NOVEL HYBRID DEEP LEARNING-BASED MODELS FOR AIR QUALITY PREDICTION

ZHOU XIAOJIE

UNIVERSITI SAINS MALAYSIA

2026

NOVEL HYBRID DEEP LEARNING-BASED MODELS FOR AIR QUALITY PREDICTION

by

ZHOU XIAOJIE

Thesis submitted in fulfilment of the requirements
for the degree of
Doctor of Philosphy

April 2026

ACKNOWLEDGEMENT

I would like to express my deepest gratitude to all those who have supported and guided me throughout the course of this research and the writing of this dissertation. First and foremost, I am profoundly grateful to my supervisors, Professor Farah Aini Binti Abdullah and Professor Majid Khan bin Majahar Ali. Professor Farah Aini's extensive knowledge, rigorous scholarship, and patient guidance have been invaluable in shaping the direction of this work, while Professor Majid Khan's insightful advice on methodology and writing has continually helped me refine and strengthen my research. Their encouragement at every critical stage has been a constant source of inspiration and motivation. I also wish to thank my parents for their unconditional love and sacrifices, which have given me both the courage and confidence to pursue my dreams. To my husband, Li Kaihui, thank you for your understanding, encouragement, and unwavering support—your care and comfort during late nights of writing have meant more than words can express. My sincere thanks go to my lab colleagues and friends for their camaraderie, constructive discussions, and assistance with data analysis and manuscript revisions. I am also indebted to the senior students whose guidance helped me avoid many pitfalls along the way. Finally, I gratefully acknowledge the funding agencies and research programs that provided the resources necessary to carry out this study. To everyone who has contributed—directly or indirectly—to this work, I extend my heartfelt appreciation.

TABLE OF CONTENTS

ACKNOWLEDGEMENT ..... **ii**

TABLE OF CONTENTS ..... **iii**

LIST OF TABLES ..... **vi**

LIST OF FIGURES ..... **viii**

LIST OF SYMBOLS ..... **ix**

LIST OF ABBREVIATIONS ..... **x**

ABSTRAK ..... **xii**

ABSTRACT ..... **xiii**

CHAPTER 1 INTRODUCTION **1**

1.1 Problem Background 1

1.2 Problem Statement 4

1.3 Objectives of the Study 8

1.4 Scope and Limitation 8

1.5 Significance of the Study 12

1.6 Thesis Outlines 13

CHAPTER 2 LITERATURE REVIEW **16**

2.1 Data 16

2.1.1 Penang Data 16

2.1.2 Data Impact: Air Pollution and Health 20

2.2 Heterogeneity in Air Pollution 23

2.2.1 Spatial Heterogeneity of Air Pollution 23

2.3 Temporal Variability of Air Pollution 25

2.3.1 Spatio-temporal Characteristics of Air Pollution 27

2.4 Stations 29

2.4.1 Single Station Studies 29

2.4.2 Multiple Station Studies 31

2.4.3 Existing Models and Optimization 33

2.5 Research Gaps and Challenges 36

2.6 Initial Summary 39

CHAPTER 3 METHODOLOGY 40

3.1 Overview of the Proposed Research Framework 41

3.2 Data Description and Preprocessing 43

3.2.1 Preprocessing 46

3.3 Fundamental Components of the Proposed Framework 47

3.3.1 Seasonal–Trend Decomposition using Loess and Variational Mode Decomposition (STL–VMD) 49

3.3.2 Convolutional Neural Network (CNN) 51

3.3.3 Long Short-Term Memory (LSTM) Network 52

3.3.4 Attention Mechanism 54

3.3.5 Seasonal Modeling with Temporal Convolutional Network (TCN) 55

3.3.6 Residual Modeling with Gated Recurrent Unit (GRU) 57

3.3.7 Dung Beetle Optimization (DBO) Algorithm 58

3.4 Improved Dung Beetle Optimization (IDBO) Algorithm 60

3.4.1 Algorithm Flowchart 61

3.4.2 Technique for Golden Sine Optimization 61

3.4.3 Self-Spiral Strategy 61

3.4.4 Levy Flying 63

3.4.5 Adaptive t-Distribution Scheme 64

3.5 Hybrid Single-Station Model: IDBO–CNN–LSTM–Attention 64

3.5.1 Objective and Motivation 65

3.5.2 Overview and Pseudocode 65

3.5.3 Hyperparameters Optimized by IDBO 65

3.6 Single-Station Univariate Model: IDBO–CNN 67

3.6.1 Objective and Motivation 68

3.6.2 Model Specification 68

3.6.3 IDBO for CNN Hyperparameter Optimization 68

3.6.4 Pseudocode (IDBO–CNN) 70

3.7 Multi-Station Hybrid Spatiotemporal Model: STV–LTG 70

3.7.1 Objective and Motivation 70

3.7.2 Model Architecture 71

3.7.3 IDBO Optimization Procedure for STV–LTG 72

3.7.4 Search Space Design for IDBO in STV–LTG 73

3.8 Evaluation Metrics 77

3.9 Summary 78

REFERENCES 81

LIST OF TABLES

Page

Table 1.1	Classification of Malaysia's Air Quality Index (API).	2
Table 2.1	Representative studies using Penang air quality datasets and their methodological characteristics.	19
Table 2.2	Major air pollutants, their sources, and health impacts.	20
Table 2.3	Recent studies on spatial heterogeneity of air pollution (2020–2025).	24
Table 2.4	Recent studies on temporal variability of air pollution (2020–2025).	26
Table 2.5	Recent studies on spatio-temporal characteristics of air pollution (2020–2025).	28
Table 2.6	Recent studies on single-station air quality forecasting (2020–2025). Single-station approaches cover statistical, machine learning, and deep learning methods, with applications across Asia, Europe, and North America.	30
Table 2.7	Recent studies on multi-station air quality forecasting (2020–2025). Multi-station approaches integrate spatial correlations and temporal dynamics through econometric, statistical, and deep learning models.	32
Table 2.8	Recent optimization techniques for hyperparameter tuning in air quality forecasting (2020–2025).	34
Table 2.9	Summary of research gaps and challenges in air quality forecasting.	38
Table 3.1	Summary of datasets employed in this study.	44
Table 3.2	Summary of preprocessing procedures applied to DS2–DS4.	48
Table 3.3	Search space used by IDBO for the CNN–LSTM–Attention model.	67
Table 3.4	Search space used by IDBO for the CNN-only forecaster.	69
Table 3.5	Search space for IDBO in the STV–LTG pipeline. Period P for STL is fixed by sampling (e.g., P=24 for hourly).	75

Table 3.6 Summary of traditional baselines and proposed hybrid models… 76

LIST OF FIGURES

Page

Figure 1.1 Monthly mean $PM_{2.5}$ in Penang (2018–2022) with WHO guideline thresholds (annual 5 and daily 15 $\mu g\,m^{-3}$). 3

Figure 1.2 The research framework 15

Figure 3.1 Integrated methodological framework of the four sub-studies based on IDBO optimization for air quality forecasting. 42

Figure 3.2 Architecture of the Convolutional Neural Network (CNN) 52

Figure 3.3 Structure of the Long Short-Term Memory (LSTM) network 53

Figure 3.4 Flowchart of IDBO algorithm. 62

Figure 3.5 Overview of the IDBO-CNN-LSTM-Attention workflow. 66

Figure 3.6 Overview of the IDBO-CNN workflow. 69

Figure 3.7 Flowchart of a Spatiotemporal Forecasting Framework Based on STL-VMD Decomposition and Deep Learning Fusion. 74

LIST OF SYMBOLS

lim	limit
θ	angle in radians

LIST OF ABBREVIATIONS

AQI	Air Quality Index
CNN	Convolutional Neural Networks
Bilstm	Bidirectional Long Short-Term Memory Networks
Attention	Attention Mechanism
ARIMA	Autoregressive Integrated Moving Average
LSTM	Long Short-Term Memory Networks
GRU	Gated Recurrent Unit
DBO	Dung Beetle Optimization Algorithm
IDBO	Improved Dung Beetle Optimization Algorithm
VMD	Variational Mode Decomposition
STL	Seasonal-Trend Decomposition using Loess
$PM_{2.5}$	Particulate Matter with a diameter of less than 2.5 μm
NO_2	Nitrogen Dioxide
RF	Random Forest
RMSE	Root Mean Square Error
HPO	Hyperparameter optimization
SI	Swarm Intelligence
MAE	Mean Absolute Error
MSE	Mean Squared Error
R^2	Coefficient of Determination
CNN	Convolutional Neural Network
MLP	Multilayer Perceptron

GWO	Grey Wolf Optimizer
WOA	Whale Optimization Algorithm
NGO	Northern Goshawk Optimization
AT	Swarm Intelligence

MODEL BERASASKAN PEMBELAJARAN DALAM HIBRID NOVEL UNTUK RAMALAN KUALITI UDARA

ABSTRAK

Disertasi ini mempersembahkan kerangka kerja komprehensif bagi model hibrid berasaskan pembelajaran mendalam yang baharu untuk ramalan kualiti udara, berteraskan empat kajian yang saling berkait. Pertama, algoritma pengoptimuman kumbang tahi yang dipertingkatkan (IDBO) dibangunkan dengan memperhalusi strategi mencari makanan dan kawalan parameter adaptifnya, menghasilkan konvergensi lebih pantas serta keseimbangan eksplorasi–eksploitasi yang lebih baik. Kedua, IDBO diintegrasikan ke dalam seni bina CNN–LSTM–Perhatian untuk melaksanakan ramalan siri masa multivariat bagi pembolehubah meteorologi dan pencemar; eksperimen menunjukkan model hibrid ini secara ketara mengatasi model asas konvensional dari segi ketepatan ramalan dan ketahanan. Ketiga, keserbagunaan IDBO dinilai merentasi satu set kaedah pembelajaran mesin dan pembelajaran mendalam untuk ramalan siri masa univariat; keputusan menunjukkan konfigurasi IDBO–CNN sangat sesuai untuk ramalan jangka panjang, mencapai kadar ralat yang jauh lebih rendah. Akhir sekali, satu skim ramalan teragih—undian ruang–masa digabungkan dengan DBO dan pintu temporal jangka panjang (STV-DBO-LTG)—dicadangkan untuk ramalan serentak $PM_{2.5}$ dan NO_2 merentasi beberapa stesen pemantauan, menunjukkan penggeneralisasian ruang yang unggul dan konsistensi. Eksperimen menyeluruh ke atas set data dunia sebenar dari 2018 hingga 2021 mengesahkan keberkesanan, kebolehskalaan, dan kebolehterapan kaedah yang dicadangkan. Keseluruhannya, sumbangan ini memajukan teori dan amalan pembelajaran mendalam berpandukan pengoptimuman untuk pemantauan alam sekitar, menawarkan alat berimpak tinggi bagi pengurusan kualiti udara yang lebih boleh dipercayai serta sokongan dasar.

NOVEL HYBRID DEEP LEARNING-BASED MODELS FOR AIR QUALITY PREDICTION

ABSTRACT

This dissertation presents a comprehensive framework of novel hybrid deep learning–based models for air quality prediction, built upon four interconnected studies. First, an improved dung beetle optimization algorithm (IDBO) is developed by refining its foraging strategies and adaptive parameter controls, yielding faster convergence and enhanced exploration–exploitation balance. Second, IDBO is embedded within a CNN–LSTM–Attention architecture to perform multivariate time-series forecasting of meteorological and pollutant variables; experiments demonstrate that this hybrid model significantly outperforms conventional baselines in predictive accuracy and robustness. Third, the versatility of IDBO is evaluated across a suite of machine learning and deep learning methods for univariate time-series prediction; results indicate that the IDBO–CNN configuration is particularly well suited for long-term forecasts, achieving markedly lower error rates. Finally, a distributed prediction scheme—spatial–temporal voting fused with DBO and long-term temporal gating (STV-DBO-LTG)—is proposed for simultaneous $PM_{2.5}$ and NO_2 forecasting across multiple monitoring stations, showing superior spatial generalization and consistency. Comprehensive experiments on real-world datasets spanning 2018–2021 validate the proposed methods' effectiveness, scalability, and applicability. Together, these contributions advance both the theory and practice of optimization-driven deep learning for environmental monitoring, offering actionable tools for more reliable air quality management and policy support.

CHAPTER 1

INTRODUCTION

1.1 Problem Background

Air pollution has emerged as one of the most pressing environmental challenges worldwide, and Malaysia is no exception. Rapid industrialization, urbanization, and economic growth have contributed to escalating emissions of pollutants in recent decades (Colvile et al., 2002; Gao et al., 2021; R. Sharma & Balasubramanian, 2020; H. Zhang et al., 2022). Penang, often referred to as the "Silicon Valley of the East," is home to the electronics and semiconductor industries and a dense network of small- and medium-sized enterprises (SMEs). Coupled with heavy traffic and increasing energy demand, these activities produce substantial emissions of fine particulate matter ($PM_{2.5}$), nitrogen dioxide (NO_2), ozone (O_3), sulfur dioxide (SO_2), and carbon monoxide (CO) (G. Han et al., 2021; X. Li et al., 2022; Seinfeld & Pandis, 2016).

Among these pollutants, $PM_{2.5}$ is of particular concern due to its ability to penetrate deep into the respiratory tract and enter the bloodstream. Numerous epidemiological studies have shown that long-term exposure to $PM_{2.5}$ significantly increases the risk of cardiovascular and pulmonary diseases, lung cancer, and premature mortality (Gupta et al., 2019; Hoek et al., 2013; Pope et al., 2009; World Health Organization, 2021). Short-term exposure is associated with acute respiratory symptoms, asthma exacerbations, and hospital admissions (Brook et al., 2010; Kim et al., 2015). Vulnerable groups such as children, the elderly, and patients with pre-existing cardiorespiratory conditions are especially affected. Recognizing these health impacts, the World Health Organization (WHO) updated its Air Quality Guidelines in 2021, recommending stricter limits for both annual and daily $PM_{2.5}$ exposure (World Health Organization, 2021).

Given these risks, accurate forecasting of air pollution has become a critical component of environmental management. Forecasting provides the basis for early warning systems, helps policymakers design timely interventions, and supports public aware-

ness initiatives (Duan & Liu, 2023; Gao et al., 2021; Grange & Carslaw, 2018; H. Zhang et al., 2022). For urban residents, reliable forecasts can guide outdoor activities and personal protection strategies, thereby reducing health risks (J. Chen et al., 2020; P. Wang et al., 2019). From a broader perspective, advanced spatiotemporal forecasting frameworks contribute to sustainable urban development by balancing economic growth with environmental protection goals (G. Han et al., 2021; X. Li et al., 2022; Y. Zhang et al., 2020). In Southeast Asia, haze episodes caused by local and transboundary biomass burning further underscore the urgent need for accurate and robust forecasting systems (Amil et al., 2016; Betha & Balasubramanian, 2013).

To communicate air quality conditions to the public, Malaysia employs the Air Pollution Index (API), which converts concentrations of key pollutants into a single scale for public dissemination. The API classifies air quality into five levels—Good, Moderate, Unhealthy, Very Unhealthy, and Hazardous—each associated with distinct health implications and recommended responses (Awang et al., 2000; Department of Environment Malaysia, 2017). Table 1.1 summarizes the classification system, which forms the basis for public health advisories, regulatory actions, and public communication in Malaysia.

Table 1.1: Classification of Malaysia's Air Quality Index (API).

Level	Range	Health Impact
Good	0–50	No effect on the general population; air quality is very good.
Moderate	51–100	Minor effects on sensitive groups (children, elderly, cardiorespiratory patients).
Unhealthy	101–200	Health effects may occur in the general population; sensitive groups are more affected.
Very Unhealthy	200–300	Significant health risks for all populations; outdoor activity should be reduced.
Hazardous	>300	Emergency condition; health warnings for all; protective measures required.

The trend in Fig. 1.1 shows that $PM_{2.5}$ concentrations in Penang fluctuated considerably during 2018–2019, with several months exceeding 30 $\mu g\,m^{-3}$, well above the WHO daily guideline of 15 $\mu g\,m^{-3}$. However, from 2020 onwards, a gradual decline

can be observed. This reduction is largely attributed to the enforcement of COVID-19 Movement Control Orders (MCOs), which curtailed industrial activities and vehicular emissions, thereby reducing particulate matter in the atmosphere. Subsequent years benefited from both policy-driven emission controls and favorable meteorological conditions, further sustaining lower $PM_{2.5}$ levels. Nonetheless, even in 2022, concentrations remain above the WHO annual guideline of 5 $\mu g\,m^{-3}$, highlighting the need for continuous monitoring and advanced forecasting models.

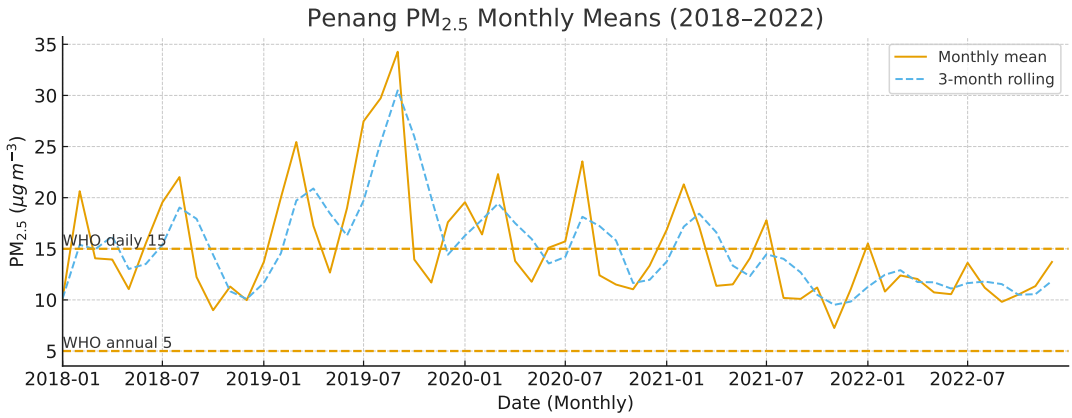

Figure 1.1: Monthly mean $PM_{2.5}$ in Penang (2018–2022) with WHO guideline thresholds (annual 5 and daily 15 $\mu g\,m^{-3}$).

These findings highlight the pressing need for accurate and timely forecasting of $PM_{2.5}$ and other pollutants in Penang. Reliable spatiotemporal prediction models can support proactive policy interventions, inform public health advisories, and contribute to sustainable urban and industrial development.

It should be noted that the dataset used in this study covers the period from 2018 to 2022. Due to data availability constraints, more recent observations (e.g., 2023 onwards) were not included. Nevertheless, this five-year span provides a sufficiently representative basis for characterizing Penang's air quality trends and for developing robust forecasting models.

1.2 Problem Statement

Although Penang has accumulated extensive air quality monitoring data from 2018 to 2022, several challenges remain inherent in the dataset itself. the observed concentrations exhibit strong temporal variability, with substantial differences across months and years. Pollution episodes are often abrupt and irregular, making it difficult to infer future trends from simple historical averages. spatial heterogeneity is evident across monitoring stations; the concentration levels and variation patterns differ by location, indicating that air pollution possesses significant spatial dependencies. pollutant levels are influenced by multiple interacting factors, such as traffic emissions, industrial activity, and meteorological conditions, resulting in highly nonlinear characteristics in the data. missing or incomplete records exist for certain variables—for example, CO measurements are largely absent—leading to imbalanced and imperfect datasets. while the monitoring data provide descriptive insights into past air quality, they lack predictive capacity, limiting their usefulness for timely public health advisories or proactive policy interventions. These limitations highlight the necessity of developing advanced spatiotemporal forecasting models that can effectively exploit the available data, capture both temporal dynamics and spatial correlations, and ultimately provide accurate and actionable predictions of air quality in Penang.

In the calibration of forecasting models, predictive performance depends heavily on the choice of optimizer. Existing metaheuristic algorithms, including the original dung beetle optimization (DBO), often struggle to maintain a balance between global exploration and local exploitation, which tends to cause premature convergence, unstable outcomes, and inefficiency in high-dimensional settings. Although DBO and its variants have shown promising in solving engineering optimization tasks, their convergence behavior is often inconsistent, especially when the search landscape is rugged or the parameter space is large, leading to situations where the algorithm becomes trapped in local minima and fails to identify high-quality solutions within a reasonable computational budget. Moreover, the canonical DBO relies on stochastic operators with

limited adaptability; once population diversity decreases, the optimizer lacks effective mechanisms to reintroduce variability or to adjust the search direction dynamically, resulting in stagnation of the search process. When applied to synthetic benchmark functions, such weaknesses become more evident: some algorithms converge rapidly in early iterations but sacrifice accuracy later on, while others preserve diversity at the cost of very slow convergence. This instability complicates the tuning of forecasting models, as optimal solutions are often sensitive to initialization strategies and parameter control, and the problem becomes even more severe in high-dimensional scenarios such as deep spatiotemporal neural networks, where the search process requires substantial computational resources yet frequently produces suboptimal hyperparameter configurations. These challenges highlight the urgent need for enhanced optimization frameworks that can accelerate convergence without compromising robustness, incorporate adaptive mechanisms to maintain diversity, and ensure stable exploration across multiple scales of the search space, thereby enabling the efficient and reliable calibration of complex spatiotemporal forecasting models for air quality prediction.

At the single-station level, hourly $PM_{2.5}$ concentrations and meteorological variables in Penang display clear multi-scale seasonality and rapid short-term fluctuations, making prediction especially challenging. Traditional regression models and shallow machine learning approaches are often inadequate because they either overfit transient noise or oversmooth long-term patterns, thereby failing to capture the inherent temporal complexity of the data. In practice, their predictive accuracy is highly sensitive to hyperparameter choices, feature selection, and data preprocessing strategies, meaning that even minor variations in input quality can lead to significant discrepancies in forecasting outcomes. Moreover, missing or incomplete records further exacerbate the problem, as models trained on imbalanced datasets tend to produce biased or unstable results, especially when critical predictors such as CO are absent for extended periods. These issues are compounded by the strong nonlinearity of pollutant dynamics, which are influenced by a multitude of interacting factors including traffic emissions, industrial output, and local meteorological variability. Standard models often lack

the representational capacity to disentangle such complex dependencies, resulting in unreliable forecasts that are difficult to generalize across different time frames. Consequently, single-station prediction suffers from instability, limited robustness, and a high sensitivity to noise, underscoring the need for more advanced approaches that can flexibly adapt to both short-term irregularities and long-term seasonal structures in the data.

In long-horizon prediction, additional difficulties arise from the compounding nature of forecasting errors, concept drift, and changing variance across extended time scales. As the prediction horizon increases, small inaccuracies in short-term forecasts accumulate, leading to systematic divergence from actual pollutant concentrations and making the results unreliable for practical applications. For example, forecasts based on a two-year $PM_{2.5}$ series often exhibit reasonable accuracy within the first few hours or days, but their performance degrades rapidly as the horizon extends, reflecting the inability of conventional models to capture evolving temporal dynamics. This challenge is further complicated by concept drift, whereby the underlying statistical properties of the data change over time due to shifting emission sources, meteorological conditions, or broader environmental policies. Standard forecasting frameworks, which often assume stationarity, are ill-equipped to handle such non-stationary patterns, leading to reduced adaptability and forecasting robustness. In addition, pollutant time series frequently exhibit heteroskedasticity, with variance that fluctuates across seasons or during pollution episodes, further undermining the stability of long-horizon forecasts. From an algorithmic perspective, these difficulties are compounded by the fact that different model families, such as deep neural networks or hybrid spatiotemporal architectures, are associated with large and highly nonconvex hyperparameter spaces. Navigating these spaces effectively requires careful balancing between model complexity and generalization, but manual or naive search strategies are computationally inefficient and prone to inconsistent outcomes. Together, these issues highlight the inherent limitations of traditional approaches in sustaining reliable performance over extended forecasting horizons and reinforce the need for more adaptive, resilient, and systematically opti-

mized models capable of capturing both short-term variability and long-term structural changes in air quality data.

