International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 03 | Mar 2019 www.irjet.net p-ISSN: 2395-0072

Air quality monitoring using CNN classification

R.KALAISELVI[1], S.PRIYANGA[2], R.GAJALAKSHMI[3], S.INDUMATHI[4], R.RAMYA[5]

[1]Assistant Professor, Computer Science and Engineering Department, Arasu Engineering College, Kumbakonam, Tamilnadu, India
[2]UG Scholar, Computer Science and Engineering Department, Arasu Engineering College, Kumbakonam, Tamilnadu, India
[3]UG Scholar, Computer Science and Engineering Department, Arasu Engineering College, Kumbakonam, Tamilnadu, India
[4]UG Scholar Computer Science and Engineering Department, Arasu Engineering College, Kumbakonam, Tamilnadu, India
[5]UG Scholar Computer Science and Engineering Department, Arasu Engineering College, Kumbakonam, Tamilnadu, India

---

Abstract: *Deep air learning analyzes the quality of air, by capturing an image in an open atmosphere. Later, the captured image will be updated in the dataset .The pixels are pointed in the gray scale conversion and the noise will be removed by using the homomorphism filter. The picture which is already updated will be compared to the new picture by using CNN classification algorithm, which will analyze the quality of air. Example oxygen, carbon-dioxide etc., Air pollution may cause many severe diseases. An efficient air quality monitoring system provides great benefit for human health and air pollution control. The proposed method uses a deep convolution Neural Network (CNN) to classify the natural image in to different categories based on their concentration. The experimental results demonstrate that this method validate the image-based concentration estimation to analyze the quality of air.*

KEYWORDS: Deep air Learning, CNN classification, Big data.

I. INTRODUCTION:

In today's world large amount of data can be generated and at the same time, it should be needed the new forms of data processing that enable enhanced insight, decision making, cost effective and process automation. Those data's are typically called as the Big Data. A massive volume of structured, semi structured and unstructured data that is so large that it's difficult to process with traditional database and software techniques. Data can be classified as structured, semi structured and unstructured based on how it is stored and managed.

In several challenges for urban air computing as the related data have some special Characteristics. There is not a universally accepted judgment to reveal the main causes of the occurrence and dissipation of air pollution. The labeled data of the air-quality-monitor-stations are incomplete, and there exist lots of missing labels. In Logistic Regression avoids data latency to its core. Chance of loss or damage of data[1].Fuzzy logic and Logistic Regression Clustering offers more efficient storage space. Integrating of grouped data may end in failure[2]. Svm algorithm competes with the interpolation data information stored are with insecure compensation[3]. Random Forest the computational cost of training a random forest it

quit low to deliver the same level of accuracy but it will benefit from massive amount of data[4]. We purposed this method, we need to implement smoke detection analysis software that includes an algorithm for image processing. The algorithm is based on image analysis[4] from thermal and visible wavelength cameras.

Thus, the proposed system obtains the images from the environment and monitors the pollution through image processing. The amount of pollution in the environment is obtained from the dialog box which displays the pollution level in the environment. Various steps have been followed to obtain the pollution level in the environment. The various steps involved are obtaining the input image,pre-processing of the input image and edge detection using canny operator. Finally, a dialog box is been displayed to show the level of pollution.

II.RELATED WORK:

2.1:Existing system:

