DEEP LEARNING MODEL TO DETECT AND LOCALIZE THE AIR POLLUTION USING SPATIOTEMPORAL DATA

Mrs. C. Sathya [1], V.Adhithyan [2], V.Jeevaprakash [3],M.Sivasakthi [4]
[1]Assistant Professor, Department of CSE, Tagore institute of Engineering and Technology, Deviyakurichi ,Salem, India
[2] UG Students, [3] UG Students, [4] UG Students ,
Department of CSE, Tagore institute of Engineering and Technology, Deviyakurichi ,Salem ,India
Av7370706@gmail.com Jeevaprakash876@gmail.com Sivasakthisakthi000@gmail.com

Abstract

Air pollution caused by the release of toxic gases from novel chemical industries has emerged as a severe threat to human health and environmental sustainability, particularly in densely populated urban areas worldwide. Addressing this escalating crisis requires effective strategies for measuring, predicting, and mitigating air quality challenges. However, traditional air quality prediction methods, such as time series analysis and conventional machine learning models, often struggle to accurately capture the intricate non-linear patterns inherent in air pollution data, including PM2.5 concentrations. To address these challenges, this project proposes a cutting edge multi-point deep learning framework based on Convolutional Long Short-Term Memory (ConvLSTM) networks for dynamic air quality forecasting. By integrating the strengths of Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks, ConvLSTM effectively captures both spatial and temporal data features, delivering highly accurate predictions of air quality metrics. The model is designed to forecast outputs from multiple monitoring nodes simultaneously within a unified framework, ensuring scalability and efficiency. The system includes real-time monitoring of toxic gas emissions from chemical industries, with automated alerts sent to the Pollution Control Board for immediate action. A cloud-based server manages and analyses the collected data, offering a centralized interface for detailed review. Integration with Google Map API provides intuitive visualization of pollution traces, empowering decision-makers with actionable location-based insights. By combining

advanced deep learning techniques with real-time monitoring and regulatory alerts, this solution enhances air quality management and contributes to safeguarding public health and the environment.

Introduction

OVERVIEW

Air pollution can be defined as the presence of toxic chemicals or compounds (including those of biological origin) in the air, at levels that pose a health risk. Air pollution is a mix of hazardous substances from both human-made and natural sources. In an even broader sense, air pollution means the presence of chemicals or compounds in the air which are usually not present and which lower the quality of the air or cause detrimental changes to the quality of life (such as the damaging of the ozone layer or causing global warming). Pollutants in the air take many forms. They can be gases, solid particles, or liquid droplets.

Pollutants in the air take many forms. They can be gases, solid particles, or liquid droplets.

Types of Pollutants

Air pollution can be classified into two sections – invisible and visible air pollution.

- Visible air pollution

As the name suggests, can be visible. The smog you see over a city is an example of visible pollution.

- Invisible air pollutants

These are less noticeable, but they can be more deadly. Good examples of invisible air pollutants are sulphur dioxide, carbon monoxide and nitrogen oxides. It can be further divided into Primarily and Secondary air pollutants.

Primarily air pollutants

It can be caused by primary sources or secondary sources. The pollutants that are a direct result of the process can be called primary pollutants. A classic example of a primary pollutant would be the sulphur-dioxide emitted from factories.

Secondary pollutants

These are the ones that are caused by the intermingling and reactions of primary pollutants. Smog created by the interactions of several primary pollutants is known as a secondary pollutant.

Sources of Air Pollution There are two types of sources that we will take a look, namely Natural sources and Man-made sources.

- Natural Sources

Natural sources of pollution include dust carried by the wind from locations with very little or no green cover, gases released from the body processes of living beings (Carbon dioxide from humans during respiration, Methane from cattle during digestion, Oxygen from plants during Photosynthesis). Smoke from the combustion of various inflammable objects, volcanic eruptions, 1 etc. along with the emission of polluted gases also makes it to the list of natural sources of pollution.

- Man-made Sources

It can be further divided into: Outdoor pollution sources, Indoor pollution sources. Outdoor Pollution Sources: The major outdoor pollution sources include power generation, vehicles, agriculture/waste incineration, industry and building heating systems. Smoke features as a prominent component. The smoke emitted from various forms of combustion, like in biomass, factories, vehicles, furnaces, etc. Waste dumped in landfills generates methane, which is harmful in several ways. The reactions of certain gases and chemicals also form harmful fumes that can be dangerous to the well-being of living creatures. Indoor Pollution Sources In low- and middle-income countries, mostly burning fuels such as dung, coal and wood in inefficient stoves or open hearths produces a variety of health-damaging pollutants. These include carbon monoxide, methane, particulate matter (PM), polyaromatic hydrocarbons (PAH) and volatile organic compounds (VOC). Even burning kerosene in simple wick lamps also produces significant emissions of fine particles and other pollutants. Exposure to smoke from cooking fires causes 3.8 million premature deaths each year.

