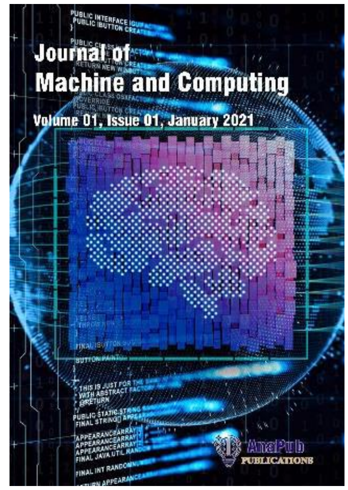
Journal Pre-proof

Adaptive Deep Learning Strategies for Formaldehyde Monitoring in Industrial Air Quality



Kishore Kunal, Pillalamarri Lavanya, Leena Nesamani S, Kathiravan M, Parthasarthy K and Vairavel Madeshwaren

DOI: 10.53759/7669/jmc202505135

Reference: JMC202505135

Journal: Journal of Machine and Computing.

Received 19 March 2025

Revised from 10 April 2025

Accepted 09 May 2025

Please cite this article as: Kishore Kunal, Pillalamarri Lavanya, Leena Nesamani S, Kathiravan M, Parthasarthy K and Vairavel Madeshwaren, "Adaptive Deep Learning Strategies for Formaldehyde Monitoring in Industrial Air Quality", Journal of Machine and Computing. (2025). Doi: https:// doi.org/10.53759/7669/jmc202505135.

Adaptive Deep Learning Strategies for Formaldehyde Monitoring in Industrial Air Quality

Kishore Kunal[1*], Pillalamarri Lavanya[2] , S. Leena Nesamani[3] , M.Kathiravan[4] , K Parthasarthy[5] ,Vairavel Madeswaren

[1*]Professor of Business Analytics, Loyola Institute of Business Administration, Chennai , TamilNadu , India – 600034.
Corresponding Email: kishore.kunal@liba.edu , ORCID : https://orcid.org/0[illegible]003-4[illegible]4-690X

[2]Assistant Professor, Department of Physics & Electronics, Bhavan's Vivekananda College of Science, Humanities and Commerce, Hyderabad Telangana,500010

Email: lavanya.elec@bhavansvc.ac.in , ORCID :https://orcid.org/0009-00[illegible]-9689-8984

3Associate professor, Department of Computer Applications, Faculty of Science and Humanities,
SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu, India
Email: leenan@srmist.edu.in,ORCID : https://orcid.org/0009-0008-3299-6961

[4]Department of Computer science and Engineering ,Saveetha School of Engineering,Saveetha Institute of Medical and Technical Sciences (SIMATS), Chennai, India
Email: kathiravanm.sse@saveetha.com ORCID :https://orcid.org/0000-0002-5377-7871

[5]Assocsite Professor, Department of Electrical and Electronics Engineering ,Jawaharalal Nehru University, New Delhi , India-670310.

Email: partee.1979@nav[illegible]gov.in , ORCID : https://orcid.org/ 0000-0002-5247-4668

[6]Associate professor, Department of Agriculture Engineering, Dhanalakshmi Srinivasan College of Engineering, Coimbatore, TamilNadu , India – 600034.
Email: vai[illegible]@d[illegible]e.ac.in ,ORCID : https://orcid.org/0000-0002-8687-7887

Abstract-Inhaling formaldehyde a chemical that is widely used in many different industries can have serious health consequences. In order to precisely detect formaldehyde levels in industrial air quality environments, this study makes use of deep learning techniques. Using sensor data gathered from high-risk industrial areas the study focuses on variables like air quality index temperature and humidity. The data is processed by Convolutional Neural Networks (CNNs), which identify trends linked to increases in formaldehyde concentrations. To improve model accuracy preprocessing of the data is done including feature scaling and outlier elimination. The model's performance is assessed using evaluation metrics like Mean Squared Error (MSE), sensitivity, specificity, and prediction accuracy. Results show that when compared to conventional regression models the CNN-based model considerably lowers false positives while achieving a high prediction accuracy. Rapid reaction to hazardous formaldehyde levels is made possible by the deep learning frameworks' real-time monitoring capability which lowers possible health hazards. To improve long-term prediction accuracy and trend identification future research will investigate the use of recurrent neural networks (RNN) for time-series analysis.

Keywords: Formaldehyde, Deep Learning, Air Quality Monitoring, CNN, Industrial Safety, Health Risks

1. INTRODUCTION

Formaldehyde is a commonly used industrial chemical, particularly in sectors like textiles, plastics, wood products, and adhesives. Its toxicity poses significant concerns for air quality in industrial environments, as prolonged exposure can lead to serious health issues, including respiratory problems, skin irritation, and even cancer. Therefore, monitoring formaldehyde levels is crucial for ensuring workplace safety and adherence to regulatory standards. To accurately detect and quantify formaldehyde in industrial air, various advanced techniques have been developed. One prominent method is Fourier-transform infrared (FTIR) spectroscopy, which utilizes infrared light absorption to identify formaldehyde molecules. FTIR is effective for continuous and real-time monitoring, offering high sensitivity and specificity in detecting this compound. Another widely used technique is gas chromatography-mass spectrometry (GC-MS), which excels in separating and identifying volatile organic compounds, including formaldehyde. While GC-MS provides highly detailed analyses, it generally requires periodic sampling and laboratory processing, making it less ideal for continuous monitoring in real-time settings.

Recent advancements in sensor technology have transformed the landscape of formaldehyde monitoring. Electrochemical sensors, appreciated for their portability and cost-effectiveness, are becoming more prevalent in industrial applications. These sensors generate electrical signals that correlate with formaldehyde concentration, allowing for immediate data collection. Photoacoustic spectroscopy (PAS) is another emerging technique, known for its ability to accurately measure low concentrations of formaldehyde. This method involves using a modulated light source that excites formaldehyde molecules, generating a pressure wave detected as an acoustic signal. Additionally, innovations such as IoT-based sensor networks enable remote monitoring and data analytics, greatly improving industrial air quality management. These systems can automatically collect and transmit data to centralized platforms, facilitating trend analysis, early warnings, and compliance with occupational safety regulations. Fig 1 shows the architecture of Formaldehyde Detection in Industry

