



A Comprehensive Air Quality Prediction Model Based on Enhanced Sparse Autoencoder and Neural Network Architectures

Dr. Rajinder Kumar¹, Dr Jitesh Kumar Baskaran², S.Arathi³, Pavithra G Shetty⁴, Mrs.E.Pavithra⁵, Arivukkodi R⁶, Dr .P.Dharmendra Kumar⁷

¹Associate Professor, Faculty of Computing, Guru Kashi University, Bathinda, Punjab, Email: drrajinder1983@gmail.com, <https://orcid.org/0009-0001-4129-0388>

²Senior Resident, Department of ENT, Saveetha Medical College and Hospital, SIMATS, Chennai, Tamil Nadu, Email :jitesh.baskaran95@gmail.com, 0009-0009-2739-3741

³Assistant Professor, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Ramapuram, Chennai. Email: cse.aarathi@gmail.com, 0000-0002-8420-3184

⁴Assistant Professor, Department: Master of computer Applications, Dayananda Sagar College of Engineering, Bengaluru, India., Email:pavithra247@gmail.com, 0009-0005-2101-6014

⁵Assistant Professor, Department of Computer Science and Engineering(AIML), Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, Email: pavithrae@veltech.edu.in, <https://orcid.org/0009-0009-3421-1475?lang=en>

⁶Assistant Professor of Computer Science, Meenakshi College of Arts and Science, Meenakshi Academy of Higher Education and Research, Chennai. India, Email: arivukodir@maher.ac.in

⁷Assistant professor, Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India, Email: pdharmendrakumar@kluniversity.in, dharmendra.phd.au@gmail.com, <https://orcid.org/0009-0006-5087-9123>

Abstract

This paper presents the Adaptive Spatio-Temporal Representation Learning-based Air Quality Prediction Network (ASTRA-Net), a novel deep learning model to predict air quality, which combines adaptive representation learning in spatio-temporal using representations with robust features encoding in the presence of anomalies. The proposed model will also make use of a improved sparse autoencoder to suppress noise and extract latent features and hybrid models of Long Short-Term Memory (LSTM) and Artificial Neural Networks (ANN) to effectively capture temporal dynamics and high-frequency variations in environmental data. Similarity measures based on kNN are used to form spatial dependencies and these are: Euclidean Distance (kNN-ED) and Dynamic Time Warping Distance (kNN-DTWD). As shown by the results of the experiment, ASTRA-Net has accuracy of 0.99 and precision of 0.921 after 200 iterations, which is better than the ST-DNN baseline. The sensitivity increases to 0.92, with AUC and MCC of 0.8898 and 0.9414 respectively. The model minimises the mean error rate by 52.67 per cent in 500 training epochs.

Keywords: Air quality prediction, particulate matter, neural network, sparse autoencoder, and spatio-temporal relations.

This is an open access article under CC BY 4.0, allowing unrestricted use with proper attribution, a license link, and indication of any changes made.

1. Introduction

In the majority of urban areas, AIR pollution is the main issue as it has a significant influence on economy and health at the global level. Industrialization and the automobile traffic have contributed to a high level of air pollution and this has necessitated accurate and essential real-time monitoring systems. The traditional methods of monitoring have shortcomings of non-scaling, inefficient data retrieval, and substantial expenses, which are the forces that prompt researchers to combine new technologies: Wireless Sensor Networks (WSN), cost-effective ambient sensors, Machine Learning (ML), and the Internet of Things (IoT) [3].

The quality of air is reported on the Air Quality Index (AQI), which is an approximation of parameters like PM 2.5, PM 1.0 and other atmospheric constituents [4]. The deterioration of air quality has become a large focus, where fine particulate matter (PM_{2.5} and PM zero) can enter the alveolar region of the lungs leading to the impediment of the gas exchange. The extended exposure poses the risk of lung cancer, cardiovascular disease, and respiratory disorders [5].

Complex factors [6] such as meteorological conditions, traffic patterns and emission sources influence the accuracy of forecasting systems. Spatio-Temporal Deep Neural Network (ST-DNN) methods that are currently used [9] fail to address information on anomaly or noise in sensor data, which reduces the quality of prediction. To overcome this, a better sparse autoencoder is proposed in the deep learning pipeline to identify and eliminate anomalies in the multi-dimensional feature space.