Various Causes of Air pollution

- The Burning of Fossil Fuels

Sulphur dioxide emitted from the combustion of fossil fuels like coal, petroleum for energy in power plants, and other factory combustibles is one the major cause of air pollution. Billions of vehicles run on roads are powered by gasoline and diesel engines that burn petroleum for releasing energy. Petroleum is made up of hydrocarbons, and engines don't burn them cleanly. As a result, pollutants such as PM, nitric oxide and NO2 (together referred to as NOx), carbon monoxide, organic compounds, and lead emit from vehicles including trucks, jeeps, cars, trains, airplanes, causing a high level of pollution. These modes of transportation form part of our daily basic needs, so we rely on them heavily.

- Industrial Emission

Industrial activities emit several pollutants in the air that affects the air quality more than we can even imagine. Particulate matter 2.5 and 10, Nitrogen dioxide, Sulphur dioxide, and carbon monoxide are key pollutants that are emitted from industries that use coal and wood as their primary energy source for production of their goods. Industrial pollution effects associated with your health can range from irritation in your eyes and throat to breathing issues, at times can even lead to chronic illness.

- Wildfires

Climate change is not just increasing wildfire but also spiking air pollution. Burning stubble and farm residue is also a major contribution to wildfire. It causes increased PM2.5 in the air which collides with other harmful substances like chemical gas and pollen creating smog. Smog makes the air hazy and people find it difficult to breathe

- Microbial Decaying Process

Manufacturing, chemical, and textiles industries release a large number of carbon monoxides, hydrocarbons, chemicals and organic compounds which contaminate our environment. Bacteria and fungi play a fundamental role in the biogeochemical cycles in nature. They are the key indicators of abnormal environmental conditions. Decaying of these microorganisms present in the surrounding releases methane gas which is highly toxic. Breathing toxic gas like methane may lead to death.

- Open Burning of Garbage Waste

Exposure to open burning of garbage waste can pose serious health risk including cancer, liver issues, impairment of immune system, reproductive functions; can also affect the developing nervous system.

- Construction and Demolition

With the rise of population in the city, construction and demolition is a part of the ever-going development phase of the national capital. Several construction sites and raw materials such as bricks and concrete cause haze and foul air which is hazardous for the people especially, children and elderly citizens.

- Agricultural Activities

Agricultural activities have had a serious impact on the decreasing air quality. To begin with pesticides and fertilizers are the main source to contaminate the surrounding air.

- Use of chemical and synthetic products

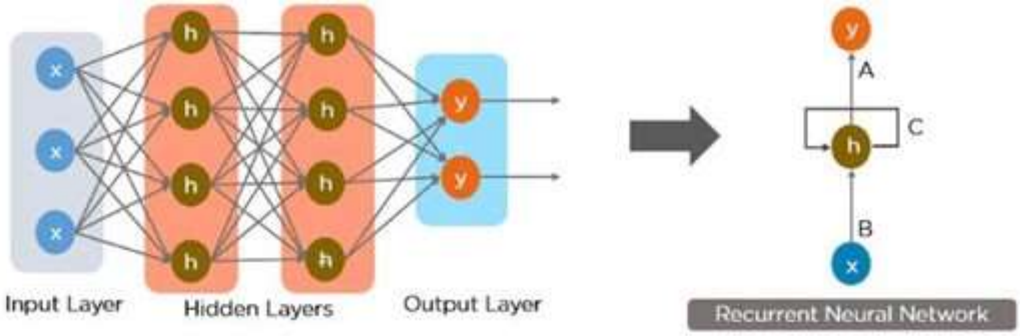
Household products cause indoor air pollution which is 10 times more harmful than outdoor air pollution. Volatile Organic Compounds (VOCs) found in paints, cleaners and personal care products such as perfume and deodorants are a reason for common heath issues.

Proposed System

In this project, we propose a deep learning model based on convolutional long short term memory (ConvLSTM) for highly dynamic air quality forecasting. ConvLSTM architectures combines long short term memory (LSTM) and convolutional neural network (CNN), which allows to mine both temporal and spatial data features. The motivation behind this choice is to benefit from the pattern recognition of Convolution Networks and the memory properties of pure LSTM networks.

