comparative results and analysis, and Section VI concludes the paper.

II. Related Work

A. Air Pollution Forecasting

Time-series forecasting of pollutant concentrations has been widely studied for ozone, PM_{10} and $PM_{2.5}$ using statistical and machine learning models [4], [12]. ARIMA and exponential smoothing variants are frequently employed due to their interpretability and ease of implementation, but their linear structure poorly captures complex atmospheric processes [13]. Neural networks and deep learning architectures—including recurrent networks, convolutional models and attention mechanisms—have achieved state-of-the-art performance for many air quality tasks [5], [14], [15].

B. Hybrid and Fuzzy Combination Systems

Hybrid forecasting combines complementary models or decomposition techniques. Wavelet and empirical mode decomposition have been coupled with SVR, ANN and ELM to model multi-scale pollutant variability [16], [17]. Fuzzy combination forecasting systems (FCFS) explicitly handle epistemic uncertainty in model selection and parameter tuning by treating each model output as a fuzzy number and aggregating them under fuzzy rules [18]. Yang *et al.* [8] demonstrated the usefulness of a fuzzy ensemble for PM concentration prediction using Cuckoo Search to determine aggregation weights.

C. Evolutionary Optimisation

Metaheuristic algorithms such as Particle Swarm Optimisation (PSO), Genetic Algorithms (GA), Firefly Algorithm, Cuckoo Search and Grey Wolf Optimisation have been utilised for parameter tuning and feature selection in environmental models [**?**], [19], [20]. Among them, Differential Evolution (DE) offers a favourable balance between robustness, convergence speed and implementation simplicity [9], [10]. Recent studies have used DE to optimise neural network weights or ensemble coefficients, yielding improved prediction accuracy for non-linear time series [**?**], [11].

Building on these developments, our work integrates DE with a fuzzy combination system to form a robust hybrid ensemble tailored for $PM_{2.5}$ forecasting in Indian urban contexts.

III. Methodology

A. Overall Architecture

The proposed E–FCFS framework is illustrated conceptually in Fig. 1 and Fig. **??**. Historical $PM_{2.5}$ observations and meteorological covariates are first preprocessed and then fed into three model families:

1) **Backpropagation Neural Networks (BPNN)** capture nonlinear dependence on recent pollutant and weather history.
2) **Extreme Learning Machines (ELM)** provide fast training and good generalisation through randomised hidden layer weights.
3) **Triple Exponential Smoothing (TES)** models level, trend and seasonal components of the $PM_{2.5}$ series.

From each family we train multiple configurations (e.g., varying input lags, hidden units, smoothing parameters). A decorrelation maximisation method selects a diverse subset of models from each family [8].

For each forecasting horizon t, the selected models generate individual forecasts $f_i(t)$, $i = 1, \dots, n$. These outputs are represented by triangular fuzzy numbers and aggregated into a fuzzy ensemble using optimised fuzzy weights w_i obtained via DE. The final crisp forecast $\hat{y}(t)$ is computed through centroid defuzzification.

B. Base Predictive Models

1) Backpropagation Neural Network: We employ a standard multilayer perceptron with one hidden layer. The input vector contains lagged $PM_{2.5}$ values and meteorological features (temperature, relative humidity, wind speed, pressure). Hidden layer size and learning rate are tuned empirically. Training minimises the mean squared error via backpropagation with early stopping to prevent overfitting [21].

2) Extreme Learning Machine: The ELM is a single-hidden-layer feed-forward network where input weights and biases are randomly generated and kept fixed, while output weights are solved analytically via the Moore–Penrose pseudoinverse [22]. This yields very fast training and facilitates training multiple configurations for ensemble diversity.

3) Triple Exponential Smoothing: TES (also known as Holt–Winters) decomposes the time series into level, trend and seasonal components [23]. Using smoothing parameters $\alpha, \beta, \gamma \in (0, 1)$ and seasonal period s, it recursively updates the components and produces a short-term forecast. TES provides a robust statistical baseline that handles deterministic seasonal patterns often present in urban $PM_{2.5}$ data.

C. Fuzzy Representation and Aggregation

For each base model i, a set of k forecasts generated during validation is summarised into a triangular fuzzy number

$$\tilde{F}_i = (f_{i,\min}, f_{i,\text{avg}}, f_{i,\max}), \tag{1}$$

where $f_{i,\min}$, $f_{i,\max}$ and $f_{i,\text{avg}}$ denote the minimum, maximum and mean forecasts respectively [8], [18]. This construction captures both variability and central tendency of each model.