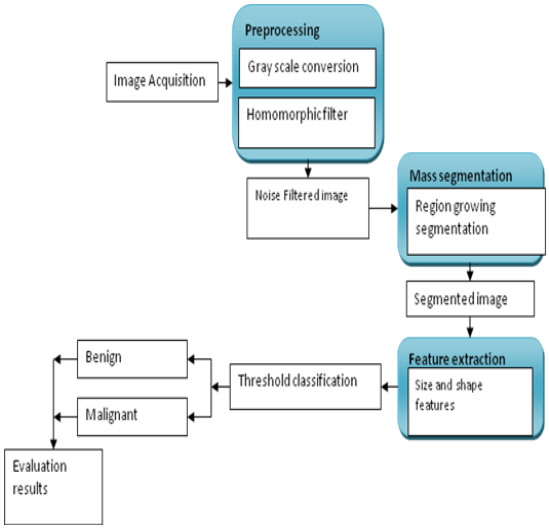
In several challenges for urban air computing as the related data have some special Characteristics. There is not a universally accepted judgment to reveal the main causes of the occurrence and dissipation of air pollution. The labeled data of the air-quality-monitor-stations are incomplete, and there exist lots of missing labels. In Logistic Regression avoids data latency to its core. Chance of loss or damage of data [1]. Fuzzy logic and Logistic Regression Clustering offers more efficient storage space. Integrating of grouped data may end in failure[2].Svm algorithm competes with the interpolation data information stored are insecure compensation[3].Random Forest the computational cost of training a random forest it quit low to deliver the same level of accuracy but it will benefit from massive amount of data[4]. Hard to know that what kinds of data are the main relevant features for interpolation and prediction, and the key factors for environment departments to prevent and control air pollution. High cost of building and maintaining such a station

2.2:PROPOSED SYSTEM:

The main objective is to monitor the air pollution present in the environment. Image processing is used to detect the pollution in the atmosphere[1]. Binary segmentation algorithm is used to segment the input image. With the help of the input images the pollution is monitored and the ratio factor and the diffusion process are obtained. In existing system, various images of the environment at different levels of smoke emission from vehicles using far infrared camera and a high resolution visible wavelength camera are obtained. These input images are preprocessed. Various Grey level images are obtained for different ranges[5] .Density analysis of pixel is done using brightness ratio test. Finally the noise is estimated. The key idea of the existing system was to investigate a means to identify smoke-exhausting vehicles from the traffic flow. For this purpose, we need to implement smoke detection analysis software (SDAS) that includes an algorithm for image processing. The algorithm is based on image analysis[4] from thermal and visible wavelength cameras. Thus, the proposed system obtains the images from the environment and monitors the pollution through image processing. The

amount of pollution in the environment is obtained from the dialog box which displays the pollution level in the environment. Various steps have been followed to obtain the pollution level in the environment. The various steps involved are obtaining the input image,pre-processing of the input image and edge detection using canny operator.Finally, a dialog box is been displayed to show the level of pollution.

III.SYSTEM ARCHITECHTURE:



IV.MODULES DESCRIPTION:

There are five different modules in our system, the most critical stage is achieving a successfully system and in giving confidence on the new system for the user that it will work efficiently and effectively. Here we discuss about the detailed description of the modules used in our system.

4.1. IMAGE ACQUISITION:

Input image is the image which is which is taken from the environment. The image obtained is used to monitor the pollution in the environment. The input image is obtained from the traffic or the road. It is not possible to obtain a satellite image.

4.2 PREPROCESSING

In this module we convert the RGB image into gray scale images. Then remove the noises from images by using filter techniques. The goal of the filter is to filter out noise that has corrupted image. It is based on a statistical approach. Typical filters are designed for a desired frequency response. Filtering is a nonlinear operation often used in image processing to reduce "salt and pepper" noise. A median filter is more effective than convolution when the goal is to simultaneously reduce noise and preserve edges.

4.3. SEGMENTATION

Edges are boundaries between different textures or the change of intensity. Edge also can be defined as discontinuities in image intensity from one pixel to another. Edge detection uses canny edge detector which is an edge detection operator that uses a multistage algorithm to detect wide range of edges in images. The canny algorithm is adaptable to various environments. Its parameters allow the recognition of edges of differing characteristics depending on the particular requirements of a given implementation.

4.4. FEATURE CLSSFICATION:

After segmentation completed they performed in classification. Segmentation can perform segment the Pollution in four parts. These parts are not clear in clarity so to apply spectrum color to segmented image. These color images define unique parts of Pollution images. Classification images used to detect Pollution.