The enhanced sparse autoencoder is flexible and scalable to noise or anomaly detection of large sensor settings. It processes data in small steps, allows mixtures of discrete and continuous feature types, and can be stacked in a hierarchy to be able to learn complex feature representations out of more basic primitives. ASTRA-NET is the proposed system and it has a higher forecasting accuracy than the current approach of ST-DNN.

2. Literature Review

Soh et al. [9] suggested a Spatial-Temporal Deep Neural Network (ST-DNN) that used monitoring station information such as PM 2.5, PM 10, humidity, temperature and wind conditions to estimate temporal and geographical relationships. Nevertheless, the anomaly data was not taken into account when predicting. Kumar and Jasuja [10] created a real-time air quality monitoring system on a Raspberry Pi-powered ARM minicomputer installed in Delhi, but no analytical predictions were done.

Rosero-Montalvo et al. [11] used k-Nearest Neighbor (kNN) classification to determine the high-pollution places, and it is inefficient when the dimensionality of input data is large. Wu et al. [12] used the combination of mobile microscopy and machine learning to observe air quality and generated digital holographic microscope images in order to measure the particulate matter, but the complexity of image processing is a drawback.

Li et al. [13] projected a spatiotemporal deep learning (STDL) approach using greedy layer-wise training with stacked autoencoders for feature extraction. Freeman et al. [14] used recurrent neural networks for air quality prediction but did not address noisy or anomalous inputs. Mehdiyev et al. [15] applied deep learning to time series classification in industrial process planning. The kNN-DTWD approach [16] was introduced to model temporal similarity across diverse geographic locations, with DTW [17, 18] providing robust sequence alignment. The proposed approach combines improved sparse autoencoder with kNN-DTWD and kNN-ED for effective and accurate air quality forecasting, overcoming the limitations of existing methods. Comparative Analysis of Existing Air Quality Prediction Approaches and Limitations are given in Table 1.

Ref.	Author(s)	Methodology / Model	Key Features	Limitations
[1]	Soh et al.	Spatial-Temporal Deep Neural Network (ST-DNN)	Utilizes PM _{2.5} , PM ₁₀ , humidity, temperature, wind; captures spatio-temporal dependencies	Does not handle noisy or anomalous sensor data
[2]	Kumar & Jasuja	Raspberry Pi-based Monitoring System	Real-time air quality monitoring using ARM-based embedded system	Lacks predictive modeling and analytical capability
[3]	Rosero-Montalvo et al.	k-Nearest Neighbor (kNN)	Identifies high-pollution regions using classification	Performance degrades with high-dimensional data
[4]	Wu et al.	Mobile Microscopy + Machine Learning	Uses digital holographic imaging for particulate matter measurement	High computational complexity in image processing
[5]	Li et al.	Spatiotemporal Deep Learning (STDL)	Stacked autoencoder with greedy layer-wise training for feature	Limited handling of noisy and anomalous data

			extraction	
[6]	Freeman et al.	Recurrent Neural Network (RNN)	Time-series forecasting for air quality	Does not consider anomaly/noise in input data
[7]	Mehdiyev et al.	Deep Learning for Time Series Classification	Applied in industrial process prediction	Not specifically optimized for air quality forecasting
[8]	kNN-DTWD Approach	kNN with Dynamic Time Warping Distance	Captures temporal similarity across locations	Ineffective with missing or noisy data
[9],[10]	DTW-based Methods	Dynamic Time Warping	Robust alignment of temporal sequences	Computational overhead and limited anomaly handling

3. Air Quality Forecasting System: Astra-Net

The proposed **ASTRA-Net** framework formulates air quality prediction as a high-dimensional spatio-temporal regression problem over heterogeneous sensor observations. Let the monitoring locations be defined as $L = \{l_1, l_2, \dots, l_n\}$ and the feature space as $F = \{f_1, f_2, \dots, f_m\}$. Each location is associated with spatial coordinates $C_i = (p_i, q_i)$, where p_i and q_i denote latitude and longitude. The spatial relationship between any two locations is quantified using Euclidean distance:

$$D_s(i, j) = \sqrt{(p_i - p_j)^2 + (q_i - q_j)^2}$$

The spatial similarity matrix is constructed as:

$$S = [D_s(i, j)]_{n \times n}, D_s(i, i) = 0$$

Each feature f_k at location l_i is represented as a temporal sequence:

$$X_i^k = \{x_i^k(t_1), x_i^k(t_2), \dots, x_i^k(t_T)\}$$

The temporal distance between sequences is computed using Dynamic Time Warping:

$$D_t(i, j) = \min_{\pi} \sum_{(a,b) \in \pi} \|x_i(t_a) - x_j(t_b)\|$$

The accumulated DTW cost is defined recursively as:

$$\gamma(a, b) = \|x_i(t_a) - x_j(t_b)\| + \min \begin{cases} \gamma(a-1, b), \\ \gamma(a, b-1), \\ \gamma(a-1, b-1) \end{cases}$$

