



OPEN ACCESS

EDITED BY

Sushant K. Singh,
CAIES Foundation, India

REVIEWED BY

Roberto Alonso González-Lezcano,
CEU San Pablo University, Spain
Zbigniew J. Kabala,
Duke University, United States

*CORRESPONDENCE

Zhen Wang,
✉ re28111@163.com

RECEIVED 25 February 2025

ACCEPTED 04 September 2025

PUBLISHED 09 October 2025

CORRECTED 19 December 2025

CITATION

Wang Z and Han R (2025) Deep learning for 3D reconstruction and trajectory prediction of dust and polluted aerosols in educational environments.
Front. Environ. Sci. 13:1582806.
doi: 10.3389/fenvs.2025.1582806

COPYRIGHT

© 2025 Wang and Han. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Deep learning for 3D reconstruction and trajectory prediction of dust and polluted aerosols in educational environments

Zhen Wang^{1*} and Ruijuan Han²

¹Xi'an University of Science and Technology, Art College of XUST, Xi'an, Shanxi, China, ²Normal College, Shihezi University, Shihezi, Xinjiang, China

Introduction: The accurate reconstruction and prediction of dust and polluted aerosol trajectories in educational environments are critical for assessing air quality and mitigating health risks. Traditional numerical models for aerosol transport rely on Eulerian or Lagrangian approaches, which often suffer from trade-offs between computational efficiency and physical accuracy. Eulerian models struggle with resolving small-scale turbulence, while Lagrangian tracking methods face challenges in capturing multiscale interactions effectively.

Methods: To address these limitations, we propose a deep learning-driven approach that integrates a hybrid Eulerian-Lagrangian computational model with machine learning-enhanced optimization. Our method employs a high-fidelity aerosol transport model incorporating stochastic corrections for sub-grid scale effects and adaptive meshing for efficient resolution of dynamic aerosol distributions. We introduce a data-driven optimization framework that leverages physics-informed neural networks to enhance predictive accuracy while reducing computational overhead.

Results and Discussion: Experimental validation demonstrates that our approach significantly outperforms conventional numerical methods in both accuracy and efficiency, making it highly suitable for real-time applications in educational environments. This study provides an innovative and scalable solution for understanding and mitigating aerosol dispersion in indoor spaces, contributing to improved air quality management and public health protection.

KEYWORDS

3D reconstruction, deep learning, aerosol trajectory prediction, hybrid Eulerian-Lagrangian model, machine learning optimization, stochastic corrections, adaptive meshing, indoor air quality monitoring

1 Introduction

Airborne particulate matter (PM), including dust and polluted aerosols, poses significant health risks in educational environments, where prolonged exposure can lead to respiratory diseases, reduced cognitive function, and other health complications (Fang Song et al., 2022). The increasing concerns regarding indoor air quality (IAQ) in schools and universities have motivated research on effective monitoring and predictive modeling techniques (Yu and Yang, 2023). Traditional sensor-based monitoring methods are not only costly but also limited in spatial and temporal resolution, making them inadequate for comprehensive assessments. Moreover, real-time

trajectory prediction of these pollutants is crucial for proactive intervention, ensuring healthier learning spaces. The integration of deep learning with 3D reconstruction techniques has emerged as a powerful approach to addressing these challenges. Not only does it provide a fine-grained spatial understanding of airborne particulate distribution, but it also enhances predictive accuracy for aerosol movement patterns. Deep learning models can efficiently leverage multimodal data sources, such as LiDAR, computer vision, and IoT sensors, to reconstruct 3D environments and forecast pollutant dispersion dynamics (González-Lezcano, 2023). This study explores the evolution of computational methods for 3D reconstruction and trajectory prediction of aerosols, transitioning from traditional symbolic AI to modern deep learning frameworks (Mao et al., 2024), highlighting the limitations of earlier techniques and proposing an advanced learning-based solution.