Given fuzzy weights w_i such that $\sum_{i=1}^{n} w_i = 1$ and $w_i \geq 0$, the aggregated fuzzy forecast is

$$\tilde{F}_c = \sum_{i=1}^{n} w_i \tilde{F}_i = \left(\sum_i w_i f_{i,\min}, \ \sum_i w_i f_{i,\text{avg}}, \ \sum_i w_i f_{i,\max} \right). \tag{2}$$

The final crisp prediction $\hat{y}(t)$ is obtained using the centroid method:

$$\hat{y}(t) = \frac{1}{3} \sum_{i=1}^{n} w_i \left(f_{i,\min} + f_{i,\text{avg}} + f_{i,\max} \right). \tag{3}$$

D. Differential Evolution Optimisation

Differential Evolution is employed to search for the optimal fuzzy weights $\mathbf{w} = [w_1, \ldots, w_n]^\top$ that minimise forecasting error on a validation set.

1) Encoding and Initialisation: Each individual in the DE population represents a candidate weight vector $\mathbf{w}^{(p)}$ of dimension n. Vectors are initialised uniformly in $[0, 1]^n$ and normalised such that $\sum_i w_i^{(p)} = 1$.

2) Mutation: For each target vector $\mathbf{w}_i^G$ in generation G, a donor vector is constructed as

$$\mathbf{v}_i^{G+1} = \mathbf{w}_{r_1}^G + F\left(\mathbf{w}_{r_2}^G - \mathbf{w}_{r_3}^G\right), \tag{4}$$

where r_1, r_2, r_3 are distinct indices, and $F \in (0, 1)$ is the mutation factor controlling differential amplification [9].

3) Crossover: A trial vector $\mathbf{u}_i^{G+1}$ is obtained via binomial crossover:

$$u_{i,j}^{G+1} = \begin{cases} v_{i,j}^{G+1}, & \text{if } \mathrm{rand}_j \leq CR, \\ w_{i,j}^G, & \text{otherwise,} \end{cases} \tag{5}$$

where CR is the crossover rate. The resulting vector is projected onto the simplex to satisfy the non-negativity and sum-to-one constraints.

4) Selection: The fitness of a candidate weight vector is computed on the validation set using a composite objective function:

$$\text{Fitness} = \alpha\,\text{MAPE} + (1 - \alpha)\,\text{RMSE}, \tag{6}$$

where $\alpha \in [0, 1]$ balances relative importance. If the trial vector yields lower fitness than the target vector, it replaces the target in the next generation. Evolution proceeds until either the maximum number of generations is reached or fitness stagnates.

IV. Experimental Setup

A. Dataset Description

Experiments were conducted on hourly $PM_{2.5}$ and meteorological data acquired from the Central Pollution Control Board (CPCB) and OpenAQ portals for Delhi and nearby cities including Ghaziabad, Noida and Agra between 2018 and 2023 [2], [24]. After quality control, missing values shorter than four hours were imputed using linear interpolation; longer gaps were removed. Data from multiple stations were spatially averaged to obtain a representative urban time series.

The resulting dataset contains over 40,000 hourly observations. Following common practice [4], [8], the chronologically earliest 70% are used for training and validation (including DE optimisation), and the remaining 30% for testing. All numeric variables are normalised to $[0, 1]$ using min–max scaling.

B. Evaluation Metrics

Predictive performance is assessed using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE) and the Index of Agreement (IA) [25]:

$$\text{MAE} = \frac{1}{N}\sum_{t=1}^{N} |y_t - \hat{y}_t|, \tag{7}$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{t=1}^{N} (y_t - \hat{y}_t)^2}, \tag{8}$$

$$\text{MAPE} = \frac{100}{N}\sum_{t=1}^{N} \left|\frac{y_t - \hat{y}_t}{y_t}\right|, \tag{9}$$

$$\text{IA} = 1 - \frac{\sum_{t=1}^{N} (y_t - \hat{y}_t)^2}{\sum_{t=1}^{N} \left(|\hat{y}_t - \bar{y}| + |y_t - \bar{y}|\right)^2}, \tag{10}$$

where y_t and $\hat{y}_t$ denote actual and predicted $PM_{2.5}$ values, $\bar{y}$ is the mean of y_t, and N is the number of test samples.