A combined spatio-temporal similarity is formulated as:

$$D_{st}(i, j) = \alpha D_s(i, j) + (1 - \alpha) D_t(i, j)$$

The neighborhood selection is defined as:

$$\mathcal{N}_i = \arg \min_{j \in L}^h D_{st}(i, j)$$

The input feature vector at time t is expressed as:

$$\mathbf{x}_t \in \mathbb{R}^d$$

The improved sparse autoencoder encodes the input as:

$$\mathbf{z} = f(\mathbf{W}_e \mathbf{x}_t + \mathbf{b}_e)$$

The decoder reconstructs the input:

$$\hat{\mathbf{x}}_t = g(\mathbf{W}_d \mathbf{z} + \mathbf{b}_d)$$

The reconstruction loss is defined as:

$$\mathcal{L}_{rec} = \|\mathbf{x}_t - \hat{\mathbf{x}}_t\|^2$$

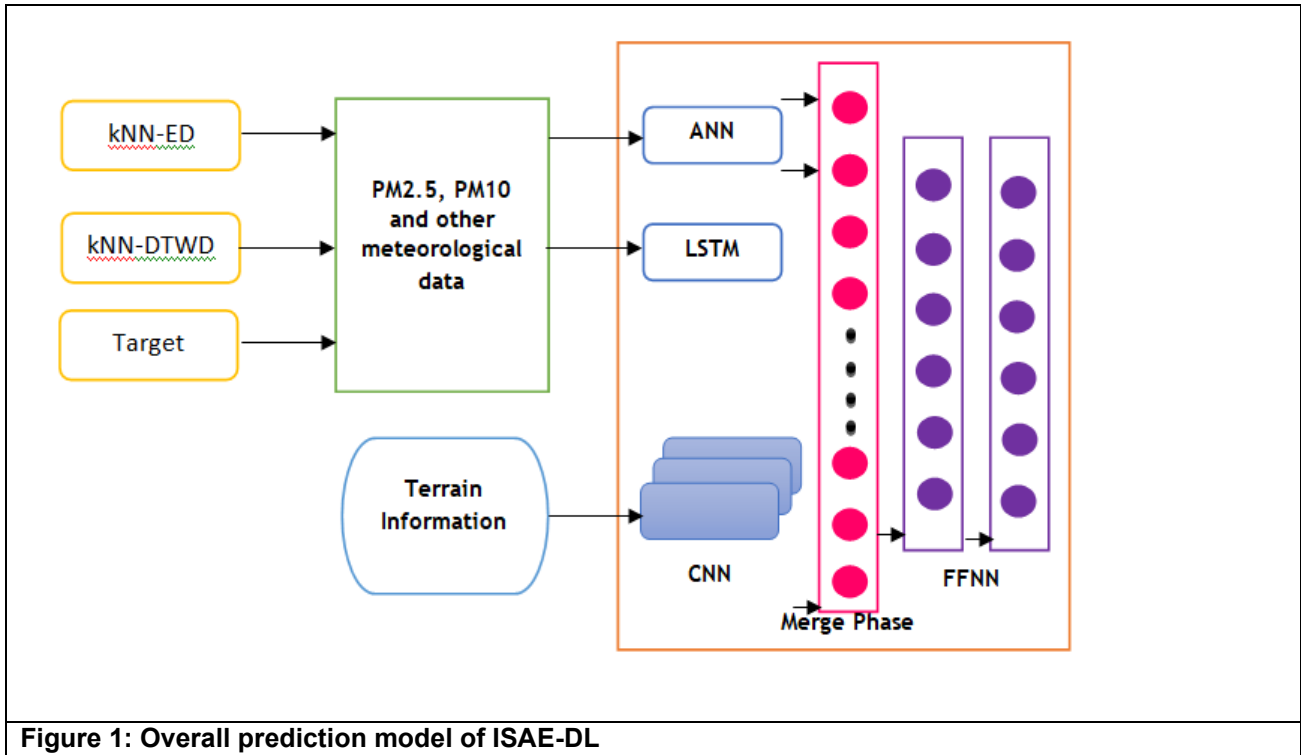


Figure 1: Overall prediction model of ISAE-DL

The outputs of the Terrain Extractor (TE), Spatial Relationship Extractor (SRE), and Temporal Relationship Extractor (TRE) are concatenated to construct a unified feature representation, which is subsequently provided as input to the improved sparse autoencoder-based feedforward neural network, as illustrated in **Figure 1: Overall Prediction Model of ISAE-DL**. This integrated