4.5. POLLUTION PREDICTION:

We can evaluate the performance of proposed algorithm. This algorithm able to perform well perform localize the renal cortex. Provide best accuracy results in pollution segmentation. In our proposed system, provide improved verification rate and Identification rate. Finally easy identify Pollution in segmentation image.

V. CNN Classification:

Convolution Neural Network is one of the artificial neural networks, with its strong adaptability and good at mining data local characteristics. The weight of sharing network structure make it more similar to the biological neural networks, reduce the complexity of the network model[2], a reduction in the number of weights, makes the CNN be applied in various fields of pattern recognition, and achieved very good results. CNN by combining local perception area, sharing the weight, the drop in space or time sampling to make full use of the data itself contains features such as locality, optimize network structure, and to ensure a degree of displacement invariability.

CNN models are applied to feature extraction in text classification [3], and filter with different lengths, which are used to convolve text matrix. Widths of the filters equal to the lengths of word vectors. Then max pooling is employed to operate extractive vectors of every filter. Finally, each filter corresponds to a digit and connects these filters to obtain a vector representing this sentence, on which the final prediction is based. In these model[5] that is used is relatively complicated, in which convolution operation of each layer is followed by a max pooling operation.To obtain appropriate values for testing phase. In order to find the optimum structure [4], the CNN network performance has been analyzed for the optimum number of hidden nodes and epochs. For this situation, the epochs will be set to a firm preset value. Then, the CNN network was trained at the appropriate range of hidden nodes. The number of hidden nodes that have given the best performance is then selected as the optimum hidden nodes.

After that, by fixing the optimum number of hidden nodes, the epochs will be analyzed in a similar way to obtain the optimum number of epochs that can give the highest or best accuracy.

5.1. MAT LAB:

MATLAB is a high-performance language for technical computing. It integrates computation, visualization, and programming in an easy-to-use environment where problems and solutions are expressed in familiar mathematical notation. Typical uses include math and computation, algorithm development, modelling, simulation and prototyping, Data analysis exploration and visualization. MATLAB features a family of application-specific solutions called toolboxes. Very important to most users of

MATLAB, toolboxes allow you to learn and apply specialized technology. Toolboxes are comprehensive collections of MATLAB functions (M-files) that extend the MATLAB environment to solve particular classes of problems. Areas in which toolboxes are available include signal processing, control systems, neural networks, fuzzy logic, wavelets, simulation, and many others.This is a vast collection of computational algorithms ranging from elementary functions like sum, sine, cosine, and complex arithmetic, to more sophisticated functions like matrix inverse, matrix eigenvalues, Bessel functions, and fast Fourier transforms.

5.1.1. MATLAB functions:

A MATLAB "function" is a MATLAB program that performs a sequence of operations specified in a text file (called an m-file because it must be saved with a file extension of *.m). A function accepts one or more MATLAB variables as inputs, operates on them in some way, and then returns one or more MATLAB variables as outputs and may also generate plots, etc. Some functions are Imread()–Reading the image from the graphics file. A= imread(filename, fmt) reads a grayscale or color image from the file specified by the stringfilename. If the file is not in the current folder, or in a folder on the MATLAB path, specify the full pathname.The text string fmt specifies the format of the file by its standard file extension. For example, specify'gif' for Graphics Interchange Format files. To see a list of supported formats, with the file extensions use imformats function. If imread CNNot find a file named filename, it looks for a file named filename.fmt.

The return value A is an array containing the image data. If the file contains a grayscale image, A is an M-by-N array. If the file contains a truecolor image, A is an M-by-N-by-3 array. For TIFF files containing color images that use the CMYK color space, A is an M-by-N-by-4 array. The class of A depends on the bits-per-sample of the image data, rounded to the next byte boundary. For example, imread returns 24-bit color data as an array of uint8 data because the sample size for each color component is 8 bits.[X, map]=imread(...) reads the indexed image in filename into X and its associated colormap into map. Colormap values in the image file are automatically rescaled into the range [0,1].