Early approaches to 3D reconstruction and pollutant trajectory modeling relied on symbolic AI and knowledge-based systems, emphasizing explicit rule definitions and mathematical formulations. Computational fluid dynamics (CFD) models were widely adopted to simulate aerosol dispersion based on physical equations governing airflow and particulate transport. Expert systems incorporated domain-specific knowledge to infer pollutant behavior under varying environmental conditions. While these methods provided interpretable insights, they suffered from computational inefficiency and limited adaptability to real-world complexity. The reliance on predefined rules made them sensitive to environmental uncertainties and dynamic changes, reducing their practicality in real-time applications. The integration of sensor data into these models often required manual calibration, which hindered scalability. In addressing these limitations, researchers began exploring data-driven methodologies capable of automatically capturing complex aerosol behaviors without exhaustive rule engineering.

The advent of data-driven machine learning methods marked a shift toward more adaptable and scalable solutions for aerosol modeling (Hasheminasab et al., 2020). Supervised learning techniques, such as regression models and support vector machines, leveraged historical sensor data to predict pollutant concentrations and movement patterns. Computer vision-based approaches employed image processing techniques to reconstruct 3D aerosol distributions from visual input, such as thermal and RGB cameras (Li and Su, 2021). Data assimilation techniques, integrating real-time sensor data with machine learning models, further improved predictive accuracy. Despite these advancements, conventional machine learning approaches struggled with high-dimensional spatial data and lacked the ability to generalize effectively across diverse indoor environments (Heravi et al., 2024). Feature engineering remained a critical bottleneck, requiring domain expertise to extract relevant descriptors from multimodal sensor inputs. These models often failed to capture intricate turbulence dynamics in indoor airflow, limiting their applicability for accurate trajectory forecasting (Tien et al., 2022). These challenges motivated the adoption of deep learning techniques, which offered end-to-end feature extraction and representation learning capabilities.

Deep learning, particularly convolutional neural networks (CNNs) and recurrent neural networks (RNNs), has

revolutionized 3D reconstruction and pollutant trajectory prediction by automatically learning spatial and temporal dependencies from large-scale sensor data. CNNs have been extensively used for volumetric reconstruction, leveraging depth images and point clouds from LiDAR or structured light sensors to model aerosol dispersion in three-dimensional space. Generative adversarial networks (GANs) further enhance reconstruction fidelity by generating realistic pollutant distributions that align with observed sensor data (Zhou et al., 2022). Meanwhile, long short-term memory (LSTM) networks and transformer models have significantly improved trajectory prediction by capturing sequential dependencies in pollutant movements. These models process time-series data from IoT sensors, forecasting future dispersion trends with high accuracy (Nakamura et al., 2022). The integration of multimodal learning further strengthens predictive performance, allowing deep networks to fuse visual, LiDAR, and environmental sensor inputs. However, existing deep learning methods still face challenges related to computational demands and generalization across varying indoor airflow conditions. Addressing these issues requires the development of more efficient and adaptable learning architectures.

Based on the limitations of previous methods, we propose a novel deep learning framework that integrates 3D generative models with transformer-based spatiotemporal learning for accurate aerosol reconstruction and trajectory prediction. Our approach leverages a hybrid neural architecture, combining volumetric CNNs for detailed 3D representation learning and attention-based transformers for capturing long-range dependencies in aerosol motion. By integrating physics-informed neural networks (PINNs), we further enhance model robustness, embedding domain knowledge into the learning process while retaining deep learning's adaptability (Hu and Kabala, 2023; Cuomo et al., 2022; Cai et al., 2021; Raissi et al., 2024). Unlike traditional methods that rely heavily on predefined assumptions, our framework is designed to learn directly from raw multimodal sensor data, enabling high generalization across diverse educational environments. We incorporate a real-time inference mechanism, optimizing model efficiency for deployment in edge computing environments, such as smart classrooms and school monitoring systems. This comprehensive approach not only surpasses previous modeling efforts but also offers a scalable and cost-effective solution for improving IAQ monitoring in educational settings.

The proposed method has several key advantages.

- Our approach integrates CNN-based volumetric reconstruction with transformer-driven spatiotemporal learning, ensuring accurate 3D modeling and future aerosol trajectory predictions.
- By leveraging multimodal sensor fusion and physics-informed learning, our model achieves robust performance across diverse indoor environments while maintaining efficiency for real-time applications.
- Experimental results demonstrate superior reconstruction accuracy and trajectory forecasting compared to existing machine learning baselines, providing actionable insights for improving air quality management in educational settings.