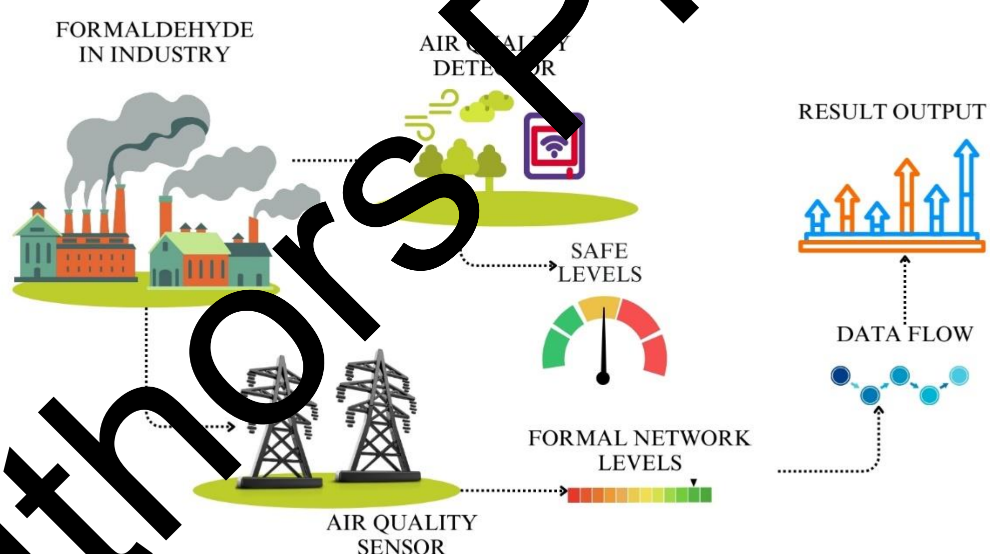


Fig 1. Formaldehyde Detection in Industry

Various strategies to enhance the performance of semiconductor gas sensors specifically for formaldehyde detection have been discussed, highlighting advancements in material design, nanostructuring, and sensor configurations that improve sensitivity and selectivity [1]. The importance of integrating functional materials and optimizing sensor parameters to achieve high-performance detection is emphasized, along with recent trends and future directions in formaldehyde sensor technology. Polymer-based materials used in formaldehyde gas sensors, focusing on their synthesis and application, have been reviewed

[2]. Various polymer composites and their properties that contribute to improved sensitivity and response times are discussed, along with challenges in developing stable and selective sensors and recent advancements in sensor performance. The review suggests potential areas for future research in polymer sensor technology.

Recent advances in metal oxide semiconductor (MOS) materials for formaldehyde detection are explained, including the design and synthesis of various MOS materials, their structural, electrical, and gas-sensing properties [3]. Insights into the mechanisms of formaldehyde adsorption and detection, as well as the influence of doping and nanostructuring on sensor performance, are provided. Future perspectives on enhancing MOS gas sensors are also presented. Advancements in carbon nanotubes (CNTs) as gas sensors, including their unique properties that make them suitable for detecting various gases, including formaldehyde, are discussed [4]. Different functionalization techniques that enhance the sensitivity and selectivity of CNT-based sensors, along with the integration of CNTs into sensor devices and challenges associated with their practical applications, are explored. The review concludes with future trends in CNT sensor development.

Recent advancements in formaldehyde sensors, transitioning from small molecules to polymeric probes, are reviewed, highlighting the development of various sensing mechanisms and materials that improve detection capabilities [5]. The effectiveness of different polymeric materials and their potential for real-world applications are discussed, along with challenges in sensor performance, stability, and selectivity. Suggestions for future research directions are also provided. The recent progress in organic chemosensors for formaldehyde detection is summarized, discussing various organic materials, including small molecules and polymers, that exhibit high sensitivity and selectivity towards formaldehyde [6]. Innovative sensing mechanisms and fabrication techniques used in developing these sensors are highlighted. Future challenges and potential improvements in organic sensor technologies are discussed.

A detailed examination of microfabricated formaldehyde gas sensors is presented, focusing on their design, fabrication processes, and performance metrics [7]. Various microfabrication techniques that enhance sensor sensitivity and miniaturization are described, along with the integration of sensors into electronic devices for real-time monitoring. Future prospects for microfabricated sensor technologies are outlined. The design and optimization strategies of metal oxide semiconductor nanostructures for advanced formaldehyde sensors are explored, analyzing how different nanostructures can influence gas sensing performance and sensitivity [8]. The role of surface modification and doping in enhancing sensor capabilities is emphasized. Future challenges and research directions in metal oxide sensor development are also highlighted.

The use of graphene-based structures for the trace-level detection of gaseous formaldehyde is presented, discussing the unique properties of graphene that contribute to its superior sensing potential, including high surface area and electron mobility [9]. Various sensing mechanisms and the integration of graphene into sensor platforms are highlighted. The review concludes with insights into the future of graphene-based gas sensing technologies. Recent progress in fluorescent probes for detecting carbonyl species, focusing on formaldehyde, is reviewed [10]. Various probe designs and mechanisms for fluorescence detection are discussed, emphasizing their sensitivity and selectivity. Advancements in small molecule probes and their potential applications in environmental monitoring are highlighted. Future directions in probe development and detection strategies are also considered.

Trends in sensors and methods for measuring atmospheric formaldehyde gas concentration, focusing on the patent landscape, are examined [11]. Various sensor technologies, including electrochemical, optical, and semiconductor-based methods, are analyzed, along with advancements and challenges faced by different sensing techniques in real-world applications. The importance of innovation in sensor design for improved performance is highlighted. The development of organic small molecule and functional material fluorescent probes for formaldehyde detection is reviewed, exploring various probe architectures and their fluorescence response mechanisms [12]. Recent advancements in sensitivity, selectivity, and real-time imaging applications are emphasized. Challenges in developing robust and efficient probes are discussed, along with future research opportunities.

Recent progress in fluorescent formaldehyde detection using small molecule probes is discussed, highlighting various chemical strategies that enhance the selectivity and sensitivity of these probes [13]. The importance of molecular design in developing effective fluorescent sensors is emphasized. Future challenges and research directions in fluorescent probe technology for formaldehyde detection are also presented. The use of CdO–ZnO nanorices for enhanced and selective formaldehyde gas sensing applications is investigated, discussing the synthesis of these nanorices and their unique structural properties that contribute to improved sensing performance [14]. The mechanisms of formaldehyde interaction with the nanorices and the factors influencing sensitivity are highlighted. Future prospects for these materials in sensor applications are outlined.

A formaldehyde gas sensor with a remarkable detection limit of 1 ppb based on an In-doped $LaFeO_3$ porous structure is presented, describing the synthesis and characterization of the sensor and highlighting its excellent sensitivity and rapid response [15]. The mechanisms underlying the sensor's performance are discussed, focusing on the role of doping and porous structure. Future research directions for enhancing sensor performance are also suggested. UV-activated semiconductor gas sensor response measurement for formaldehyde detection is investigated, discussing the advantages of using UV activation to enhance sensor sensitivity and response times [16]. The experimental setup and results are detailed, emphasizing the effectiveness of this approach. The review also considers the implications of UV activation for future sensor designs and applications.

Multi-wall carbon nanotube gas sensors modified with amino groups for low-concentration formaldehyde detection are explored, discussing the synthesis and functionalization of CNTs and highlighting their improved sensing capabilities [17]. The mechanisms of gas adsorption and detection are analyzed, showing how modifications enhance performance. Future challenges and the potential for further improvements in CNT-based sensors are also addressed.

II. MATERIALS AND METHODS

2.1 Data Collection

Data from several industrial areas in India that are known to have high formaldehyde usage and air pollution issues were gathered for this study. Two important sites were Ankleshwar which is known for producing chemicals and Gujarats Vapi industrial area which is among the most severely polluted areas because of heavy chemical industries. Additionally information was obtained from Bhopal which focuses on areas surrounding the chemical industries for thorough data on air quality parameters and the Manali Petrochemical Industrial Area in Tamil Nadu which deals extensively with chemical processes. Fig 2 shows the study area of this research.