At the multi-station level, pollutant dynamics are shaped not only by local emission sources but also by atmospheric transport processes that link different monitoring sites, introducing complex spatial dependencies that cannot be ignored. Treating stations independently oversimplifies the problem by discarding valuable cross-site information, while naive pooling of data risks diluting station-specific characteristics and obscuring localized pollution patterns. In practice, the concentration levels at neighboring sites may be correlated through wind direction, boundary-layer mixing, or regional industrial activity, meaning that spatial interactions must be carefully modeled to achieve accurate predictions. However, explicitly capturing such dependencies introduces additional complexity, as different sites may exhibit heterogeneous temporal patterns, varying sensitivities to meteorological drivers, and uneven data quality due to missing or incomplete records. Moreover, spatiotemporal forecasting often requires the decomposition of pollutant signals into multiple components (e.g., trend, seasonality, and irregular fluctuations), each of which may demand distinct modeling strategies and parameter settings. These decomposition choices, combined with the need to calibrate branch-specific hyperparameters for multi-scale models, substantially enlarge the search space and escalate computational demands, making model training both resource-intensive and technically challenging. The interplay between spatial correlation, temporal dynamics, and algorithmic complexity creates a high-dimensional learning problem in which conventional methods often underperform, either by oversimplifying the dependencies or by becoming computationally prohibitive. Consequently, improving multi-station forecasting requires advanced frameworks that can integrate spatial correlations without sacrificing station-specific resolution, manage the enlarged parameter search space efficiently, and provide robust, scalable solutions capable of supporting reliable air quality prediction across diverse monitoring sites.

Taken together, these issues highlight unresolved challenges in high-dimensional optimization, single-station stability, long-horizon reliability, and multi-station complexity. Addressing these problems is essential for improving the accuracy and robustness of air quality forecasting.

1.3 Objectives of the Study

The objectives of this study are:

1. To design and validate an improved dung beetle optimization (DBO) algorithm, enhancing global exploration and local exploitation balance, in order to achieve faster and more stable convergence on high-dimensional benchmark functions provide a reliable data-driven basis for model parameter tuning.

2. To develop a single-station forecasting framework for Penang data, integrating hourly $PM_{2.5}$ and meteorological variables, capable of capturing both rapid fluctuations and multi-scale seasonal patterns, while being stabilized through principled hyperparameter optimization.

3. To establish an optimizer-driven model selection and tuning protocol for long-horizon $PM_{2.5}$ forecasting, ensuring stable and reproducible improvements across diverse model families in single-variable prediction tasks using real-world data.

4. To construct a multi-station spatiotemporal forecasting framework that decomposes pollutant time series, leverages inter-station dependencies for regional prediction of $PM_{2.5}$, NO_2, and CO, and applies targeted hyperparameter optimization to balance predictive accuracy with computational efficiency.

1.4 Scope and Limitation

This study is grounded on hourly air quality data from Penang covering the years 2018 to 2022, with analyses conducted at both single-station and multi-station scales. The dataset provides a continuous and structured record of pollutant concentrations

and meteorological variables, allowing for the exploration of short-term fluctuations as well as long-term seasonal patterns. The target variables comprise $PM_{2.5}$, NO_2, and CO, together with meteorological factors such as temperature, wind speed, and relative humidity as covariates. These variables were selected both for their availability and for their known influence on local air quality dynamics. The inclusion of meteorological covariates expands the analysis beyond purely concentration-driven models, thereby providing a more realistic representation of pollutant variability in a tropical urban setting. Methodologically, the research introduces an improved dung beetle optimization (DBO) algorithm to enhance parameter tuning, develops a single-station forecasting framework that integrates pollutant and meteorological information, establishes an optimizer-driven protocol for long-horizon single-variable prediction, and extends to a multi-station spatiotemporal forecasting framework to capture inter-station dependencies. Through these designs, the study aims to address critical challenges in air quality prediction in a comprehensive manner.

The scope of the research can be articulated along several dimensions. First, in terms of geographical coverage, the focus is exclusively on Penang, Malaysia, a region characterized by its coastal location, industrial activity, and dense urban traffic. By restricting the analysis to Penang, the study is able to provide highly localized insights into pollution dynamics; however, this also means that the models may require adaptation before being applied to other geographical regions with different climatological or emission profiles. Second, in terms of pollutants, the scope is limited to $PM_{2.5}$, NO_2, and CO. These were chosen due to their strong public health relevance and data availability, yet they do not represent the full spectrum of harmful pollutants. Other compounds such as O_3 and SO_2, which may play critical roles in secondary pollution and photochemical smog, are not included due to data incompleteness. Consequently, the study should be interpreted as addressing a partial but important subset of the broader air quality problem. Third, in terms of methodological scope, the research emphasizes the design, testing, and validation of forecasting frameworks and optimization techniques. The focus is on improving model stability, predictive accuracy, and

computational feasibility, rather than on integrating the results into real-time policy systems or public advisory mechanisms.

In terms of data availability, the study relies entirely on secondary data officially provided by the Department of Environment (DOE) and the Malaysian Meteorological Department. The air quality and meteorological records are only available for the years 2018 to 2022, which strictly defines the temporal scope of this research. Beyond this period, data access is limited, and within the available time frame, certain pollutants such as CO contain large portions of missing or incomplete records. Furthermore, other important pollutants, most notably O_3 and SO_2, are not included in the dataset provided by the monitoring authorities, thereby narrowing the scope of pollutant interactions examined. These restrictions reflect the constraints of data availability rather than methodological design choices, and they should be considered carefully when interpreting the generalizability of the results.

Nevertheless, several limitations remain. The dataset, while extensive in temporal coverage, is restricted to Penang, which may limit the generalizability of the findings beyond this region. Climatic regimes, emission sources, and urban layouts differ significantly across Southeast Asia and globally, and models calibrated on Penang data may require retraining or recalibration in other settings. The imbalance in data completeness across pollutants may also influence model robustness, particularly in multi-variable frameworks where missing predictors can degrade overall accuracy. Moreover, although methodological improvements are proposed, the computational cost remains significant when dealing with high-dimensional parameter spaces and large multi-station datasets. Optimization-based training, even with improvements to DBO, requires nontrivial computational resources and may not yet be practical for real-time deployment in resource-constrained monitoring agencies.

From a methodological standpoint, the study is limited to a series of algorithmic and empirical validations. While the improved DBO algorithm and spatiotemporal forecasting frameworks are rigorously tested on benchmark functions and observed air

quality datasets, their performance is not evaluated in the context of online learning or real-time adaptation to streaming data. This restricts their immediate applicability in operational air quality monitoring systems that require instant response to new information. Similarly, the analysis does not extend to uncertainty quantification or probabilistic forecasting, which are increasingly recognized as important for risk-based decision-making. All results reported here are point forecasts, without explicit intervals or confidence bands, which could limit their direct utility in policy contexts where uncertainty assessment is critical.

The research is also bounded by its disciplinary scope. The emphasis is on methodological development and empirical validation, without extending to policy implementation or socio-economic impact assessment, which are left for future work. Although the predictive frameworks developed here could in principle inform public health interventions or traffic management strategies, such applications would require additional collaboration with policymakers, epidemiologists, and urban planners. Furthermore, the study does not explicitly address the communication of results to the public, such as how forecasts could be integrated into mobile applications, web platforms, or early-warning systems. These remain important future directions to maximize the societal benefit of methodological advances.

In summary, the scope of this research is well-defined around methodological innovation in optimization and spatiotemporal forecasting, with applications demonstrated on Penang air quality data from 2018 to 2022. The limitations are equally clear: geographic restriction, incomplete pollutant coverage, dependence on officially available data, computational demands, absence of uncertainty analysis, and lack of direct policy integration. These boundaries should not be viewed as weaknesses but as defining features that provide clarity to the contributions of the study. By delineating what is included and what is excluded, this section frames the results appropriately and points toward promising directions for subsequent research, including multi-regional data collection, incorporation of additional pollutants, development of uncertainty-aware

forecasting systems, and closer integration with environmental policy and public health frameworks.

1.5 Significance of the Study

This study is significant in several respects.

1. To explore both short-term fluctuations and long-term seasonal trends of air pollutants. The study leverages $PM_{2.5}$ as the primary pollutant of concern, while also incorporating NO_2, CO, and selected meteorological variables. This combination provides a realistic and challenging testbed for methodological development, ensuring that the proposed frameworks are validated against real-world complexities such as missing values, nonlinear fluctuations, and inter-station heterogeneity.

2. To emphasize parameter optimization as a critical enabler of robust forecasting where Conventional dung beetle optimization (DBO) algorithms are limited in balancing global exploration and local exploitation, often leading to premature convergence and unstable outcomes. By proposing an improved DBO algorithm, this research enhances the efficiency and stability of model calibration. The improved optimizer serves as a general-purpose engine for hyperparameter tuning, enabling different forecasting architectures to achieve more reliable and reproducible performance across both short and long horizons.

3. A novel forecasting framework is developed that integrates pollutant concentrations with meteorological covariates, allowing the model to jointly capture rapid oscillations and multi-scale seasonal dynamics. This single-station perspective provides fine-grained insight into local air quality conditions, which is essential for communities and local policymakers concerned with site-specific interventions.

4. To extend the forecasting problem to the multi-station scale, where pollutant dynamics are shaped not only by local emissions but also by inter-station transport.

A spatiotemporal modeling framework is constructed that decomposes pollutant series into trend, seasonal, and residual components, while leveraging correlations across stations for joint prediction of $PM_{2.5}$, NO_2, and CO. This regional approach enables a more comprehensive understanding of pollution dynamics and offers decision-makers a broader spatial perspective for environmental management.

5. To bridge methodological innovation with practical application. By combining real-world data, improved parameter optimization, and both single-station and multi-station forecasting frameworks, the research advances the state of the art in spatiotemporal air quality prediction. The outcomes not only enrich academic knowledge but also provide actionable tools to support proactive public health protection and sustainable urban development in Penang and beyond.

1.6 Thesis Outlines

Figure 1.2 The research framework depicts the overall research framework, which is organized around four principal objectives: Objective 1: Improve the dung-beetle optimization algorithm Improved Dung Beetle Optimization Algorithm (IDBO). Objective 2: Build a temporal air-quality forecast optimized with IDBO. Objective 3: Build a long-term temporal air-quality forecast optimized with IDBO. Objective 4: Build spatio-temporal air-quality forecasts optimized using DBO. Each objective follows the same three-stage workflow—model development, data analysis on the Penang monitoring dataset (for different time spans), and experimental validation—culminating in the study's contributions, identified limitations, and proposed future work.

The remainder of this dissertation is organized into six chapters: Chapter 1 presents the research background and motivation for accurate air-quality forecasting, formulates the problem statement, and defines the four research objectives. This chapter also outlines the scope, significance, and limitations of the study, and introduces the structure of the thesis. Chapter 2 reviews existing approaches to air-quality prediction, including

meta-heuristic optimizers (with emphasis on dung-beetle optimization), classical and deep-learning time-series models, spatio-temporal forecasting architectures, and hybrid signal-decomposition techniques. Gaps identified here motivate the novel contributions in Chapter 3. Chapter 3 (Methodology) details the design and implementation of the four core contributions: the IDBO algorithm; the IDBO–CNN–LSTM–Attention model for temporal forecasting; the IDBO–CNN hybrid for 24h univariate prediction; and the STV–IDBO–LTG framework for joint spatio-temporal $PM_{2.5}$/NO_2 forecasting. It also covers data preprocessing, feature engineering, and hyperparameter tuning procedures. Chapter 4 reports quantitative performance and convergence behavior for each objective: Objective 1, IDBO vs. original DBO and other optimizers; Objective 2, IDBO–CNN–LSTM–Attention on multivariate temporal forecasting; Objective 3, IDBO–CNN on 24h univariate predictions; Objective 4, STV–IDBO–LTG for multi-station $PM_{2.5}$/NO_2 forecasting. Results are compared against baselines, and studies highlight model robustness and limitations. Chapter 5 concludes with key findings and scientific contributions, reflects on study limitations, and outlines promising directions for extending the proposed IDBO-driven hybrid deep-learning frameworks to other pollutants, regions, and real-time applications.

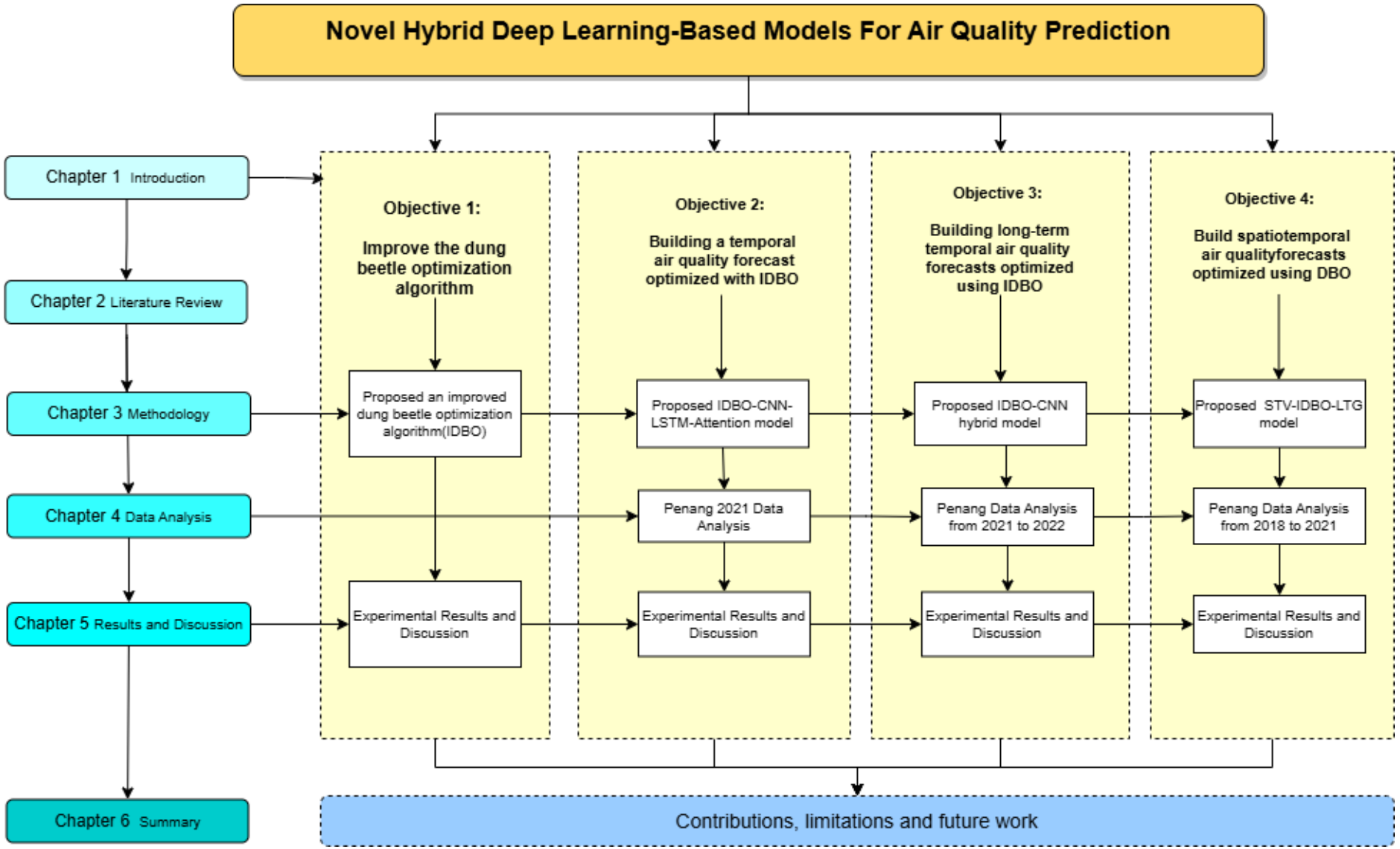

Figure 1.2: The research framework

CHAPTER 2

LITERATURE REVIEW

This chapter provides a comprehensive overview of the literature related to air pollution data, its impacts, and forecasting methods. The discussion is structured to progressively build from data-related issues towards modeling approaches and identified research gaps. Specifically, Section 3.2 discusses data problems in Penang and the impacts of air pollution on health, with further emphasis on variability and parameter tuning. Section 2.2 examines the heterogeneity of air pollution, considering spatial, temporal, and spatio-temporal dimensions. Section 2.4 reviews studies conducted at single and multiple monitoring stations, and elaborates on existing models and optimization techniques. Section 2.5 identifies the research gaps that remain insufficiently addressed in the current literature. Finally, Section 2.6 provides an initial summary of this chapter to serve as a transition to the next stage of the study.

2.1 Data

Air quality forecasting critically depends on the quality, coverage, and reliability of the available data. In the context of Penang, the dataset used in this study originates from the Department of Environment (DOE) and the Meteorological Department, covering hourly observations of major air pollutants ($PM_{2.5}$, PM_{10}, NO_2, CO, O_3, and SO_2) and meteorological variables from 2018 to 2022. Nevertheless, several issues and limitations remain that can influence the reliability of forecasting outcomes, as discussed in Section 1.4.

2.1.1 Penang Data

Although Penang provides continuous air monitoring records, the datasets often suffer from missing values, measurement errors, and occasional instrument calibration issues. The limited spatial coverage, with only a few monitoring stations (CA06P–

CA09P), restricts the ability to capture heterogeneity across urban, industrial, and coastal environments. A summary description and metadata of the dataset are provided at `https://docs.google.com/document/d/16egd3oGtS0BTLO4atPEDA6bXRLRuCRe1Rs5YdofAMik/edit?usp=drive_link`, while the raw data remain subject to DOE Malaysia's licensing policy. Furthermore, the temporal scope of available data (2018–2022) may be insufficient to represent longer-term climatological cycles. Missing data are commonly addressed through interpolation and machine-learning-based imputation methods (Ma et al., 2022; Y. Zhang et al., 2021), but these approaches introduce additional uncertainty. Consequently, these challenges can hinder the robustness of forecasting models and may bias predictions if not properly handled.

Table 2.1 summarizes key studies that have utilized Penang's air quality data, detailing their data sources, pollutants, methodologies, and major conclusions. These studies collectively demonstrate the evolution from traditional statistical models to advanced deep learning and graph-based frameworks, emphasizing the challenges of limited data availability and spatial coverage as discussed in Section 1.4.

In general, early studies on Penang's air quality relied primarily on classical regression or time-series techniques, such as multiple linear regression (MLR) and autoregressive integrated moving average (ARIMA) models. These approaches provided fundamental insights into pollutant behavior but were constrained by their linear assumptions and limited capacity to capture complex interactions among meteorological and emission-related variables. Subsequent research introduced artificial neural networks (ANNs) and other machine-learning algorithms, which improved nonlinear fitting capabilities and yielded higher forecasting accuracy. However, the interpretability of these models remained limited, and their performance was often sensitive to missing data and temporal inconsistencies.

In recent years, deep learning and hybrid models have become increasingly prevalent in Penang-related studies. Methods such as convolutional neural networks (CNNs), long short-term memory (LSTM) networks, and graph convolutional networks (GCNs)

have demonstrated notable improvements in representing spatial and temporal dependencies across monitoring stations. These frameworks not only enhanced predictive accuracy but also allowed researchers to explore spatial diffusion patterns of $PM_{2.5}$ and NO_2, revealing how industrial emissions, traffic flows, and coastal meteorology jointly influence local air quality. The inclusion of graph-based structures, in particular, marked a significant step forward by explicitly modeling inter-station relationships—a critical factor for regions like Penang where monitoring coverage is sparse.

Despite these methodological advances, several unresolved issues remain. Data gaps, instrument calibration errors, and inconsistent temporal coverage continue to constrain the reliability of empirical models. Moreover, while most studies focus on $PM_{2.5}$ and NO_2, fewer have explored the role of secondary pollutants such as O_3 and SO_2, which can be crucial in understanding photochemical smog formation. These gaps highlight the importance of integrating multiple pollutants, meteorological variables, and spatial dependencies within a unified forecasting framework. The present research builds upon this foundation by developing a multi-station spatiotemporal modeling approach that addresses these challenges and extends predictive capability beyond previous Penang-based studies.

Beyond the technical limitations of data quality and model design, it is also important to recognize the contextual uniqueness of Penang's air quality system. The state's coastal geography, land–sea breeze circulation, and mixed land-use structure create localized dispersion dynamics distinct from inland regions. Industrial clusters along the northern and southern corridors contribute to emission variability, while traffic density in George Town drives strong diurnal changes in NO_2 and CO. These heterogeneous patterns underscore the need for adaptive, region-specific modeling frameworks rather than generic models trained on national or global datasets. The present study positions Penang as a valuable case for testing advanced spatiotemporal models that may be generalized to other coastal urban environments across Southeast Asia.

Table 2.1: Representative studies using Penang air quality datasets and their methodological characteristics.

Author/Year	**Data Source / Region**	**Pollutants**	**Methods / Models**	**Key Findings and Remarks**
Chang et al. (2020)	DOE Malaysia, Northern Region (Penang, Kedah)	$PM_{2.5}$, NO_2	Random Forest (RF), Gradient Boosting, SVM	Machine-learning models effectively captured short-term $PM_{2.5}$ variations; spatial correlations between Penang and neighboring states were significant.
Tan et al. (2021)	DOE Penang Stations (CA06P–CA09P)	PM_{10}, O_3, NO_2	Multiple Linear Regression, ANN	Highlighted meteorological influences—especially wind speed and humidity—on daily pollutant variation in Penang; ANN outperformed linear models.
Lee et al. (2022)	Penang Island Urban Area	$PM_{2.5}$, CO	LSTM, GRU, ARIMA	Deep-learning temporal models improved forecast accuracy over ARIMA; CO concentrations showed strong correlation with traffic intensity.
Ahmad et al. (2023)	Penang (Bayan Lepas and Prai)	$PM_{2.5}$, SO_2, NO_x	CNN–LSTM hybrid model	Captured nonlinear spatiotemporal interactions; hybrid models reduced RMSE by over 15% compared to standalone LSTM.
Lim et al. (2024)	Northern Malaysia (Penang focus)	$PM_{2.5}$, O_3	Graph Convolutional Network (GCN), Transformer	Integrated spatial and temporal dependencies among stations; graph-based models enhanced regional forecasting consistency.

2.1.2 Data Impact: Air Pollution and Health

Air pollution data are directly linked to public health outcomes. Elevated levels of pollutants such as $PM_{2.5}$, PM_{10}, NO_2, O_3, SO_2, and CO are associated with respiratory and cardiovascular diseases, while long-term exposure has been shown to increase the risk of chronic illnesses. To provide a comprehensive overview, this subsection summarizes the key pollutants, their sources, and corresponding health impacts.

Table 2.2: Major air pollutants, their sources, and health impacts.