SYSTEM REQUIREMENTS:

HARDWARE REQUIREMENTS:

Processor 2.6.0 GHZ	: Dual core processor
RAM	: 1GB
Hard disk	: 160 GB
Compact Disk	: 650 Mb
Keyboard	: Standard keyboard
Monitor	: 15 inch color monitor

SOFTWARE REQUIREMENTS:

Operating system 2007, 2008)	:WindowsOS(XP,
IDE	: MAT LAB

VI. Testing Phases:

Testing is the process of detecting errors. Testing performs a very critical role for quality assurance and for ensuring the reliability of software. Testing is the process of executing a program with the intent of finding errors. Stating formally we can

say Testing is the process of executing a program with the intent of finding errors.

1. A successful test is one that uncovers an as yet undiscovered error.
2. A good test case is one that has a high probability of finding errors, if exists.
3. The tests are adequate to detect possibly present error
4. The software more or less confirms to the quality and reliable standards.

6.1. LEVELS OF TESTING:

In order to uncover the error present in different phases we have the concept of levels of testing. The basic levels of testing are discuss about the detailed description of the testing used in our system.

6.1.1UNIT TESTING:

Unit testing focuses verification effort on the smallest unit of software i.e. the module. Using the detailed design and the process specifications testing is done to uncover error within the boundary of the module. All modules must be successful in the unit test before the start of the integration testing begins. Big data Analysis for Real time investigation and classification system for air quality each service through of a module. There are so many modules like Zone separation and data feeding, Gas data acquisition and predicting, clustering and classification, gas data integration and reporting. Each module has been tested by giving different set of inputs (gathering data from sensor nodes, patient's details, etc…). When developing the modules as well as finishing the development so that each module works without any error. The inputs are validated when accepting from user.

6.1.2 INTEGRATION TESTING

To perform a integration testing activity can be considered as testing the design and hence the emphasis on testing module interactions. The Real time investigation and classification system of air quality the main system performed by integrating all the modules. When integrating all the modules checked integration effects working of any of the services by giving different combinations of input with which the four services run perfectly before integration.

6.1.3 SYSTEM TESTING

The software system is tested for this process is the requirements documents, and goal so to see if software meets its requirements. If the Real time investigation and classification system for air quality using Big data Analysis has been tested against requirements of projects and it is checked whether all requirements of project have been satisfied or not.

6.1.4 ACCEPTANCE TESTING

If an Acceptance test is performed with realistic data of the client to documents that the software is working satisfactorily. Testing is to be focused on external behavior of the system: the internal logic of program is not emphasized. The Real time investigation and classification system for air quality have been collected with some data and tested whether project is working correctly or not. Test cases should be selected so that the largest number of attributes of an equivalence class is exercised at

once. Testing phase is an important part of development. It is the process of finding errors and missing operations and also a complete verification to determine whether objectives are met and the user requirements are satisfied.

6.1.5 WHITE BOX TESTING

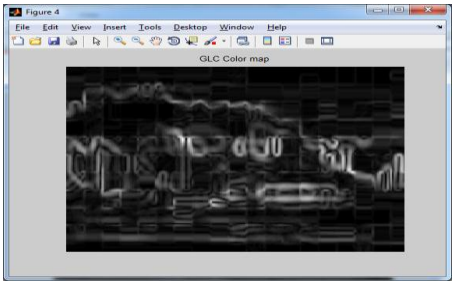
This is a unit testing method a unit will be at a time and tested thoroughly at a statement level2 find the maximum possible errors. The tested step wise every piece of code, taking care that every statement in the code is executed at least once. The white box testing is also called glass box testing. The list of test cases, sample data, which is used to check all possible combinations of execution paths through the code at every module level.