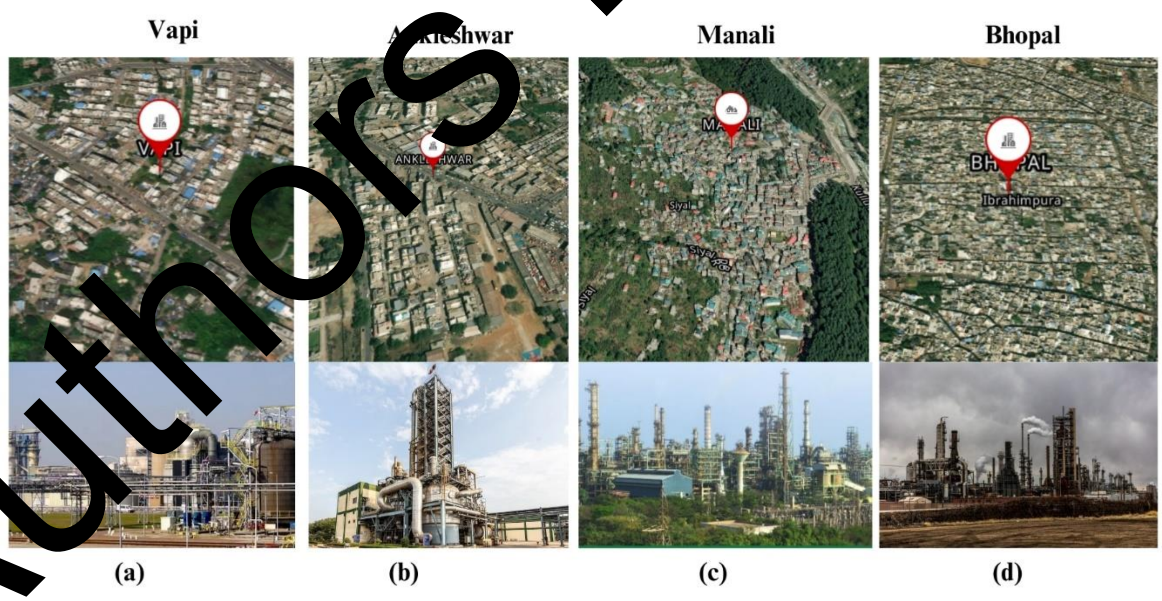


Fig 2 Data Collection locations

High-precision sensors were installed to track formaldehyde levels and real-time data was recorded every minute for six months. The air quality index (AQI) which highlights spikes in formaldehyde concentrations and acts as a stand-in for overall pollution levels was one of the parameters that was continuously monitored along with temperature and humidity. Within each industrial zone, each sensor node was positioned strategically at several points to guarantee both data accuracy and wide coverage.

2.2 Data measurement

To guarantee the validity and dependability of formaldehyde detection, the data needed to be precisely measured after it was gathered. Using parts per billion (ppb) units the high-resolution sensors measured formaldehyde concentrations offering a fine-grained level of measurement for even the smallest changes. A thorough dataset covering different seasons changes in industrial activity and variations in environmental factors was ensured by the raw data sets which contained thousands of data points from various times and conditions. These exacting measurements offered a strong basis for assessing how well the deep-learning model predicted formaldehyde concentrations.

2.3 Data preprocessing

Processing the data was a crucial step in improving the predictive power of the model. Feature scaling was done during the preprocessing stage to standardize all variables and enhance model convergence during training. By scaling temperature humidity AQI and formaldehyde levels to a range between 0 and 1, a Min-Max normalization technique was used to ensure that each feature had an equal impact on the model. Fig 3 demonstrates the molecules of formaldehyde.

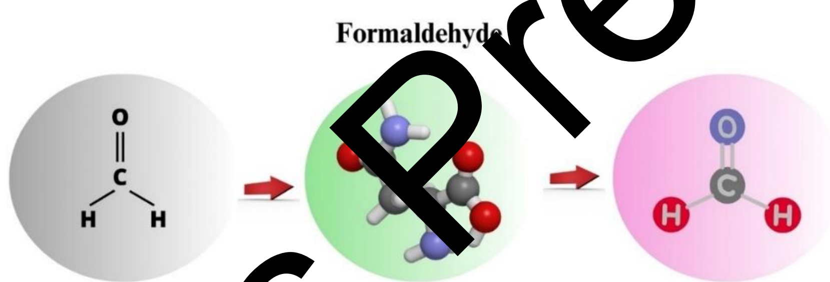


Fig 3. Air pollutant molecules of chemical formaldehyde

The Interquartile Range (IQR) approach was used to find and eliminate outliers. Any data points that were more than 1 or 5 times, the IQR were filtered out in order to remove anomalies that might distort the results. In order to preserve continuity missing data was imputed using linear interpolation taking advantage of patterns in the current dataset. The dataset was artificially expanded using data augmentation techniques such as small changes to temperature and AQI values to replicate real-world conditions and improve training accuracy.

III.PROPOSED TECHNIQUE

3.1 Deep Learning Technique

3.1.1 Convolutional Neural Network

The suggested method uses a Convolutional Neural Network (CNN) architecture which was created especially to find formaldehyde patterns in data on air quality. Each layer of the CNN has three-layer architecture, which consists of convolution pooling and activation functions optimized to detect minute changes in formaldehyde concentrations in the face of shifting environmental conditions. Fig 4 provides the architecture of CNN. The input matrix denoted by X contains variables such as temperature (T) humidity (H) and AQI (A). The convolution process is described as follows in equation (1).

$$F(i,j) = \sum_{m=0}^{M} \sum_{n=0}^{N} X(i+m, j+n) \cdot K(m,n) \tag{1}$$

where X is the input matrix, K is the kernel matrix, M, NM , NM,N are kernel dimensions and F(ij) is the filtered output. This method is used to find local patterns linked to elevated formaldehyde levels. The feature maps are down-sampled using max pooling after convolution which lowers dimensionality while preserving important information.

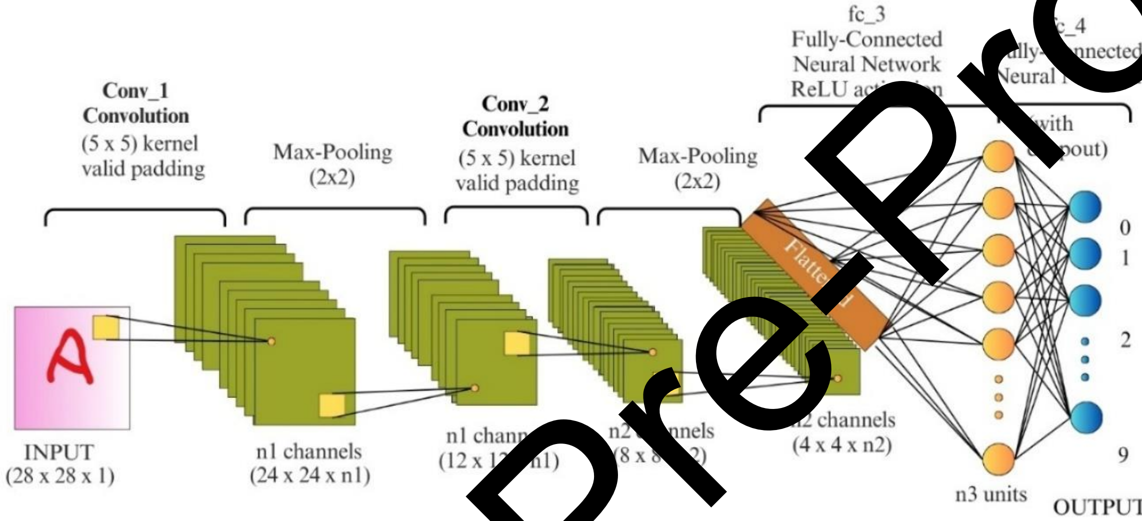


Fig 4 CNN architecture

The Rectified Linear Unit (ReLU) activation function: f(x)=max(0,x)

Several metrics were used to assess the CNN model's performance offering a thorough examination of its predictive power. One important metric of overall model performance was prediction accuracy which quantifies the percentage of correctly predicted data points among all observations. For evaluating the CNN's dependability sensitivity the model's capacity to precisely identify true positives and specificity and the capacity to accurately identify true negatives were also essential. The prediction errors were also quantified using the Mean Squared Error (MSE) where a lower MSE denotes a better model fit.

