.

5. 1 Air Quality Data Set

Metrics	Smart SVM	SVM	Smart KNN	KNN
Sensitivity	0.97	0.83	0.93	0.76
Specificity	0.98	0.82	0.89	0.81
F-Measure	0.97	0.84	0.92	0.83

Table 1: API Level

This dataset 824 instances and 9 features. The features are Country information , state, city, place, last update, minimum, maximum, average, and pollutants. The missing values in the data are imputed by K- means clustering imputation [5]. The Principal Component Analysis method (PCA) [8] is used to select the important attributes in the dataset and these features are discretized. The Table 1 shows the predicted AQI values.

SNO	API	Air Quality	Causes
1.	51-100	Moderate	Respiratory disease for prolonged outdoor exertion.
2.	101-150	Unhealthy	Members of sensitive groups may experience health effects.
3.	151-200	Unhealthy	Everyone may begin to experience health effects
4.	201-300	Very Unhealthy	Health warnings of emergency conditions
5.	300+	Hazardous	Health alert: everyone may experience more serious health effects

Table 2. Performance of Smart Air pollution System.

The Table 2 shows the sensitivity and specificity of the Smart air pollution monitoring system using SVM and KNN classifier.

This table indicates that SmartSVM achieved higher sensitivity and specificity values when compared to KNN. The improvement in sensitivity value by SmartSVM is 14% and 17% by SmartKNN. The specificity increment by Smart SVM is 16% and by SmartKNN is 8%. The f-measure value of Smart systems is higher than the normal classifiers by 13% and 9%.

This table shows that Smart Air pollution monitoring System achieved higher sensitivity and specificity values using Smart classifiers when compared to SVM and KNN while monitoring the pollution.

The Fig.5 shows the accuracy of Smart Air pollution monitoring System when designed using SmartSVM and SmartKNN methods.

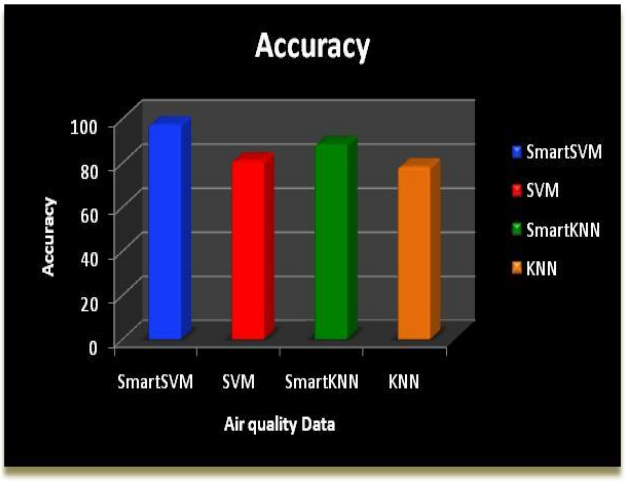

Fig. 5. Accuracy of Smart Air Pollution System

This graph depicts that SmartSVM and SmartKNN achieved high accuracy when compared to SVM and KNN. The accuracy improvements by smart systems are 16.56% and 10.09%.

The Fig.6 shows the runtime of Smart Air pollution monitoring System against SVM and KNN. This figure indicates that the runtime of Smart Air pollution monitoring System is lesser than the normal classifiers.

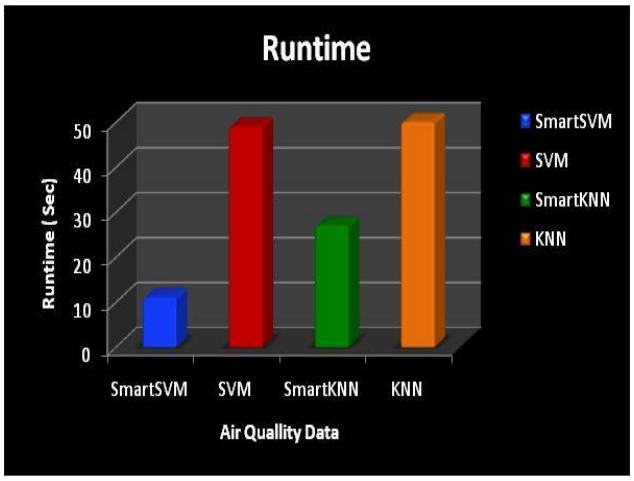

Fig. 6. Runtime of Smart Air Pollution System

In all the experiments, the Smart Air pollution monitoring system performs well than the normal SVM and KNN classifiers based on various measures like sensitivity, specificity, accuracy and runtime and f-measure. When comparing Smart Air pollution monitoring systems using SmartSVM and SmartKNN , the SmartSVM is better than SmartKNN in terms of all measures.

