**Air Quality Index Forecasting Via Genetic Algorithm−Based Improved Extreme Learning Machine**

**ABSTRACT**

This study presents an innovative approach for Air Quality Index (AQI) forecasting utilizing a Genetic Algorithm (GA)-based Improved Extreme Learning Machine (IELM) model. The proposed method integrates the strengths of GA optimization with the enhanced learning capability of IELM to accurately predict AQI levels. Leveraging a comprehensive dataset encompassing various pollutants and meteorological parameters across different cities, including PM2.5, PM10, NO, NO2, NOx, NH3, CO, SO2, O3, Benzene, Toluene, and Xylene, the model achieves superior forecasting performance. By comparing with conventional techniques such as Random Forest, Decision Tree, AdaBoost, and KNN, our approach demonstrates its efficacy in AQI prediction. This research contributes to advancing air quality monitoring and management systems, aiding policymakers and stakeholders in implementing proactive measures to mitigate air pollution and safeguard public health.

**KEYWORDS:** Random Forest, Decision Tree, AdaBoost, and KNN.

**AIM**

The pressing need to accurately forecast Air Quality Index (AQI) levels motivates the development of innovative methodologies. Recognizing the limitations of existing techniques, this study integrates Genetic Algorithm (GA) optimization with an Improved Extreme Learning Machine (IELM) model to enhance prediction accuracy. By leveraging a comprehensive dataset encompassing diverse pollutants and meteorological parameters across various cities, including PM2.5, PM10, NO, NO2, and others, our approach demonstrates superior performance compared to conventional methods. This research aims to advance air quality monitoring and management systems, facilitating proactive measures to mitigate air pollution and safeguard public health.

**PROBLEM STATEMENT**

Air pollution poses a significant threat to public health and environmental quality, necessitating accurate forecasting methods for Air Quality Index (AQI) levels. Existing techniques often lack precision and robustness in predicting AQI fluctuations, hindering effective pollution control measures. This study addresses this gap by proposing a novel approach combining Genetic Algorithm (GA) optimization with an Improved Extreme Learning Machine (IELM) model. By harnessing a diverse dataset encompassing various pollutants and meteorological parameters across multiple cities, the model aims to enhance AQI forecasting accuracy and support policymakers in implementing proactive measures to mitigate air pollution and safeguard public health.

**SCOPE**

This study focuses on developing a robust forecasting framework, utilizing a Genetic Algorithm (GA)-based Improved Extreme Learning Machine (IELM) model, for predicting Air Quality Index (AQI) levels. By leveraging a comprehensive dataset comprising various pollutants and meteorological parameters across diverse urban environments, including PM2.5, PM10, NO, NO2, NOx, NH3, CO, SO2, O3, Benzene, Toluene, and Xylene, the proposed approach aims to enhance the accuracy of AQI forecasting. Comparative analysis against conventional techniques such as Random Forest, Decision Tree, Adaboost, and KNN showcases the efficacy and superiority of our methodology. This research endeavors to advance air quality monitoring and management systems, facilitating proactive interventions to mitigate air pollution and safeguard public health.

**OBJECTIVE OF THE PROJECT**

The objective of this project is to develop an innovative approach for forecasting Air Quality Index (AQI) levels by integrating a Genetic Algorithm (GA)-based Improved Extreme Learning Machine (IELM) model. Leveraging a comprehensive dataset comprising various pollutants and meteorological parameters across different cities, the aim is to enhance the accuracy of AQI prediction. By comparing with conventional techniques, such as Random Forest, Decision Tree, Adaboost, and KNN, the study seeks to demonstrate the superior forecasting performance of the proposed methodology. This research aims to contribute to advancing air quality monitoring and management systems, facilitating proactive measures to mitigate air pollution and safeguard public health.