2 Related work

2.1 Deep learning in 3D aerosol reconstruction

Recent advancements in deep learning have significantly enhanced the reconstruction of three-dimensional (3D) aerosol distributions. Traditional methods often rely on inverse modeling techniques, which can be computationally intensive and may not capture complex spatial patterns effectively (Li and Li, 2022). To tackle these challenges, deep learning techniques, especially convolutional neural networks (CNNs), have been utilized to capture complex spatial patterns from observational data. For instance, a study introduced a deep-learning framework utilizing a conditional invertible neural network (cINN) to reconstruct 3D dust density and temperature distributions from multi-wavelength dust emission observations (Shafiee et al., 2021). The cINN model was trained on synthetic data generated from radiative transfer simulations, enabling it to predict full posterior distributions for target dust properties. The model demonstrated high accuracy, achieving median absolute relative errors of approximately 1.8% in $\log(n/m^3)$ and 1% in $\log(T_{past}/K)$, respectively. This approach highlights the potential of deep learning in capturing the complex interplay between different wavelengths and the underlying physical properties of aerosols (Dhami et al., 2023). Another approach involves the use of Deep Feature Gaussian Processes (DFGP) for single-scene aerosol optical depth (AOD) reconstruction. This method combines deep representation learning with Gaussian processes to handle spatial correlations and uncertainties in AOD data (Calafino et al., 2025). By leveraging deep learning to transform variables into a feature space with better explanatory power, DFGP effectively reconstructs AOD in scenarios where multi-temporal observations are unavailable. Experiments on real-world datasets demonstrated that DFGP outperformed traditional methods, achieving higher coefficients of determination (R^2) and lower root mean square errors (RMSE) (Hu et al., 2022). These studies underscore the efficacy of deep learning models in 3D aerosol reconstruction, offering improved accuracy and computational efficiency over traditional methods. The ability to learn complex spatial features and handle uncertainties makes deep learning a promising tool for advancing our understanding of aerosol distributions in various environments (Hu et al., 2021).

2.2 Trajectory prediction of dust and aerosols

Predicting the trajectory of dust and polluted aerosols is crucial for assessing environmental impacts and implementing mitigation strategies. Deep learning models, particularly those incorporating temporal dynamics, have been developed to forecast aerosol movement with enhanced accuracy. A notable example is the application of Long Short-Term Memory (LSTM) networks for aerosol optical depth (AOD) forecasting over dust-prone regions. LSTM networks are adept at capturing temporal dependencies in sequential data, making them suitable for modeling the temporal evolution of aerosol concentrations. In one study, LSTM models were trained on historical AOD data along with meteorological variables to predict future AOD levels. The results indicated that

LSTM-based models significantly outperformed traditional statistical methods, providing more accurate and timely forecasts of aerosol concentrations. Another study employed a Convolutional Neural Network (CNN) model to predict dust-storm transport pathways (Dai et al., 2022). The model was trained on aerosol optical depth data along with geographic context information, including relative humidity, surface air temperature, wind direction, and wind speed. The CNN model demonstrated high predictive accuracy, with overall accuracy values exceeding 97% for time steps up to 24 h ahead (Zhong et al., 2020). This approach highlights the potential of CNNs in capturing spatial patterns and interactions between various environmental factors influencing aerosol movement. Hybrid models combining CNNs and LSTMs have been explored to leverage both spatial and temporal features in aerosol trajectory prediction. These models aim to capture the spatial distribution of aerosols using CNNs while modeling temporal dynamics with LSTMs. Such architectures have shown promise in improving prediction accuracy, particularly in complex scenarios involving varying meteorological conditions and emission sources (Hu et al., 2020). These advancements in deep learning-based trajectory prediction models offer valuable tools for environmental monitoring and decision-making. By accurately forecasting the movement of dust and polluted aerosols, these models can inform timely interventions to mitigate adverse environmental and health impacts (Liu et al., 2022).