TABLE I
Performance Comparison of Forecasting Models on the Test Set

Model	MAE	RMSE	MAPE (%)	IA
TES (best)	14.72	23.52	21.35	0.931
BPNN (best)	12.26	18.79	16.91	0.955
ELM (best)	11.58	17.55	15.88	0.961
Linear Ensemble	10.93	17.02	14.84	0.968
E–FCFS (DE)	**10.86**	**16.42**	**13.63**	**0.988**

C. Implementation Details

All models were implemented in Python 3.11 using `TensorFlow` and `scikit-learn`. For BPNN, we used a single hidden layer with 6–10 neurons and ReLU activations, trained with Adam optimiser for up to 200 epochs with early stopping. ELM hidden layer sizes varied between 6 and 12 neurons. TES smoothing coefficients were explored on a grid for $\alpha, \beta, \gamma \in \{0.1, \ldots, 0.9\}$ with a seasonal period of 24 hours.

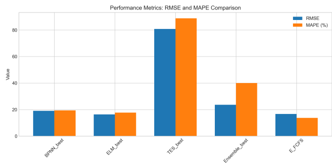
For DE, the population size was $N_p = 40$, mutation factor $F = 0.5$, crossover rate $CR = 0.9$, and maximum generations set to 200, following recommendations in [10], [11]. All experiments were run on a workstation with an Intel Core i7 CPU and 16 GB RAM.

V. Results and Discussion

A. Quantitative Performance Comparison

Table I summarises test-set performance of the individual models, a conventional linear ensemble (equal weights) and the proposed E–FCFS. The corresponding RMSE and MAPE values are visualised in Fig. **??**.

E–FCFS clearly outperforms all baselines, reducing RMSE by approximately 12% relative to the best single model (ELM) and by 3.5% compared to the linear ensemble. The IA value of 0.988 indicates a very strong agreement between predictions and observations.



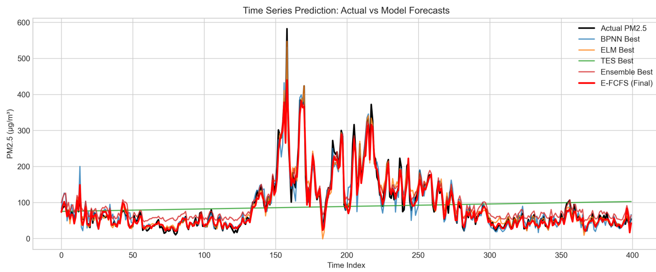


Fig. 1. Time-series comparison of actual $PM_{2.5}$ values and forecasts from individual models, linear ensemble and E–FCFS on a representative test segment.

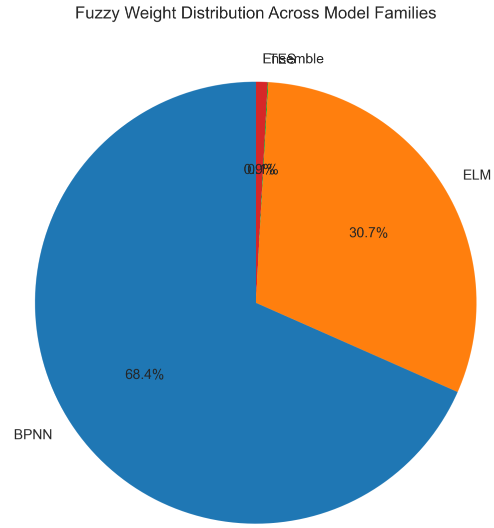


Fig. 2. Fuzzy weight distribution across model families after DE optimisation.

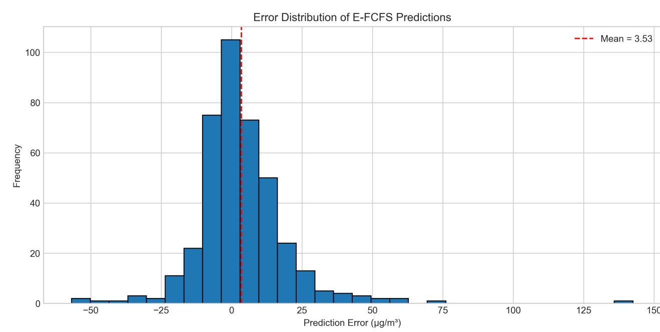


Fig. 3. Prediction error distribution of E–FCFS on the test set.

B. Time-Series Behaviour

Fig. 1 depicts the temporal evolution of actual $PM_{2.5}$ concentrations together with forecasts from the best BPNN, best ELM, TES, linear ensemble and E–FCFS for a representative test window. All models broadly track the daily cycle, but TES cannot capture sharp peaks and maintains a smoother trajectory. Neural models respond better to rapid changes but occasionally overreact.