6.1.6 BLACK BOX TESTING

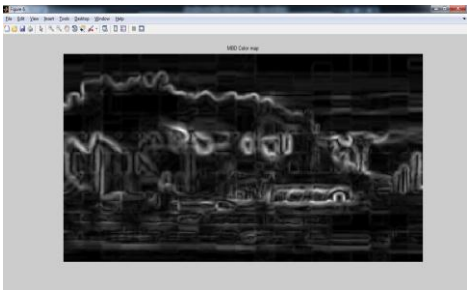
This testing method considers a module as a single unit and checks the unit at interface and communication with other modules rather getting into details as statement level. The module will be treated as black box that will take some input and generate output. Output for a given set of inputs combinations are forwarded to other modules.

a. Input image

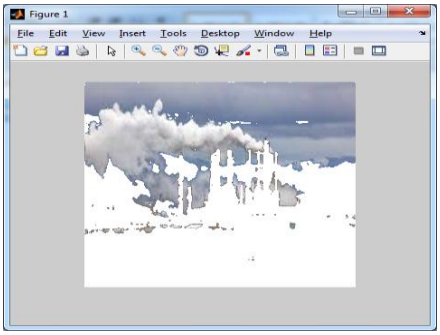
b. Gray scale image



c. GLC colour map



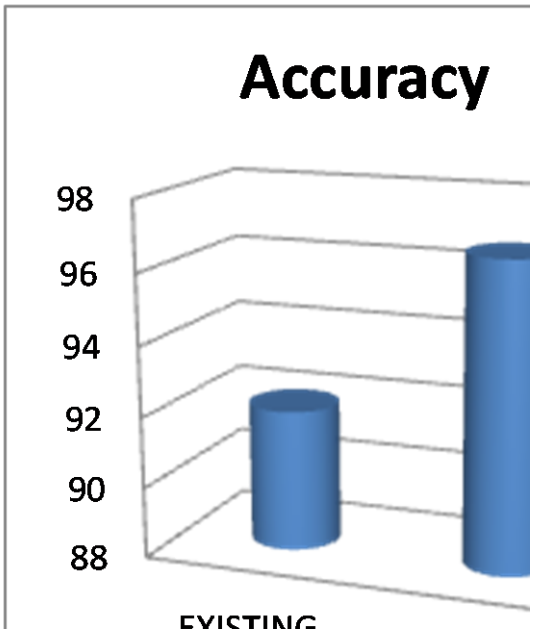
d. GD colour map



e. Air polluted area

VII. PERFORMANCE ANALYSIS:

In this section, the performance evaluation of this work is done to prove the performance improvement over the proposed methodology than the existing system in terms of time and quality of generated rules with maximizing the fitness function. The proposed application concentrates on the CNN classification based on the accuracy value derived from feature values.



CONCLUSION:

In this paper we proposed Deep air learning analyzes the quality of air, by capturing an image in an open atmosphere the captured image will be updated in the dataset. This system can provide crucial information to support air pollution control, and consequently generate great societal and technical impacts. To solving the three problems separately by establishing different models The pixels are pointed in the gray scale conversion and the noise will be removed by using the homomorphism filter. The picture which is already executed time and accuracy with 92% updated will be compared to the new picture by using CNN classification algorithm, which will analyze the quality of air. Example oxygen, carbon-dioxide etc., Air pollution may cause many severe diseases. An efficient air quality monitoring system provides great benefit for human health and air pollution control. The proposed method has been implemented on dataset of CNN classification of time and accuracy analyzed with 96.67% of uses a deep convolution Neural Network (CNN) to classify the natural image in to different categories based on their concentration. The experimental results demonstrate that this method validate the image-based concentration estimation to analyze the quality of air.

REFERENCE:

1. R. Tibshirani, "Regression shrinkage and selection via the lasso," Journal of the Royal Statistical Society. Series B (Methodological), vol. 58, no. 1, pp. 267–288, 1996.

2. M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," Journal of the Royal Statistical Society:Series B (Statistical Methodology), vol. 68, pp. 49–67, 2006.

3. L. Li, X. Zhang, J. Holt, J. Tian, and R. Piltner, "Spatiotemporal interpolation methods for air pollution exposure," in Symposium on Abstraction, Reformulation, and Approximation, 2011.

4. Y. Zheng, F. Liu, and H.-P. Hsieh, "U-air: When urban air quality inference meets big data," in Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ser. KDD '13, 2013.

.