ISSN 2455 – 9091
IJCSET APRIL Volume 9 Issue 4
International Journal of Computer Science Engineering and Technology (IJCSET)
https://www.doi.org/10.5281/zenodo.8382254

Air Pollution Hotspot Detection using Machine Learning and IoT

[1] B. SURYA NARAYANA REDDY
[1] Assistant Professor,
[1] Department of Computer Science Engineering,
[1] St. Martins Engineering College,
[1] Secunderabad, Telangana.

ABSTRACT: Air pollution rate snow a daysare drastically increasing in all the developed and developing countries which require a more portable and cost effective solution. The proposedsystemincludesthedesignformonitoringairpolluti onandcreating awareness among the public .The proposed system is installed in a particular locality where there is acute air pollution. The level of each hazardous pollutant is monitored at periodic intervals. The Air Quality Sensor for the observed pollutants is determined and awareness is created among the public through a proposed system that display sthelevelofeachobservedpollutantandalsotheairqualitysen sorinthatparticularlocation. Thus, the quality of air in that area can be under stood numerical and graphical formats. Further, this system is to be the mselves in an app that pusher's weekly or monthly air quality notification that is more comfortable to access. In this proposed Quality Sensor and Arduino controller to detect air pollution and use machine learning algorithm to predict whether it is an Air Pollution Hotspot or not.

Keywords: [Air Pollution; Air Quality Sensor; Air Pollution Hotspot Prediction; Machine Learning & Classification; IoT; Decision Tree Algorithm; Arduino Controller; and Dataset; etc.]

1. INTRODUCTION

As per World Health Organization [WHO], air pollution isthe infectivity of the indoor or outdoor environment by anychemicalandbiologicalagentwhichchangesthecharacteristic softheenvironment.Householdcombustiondevices,vehicles,and forestfiresarethecommonoriginsofairpollutionandnoisepollutio n.Pollutantsthatareresponsibleforhealth concerns include particulate matter, carbon monoxide, ozone, nitrogendioxide, andsulfurdioxide. Airpollutioncauses respiratory and other diseases, which can be deadly. WHO has measured the quality of air in approximately 1500cities and the Indian capital city was one of the most polluted cities in the world. Pune is having the highest concentration of particulatematterwhichissmallerthan 2.5 micrometers [1].
Airpollutionandlackofairqualitymonitoringpointsrepresentenvi ronmentalandtechnologicalchallengesforcitiesand environments around the world. To face this issue, theindustryhasfocuseditseffortsonfindingaversatiletechnologic alalternativethatallowstheimprovementoftheairquality measuring process and provides reference values innetwork sites where conventional monitoring fails to coverappropriately.Unfortunately,existingproductsandthegener atedresultsdonotrepresentlow-costsolutions.
Ascivilization proceeds,theamount of pollution in allmediums also increases exponentially. Whilst the populationincreases, there are more factories, transport, and fossil fuelconsumptions,andtheamountofairpollutionisatlarge.Accordin g to the 2016 World Health Organization report [1],air pollution in 2012 caused the deaths of around 7 millionpeopleworldwide,anestimateroughlyechoedbytheInternati onalEnergyAgency.India,whichhasthelargestdemocracy andthesecondlargestpopulationinthewholeworld,issufferingverym uchfromcontaminationintheair.
Airpollutionisthecauseof1.2millionprematuredeathsinIndia[6].
ThemainagentsofchaosinairpollutioninIndiaareindustrial pollution (51%), vehicular pollution (27%), and therest caused by crops and waste burning, fireworks, etc. (22%).Already 13 cities in India place in the top 20s in WHO's list ofmost polluted cities. But to reduce pollution, there must be areliable system that can detect pollution. In the case of metrocities like Delhi, Mumbai, Chennai, and Kolkata, the mainsource of pollution is vehicular pollution [4]. So, to reducevehicularpollution,asystemhastobemadethatcandetectthem ainpollutedareasorroutesinacity[7].