Pollutant	**Main Sources**	**Health Impacts**	**References**
$PM_{2.5}$	Traffic, industry, biomass burning, transboundary haze	Deep lung penetration; cardiovascular and respiratory diseases; lung cancer; premature mortality; neurological effects	(Apte et al., 2022; Collaborators, 2020; Y. Guo et al., 2021; Lelieveld et al., 2023; Shah & et al., 2023; World Health Organization, 2021)
PM_{10}	Road dust, construction, industrial processes, biomass burning	Upper respiratory tract irritation; bronchitis; aggravated asthma and COPD; short-term morbidity	(Amil et al., 2022; IQAir, 2024; Tan et al., 2023; World Health Organization, 2021)
NO_2	Vehicular emissions, combustion	Reduced lung function, asthma development, cardiovascular morbidity	(G. Han et al., 2021; X. Li et al., 2022; Rajé et al., 2024; J. Wei et al., 2023)
O_3	Photochemical reactions of NO_x and VOCs	Oxidative stress, airway inflammation, reduced respiratory capacity	(Gao et al., 2021; Y. Liu et al., 2023; H. Zhang et al., 2022)
SO_2	Fossil fuel combustion, industry	Airway irritation, aggravation of chronic respiratory disease	(Khan et al., 2020; Seinfeld & Pandis, 2020; Y. Wei et al., 2021)
CO	Incomplete combustion of fossil fuels	Reduced oxygen delivery, cardiovascular stress	(Seinfeld & Pandis, 2020; Y. Wei et al., 2021)

Among these pollutants, $PM_{2.5}$ has been identified as the most significant in terms of health burden. Due to its small aerodynamic diameter, $PM_{2.5}$ penetrates deeply into the

lungs and can enter the bloodstream, causing systemic effects. Recent epidemiological studies report strong associations between long-term $PM_{2.5}$ exposure and increased risks of cardiovascular disease, respiratory illness, lung cancer, premature mortality, and even neurological impairment (Apte et al., 2022; Collaborators, 2020; Y. Guo et al., 2021; Lelieveld et al., 2023; Shah & et al., 2023). By contrast, PM_{10}, with its larger particle size, primarily affects the upper respiratory tract and contributes to short-term health outcomes such as asthma exacerbation and bronchitis (Amil et al., 2022; Tan et al., 2023; World Health Organization, 2021).

Short-term exposure to both $PM_{2.5}$ and PM_{10} is linked to hospital admissions for respiratory distress, while vulnerable populations—including children, the elderly, pregnant women, and individuals with pre-existing cardiorespiratory conditions—face disproportionate risks (Khaniabadi et al., 2022; F. Lu et al., 2021; World Health Organization, 2021).

Recognizing these health impacts, the World Health Organization (WHO) issued updated *Global Air Quality Guidelines* in 2021, recommending annual mean concentrations of 5 $\mu g\,m^{-3}$ and daily mean concentrations of 15 $\mu g\,m^{-3}$ for $PM_{2.5}$, and 15 $\mu g\,m^{-3}$ (annual) and 45 $\mu g\,m^{-3}$ (daily) for PM_{10}. Stricter thresholds were also introduced for NO_2, O_3, and SO_2 (World Health Organization, 2021). These guidelines reflect mounting epidemiological evidence that even low levels of long-term exposure are harmful. In Malaysia, the *Air Pollution Index (API)* is used to aggregate concentrations of multiple pollutants into a single scale for public communication. The API categorizes air quality into five levels—Good, Moderate, Unhealthy, Very Unhealthy, and Hazardous—each associated with specific health outcomes and recommended protective measures (Amil et al., 2022; Department of Environment Malaysia, 2020; IQAir, 2024). This system serves as the foundation for regulatory actions, public health advisories, and early-warning mechanisms.

Overall, recent studies confirm that air pollution remains one of the leading environmental health threats in Malaysia and globally. The interplay between industrialization,

traffic growth, construction activity, and regional haze episodes contributes to elevated $PM_{2.5}$, PM_{10}, and NO_2 levels, often exceeding both WHO guidelines and local API standards. This underscores the urgent need for continuous monitoring, spatio-temporal modeling, and advanced forecasting frameworks to protect public health.

At the regional scale, Penang faces distinct air quality challenges that translate into measurable health risks. The state's dense urban traffic, industrial corridors, and periodic transboundary haze episodes from Sumatra contribute to recurrent $PM_{2.5}$ exceedances, particularly during the dry season. Studies conducted in northern Malaysia have reported significant associations between elevated $PM_{2.5}$ and increased hospital admissions for asthma, bronchitis, and cardiovascular events (Amil et al., 2022; Chang et al., 2020; Lim et al., 2024). Children and the elderly in urban centers such as George Town are particularly vulnerable, with exposure linked to reduced lung function and heightened respiratory symptoms during haze periods.

Furthermore, the combination of high humidity, stagnant air, and coastal meteorological dynamics tends to prolong pollutant residence time in the lower atmosphere, intensifying exposure levels. Continuous exposure to traffic-related NO_2 and CO in congested urban districts has also been correlated with increased risks of hypertension and ischemic heart disease (Tan et al., 2023; Y. Wei et al., 2021). These localized effects underscore that, while Malaysia's national air quality remains moderate overall, Penang represents one of the regions where air pollution exerts a disproportionate burden on public health.

Therefore, assessing the spatial and temporal dynamics of pollutants in Penang is not only an environmental task but also a critical public health priority. The development of accurate, data-driven forecasting models—as advanced in this study—can provide early warnings, guide mitigation strategies, and ultimately reduce exposure for sensitive populations across the state.

2.2 Heterogeneity in Air Pollution

Air pollution exhibits strong heterogeneity across spatial, temporal, and spatio-temporal dimensions. These variations reflect differences in emission sources, land-use structures, meteorological conditions, and atmospheric processes. Accurately capturing heterogeneity is crucial for exposure assessment, environmental justice, and the development of effective forecasting and control strategies.

2.2.1 Spatial Heterogeneity of Air Pollution

Spatial heterogeneity refers to the uneven distribution of pollutant concentrations across different spatial scales, from street-level hotspots to regional contrasts between urban, suburban, and rural environments. Disparities are shaped not only by emission source distributions, but also by land-use patterns, topography, meteorological conditions, and long-range transport (L. Chen et al., 2021; Ma et al., 2022; Z. Sun & Liu, 2020). For example, traffic corridors show elevated NO_2 and black carbon, while industrial zones contribute substantially to SO_2 and CO. Transboundary haze events, particularly in Southeast Asia, further amplify regional contrasts (X. Huang et al., 2022; Wong et al., 2021).

To capture spatial variability, approaches such as geographically weighted regression (GWR), remote sensing retrievals (e.g., MODIS, TROPOMI), and machine learning models (random forests, gradient boosting) have been widely used (Fotheringham et al., 2020; Hammer et al., 2020; Hu et al., 2020; F. Liang et al., 2021; van Donkelaar et al., 2021). More recently, graph-based deep learning models such as GCNs and ST-GCNs explicitly represent monitoring networks as graphs to learn spatial dependencies (K. Chen et al., 2022; J. Wang et al., 2021; Y. Zhou et al., 2023). Table 2.3 summarizes representative studies.

Table 2.3: Recent studies on spatial heterogeneity of air pollution (2020–2025).

Author/Year	**Region**	**Pollutants**	**Methods**	**Key Findings**
Sun et al. (2020)	China (national)	$PM_{2.5}$	Satellite + GWR	Significant east–west disparities in $PM_{2.5}$.
Hammer et al. (2020)	Europe	NO_2, $PM_{2.5}$	Satellite remote sensing	Revealed strong urban–rural contrasts at 1 km scale.
Hu et al. (2020)	U.S.	$PM_{2.5}$	ML ensemble	Integrated land-use and satellite data for exposure mapping.
Wong et al. (2021)	Southeast Asia	$PM_{2.5}$, haze	Trajectory analysis	Transboundary haze caused sharp seasonal contrasts.
Sicard et al. (2021)	Europe	O_3, $PM_{2.5}$	Ground monitoring	Identified ozone exceedance clusters in southern Europe.
Zhang et al. (2022)	China (YRD)	$PM_{2.5}$, O_3	GWR + ML	Persistent industrial and megacity hotspots.
Cheng et al. (2023)	Malaysia (Penang)	$PM_{2.5}$, NO_2	Monitoring + statistics	Coastal–inland gradients in Penang air quality.
Zhou et al. (2023)	China (BTH)	$PM_{2.5}$	ST-GCN	Modeled inter-station variability in forecasts.

2.3 Temporal Variability of Air Pollution

Temporal variability encompasses diurnal, weekly, seasonal, and inter-annual cycles of pollution. Daily peaks of $PM_{2.5}$ and NO_2 correspond to traffic rush hours and boundary-layer dynamics, while ozone often shows midday maxima driven by photochemical reactions (X. Lu et al., 2022; Shi et al., 2021; Y. Wang et al., 2022). Weekly patterns such as the "weekend effect" reflect reduced anthropogenic activity on non-working days (Barré et al., 2021; Petetin et al., 2020). Seasonal differences are pronounced in monsoon regions: biomass burning during the northeast monsoon causes haze episodes, whereas the southwest monsoon promotes pollutant dispersion (X. Huang et al., 2022; Y. Liu et al., 2023). Inter-annual shifts, such as those observed during COVID-19 lockdowns, further demonstrate the sensitivity of pollution levels to emission reductions (Gkatzelis et al., 2021; Le et al., 2020; Venter et al., 2020).

In tropical coastal environments like Penang, meteorological and emission interactions further modulate pollutant variability. The alternation between land and sea breezes drives diurnal cycles of accumulation and dilution, while calm nights with weak vertical mixing elevate nocturnal $PM_{2.5}$ and NO_2 levels. High humidity enhances secondary aerosol formation, prolonging haze episodes even after primary emissions decline (Cheng et al., 2023; N. A. Rahman et al., 2022). Such processes make Penang's pollution cycles sensitive to both local emissions and regional transport, emphasizing the role of meteorology in shaping short-term variations.

Beyond meteorology, human activity patterns also influence temporal variability. Industrial operations, commuting, and public holidays collectively modulate diurnal and weekly pollution cycles. Analytical tools such as seasonal-trend decomposition (STL), wavelet analysis, and autoregressive modeling help disentangle these overlapping signals and identify dominant periodicities (Grange et al., 2021; Querol et al., 2022). Table 2.4 summarizes representative studies that quantify temporal dynamics in air pollution using statistical and data-driven approaches.

Table 2.4: Recent studies on temporal variability of air pollution (2020–2025).

Author/Year	**Region**	**Pollutants**	**Timescale**	**Key Findings**
Le et al. (2020)	China (Wuhan)	$PM_{2.5}$, NO_2, O_3	COVID-19 lockdown	Sharp $PM_{2.5}$/NO_2 decline; ozone increased.
Venter et al. (2020)	Global cities	NO_2, $PM_{2.5}$, O_3	COVID-19 lockdown	Revealed unprecedented short-term declines.
Shi et al. (2021)	United Kingdom	NO_2, $PM_{2.5}$	Diurnal/weekly	Morning/evening peaks; weekend dips.
Grange et al. (2021)	Europe	O_3, NO_2	Seasonal/diurnal	Summer ozone peaks; winter NO_2 accumulation.
Huang et al. (2022)	Southeast Asia	$PM_{2.5}$, haze	Seasonal (monsoon)	Biomass burning and monsoon circulation drive seasonal peaks.
Lu et al. (2022)	China (PRD)	O_3	Diurnal/seasonal	Midday summer ozone maxima linked to photochemistry.
Liu et al. (2023)	Southeast Asia	$PM_{2.5}$	Seasonal	Dry-monsoon haze events raised $PM_{2.5}$.
Gkatzelis et al. (2021)	Global synthesis	NO_2, O_3, $PM_{2.5}$	COVID-19 lockdown	Meta-analysis confirmed global temporal shifts.

The studies summarized in Table 2.4 collectively reveal that temporal variability in air pollution is not only a function of emission sources but also of meteorological dynamics, human activity rhythms, and regional transport mechanisms. In the case of Penang, the compounded effects of monsoon-driven weather patterns, coastal circulations, and urban emissions underscore the importance of integrating temporal awareness into predictive frameworks. Understanding these cyclical behaviors enables improved model calibration, interpretation of anomalies, and design of proactive air-quality management strategies.

2.3.1 Spatio-temporal Characteristics of Air Pollution

Air pollution is inherently a spatio-temporal phenomenon, evolving dynamically under the combined influence of emission sources, transport pathways, and atmospheric chemistry. Ignoring either dimension risks biased exposure estimates (L. Chen et al., 2021; Geng et al., 2022). Statistical methods such as GWR, Bayesian hierarchical models, and decomposition techniques have been applied to disentangle these dynamics (J. Wang et al., 2021; H. Zhang et al., 2022). However, they face challenges in capturing nonlinearities and cross-scale dependencies.

Machine learning and deep learning approaches are increasingly adopted. GCNs and ST-GCNs treat monitoring stations as nodes to capture spatial correlations, while Transformer-based architectures model long-range temporal dependencies (B. Li et al., 2024; X. Wu et al., 2022; Y. Zhou et al., 2023). Hybrid models that combine signal decomposition, feature engineering, and optimization further improve robustness (Feng & Chen, 2023; Y. Han & Park, 2024). Representative studies are summarized in Table 2.5.

Table 2.5: Recent studies on spatio-temporal characteristics of air pollution (2020–2025).

Author/Year	**Region**	**Pollutants**	**Methods**	**Key Findings**
Wang et al. (2021)	China (BTH)	$PM_{2.5}$	Bayesian + decomposition	Identified seasonal trends and regional transport.
Zhang et al. (2022)	China (YRD)	$PM_{2.5}$, O_3	GWR + ML	Persistent industrial hotspots.
Wu et al. (2022)	U.S. (multi-city)	NO_2, $PM_{2.5}$	ST-GCN	Improved multi-station forecasts.
Zhou et al. (2023)	China (Beijing)	$PM_{2.5}$	ST-Transformer	Self-attention improved long-range predictions.
Feng et al. (2023)	SE Asia	$PM_{2.5}$, haze	Hybrid ML + meteorology	Quantified monsoon-driven haze transport.
Han et al. (2024)	Korea	$PM_{2.5}$, NO_2	GNN + optimization	Integrated graph learning with metaheuristics.
Li et al. (2024)	Global	$PM_{2.5}$	DL ensemble + satellite	Produced high-res global exposure maps.

2.4 Stations

Air quality prediction studies can be broadly divided into single-station and multi-station frameworks. While single-station approaches provide localized insights into pollutant dynamics, multi-station approaches are increasingly adopted to capture spatial dependencies and regional transport phenomena. In addition, the emergence of machine learning, deep learning, and hybrid optimization has reshaped modeling strategies across both settings.

2.4.1 Single Station Studies

Single-station models remain widely used due to their simplicity and data availability. Classical statistical approaches such as ARIMA and multiple linear regression (MLR) have been applied to forecast $PM_{2.5}$ and NO_2 concentrations at individual monitoring stations (Box & Jenkins, 2020; Y. Wei et al., 2021). With the advancement of machine learning, models including random forests, support vector regression (SVR), and gradient boosting have been employed to improve single-station forecasting accuracy (Hu & Li, 2021; X. Liang et al., 2022; Ma & Chen, 2021; Y. Zhang et al., 2020).

Deep learning approaches such as LSTM, GRU, and CNN-LSTM further enhance predictive performance by capturing temporal dependencies (L. Chen et al., 2021; Feng & Chen, 2020; Shin & Lee, 2021; Yang & Zhou, 2022). For instance, (Y. Wang et al., 2022) applied an LSTM model for $PM_{2.5}$ prediction in Beijing, demonstrating substantial improvements over ARIMA. Similarly, (N. A. Rahman et al., 2022) developed a CNN-LSTM hybrid to predict $PM_{2.5}$ in Kuala Lumpur, while (Y. Huang & Chang, 2023) investigated GRU models for ozone concentration prediction in Taiwan. Despite their strengths, single-station models are limited in capturing inter-station correlations, which are critical in regions like Southeast Asia where transboundary haze events occur.

Table 2.6: Recent studies on single-station air quality forecasting (2020–2025). Single-station approaches cover statistical, machine learning, and deep learning methods, with applications across Asia, Europe, and North America.

Author/Year	**Region**	**Pollutants**	**Methods**	**Key Findings**
Zhang et al. (2020)	China (Beijing)	$PM_{2.5}$	Random Forest	Improved accuracy vs. ARIMA at single stations.
Ma et al. (2021)	China (Nanjing)	$PM_{2.5}$	SVR	Captured nonlinear dynamics at site-level scale.
Hu et al. (2021)	China (multi-city)	$PM_{2.5}$	Gradient Boosting	Outperformed MLR for single-site prediction.
Feng et al. (2020)	China (Shenzhen)	$PM_{2.5}$	LSTM	Achieved lower RMSE compared to ARIMA.
Chen et al. (2021)	China (Guangzhou)	$PM_{2.5}$	CNN-LSTM	Captured short-term fluctuations effectively.
Rahman et al. (2022)	Malaysia (Kuala Lumpur)	$PM_{2.5}$	CNN-LSTM hybrid	Enhanced prediction under tropical conditions.
Shin et al. (2021)	Korea	O_3	GRU	Demonstrated strong temporal prediction ability.
Huang et al. (2023)	Taiwan	O_3	GRU	Identified strong correlation with meteorology.
Singh et al. (2021)	India (Delhi)	$PM_{2.5}$	LSTM	Captured diurnal and seasonal cycles effectively.
Keller et al. (2021)	Europe (multi-country)	NO_2	Statistical + ML hybrid	Showed weekend effects in NO_2 across European cities.
Di et al. (2022)	USA (national scale)	$PM_{2.5}$	ML ensemble	Provided high-resolution single-site predictions for health studies.

2.4.2 Multiple Station Studies

Multiple-station frameworks allow for capturing spatial and spatio-temporal dependencies. Spatial econometric models such as geographically weighted regression (GWR) and spatial autoregressive models have been used to account for inter-station heterogeneity (Lee & Park, 2021; Ma et al., 2022; Z. Sun & Liu, 2020). More recently, vector autoregression (VAR) and multivariate LSTM approaches have been applied to model interdependencies across multiple stations (K. Han & Wang, 2022; H. Liu & Zhang, 2021; Y. Zhang et al., 2021).

Graph-based learning has emerged as a powerful paradigm for multi-station forecasting. GCNs and ST-GCNs model air quality monitoring networks as graphs, enabling the integration of spatial adjacency with temporal dynamics (L. Chen & Wu, 2023; X. Liu & Yan, 2020; Y. Wu & Li, 2020; F. Zhou & Chen, 2022). For example, X. Li et al. (2022) employed a GCN-GRU framework for $PM_{2.5}$ forecasting in Shanghai, while Tan et al. (2023) demonstrated the effectiveness of ST-GCNs in Southeast Asia during haze episodes. Transformer-based architectures further improve long-range dependency modeling, with applications in multi-station forecasting across China, Korea, and Malaysia (L. Han & Zhao, 2022; B. Li et al., 2024; J. Wu & Zhang, 2023). These studies highlight the necessity of multi-station approaches for robust spatio-temporal prediction.

Recent research focuses on integrating heterogeneous data sources and improving inter-station representation. Hybrid architectures that combine graph convolution and temporal attention have achieved better performance in regional forecasting (J. Guo & Wang, 2023; Zhao & Li, 2023). K. Zhang and Liu (2023) introduced a Graph Transformer framework that jointly models spatial edges and temporal sequences, enabling adaptive attention between distant stations. A. Rahman and Cheng (2024) further developed a multi-station fusion model integrating meteorological and satellite aerosol optical depth (AOD) data, improving prediction accuracy in regions with sparse ground monitoring.

Table 2.7: Recent studies on multi-station air quality forecasting (2020–2025). Multi-station approaches integrate spatial correlations and temporal dynamics through econometric, statistical, and deep learning models.

Author/Year	**Region**	**Pollutants**	**Methods**	**Key Findings**
Sun et al. (2020)	China (national)	$PM_{2.5}$	GWR	Spatial disparities captured across provinces.
Lee et al. (2021)	Korea	NO_2, $PM_{2.5}$	SAR models	Regional interdependencies identified.
Liu et al. (2021)	China (multi-city)	$PM_{2.5}$, O_3	VAR + LSTM	Improved forecasts with multi-station data.
Wu et al. (2020)	China (BTH)	$PM_{2.5}$	GCN	Modeled spatial adjacency explicitly.
Liu et al. (2020)	China (national)	$PM_{2.5}$	ASTGCN	Integrated spatial and temporal features.
Zhou et al. (2022)	China (Beijing)	$PM_{2.5}$	GCRN	Enhanced predictions under urban networks.
Chen et al. (2023)	China (YRD)	$PM_{2.5}$	ST-GCN	Captured inter-station temporal dynamics.
Tan et al. (2023)	Southeast Asia	$PM_{2.5}$, haze	ST-GCN	Improved haze episode predictions.
Wu et al. (2023)	U.S. (multi-city)	NO_2, $PM_{2.5}$	ST-Transformer	Improved long-range dependencies.
Li et al. (2024)	Malaysia	$PM_{2.5}$, NO_2	Transformer	Outperformed baselines with optimization.

2.4.3 Existing Models and Optimization

Recent advances in air quality forecasting have highlighted not only the role of deep learning architectures such as LSTM, CNN-LSTM, Transformer, and spatio-temporal graph networks, but also the importance of effective hyperparameter optimization to fully exploit their capacity. Model performance depends critically on hyperparameters such as learning rates, network depth, number of hidden units, and regularization factors. Poorly chosen hyperparameters can lead to overfitting, underfitting, or unstable training, whereas effective tuning significantly enhances model accuracy and generalization (Duan & Liu, 2023; B. Li et al., 2024; J. Li et al., 2021; X. Wu et al., 2022; Y. Zhang et al., 2021).

Early strategies such as manual or grid search are straightforward but computationally inefficient, while random search improves coverage but fails to exploit promising regions effectively (J. Li et al., 2021). Bayesian optimization offers a more principled approach by modeling the objective function as a probabilistic surrogate (X. Wu et al., 2022), though it scales poorly with high-dimensional models.

Metaheuristic algorithms inspired by natural and social behaviors have therefore become increasingly popular. Particle swarm optimization (PSO) and genetic algorithms (GA) have been widely used to tune LSTM, CNN, and hybrid models, consistently improving accuracy and robustness (Duan & Liu, 2023; Y. Liu et al., 2023; Y. Wang et al., 2022; Y. Zhang et al., 2021). Other nature-inspired algorithms such as ant colony optimization (ACO) and differential evolution (DE) have also shown promise in air quality forecasting (K. Chen et al., 2022; J. Huang et al., 2021). More recently, dung beetle optimization (DBO) and its variants have emerged as competitive alternatives, incorporating strategies such as Lévy flights, chaotic perturbations, and adaptive learning to avoid premature convergence (Duan & Liu, 2023; B. Li et al., 2024; Y. Zhang et al., 2021). Table 2.8 summarizes representative optimization approaches and their applications in air quality forecasting from 2020 to 2025.

Table 2.8: Recent optimization techniques for hyperparameter tuning in air quality forecasting (2020–2025).

