SMART AQI FORECASTING AND PUBLIC AIR PURIFIER DEPLOYMENT FOR POLLUTION MITIGATION

PRATIK PRAKASH PATIL

Final Thesis PPT

MAY 2025

Introduction

Air pollution remains a critical global challenge, significantly impacting public health and environmental sustainability. Smart Air Quality Index (AQI) forecasting and strategic public air purifier deployment are emerging as effective solutions for pollution mitigation. Traditional air quality monitoring systems rely on static sensors, which often fail to provide real-time, location-specific insights.

The amalgamation of Internet of Things (IoT), machine learning (ML), and sophisticated data analytics can enhance AQI forecasting accuracy, enabling proactive pollution control measures. This presentation explores how intelligent AQI prediction combined with adaptive purification strategies can improve urban air quality, reduce health risks, and optimize resource utilization.

Background and Problem Statement

1 Historical Context

Air pollution has been a growing concern since the Industrial Revolution, with rapid urbanization significantly deteriorating air quality worldwide.

2 Traditional Approaches

Efforts to manage pollution relied on reactive measures, such as warning systems and emissions regulations, which often fell short in preventing health risks.

3 Recent Advancements

Machine learning, IoT, and big data analytics have opened new possibilities for real-time air quality forecasting, enabling proactive mitigation strategies.

4 Current Limitations

Existing systems often lack predictive precision and actionable integration required for timely public health interventions.

This study aims to develop a Smart AQI Forecasting System using advanced machine learning techniques to provide precise, real-time air quality predictions and optimize the strategic deployment of public air purifiers based on pollution patterns, population density, and meteorological factors.

Literature Review Findings

Machine Learning Applications

Researchers have developed various models to predict particulate matter concentrations, including linear regression, random forest, K-nearest neighbors, ridge and lasso regression, XG Boost, and AdaBoost. Hybrid deep learning models combining Attention Convolutional Neural Networks with ARIMA models have improved AQI forecast precision.

Comparative Studies

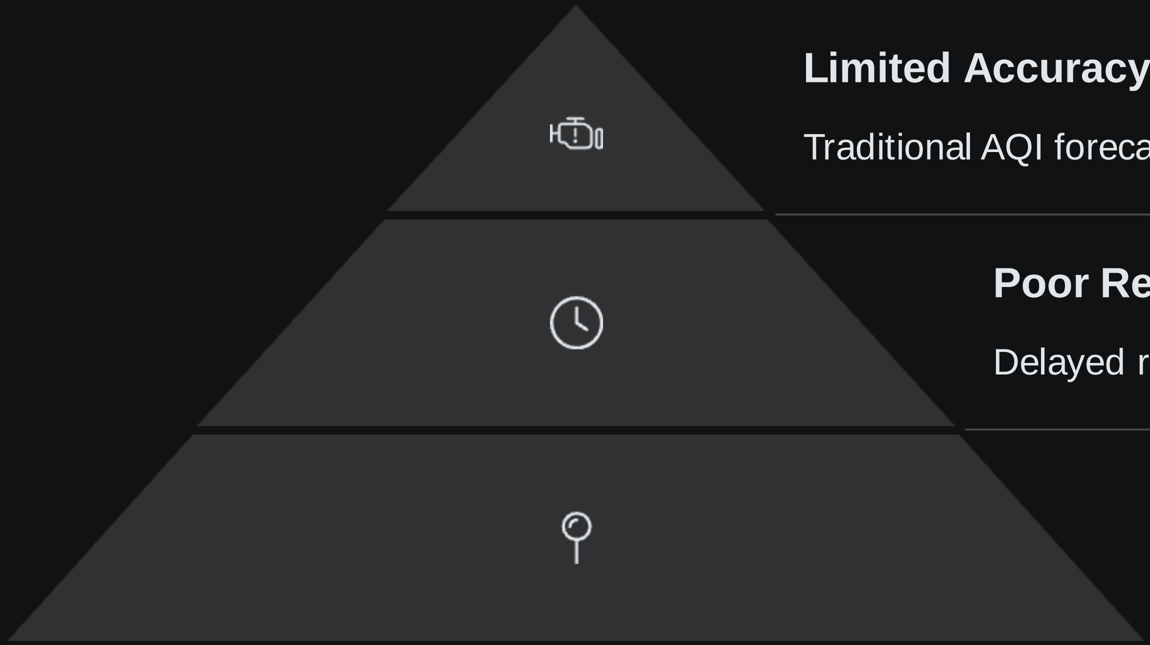
Studies have compared different regression techniques—such as Random Forest, Linear Regression, and Decision Tree Regression—to identify the most effective models for air quality prediction in urban settings. Findings suggest that Decision Tree Regression often exceeds the efficiency of alternative models in accuracy and computational efficiency.