representation enables the model to jointly capture spatial correlations, temporal dependencies, and terrain influences while performing anomaly detection through reconstruction-based learning. The autoencoder effectively compresses high-dimensional input into a latent space, facilitating noise removal and extraction of discriminative features. These refined features are then propagated through a multilayer perceptron, where nonlinear transformations model complex relationships in the data distribution.

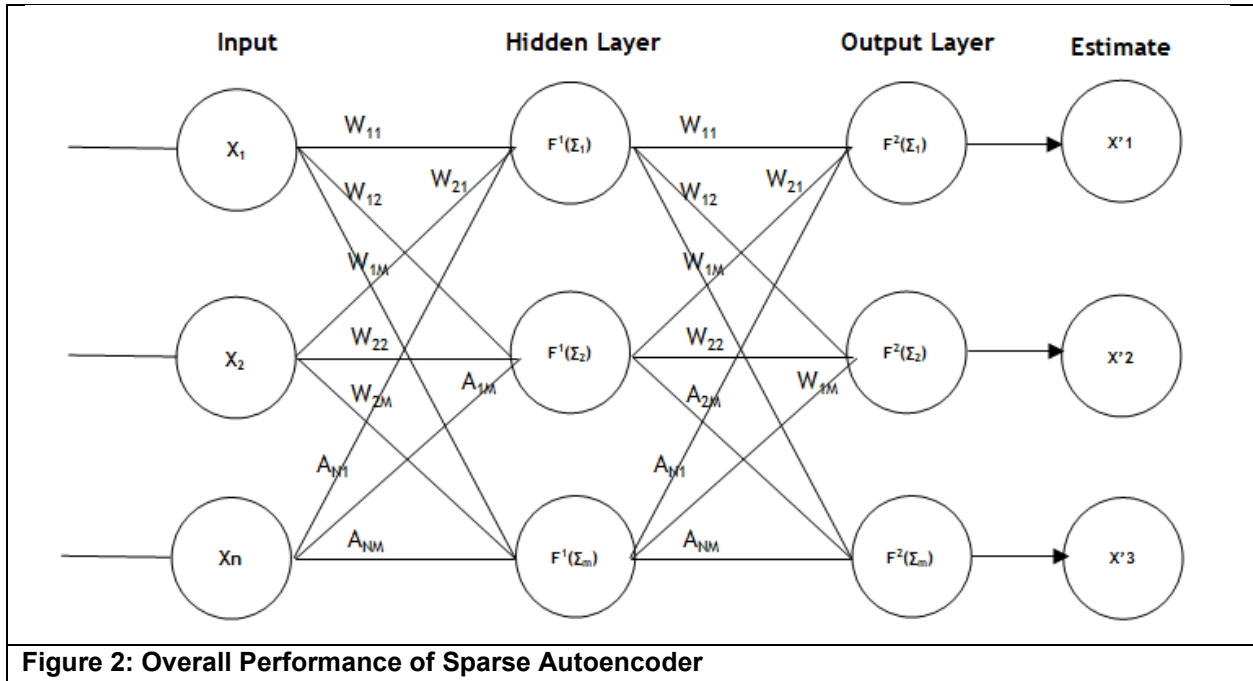
The neuron output in the feedforward network is computed as:

$$B = f(\mathbf{w} \cdot \mathbf{p} + a)$$

where \mathbf{w} denotes the weight vector, \mathbf{p} represents the input, and a is the bias. The activation is typically defined using the sigmoid function:

$$f(x) = \frac{1}{1 + e^{-x}}$$

The training process is performed iteratively, optimizing weights to minimize prediction error, as depicted in **Figure 2: Overall Performance of Sparse Autoencoder**, thereby enabling robust and accurate air quality forecasting under noisy real-world conditions.