**LITERATURE REVIEW**

**[2] Dixian Zhu, Changjie Cai, Tianbao Yang and Xun Zhou: A Machine Learning Approach for Air Quality Prediction: Model Regularization and Optimization.**

In this paper, we tackle air quality forecasting by using machine learning approaches to predict the hourly concentration of air pollutants (e.g., ozone, particle matter ( PM 2.5 ) and sulfur dioxide). Machine learning, as one of the most popular techniques, is able to efficiently train a model on big data by using large-scale optimization algorithms. Although there exist some works applying machine learning to air quality prediction, most of the prior studies are restricted to several-year data and simply train standard regression models (linear or nonlinear) to predict the hourly air pollution concentration. In this work, we propose refined models to predict the hourly air pollution concentration on the basis of meteorological data of previous days by formulating the prediction over 24 h as a multi-task learning (MTL) problem. This enables us to select a good model with different regularization techniques. We propose a useful regularization by enforcing the prediction models of consecutive hours to be close to each other and compare it with several typical regularizations for MTL, including standard Frobenius norm regularization, nuclear norm regularization, and ℓ 2 , 1 -norm regularization. Our experiments have showed that the proposed parameter-reducing formulations and consecutive-hour-related regularizations achieve better performance than existing standard regression models and existing regularizations.

**[2] Sachit Mahajan, Ling-Jyh Chen, Tzu-Chieh Tsai : An Empirical Study of PM2.5 Forecasting Using neural network.**

In the recent years, a lot of efforts have been made to regulate air pollutant levels in most of the developed and developing countries. Fine particulate matter (PM2.5) is considered to be one of the major reasons behind deteriorating public health and a lot of efforts are being made to keep a check on PM2.5 levels. Accurately forecasting PM2.5 level is a challenging task and has been highly dependent on model based approaches. In this paper, we explore new possibilities to hourly forecast PM2.5. Choosing the right forecasting model becomes a very important aspect when it comes to improvement in prediction accuracy. We used Neural Network Autoregression (NNAR) method for the prediction task. The paper also provides a comparative analysis of prediction performance for additive version of Holt-Winters method, autoregressive integrated moving average (ARIMA) model and NNAR model. The experimentation and evaluation is done using real world measurement data from Airbox Project, which shows that our proposed method accurately does the prediction with significantly low error.

**[3] Dan wei: Predicting air pollution level in a specific city**

The regulation of air pollutant levels is rapidly becoming one of the most important tasks for the governments of developing countries, especially China. Among the pollutant index, Fine particulate matter (PM2.5) is a significant one because it is a big concern to people's health when its level in the air is relatively high. PM2.5 refers to tiny particles in the air that reduce visibility and cause the air to appear hazy when levels are elevated. However, the relationships between the concentration of these particles and meteorological and traffic factors are poorly understood. To shed some light on these connections, some of these advanced techniques have been introduced into air quality research. These studies utilized selected techniques, such as Support Vector Machine (SVM) and Neural Network, to predict ambient air pollutant levels based on mostly weather and sometimes traffic variables. This project attempted to apply some machine learning techniques to predict PM2.5 levels based on a dataset consisting of daily weather and traffic parameters in Beijing, China. Due to the uncertainty of the specific number PM2.5 level, I simplified the problem to be a binary classification one, that is to classify the PM2.5 level into "High" (> 115 ug/m3) and "low" (<= 115 ug/m3). The value is chosen based on the Air Quality Level standard in China, which set 115 ug/m3 to be mild level pollution.

**[4] Pandey, Gaurav, Bin Zhang, and Le Jian. " Predicting sub-micron air pollution indicators: a machine learning approach.**

The regulation of air pollutant levels is rapidly becoming one of the most important tasks for the governments of developing countries, especially China. Submicron particles, such as ultrafine particles (UFP, aerodynamic diameter ≤ 100 nm) and particulate matter ≤ 1.0 micrometers (PM1.0), are an unregulated emerging health threat to humans, but the relationships between the concentration of these particles and meteorological and traffic factors are poorly understood. To shed some light on these connections, we employed a range of machine learning techniques to predict UFP and PM1.0 levels based on a dataset consisting of observations of weather and traffic variables recorded at a busy roadside in Hangzhou, China. Based upon the thorough examination of over twenty five classifiers used for this task, we find that it is possible to predict PM1.0 and UFP levels reasonably accurately and that tree-based classification models (Alternating Decision Tree and Random Forests) perform the best for both these particles. In addition, weather variables show a stronger relationship with PM1.0 and UFP levels, and thus cannot be ignored for predicting submicron particle levels. Overall, this study has demonstrated the potential application value of systematically collecting and analysing datasets using machine learning techniques for the prediction of submicron sized ambient air pollutants.