2. LITERATURESURVEY

K B Gurumoorthy, S P Vimal, N Sathish Kumar and MKasiselvanathan,"AirPollutionHotspotDetectionandIdentificati onofTheirSourceTrajectory."[1]Thispaperpredicts the accumulation of PM2.5 from wind (velocity anddirection)andprecipitationlevels.Itimbibesamachinelearning(ML)algorithmsupportedbysixyearsofearthscienceandpollutioninf ormationinferences.Atpresent,pollution may be a world downside. The Republic of India is additionally an enormous sufferer of this downside. Thus, it'snecessarytospottherecentspotsofpollutantsandtheirtransportspe cificallycarbonmonoxidegas(CO),sulfurdioxide(SO2),andoxideso felement(NO+NO2)victimizationadvancedinformationanalysistec hniques.Challengesconcernedduringthiscurrentstatementareminin gthedatasetsfromcompletelydifferentparametersandprovidingtheu ltimateoutputwithmoderateabstractionresolutiononpollutioninfo.T herefore,thestudyillustratesthattheemployment of applied mathematics models supported by theML algorithm is most relevant to predict PM2.5 accumulationfromearthscienceinformation.
Soumya deep Sur, Rohit Ghosal, Rittik Mondal," Air Pollution Hotspot Identification and Pollution Level Predictionin the City of Delhi."[2] In this paper, we use various methodsandalgorithmstodetectairpollutionhotspotsandpredictpoll ution levels in a selected area in the city of Delhi. TimeseriesAQIdataiscollectedthroughtheCPCBsensorsinDelhi.
Classification of hotspots is done using SVM, and thetime series analysis based on pollutants like PM2.5, PM10,CO, NO data

samples is done using LSTM and PROPHET.Pollution levels of a day in the future are predicted using thesaidmodels.

RohitAdke, SuyogBachhav, AkashBambale, BhushanWawre,"AirPollutionPredictionusingMachineLearning."[3]The urban air pollution rate has grown at an alarming stateacross India. Most of the cities are facing the issue of poor airqualitywhichfailstomeetstandardsofairforgoodhealth.Itisindeed necessary to develop an air pollution measurement andprediction system for the smart city. Nowadays almost everycity has its prediction system whichuses linear regression.Exploitationbrowsertothevisualizedaccurateresultof prediction. For improving the linear regression algorithm withmaximum accuracy, we are using a neural network. WhichusestheMultilayerPerceptronalgorithm.

Aditya C R, Chandana R Deshmukh, Nayana D K, PraveenGandhi Vidyavastu, "Detection and Prediction of Air PollutionusingMachineLearningModels."[4]Inpopulatedanddeveloping countries, governments consider the regulation ofair as a major task. The meteorological and traffic factors,burningoffossilfuels,andindustrialparameterssuchaspowerplant emissions play significant roles in air pollution. Amongall the particulate matter that determines the quality of the air,Particulate matter (PM 2.5) needs more attention. When itslevel is high in the air, it causes serious issues to people'shealth. Hence, controlling it by constantly keeping a check onitslevelintheairisimportant.Inthispaper,Logisticregressionisemployedtodetectwhetheradatasampleiseitherpolluted or not polluted. Autoregression is employed to predictfuture values of PM2.5 based on the previous PM2.5 readings.Knowledge of the level of PM2.5 in nearing years, month orweek, enables us to reduce its level to a lesser than the harmfulrange. This system attempts to predict PM2.5 levels and detectair quality based on a data set consisting of daily atmosphericconditionsinaspecificcity.

Yawen Zhang, Michael Hannigan, Qin Lv, "Air PollutionHotspot Detection and Source Feature Analysis using Cross-domainUrbanData."[5]Airpollutionisamajorglobalenvironmentalhealththreat,inparticularforpeoplewholiveorworknearpollutionsources.Areasadjacenttopollutionsources often have high ambient pollution concentrations, andthose areas are commonly referred to as air pollution hotspots.Detecting and characterizing pollution hotspots are of greatimportanceforairqualitymanagement,butarechallengingdueto the high spatial and temporal variability of air pollutants. Inthis work, we explore the use of mobile sensing data (i.e., airqualitysensorsinstalledonvehicles)todetectpollutionhotspots. One major challengewith mobile sensing data isuneven sampling, i.e., data collection can vary by both spaceand time. To address this challenge, we propose a two-stepapproach to detect hotspots from mobile sensing data, whichincludes local spike detection and sample-weighted clustering.Essentially,thisapproachtacklestheunevensamplingissuebyweightingsamplesbasedontheirspatialfrequencyandtemporal hit rate, to identify robust and persistent hotspots. Tocontextualizethehotspotsanddiscoverpotentialpollutionsource characteristics, we explore a variety of cross-domainurban data and extract features from them. As a soft validationoftheextractedfeatures,webuildhotspotinferencemodelsforcities with and without mobile sensing data. Evaluation resultsusing real-world mobile sensing air quality data as well ascross-domain urban data demonstrate the electiveness of ourapproachindetectingandinferringpollutionhotspots.Furthermore,theempiricalanalysisofhotspotsandsourcefeaturesyieldsusefulinsightsregardingneighborhoodpollutionsources.