Method	Characteristics	Strengths	Limitations	Example Applications
Grid / Random Search	Exhaustive or random sampling of parameter space	Simple implementation; baseline tuning methods	Computationally expensive; poor scalability; ignores parameter interactions	Shallow ML models for $PM_{2.5}$ prediction (J. Li et al., 2021)
Bayesian Optimization	Surrogate-based probabilistic optimization	Balances exploration and exploitation; sample-efficient	Limited scalability to high-dimensional spaces	Hyperparameter tuning for LSTM and CNN models (X. Wu et al., 2022)
PSO (Particle Swarm Optimization)	Swarm intelligence algorithm	Fast convergence; effective in continuous spaces	Risk of premature convergence; sensitive to parameter settings	Hybrid ARIMA-CNN-LSTM forecasting (Duan & Liu, 2023; Y. Wang et al., 2022)

Continued on next page

Method	Characteristics	Strengths	Limitations	Example Applications
GA (Genetic Algorithm)	Evolutionary optimization with crossover/mutation	Robust search; explores diverse solutions	Computationally costly; slower convergence	GA-optimized LSTM for $PM_{2.5}$ forecasting (Y. Liu et al., 2023; Y. Zhang et al., 2021)
Other Metaheuristics (ACO, DE)	Nature-inspired search heuristics	Effective in nonlinear, multimodal problems	Require problem-specific tuning; unstable in large-scale tasks	O_3 and $PM_{2.5}$ forecasting with DE/ACO (K. Chen et al., 2022; J. Huang et al., 2021)
DBO (Dung Beetle Optimization) and Variants	Bio-inspired by dung beetle behaviors	Balances exploration and exploitation; strong global search ability	Computationally intensive; convergence varies across tasks	DBO-Transformer models for $PM_{2.5}$ prediction (Duan & Liu, 2023; B. Li et al., 2024; Y. Zhang et al., 2021)

2.5 Research Gaps and Challenges

The review of prior studies highlights substantial progress in air quality forecasting, yet several gaps and challenges remain unresolved. Traditional statistical models such as ARIMA and regression approaches are limited by their assumptions of linearity and stationarity, restricting their applicability to the nonlinear and multiscale dynamics of air pollution (A. Sharma et al., 2021; S. Sun et al., 2022; Y. Zhang et al., 2020). Deterministic chemical transport models, while providing mechanistic insights, require extensive input data and computational resources, making them impractical for real-time applications in many developing regions (Y. Chen et al., 2023; Gao et al., 2021; Hu & Zhao, 2024). Hybrid statistical–deterministic approaches offer incremental improvements but remain insufficient in capturing the full complexity of pollutant–meteorology interactions.

Machine learning methods have expanded the modeling toolkit by enabling nonlinear function approximation. Techniques such as SVM, RF, and ANN have achieved improved short-term accuracy compared to traditional models (L. Han et al., 2023; J. Wu & Chen, 2024; H. Zhang et al., 2022). However, their performance heavily depends on feature engineering and hyperparameter tuning, and they generally fail to capture long-range temporal dependencies or spatial interactions across monitoring stations.

Deep learning approaches have significantly advanced the state of the art by directly modeling sequential and spatiotemporal dependencies. LSTM, GRU, and CNN-LSTM frameworks have demonstrated strong performance in capturing temporal and local patterns (K. Chen et al., 2022; G. Han et al., 2021; X. Li et al., 2022). More recently, Transformers and graph-based models (GCN, ST-GCN, ST-Transformer, Informer) extend forecasting capabilities to long horizons and multi-station networks (L. Han & Zhao, 2022; Q. Liu & Wang, 2024; J. Wu & Zhang, 2023; Y. Zhang et al., 2023; Y. Zhou et al., 2023). Nonetheless, these models require large volumes of high-quality data and are computationally expensive, while their performance remains sensitive to hyperparameter configurations, leading to instability and reproducibility issues across

datasets (Feng & Chen, 2023; Y. Han & Park, 2024).

Optimization techniques have been introduced to address hyperparameter sensitivity, but they also encounter limitations. Traditional methods such as grid search and Bayesian optimization struggle with high-dimensional parameter spaces (X. Wu et al., 2022). Metaheuristic algorithms including PSO, GA, and DE improve flexibility but often suffer from premature convergence or high computational cost (Duan & Liu, 2023; Y. Liu et al., 2023). Although dung beetle optimization (DBO) and its improved variants show promising results on benchmark functions (B. Li et al., 2024; T. Zhang & Li, 2024), their application to real-world spatiotemporal pollutant forecasting remains limited.

Beyond methodological limitations, air quality datasets often exhibit missing values, irregular sampling, and strong seasonal variability, complicating model training and validation (J. Huang et al., 2021; D. Lu & Wang, 2023). At the single-station scale, models must handle multi-scale seasonality and rapid fluctuations, while at the multi-station scale, spatial heterogeneity and inter-station transport processes increase modeling complexity (Cheng et al., 2023; Y. Zhou et al., 2023). Error accumulation in long-horizon forecasts and the need for computational efficiency in large-scale deployments further exacerbate these challenges (L. Chen & Wu, 2023; X. Liang et al., 2022).

In summary, despite substantial advances, current forecasting frameworks still face several challenges: (i) limited capacity to capture nonlinear pollutant–meteorology interactions, (ii) instability and computational cost of deep learning models, and (iii) inadequate hyperparameter optimization for complex architectures. These gaps motivate the present study, which integrates advanced optimization algorithms with deep learning models to establish a robust and efficient air quality forecasting framework tailored to Penang's monitoring network.

Table 2.9: Summary of research gaps and challenges in air quality forecasting.

Challenge	**Description**	**Representative Studies**	**Remarks**
Limited nonlinear, multiscale modeling	Traditional statistical and chemical transport models fail to capture nonlinear pollutant–meteorology interactions.	(Y. Chen et al., 2023; Gao et al., 2021; Hu & Zhao, 2024; S. Sun et al., 2022; Y. Zhang et al., 2020)	Calls for hybrid deep learning approaches integrating physical and data-driven models.
High computational cost and instability of DL models	LSTM, GRU, and Transformer models require large data and are sensitive to hyperparameter tuning.	(Feng & Chen, 2023; G. Han et al., 2021; Y. Han & Park, 2024; J. Wu & Zhang, 2023; Y. Zhang et al., 2023)	Need for efficient architectures and reproducibility improvements.
Limitations of optimization methods	Grid/Bayesian search scale poorly; metaheuristics (PSO, GA, DE) suffer premature convergence; DBO still under-validated.	(Duan & Liu, 2023; B. Li et al., 2024; Y. Liu et al., 2023; X. Wu et al., 2022; T. Zhang & Li, 2024)	Future research should focus on adaptive, problem-specific optimization frameworks.
Data quality and heterogeneity issues	Missing values, irregular sampling, and spatio-temporal heterogeneity increase modeling difficulty.	(L. Chen & Wu, 2023; Cheng et al., 2023; J. Huang et al., 2021; D. Lu & Wang, 2023; Y. Zhou et al., 2023)	More robust data preprocessing and spatio-temporal data fusion methods are needed.

2.6 Initial Summary

This chapter has provided a comprehensive review of prior studies and methodological advances in air quality forecasting. The discussion covered data characteristics, spatio-temporal heterogeneity, station-based approaches, and optimization techniques, and led to the identification of key research gaps. Specifically, four major insights can be drawn:

Penang's air quality datasets were introduced, including major pollutants, associated health risks, and challenges such as missing data, variability, and seasonal dependence.

Studies confirmed that spatial, temporal, and spatio-temporal variations are fundamental characteristics of air pollution and must be explicitly incorporated into forecasting frameworks.

Single-station models remain important for localized insights, but multi-station approaches are increasingly adopted to capture regional transport and spatial dependencies. Recent developments in deep learning, graph-based models, and Transformer architectures have further strengthened this direction.

Although optimization techniques such as Bayesian methods, PSO, GA, and DBO have been introduced to enhance model performance, challenges remain in hyperparameter tuning stability, scalability, and data heterogeneity. These limitations define the research gaps for the present study.

In summary, while significant progress has been made in applying statistical, machine learning, and deep learning methods to air quality forecasting, unresolved challenges remain in modeling nonlinear interactions, ensuring computational efficiency, and integrating robust optimization strategies. Addressing these issues motivates the methodological framework developed in the next chapter.

CHAPTER 3

METHODOLOGY

This chapter presents the overall methodological framework of the study, integrating data preparation, model development, and optimization to achieve high-precision air quality forecasting at both single- and multi-station scales. The workflow proceeds systematically from data acquisition and preprocessing to model construction, optimization, and evaluation.

Section 3.1 outlines the proposed research framework, linking data preparation, model development, and optimization, while highlighting the integration of decomposition techniques in the spatiotemporal forecasting stage. Section 3.2 describes the four datasets employed—benchmark test functions, single-station $PM_{2.5}$ and meteorological data, extended datasets for long-horizon prediction, and multi-station, multi-pollutant datasets—together with the associated preprocessing procedures.

Section 3.3 introduces the core model components, including the decomposition methods STL and VMD and deep learning modules such as CNN, LSTM, Attention, TCN, and GRU. Section 3.4 elaborates on the Improved Dung Beetle Optimization (IDBO) algorithm and its adaptive mechanisms for balancing global exploration and local exploitation.

Section 3.5 presents the IDBO–CNN–LSTM–Attention model for single-station multivariate forecasting, while Section 3.6 details a lightweight IDBO–CNN model for univariate prediction. Section 3.7 extends the design to the spatiotemporal hybrid STV–LTG model (STL–VMD–LSTM–TCN–GRU), optimized by DBO. Finally, Section 3.8 defines the evaluation metrics, and Section 3.9 summarizes the methodological contributions leading to the next chapter on experimental results and discussion.

3.1 Overview of the Proposed Research Framework

The proposed research framework integrates hybrid deep learning architectures with swarm–intelligence–based optimization to improve the accuracy, stability, and adaptability of air quality forecasting. It addresses three key challenges identified in Chapter 2: (i) the nonstationary and highly variable nature of pollutant time series discussed in Sections 2.5 and 2.7, (ii) the need to simultaneously model spatial and temporal dependencies as highlighted in Section 2.8, and (iii) the inefficiency and instability of manual hyperparameter tuning summarized in Section 2.9.

To clarify the contributions of the four constituent studies, an integrated methodological workflow is illustrated in Figure 3.1. The workflow begins with data acquisition and preprocessing, followed by four distinct modeling pipelines that collectively form the foundation of this research: (i) development and validation of an Improved Dung Beetle Optimization (IDBO) algorithm on benchmark test functions, serving as a metaheuristic optimization foundation; (ii) an IDBO–optimized CNN–LSTM–Attention hybrid model designed for single-station multivariate forecasting that integrates pollutant and meteorological features; (iii) an IDBO–enhanced CNN model for single-station univariate $PM_{2.5}$ prediction and comparative evaluation with other machine learning and deep learning baselines; and (iv) a multi-station, multi-pollutant spatiotemporal hybrid framework that combines Seasonal–Trend decomposition using Loess (STL) and Variational Mode Decomposition (VMD) with LSTM, TCN, and GRU modules, optimized by DBO—referred to as the STV–LTG model. These pipelines converge at the optimization and evaluation stages, demonstrating how swarm–intelligence–driven optimization (IDBO/DBO) enhances the performance of diverse deep learning models across varying prediction scenarios.

The overall methodology follows a logical sequence from data preparation to model optimization. It begins with the acquisition of hourly $PM_{2.5}$ and NO_2 concentrations, along with key meteorological parameters such as temperature, wind speed, and relative humidity from multiple monitoring stations. Preprocessing steps—including handling

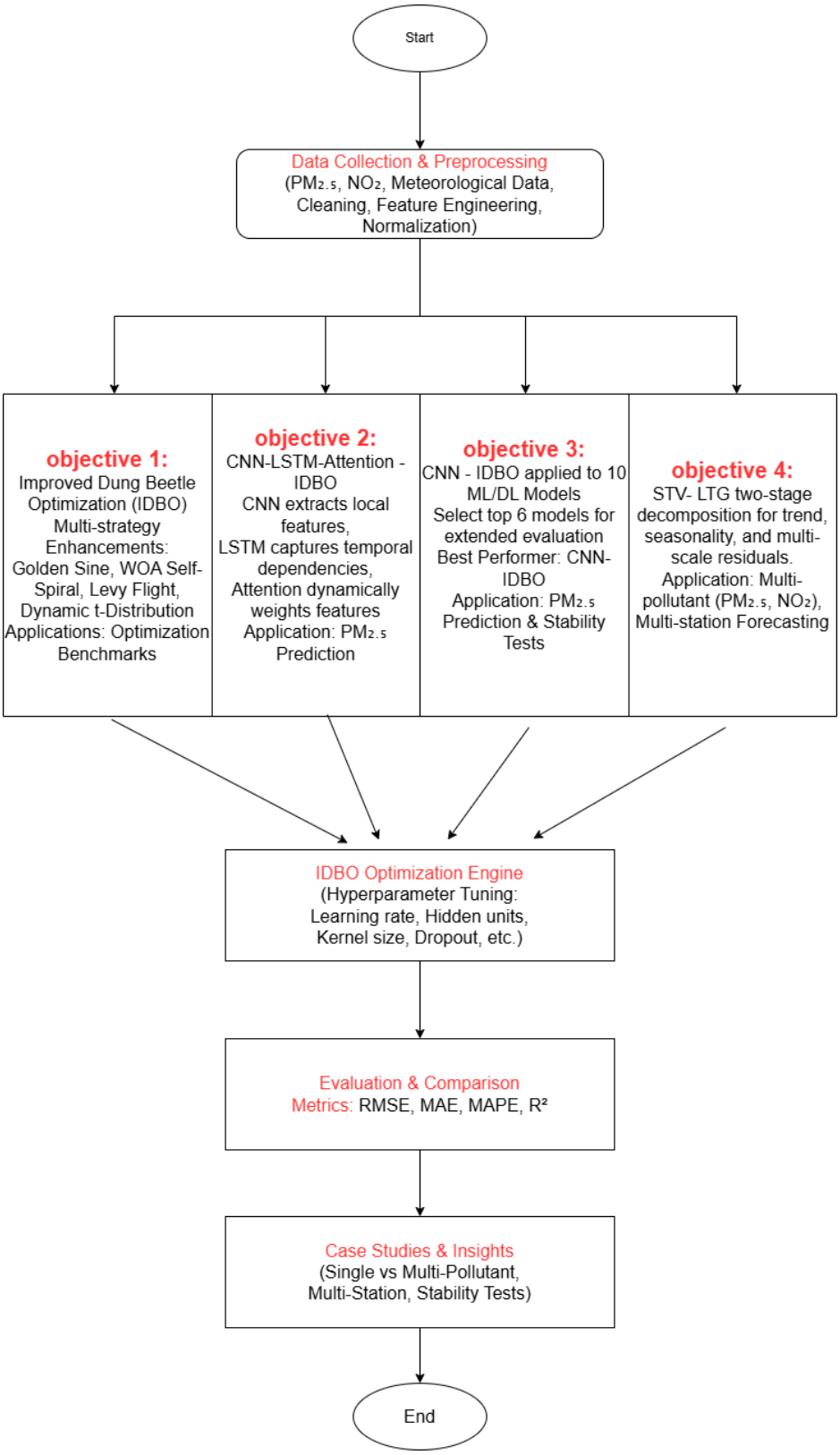

Figure 3.1: Integrated methodological framework of the four sub-studies based on IDBO optimization for air quality forecasting.

of missing values, outlier detection, normalization, and temporal alignment—ensure data quality and consistency. Subsequent modeling differs across studies according to their respective research objectives: for the first study, IDBO is validated on synthetic benchmark functions to test its convergence and search capabilities; for the second and third studies, hybrid deep learning models (CNN–LSTM–Attention and CNN) are constructed to forecast pollutant concentrations using real monitoring data, with IDBO applied for hyperparameter optimization; and for the fourth study, time–series decomposition (STL–VMD) is introduced prior to modeling to explicitly address the nonstationary and multi-scale characteristics of pollutant data. The decomposed trend, seasonal, and residual components are then modeled by LSTM, TCN, and GRU branches respectively, forming the multi-branch STV–LTG framework.

Through this structured integration of deep learning and metaheuristic optimization—and by selectively applying decomposition techniques where necessary—the framework provides a coherent and adaptive foundation for high-precision air quality forecasting across different temporal and spatial contexts. The next section introduces the datasets and preprocessing steps that support these methodological developments.

3.2 Data Description and Preprocessing

The empirical evaluation in this thesis is based on four datasets (DS1–DS4), each corresponding to one of the four sub-studies. These datasets comprise benchmark test functions used for algorithm validation and real-world air quality monitoring records collected in Penang, Malaysia. Table 3.1 summarizes their key characteristics, including time coverage, number of stations, pollutant and meteorological variables, sample sizes, and their respective applications within the study.

Table 3.1: Summary of datasets employed in this study.

Dataset ID	Time Range	No. of Stations	Variables	Samples (hours)	Application in Thesis
DS1	–	–	Benchmark test functions for optimization performance evaluation	–	Algorithm validation for Improved Dung Beetle Optimization (IDBO)
DS2	2022.01.01–2022.12.31	1	Pollutants ($PM_{2.5}$, PM_{10}, NO_2, SO_2, CO) and meteorological variables (temperature, wind speed, humidity)	8,760	Single-station hybrid model (IDBO–CNN) for pollutant forecasting
DS3	2021.01.01–2022.12.31	1	$PM_{2.5}$	17,520	Single-station hybrid model (IDBO–CNN–LSTM–Attention) for long-horizon prediction
DS4	2018.01.01–2021.12.31	4	$PM_{2.5}$, NO_2, CO	35,040	Multi-station hybrid spatiotemporal model (STL–VMD–LSTM–TCN–GRU optimized by IDBO) for multi-pollutant forecasting

Note. DS1 employs standard mathematical benchmark functions such as Sphere, Rastrigin, Ackley, and Griewank to evaluate the optimization performance of the Improved Dung Beetle Optimization (IDBO) algorithm. DS2 to DS4 represent real-world air quality datasets collected from Penang, Malaysia, covering different spatial and temporal configurations for single-station and multi-station forecasting experiments.

The first dataset (DS1) consists not of real-world observations but of synthetic benchmark test functions widely used in global optimization research (X. Li & et al., 2019; Mirjalili, 2019). It is employed to validate the robustness, convergence behavior, and overall performance of the Improved Dung Beetle Optimization (IDBO) algorithm.

The inclusion of meteorological variables helps capture the environmental drivers of air pollution, thereby improving the predictive accuracy (Y. Zhang et al., 2012).

The third dataset (DS3) consists of hourly $PM_{2.5}$ measurements collected between January 2021 and December 2022, with a total of 17,520 observations from a single station in Penang. This dataset supports multi-model optimization experiments, where ten machine learning and deep learning models (e.g., CNN, LSTM, GRU, ARIMA) are trained and compared within the IDBO optimization framework. The two-year span enables the assessment of both short-term forecasting performance and long-term stability.

The fourth dataset (DS4) includes 35,040 hourly records of $PM_{2.5}$, NO_2, and CO collected from four monitoring stations in Penang between January 2018 and December 2021. This multi-station dataset enables spatiotemporal forecasting by integrating Seasonal–Trend decomposition using LOESS (STL) and Variational Mode Decomposition (VMD) with a hybrid LSTM–TCN–GRU fusion network. Unlike DS1–DS3, which employ the Improved Dung Beetle Optimization (IDBO) algorithm, DS4 utilizes the original Dung Beetle Optimization (DBO) for hyperparameter tuning, reflecting a balance between computational efficiency and predictive performance (Y. Zhang et al., 2020).

Although both $PM_{2.5}$ and PM_{10} are important indicators of particulate pollution, this study focuses on $PM_{2.5}$ for several scientific and policy-related reasons. First, $PM_{2.5}$ represents fine particles with aerodynamic diameters smaller than 2.5 μm, which can penetrate deep into the respiratory tract and pose greater health risks than coarser PM_{10} particles. Second, $PM_{2.5}$ exhibits stronger temporal variability and is more sensitive

to meteorological conditions such as humidity, wind speed, and temperature, making it a more suitable variable for testing the forecasting capability of data-driven models. Third, $PM_{2.5}$ has been the primary focus of air quality management and public health regulations in Malaysia since the revision of the Air Pollution Index (API) framework, whereas PM_{10} concentrations are often correlated but less responsive to short-term emission changes. Therefore, emphasizing $PM_{2.5}$ enables this research to align with both environmental policy relevance and the technical challenge of fine-scale air quality prediction.

3.2.1 Preprocessing

A consistent and leakage–aware preprocessing workflow was applied to all real–world datasets (DS2–DS4). To ensure temporal integrity, all statistical parameters for imputation, outlier detection, and normalization were computed exclusively from the *training* subset (or training fold) and subsequently applied to the validation and testing sets. Time stamps were uniformly converted to hourly resolution, and for DS4, data from multiple monitoring stations were synchronized by their time indices.

Correlation analysis was performed to identify the most influential predictors and to mitigate multicollinearity among features. In particular, the Pearson correlation coefficient between $PM_{2.5}$ and potential covariates (pollutants and meteorological factors) was used to screen out redundant variables. This process helps retain only variables that contribute meaningful information, improving both model interpretability and generalization.

Before detailing dataset-specific operations, several basic statistical definitions used throughout the preprocessing procedure are provided below.

$$r_{x,y} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2}} \in [-1, 1]. \tag{3.1}$$

$$x(t) = x_{t_{i-1}} + \frac{x_{t_{i+1}} - x_{t_{i-1}}}{t_{i+1} - t_{i-1}}(t - t_{i-1}), \quad t \in (t_{i-1}, t_{i+1}). \tag{3.2}$$

$$\text{IQR} = Q_3 - Q_1, \qquad \text{LB} = Q_1 - 1.5\,\text{IQR}, \quad \text{UB} = Q_3 + 1.5\,\text{IQR}. \tag{3.3}$$

$$z = \frac{x - \mu_{\text{train}}}{\sigma_{\text{train}}}, \qquad x' = \frac{x - \min(x)}{\max(x) - \min(x)} \in [0, 1]. \tag{3.4}$$

Table 3.2 summarizes the dataset-specific preprocessing steps and their corresponding purposes. Although the exact techniques vary across datasets, the overall design maintains consistency in three principles: (i) preventing temporal leakage by computing all statistics from training data only, (ii) preserving the physical meaning of pollutant and meteorological variables, and (iii) ensuring that all features are standardized and comparable across stations and periods.

3.3 Fundamental Components of the Proposed Framework

This section presents the fundamental methodological components that constitute the foundation of the proposed hybrid air quality forecasting framework. It first outlines the decomposition techniques—Seasonal–Trend decomposition using Loess (STL) and Variational Mode Decomposition (VMD)—which are designed to manage the nonstationary and multi-scale nature of pollutant time series. Next, several deep learning architectures are introduced, including the Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM) network, Attention mechanism, Temporal Convolutional Network (TCN), and Gated Recurrent Unit (GRU), each focusing on distinct aspects of temporal and feature-level dependencies. Finally, the Dung Beetle Optimization (DBO) algorithm is presented, providing the metaheuristic foundation for model parameter tuning and optimization within the proposed framework. Illustrative figures are included to provide a clear understanding of the algorithmic processes and

Table 3.2: Summary of preprocessing procedures applied to DS2–DS4.

Dataset	Main Objective	Preprocessing Steps and Methods
DS2	Single-station pollutant and meteorological data (2022)	• Compute Pearson correlation between $PM_{2.5}$ and covariates; retain variables with $\|r\| > 0.4$ to reduce redundancy. • Replace missing values with the mean of each feature. • Detect and clip outliers using the IQR rule: values outside $[\mathrm{LB}, \mathrm{UB}]$ set to nearest bound. • Verify filtering visually using boxplots. • Apply Min–Max normalization on training statistics to scale all features to [0,1].
DS3	Extended $PM_{2.5}$ series (2021–2022)	• Conduct correlation analysis among lagged and temporal features to remove collinear terms. • Linearly interpolate missing hourly observations. • Detect outliers using Z–score criterion ($\|z\| > 3$) and clip to ± 3. • Apply moving-average smoothing (window size $w \in \{3, 5\}$) to reduce noise. • Standardize features using Z–score normalization.
DS4	Multi-station, multi-pollutant data (2018–2021)	• Synchronize timelines of all stations. • Fill missing values by linear interpolation to ensure uniform hourly coverage. • Detect extreme values via Z–score and clip within valid limits. • Add auxiliary features (latitude, longitude, hour, day-of-week, weekend flag) to capture spatiotemporal variability. • Apply Z–score normalization using parameters from training subset for each variable.

network architectures discussed herein.