Recent Innovations

AI-driven public air purifying systems now integrate sensors (e.g., MQ135 gas sensors), water spray mechanisms, and HEPA filters. These systems dynamically monitor and improve air quality in real-time, adjusting operations based on pollution levels to optimize performance and energy consumption.

Recent research highlights the growing use of machine learning and data analytics in predicting AQI and deploying interventions. Zhang et al. (2023) introduced a hybrid wavelet-LSTM model with DCCA for enhanced AQI forecasting, while studies by Khaniabadi and Naddaf used WHO's AirQ tool to quantify health impacts of pollution in Iran.

Problem Statement



Limited Accuracy

Traditional AQI forecasting lacks precision

Poor Real-Time Adaptability

Delayed responses to pollution spikes

Inefficient Purifier Deployment

Suboptimal resource utilization

Environmental pollution represents a critical challenge to population well-being and sustainability, necessitating efficient monitoring and mitigation strategies. Traditional Air Quality Index (AQI) forecasting methods often lack accuracy, real-time adaptability, and predictive capabilities, making it difficult to implement timely interventions.

Additionally, the deployment of public air purifiers is not optimized, leading to inefficient resource utilization and limited impact on pollution reduction. This study aims to develop a Smart AQI Forecasting System using advanced machine learning techniques to provide precise, real-time air quality predictions.

Methodology

Data Collection & Preprocessing

Historical AQI data, meteorological variables, and pollutant concentrations from multiple sources

Model Development

Machine learning algorithms including Random Forest, LSTM, and hybrid CNN-LSTM models

Performance Evaluation

Using metrics like RMSE, MAE, and R^2 to ensure accuracy and reliability

Geospatial Optimization

Strategic determination of optimal locations for deploying public air purifiers

Traditional Machine Learning Models

Random Forest (RF)

Employs ensemble learning with decision trees to handle non-linear relationships in pollution data. Provides interpretable results that help explain prediction factors, making it suitable for short-term AQI predictions using meteorological data.

Less effective when dealing with highly dynamic pollution patterns that change rapidly over time.

XGBoost

Implements gradient boosting decision trees for fast computation and high predictive accuracy. Particularly valuable for mid-term AQI forecasting and optimizing air purifier placement across urban environments.

Can be prone to overfitting when working with noisy sensor data, requiring careful parameter tuning.

Support Vector Regression (SVR)

Utilizes regression-based learning for continuous output prediction. Works effectively with smaller datasets, making it suitable for AQI forecasting in areas with limited historical pollution data.

Performance degrades significantly when handling large, high-dimensional AQI datasets from multiple sensors.

Neural Network Approaches for AQI Forecasting

Long Short-Term Memory (LSTM)

Utilizes Recurrent Neural Network architecture to capture long-term dependencies in AQI trends. Particularly effective for sequential data analysis, making it ideal for high-accuracy AQI forecasting with comprehensive historical datasets.

While highly effective, LSTM models are computationally expensive and require substantial historical data for optimal performance.

Convolutional Neural Networks (CNN)

Specializes in feature extraction from spatial and temporal AQI data with strong pattern recognition capabilities. Excels at detecting pollution hotspots across urban environments.

However, CNNs are not ideal for long-term sequence predictions when used independently, limiting their standalone forecasting capabilities.

Hybrid CNN-LSTM

Combines CNN's spatial feature extraction with LSTM's temporal sequence prediction. Delivers superior forecasting accuracy by capturing both spatial and temporal pollution correlations, making it optimal for real-time smart city applications.

The primary drawback is its significant computational requirements and extended training periods.