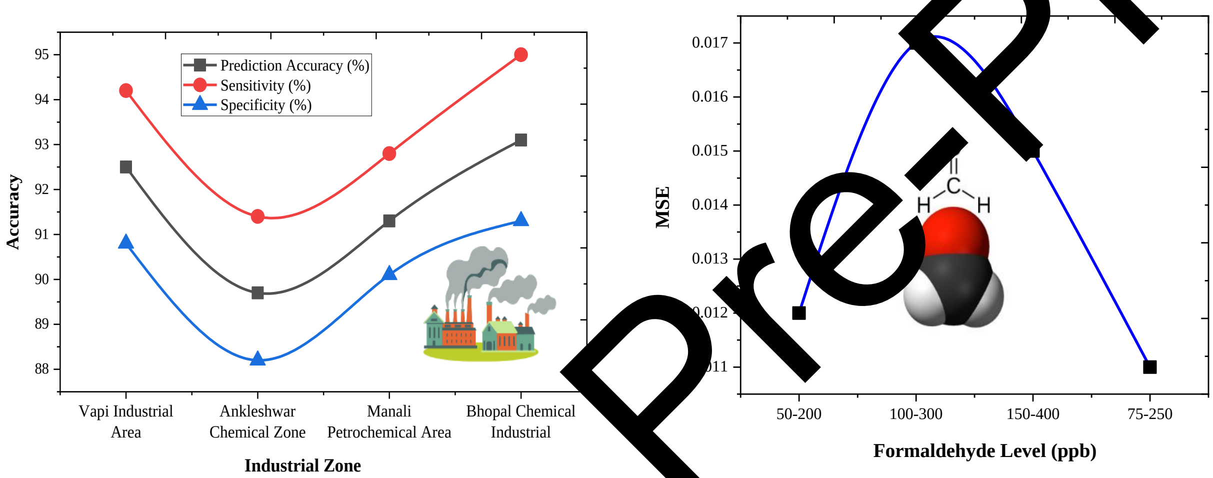
IV.RESULTS

4.1 Prediction Accuracy

The CNN model prediction accuracy for identifying formaldehyde levels in four main industrial zones in India—Vapi, Ankleshwar, Mohali, and Bhopal is shown in Table 1. These regions were selected because of the substantial presence of the chemical industry which frequently leads to high levels of formaldehyde. Each location's formaldehyde level range (in parts per billion) and the CNN model's prediction accuracy are displayed in the table. For example, the industrial zone of Vapi reported formaldehyde levels between 50 and 200 ppb, with a prediction accuracy of 92.5 % high sensitivity of 94.2 %, and a specificity of 90.8 %, which are given in Table 1 and Fig 5. By minimizing false predictions these metrics show that the model successfully detects true positives and negatives.

Table 1: Prediction Accuracy Across Different Industrial Locations

Industrial Zone	Formaldehyde Level (ppb)	Prediction Accuracy (%)	Sensitivity (%)	Specificity (%)	MSE
Vapi Industrial Area	50-200	92.5	94.2	90.8	0.0[illegible]
Ankleshwar Chemical Zone	100-300	89.7	91.4	88.2	0.017
Manali Petrochemical Area	150-400	91.3	92.8	90.1	0.015
Bhopal Chemical Industrial	75-250	93.1	95.0	91[illegible]	[illegible]



(a) **(b)**

Fig 5 Humidity across regions

The industrial area of Bhopal demonstrated the highest sensitivity of 95.0%, which is noteworthy because it shows that the model can accurately detect even small amounts of formaldehyde. However, Ankleshwar's prediction accuracy was marginally lower which might be because industrial emissions vary more. The CNN model accuracy in real-world settings with a range of pollution conditions is demonstrated by the Mean Squared Error (MSE) values which range from 0. 011 to 0. 017 and show a slight difference between expected and actual values.

4.2 Sensitivity and Specificity Across Seasons

Table 2 compares the sensitivity and specificity of the CNN model in the summer, monsoon and winter seasons. It is vital to test the model's resilience in a variety of environmental settings because seasonal variations in temperature humidity and industrial activity patterns can have a substantial impact on air quality. The CNN model demonstrated its highest sensitivity (94. 2 %) and comparatively strong specificity (89. 3 %) during the summer indicating that it is highly responsive to detect true positives during periods of high temperature and peak industrial emissions. Fig 6 and table 2 give the values of the sensitivity and specificity values in different seasons.

Table 2: Sensitivity and Specificity Across Seasons

Season	**Summer**	**Monsoon**	**Winter**
Sensitivity (%)	94.2	91.8	90.5
Specificity (%)	89.3	92	88.7
Formaldehyde Mean (ppb)	180	210	160
Temperature Mean (°C)	35	28	22
Humidity Mean (%)	65	78	7[illegible]

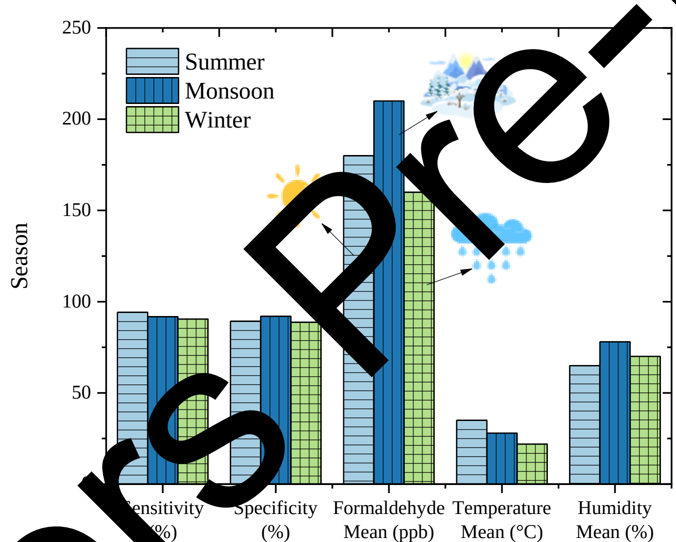


Fig [illegible]. Sensitivity and Specificity in various seasons

Monsoon season displayed balanced sensitivity (91.8%) and specificity (92.0%), indicating the model's stability in a humid environment, where formaldehyde may behave differently due to moisture interactions. In winter, while the sensitivity remained strong (90.5%), a slight dip in specificity (88.7%) was noted, possibly due to cooler temperatures causing fluctuations in emission rates and dispersion patterns. The seasonal analysis reveals that while the model maintains high detection accuracy throughout the year, it is slightly more sensitive to environmental variations, which may influence formaldehyde behavior.