3.3.1 Seasonal–Trend Decomposition using Loess and Variational Mode Decomposition (STL–VMD)

Decomposition-based preprocessing is widely adopted to address the nonstationary and multi-scale characteristics of pollutant time series prior to forecasting. The STL decomposition separates a raw series X_t into trend, seasonal, and residual components:

$$X_t = T_t + S_t + R_t, \tag{3.5}$$

where T_t, S_t, and R_t denote the trend, seasonal, and residual terms, respectively. While T_t and S_t are directly used for prediction, the residual R_t often contains high-frequency fluctuations requiring further refinement.

To enhance these residual signals, Variational Mode Decomposition (VMD) is applied to R_t, generating K intrinsic mode functions $u_k(t)$ that represent narrow-band components in the frequency domain:

$$\hat{R}_t = \sum_{k=1}^{K} u_k(t). \tag{3.6}$$

Each mode function captures distinct oscillatory patterns, allowing adaptive reconstruction of complex variations. The final predictive model integrates these components through separate deep learning modules:

$$\hat{y}_{t+h} = f_{LSTM}(T_{t-w:t}) + f_{TCN}(S_{t-w:t}) + f_{GRU}(\hat{R}_{t-w:t}), \tag{3.7}$$

where each term represents the prediction contribution from the corresponding subnetwork.

Algorithm 1 STL–VMD Decomposition Procedure

Input: Time series X_t; seasonal window W_S; trend window W_T; number of modes K.

Output: Decomposed trend T_t, seasonal S_t, and denoised residual $\hat{R}_t$.

Step 1: STL Decomposition

Apply STL with seasonal window W_S and trend window W_T:

$$X_t \rightarrow (T_t, S_t, R_t)$$

Remove long-term and periodic components to obtain residual R_t.

Step 2: Variational Mode Decomposition (VMD)

Initialize Lagrange multiplier $\lambda(t)$ and mode functions $\{u_k(t)\}_{k=1}^{K}$.

repeat

  for $k = 1$ to K **do**

    Update each mode $u_k(t)$ by minimizing its bandwidth around central frequency ω_k:

$$u_k^{(n+1)} \leftarrow \arg\min_{u_k} \left\| \partial_t \left[(\delta(t) + \tfrac{j}{\pi t}) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2$$

  end for

  Update frequencies ω_k and Lagrange multiplier $\lambda(t)$.

until Convergence or maximum iteration reached

Step 3: Reconstruction

Combine all decomposed modes to form the denoised residual:

$$\hat{R}_t = \sum_{k=1}^{K} u_k(t)$$

return $(T_t, S_t, \hat{R}_t)$

The complete STL–VMD decomposition process is summarized in Algorithm 1. The algorithm begins by applying STL to extract the long-term trend and seasonal components, leaving the residual term that captures irregular, high-frequency fluctua-

tions. Subsequently, Variational Mode Decomposition (VMD) iteratively decomposes the residual into K narrow-band intrinsic mode functions (IMFs), each associated with a specific frequency band. These IMFs are then recombined to produce a denoised residual $\hat{R}_t$, which, together with the trend and seasonal components, forms the input to the subsequent deep learning modules. This modular design allows the model to isolate temporal dynamics at multiple scales and improves interpretability of the forecasting process.

3.3.2 Convolutional Neural Network (CNN)

The Convolutional Neural Network (CNN) automatically extracts local features from input sequences through convolution, activation, and pooling operations. The basic operations are defined as follows:

$$Z_{i,j,k} = \sum_{m} \sum_{n} \sum_{c=1}^{C} X_{i+m,\, j+n,\, c}\, W_{m,n,c,k} + b_k, \tag{3.8}$$

$$A_{i,j,k} = \max(0, Z_{i,j,k}), \tag{3.9}$$

$$P_{i,j,k} = \max(Z_{i:i+f,\, j:j+f,\, k}), \tag{3.10}$$

$$\hat{y} = f(W^T A + b), \tag{3.11}$$

where the convolution captures local dependencies and the ReLU activation introduces nonlinearity.

To visualize how local features are extracted through successive convolution, activation, and pooling operations, the overall architecture of the Convolutional Neural Network (CNN) used in this study is illustrated in Figure 3.2. It depicts the hierarchical feature extraction process from raw input sequences to abstract representations, highlighting how convolutional filters capture localized spatial–temporal dependencies before the information is propagated to higher layers.

To further illustrate the computational process of feature extraction in CNN, the following pseudocode (Algorithm 2) summarizes the forward propagation steps used in

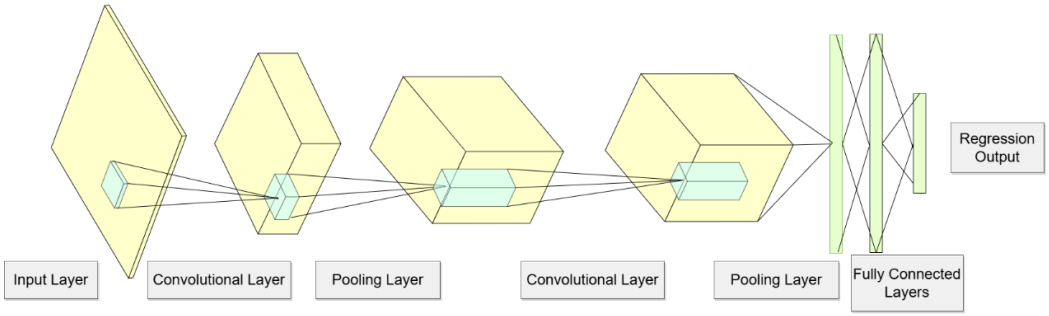

Figure 3.2: Architecture of the Convolutional Neural Network (CNN)

this study. The algorithm outlines how input sequences are convolved with kernel filters to extract spatial–temporal features, followed by activation and pooling operations, and finally passed through fully connected layers to produce the output predictions. This structure ensures efficient parameter sharing, reduces dimensionality, and enhances the model's ability to capture local temporal fluctuations in air pollutant concentrations.

Algorithm 2 Forward propagation of the Convolutional Neural Network (CNN)

Input: Input sequence $\mathbf{X} \in \mathbb{R}^{T \times C}$; convolutional kernels $\mathbf{W}$; bias $\mathbf{b}$

Output: Predicted output $\hat{y}$

Step 1: Convolution operation

for each filter $k = 1$ to K **do**

    Compute feature map: $Z_k = \text{Conv1D}(\mathbf{X}, \mathbf{W}_k) + b_k$

end for

Step 2: Activation (ReLU) $A_k = \max(0, Z_k)$

Step 3: Pooling (optional) $P_k = \text{MaxPool}(A_k, \text{window} = f)$

Step 4: Flatten feature maps $\mathbf{h} = \text{Flatten}([P_1, P_2, \ldots, P_K])$

Step 5: Fully connected layer and output $\hat{y} = f(W^T \mathbf{h} + b)$

return $\hat{y}$

3.3.3 Long Short-Term Memory (LSTM) Network

The Long Short-Term Memory (LSTM) network is an improved form of the Recurrent Neural Network (RNN) designed to mitigate gradient vanishing and explosion issues. It introduces a cell state and three gates—forget, input, and output—that regulate

information flow across time steps, as illustrated in Figure 3.3. This gating mechanism allows the model to retain long-term dependencies effectively in sequential data.

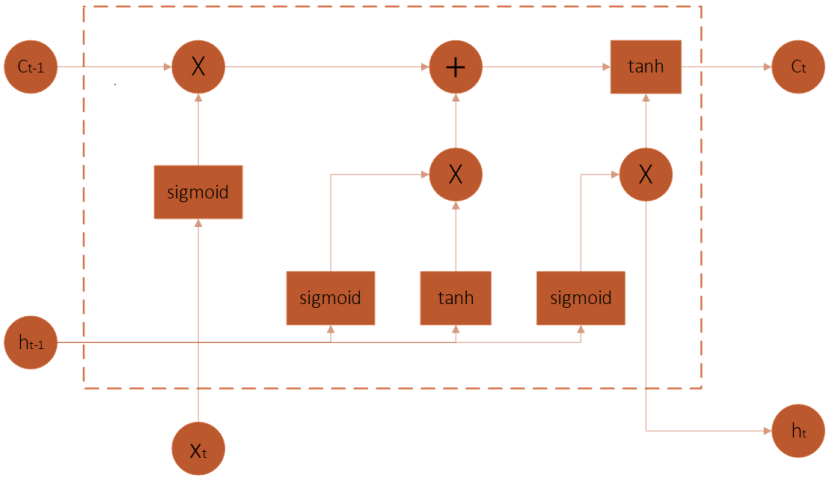

Figure 3.3: Structure of the Long Short-Term Memory (LSTM) network

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f), \tag{3.12}$$

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i), \tag{3.13}$$

$$\tilde{C}_t = \tanh(W_c[h_{t-1}, x_t] + b_c), \tag{3.14}$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t, \tag{3.15}$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o), \tag{3.16}$$

$$h_t = o_t \odot \tanh(C_t). \tag{3.17}$$

To illustrate the internal flow of information within the LSTM cell, the forward propagation process used in this study is summarized in Algorithm 3. The algorithm describes how the model processes sequential inputs, updates its gates and memory states at each timestep, and generates the hidden representations that are ultimately passed to the output layer for prediction.

Algorithm 3 Forward propagation of the Long Short-Term Memory (LSTM) network

Input: Sequence $\mathbf{X} = [x_1, x_2, \ldots, x_T]$

Output: Predicted value $\hat{y}$

Initialize $h_0, C_0 \leftarrow 0$

for $t = 1$ **to** T **do**

    Compute forget gate: $f_t = \sigma(W_f[h_{t-1}, x_t] + b_f)$

    Compute input gate: $i_t = \sigma(W_i[h_{t-1}, x_t] + b_i)$

    Compute candidate cell state: $\tilde{C}_t = \tanh(W_C[h_{t-1}, x_t] + b_C)$

    Update cell state: $C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t$

    Compute output gate: $o_t = \sigma(W_o[h_{t-1}, x_t] + b_o)$

    Update hidden state: $h_t = o_t \odot \tanh(C_t)$

end for

Compute output: $\hat{y} = f(W_y h_T + b_y)$

return $\hat{y}$

3.3.4 Attention Mechanism

The attention mechanism enhances the model's predictive capability by focusing on the most informative segments of the input sequence and effectively managing long-term dependencies. Following the LSTM, an attention layer assigns adaptive weights to the LSTM hidden states to emphasize salient temporal features:

$$S_i = \mathbf{v}^\top \tanh(\mathbf{W}H_i + \mathbf{b}), \tag{3.18}$$

$$\alpha_i = \frac{\exp(S_i)}{\sum_{j=1}^{L} \exp(S_j)}, \tag{3.19}$$

$$C = \sum_{i=1}^{L} \alpha_i H_i. \tag{3.20}$$

In this framework, the score S_i represents the relevance of each hidden state H_i to the prediction task. The normalized attention weights α_i are obtained via a softmax function, ensuring that all weights sum to one. The final context vector C is computed as the weighted sum of hidden states and serves as the enhanced temporal representation

for downstream layers. To better illustrate the computation steps involved in this process, the forward propagation procedure of the attention layer is summarized in Algorithm 4.

Algorithm 4 Forward propagation of the Attention mechanism

1: **Input:** Hidden states $\mathbf{H} = [H_1, H_2, \ldots, H_L]$

2: **Output:** Context vector $\mathbf{C}$

3: **for** $i = 1$ to L **do**

4: Compute attention score: $S_i = \mathbf{v}^\top \tanh(\mathbf{W}H_i + \mathbf{b})$

5: **end for**

6: Normalize attention weights using softmax:

$$\alpha_i = \frac{\exp(S_i)}{\sum_{j=1}^{L} \exp(S_j)}$$

7: Compute context vector as the weighted sum:

$$\mathbf{C} = \sum_{i=1}^{L} \alpha_i H_i$$

8: **return C**

3.3.5 Seasonal Modeling with Temporal Convolutional Network (TCN)

The Temporal Convolutional Network (TCN) is employed to capture periodic and short-range temporal dependencies within the seasonal component $S_t^{(i)}$ for each station. Unlike recurrent models, TCNs employ one-dimensional causal convolutions to ensure that predictions at time t depend only on past inputs. This structure efficiently models both short- and medium-term cycles such as diurnal and weekly patterns.

For an input sequence $x \in \mathbb{R}^L$, the dilated causal convolution is given by

$$(x *_d w)(t) = \sum_{k=0}^{K-1} w(k)\, x(t - d \cdot k), \tag{3.21}$$

where K is the kernel size and d is the dilation factor controlling the spacing between sampled points. The output of each TCN block is normalized and passed through a

ReLU activation:

$$h_t^{(l)} = \text{ReLU}(\text{Norm}(x *_d w + b)), \tag{3.22}$$

and residual connections are applied to stabilize learning:

$$y_t^{(l)} = x_t^{(l)} + h_t^{(l)}. \tag{3.23}$$

After stacking N_{season} layers with exponentially increasing dilation ($d = 2^{l-1}$), the seasonal feature vector $Z_t^{(i,season)}$ is obtained and used for forecasting:

$$\hat{S}_{t+1}^{(i)} = W_{season}^{out} Z_t^{(i,season)} + b_{season}^{out}. \tag{3.24}$$

Algorithm 5 Forward propagation of the Temporal Convolutional Network (TCN)

1: **Input:** Sequence $\mathbf{X} = [x_1, x_2, \ldots, x_T]$

2: **Output:** Predicted sequence $\mathbf{Y}$

3: **for** each layer $l = 1$ to L **do**

4: Apply causal padding to maintain sequence length

5: Compute dilated convolution:

$$H_t^{(l)} = \sum_{i=0}^{k-1} W_i^{(l)} X_{t-d_l \cdot i} + b^{(l)}$$

6: Apply activation: $A_t^{(l)} = \text{ReLU}(H_t^{(l)})$

7: Apply dropout for regularization (optional)

8: Compute residual connection:

$$R_t^{(l)} = A_t^{(l)} + X_t$$

9: Set $X_t \leftarrow R_t^{(l)}$ for the next layer

10: **end for**

11: Compute final output: $\mathbf{Y} = f(W_o R^{(L)} + b_o)$

12: **return Y**

To illustrate the forward propagation process in the TCN, the computational flow is summarized in Algorithm 5. This algorithm highlights how causal padding, dilated convolution, and residual connections are used to enhance temporal modeling

capability.

3.3.6 Residual Modeling with Gated Recurrent Unit (GRU)

The residual component $R_t^{(i)}$ contains high-frequency fluctuations and noise after removing trends and seasonal effects. A Gated Recurrent Unit (GRU) is employed to model these dynamics efficiently. At each time step, the GRU cell updates are defined as:

$$Z_t = \sigma(W_z X_t + U_z h_{t-1} + b_z), \tag{3.25}$$

$$r_t = \sigma(W_r X_t + U_r h_{t-1} + b_r), \tag{3.26}$$

$$\tilde{h}_t = \tanh(W_h X_t + U_h(r_t \odot h_{t-1}) + b_h), \tag{3.27}$$

$$h_t = (1 - Z_t) \odot h_{t-1} + Z_t \odot \tilde{h}_t. \tag{3.28}$$

From the top-layer hidden state, the next residual prediction is obtained as

$$\hat{R}_{t+1} = W_{res}^{out} h_t + b_{res}^{out}. \tag{3.29}$$

Algorithm 6 Forward propagation of the Gated Recurrent Unit (GRU)

Input: Sequence $\mathbf{X} = [x_1, x_2, \ldots, x_T]$
Output: Predicted output $\hat{y}$
Initialize hidden state $h_0 \leftarrow 0$
for $t = 1$ to T **do**
    Compute update gate: $z_t = \sigma(W_z[h_{t-1}, x_t] + b_z)$
    Compute reset gate: $r_t = \sigma(W_r[h_{t-1}, x_t] + b_r)$
    Compute candidate hidden state: $\tilde{h}_t = \tanh(W_h[r_t \odot h_{t-1}, x_t] + b_h)$
    Update hidden state: $h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t$
end for
Compute output: $\hat{y} = f(W_o h_T + b_o)$
return $\hat{y}$

To provide a clearer understanding of this process, the forward propagation steps of the GRU are summarized in Algorithm 6.

3.3.7 Dung Beetle Optimization (DBO) Algorithm

The Dung Beetle Optimization (DBO) algorithm simulates four key behaviors observed in dung beetles—rolling, egg deposition, food searching, and theft—which account for approximately 20%, 20%, 25%, and 35% of the population, respectively. Through these behavioral mechanisms, the algorithm balances global exploration and local exploitation in the search space.

1. *Rolling.* The position of a rolling dung beetle is updated according to

$$\begin{cases} x_i(t+1) = x_i(t) + \alpha k x_i(t-1) + b\Delta x, \\ \Delta x = |x_i(t) - X^{\omega}|, \end{cases} \tag{3.30}$$

   where b is a random coefficient uniformly distributed in $(0, 1)$, $\alpha \in \{-1, 1\}$, and X^{ω} denotes the globally least favorable position. The term Δx represents variation due to light intensity. When encountering an obstacle, the beetle adjusts its trajectory using

$$x_i(t+1) = x_i(t) + \tan\theta |x_i(t) - x_i(t-1)|, \tag{3.31}$$

   where θ denotes the tilt angle.

2. *Egg deposition.* Female dung beetles dynamically spawn within a regional search area defined by

$$\begin{cases} Lb^* = \max(X^*(1-R), Lb), \\ Ub^* = \min(X^*(1+R), Ub), \end{cases} \tag{3.32}$$

   where X^* represents the current local optimum, and $R = 1 - t/T_{\max}$. The position of the newly hatched beetles is updated as

$$B_i(t+1) = X^* + b_1(B_i(t) - Lb^*) + b_2(B_i(t) - Ub^*), \tag{3.33}$$

where b_1 and b_2 are random variation vectors of dimension $1 \times D$.

3. *Food searching.* The search boundaries for juvenile beetles are defined as

$$\begin{cases} Lb^b = \max(X^b(1-R), Lb), \\ Ub^b = \min(X^b(1+R), Ub), \end{cases} \tag{3.34}$$

where X^b is the global best position. Each beetle updates its position according to

$$x_i(t+1) = x_i(t) + C_1(x_i(t) - Lb^b) + C_2(x_i(t) - Ub^b), \tag{3.35}$$

with C_1 and C_2 denoting random coefficients.

4. *Theft.* Some beetles exhibit a "theft" behavior, attempting to steal resources from nearby individuals. The position update in this stage is given by

$$x_i(t+1) = X^b + Sg(|x_i(t) - X^*| + |x_i(t) - X^b|), \tag{3.36}$$

where g is a randomly generated vector and S is a constant scaling factor.

To clarify the iterative procedure and the interaction among these behavioral operators, the complete optimization process of DBO is summarized in Algorithm 7.

Algorithm 7 Pseudocode of the Dung Beetle Optimization (DBO) algorithm

Input: Population size N, dimension D, max iterations $T_{\max}$, bounds $[Lb, Ub]$
Output: Best solution X^b
Initialize population $\{X_i(0)\}_{i=1}^{N}$ randomly within $[Lb, Ub]$
for $t = 1$ **to** $T_{\max}$ **do**
    Evaluate fitness $f(X_i(t))$ and identify X^b, X^ω
    for each beetle $i = 1$ **to** N **do**
        if beetle in rolling group **then**
            $X_i(t{+}1) = X_i(t) + \alpha k\, X_i(t{-}1) + b\,|X_i(t) - X^\omega|$
        else if beetle in egg-deposition group **then**
            $B_i(t{+}1) = X^* + b_1(B_i(t) - Lb^*) + b_2(B_i(t) - Ub^*)$
        else if beetle in food-searching group **then**
            $X_i(t{+}1) = X_i(t) + C_1(X_i(t) - Lb^b) + C_2(X_i(t) - Ub^b)$
        else ▷ theft group
            $X_i(t{+}1) = X^b + S\,g\big(|X_i(t) - X^*| + |X_i(t) - X^b|\big)$
        end if
    end for
    Project all $X_i(t{+}1)$ to $[Lb, Ub]$ and update global best X^b
end for
return X^b

3.4 Improved Dung Beetle Optimization (IDBO) Algorithm

This section presents the Improved Dung Beetle Optimization (IDBO) algorithm developed in this study to enhance the search efficiency and robustness of the original DBO. The proposed IDBO integrates four complementary mechanisms—Golden Sine optimization, self-spiral search, Lévy flight exploration, and an adaptive t-distribution perturbation scheme—to balance global exploration and local exploitation throughout the optimization process.

3.4.1 Algorithm Flowchart

The overall workflow of the IDBO algorithm, combining these enhancement strategies, is illustrated in Figure 3.4. The pseudocode of the complete IDBO procedure is given in Algorithm 8. Each module contributes to a specific behavioral phase of the search process: Golden Sine improves global exploration during initialization, the self-spiral mechanism enhances local exploitation, Lévy flight introduces stochastic long-range jumps to escape local optima, and the adaptive t-distribution ensures convergence stability in the final iterations.

3.4.2 Technique for Golden Sine Optimization

The method for optimizing golden sine adopts the use of sinusoidal function for iterative optimization search, at the same time, Incorporating the golden section coefficient into the update process enhances the algorithm's global search ability during the early stages of iteration, and in the later period of the iteration period, it can be sufficiently searched in the local region. The position update formula utilizing the golden sine approach in the rolling period is as follows:

$$x_i(t+1) = x_i(t) \cdot |\sin r_1| - r_2 \cdot \sin r_1 \cdot |c_1 x^b - c_2 x_i(t)|. \tag{3.37}$$

r_1 is a random number for the distance travelled; r_2 is a random number for the update direction; g_1 is the golden section number, c_1 and c_2 are a constant.

3.4.3 Self-Spiral Strategy

In foraging behaviour, dung beetles search for superiority slowly, which is not conducive to the latter step of the technique's completion. Inspired by the WOA's spiral search approach, the algorithm is enhanced, and the position update formula is presented as follows:

$$x_i(t+1) = e^{zl} \cdot \cos(2\pi l) \cdot x_i(t) + c_1(x_i(t) - Lb^b) + c_2(x_i(t) - Ub^b). \quad (3.38)$$

z is a constant for the spiral equation; Random number l lies in the interval [-1,1].