3. PROPOSEDSYSTEM

3.1 Problem Statement:

Air pollution rates nowadays are drastically increasing inall the developed and developing countries which require amore portable and cost-effective solution. In this proposedsystem we design and develop a system with the help of anAir Quality Sensor and Arduino controller to detect airpollution and use a Machine Learning algorithm to predictwhetheritisanAirPollutionHotspotornot.

3.2 Architecture Design:

We propose a two-step approach to automatically detect airpollutionhotspotsfrommobilesensingdata,whichaddresses the uneven sampling issue with mobile sensing.The discovery of hotspots can help in narrowing down theregionsofinterest,whicharefurtherinvestigatedbydomainexperts andregulatoryagencies.

Our analysis of hotspots and source features yields usefulinsights regarding neighborhood pollution sources. Theseinsights may be directly used to complement conventionalpollutioninventories.Asasupplementtoempiricalanalysis, we build hotspot inference models for multiplecities to validate the effectiveness and generalizability oftheextractedfeaturesinrepresentinglocalpollutionsources.

Weevaluateourapproachesusingreal-worldmobilesensingairqualitydatacollectedbyGSVvehiclesinCaliforniaandcross-domainurbandatafromvariousplatforms[8].

Ourmethodologyconsistsofthreekeycomponents:hotspotdetection, source-relatedfeatureextraction,andhotspot inference. Firstly, with mobile sensing air qualitydata,wedetectpollutionhotspotswithatwo-stepapproach[9]. Step1involvesdetectinglocalspikesfromrawobservations to pick out locations with elevated pollutionconcentrations. Step 2 aggregates the results of Step 1 byapplying sample-weighted clustering to identify robust andpersistenthotspots.

Thesample-weightedclusteringtechniquetakesintoaccount the spatial and temporal coverage of each sample.Secondly,withcross-domainurbandata,suchasPOIs,trac,land use, and elevation, we extract source-related featuresand create a feature vector representation for each hotspot.Finally,withfeaturevectorrepresentation,webuildmachine learning (ML) models to infer hotspots in citieswithandwithoutmobilesensingdata.Forcitieswithmobile sensing data, the hotspot detection results are usedaslabelsformodeltraining.Forcitieswithoutmobilesensing data, we perform domain adaption between citiesforhotspotinference[10].

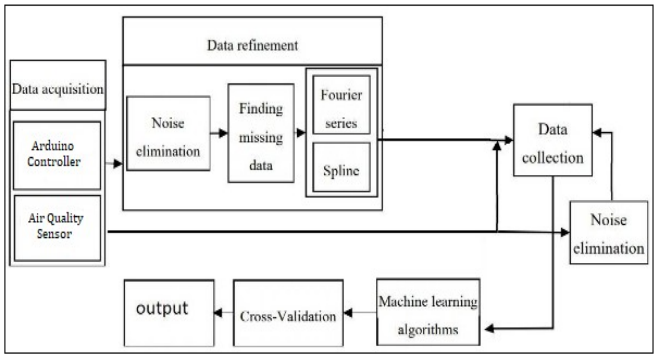

Figure.1: System Architecture

4. RESULTSANDDISCUSSIONS

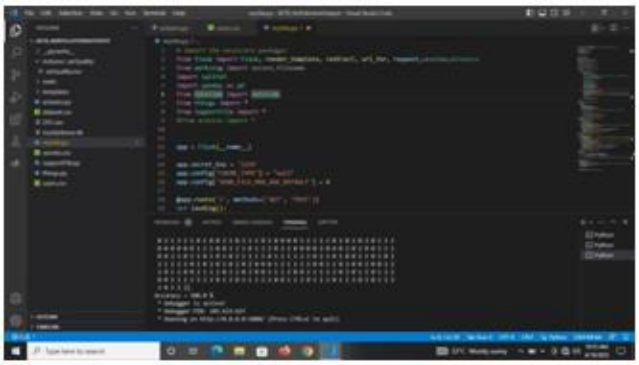

Figure-2Pythoncode

This is our source code for this project in Python. We haveusedseverallibrariessuchasaurdino,metplotlib,etc.Asshow n inthe screenshots