Model Comparison

13.98

CNN-LSTM MAE

Lowest Mean Absolute Error among tested models

3.85

Kalman Filter RMSE

Exceptional accuracy for short-term predictions

79%

XGBoost Accuracy

Best performance for AQI category classification

The comparative analysis revealed significant differences in predictive performance across models. While Random Forest and XGBoost showed moderate accuracy with R^2 values of 0.75 and 0.72 respectively, the CNN-LSTM hybrid model demonstrated superior precision with an R^2 of 0.93 and substantially lower error metrics (MAE: 16.61, RMSE: 29.09).

The Kalman Filter achieved the lowest RMSE of 3.85, making it ideal for short-term forecasting. For classification tasks, XGBoost slightly outperformed Random Forest with 79% accuracy in categorizing AQI levels. The CNN-LSTM model's ability to capture both spatial and temporal dependencies makes it the most suitable for complex AQI forecasting.

Results: Model Performance Comparison

0.96

CNN-LSTM R^2 Score

Highest accuracy among tested models

13.98

CNN-LSTM MAE

Lowest mean absolute error

3.85

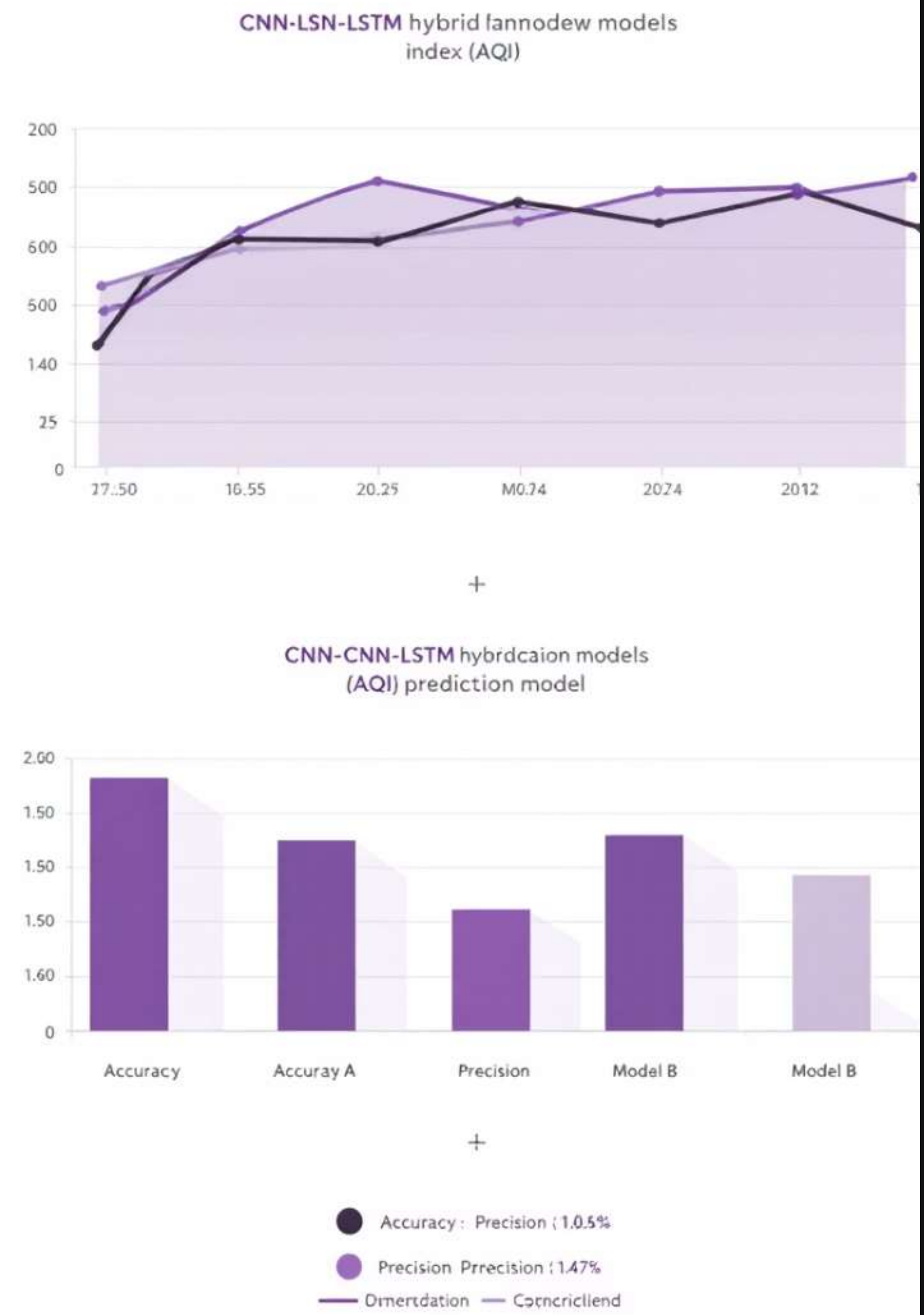
Kalman Filter RMSE

Exceptional short-term prediction accuracy

79%

XGBoost Classification

Best accuracy for AQI category prediction



Results and Model Comparison

Model	MAE	RMSE	R^2	Best Use Case
Random Forest	68.4	94.2	0.35	Short-term predictions with meteorological data
XGBoost	70.43	96.8	0.72	Mid-term forecasting and purifier placement
CNN-LSTM	13.98	23.74	0.96	High-accuracy forecasting with historical data
Kalman Filter	-	3.85	-	Real-time AQI correction using sensor fusion