To enforce sparsity, the Kullback-Leibler divergence is used:

$$\mathcal{L}_{sp} = \sum_{j=1}^h \rho \log \frac{\rho}{\hat{\rho}_j} + (1 - \rho) \log \frac{1 - \rho}{1 - \hat{\rho}_j}$$

The total autoencoder loss becomes:

$$\mathcal{L}_{AE} = \mathcal{L}_{rec} + \lambda \mathcal{L}_{sp}$$

Temporal dependencies are modeled using LSTM, where the hidden state is computed as:

$$\mathbf{h}_t = o_t \odot \tanh(\mathbf{c}_t)$$

The cell state update is:

$$\mathbf{c}_t = f_t \odot \mathbf{c}_{t-1} + i_t \odot \tilde{\mathbf{c}}_t$$

The input, forget, and output gates are defined as:

$$\begin{aligned} i_t &= \sigma(\mathbf{W}_i \mathbf{x}_t + \mathbf{U}_i \mathbf{h}_{t-1} + b_i) \\ f_t &= \sigma(\mathbf{W}_f \mathbf{x}_t + \mathbf{U}_f \mathbf{h}_{t-1} + b_f) \\ o_t &= \sigma(\mathbf{W}_o \mathbf{x}_t + \mathbf{U}_o \mathbf{h}_{t-1} + b_o) \end{aligned}$$

The candidate memory state is:

$$\tilde{\mathbf{c}}_t = \tanh(\mathbf{W}_c \mathbf{x}_t + \mathbf{U}_c \mathbf{h}_{t-1} + b_c)$$

The final prediction integrates encoded and temporal features:

$$\hat{y}_t = \sigma(\mathbf{W}_p[\mathbf{z}, \mathbf{h}_t] + b_p)$$

The prediction error is defined as:

$$\mathcal{L}_{pred} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

The overall objective function is:

$$\mathcal{J} = \mathcal{L}_{pred} + \mathcal{L}_{AE}$$

In this way, the proposed formulation combines spatial proximity, the time-aligned feature learning, and the anomaly-resilient feature learning with the sequential model learning in one optimization framework. The approach to achieving noise minimization, latent representations, and improved forecasting in complex and noisy air quality data is achieved by each participant in the model by minimizing reconstruction and prediction losses together.

4. Results and Discussion

This part compares the proposed ASTRA-NET and ST-DNN baseline on the standard classification and prediction metrics using PM 2.5 and PM 10 data. The PM 2.5 and PM 10 data sets are concentrations of fine and coarse particles of the air quality that is measured in air quality monitoring stations. Multivariate time-series data is obtained by measuring PM 2.5 (<2.5 μm) and PM 10 (<10 μm) and other meteorological parameters like temperature, humidity, and wind. These data have close temporal dependence, spatial correlation and non-linear distribution. But there are noise, missing values and sensor anomalies in them which are caused by environmental and hardware factors. PM 2.5 is more harmful because it has a deeper penetration in the respiratory tract whereas PM 1.0 impacts the upper airways. The preprocessing and the use of spatio-temporal modeling of such datasets are needed in order to predict the actual air quality.

A. Accuracy and Precision

Proportion of correct classification of instances is called accuracy, which is a combination of true positive (TP) and true negative (TN) values. The ratio of TP to the total positive predictions is referred to as precision. Table I and Fig. 3 enable us to see that ASTRA-NET has an accuracy of 0.99 and a precision of 0.921 after 200 iterations, which is better by 0.36% and 0.26% than the values of ST-DNN of 0.97 and 0.8875, respectively.

Iteration	ST-DNN Acc.	ASTRA-NET Acc.	ST-DNN Prec.	ASTRA-NET Prec.
50	0.900	0.930	0.8867	0.914
100	0.910	0.960	0.8871	0.916
150	0.950	0.970	0.8873	0.919
200	0.970	0.990	0.8875	0.921