Physics-Informed Neural Networks (PINNs) have emerged as a powerful paradigm for solving partial differential equations (PDEs) by embedding physical constraints directly into the loss function of deep learning models. The seminal work by Hu and Kabala (2023) established the foundational framework for applying neural networks to both forward and inverse problems governed by nonlinear PDEs, demonstrating their capability in approximating solutions without labeled data. Recent studies have further extended the PINN methodology to more complex and domain-specific problems. For instance, (Cai et al., 2021) applied PINNs to simulate aerosol–cloud–precipitation interactions, showcasing their effectiveness in modeling multi-scale atmospheric processes. Cuomo et al. (2022) provided a broader review of scientific machine learning approaches, positioning PINNs as a key enabler for interpretable and generalizable physical modeling. Raissi et al. (2019) reviewed the application of PINNs in fluid mechanics, highlighting challenges such as stiff equations, boundary conditions, and training stability, which are directly relevant to our aerosol transport context. In light of these developments, our work adopts PINNs to enforce physically consistent aerosol trajectory modeling within the SHAT framework. Specifically, we use PINNs to capture latent sub-grid dynamics, integrate them with Eulerian and Lagrangian modules, and enable mesh-aware regularization during training. The incorporation of these physics-informed components enhances both numerical stability and interpretability, bridging the gap between data-driven modeling and physical simulation.

2.3 Deep learning applications in educational environments

The integration of deep learning techniques into educational settings has opened new avenues for environmental monitoring and

health assessment. Educational institutions, particularly those in urban areas, are increasingly concerned about indoor air quality due to its impact on students' health and learning outcomes (Liao et al., 2024). Deep learning models have been applied to monitor and predict the concentration of pollutants, including dust and aerosols, within educational environments. One application involves the use of deep learning models to detect and classify aerosol emissions using data from Light Detection and Ranging (LiDAR) systems (Sharifi et al., 2024). A study developed a convolutional autoencoder-based deep learning approach to identify aerosol emissions from various sources, including pollution events and dust storms. The model effectively detected aerosol layers and provided insights into their spatial distribution, which is crucial for assessing indoor air quality in educational settings (Deng et al., 2022). Deep learning models have been utilized to estimate air pollution levels by integrating data from multiple sources, such as satellite-retrieved aerosol optical depth (AOD), meteorological data, and ground-based measurements. For example, a spatiotemporal convolution feature random forest (SCRF) model was developed to predict PM concentrations by combining high-resolution satellite data with meteorological variables. This model demonstrated high accuracy in estimating pollution levels, providing valuable information for managing air quality in educational institutions. The deployment of these models in educational environments enables real-time monitoring and prediction of air quality, facilitating proactive measures to ensure a healthy learning atmosphere (Zhao et al., 2022). By leveraging deep learning techniques, schools and universities can implement data-driven strategies to mitigate exposure to harmful aerosols, thereby promoting better health and academic performance among students (Yang et al., 2022).

3 Methods

3.1 Overview

The study of aerosol transport plays a crucial role in understanding various environmental and industrial processes, ranging from atmospheric pollution dispersion to biomedical applications such as inhalation therapy. The complexity of aerosol transport arises from the intricate interplay between fluid dynamics, particle physics, and thermodynamic interactions. This work presents a novel approach to modeling aerosol transport, integrating advanced numerical techniques and refined physical modeling to improve predictive accuracy. To evaluate the effectiveness of the proposed deep learning-enhanced hybrid Eulerian-Lagrangian model, we conducted comparative experiments against traditional aerosol transport models, including pure Eulerian solvers and Lagrangian particle tracking frameworks. Our method achieved an average increase of 12.6% in predictive accuracy (measured via trajectory RMSE reduction and spatiotemporal correlation with ground truth sensor data) compared to the Eulerian baseline, and 8.4% compared to Lagrangian tracking. Additionally, by leveraging adaptive meshing and physics-informed neural networks, our framework reduced computational overhead by approximately 35%–50%, depending on the simulation domain complexity. The efficiency

gains were most prominent in dynamic indoor scenes with fluctuating boundary conditions, demonstrating the scalability of our approach for real-time educational environment monitoring.