- *Recurrent Neural Network*

RNNs are a type of neural network that can be used to model sequence data. RNNs, which are formed from feedforward networks, are similar to human brains in their behaviour. Simply said, recurrent neural networks can anticipate sequential data in a way that other algorithms can't. RNN works on the principle of saving the output of a particular layer and feeding this back to the input in order to predict the output of the layer. Below is how you can convert a Feed-Forward Neural Network into a Recurrent Neural Network:

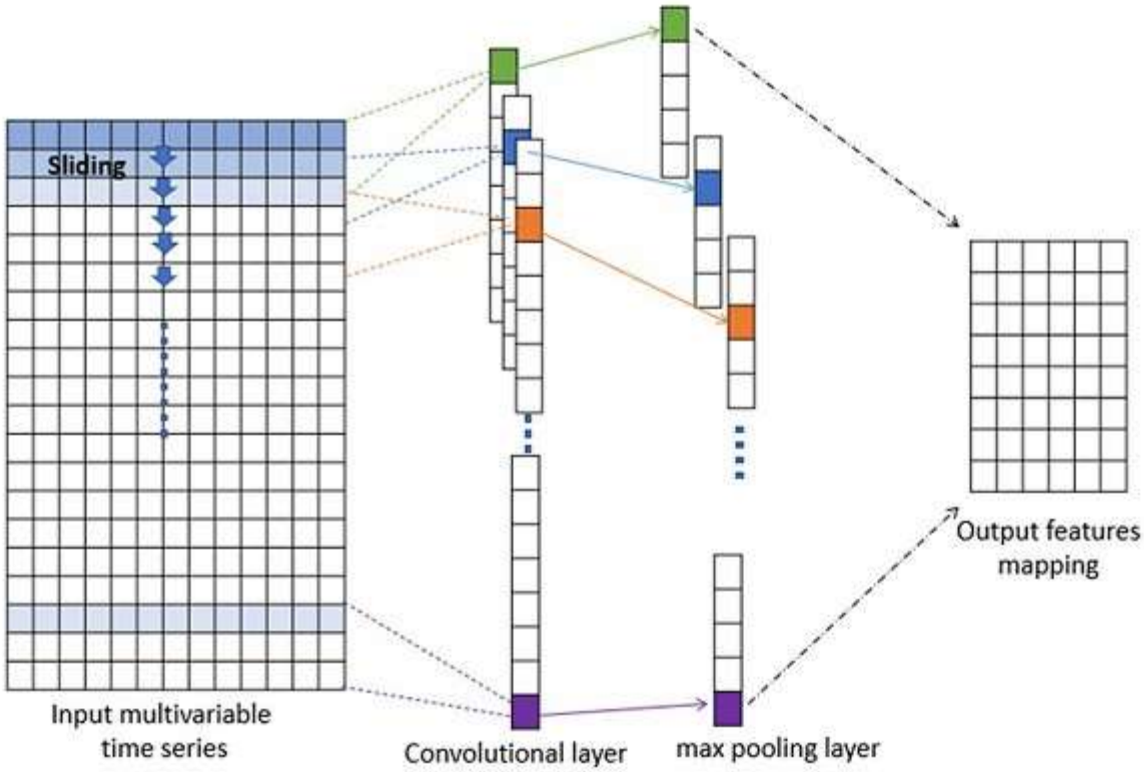


Simple RNN

All of the inputs and outputs in standard neural networks are independent of one another, however in some circumstances, such as when predicting the next word of a phrase, the prior words are necessary, and so the previous words must be remembered. As a result, RNN was created, which used a Hidden Layer to overcome the problem. The most important component of RNN is the Hidden state, which remembers specific information about a sequence. RNNs have a Memory that stores all information about the calculations. It employs the same settings for each input since it produces the same outcome by performing the same task on all inputs or hidden layers.

- *CNN*

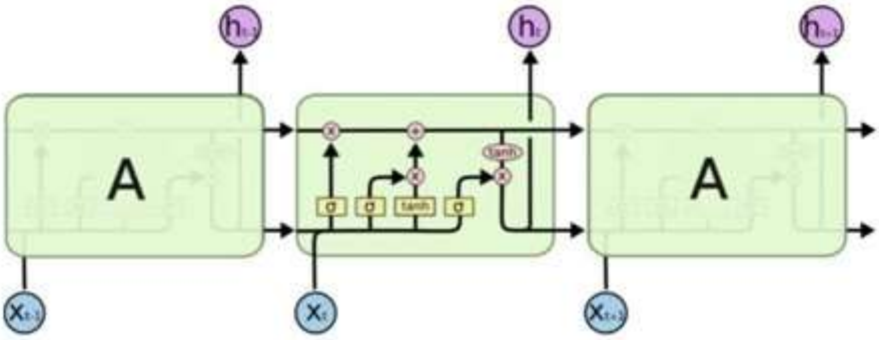
CNN has been successfully applied to computer vision and medical image analysis. Moreover, in this project proposes a multiscale fully convolutional neural network (MFCN) for change detection in high-resolution remote sensing images. In our model, the convolutional layers are constructed using one-dimensional kernels that move through the sequence (unlike images where 2D convolutions are utilized). These kernels act as filters that are learned during training. As in many CNN architectures, the deeper the layers get, the higher the number of filters.