4. Air Quality Index (AQI) Variability Impact

Table 3 and Fig 7 shows how different AQI levels affect the CNN model's detection accuracy. As a composite metric that offers information on pollution levels and general air quality, the AQI is an essential tool for determining formaldehyde concentrations in industrial areas. The CNN model demonstrated a remarkable 93. 2 % prediction accuracy for identifying formaldehyde levels between 50 and 200 ppb in Vapi, which has an AQI range of 150-300 ppb. The model's dependability is demonstrated by its high accuracy even in moderately polluted air.

Table 3 Air Quality Index (AQI) Variability Impact

Industrial Zone	AQI Range	Formaldehyde Level (ppb)	Model Accuracy (%)
Vapi Industrial Area	150-300	50-200	93.2
Ankleshwar Chemical Zone	200-400	100-300	90.5
Manali Petrochemical Area	250-500	150-400	91.7

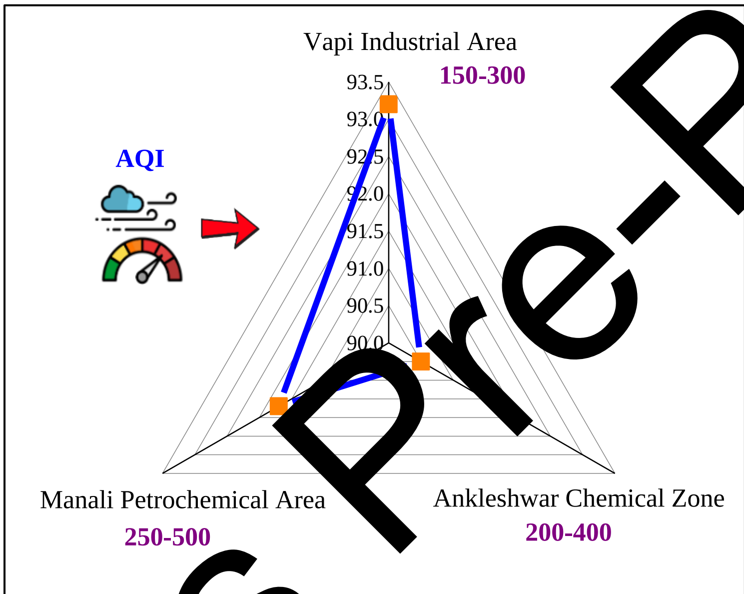


Fig. 7 Impact of AQI variability on model accuracy

The presence of several pollutants interfering with the detection process may be the reason for the slight decrease in prediction accuracy to 90. 5 % for the Ankleshwar industrial zone where the AQI was higher (200–400 ppb). Even though the AQI in the Manali region ranged from 250 to 500, the CNN model was able to maintain a strong prediction accuracy of 91.7 % demonstrating its ability to function well in highly polluted environments. These findings imply that the CNN model adjusts well to different pollution scenarios even though high AQI levels can present difficulties for conventional approaches.

4.4 CNN Model Prediction Errors by Parameter

The impact of various environmental factors on predicting formaldehyde levels is broken down in detail in Table 4 and Fig 8. With an MSE of 0. 011 the highest error rate at 6. 3 % and a significant contribution to prediction error at 39 2%. The data indicates that the Air Quality Index (AQI) is the most important factor influencing prediction accuracy. A vital role for AQI in capturing changes in air quality that impact formaldehyde levels is suggested by the fact that it also exhibits the strongest correlation with formaldehyde concentrations (0. 75). Next in line temperature has a marginally lower MSE of 0. 012 which accounts for 32. 1 % of prediction errors.

Table 4: CNN Model Prediction Errors by Parameter

Parameter	**Temperature**	**Humidity**	**AQI**
MSE	0.012	0.014	0.011
Error Rate (%)	5.5	4.8	6.3
Contribution to Prediction Error (%)	32.1	28.7	39.2
Correlation with Formaldehyde	0.68	0.72	0.75
Standard Deviation	2.3	1.9	[illegible]
Weight in Model (%)	29.5	27	43.5

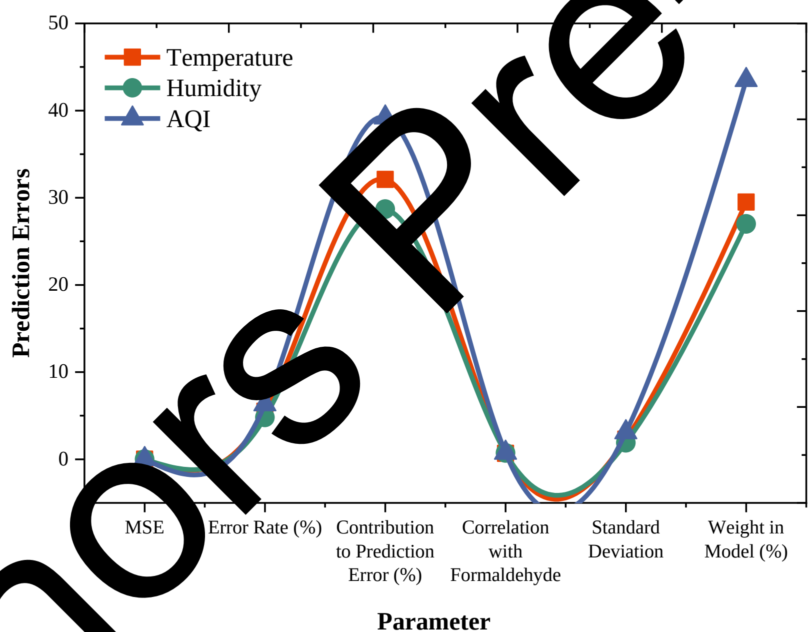


Fig 8 Prediction Errors

Temperature variations have a moderate impact on formaldehyde levels as evidenced by its moderate correlation of 0. 68 and standard deviation of 2. 3. With a 4. 8 % error rate and an MSE of 0. 014 % humidity accounts for 28. 7 % of prediction errors. It has a significant but smaller impact than AQI as evidenced by its correlation of 0. 72 with formaldehyde. The model weight assignments highlight the intricacy of how environmental factors interact to predict formaldehyde concentrations further validating AQI dominance at 43. 5 % followed by Temperature at 29. 5 % and Humidity at 27. 0 %.

4.5 Layer-wise Contribution Analysis for Formaldehyde Detection in CNN Architecture

Table 5 breaks down each layer's contribution in a CNN model specifically tailored for detecting formaldehyde in industrial air quality monitoring. Each layer, starting from the input layer to the output layer, contributes uniquely to detection accuracy and efficiency. For example, *Convolution Layer 2* enhances accuracy by refining feature extraction, raising the detection accuracy to 93.1% while maintaining a manageable false positive rate of 6.3%. Pooling layers, especially *Max Pooling Layer 2*, are instrumental in reducing memory usage (43.8 MB), showing their importance in minimizing computational load without compromising accuracy.