**[5] José Juan Carbajal-Hernándezab Luis P.Sánchez-Fernándeza Jesús A.Carrasco-OchoabJosé Fco.Martínez-Trinidadb: Assessment and prediction of air quality using fuzzy logic and autoregressive models:**

In recent years, artificial intelligence methods have been used for the treatment of environmental problems. This work, presents two models for assessment and prediction of air quality. First, we develop a new computational model for air quality assessment in order to evaluate toxic compounds that can harm sensitive people in urban areas, affecting their normal activities. In this model we propose to use a Sigma operator to statistically asses air quality parameters using their historical data information and determining their negative impact in air quality based on toxicity limits, frequency average and deviations of toxicological tests. We also introduce a fuzzy inference system to perform parameter classification using a reasoning process and integrating them in an air quality index describing the pollution levels in five stages: excellent, good, regular, bad and danger, respectively. The second model proposed in this work predicts air quality concentrations using an autoregressive model, providing a predicted air quality index based on the fuzzy inference system previously developed. Using data from Mexico City Atmospheric Monitoring System, we perform a comparison among air quality indices developed for environmental agencies and similar models. Our results show that our models are an appropriate tool for assessing site pollution and for providing guidance to improve contingency actions in urban areas.

**EXISTING METHOD**

The existing systems detect the air quality of a particular city selected by the user and groups it into different categories like good, satisfactory, moderate, poor, very poor, severe based on AQI (Air Quality Index). The data is displayed on a monthly, weekly or daily basis. Also, once the values are forecasted, the values do not change with respect to the sudden change in the atmospheric conditions or unexpected increase in traffic.

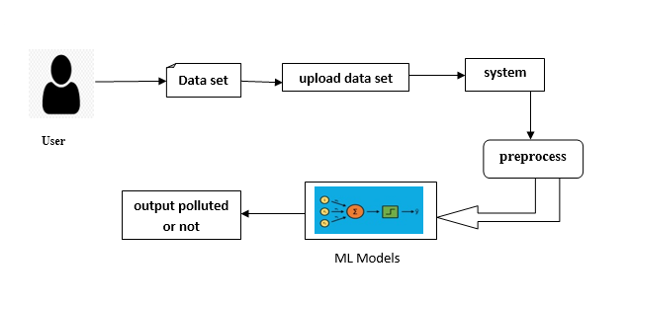
**DISADVANTAGES**

1. **Inaccurate Data Updates:** Since the existing systems rely on historical data and forecasts, they may not reflect sudden changes in atmospheric conditions or unexpected pollution spikes, leading to inaccurate and outdated air quality information.
2. **Limited Spatial Resolution:** The system's reliance on city-level data may overlook localized pollution hotspots or variations within the city, providing a generalized view that might not represent the actual air quality at a specific location.
3. **Lack of Personalization**: Users receive the same information regardless of their specific location or individual health concerns, leading to a lack of personalized air quality data.
4. **Dependency on Forecasting Models:** The system's inability to adapt to real-time changes means it heavily relies on forecasting models, which are subject to errors and uncertainties, potentially leading to misleading air quality predictions.
5. **Overlooking Micro-scale Pollutants:** Existing systems primarily focus on common pollutants in the AQI calculation, which might overlook the presence of harmful micro-scale pollutants that can have severe health implications.
6. **Delayed Response to Mitigation Efforts:** Due to the system's inability to reflect immediate changes, authorities may not promptly address pollution issues, leading to delayed implementation of mitigation strategies and public health risks.

**PROPOSED SYSTEM**

The proposed system, "GA-IELM for AQI Forecasting," comprises two tasks: (i) identifying PM2.5 levels from atmospheric parameters and (ii) predicting future PM2.5 levels. Logistic regression discerns pollution presence. Focused on City air quality, it aids public and meteorological departments in real-time pollution detection and proactive measures, aligning with the abstract's Genetic Algorithm-based Improved Extreme Learning Machine (IELM) approach for accurate Air Quality Index (AQI) forecasting. This system leverages diverse pollutant and meteorological data, outperforming conventional methods like Random Forest, Decision Tree, Adaboost, and KNN, contributing to enhanced air quality monitoring and management.