Architecture of CNN

• *LSTM – Long Short-Term Memory LSTMS are a special kind of RNN which is capable of learning long-term dependencies. LSTMs are designed to dodge long-term dependency problem as they are capable of remembering information for longer periods of time. Long short-term memory (LSTM) units (or blocks) are a building unit for layers of a recurrent neural network (RNN). A RNN composed of LSTM units is often called an LSTM network. A common LSTM unit is composed of a cell, an input gate, an output gate and a forget gate. The cell is responsible for "remembering" values over arbitrary time intervals; hence the word "memory" in LSTM. Each of the three gates can be thought of as a "conventional" artificial neuron, as in a multi-layer (or feedforward) neural network: that is, they compute an activation (using an activation function) of a weighted sum. Intuitively, they can be thought as regulators of the flow of values that goes through the connections of the LSTM; hence the denotation "gate". There are connections between these gates and the cell. The expression long short-term refers to the fact that LSTM is a model for the short-term memory which can last for a long period of time. An LSTM is well-suited to classify, process and predict time series given time lags of unknown size and duration between important events. LSTMs were developed to deal with the exploding and vanishing gradient problem when training traditional RNNs.*

Real-time Alert System Module



The popularity of LSTM is due to the Getting mechanism involved with each LSTM cell. In a normal RNN cell, the input at the time stamp and hidden state from the previous time step is 17 passed through the activation layer to obtain a new state. Whereas in LSTM the process is slightly complex, as you can see in the above architecture at each time it takes input from three different states like the current input state, the short-term memory from the previous cell and lastly the long-term memory. These cells use the gates to regulate the information to be kept or discarded at loop operation before passing on the long term and short-term information to the next cell. We can imagine these gates as Filters that remove unwanted selected and irrelevant information. There are a total of three gates that LSTM uses as Input Gate, Forget Gate, and Output Gate.

Input Gate

The input gate decides what information will be stored in long term memory.

It only works with the information from the current input and short-term memory from the previous step. At this gate, it filters out the information from variables that are not useful.

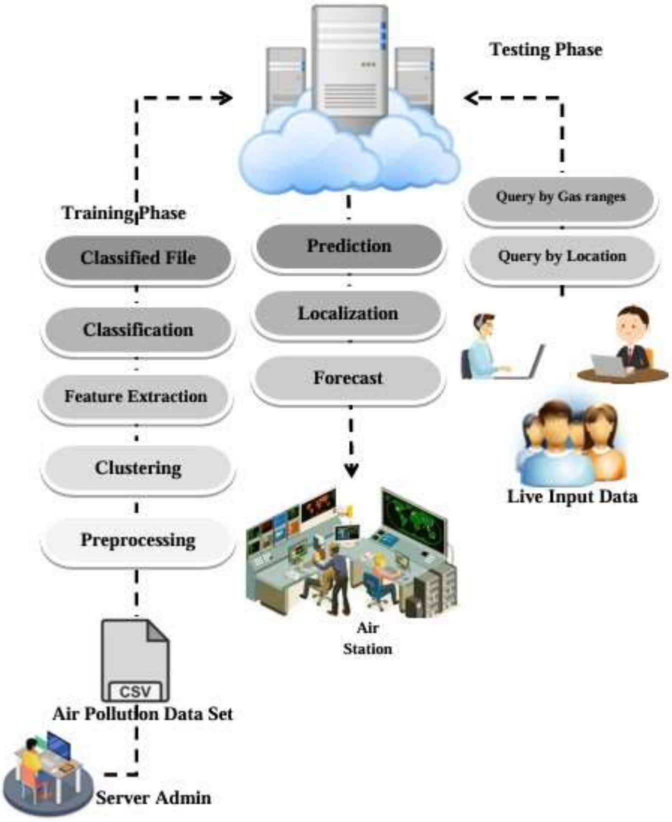
Forget Gate

The forget decides which information from long term memory be kept or discarded and this is done by multiplying the incoming long-term memory by a forget vector generated by the current input and incoming short memory.