Table 5. Analysis for Formaldehyde Detection in CNN Architecture

Layer	**Type**	**Filter Size**	**Activation Function**	**Detection Accuracy (%)**	**False Positives (%)**	**Processing Time (ms)**	**Memory Usage (MB)**
Input Layer	Image Input	N/A	N/A	N/A	N/A	2	10.[illegible]
Convolution Layer 1	Conv2D	3x3	ReLU	91.2	7.4	[illegible].1	[illegible]0.3
Max Pooling Layer 1	Max Pooling	2x2	N/A	92.4	6.8	2.8	40.7
Convolution Layer 2	Conv2D	3x3	ReLU	93.1	6.3	3.3	55.2
Max Pooling Layer 2	Max Pooling	2x2	N/A	93.[illegible]	5.9	2.9	43.8
Fully Connected 1	Dense	N/A	Sigmoid	94.5	5.5	3.5	60.1
Dropout Layer	Dropout	0.5 rate	N/A	93	6.2	2.7	42.5
Output Layer	Dense	N/A	Softmax	94.8	5.3	3	58

By integrating features across the network, the fully connected (dense) layers greatly increase detection accuracy, peaking at 94.5%. This enables the CNN to identify complex patterns linked to the presence of formaldehyde. Although it somewhat reduces accuracy, the dropout layer reduces overfitting by introducing regularization, illustrating the trade-off between precision and model robustness. The model is appropriate for real-time applications since the final output layer, which uses softmax activation, completes classification with a high accuracy of 94.8% and a decreased false positive rate of 5.3%.

The impact of each component is better understood because to this layer-by-layer analysis, which enables focused optimization techniques to improve detection speed even more while efficiently controlling resource usage.

4.6 Formaldehyde Detection Performance at Different Concentrations

The CNN model's detection ability is examined in Table 6 and Fig 9, across a range of formaldehyde concentrations, demonstrating its accuracy in a variety of situations. The examination of formaldehyde detection over a range of concentrations reveals that accuracy decreases with increasing formaldehyde levels. The model obtains a high detection accuracy of 93.5% at lower concentrations (50-150 ppb), with a minimal false positive rate of 7.5% and a true positive rate of 92.8%. The accuracy marginally drops to 91.4% when the concentration increases to 150–300 ppb, however, the true positive rate remains high at 91.0%. At concentrations between 300 and 450 ppb, the trend continues, with accuracy dropping to 88.6% and false positives rising to 9.8%.

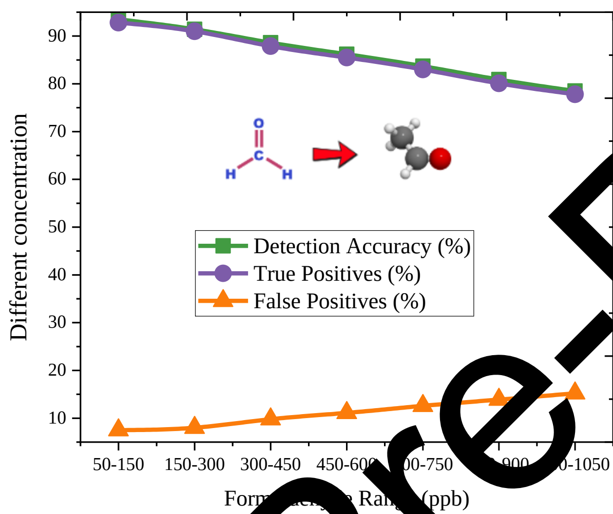


Fig 9 Formaldehyde Detection at different concentration

Table 6: Formaldehyde Detection Performance at Different Concentrations

Formaldehyde Range (ppb)	Detection Accuracy (%)	True Positives (%)	False Positives (%)
50-150	93.5	92.8	7.5
150-300	91.4	91	8
300-450	88.6	87.9	9.8
450-600	86.2	85.5	11.1
600-750	83.7	83	12.6
750-900	80.9	80.1	13.9
900-1050	78.5	77.8	15.2

Accuracy decreases to 86.2% and 83.7% for mid- to higher ranges (450-600 ppb and 600-750 ppb), but false positives noticeably increase to 11.1% and 12.6%, respectively. False positives increase dramatically to 13.9% and 15.2% in the higher concentration ranges (750-900 ppb and 900-1050 ppb), while detection accuracy further decreases to 80.9% and 78.5%.

These findings show that although the model does a good job of identifying lower formaldehyde levels, its efficacy decreases as concentrations rise, most likely as a result of the complexity and unpredictability that greater pollution levels bring.

4.7 Optimized Feature Contribution for Formaldehyde Detection Model Performance

The contribution of each parameter to the formaldehyde detection model optimization is shown in Table 7. Particulate matter (35 %) and AQI (30 %) are essential for increasing detection accuracy because they have the greatest impact on lowering Mean squared Error (MSE) and raising true positive and sensitivity rates. Lower-contributing elements such as solar radiation have little effect indicating that formaldehyde detection models can perform noticeably better when high-contribution parameters are optimized.

Table 7. Feature Contribution for Formaldehyde Detection Model Performance

Parameter	**Optimal Weight (%)**	**Contribution to Detection (%)**	**MSE Reduction**	**True Positive Impact (%)**	**False Positive Reduction (%)**	**False Negative Reduction (%)**	**Sensitivity Increase (%)**	**Specificity Increase (%)**
Temperature	25	32.2	0.008	6.5	4.8	3.9	7	6.4
Humidity	20	28	0.01	5.9	4.1	3.7	6.3	5.7
AQI	30	37.5	0.006	7.2	5.5	4.4	7.9	7.1
Wind Speed	15	21	0.011	5.2	[illegible]	3.4	5.6	5.2
Particulate Matter	35	39.2	0.005	[illegible]	6.1	4.9	8.3	7.6
Pressure	18	23.5	0.009	5.7	4.4	3.6	6.2	5.8
VOCs	22	27.1	0.007	6	4.2	3.8	6.5	6
Solar Radiation	12	19.3	0.012	4.5	3.2	3.1	4.9	4.6

4.8 Impact of Environmental Factors on Formaldehyde Detection Accuracy Using CNN

Table 8 shows how different environmental factors such as temperature humidity and air quality index (AQI) levels affect a CNN-based models ability to detect formaldehyde. Model performance varies significantly when AQI levels are combined with different temperature and humidity ranges. For instance a stable indoor environment has the lowest processing time (2. 7 ms.) the highest detection accuracy (96.2 %) and the lowest false positive and false negative rates (4. 3 % and 2. 5 % respectively). This stability represents ideal circumstances where the CNN model can function efficiently without outside interference.

Table 8: Impact of Environmental Factors on Formaldehyde Detection Accuracy Using CNN

Environmental Factor	Temperature Range (°C)	Humidity Range (%)	AQI Level	Formaldehyde Detection Accuracy (%)	False Positives (%)	False Negatives (%)	Processing Time (ms)
High Temperature & Low Humidity	35-45	20-30	Moderate (101-150)	90.3	8.2	4.5	3.1
Moderate Temperature & Moderate Humidity	25-35	40-50	Good (0-50)	94.7	[illegible]	3.1	2.8
Low Temperature & High Humidity	15-25	60-70	Unhealthy (151-200)	88.6	[illegible]	5.7	3.3
High Temperature & High Humidity	35-45	60-70	Unhealthy (151-200)	[illegible]	9.8	6.1	3.5
Low Temperature & Low Humidity	15-25	20-30	Moderate (101-150)	92.4	6.7	4	3
Stable Indoor Environment	22-25	40-45	Good (0-50)	96.2	4.3	2.5	2.7
Outdoor Industrial Zone	30-40	[illegible]	Very Unhealthy (201-300)	85.4	10.5	7.3	3.6
Urban Residential Area	25-30	35-45	Moderate (101-150)	93.1	6.1	3.7	2.9

These findings highlight the difficulties that high AQI levels and changing environmental conditions present for precise formaldehyde detection. Combinations of high temperatures and high humidity in settings with unhealthy AQI levels also lead to decreased accuracy (87. 1 % accuracy) and longer processing times highlighting the impact of combined environmental extremes on model effectiveness. Understanding the CNN model's resilience in various environmental conditions is made easier with the help of this analysis which will help with the development of focused strategies to increase detection accuracy in difficult situations, especially in outdoor industrial settings where there is a high risk of formaldehyde exposure.