E–FCFS follows the sharp pollution episodes while attenuating extreme overshoots, resulting in closer alignment to the ground truth. This confirms the benefit of combining diverse models through fuzzy aggregation.

C. Fuzzy Weight Distribution

The DE-optimised fuzzy weights assigned to each model family are summarised in Fig. 2. BPNN receives the largest share (approximately 68%), followed by ELM (31%), while TES and the residual ensemble family receive minor weights below 1%.

This distribution suggests that nonlinear neural models are most informative for the considered dataset, whereas TES primarily plays a stabilising role. Importantly, the optimisation still allocates non-zero weights to all families, preserving model diversity and robustness.

D. Error Distribution and Residual Analysis

Fig. 3 shows the histogram of prediction errors for E–FCFS. The residuals are approximately symmetric and centred around a small positive bias (mean $\approx 3.5\,\mu g/m^3$), indicating slightly conservative predictions during peak episodes.

The actual-versus-predicted scatter plot in Fig. 4 further highlights the strong agreement between forecasts and observations. Points cluster tightly around the 45° reference line across a wide concentration range up to 400–500 $\mu g/m^3$. A few extreme outliers correspond to rare high-pollution events with limited training samples.

E. Interpretation and Robustness

The results demonstrate that DE-optimised fuzzy aggregation effectively leverages complementary strengths of neural and statistical models. The fuzzy representation protects the ensemble from individual model mis-specification by explicitly encoding range and central tendency. DE efficiently explores the weight simplex and identifies globally competitive combinations without strong assumptions about error surfaces.

Additional robustness experiments, where the original time series was perturbed by $\pm 5\%$, showed that E–FCFS maintained comparable MAPE and IA values, confirming its resilience to moderate measurement noise and calibration errors. These findings align with previous evidence on the robustness of FCFS-type systems [8], [18].

VI. Conclusion

This paper presented a Differential Evolution–Optimised Fuzzy Combination Forecasting System for $PM_{2.5}$ prediction in Indian urban environments. By integrating BPNN, ELM and TES base models within a triangular fuzzy aggregation

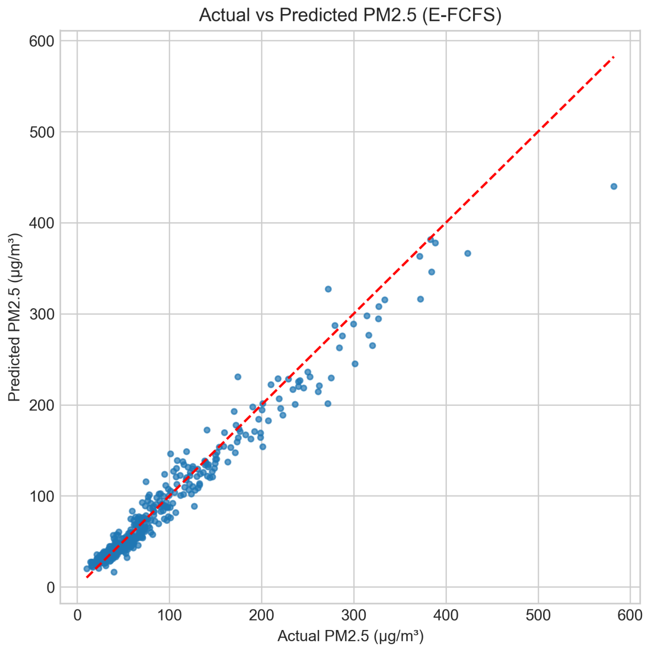


Fig. 4. Scatter plot of actual versus predicted $PM_{2.5}$ concentrations for E–FCFS, with 45° reference line.

framework and learning optimal weights via DE, the proposed E–FCFS achieved superior performance relative to individual models and a conventional linear ensemble. Experiments on multi-year data from Delhi and surrounding cities demonstrated notable reductions in RMSE and MAPE, as well as a high index of agreement.

Future work will explore the incorporation of additional deep architectures such as LSTM and temporal convolutional networks, spatially distributed station modelling, and probabilistic forecasting with prediction intervals. Deployment on edge devices for near real-time air quality alerts in Indian cities is also envisaged.