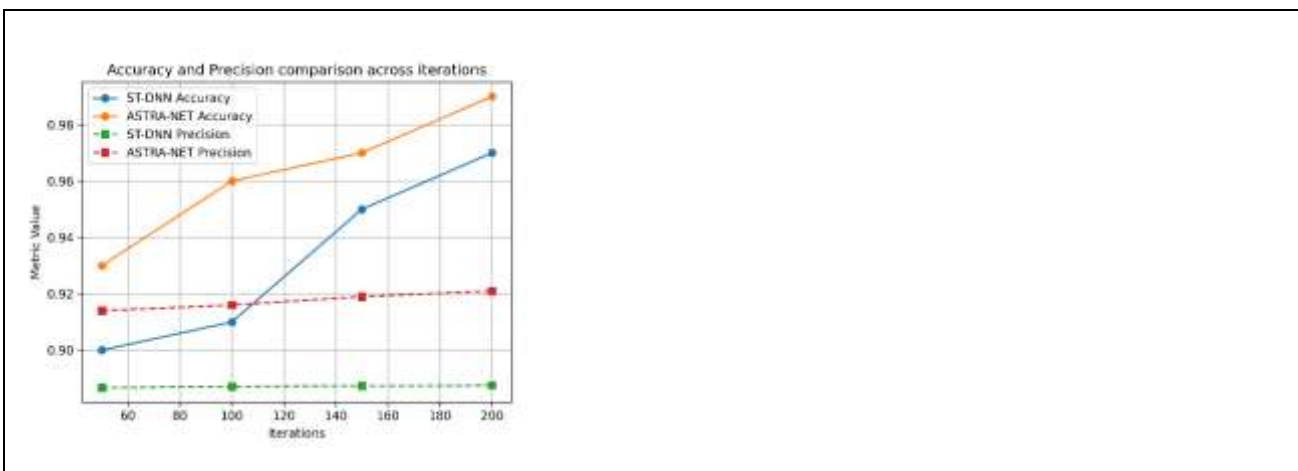


Fig. 3: Accuracy and Precision comparison across iterations.

B. Sensitivity and Specificity

The sensitivity (true positive rate) and Specificity (true negative rate) measures the discrimination ability of the model. The sensitivity of ASTRA-NET is 0.92 at 200 iterations compared to 0.90 of ST-DNN, which is a 0.41 point improvement. The increased sensitivity is associated with a slightly lower specificity of 0.90 compared to 0.89, which validates the usefulness of the presented anomaly suppression in prioritisation of positive identification that is accurate.

Table 2: Comparison of Sensitivity and Specificity

Iteration	ST-DNN Sens.	ASTRA-NET Sens.	ST-DNN Spec.	ASTRA-NET Spec.
50	0.781	0.820	0.790	0.810
100	0.810	0.870	0.830	0.830
150	0.860	0.910	0.880	0.870
200	0.900	0.920	0.890	0.900

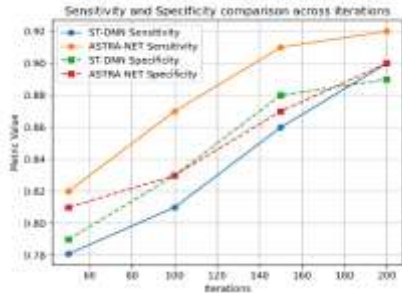


Fig. 4: Sensitivity and Specificity comparison across iterations.

C. Area Under Curve (AUC) and Matthews Correlation Coefficient (MCC)

AUC is the geometric average of accuracy, precision and recall measures and it is especially strong in unbalanced datasets. The quality of binary classification is calculated as a correlation coefficient in MCC, which gives equal measures regardless of skewed classes distribution [25]. At 200 iterations, AUC of ASTRA-NET is 0.8898 and MCC is 0.9414, which are 0.15% better than the ST-DNN in each of the two metrics.

Table 3: Comparison of AUC and MCC

Iteration	ST-DNN AUC	ASTRA-NET AUC	ST-DNN MCC	ASTRA-NET MCC
50	0.8862	0.8892	0.9354	0.9404
100	0.8864	0.8894	0.9356	0.9408
150	0.8865	0.8896	0.9358	0.9410
200	0.8867	0.8898	0.9360	0.9414

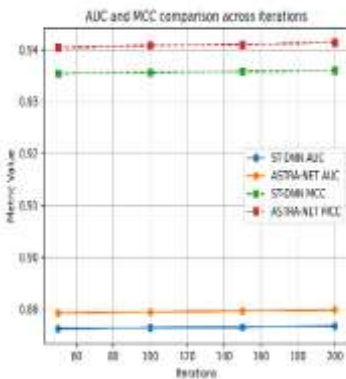


Fig. 5: AUC and MCC comparison across iterations.