6 Conclusion

Precise forecasting of air pollution benefits people to plan ahead and lessening the effects on health. Predicting the air quality is a multifaceted task due to the dynamic nature, unpredictability, and high inconsistency in space and time of pollutants and particulates. This work presented a study of air pollution to forecast pollutants and particulates' levels and to correctly identify the AQI. The most efficient algorithm among the four algorithms for air quality prediction is SmartSVM. The result proves that the proposed methods are effective and reliable for use.

References

1. C.J. Lin and C. Chang , "LIBSVM: A Library for Support Vector Machines," 2005. http://www.csie.ntu.edu.tw/~cjlin/libsvm
2. H. Yang, Y. Qin, G. Feng, and H. Ci, "Online Monitoring of Geological CO2 Storage and Leakage Based on Wireless Sensor Networks," Sensors Journal, IEEE, vol. 13, no. 2, pp. 556–562, Feb. 2013.
3. H.W. Ian, and E. Frank, (2000) *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations.* Morgan Kaufmann Publishers.
4. https://www.kaggle.com/venky73/airquality.
5. J. Deogun, , W.Spaulding., B. Shuart et al." Towards missing data imputation: A study of fuzzy k-means clustering method," Proc. of 4th Int. Conf. on Rough Sets and Current Trends in Computing (RSCTC04), Vol. 3066 of LNCS, pp. 573-579.2004.
6. J.Muhammad Saqib, et al. "Smart environment monitoring system by employing wireless sensor networks on vehicles for pollution free smart cities." *Procedia Engineering* 107 (2015): 480-484.
7. J.Y. Kim, C.-H. Chu, and S.-M. Shin, "ISSAQ: An Integrated Sensing Systems for Real-Time Indoor Air Quality Monitoring," Sensors Journal, IEEE, vol. 14, no. 12, pp. 4230–4244, Dec. 2014.
8. L,I. Smith,,2002, " Tutorial on principal component analysis," February.
9. M.Thangaraj,P.PunithaPonmalar 'A survey on data aggregation techniques in wireless sensor networks" International Journal of Research and reviews in wireless sensor networks, 2011.
10. R. Hafizur, N. Ahmed, and I. Hussain. "Comparison of data aggregation techniques in Internet of Things (IoT)." 2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET). IEEE, 2016.
11. R.. Yang , D. Ding and Feng, Yan. (2019). Application of Improved KNN Algorithm in Air Quality Assessment. HPCCT 2019: Proceedings of the 2019 3rd High Performance Computing and Cluster Technologies Conference. 108-112. 10.1145/3341069.3342976.
12. S. K. Sowe , Takashi Kimata , Mianxiong Dong , Koji Zettsu, "ManagingHeterogeneous Sensor Data on a Big Data Platform: IoT Services for Data-Intensive Science", 2014.
13. S. Kelly, N. Suryadevara, and S. Mukhopadhyay, "Towards the Implementation of IoT for Environmental Condition Monitoring in Homes," Sensors Journal, IEEE, vol. 13, no. 10, pp. 3846–3853, Oct. 2013.
14. S. Khaled Bashir, Abdullah Kadri, and EmanRezk. "Urban air pollution monitoring system with forecasting models." *IEEE Sensors Journal* 16.8 (2016): 2598-2606.
15. S. Olmedo, M. A. Aceves-Fernández, E.F. Hurtado et al., Forecast Urban Air Pollution in Mexico City by Using Support Vector Machines: A Kernel Performance Approach
16. V. N. Vapnik. An Overview of Statistical Learning Theory. IEEE Transactions of Neural Networks, Vol. 10, No. 5, (1999).
17. W.Lu, Wenjian Wang, A. Y. T. Leung, Siu-Ming Lo, R. K. K. Yuen, Zongben Xu, Huiyuan Fan. Air Pollutant Parameter Forecasting Using Support Vector Machines. IJCNN, (Volume: 1), (2002).
18. X. Chen, L. Xianpeng, and X. Peng. "IOT-based air pollution monitoring and forecasting system." *2015 International Conference on Computer and Computational Sciences (ICCCS)*. IEEE, 2015.
19. Y. Zhao, Y. Abu Hasan, Comparison of three classification algorithms for predicting PM2.5 in Hong kong Rural area, Asian Journal of Scientific Research 3(7):715-728, 2013.
20. Yi, Wei Ying, et al. "A survey of wireless sensor network based air pollution monitoring systems." *Sensors* 15.12 (2015): 31392-31427.