Figure-3PollutionfromIndustries

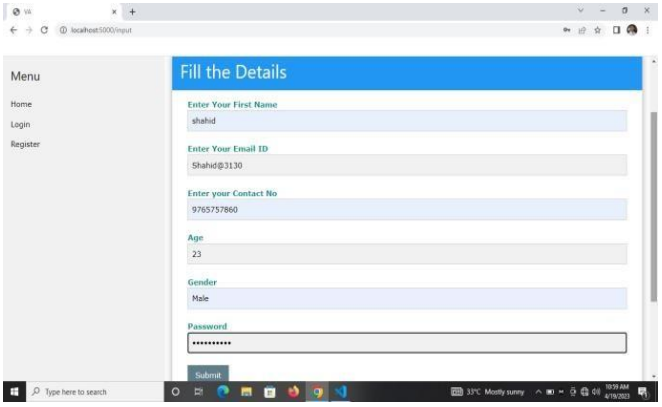

Figure-4MenuDetails

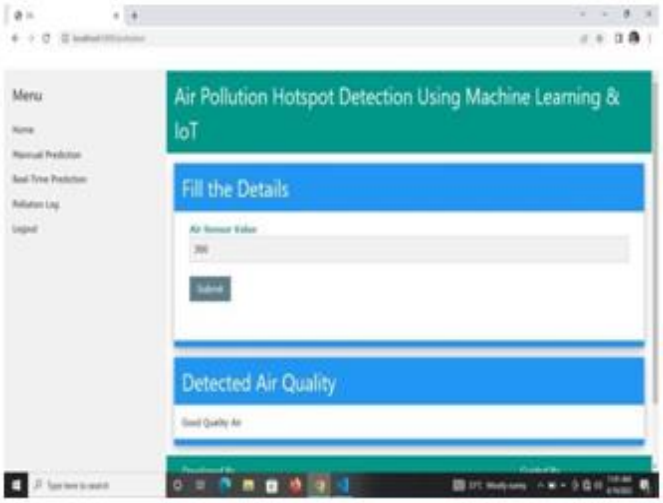

Figure-5 Menu Details with detected Air Quality

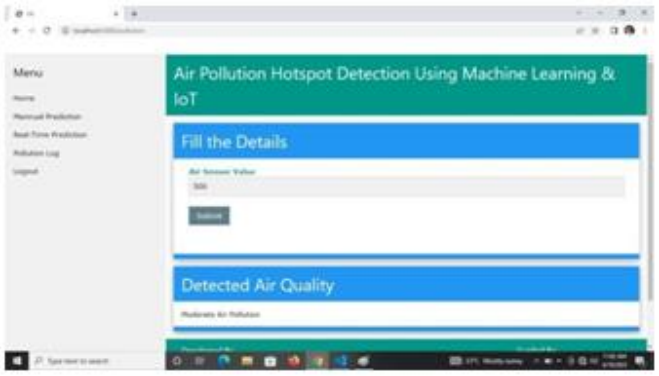

Figure-6MenuDetails withDetectedAirQuality

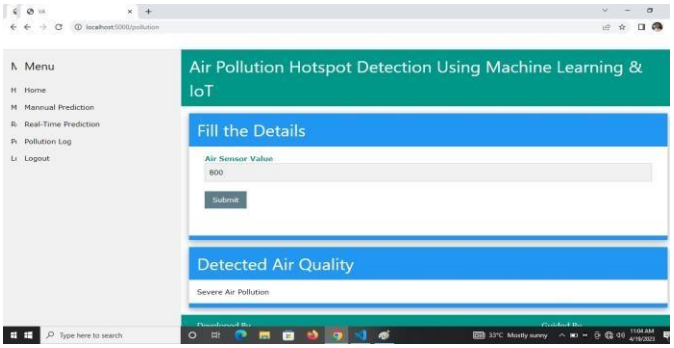

Figure-7MenuDetails withdetectedAirQuality

This figure shows the output for the input given manually tothesystemthatwillclassifythequalityofAirinGood,Moderate and Severe Quality manually. Thiswill help us toget the Air quality of any location just by putting their PMvalueinit.

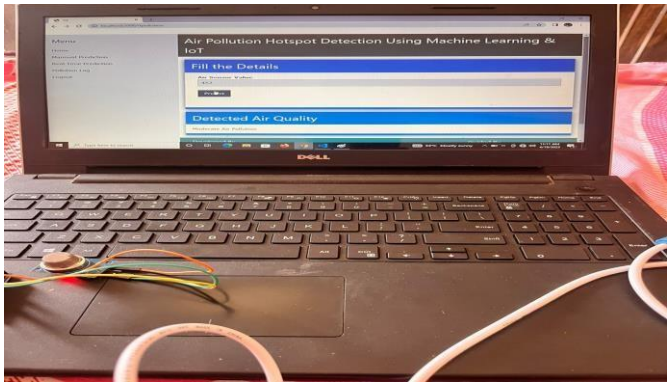

Figure-8HardwareusedinLaptopfordetectionofAirQuality

CONCLUSION

The regulation of air pollutant levels is rapidly becomingone of the most important tasks. It is important that peopleknow what the level of pollution in their surroundings is andtakes a step towards fighting against it. The results show thatmachine learning models (Decision Trees) can be efficientlyusedtodetectthequalityofairandpredictAirPollutionhots pots.Thisworkcanbefurtherextendedbyaddingasimulationmechani smthatcansimulatecongestionandpollution levels in a city. If various important routes can besimulated, we will get an approximate idea about a real-lifescenario&caninferarelationshipbetweenthetrafficpollutionleve lsinthecity.TheprojectonAirPollutionHotspotDetectionusingMach ineLearninghassuccessfullydemonstrated the potential of machine learning techniques toidentify and predict the quality of air in specified area. Byanalyzingdatafromvarioussources,includingairqualitysensorsa