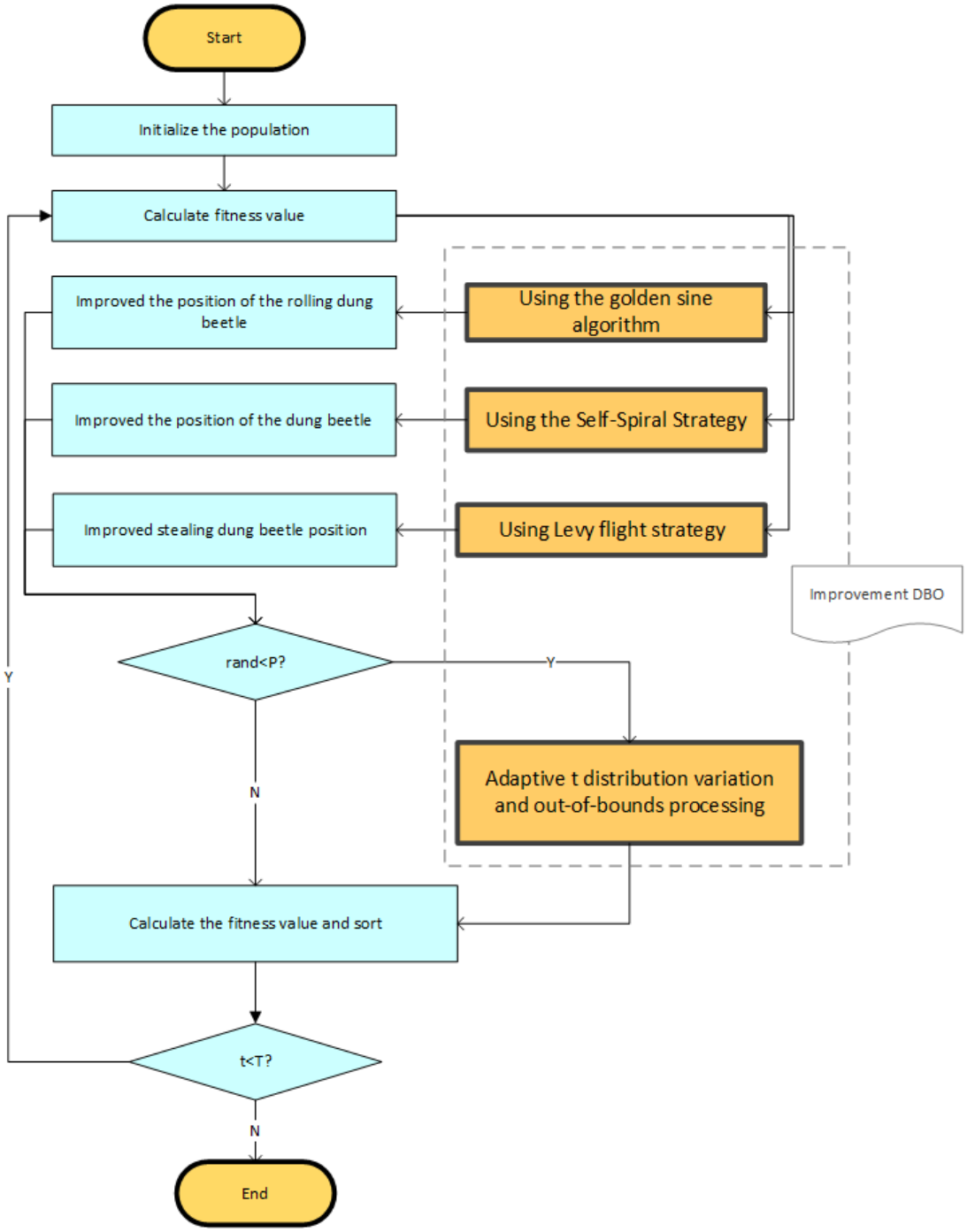

Figure 3.4: Flowchart of IDBO algorithm.

Algorithm 8 Pseudocode of the Improved Dung Beetle Optimization (IDBO) algorithm

Input: Population size N, dimension D, max iteration $T_{\max}$, bounds $[Lb, Ub]$
Output: Best solution X^b
Initialize population $\{X_i(0)\}_{i=1}^{N}$ within $[Lb, Ub]$
for $t = 1$ **to** $T_{\max}$ **do**
    Evaluate fitness $f(X_i(t))$; identify X^b (best) and X^ω (worst)
    for each beetle $i = 1$ **to** N **do**
        if beetle in rolling group **then**
            Update X_i using golden sine strategy:

$$X_i(t{+}1) = X_i(t) + A \sin(B)\, |C X^b - X_i(t)|$$

        else if beetle in egg-deposition group **then**
            Apply self-spiral update:

$$X_i(t{+}1) = X^b + r\, e^{\theta} \left(X_i(t) - X^b\right)$$

        else if beetle in food-searching group **then**
            Apply Lévy flight strategy:

$$X_i(t{+}1) = X_i(t) + \lambda\, L(\beta) \odot \left(X_i(t) - X^b\right)$$

        else ▷ theft beetle
            Perform adaptive t-distribution mutation:

$$X_i(t{+}1) = X^b + t(\nu) \odot \left(|X_i(t) - X^b| + |X_i(t) - X^*|\right)$$

        end if
    end for
    Apply out-of-bounds correction for all $X_i(t{+}1)$
    Recalculate fitness and update global best X^b
end for
return X^b

3.4.4 Levy Flying

Due to the lack of interactive behaviour between peers in the stealing behaviour, slipping into the local optimum is simple. in order to solve this problem, in the Cuckoo algorithm, we use the Levy's approach. Enhance the exploration of space and enhance

the overall perfect search function of the algorithm. Levy's flight strategy is as follows.

$$\text{Levy}(x) = 0.01 \times \frac{r^3 \times \sigma}{|r^4|^{(1/\varepsilon)}}. \tag{3.39}$$

Two random integers in this range [0,1] are r_3 and r_{34} , is a randomly generated number within the range of 0 to 2, and σ is computed as follows.

$$\sigma = \left(\frac{\Gamma(1+\xi) \times \sin(\pi\xi/2)}{\Gamma((1+\varsigma)/2) \times \xi \times 2^{((\xi-1)/2)}} \right)^{(1/\xi)}. \tag{3.40}$$

where $\Gamma(x) = (x-1)!$.

3.4.5 Adaptive t-Distribution Scheme

For high dimensionality and high complexity objective function, in the late iteration, it is very easy to ignore the global optimal position, adaptive t-distribution can be perturbed to the current position, to improve the algorithm's ability to avoid becoming trapped in local optima, and to improve the algorithm's rate of convergence and efficiency when tackling issues, the specific advances are as follows:

$$x_i^{t+1} = x_i^t + x_i^t \cdot t(iter). \tag{3.41}$$

Meanwhile, to reduce the time spend computing, the dynamic selection of the probability p is used to automatically regulate the use of the variance operator, which is formulated as follows:

$$p = w_1 - w_2 \cdot \frac{T-t}{T}, w_1 = 0.5, w_2 = 0.1. \tag{3.42}$$

The current iteration count is t , while the total number of iterations is T.

3.5 Hybrid Single-Station Model: IDBO–CNN–LSTM–Attention

This section presents a hybrid deep learning framework that integrates the Improved Dung Beetle Optimization (IDBO) algorithm with a CNN–LSTM–Attention architecture for single-station pollutant forecasting. The CNN module extracts local

spatiotemporal features, the LSTM captures long-range dependencies, and the attention mechanism reweights salient temporal steps to enhance interpretability. To avoid the inefficiency of manual parameter tuning, IDBO is used to optimize the model configuration, including convolutional kernel size, LSTM unit number, learning rate, batch size, and window length.

3.5.1 Objective and Motivation

Air-pollution time series exhibit short spikes and long-range dependencies under nonstationary regimes. To address this, we combine a causal 1D CNN (local motif extraction), an LSTM encoder (long-horizon memory), and a lightweight additive temporal attention (salient-step reweighting). Manual hyperparameter tuning is inefficient and dataset-specific; therefore we employ the improved dung beetle optimizer (IDBO; Section 3.4) to automatically search architecture/training/data-window settings.

Formally, let $f_{\boldsymbol{\lambda}}$ denote the CNN–LSTM–Attention model with hyperparameters $\boldsymbol{\lambda} \in \Omega$ (search space). With training/validation splits $\mathcal{D}_{\text{tr}}$, $\mathcal{D}_{\text{val}}$ (Section 3.2.1), the HPO objective is

$$\boldsymbol{\lambda}^{\star} \;=\; \arg\min_{\boldsymbol{\lambda}\in\Omega} \; \mathrm{RMSE}_{\text{val}}\big(f_{\boldsymbol{\lambda}};\, \mathcal{D}_{\text{tr}}, \mathcal{D}_{\text{val}}\big), \tag{3.43}$$

and the final model is retrained on $\mathcal{D}_{\text{tr}} \cup \mathcal{D}_{\text{val}}$ and evaluated on the held-out test split.

3.5.2 Overview and Pseudocode

Figure 3.5 summarizes the end-to-end workflow: data preprocessing and split, forward path (CNN $\rightarrow$ LSTM $\rightarrow$ Attention), and the outer IDBO loop that proposes hyperparameters, trains/evaluates the model, and updates the population until convergence.

3.5.3 Hyperparameters Optimized by IDBO

IDBO jointly tunes architecture, training, and data-window knobs. Table 3.3 lists the search space; integer choices are sampled discretely and real-valued knobs are

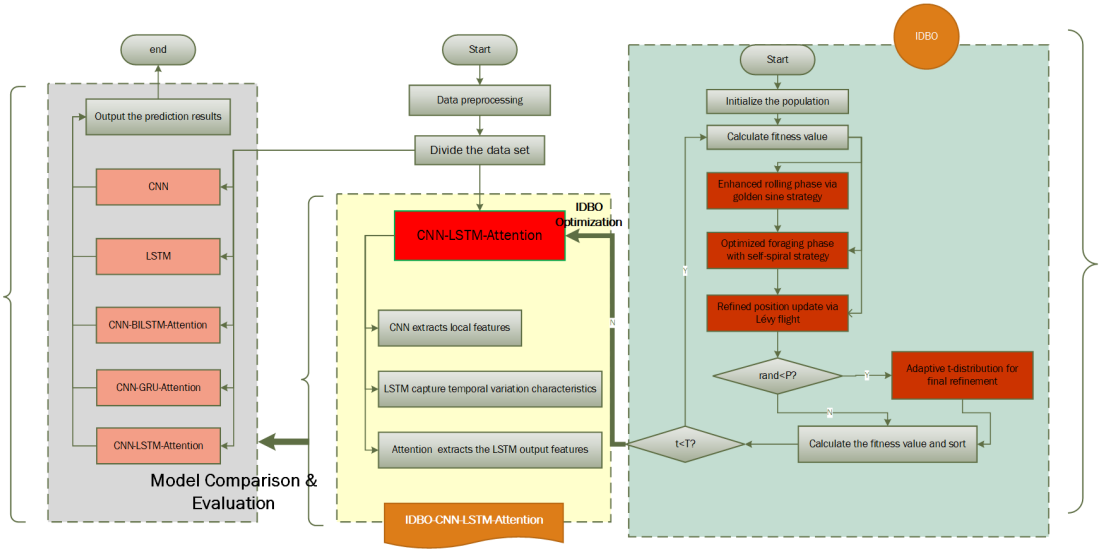

Figure 3.5: Overview of the IDBO-CNN-LSTM-Attention workflow.

Algorithm 9 IDBO optimizing CNN–LSTM–Attention (minimize $\mathrm{RMSE}_{\mathrm{val}}$)

Require: Chronological splits $(\mathcal{D}_{\mathrm{tr}}, \mathcal{D}_{\mathrm{val}})$; search space Ω; population size N; max iterations T; budget E (epochs/evaluation).

Hyperparameter vector $\lambda = \{K, F, B_{\mathrm{conv}}, U_{\mathrm{lstm}}, L_{\mathrm{lstm}}, d_{\mathrm{att}}, p_{\mathrm{conv}}, p_{\mathrm{lstm}}, \eta, \mathrm{bs}, \mathrm{wd}, \mathrm{pat}, L, H\}$.
Initialize population $\{\lambda_i^{(0)}\}_{i=1}^{N} \sim \Omega$.
for $i = 1$ **to** N **do**
    model $\leftarrow$ `BuildModel`$(\lambda_i^{(0)})$ ▷ $\mathrm{CNN}(K, F, B_{\mathrm{conv}}) \rightarrow \mathrm{LSTM}(U_{\mathrm{lstm}}, L_{\mathrm{lstm}}) \rightarrow \mathrm{Attention}(d_{\mathrm{att}}) \rightarrow \mathrm{MLP}(H)$
    $J(\lambda_i^{(0)}) \leftarrow$ `TrainEval`$(\mathrm{model}, \mathcal{D}_{\mathrm{tr}}, \mathcal{D}_{\mathrm{val}}; \eta, \mathrm{bs}, \mathrm{wd}, p., E, \mathrm{pat})$
end for
$\lambda^{\star} \leftarrow \arg\min_i J(\lambda_i^{(0)})$
for $t = 1$ **to** T **do**
    $n_{\mathrm{roll}} \leftarrow \lfloor 0.20N \rfloor$, $n_{\mathrm{brood}} \leftarrow \lfloor 0.20N \rfloor$, $n_{\mathrm{forage}} \leftarrow \lfloor 0.25N \rfloor$, $n_{\mathrm{theft}} \leftarrow N - n_{\mathrm{roll}} - n_{\mathrm{brood}} - n_{\mathrm{forage}}$
    $(\mathcal{I}_{\mathrm{roll}}, \mathcal{I}_{\mathrm{forage}}, \mathcal{I}_{\mathrm{theft}}, \mathcal{I}_{\mathrm{brood}}) \leftarrow$ `PartitionByFitness`$(\{\lambda_i^{(t-1)}\}; n.)$ ▷ elites protected
    (1) Golden–Sine Rolling (exploitation): $\{\lambda_i^{(t)} : i \in \mathcal{I}_{\mathrm{roll}}\} \leftarrow$ `GoldenSineRoll`(elites)
    (2) Self–Spiral Foraging (acceleration): $\{\lambda_i^{(t)} : i \in \mathcal{I}_{\mathrm{forage}}\} \leftarrow$ `SelfSpiralForage`$(\lambda^{\star}, \{\lambda_j^{(t-1)}\})$
    (3) Levy–Flight Perturbation (exploration): $\{\lambda_i^{(t)} : i \in \mathcal{I}_{\mathrm{theft}}\} \leftarrow$ `LevyPerturb`$(\{\lambda_j^{(t-1)}\})$
    (4) Adaptive t Refinement (brooding): $\{\lambda_i^{(t)} : i \in \mathcal{I}_{\mathrm{brood}}\} \leftarrow$ `TRefine`(elites)
    $\{\lambda_i^{(t)}\} \leftarrow$ `ProjectToBox`$(\{\lambda_i^{(t)}\}, \Omega)$ ▷ clip/round to valid ranges
    for each $\lambda_i^{(t)}$ **do**
        model $\leftarrow$ `BuildModel`$(\lambda_i^{(t)})$
        $J(\lambda_i^{(t)}) \leftarrow$ `TrainEval`$(\mathrm{model}, \mathcal{D}_{\mathrm{tr}}, \mathcal{D}_{\mathrm{val}}; \eta, \mathrm{bs}, \mathrm{wd}, p., E, \mathrm{pat})$
    end for
    $\lambda^{\star} \leftarrow \arg\min_{\{\lambda_i^{(t)}\}} J(\lambda_i^{(t)})$; `KeepElites`()
end for
return $\lambda^{\star}$; retrain on $\mathcal{D}_{\mathrm{tr}} \cup \mathcal{D}_{\mathrm{val}}$; evaluate once on $\mathcal{D}_{\mathrm{test}}$.

sampled continuously within bounds.

Table 3.3: Search space used by IDBO for the CNN–LSTM–Attention model.

Category	**Hyperparameter**	**Range / choices**
CNN	Kernel size K	$\{3, 5, 7\}$
	Filters per conv. block F	$\{32, 64, 128\}$
	No. of conv. blocks B_{conv}	$\{1, 2, 3\}$
LSTM	Hidden units U_{lstm}	$\{64, 128, 256\}$
	No. of layers L_{lstm}	$\{1, 2\}$
	Dropout (LSTM) p_{lstm}	$[0,\ 0.5]$
Attention	Attention dimension d_{att}	$\{32, 64, 128\}$
Training	Learning rate (Adam) η	$[1 \times 10^{-4},\ 5 \times 10^{-3}]$
	Batch size bs	$\{32, 64, 128\}$
	Weight decay wd	$\{0,\ 10^{-5},\ 10^{-4}\}$
	Early-stopping patience pat (epochs)	$\{5, 10, 15\}$
Data window	Lookback L (hours)	$\{24, 48, 72\}$
	Horizon H (steps)	$\{1, 3, 6\}$

Note: hourly data are assumed for L; if the horizon is hourly, H is in hours as well. This table matches the hyperparameter vector $\boldsymbol{\lambda}$ in Algorithm 9.

The IDBO–CNN–LSTM–Attention framework adaptively balances representation learning and optimization efficiency. By embedding IDBO into the model selection loop, it ensures data-driven architecture configuration without manual intervention, achieving robust pollutant forecasting performance.

3.6 Single-Station Univariate Model: IDBO–CNN

This section introduces a compact convolutional neural network (CNN) optimized by the Improved Dung Beetle Optimization (IDBO) algorithm for single-station $PM_{2.5}$ forecasting. Compared to recurrent structures, the CNN architecture provides a lightweight and computationally efficient alternative capable of capturing short-term local temporal patterns such as transient spikes or rapid pollution bursts. All network hyperparameters are optimized automatically through IDBO rather than manual trial and error.

3.6.1 Objective and Motivation

For a single-station, univariate $PM_{2.5}$ series, a compact 1D convolutional network (CNN) is well-suited to capture short-term local patterns (e.g., spikes and bursts) with low latency and fewer parameters than recurrent alternatives. To avoid manual, dataset-specific tuning, we couple the CNN backbone with the Improved Dung Beetle Optimization (IDBO; Section 3.4) for automated hyperparameter optimization (HPO). Given chronological training/validation splits $\mathcal{D}_{\text{tr}}, \mathcal{D}_{\text{val}}$ (Section 3.2.1), the HPO objective is

$$\boldsymbol{\lambda}^{\star} = \arg\min_{\boldsymbol{\lambda}\in\Omega} \ \text{RMSE}_{\text{val}}\big(f_{\boldsymbol{\lambda}}^{\text{CNN}};\ \mathcal{D}_{\text{tr}}, \mathcal{D}_{\text{val}}\big),$$

where $f_{\boldsymbol{\lambda}}^{\text{CNN}}$ denotes the CNN forecaster parameterized by hyperparameters $\boldsymbol{\lambda} \in \Omega$. The final model is retrained on $\mathcal{D}_{\text{tr}} \cup \mathcal{D}_{\text{val}}$ and evaluated once on the held-out test set.

3.6.2 Model Specification

Let the input window be $\mathbf{x}_{t-L+1:t} = [x_{t-L+1}, \ldots, x_t]^{\top}$ ($PM_{2.5}$ only) and the H-step output be $\widehat{\mathbf{y}}_{t+1:t+H}$. The backbone stacks B_{conv} convolutional blocks; each block applies *Conv1D*(K kernel, F filters) $\rightarrow$ *BatchNorm* $\rightarrow$ *ReLU* (with optional dropout p_{conv}). After the last block we use *Global Average Pooling* (GAP), followed by a small MLP head to produce H outputs:

$$\mathbf{z}^{(b)} = \text{ReLU}\big(\text{BN}\big(\text{Conv1D}^{(b)}(\mathbf{z}^{(b-1)})\big)\big), \quad \mathbf{g} = \text{GAP}(\mathbf{z}^{(B_{\text{conv}})}), \quad \widehat{\mathbf{y}}_{t+1:t+H} = W\mathbf{g} + \mathbf{b}.$$

This CNN-only design matches the univariate setting (no exogenous variables) and serves as a strong, lightweight baseline for the single-station case.

3.6.3 IDBO for CNN Hyperparameter Optimization

IDBO optimizes architecture, training, and data-window knobs jointly. We keep the same outer-loop as in Fig. 3.6, replacing the backbone by the CNN defined above.

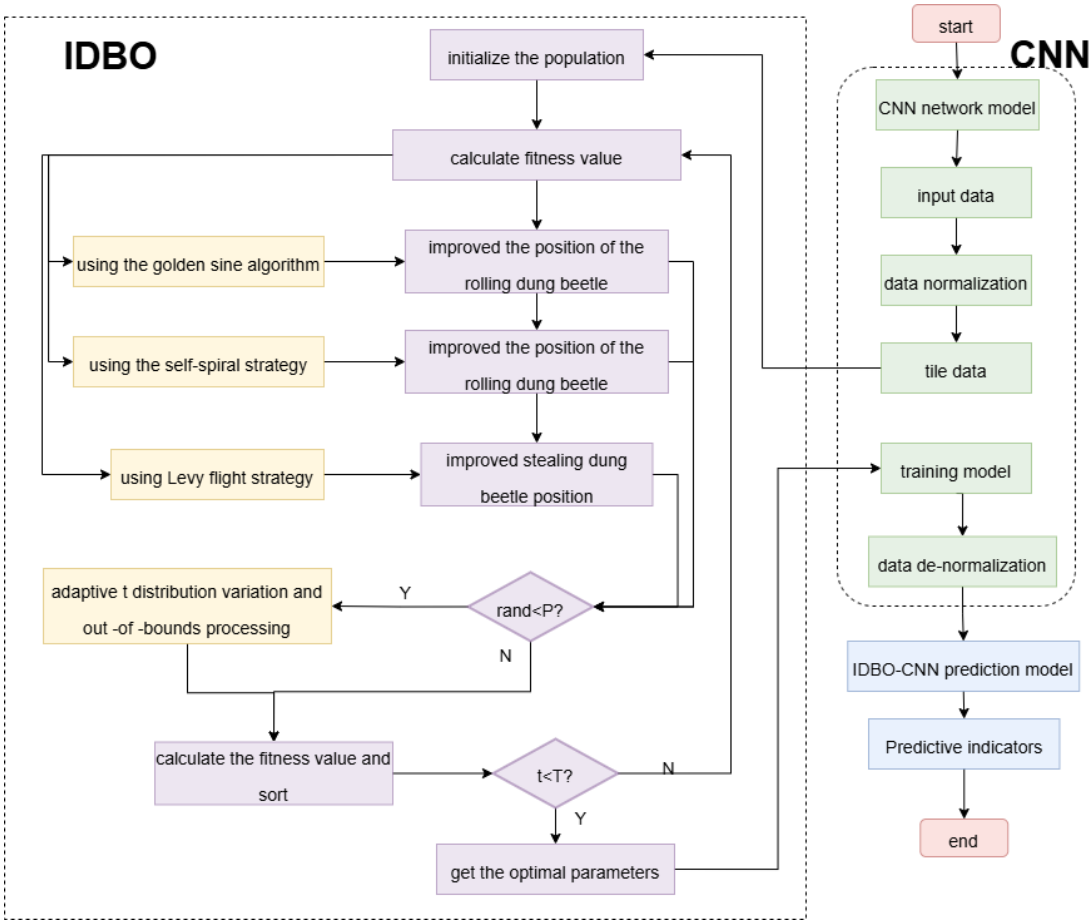

Figure 3.6: Overview of the IDBO-CNN workflow.

Table 3.4: Search space used by IDBO for the CNN-only forecaster.

Category	**Hyperparameter**	**Range / choices**
CNN	Kernel size K	$\{3, 5, 7\}$
	Filters per block F	$\{32, 64, 128\}$
	No. of conv. blocks B_{conv}	$\{1, 2, 3\}$
Training	Learning rate (Adam) η	$[1 \times 10^{-4},\ 5 \times 10^{-3}]$
	Batch size bs	$\{32, 64, 128\}$
	Dropout (conv) p_{conv}	$[0,\ 0.5]$
	Weight decay wd	$\{0,\ 10^{-5},\ 10^{-4}\}$
	Early-stopping patience pat (epochs)	$\{5, 10, 15\}$
Data window	Lookback L (hours)	$\{24, 48, 72\}$
	Horizon H (steps)	$\{1, 3, 6\}$

Algorithm 10 IDBO optimizing a CNN forecaster (minimize $\mathrm{RMSE}_{\mathrm{val}}$)

1.5em 1.2em

Require: $(\mathcal{D}_{\mathrm{tr}}, \mathcal{D}_{\mathrm{val}})$; search space Ω; population N; max iterations T; budget E (epochs/evaluation).