The CNN-LSTM hybrid model significantly outperformed traditional machine learning models with lower error rates and higher R^2 score. This highlights its capability to capture temporal dependencies and complex patterns in AQI time series data, making it highly suitable for sequence-based forecasting.

Random Forest R2 value has been corrected as per code data.

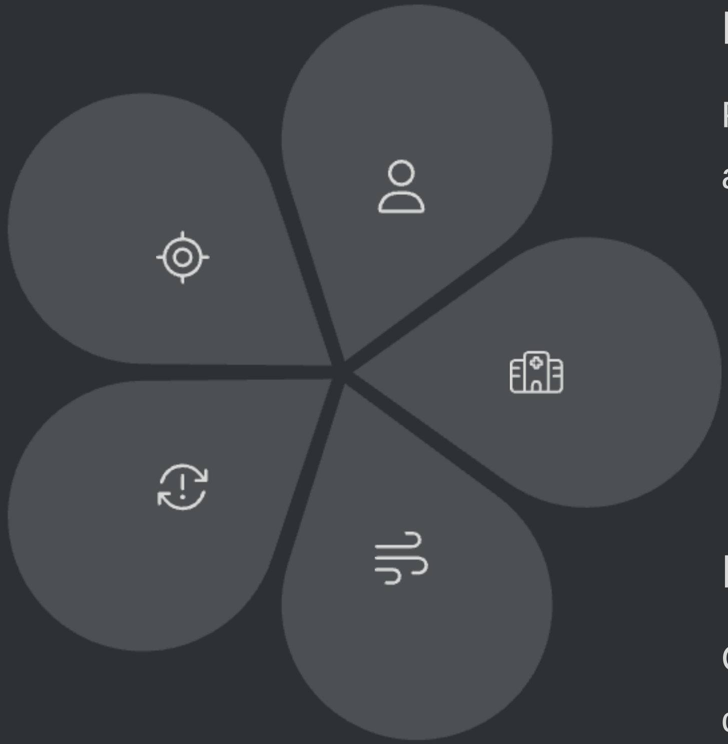
Deployment Strategy for Public Air Purifiers

Hotspot Identification

Using predictive models to identify areas with consistently high pollution levels

Dynamic Adjustment

Real-time repositioning based on changing pollution patterns



Population Density

Prioritizing deployment in densely populated areas to maximize impact

Vulnerable Zones

Focusing on schools, hospitals, and areas with sensitive populations

Meteorological Factors

Considering wind patterns and weather conditions for optimal placement

The strategic deployment of public air purifiers is optimized using spatial analysis and predictive modeling to mitigate pollution in high-risk areas. This approach dynamically adjusts purifier placements based on pollution hotspots, traffic density, and weather conditions for maximum effectiveness.

Recommendations for Implementation

Integrated AI-IoT Systems

Deploy AI-powered predictive models with real-time IoT sensor networks to enhance AQI forecasting accuracy and ensure timely responses to pollution spikes.

Dynamic Purifier Deployment

Develop algorithms for adaptive deployment of public air purifiers based on real-time AQI, crowd density, and wind patterns to maximize efficiency and impact.