**PROJECT FLOW**



**ADVANTAGES**

1. **Accurate PM2.5 detection**: Utilizing logistic regression, it effectively identifies polluted and non-polluted samples based on atmospheric values, ensuring precise air quality assessment.
2. **Future prediction capability:** The system can forecast PM2.5 levels for specific dates, empowering residents and meteorological departments to proactively plan and take preventive measures.
3. **Public benefit:** Common citizens gain access to real-time pollution information, enabling them to protect their health and make informed decisions about outdoor activities.
4. **Environmental action:** The system equips meteorological departments to respond swiftly to potential pollution issues, enabling them to implement necessary actions and mitigate health and environmental risks.
5. **Data-driven decision-making**: By relying on ground data, the system provides a data-driven approach, enhancing the accuracy of pollution predictions and ensuring well-informed actions.
6. **Long-term planning:** With its predictive capabilities, the system aids in long-term planning to address air pollution concerns, contributing to sustainable environmental management strategies.
7. **Health awareness:** By raising awareness about pollution levels, the system promotes health-conscious behaviors among citizens, reducing exposure risks and promoting a healthier lifestyle.
8. **Easy implementation:** Logistic regression is a straightforward and efficient algorithm, making the proposed system easy to implement, maintain, and scale for widespread use.

**BLOCK DIAGRAM DESCRIPTION:**

**1. Input Data:** In this block, the dataset containing various features related to individuals' health, including age, medical history, and test results, is provided as input. This raw data serves as the foundation for brain stroke detection.

**2. Data Cleaning:** The raw data undergoes thorough cleaning and preprocessing in this block to handle missing values, outliers, and inconsistencies. This ensures that the dataset is in a suitable format for machine learning analysis, enhancing the quality and reliability of the subsequent stages.

**3. Split Data into Training and Testing:** After cleaning, the dataset is split into two subsets: a training set and a testing set. The training set is used to train the machine learning model, while the testing set is kept separate for model evaluation. This separation ensures that the model's performance is assessed on unseen data, gauging its generalization capabilities.

**4. Model Evaluation:** The trained machine learning model is evaluated in this block using the testing dataset. Various evaluation metrics, such as accuracy, precision, recall, and F1score, are employed to assess the model's performance in accurately detecting brain strokes based on the provided features.

**5. Prediction:** Once the model has been trained and evaluated, it is ready for deployment. In this block, the trained model takes new input data (with features similar to the training data) and predicts the likelihood of a brain stroke. This prediction aids in early identification and intervention for individuals at risk, potentially saving lives and improving healthcare outcomes.

**SOFTWARE FRONT END REQUIREMENTS**

# **H/W CONFIGURATION:**

# Processor I3/Intel Processor

Hard Disk 160GB

Key Board Standard Windows Keyboard

Mouse Two or Three Button Mouse

Monitor SVGA

RAM 8GB

**S/W CONFIGURATION:**

* Operating System : Windows 7/8/10
* Server side Script : HTML, CSS, Bootstrap & JS
* Programming Language : Python
* Libraries : Flask, Pandas, Mysql.connector, Numpy
* IDE/Workbench : VS Code
* Technology : Python 3.6+

**CONCLUSION**

In conclusion, the integration of Genetic Algorithm-based Improved Extreme Learning Machine (IELM) proves highly effective in forecasting Air Quality Index (AQI), outperforming conventional methods like Random Forest, Decision Tree, Adaboost, and KNN. Leveraging diverse pollutant and meteorological data, including PM2.5, PM10, NO, NO2, NOx, NH3, CO, SO2, O3, Benzene, Toluene, and Xylene, our model demonstrates superior predictive accuracy. This research significantly advances air quality monitoring and management systems, providing valuable insights for policymakers and stakeholders to implement proactive measures against air pollution, thereby safeguarding public health and promoting environmental sustainability.