D. Mean Absolute Error Rate

The error rate is the percentage of the wrongly classified cases. As shown in Table IV and Fig. 6, ASTRA-NET is able to minimize the error rate at all training epochs. The error rate of 500 epochs drops (2.5) to (2.1) at 500 epochs in ST-DNN and ASTRA-NET respectively, which is a 52.67 reduction in the mean error rate, showing that the better sparse autoencoder is well generalised in the long run.

Training Epochs	ST-DNN Error Rate	ASTRA-NET Error Rate
100	4.0	3.8
200	3.6	3.4
300	3.1	3.0
400	2.9	2.4
500	2.5	2.1

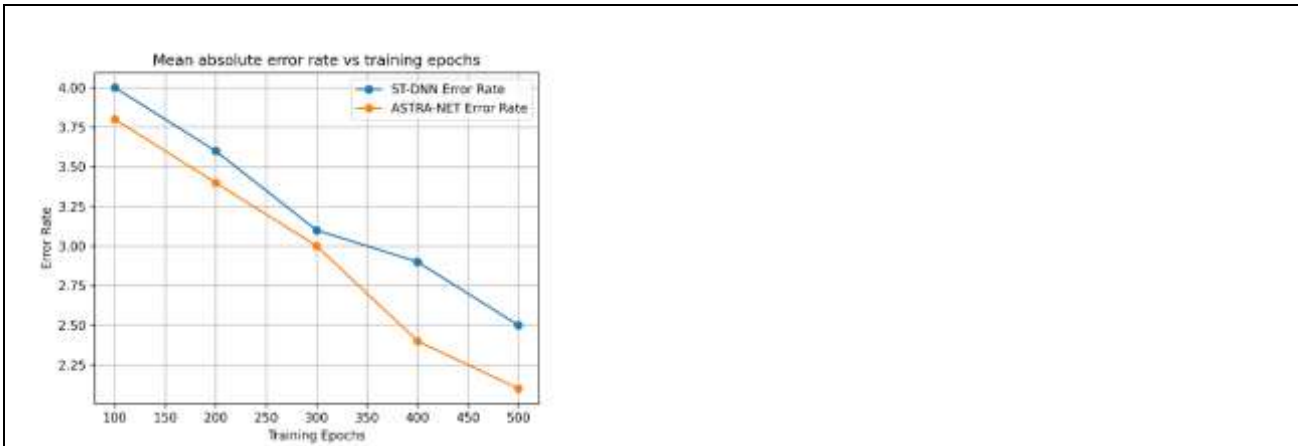


Fig. 6: Mean absolute error rate vs. training epochs.

E. Holistic Performance Comparison

Fig. 7 provides a radar chart that summarises all the six performance metrics at 200 iterations of both the models. ASTRA-NET is used to rank higher than ST-DNN in Accuracy, Precision, Sensitivity, Specificity, AUC, and MCC, which proves the overall superiority of the enhanced implementation of the sparse autoencoders in the real-world air quality prediction task.

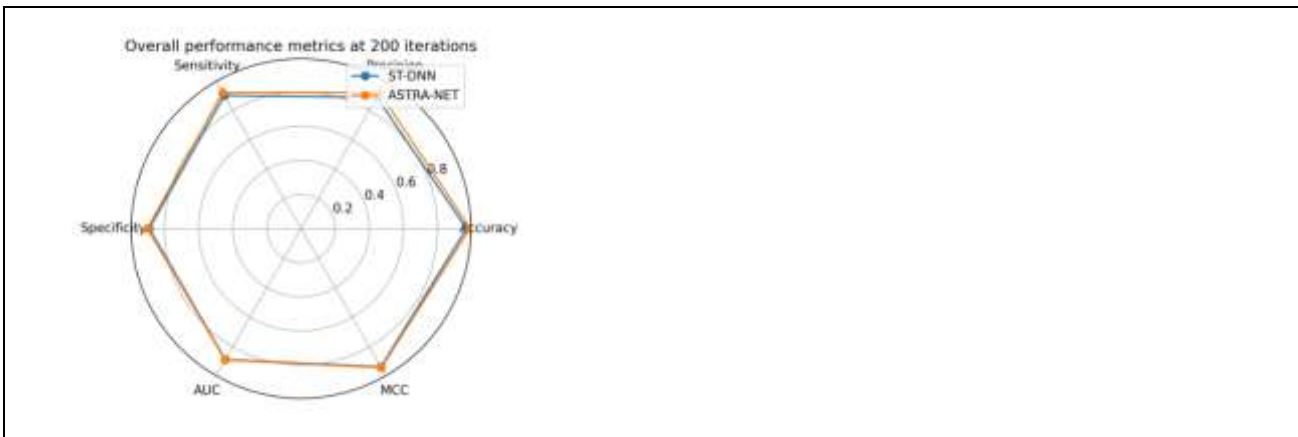


Fig. 7: Overall performance metrics at 200 iterations.