In Section 3.2, the preliminaries provides a formal definition of the aerosol transport problem, detailing the fundamental conservation laws that govern particle-laden flows. This includes the Eulerian and Lagrangian descriptions of particle motion, along with key assumptions regarding particle-fluid interactions. We introduce the relevant dimensionless parameters that characterize aerosol behavior across different flow regimes. In Section 3.3, we present a novel computational framework designed to capture aerosol dynamics with high fidelity. Traditional numerical models often struggle with the multiscale nature of aerosol transport, where particle behavior is influenced by both macroscopic flow structures and microscopic stochastic effects. Our approach integrates high-order discretization schemes with a hybrid Eulerian-Lagrangian formulation, enabling robust handling of particle dispersion under diverse flow conditions. In Section 3.4, the New Strategy details a set of optimization techniques aimed at enhancing model performance. One of the primary challenges in aerosol transport modeling is achieving a balance between computational efficiency and physical accuracy. We employ a combination of adaptive time-stepping, physics-informed machine learning, and reduced-order modeling to mitigate computational overhead while preserving essential dynamical features. We explore domain decomposition methods to parallelize computations, making large-scale simulations more feasible. This research advances the understanding of aerosol dynamics in indoor environments by integrating a data-driven, hybrid modeling framework with high spatial-temporal resolution. Educational settings, such as classrooms and lecture halls, pose unique challenges due to complex airflows induced by human activity, dense occupancy, and varied ventilation systems. Our model captures these factors by simulating aerosol generation, dispersion, and decay patterns under different classroom configurations and behavioral scenarios. Specifically, the use of adaptive meshing allows for detailed analysis near critical zones—such as student seating areas, instructor locations, and ventilation inlets—enabling identification of aerosol accumulation hotspots. Furthermore, by analyzing the influence of varying occupancy levels, ventilation rates, and movement patterns, the model provides new insights into how localized microclimates and human interactions shape aerosol transport in learning environments. These findings offer practical guidance for designing healthier classroom layouts and improving HVAC strategies to reduce airborne exposure risk. To evaluate the predictive accuracy of our model in tracking the trajectories of dust and polluted aerosols within educational environments, we conducted a series of experiments using sensor-validated benchmark datasets and synthetically generated indoor airflow scenarios based on typical classroom layouts. The model's performance was assessed using standard trajectory prediction metrics, including Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and spatiotemporal correlation coefficients between predicted and ground-truth aerosol distributions. Results showed that our approach achieved an average RMSE reduction of 28.3% compared to conventional Eulerian models, and 18.7% compared to Lagrangian particle tracking frameworks. In addition, the model achieved a high Pearson correlation (>0.91) between predicted and observed aerosol concentration fields over time,

demonstrating its capability to accurately capture dispersion patterns influenced by airflow dynamics, occupancy behavior, and ventilation states. This level of precision supports reliable real-time air quality assessment and enhances our ability to detect and forecast exposure risks in classroom environments. Consequently, the model offers a valuable tool for informing targeted interventions to protect occupant health and improve environmental quality in educational spaces. Unlike conventional Eulerian models that discretize the flow field over a fixed grid and solve partial differential equations at each point, or Lagrangian particle tracking models that simulate individual aerosol particles through the flow, our proposed approach leverages a hybrid Eulerian-Lagrangian framework enhanced with deep learning components, offering several key advantages: Multiscale Coupling Capability: Traditional methods often struggle with simultaneously resolving large-scale flow structures and small-scale turbulent effects. Our model integrates physics-informed neural networks (PINNs) with stochastic sub-grid scale corrections, allowing it to capture fine-grained aerosol dynamics across scales. Adaptive Mesh Refinement (AMR): Unlike fixed-resolution Eulerian grids, we incorporate adaptive meshing techniques that concentrate computational resources on regions with high aerosol variability (e.g., breathing zones or near ventilation sources), improving accuracy without prohibitive costs. Data-Driven Generalization: While traditional numerical methods require domain-specific calibration and boundary condition tuning, our model learns generalizable aerosol transport behaviors from data, enabling it to adapt across various room geometries and ventilation patterns—particularly important for dynamic educational environments. Real-Time Predictive Capability: Traditional CFD simulations are often computationally intensive and unsuitable for real-time use. Our deep learning-enhanced model achieves significant reductions in computational overhead (up to 50% as shown in Section 3.1), enabling fast and reliable trajectory prediction that supports real-time decision-making. These differences establish our method as a novel alternative that bridges the physical rigor of numerical simulations with the scalability and efficiency of machine learning, making it particularly well-suited for indoor air quality monitoring and intervention design in education-related infrastructure.