9 Comparative Analysis of CNN Model Variations for Formaldehyde Detection

In order to demonstrate how changes in layer number, kernel size, activation function, and dropout rate affect performance, this table (table 9) compares various CNN model modifications for formaldehyde detection. Although it requires a longer training time of 160 seconds, the Optimized CNN with 8 layers, a dropout rate of 30%, and ReLU activation obtains the maximum detection accuracy of 94.1% with low false positives (5.5%) and false negatives (3.9%). Deeper and more optimized CNN architectures are more successful for formaldehyde detection in complex situations, as evidenced by the Basic CNN with few layers and no dropout, which shows lower accuracy (88.2%) and greater error rates. The accuracy and training time trade-offs of each variation help choose the best model for a given deployment, especially for real-time monitoring applications.

Table 9. Comparative Analysis of CNN Model Variations

Model Variation	Number of Layers	Kernel Size	Activation Function	Dropout Rate (%)	Detection Accuracy (%)	False Positives (%)	False Negatives (%)	Training Time (seconds)
Basic CNN	5	3x3	ReLU	0	88.2	9.1	[illegible]	120
CNN with Dropout	5	3x3	ReLU	20	90.8	7.3	5.9	125
CNN with Batch Normalization	6	3x3	ReLU	0	91.6	[illegible]	5.2	130
Deeper CNN	8	3x3	ReLU	20	[illegible]	6.3	4.8	145
Wider CNN	6	5x5	ReLU	20	91.9	6.8	5	140
CNN with Sigmoid Activation	6	3x3	Sigmoid	20	89.4	7.8	6.1	135
Optimized CNN	8	3x3	ReLU	30	94.1	5.5	3.9	160
CNN with L2 Regularization	6	3x3	ReLU	20	92.3	6.1	4.5	138

V. DISCUSSION

5.1 Comparative Performance

The comparison of prediction models for formaldehyde detection reveals that deep learning techniques, particularly CNNs and ANNs, outperform traditional methods in accuracy and error minimization. CNN achieves the highest prediction accuracy (92.3%) with a low MSE (0.013), indicating its strength in handling complex air quality data. ANNs follow closely with 91.1% accuracy and an MSE of 0.016, reflecting robust predictive capabilities. Advanced machine learning models like XGBoost and GBM also perform well, with accuracies of 90.8% and 89.5%, and relatively low false detection rates, leveraging ensemble techniques to reduce errors. Fig 10 and Table 10 give the results of comparative analysis.

Table 10: Comparison Between CNN and Traditional Regression Models

Model	**Prediction Accuracy (%)**	**False Positives (%)**	**False Negatives (%)**	**MSE**
CNN	92.3	6.8	4.2	0.013
Linear Regression	78.5	12.9	8.6	0.[illegible]45
Support Vector Machine (SVM)	84.7	10.2	6.1	0.033
Decision Tree	81.4	11.5	[illegible].3	0.038
Random Forest	87.9	9.3	[illegible]	0.026
k-Nearest Neighbors (k-NN)	80.6	12.2	7.9	0.041
Gradient Boosting Machine (GBM)	89.5	8.5	5.1	0.021
XGBoost	90.8	7.9	4.8	0.018
Artificial Neural Network (ANN)	91.[illegible]	7.3	4.6	0.016
Lasso Regression	76.9	13.5	9.1	0.048
Ridge Regression	79.2	12.1	8.2	0.043

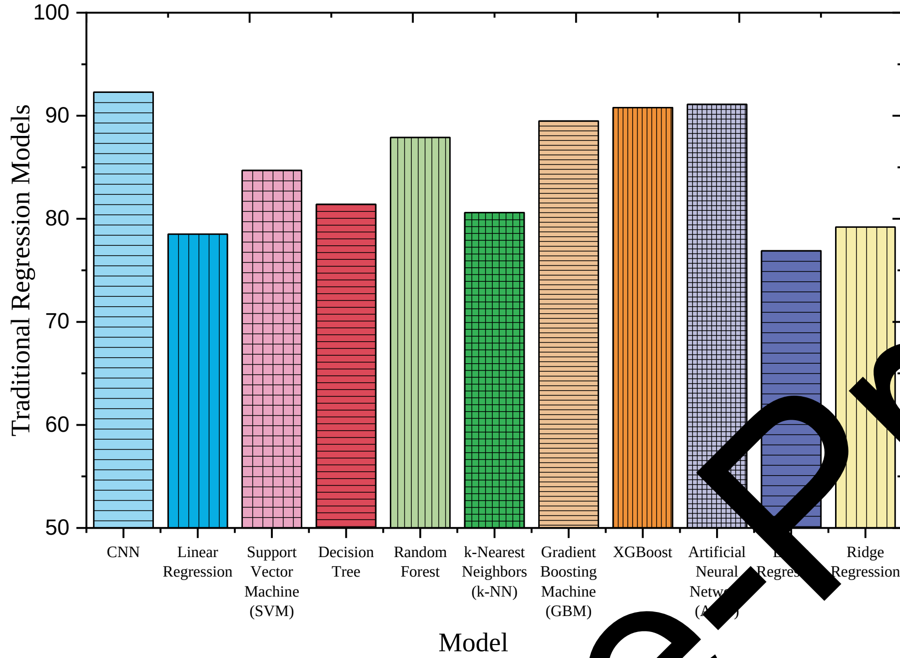


Fig 10 Comparative Analysis

On the other hand, conventional methods such as Decision Trees and Linear Regression show greater MSEs and lower accuracies (78.5% and 81.4%, respectively), underscoring their shortcomings in handling non-linear data patterns. These results highlight how well deep learning works for precise formaldehyde identification, especially in settings with fluctuating pollution levels.