Output Gate

The ouGtput gate will take the current input, the previous short-term memory and newly computed long-term memory to produce new short-term memory which will be passed on to the cell in the next time step. The output of the current time step can also be drawn from this hidden state.

System Architecture



Conclusion

In conclusion, the proposed system for predicting and localizing toxic emissions from novel chemical industries offers an innovative and integrated solution for addressing urban air quality challenges. The development focuses on using advanced deep learning techniques, particularly the ConvLSTM model, to predict air quality accurately by capturing both spatial and temporal pollution patterns. The system's ability to monitor toxic gas emissions in real-time, combined with cloud-based data analysis and Google Maps API integration, ensures proactive management and informed decision-making. The automated alert mechanism for the Pollution Control Board enhances regulatory response times, helping to mitigate the health risks associated with industrial emissions. Additionally, the scalability and efficiency of the system make it adaptable to various urban settings, allowing for broad implementation. This project successfully combines machine learning, real-time monitoring, and interactive visualization, providing a comprehensive solution for urban air quality management. Its impact on public health and environmental sustainability is significant, as it supports timely interventions and long-term improvements in air quality. As the system continues to evolve, it has the potential to contribute significantly to urban planning and environmental policy, creating healthier, more sustainable cities.

Future Enhancement

The future scope of the project focuses on refining the predictive and localization capabilities of the system for better urban air quality management.

Integration of Additional Environmental Data: Incorporating data from additional sources like traffic, weather patterns, and industrial activities for more accurate pollution prediction.

Advanced Machine Learning Models: Implementing newer deep learning models and techniques for enhanced prediction accuracy and real-time forecasting.

Public Health Correlation: Integrating health-related data to predict and mitigate the impact of air pollution on public health, offering actionable insights for intervention.

Mobile Application for Public Awareness: Developing a mobile app to provide real-time pollution alerts and public health warnings to citizens.

Expansion to Other Urban Areas: Scaling the system to monitor and predict air quality in multiple cities, enhancing the system's utility on a global scale.

Reference

1. S. Li, G. Xie, J. Ren, L. Guo, Y. Yang, and X. Xu, ``Urban PM2.5 concentration prediction via attention-based CNNLSTM," Appl. Sci., vol. 10, no. 6, p. 1953, Mar. 2020.
2. M. A. Fekih, I. Mokhtari, W. Bechkit, Y. Belbaki, and H. Rivano, ``On the regression and assimilation for air quality mapping using dense low-cost wsn," in Proc. Int. Conf. Adv. Inf. Netw. Appl. Cham, Switzerland: Springer, 2020, pp. 566-578.
3. M. Yang, H. Fan, and K. Zhao, ``PM2.5 predictions with a novel multistep ahead forecasting model based on dynamic wind field distance," Int. J. Environ. Res. Public Health, vol. 16, no. 22, p. 4482, Nov. 2019.
4. J. Ma, J. C. P. Cheng, C. Lin, Y. Tan, and J. Zhang, ``Improving air quality prediction accuracy at larger temporal resolutions using deep learning and transfer learning techniques," Atmos. Environ., vol. 214, Oct. 2019, Art. no. 116885.
5. S. M. Cabaneros, J. K. Calautit, and B. R. Hughes, ``A review of artificial neural network models for ambient air pollution prediction," Environ. Model. Softw., vol. 119, pp. 285 304, Sep. 2019.
6. B. Liu, S. Yan, J. Li, G. Qu, Y. Li, J. Lang, and R. Gu, ``A sequence-to sequence air quality predictor based on the n-step recurrent prediction," IEEE Access, vol. 7, pp. 43331-43345, 2019.

7. Q. Tao, F. Liu, Y. Li, and D. Sidorov, ``Air pollution forecasting using a deep learning model based on 1D convnets and bidirectional GRU," IEEE Access, vol. 7, pp. 76690 76698, 2019.

8. C.-J. Huang and P.-H. Kuo, ``A deep CNN-LSTM model for particulate matter (PM2.5) forecasting in smart cities," Sensors, vol. 18, no. 7, p. 2220, Jul. 2018.

9. P.-W. Soh, J.-W. Chang, and J.-W. Huang, ``Adaptive deep learning-based air quality prediction model using the most relevant spatial-temporal relations," IEEE Access, vol. 6, pp. 38186-38199, 2018.

10. W. Sun and J. Sun, ``Daily PM 2.5 concentration prediction based on principal component analysis and LSSVM optimized by cuckoo search algorithm," J. Environ. Manage., vol. 188, pp. 144-152, Mar. 2017.