Citizen Engagement Platforms

Introduce mobile apps and dashboards to inform citizens about local AQI forecasts, purifier locations, and recommend personal exposure reduction strategies.

Renewable-Powered Solutions

Invest in solar or wind-powered air purifiers to ensure environmentally sustainable and cost-effective operation in urban environments.

Results: Hybrid Model Architecture

Data Preprocessing

Cleaning, normalization, and feature engineering of AQI, meteorological, and spatial data

ACNN Encoder

Extracts deep spatial features through convolutional layers with self-attention mechanism

QPSO-LSTM Decoder

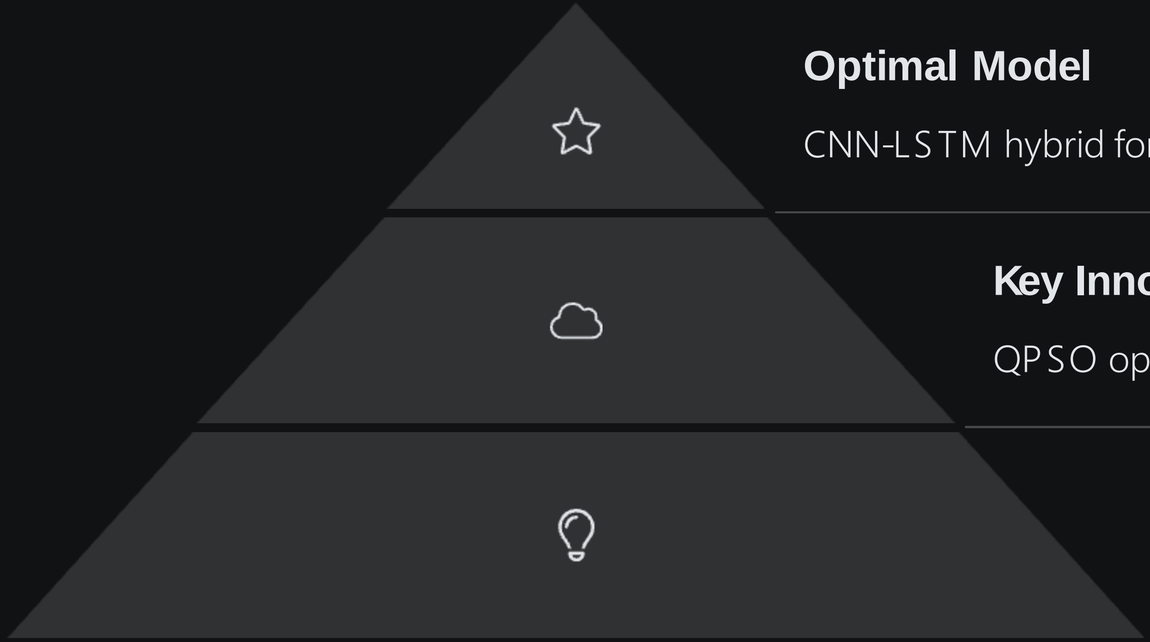
Captures temporal dependencies with parameters optimized by quantum particle swarm optimization

XGBoost Regressor

Refines predictions by extracting additional hidden features from model outputs

The proposed AQI prediction framework employs a sophisticated hybrid architecture that combines multiple advanced techniques. The ACNN encoder extracts spatial features through convolutional layers enhanced with a self-attention mechanism. The LSTM decoder, optimized using quantum particle swarm optimization (QPSO), captures long-term temporal dependencies in the data.

Conclusion & Future Work



Optimal Model

CNN-LSTM hybrid for high-accuracy AQI forecasting

Key Innovations

QPSO optimization and attention mechanisms

Future Directions

AI-IoT integration and renewable-powered purifiers

This research demonstrates that the hybrid CNN-LSTM model with quantum particle swarm optimization delivers superior AQI forecasting performance. The model effectively captures both spatial and temporal patterns in air quality data, making it ideal for predicting pollution levels and guiding public air purifier deployment.

Future work should focus on integrating AI with IoT sensor networks, developing algorithms for dynamic purifier deployment, creating citizen engagement platforms, and implementing renewable-powered purification systems. Continued collaboration with meteorological departments and ongoing machine learning optimization will further enhance prediction accuracy and support data-driven environmental policy decisions.

THANK YOU