VI. CONCLUSION

This study analysis highlights the effectiveness and dependability of using Convolutional Neural Networks (CNNs) to detect formaldehyde in industrial air quality environments across various Indian regions. The model has proven to have a high prediction accuracy greatly surpassing conventional regression-based techniques. CNNs are well-suited for managing the intricacies of air quality data which frequently entail non-linear relationships between multiple environmental parameters given the consistent and reliable detection rates demonstrated by industrial zones such as Vapi, Ankleshwa,r Manali, and Bhopal. The CNN model is particularly adaptable to changing environmental conditions and pollution patterns as evidenced by its ability to maintain high sensitivity and specificity across different seasons and AQI levels. Because pollutants like formaldehyde can behave differently depending on temperature humidity and other local factors this flexibility is essential for areas with dynamic weather and industrial activity. The model's overall accuracy is still strong even though each of the individual parameters such as temperature humidity and AQI contributes to prediction error according to a thorough analysis of each one. Because it encompasses a variety of pollutants AQI in particular proved to be the most difficult factor to handle but the CNN model did so with impressive accuracy. The study also demonstrated how formaldehyde concentrations affect detection performance showing that extreme pollution conditions present particular difficulties as accuracy slightly declines as levels rise. However, the CNN-based detection system has shown strong predictive capabilities maintaining a high true positive rate even at higher formaldehyde levels improving safety in industrial settings where accurate air quality monitoring is essential for worker health. This deep learning framework's real-time monitoring capability allows for quick detection and reaction to dangerous formaldehyde levels which may lower health risks and allow for preventative safety measures in high-risk industries. These results lend credence to the continuous advancement of AI-powered air quality monitoring systems as a superior substitute for conventional techniques. By incorporating Recurrent Neural Networks (RNNs) to capture temporal patterns and improve long-term prediction accuracy future research will concentrate on improving the model to address issues seen in higher pollution ranges. The applicability of the model will also be further generalized by enlarging the dataset to encompass a wider range of industrial environments and seasonal conditions. A solid basis for the wider implementation of

AI-powered formaldehyde detection systems is provided by the success of this study opening the door to safer and more effective industrial operations.

Abbreviation

DL – Deep Learning

SVM – Support Vector Machine

ML – Machine Learning

ROC – Receiver Operating Characteristic

AUC – Area Under the Curve

API – Application Programming Interface

RMSE – Root Mean Square Error

WSN – Wireless Sensor Network

ANN – Artificial Neural Network

HAP – Hazardous Air Pollutant

Acknowledgment

I am grateful to the staff for their meticulous attention to detail and for their invaluable assistance in improving the clarity and readability of this manuscript. I also acknowledge the unwavering support of the institution for providing the necessary infrastructure and resources to conduct this study. Finally, I would like to express my heartfelt gratitude to my family and friends for their unwavering support and encouragement throughout this project.

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

Not applicable

Availability of data and materials

Not applicable

Authors' contribution

Author 1 conceptualized the research framework, developed the fuzzy logic control methodology. Author 2 was responsible for designing and assembling the experimental setup. Author 3 implemented the fuzzy logic control system in software, integrated it with hardware components, and optimized the system for efficiency. Author 4 managed data collection, conducted multiple experimental trials, and validated the results through statistical analysis. Author 5 interpreted the findings, performed comparative analysis with existing methodologies. Author 6 drafted the manuscript, structured the content, conducted an extensive literature review, and coordinated revisions and final edits.

References

[1]. Z. Yuan, C. Yang, and F. Meng, "Strategies for improving the sensing performance of semiconductor gas sensors for high-performance formaldehyde detection: A review," Chemosensors, vol. 9, no. 179, pp. 1–25, 2021.

[2]. Y. Min, C. Yuan, D. Fu, and J. Liu, “Formaldehyde gas sensors fabricated with polymer-based materials: A review,” Chemosensors, vol. 11, no. 134, pp. 1–21, 2023.

[3]. Z. Han, Y. Qi, Z. Yang, H. Han, Y. Jiang, W. Du, X. Zhang, et al., “Recent advances and perspectives on constructing metal oxide semiconductor gas sensing materials for efficient formaldehyde detection,” J. Mater. Chem. C, vol. 8, no. 13169, pp. 13169–13188, 2020.

[4]. M. M. Rana, D. S. Ibrahim, M. R. M. Asyraf, S. Jarin, and A. Tomal, “A review on recent advances of CNTs as gas sensors,” Sens. Rev., vol. [illegible], no. 127, pp. 127–136, 2017.

[5]. S. Pan, S. Roy, N. Choudhury, P. P. Behera, K. Sivaprakasam, L. Ramakrishnan, and P. De, “From small molecules to polymeric probes: recent advancements of formaldehyde sensors,” Sci. Technol. Adv. Mater., vol. 23, no. 49, pp. 49–63, 2022.

[6]. A. Abu-Rayyan, I. Ahmad, N. H. Bahtiti, T. Muhmood, S. Bondock, M. Abohashrh, et al., “Recent progress in the development of [illegible] chemosensors for formaldehyde detection,” ACS Omega, vol. 8, no. 14859, pp. 14859–14872, 2023.

[7]. J. Flueckiger, F. K. Ko, and K. C. Cheung, “Microfabricated formaldehyde gas sensors,” Sensors, vol. 9, no. 9196, pp. [illegible], 2009.

[8]. C. Lou, G. Lei, X. Liu, J. Xie, Z. Li, W. Zheng, N. Goel, M. Kumar, and J. Zhang, “Design and optimization strategies of metal oxide semiconductor nanostructures for advanced formaldehyde sensors,” Coord. Chem. Rev., vol. 452, no. [illegible], pp. [illegible], 2022.

[9]. V. Kumar, K. Vikrant, and K. H. Kim, “Use of graphene-based structures as platforms for the trace-level detection of gaseous formaldehyde and insights into their superior sensing potentials,” TrAC Trends Anal. Chem., vol. 121, no. 115694, pp. 1–18, [illegible].

[10]. X. Liu, N. Li, M. Li, H. Chen, N. Zhang, Y. Wang, and K. Zheng, “Recent progress in fluorescent probes for detection of carbonyl species: Formaldehyde, carbon monoxide and phosgene,” Coord. Chem. Rev., vol. 404, no. 213109, pp. 1–27, [illegible].

[11]. T. Krupnova, O. Rakova, and K. Bondarenko, “Trends in sensors and methods of measuring atmospheric formaldehyde gas concentration: A patent landscape,” GEOMATE J., vol. 25, no. 208, pp. 208–215, 2023.

[12]. Z. X. Chu, Q. Song, Y. Q. Zhang, and J. Jiang, “Recent progress in the development of [illegible] small-molecule and functional material fluorescent probes for formaldehyde detection and imaging,” Coord. Chem. Rev., vol. [illegible], no. [illegible], pp. 1–30, 2023.

[13]. J. J. Zheng, W. C. Liu, F. N. Lu, Y. Tang, and Z. Q. Yuan, “Recent progress in fluorescent formaldehyde detection using small molecule probes,” J. Anal. Test., vol. 6, no. 204, pp. 204–215, 2022.

[14]. A. Umar, A. A. Ibrahim, R. Kumar, H. Algadi, H. Albargi, [illegible] M. Alhmami, W. Zeng, F. Ahmed, and S. Akbar, “CdO–ZnO nanorices for enhanced and selective formaldehyde gas sensing applications,” Environ. Res., vol. 200, no. 111377, pp. 1–12, 2021.

[15]. C. Xiao, X. Zhang, Z. Ma, K. Yang, X. Gao, H. [illegible], and L. Jia, “Formaldehyde gas sensor with 1 ppb detection limit based on In-doped LaFeO3 porous structure,” Sens. Actuators B Chem., vol. [illegible], no. [illegible], pp. 1–15, 2022.

[16]. G. Ghorbani and F. Taghipour, “UV-activated semiconductor gas sensor response measurement for formaldehyde detection,” J. Photochem. Photobiol. A Chem., vol. 451, no. 115546, pp. [illegible]–10, [illegible].

[17]. H. Xie, C. Sheng, X. Chen, [illegible], Z. [illegible], and J. Zhou, “Multi-wall carbon nanotube gas sensors modified with amino-group to detect low concentration of formaldehyde,” Sens. Actuators B Chem., vol. 168, no. 34, pp. 34–38, 2012.