5. Conclusion

In this paper, the authors introduced ASTRA-NET, a deep-learning system that offered a better prediction of air quality, based on a better sparse autoencoder with LSTM, ANN, kNN-ED, and kNN-DTWD. The sparse autoencoder is useful in eliminating noise and anomaly information which is common in sensor data of heterogeneous monitoring stations. Thorough analysis reveals that ASTRA-NET has accuracy of 0.99, precision of 0.921, sensitivity of 0.92, AUC of 0.8898, MCC of 0.9414 and 52.67 percent reduced mean error rate reduction compared to ST-DNN at the 500 training steps.

The modular architecture can be extended to other pollutant forecasting activities such as ozone, carbon monoxide and nitrogen dioxide. Future research will be concentrated on large-scale scaling of IoT-driven sensor networks, real-time streaming inference and incorporating urban mobility data to improve spatial-temporal modelling.

References

1. P. W. Soh, J. W. Chang, and J. W. Huang, "Adaptive deep learning-based air quality prediction model," *IEEE Access*, vol. 6, pp. 38186–38199, 2018.
2. S. Kumar and A. Jasuja, "Air quality monitoring system based on IoT using Raspberry Pi," in *Proc. ICCCA*, 2017, pp. 1341–1346.
3. P. D. Rosero-Montalvo et al., "Air quality monitoring intelligent system using machine learning," in *Proc. INCISCOS*, 2018, pp. 75–80.
4. [12] Y. C. Wu et al., "Air quality monitoring using mobile microscopy and machine learning," *Light Sci. Appl.*, vol. 6, no. 9, p. e17046, 2017.
5. X. Li et al., "Deep learning architecture for air quality predictions," *Environ. Sci. Pollut. Res.*, vol. 23, no. 22, pp. 22408–22417, 2016.
6. B. S. Freeman et al., "Forecasting air quality time series using deep learning," *J. Air Waste Manag. Assoc.*, vol. 68, no. 8, pp. 866–886, 2018.
7. N. Mehdiyev et al., "Time series classification using deep learning for process planning," *Procedia Comput. Sci.*, vol. 114, pp. 242–249, 2017.
8. H. Arora and A. Solanki, "Prediction of air quality index in metro cities using time series forecasting models," unpublished.
9. T. Rakthanmanon et al., "Searching and mining trillions of time series subsequences under dynamic time warping," in *Proc. ACM SIGKDD*, 2012, pp. 262–270.
10. E. Keogh and C. A. Ratanamahatana, "Exact indexing of dynamic time warping," *Knowl. Inf. Syst.*, vol. 7, no. 3, pp. 358–386, 2005.
11. Y. Zheng et al., "Forecasting fine-grained air quality based on big data," in *Proc. ACM SIGKDD*, 2015, pp. 2267–2276.
12. A.Yamini. (2026). AI-Augmented Fault-Tolerant Control Architecture for High-Reliability Electric Drive Systems in Smart Manufacturing. *National Journal of Electric Drives and Control Systems*, 16-23.
13. Olivia Myles, Julian Hartsfield. (2026). Impact of Warm-Start Initialization Policies on Multi-Run Model Training Dynamics. *Journal of Engineering, Signal, and Computer Technologies*, 13(2), 1–5.
14. Namrata Mishra. (2025). A Multidisciplinary Exploration of Intelligent Systems: Technical Advances and Societal Implications. *Bridge: Journal of Multidisciplinary Explorations*, 1(3), 27-33.
15. K. Geetha, &P.Dineshkumar. (2025). Multi-Objective Mathematical Optimization Framework for Renewable Energy Systems. *Journal of Applied Mathematical Models in Engineering*, 1(3), 8–13.
16. P.Dineshkumar. (2025). Managing Cultural Authenticity in Tourism Destinations: Impacts on Tourist Satisfaction and Revisit Intention. *Journal of Tourism, Culture, and Management Studies*, 2(3), 1–9.