ndsatelliteimagery,theprojecthasdevelopedamodelthatcanaccurately detectandclassifyairpollutionhotspots in real-time.Theresult ofthis study can beusedinform policy decisions aimed at reducing air pollution andimprovingpublichealth. Furthermore, there are several avenues for future researchin the area of air pollution hotspot detection using machinelearning. One possible direction is to explore the use of moreadvanced machine learning algorithms, such as deep learning,,toimprovetheaccuracyandrobustnessofthemodels.Another possible direction is to incorporate more diverse data sources,such as weather patterns and traffic data, to further refine themodels. Additionally, future studies could focus on developingpredictivemodelsthatcanforecastairpollutionhotspots,enablingproactiveinterventionsandpreventivemeasures.Overall, the potential applications of machine learning in airpollution hotspot detection are vast and offer more promisingopportunities for improving public health and environmentalsustainability.

ACKNOWLEDGMENTS

We would prefer to give thanks to the researchers likewisepublishersforcreatingtheirresourcesavailable.Weareconjointly grateful to the guide, and reviewer for their valuablesuggestionsandalsothankthecollegeauthoritiesforprovidingtherequiredinfrastructureandsupport.

REFERENCES

[1]. K B Gurumoorthy, S P Vimal, N Sathish Kumar and MKasiselvanathan,"AirPollutionHotspotDetectionandIdentification of Their Source Trajectory". InternationalJournalofAdvancedResearchinScience,CommunicationandTechnology(IJARSCT)Volume2, Issue6,May2022

[2]. SoumyadeepSur,RohitGhosal,RittikMondal,"AirPollutionHotspotIdentificationandPollutionLevelPredictionintheCityofDelhi".JournalofPhysics:ConferenceSeries, 2019.

[3]. RohitAdke, SuyogBachhav, AkashBambale, BhushanWawre,"AirPollutionPredictionusingMachineLearning".2020IEEEInternationalConferenceforConvergenceinEngineering.

[4]. AdityaCR,ChandanaRDeshmukh,NayanaDK,Praveen Gandhi Vidyavastu, "Detection and Prediction ofAirPollutionusingMachineLearningModels".InternationalJournalofEngineeringTrendsandTechnology(IJETT)–volume59 Issue4– May2018.

[5]. Yawen Zhang, Michael Hannigan, Qin Lv, "Air PollutionHotspotDetectionandSourceFeatureAnalysisusingCross-domainUrbanData". IEEE, 2022

[6]. Jan Kleine Deters, Rasa Zalakeviciute, Mario GonzalezandYvesRybarczyk2017ModelingPM2.5UrbanPollutionUsingMachineLearningandSelectedMeteorologicalParametersJournalofElectricalandComputer Engineering, 5,pp.1-14.

[7]. X. Xi, Z. Wei, R. Xiaoguang et al 2015 A comprehensiveevaluation of air pollution prediction improvement by amachine learning method in the 10th IEEE InternationalConferenceonServiceOperationsandLogistics,andInformatics,pp.176–181,Hammamet,Tunisia.

[8]. Y. Li, Q. Chen, H. Zhao, L. Wang, and R. Tao 2015Variations in pm10, pm2.5 and pm1.0 in an urban area ofthesichuanbasinandtheirrelationtometeorologicalfactorsAtmosphere,6, no.1,pp. 150–163.

[9]. M.Mead,O.Popoola,G.Stewart,P.Landshoff,M.Calleja,M.Hayesb,J.Baldovi,M.McLeod,T.Hodgson, J. Dicks, A. Lewis, J. Cohen, R. Baron, J. Saffell, and R.Jones2013TheUseofElectrochemicalSensorsforMonitoringUrbanAirQualityinLow-Cost,High-DensityNetworksAtmosphericEnvironment,70,pp.186–203.

[10]. Djebbri,NandRouainia,M.2017Artificialneuralnetworks based air pollution monitoring in industrial sitesintheInternationalConferenceonEngineeringandTechnology(ICET),Antalya,Turkey,21–23.