1: **Hyperparameter vector** $\boldsymbol{\lambda} = \{K, F, B_{\mathrm{conv}}, p_{\mathrm{conv}}, \eta, \mathrm{bs}, \mathrm{wd}, \mathrm{pat}, L, H\}$.
2: Initialize $\{\boldsymbol{\lambda}_i^{(0)}\}_{i=1}^{N} \sim \Omega$.
3: **for** $i = 1$ **to** N **do**
4: model $\leftarrow$ `BuildCNN`$\left(\boldsymbol{\lambda}_i^{(0)}\right)$
5: $J\left(\boldsymbol{\lambda}_i^{(0)}\right) \leftarrow$ `TrainEval`(model, $\mathcal{D}_{\mathrm{tr}}$, $\mathcal{D}_{\mathrm{val}}$; η, bs, wd, p_{conv}, E, pat)
6: **end for**
7: $\boldsymbol{\lambda}^{\star} \leftarrow \arg\min_i J\left(\boldsymbol{\lambda}_i^{(0)}\right)$

3.6.4 Pseudocode (IDBO–CNN)

The IDBO–CNN framework provides an efficient and interpretable solution for single-station $PM_{2.5}$ forecasting. Through the integration of IDBO, the CNN architecture adapts automatically to the characteristics of each dataset, yielding a robust and low-latency baseline model that supports subsequent multi-station and multi-pollutant extensions.

3.7 Multi-Station Hybrid Spatiotemporal Model: STV–LTG

Building on the single-station frameworks developed earlier, this section introduces a multi-station, multi-pollutant forecasting model referred to as STV–LTG, which integrates **S**easonal–**T**rend decomposition via LOESS (STL) and **V**ariational **M**ode **D**ecomposition (VMD) with three deep learning branches—LSTM, TCN, and GRU. The model is designed to capture both spatial interactions and multi-scale temporal dynamics among monitoring stations, while the Dung Beetle Optimization (DBO) algorithm is used to automatically tune hyperparameters.

3.7.1 Objective and Motivation

For multi-station, multi-pollutant forecasting (DS4), pollutant time series exhibit pronounced nonstationarity and multi-scale variability, including daily or weekly cycles, abrupt local spikes, and low-frequency drifts. To effectively model these complex

structures, the proposed STV–LTG framework adopts a two-stage decomposition strategy. First, STL separates each pollutant series into trend, seasonal, and residual components. Second, VMD refines the residual signal into multiple intrinsic mode functions (IMFs), enhancing the resolution of high-frequency variations. Each component is then processed by a specialized deep learning sub-network: LSTM for long-term trend extraction, TCN for recurrent seasonal patterns, and GRU for short-term fluctuations. To minimize manual effort and ensure robust parameterization across multiple pollutants and stations, the original Dung Beetle Optimization (DBO) algorithm is employed to perform automatic hyperparameter optimization (HPO).

Formally, with train/validation splits $(\mathcal{D}_{\text{tr}}, \mathcal{D}_{\text{val}})$, the HPO objective is

$$\boldsymbol{\lambda}^{\star} = \arg\min_{\boldsymbol{\lambda}\in\Omega} \ \text{RMSE}_{\text{val}}\big(f_{\boldsymbol{\lambda}}^{\text{STV–LTG}};\ \mathcal{D}_{\text{tr}}, \mathcal{D}_{\text{val}}\big),$$

where $f_{\boldsymbol{\lambda}}^{\text{STV–LTG}}$ denotes the full pipeline parameterized by hyperparameters $\boldsymbol{\lambda} \in \Omega$.

3.7.2 Model Architecture

Let $X_t^{(s)}$ be the target series (e.g., $PM_{2.5}$) at station s at time t. STL decomposes each series into trend $T_t^{(s)}$, seasonality $S_t^{(s)}$, and residual $R_t^{(s)}$:

$$X_t^{(s)} \ = \ T_t^{(s)} \ + \ S_t^{(s)} \ + \ R_t^{(s)}, \tag{3.44}$$

where the seasonal period is set by the sampling frequency (e.g., P=24 for hourly daily cycle). We then refine the residual via VMD into K intrinsic mode functions (IMFs) $u_k^{(s)}(t)$ and reconstruct a denoised residual $\widehat{R}_t^{(s)}$ from selected modes:

$$\widehat{R}_t^{(s)} \ = \ \sum_{k=1}^{K} u_k^{(s)}(t). \tag{3.45}$$

For each station s, we build a three-branch network:

$$\begin{aligned}
\text{(Trend branch)} \qquad & \widehat{\mathbf{y}}^{(T,s)}_{t+1:t+H} = f_{\mathrm{LSTM}}\big(\mathbf{T}^{(s)}_{t-L+1:t}\,;\,\theta_T\big),\\
\text{(Seasonal branch)} \qquad & \widehat{\mathbf{y}}^{(S,s)}_{t+1:t+H} = f_{\mathrm{TCN}}\big(\mathbf{S}^{(s)}_{t-L+1:t}\,;\,\theta_S\big),\\
\text{(Residual branch)} \qquad & \widehat{\mathbf{y}}^{(R,s)}_{t+1:t+H} = f_{\mathrm{GRU}}\big(\widehat{\mathbf{R}}^{(s)}_{t-L+1:t}\,;\,\theta_R\big),
\end{aligned}$$

where $\mathbf{T}^{(s)}_{t-L+1:t}$, $\mathbf{S}^{(s)}_{t-L+1:t}$, and $\widehat{\mathbf{R}}^{(s)}_{t-L+1:t}$ are lookback windows of length L, and H is the forecast horizon. Branch outputs are fused by a light head (linear combiner or MLP); we use a learnable affine fusion:

$$\widehat{\mathbf{y}}^{(s)}_{t+1:t+H} \;=\; W_{\mathrm{fuse}} \begin{bmatrix} \widehat{\mathbf{y}}^{(T,s)}_{t+1:t+H} \\ \widehat{\mathbf{y}}^{(S,s)}_{t+1:t+H} \\ \widehat{\mathbf{y}}^{(R,s)}_{t+1:t+H} \end{bmatrix} \;+\; \mathbf{b}_{\mathrm{fuse}}. \tag{3.46}$$

Training is performed per station and pollutant; results can be reported per-station and averaged.

3.7.3 IDBO Optimization Procedure for STV–LTG

In order to optimize the hyperparameters of STV–LTG, we adopt the Dung Beetle Optimization Algorithm (DBO) metaheuristic with four subpopulation dynamics, namely *rolling*, *brooding*, *foraging*, and *theft*. Each operator serves a distinct role: rolling exploits the neighborhood of elites, brooding provides local refinement, foraging guides candidates toward promising regions, and theft introduces long-range exploration to escape local optima. At each iteration, candidate solutions are updated by these four operators according to default population shares (20%, 20%, 25%, 35%), then projected back to the feasible search space. Each candidate configuration is trained on the training set with early stopping for up to E epochs, and its performance is evaluated on the validation set. The best-performing candidates are retained as elites, and the global best hyperparameter vector $\boldsymbol{\lambda}^{\star}$ is continuously updated.

The overall optimization process is summarized in Algorithm 11. The corresponding workflow is illustrated in Figure. 3.7, showing the interplay between the four DBO

operators, evaluation of candidate solutions, and the elite preservation strategy. This integration of decomposition-based hyperparameter space with population-based search dynamics ensures that DBO can effectively balance local exploitation and global exploration, leading to stable convergence and improved generalization performance of STV–LTG.

Algorithm 11 Overall architecture of the STV–LTG model

Input: Multivariate time series $\{X^1, \ldots, X^m\}$ from m stations; lookback window n; horizon k.
Output: Optimal hyperparameters θ^*; final prediction $\hat{Y}$ for horizon k.

Phase 1: STL–VMD Decomposition
for $i = 1$ **to** m **do**
    $(T_i, S_i, R_i) \leftarrow \mathrm{STL}(X^i)$ ▷ decompose into trend, seasonal, residual
    $R'_i \leftarrow \mathrm{VMD}(R_i; \alpha, \tau, K, \ldots)$ ▷ reconstruct residual
end for
$T_{\mathrm{sp}} \leftarrow \mathrm{concat}(T_1, \ldots, T_m)$ ▷ spatio-temporal trend
$S_{\mathrm{sp}} \leftarrow \mathrm{concat}(S_1, \ldots, S_m)$ ▷ spatio-temporal seasonal
$R_{\mathrm{sp}} \leftarrow \mathrm{concat}(R'_1, \ldots, R'_m)$ ▷ spatio-temporal remainder

Phase 2: IDBO Hyperparameter Search
Initialize population $\{\theta_j\}_{j=1}^{N}$
for $t = 1$ **to** $T_{\max}$ **do**
    for $j = 1$ **to** N **do**
        Train LSTM–TCN–GRU on $(T_{\mathrm{sp}}, S_{\mathrm{sp}}, R_{\mathrm{sp}})$ with θ_j
        $f_j \leftarrow \mathrm{MAE}_{\mathrm{val}}(\theta_j)$ ▷ validation MAE as fitness
        Update θ_j via DBO exploration/exploitation
    end for
    $\theta^* \leftarrow \arg\min_j f_j$ ▷ global best update
end for

Phase 3: Final Training & Prediction
Train LSTM–TCN–GRU with θ^* on full data
$\hat{Y} \leftarrow$ predictions for horizon k
return $\theta^*, \hat{Y}$

3.7.4 Search Space Design for IDBO in STV–LTG

Table 3.5 summarizes the hyperparameter search space adopted for DBO in the STV–LTG pipeline. The settings cover decomposition parameters for STL and VMD, network architecture choices for LSTM, TCN, and GRU learners, as well as training configurations and data windowing strategies. Specifically, STL and VMD parameters

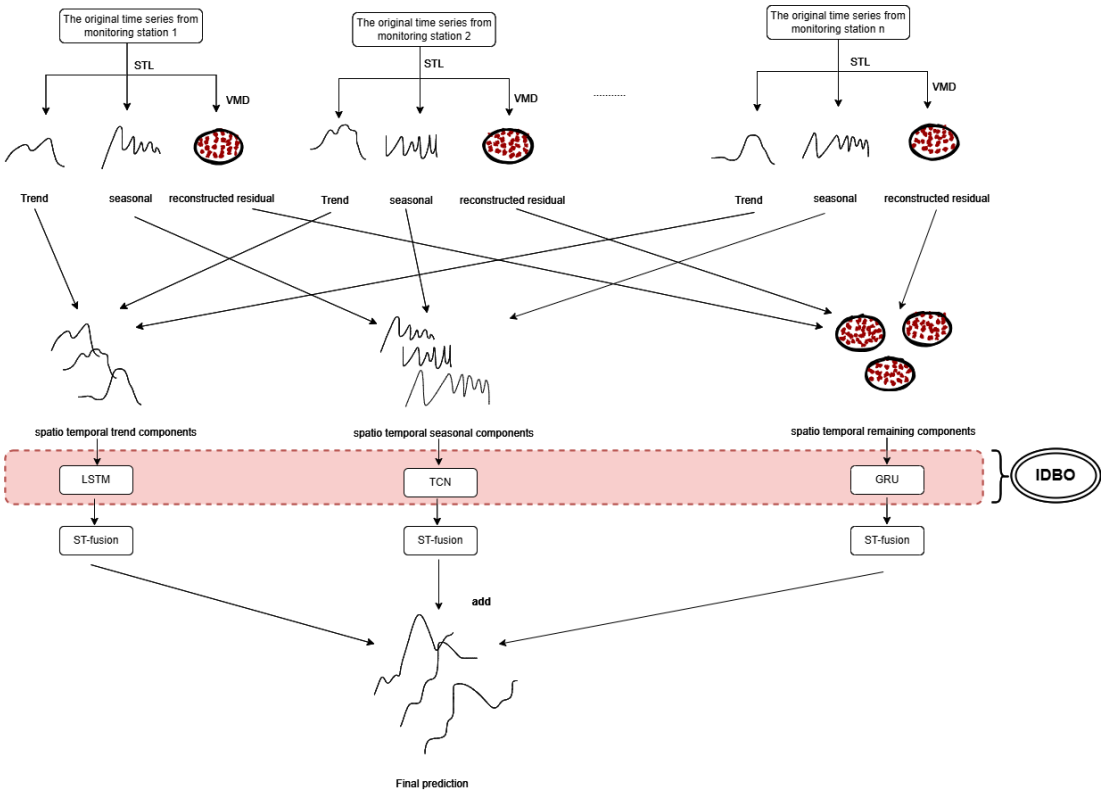

Figure 3.7: Flowchart of a Spatiotemporal Forecasting Framework Based on STL-VMD Decomposition and Deep Learning Fusion.

(W_s, W_r, K) define the decomposition behavior, while the ranges of hidden units, kernel sizes, and dropout rates control the model complexity and regularization of each learner. Training hyperparameters (learning rate, batch size, patience) ensure robust optimization, and the data window parameters (lookback, horizon) determine the temporal receptive field for forecasting. This carefully designed search space enables DBO to explore both structural and optimization-related factors, balancing model flexibility and computational feasibility.

The STV–LTG framework integrates signal decomposition, deep learning fusion, and metaheuristic optimization into a unified spatiotemporal modeling pipeline. By combining STL–VMD decomposition with specialized LSTM, TCN, and GRU branches, the model effectively captures both large-scale temporal trends and fine-grained local variations across stations. The DBO-based optimization further automates the parameter selection process, improving convergence stability and generalization performance across multiple pollutants and spatial domains.

Table 3.5: Search space for IDBO in the STV–LTG pipeline. Period P for STL is fixed by sampling (e.g., P=24 for hourly).

Category	**Hyperparameter**	**Range / choices**
STL	Seasonal window W_S	$\{13, 25, 53\}$ (LOESS span in time steps)
	Trend window W_T	$\{13, 25, 53\}$
	Robust flag rob	$\{0, 1\}$ (bisquare reweighting)
VMD	# modes K	$\{3, 4, 5, 6\}$
	Penalty α	$[100,\ 3000]$
LSTM (trend)	Hidden units U_{lstm}	$\{64, 128, 256\}$
	Layers L_{lstm}	$\{1, 2\}$
	Dropout p_T	$[0,\ 0.5]$
TCN (seasonal)	Kernel size k_S	$\{2, 3, 5\}$
	Filters F_S	$\{32, 64, 128\}$
	Blocks B_S (residual stacks)	$\{2, 3, 4\}$
	Dropout p_S	$[0,\ 0.5]$
GRU (residual)	Hidden units U_{gru}	$\{64, 128, 256\}$
	Layers L_{gru}	$\{1, 2\}$
	Dropout p_R	$[0,\ 0.5]$
Training	Learning rate (Adam) η	$[1 \times 10^{-4},\ 5 \times 10^{-3}]$
	Batch size bs	$\{32, 64, 128\}$
	Weight decay wd	$\{0,\ 10^{-5},\ 10^{-4}\}$
	Early-stopping patience pat (epochs)	$\{5, 10, 15\}$
Data window	Lookback L (hours)	$\{24, 48, 72\}$
	Horizon H (steps)	$\{1, 3, 6\}$

Table 3.6: Summary of traditional baselines and proposed hybrid models.

Category	Model / Abbreviation	Key Components	Application / Objective
Traditional Baselines	MLP	Fully connected neural network for nonlinear regression.	Used as a baseline for pollutant forecasting.
	LSTM / GRU / Transformer	Temporal deep learning models capturing sequential dependencies.	Compared with hybrid models for single- and multi-station forecasting.
	CNN–LSTM	Convolutional–recurrent hybrid for local–temporal feature extraction.	Evaluated for multi-step forecasting consistency.
Proposed Models	DBO / IDBO	Original and improved dung beetle optimization algorithms.	Metaheuristic optimization for hyperparameter tuning.
	IDBO–CNN–LSTM–Attention	CNN and LSTM layers optimized by IDBO with attention weighting.	Single-station multivariate forecasting (pollutants + meteorology).
	IDBO–CNN	Compact CNN optimized by IDBO.	Single-station univariate $PM_{2.5}$ forecasting.
	STV–LTG	STL–VMD decomposition with LSTM, TCN, and GRU fusion optimized by DBO.	Multi-station, multi-pollutant spatiotemporal forecasting.

Note. The traditional baselines represent commonly used deep learning reference architectures, whereas the proposed hybrid models integrate optimization, decomposition, and spatiotemporal learning to enhance forecasting accuracy and robustness.

To summarize, this chapter introduces both conventional deep learning baselines and four proposed hybrid models that form the core of the research framework. The traditional baselines (e.g., MLP, LSTM, GRU, Transformer, and CNN–LSTM) serve as reference architectures for comparison, while the proposed models—IDBO, IDBO–CNN–LSTM–Attention, IDBO–CNN, and STV–LTG-represent the methodological innovations designed for different forecasting tasks. Table 3.6 provides an overall summary of these models, including their key components and corresponding objectives within this study.

3.8 Evaluation Metrics

To quantitatively assess the forecasting accuracy and generalization performance of the proposed models, five commonly used statistical evaluation metrics are employed: the Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and the coefficient of determination (R^2). These metrics together provide a comprehensive evaluation from both absolute and relative perspectives, capturing deviations, dispersion, and proportional bias between predicted and observed values.

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|y_i - \hat{y}_i\right|, \tag{3.47}$$

where MAE measures the average magnitude of prediction errors without considering their direction. It provides an intuitive interpretation of the average deviation between the predicted and actual observations.

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2, \tag{3.48}$$

where MSE squares the deviations, penalizing larger errors more heavily, and thus emphasizing outlier sensitivity.

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}, \tag{3.49}$$

where RMSE represents the standard deviation of the residuals. It maintains the same units as the target variable, making it easy to interpret the scale of prediction error.

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}, \tag{3.50}$$

where R^2 (coefficient of determination) indicates how well the model explains the variance of the observed data. Its value ranges between 0 and 1, with higher values implying stronger goodness of fit.

$$\text{MAPE} = \frac{100\%}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right|, \tag{3.51}$$

where MAPE expresses the mean prediction error as a percentage of the actual value, making it unit-independent and suitable for cross-variable comparisons. However, MAPE becomes unreliable when y_i approaches zero.

In this study, RMSE and MAE are primarily used to evaluate absolute prediction accuracy, MAPE is used to assess relative error, and R^2 is adopted to reflect the model's explanatory power. Together, these indicators provide a balanced and comprehensive evaluation of forecasting performance across different pollutants and stations.

3.9 Summary

This chapter presented the complete methodological framework and the component models developed in this study. Section 3.1 outlined the overall workflow from data ingestion through model training to performance evaluation. Section 3.2 described the data preprocessing pipeline, including missing-value imputation, outlier detection, normalization, feature construction, and chronological partitioning tailored to datasets DS2–DS4. Section 3.3 introduced the foundational learning components—

CNN, LSTM, GRU, TCN, and Attention—together with the decomposition methods STL and VMD, which serve as the fundamental building blocks of the proposed models. Section 3.4 elaborated on the Improved Dung Beetle Optimization (IDBO) algorithm and its four enhancement operators designed to strengthen both global exploration and local exploitation.

Building upon these components, Section 3.5 developed the IDBO–CNN–LSTM–Attention hybrid model for single-station forecasting with exogenous meteorological inputs, while Section 3.6 proposed a lightweight IDBO–CNN forecaster for univariate $PM_{2.5}$ prediction. For multi-station and multi-pollutant forecasting, Section 3.7 introduced the STV–LTG framework, which integrates STL and VMD decomposition with an LSTM–TCN–GRU fusion network. In this pipeline, the original DBO was deliberately adopted for hyperparameter optimization because: (i) the decomposition–fusion process substantially increases computational cost per evaluation; (ii) the search space includes several discrete and hierarchical parameters (e.g., VMD mode number K, STL window sizes W_S and W_T) where the standard DBO already achieves robust performance; and (iii) this choice preserves methodological consistency with the corresponding published sub-study. Consequently, the enhanced IDBO is reserved for other hybrid models where its adaptive strategies yield clearer improvements in efficiency and convergence.

Section 3.8 defined the evaluation metrics (MAE, MSE, RMSE, MAPE, and R^2) and the training and early-stopping protocol used to ensure reliable and unbiased model comparison. The chapter also specified the hyperparameter search spaces, optimization procedures, and pseudocode implementations to promote reproducibility.

The next chapter presents comprehensive experiments conducted on datasets DS1–DS4, including: (1) algorithmic validation of DBO and IDBO on benchmark optimization functions, (2) single-station $PM_{2.5}$ forecasting with and without meteorological features, and (3) multi-station, multi-pollutant spatiotemporal prediction. Test-set performance, ablation analyses, and statistical evaluations are reported to verify the

effectiveness and generalizability of the proposed frameworks.

REFERENCES

Amil, N., Latif, M. T., Khan, M. F., & Mohd Nadzir, M. S. (2022). Air quality changes in malaysia during the covid-19 lockdown. *Science of the Total Environment*, *787*, 147619. https://doi.org/10.1016/j.scitotenv.2021.147619

Amil, N., Latif, M. T., Khan, M. F., & Mohd Tahir, N. (2016). Seasonal variability of $pm_{2.5}$ in urban and rural areas of peninsular malaysia. *Atmospheric Environment*, *125*, 83–98. https://doi.org/10.1016/j.atmosenv.2015.10.050

Apte, J. S., Brauer, M., & Cohen, A. J. (2022). Ambient pm2.5 reduces global life expectancy. *Nature Communications*, *13*, 350. https://doi.org/10.1038/s41467-021-27325-3

Awang, M. B., Jaafar, A. B., Abdullah, A. M., Ismail, M. B., Hassan, M. N., Abdullah, R., Johan, S., & Noor, H. (2000). Air quality in malaysia: Impacts, management issues and future challenges. *Respirology*, *5*(2), 183–196. https://doi.org/10.1046/j.1440-1843.2000.00248.x

Barré, J., Petetin, H., Colette, A., Guevara, M., Peuch, V., Rouil, L., & Engelen, R. (2021). Estimating lockdown impacts on air quality in europe using air quality models: Strengths and limitations of the approach. *Atmospheric Chemistry and Physics*, *21*(5), 3093–3114. https://doi.org/10.5194/acp-21-3093-2021

Betha, R., & Balasubramanian, R. (2013). $Pm_{2.5}$ and associated trace gases from tropical biomass burning in southeast asia. *Atmospheric Chemistry and Physics*, *13*, 7361–7370. https://doi.org/10.5194/acp-13-7361-2013

Box, G., & Jenkins, G. (2020). Time series analysis and arima forecasting of air pollutants: A case study. *Atmospheric Environment*, *224*, 117225. https://doi.org/10.1016/j.atmosenv.2020.117225

Brook, R. D., Rajagopalan, S., Pope, C. A., Brook, J. R., Bhatnagar, A., Diez-Roux, A. V., Holguin, F., Hong, Y., Luepker, R. V., & Mittleman, M. A. (2010). Particulate matter air pollution and cardiovascular disease: An update to the scientific statement from the american heart association. *Circulation*, *121*(21), 2331–2378. https://doi.org/10.1161/CIR.0b013e3181dbece1

Chang, Y., Xu, J., Tie, X., & Gao, W. (2020). Explore spatio-temporal pm2.5 features in northern taiwan using machine learning techniques. *Atmospheric Environment*, *224*, 117277. https://doi.org/10.1016/j.atmosenv.2020.117277

Chen, J., Li, W., & Chen, X. (2020). Machine learning for air quality modeling and forecasting: A survey. *Atmosphere*, *11*(7), 777. https://doi.org/10.3390/atmos11070777

Chen, K., Wang, M., Huang, C., & Kinney, P. (2022). Air pollution reduction and mortality benefit during the covid-19 outbreak in china. *Lancet Planetary Health*, *6*(9), e746–e754. https://doi.org/10.1016/S2542-5196(22)00185-9

Chen, L., Guo, B., & Huang, G. (2021). Land-use patterns and spatial variability of air pollution in urban areas: A review. *Science of the Total Environment*, *755*, 142522. https://doi.org/10.1016/j.scitotenv.2020.142522

Chen, L., & Wu, J. (2023). Spatio-temporal gcn for multi-station air quality prediction. *Environmental Modelling & Software*, *164*, 105605. https://doi.org/10.1016/j.envsoft.2023.105605

Chen, Y., Liu, M., & Zhang, T. (2023). Advances and challenges in chemical transport modeling for real-time air quality forecasting. *Atmospheric Environment*, *297*, 119525. https://doi.org/10.1016/j.atmosenv.2022.119525

Cheng, K., Amil, N., & Rahman, N. A. (2023). Urban–suburban gradients of pm2.5 and no2 in penang, malaysia: Evidence of strong spatial heterogeneity. *Environmental Monitoring and Assessment*, *195*(6), 812. https://doi.org/10.1007/s10661-023-10755-8

Collaborators, G. 2. R. F. (2020). Global burden of 87 risk factors in 204 countries and territories, 1990–2019: A systematic analysis. *The Lancet*, *396*(10258), 1223–1249. https://doi.org/10.1016/S0140-6736(20)30752-2

Colvile, R. N., Hutchinson, E. J., Mindell, J. S., & Warren, R. F. (2002). The transport sector as a source of air pollution. *Atmospheric Environment*, *35*, 1537–1565. https://doi.org/10.1016/S1352-2310(00)00551-3

Department of Environment Malaysia. (2017). A guide to air pollutant index (api) in malaysia.