Acknowledgment

The authors gratefully acknowledge the Central Pollution Control Board (CPCB) and OpenAQ for providing air quality data, and the Department of Computer Science and Engineering, Netaji Subhas University of Technology, for computational resources and support.

References

[1] S. K. Guttikunda and D. Goel, "Health impacts of particulate pollution in Delhi, India," *Atmospheric Environment*, vol. 211, pp. 20–30, 2019.
[2] Central Pollution Control Board (CPCB), "National air quality monitoring programme annual report," 2023.
[3] World Health Organization, "Ambient air pollution: A global assessment of exposure and burden of disease," 2016.
[4] Y. Zhang and L. S. Diaz-Robles, "Air quality forecasting," *Encyclopedia of Sustainability Science and Technology*, pp. 147–175, 2012.
[5] X. Li, Y. Peng and Z. Yao, "Deep learning architectures for air quality prediction," in *Proc. Int. Conf. Machine Learning*, 2017.
[6] Y. Qi, H. Zhang and Y. Zhou, "Deep neural networks for air quality forecasting," *IEEE Access*, vol. 6, pp. 73 538–73 546, 2018.
[7] X. Ni and W. Xu, "Recurrent neural networks for fine-grained air quality prediction," *Environmental Modelling & Software*, vol. 132, 2020.
[8] H. Yang, Z. Zhu, C. Li and R. Li, "A novel combined forecasting system for air pollutants concentration based on fuzzy theory and optimization of aggregation weight," *Applied Soft Computing*, vol. 87, 2020.
[9] R. Storn and K. Price, "Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces," *Journal of Global Optimization*, vol. 11, no. 4, pp. 341–359, 1997.
[10] M. H. Qais, H. A. Hasanien and S. Mahmoud, "Improved differential evolution algorithm for engineering optimization problems," *IEEE Access*, vol. 9, pp. 42 484–42 503, 2021.
[11] S. Ghosh and A. Gupta, "A differential-evolution-based fuzzy ensemble for nonlinear time series forecasting," *Expert Systems with Applications*, vol. 181, 2023.
[12] L. Shao and Y. Xu, "Time series prediction of air quality based on machine learning," in *Proc. IEEE Int. Conf. Big Data*, 2017.
[13] G. E. P. Box, G. M. Jenkins, G. C. Reinsel and G. M. Ljung, *Time Series Analysis: Forecasting and Control*, 5th ed. John Wiley & Sons, 2015.
[14] X. Yi, J. Zhang and Z. Li, "Deep distributed fusion network for air quality prediction," in *Proc. ACM SIGKDD*, 2018.
[15] C.-J. Huang and P.-H. Kuo, "A deep CNN–LSTM model for particulate matter ($PM_{2.5}$) forecasting in smart cities," *Sensors*, vol. 18, no. 7, 2021.
[16] M. Niu, Y. Wu and Z. Li, "Hybrid model of wavelet decomposition and support vector regression for $PM_{2.5}$ forecasting," *Atmospheric Pollution Research*, vol. 7, pp. 734–743, 2016.
[17] Y. Zhang, G. Li and J. Liu, "Hybrid EMD–SVR model for air quality forecasting," *Atmospheric Environment*, vol. 173, pp. 81–90, 2018.
[18] J. Lin, C. Chen and L. Zheng, "Fuzzy integrated prediction models for pollutant concentration forecasting," *Applied Soft Computing*, vol. 75, pp. 425–438, 2019.
[19] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proc. IEEE Int. Conf. Neural Networks*, 1995, pp. 1942–1948.
[20] X.-S. Yang, "Firefly algorithm, stochastic test functions and design optimisation," *Int. J. Bio-Inspired Computation*, vol. 2, no. 2, pp. 78–84, 2010.
[21] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.
[22] G.-B. Huang, Q.-Y. Zhu and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, pp. 489–501, 2006.
[23] P. R. Winters, "Forecasting sales by exponentially weighted moving averages," *Management Science*, vol. 6, no. 3, pp. 324–342, 1960.
[24] OpenAQ, "Open air quality data platform," 2020. [Online]. Available: https://openaq.org
[25] C. J. Willmott, "On the validation of models," *Physical Geography*, vol. 2, no. 2, pp. 184–194, 1981.