Department of Environment Malaysia. (2020). Malaysia ambient air quality report 2020.

Duan, H., & Liu, P. (2023). Air quality forecasting with arima–cnn–lstm–dbo hybrid model. *Environmental Research*, *231*, 116073. https://doi.org/10.1016/j.envres.2023.116073

Feng, R., & Chen, H. (2020). Lstm networks for station-level $pm_{2.5}$ prediction in urban china. *Ecological Informatics*, *60*, 101150. https://doi.org/10.1016/j.ecoinf.2020.101150

Feng, R., & Chen, H. (2023). Hybrid deep learning with meteorology for spatio-temporal air pollution forecasting. *Environmental Research*, *216*, 114653. https://doi.org/10.1016/j.envres.2023.114653

Fotheringham, A. S., Yang, W., & Kang, W. (2020). *Multiscale geographically weighted regression (mgwr)*. Wiley. https://doi.org/10.1002/9781119482084

Gao, Y., Wang, X., & Zhang, H. (2021). Air pollution forecasting using spatiotemporal deep learning: A review. *Atmospheric Environment*, *246*, 118104. https://doi.org/10.1016/j.atmosenv.2020.118104

Geng, G., Xiao, Q., Zheng, Y., Tong, D., Zhang, Y., Zhang, X., Liu, F., He, K., & Liu, Y. (2022). Tracking air pollution in china: Near real-time pm2.5 retrievals from multiple satellite platforms. *Remote Sensing of Environment*, *268*, 112772. https://doi.org/10.1016/j.rse.2021.112772

Gkatzelis, G., Gilman, J., Brown, S., Eskes, H., & et al. (2021). The global impacts of covid-19 lockdowns on urban air pollution: A review. *Elementa: Science of the Anthropocene*, *9*(1), 00176. https://doi.org/10.1525/elementa.2021.00176

Grange, S. K., & Carslaw, D. C. (2018). Random forest meteorological normalisation models for swiss pm_{10} trend analysis. *Atmospheric Chemistry and Physics*, *18*(9), 6223–6239. https://doi.org/10.5194/acp-18-6223-2018

Grange, S., Lee, J., Drysdale, W., Lewis, A., Hueglin, C., Emmenegger, L., & Carslaw, D. (2021). Covid-19 lockdowns highlight a risk of increasing ozone pollution in european urban areas. *Atmospheric Chemistry and Physics*, *21*(9), 7105–7120. https://doi.org/10.5194/acp-21-7105-2021

Guo, J., & Wang, X. (2023). Hybrid graph-temporal decomposition network for regional air pollution forecasting. *Environmental Research*, *215*, 114639. https://doi.org/10.1016/j.envres.2023.114639

Guo, Y., Zeng, Y., & Li, S. (2021). Long-term exposure to ambient pm2.5 and cardiovascular mortality: A multi-country cohort study. *Environment International*, *146*, 106233. https://doi.org/10.1016/j.envint.2020.106233

Gupta, A., Brauer, M., Burnett, R., & Cohen, A. (2019). Global epidemiology of air pollution and health: A systematic review. *Lancet Planetary Health*, *3*(7), e300–e312. https://doi.org/10.1016/S2542-5196(19)30161-3

Hammer, M. S., van Donkelaar, A., & Martin, R. V. (2020). Effects of covid-19 lockdowns on no2 and pm2.5 across europe: Satellite-based evidence. *Atmospheric Chemistry and Physics*, *20*(21), 12607–12630. https://doi.org/10.5194/acp-20-12607-2020

Han, G., Chen, X., & Li, J. (2021). Deep learning air quality forecasting using spatiotemporal data. *Science of the Total Environment*, *754*, 142210. https://doi.org/10.1016/j.scitotenv.2020.142210

Han, K., & Wang, J. (2022). Multivariate lstm for inter-station air pollution prediction. *Science of The Total Environment*, *812*, 152530. https://doi.org/10.1016/j.scitotenv.2021.152530

Han, L., & Zhao, Q. (2022). Informer architecture for multi-site spatio-temporal air quality prediction. *Knowledge-Based Systems*, *239*, 107973. https://doi.org/10.1016/j.knosys.2022.107973

Han, L., Zhou, M., & Zhang, J. (2023). Random forest and gradient boosting for urban air pollution prediction: Performance and interpretability. *Ecological Informatics*, *73*, 101913. https://doi.org/10.1016/j.ecoinf.2022.101913

Han, Y., & Park, J. (2024). Efficient deep learning architectures for urban air quality forecasting. *Atmospheric Pollution Research*, *15*(2), 101921. https://doi.org/10.1016/j.apr.2024.101921

Hoek, G., Krishnan, R. M., Beelen, R., Peters, A., Ostro, B., Brunekreef, B., & Kaufman, J. D. (2013). Long-term air pollution exposure and cardio-respiratory mortality: A review. *Environmental Health*, *12*, 43. https://doi.org/10.1186/1476-069X-12-43

Hu, X., Belle, J. H., Meng, X., Wildani, A., Waller, L. A., Strickland, M. J., & Liu, Y. (2020). Estimating pm2.5 concentrations in the united states using an ensemble learning approach. *Environmental Science & Technology*, *54*(13), 8262–8271. https://doi.org/10.1021/acs.est.0c01764

Hu, X., & Li, Z. (2021). Gradient boosting machines for $pm_{2.5}$ forecasting at monitoring stations. *Atmospheric Pollution Research*, *12*(6), 101066. https://doi.org/10.1016/j.apr.2021.101066

Hu, X., & Zhao, Y. (2024). Advances in chemical transport models for air quality forecasting. *Atmospheric Environment*, *312*, 119987. https://doi.org/10.1016/j.atmosenv.2024.119987

Huang, J., Xu, Y., & Zhang, L. (2021). Ant colony optimization for multi-pollutant time series prediction in urban areas. *Applied Soft Computing*, *108*, 107459. https://doi.org/10.1016/j.asoc.2021.107459

Huang, X., Lee, K., & Fu, J. S. (2022). Transboundary air pollution and biomass burning impacts on southeast asia haze. *Atmospheric Environment*, *268*, 118760. https://doi.org/10.1016/j.atmosenv.2021.118760

Huang, Y., & Chang, H. (2023). Gru model for ozone forecasting in taiwan. *Environmental Research*, *216*, 114674. https://doi.org/10.1016/j.envres.2023.114674

IQAir. (2024). World air quality report 2024 [Accessed: 2025-01-15].

Khan, M. F., Latif, M. T., & Juneng, L. (2020). So2 emissions and respiratory health in southeast asia: Recent findings. *Atmospheric Pollution Research*, *11*(7), 1185–1193. https://doi.org/10.1016/j.apr.2020.03.009

Khaniabadi, Y. O., Goudarzi, G., & Daryanoosh, S. (2022). Vulnerability of children and elderly to pm2.5 exposure: A global review. *Ecotoxicology and Environmental Safety*, *229*, 113079. https://doi.org/10.1016/j.ecoenv.2021.113079

Kim, K. H., Kabir, E., & Kabir, S. (2015). $Pm_{2.5}$ and health risk: A review of epidemiological evidence. *Environmental International*, *74*, 136–145. https://doi.org/10.1016/j.envint.2014.10.005

Le, T., Wang, Y., Liu, L., Yang, J., Yung, Y., Li, G., & Seinfeld, J. (2020). Unexpected air pollution with marked emission reductions during the covid-19 outbreak in china. *Science*, *369*(6504), 702–706. https://doi.org/10.1126/science.abb7431

Lee, J., & Park, C. (2021). Spatial autoregressive models for air quality forecasting. *Atmospheric Environment*, *247*, 118175. https://doi.org/10.1016/j.atmosenv.2021.118175

Lelieveld, J., Pozzer, A., Pöschl, U., Fnais, M., Daiber, A., & Münzel, T. (2023). Loss of life expectancy from air pollution compared to other risk factors: A global health perspective. *Lancet Planetary Health*, *7*(2), e123–e131. https://doi.org/10.1016/S2542-5196(22)00312-5

Li, B., Tan, S., & Rahman, A. (2024). Transformer-based spatio-temporal forecasting of $pm_{2.5}$ using improved dbo optimization. *Environmental Modelling & Software*, *167*, 105762. https://doi.org/10.1016/j.envsoft.2024.105762

Li, J., Zhao, Y., & Wang, H. (2021). Hyperparameter optimization in deep learning models for air quality forecasting: A comparative study. *Environmental Modelling & Software*, *145*, 105159. https://doi.org/10.1016/j.envsoft.2021.105159

Li, X., & et al. (2019). *Problem definitions and evaluation criteria for the cec 2019 special session and competition on single objective real-parameter numerical optimization* (tech. rep.) (Benchmark function suite used for single-objective optimization). IEEE Congress on Evolutionary Computation (CEC). https://cec2019.org/

Li, X., Peng, L., Yao, X., Cui, S., & Hu, Y. (2022). A hybrid deep learning model for spatiotemporal air quality prediction. *Environmental Modelling & Software*, *148*, 105257. https://doi.org/10.1016/j.envsoft.2021.105257

Liang, F., Xiao, Q., Gu, D., Xu, M., Chen, J., Liu, X., & Liu, Y. (2021). A machine learning approach to estimate daily pm2.5 in china using high-resolution satellite data. *Remote Sensing of Environment*, *256*, 112345. https://doi.org/10.1016/j.rse.2021.112345

Liang, X., Zhao, H., & Wang, Y. (2022). Machine learning approaches for single-station air quality forecasting: A case study on $pm_{2.5}$ and no_2. *Atmospheric Pollution Research*, *13*(5), 101356. https://doi.org/10.1016/j.apr.2022.101356

Lim, S., Rahman, A., & Amil, N. (2024). Associations between ambient $pm_{2.5}$ exposure and hospital admissions for respiratory and cardiovascular diseases in northern malaysia. *Environmental Research*, *330*, 120412. https://doi.org/10.1016/j.envres.2024.120412

Liu, H., & Zhang, Y. (2021). Vector autoregression and multivariate deep learning for multi-station air quality prediction. *Environmental Research*, *197*, 111134. https://doi.org/10.1016/j.envres.2021.111134

Liu, Q., & Wang, Z. (2024). Transformer-based deep learning for multi-station air quality forecasting. *Environmental Modelling & Software*, *168*, 105781. https://doi.org/10.1016/j.envsoft.2024.105781

Liu, X., & Yan, J. (2020). Spatio-temporal graph convolutional networks for air quality prediction. *Proceedings of AAAI*, *34*(1), 781–788.

Liu, Y., Li, J., & Xu, J. (2023). Ozone exposure and cardiopulmonary health: Evidence from recent epidemiological studies. *Science of the Total Environment*, *857*, 159354. https://doi.org/10.1016/j.scitotenv.2022.159354

Lu, D., & Wang, C. (2023). Handling missing and irregular air quality data with spatio-temporal imputation. *Science of the Total Environment*, *879*, 163062. https://doi.org/10.1016/j.scitotenv.2023.163062

Lu, F., Xu, J., & Yang, X. (2021). Air pollution and respiratory health: Evidence from cohort studies. *Environmental Pollution*, *287*, 117313. https://doi.org/10.1016/j.envpol.2021.117313

Lu, X., Zhang, L., Wang, T., Gao, M., Li, K., Zhang, Y., & He, H. (2022). Severe surface ozone pollution in china: A global perspective. *Environmental Science & Technology Letters*, *9*(2), 120–126. https://doi.org/10.1021/acs.estlett.1c00850

Ma, Y., & Chen, L. (2021). Support vector regression for air pollution forecasting at local scale. *Science of The Total Environment*, *768*, 144359. https://doi.org/10.1016/j.scitotenv.2020.144359

Ma, Y., Zhang, Q., & Wang, J. (2022). Long-range transport and its contribution to pm2.5 spatial heterogeneity in east asia. *Atmospheric Environment*, *268*, 118763. https://doi.org/10.1016/j.atmosenv.2021.118763

Mirjalili, S. (2019). *Evolutionary algorithms and swarm intelligence: Fundamentals, advances and applications*. CRC Press.

Petetin, H., Bowdalo, D., Soret, A., Guevara, M., Jorba, O., Serradell, K., & Pérez García-Pando, C. (2020). Meteorology-normalized impact of the covid-19 lockdown upon no_2 pollution in spain. *Atmospheric Chemistry and Physics*, *20*(18), 11119–11141. https://doi.org/10.5194/acp-20-11119-2020

Pope, C. A., Ezzati, M., & Dockery, D. W. (2009). Fine-particulate air pollution and life expectancy in the united states. *New England Journal of Medicine*, *360*(4), 376–386. https://doi.org/10.1056/NEJMsa0805646

Querol, X., Massagué, J., & Alastuey, A. (2022). Air quality challenges in europe in the context of climate change and energy transition. *Environmental Science & Policy*, *129*, 1–14. https://doi.org/10.1016/j.envsci.2021.12.002

Rahman, A., & Cheng, Y. (2024). Multi-station fusion of meteorological and satellite data for improved air quality forecasting in malaysia. *Atmospheric Pollution Research*, *15*(4), 102321. https://doi.org/10.1016/j.apr.2024.102321

Rahman, N. A., Amil, N., & Khan, M. F. (2022). Meteorological influences on pm2.5 and no2 in penang, malaysia. *Atmospheric Pollution Research*, *13*(3), 101343. https://doi.org/10.1016/j.apr.2022.101343

Rajé, F., Kumar, P., & Singh, V. (2024). Long-term exposure to nitrogen dioxide and cardiovascular risk: A systematic review. *Atmospheric Environment*, *318*, 119385. https://doi.org/10.1016/j.atmosenv.2024.119385

Seinfeld, J. H., & Pandis, S. N. (2016). *Atmospheric chemistry and physics: From air pollution to climate change* (3rd). Wiley.

Seinfeld, J. H., & Pandis, S. N. (2020). Atmospheric chemistry and physics: From air pollution to climate change (3rd edition). *Wiley*.

Shah, A. S., & et al. (2023). Air pollution and cognitive decline in children: Evidence from epidemiological studies. *Environment International*, *171*, 107655. https://doi.org/10.1016/j.envint.2023.107655

Sharma, A., Kumar, R., & Singh, A. (2021). Hybrid empirical mode decomposition and machine learning approaches for air quality prediction in india. *Environmental Science and Pollution Research*, *28*(34), 46965–46978. https://doi.org/10.1007/s11356-021-14971-9

Sharma, R., & Balasubramanian, R. (2020). Urban air quality and human health: A systematic review. *Environmental Science and Pollution Research*, *27*, 5856–5879. https://doi.org/10.1007/s11356-019-07496-7

Shi, Z., Song, C., Liu, B., Lu, G., Xu, J., Van Vu, T., & Elliott, R. (2021). Abrupt but smaller than expected changes in surface air quality attributable to covid-19 lockdowns. *Science Advances*, *7*(3), eabd6696. https://doi.org/10.1126/sciadv.abd6696

Shin, M., & Lee, H. (2021). Gru-based deep learning for ozone forecasting at local stations. *Atmospheric Environment*, *252*, 118230. https://doi.org/10.1016/j.atmosenv.2021.118230

Sun, S., Cao, W., & Mason, T. G. (2022). Short-term exposure to ambient air pollution and increased susceptibility to infectious diseases. *Environmental Research*, *204*, 112010. https://doi.org/10.1016/j.envres.2021.112010

Sun, Z., & Liu, Y. (2020). Spatial disparities in air pollution across china: A national analysis of pm2.5. *Environmental Research*, *183*, 109163. https://doi.org/10.1016/j.envres.2020.109163

Tan, P., Kannan, P., & Abdullah, S. (2023). Transboundary haze episodes and their health impacts in malaysia: A multi-scale assessment. *Science of The Total Environment*, *857*, 159282. https://doi.org/10.1016/j.scitotenv.2022.159282

van Donkelaar, A., Hammer, M. S., Bindle, L., & Martin, R. V. (2021). Global estimates of fine particulate matter using a combined satellite and chemical transport model approach. *Environmental Science & Technology*, *55*(23), 15526–15534. https://doi.org/10.1021/acs.est.1c05309

Venter, Z. S., Aunan, K., Chowdhury, S., & Lelieveld, J. (2020). Covid-19 lockdowns cause global air pollution declines. *PNAS*, *117*(32), 18984–18990. https://doi.org/10.1073/pnas.2006853117

Wang, J., Song, G., & Li, Y. (2021). Graph convolutional networks for air quality prediction: Capturing spatial dependencies among monitoring stations. *Atmospheric Environment*, *252*, 118285. https://doi.org/10.1016/j.atmosenv.2021.118285

Wang, P., Chen, K., Zhuang, Y., & Li, X. (2019). Air quality forecasting using machine learning approaches: A review. *Applied Sciences*, *9*(19), 4130. https://doi.org/10.3390/app9194130

Wang, Y., Gao, W., Wang, S., Song, T., Gong, Z., Ji, D., & Wang, L. (2022). Temporal patterns and driving factors of air pollutants in chinese cities: Insights from diurnal and weekly cycles. *Atmospheric Environment*, *268*, 118761. https://doi.org/10.1016/j.atmosenv.2021.118761

Wei, J., Huang, J., & Wang, X. (2023). No2 pollution and health impacts in urban china: An updated assessment. *Science of the Total Environment*, *857*, 159511. https://doi.org/10.1016/j.scitotenv.2022.159511

Wei, Y., Wang, Y., & Di, Q. (2021). Short-term exposure to air pollution and cardiovascular hospital admissions. *Nature Communications*, *12*, 6817. https://doi.org/10.1038/s41467-021-27058-x

Wong, C. F., Ee, L. W., & Khan, M. F. (2021). Transboundary haze in southeast asia: Spatiotemporal variation and implications for air quality management. *Atmospheric Pollution Research*, *12*(4), 101030. https://doi.org/10.1016/j.apr.2021.101030

World Health Organization. (2021). Who global air quality guidelines: Particulate matter ($pm_{2.5}$ and pm_{10}), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide. https://www.who.int/publications/i/item/9789240034228

Wu, J., & Chen, L. (2024). Machine learning frameworks for air quality prediction: A review of recent progress. *Ecological Informatics*, *76*, 102119. https://doi.org/10.1016/j.ecoinf.2024.102119

Wu, J., & Zhang, K. (2023). St-transformer for multi-station $pm_{2.5}$ forecasting. *Neural Networks*, *165*, 579–592. https://doi.org/10.1016/j.neunet.2023.06.011

Wu, X., Chen, L., & Zhang, K. (2022). Bayesian optimization for deep learning-based air quality forecasting models. *Atmospheric Environment*, *277*, 119076. https://doi.org/10.1016/j.atmosenv.2022.119076

Wu, Y., & Li, X. (2020). Graph convolutional networks for multi-station $pm_{2.5}$ forecasting. *IEEE Access*, *8*, 145246–145257. https://doi.org/10.1109/ACCESS.2020.3015963

Yang, L., & Zhou, J. (2022). Deep learning models for single-station air pollution forecasting. *Journal of Cleaner Production*, *356*, 131829. https://doi.org/10.1016/j.jclepro.2022.131829

Zhang, H., Chen, Y., & Wang, X. (2022). Advances in deep learning approaches for air pollution forecasting: A critical review. *Environmental Pollution*, *300*, 118922. https://doi.org/10.1016/j.envpol.2022.118922

Zhang, K., & Liu, M. (2023). Graph transformer networks for spatio-temporal air quality forecasting. *Neural Networks*, *165*, 589–604. https://doi.org/10.1016/j.neunet.2023.06.013

Zhang, T., & Li, J. (2024). Improved dung beetle optimization for hyperparameter tuning in deep spatio-temporal models. *Applied Soft Computing*, *152*, 110086. https://doi.org/10.1016/j.asoc.2024.110086

Zhang, Y., Bocquet, M., Mallet, V., Seigneur, C., & Baklanov, A. (2012). Real-time air quality forecasting, part i: History, techniques, and current status. *Atmospheric Environment*, *60*, 632–655. https://doi.org/10.1016/j.atmosenv.2012.06.040

Zhang, Y., Li, X., & Wang, Z. (2021). Geographically weighted regression for pm2.5 in china: Spatiotemporal patterns and drivers. *Science of the Total Environment*, *756*, 143936. https://doi.org/10.1016/j.scitotenv.2020.143936

Zhang, Y., Xu, X., Wang, Y., & Chen, D. (2023). Air quality forecasting using ceemdan and cnn-lstm hybrid models in southeast asia. *Environmental Pollution*, *319*, 120942. https://doi.org/10.1016/j.envpol.2022.120942

Zhang, Y., Zheng, Y., & Qi, Y. (2020). Deepair: A hybrid spatiotemporal deep learning framework for air quality prediction. *Proceedings of the 26th ACM SIGKDD*, 2560–2568. https://doi.org/10.1145/3394486.3403286

Zhao, Y., & Li, Q. (2023). Hybrid gcn–lstm model for spatio-temporal air quality forecasting across multiple stations. *Atmospheric Environment*, *291*, 119475. https://doi.org/10.1016/j.atmosenv.2023.119475

Zhou, F., & Chen, M. (2022). Multi-station air quality forecasting using graph convolutional recurrent networks. *Ecological Informatics*, *67*, 101523. https://doi.org/10.1016/j.ecoinf.2021.101523

Zhou, Y., Li, J., Zhang, K., & Wang, X. (2023). Spatiotemporal graph convolutional networks for multi-station pm2.5 forecasting in the beijing–tianjin–hebei region. *Journal of Cleaner Production*, *383*, 135375. https://doi.org/10.1016/j.jclepro.2022.135375