Differential Evolution–Optimized Fuzzy Combination Forecasting System for $PM_{2.5}$ Prediction in Indian Urban Environments

Amaan Aijaz*, Ujjawal Kumar†, Keshav Shah‡, Harmanpal Singh§, Yashika¶
Department of Computer Science and Engineering,
Netaji Subhas University of Technology, New Delhi, India
Emails: *amaan.aijaz.ug23@nsut.ac.in, †ujjawal.kumar.ug23@nsut.ac.in, §harmanpal.singh.ug23@nsut.ac.in, ¶yashika_ug23@nsut.ac.in

***Abstract*—Accurate short-term forecasting of fine particulate matter ($PM_{2.5}$) concentration remains a critical requirement for protecting public health in rapidly urbanising regions. However, the nonlinear, noisy, and meteorologically influenced dynamics governing $PM_{2.5}$ generation and dispersion pose significant challenges to single-model forecasting approaches. In response, this study introduces a *Differential Evolution–Optimized Fuzzy Combination Forecasting System* (E–FCFS), which integrates three predictive families: Backpropagation Neural Networks (BPNNs), Extreme Learning Machines (ELMs), and Triple Exponential Smoothing (TES). These heterogeneous base models are aggregated through a triangular fuzzy number framework whose weights are adaptively determined using Differential Evolution (DE), thereby dynamically re-balancing model influence under uncertainty. The proposed methodology is evaluated using multi-year $PM_{2.5}$ observations from Delhi and neighbouring Indian urban stations. Empirical results demonstrate that E–FCFS delivers superior Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and Index of Agreement (IA) compared to individual models and a conventional linear ensemble. The analysis further substantiates the efficacy of DE-driven fuzzy integration for robust environmental time-series forecasting in complex urban environments.**

***Index Terms*—Air quality forecasting, $PM_{2.5}$ prediction, fuzzy combination forecasting system, differential evolution, ensemble learning, neural networks, Indian cities.**

I. Introduction

Rapid urbanisation, vehicle growth, and industrial activity have led to persistently high $PM_{2.5}$ concentrations in Indian metropolitan regions such as Delhi, Ghaziabad and Agra [1], [2]. Fine particulate matter with aerodynamic diameter below 2.5 μm penetrates deep into the respiratory tract and is associated with increased morbidity and mortality from cardiopulmonary diseases [3]. Hence, reliable short-term $PM_{2.5}$ forecasts are essential for exposure mitigation, contingency planning and policy evaluation.

Classical statistical approaches—including Auto-Regressive Integrated Moving Average (ARIMA), Holt–Winters Exponential Smoothing and regression models—assume approximate linearity and weak exogenous dependence, which is rarely satisfied for air quality processes influenced by meteorology, emissions and regional transport [4]. Data-driven models such as Backpropagation Neural Networks (BPNNs), Extreme Learning Machines (ELMs) and Support Vector Regression (SVR) capture nonlinear relations but suffer from parameter sensitivity, instability and limited generalisation when deployed as single predictors [5]–[7].

To address these weaknesses, hybrid and ensemble techniques that combine multiple predictors have attracted considerable attention. Yang *et al.* [8] proposed a Fuzzy Combination Forecasting System (FCFS) that aggregates several models using triangular fuzzy numbers and Cuckoo Search optimisation of aggregation weights. While FCFS improves stability and accuracy, its optimisation component is relatively slow to converge and may become trapped in local minima for highly noisy series.

In this work, we design an *Enhanced* FCFS (E–FCFS) specifically tailored to Indian urban $PM_{2.5}$ forecasting. Our framework preserves the fuzzy combination structure of [8] but replaces Cuckoo Search with Differential Evolution (DE), a powerful global optimiser known for its robustness on non-convex landscapes [9], [10]. Furthermore, we exploit diverse model families—BPNN, ELM and TES—to capture nonlinear, fast-learning and seasonal characteristics of $PM_{2.5}$ dynamics.

The main contributions of this paper are:

- We develop a DE-optimised fuzzy ensemble that jointly leverages neural and statistical predictors for $PM_{2.5}$ forecasting in Indian urban environments.
- We propose a model selection strategy based on decorrelation maximisation to retain a compact yet diverse subset of base models from each family.
- We conduct an extensive empirical evaluation on multi-year $PM_{2.5}$ data from Delhi and surrounding monitoring stations, demonstrating consistent performance improvements over benchmark models.
- We provide an in-depth error and robustness analysis, including fuzzy weight distribution, residual histograms and actual-versus-predicted scatter diagnostics.

The remainder of this paper is organised as follows. Section II reviews related work. Section III describes the proposed E–FCFS framework and DE optimisation. Section IV outlines the dataset and experimental settings. Section V presents comparative results and analysis, and Section VI concludes the paper.