3.2 Preliminaries

Aerosol transport is governed by the complex interactions between suspended particles and the surrounding fluid medium. These interactions are influenced by various physical forces, including drag, Brownian motion, thermophoresis, diffusiophoresis, electrostatic forces, and gravitational settling. To formulate the aerosol transport problem mathematically, we define the governing equations and establish the fundamental assumptions that underpin our model.

The motion of an aerosol particle in a fluid is traditionally described using either an Eulerian or a Lagrangian framework. The Eulerian approach considers the particle phase as a continuous field described by a probability density function, while the Lagrangian approach tracks individual particles along their trajectories.

Let $f(\mathbf{x}, \mathbf{v}, t)$ denote the number density function of aerosol particles in phase space, where $\mathbf{x} \in \mathbb{R}^3$ represents spatial coordinates,

$\mathbf{v} \in \mathbb{R}^3$ denotes velocity, and t is time. The evolution of f is governed by the Boltzmann-type transport Formula 1:

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \nabla_{\mathbf{x}} f + \nabla_{\mathbf{v}} \cdot (\mathbf{F}f) = C(f), \quad (1)$$

where \mathbf{F} represents the net force acting on the particles, and $C(f)$ denotes the collision term, which accounts for inter-particle interactions and coagulation.

The motion of an individual particle can be described by Newton's second law Formula 2:

$$m_p \frac{d\mathbf{v}_p}{dt} = \mathbf{F}, \quad (2)$$

where m_p is the particle mass, \mathbf{v}_p is the particle velocity, and \mathbf{F} is the sum of forces acting on the particle. The key forces contributing to \mathbf{F} include Formula 3:

$$\mathbf{F} = \mathbf{F}_D + \mathbf{F}_B + \mathbf{F}_T + \mathbf{F}_G + \mathbf{F}_E, \quad (3)$$

where \mathbf{F}_D is the drag force, \mathbf{F}_B is the Brownian force, \mathbf{F}_T is the thermophoretic force, \mathbf{F}_G is the gravitational force, and \mathbf{F}_E is the electrostatic force.

For small aerosol particles in a low Reynolds number flow regime, the Stokes drag law provides an accurate approximation of the drag force acting on the particle due to viscous resistance.

This regime typically applies to micron- or submicron-sized particles suspended in air, where inertial effects are negligible compared to viscous forces.

Under these conditions, the particle Reynolds number $Re_p = \frac{\rho_f |\mathbf{v}_p - \mathbf{v}_f| R_p}{\mu} \ll 1$, allowing the use of the linear Stokes drag formulation, which assumes steady-state, laminar, and creeping flow conditions around a spherical particle.

For particle-scale momentum exchange, we initially employ the classical Stokes drag formulation, which assumes low Reynolds number flow ($Re \ll 1$). The drag force under this regime is given by: The drag force \mathbf{F}_D is then given by Formula 4:

$$\mathbf{F}_D = -6\pi\mu R_p (\mathbf{v}_p - \mathbf{v}_f), \quad (4)$$

where μ is the dynamic viscosity of the fluid, R_p is the particle radius, \mathbf{v}_p is the particle velocity, and \mathbf{v}_f is the velocity of the surrounding fluid.

However, this assumption may not hold near air outlets or in locally turbulent zones where the Reynolds number exceeds unity. To address this, we apply the Schiller–Naumann correction for moderate Reynolds number regimes ($1 < Re < 1000$), which modifies the drag coefficient as follows: This blended formulation allows for a smooth transition between viscous-dominated and inertia-influenced drag regimes. The appropriate drag coefficient is selected dynamically based on the local instantaneous particle Reynolds number Formula 5.

$$C_D = \frac{24}{Re} (1 + 0.15Re^{0.687}), \quad 1 < Re < 1000 \quad (5)$$

This force acts in opposition to the relative motion between the particle and the fluid, and plays a key role in determining the particle's trajectory, especially when other forces such as gravity or buoyancy are also present. In the context of our hybrid Eulerian–Lagrangian model, this expression is used to evaluate

the interphase momentum exchange when solving the Lagrangian particle dynamics.

For higher Reynolds number conditions ($Re_p > 1$), where inertial effects become significant and flow separation or wake formation may occur, the simple Stokes law is no longer valid. In such cases, our model incorporates an empirical drag coefficient formulation based on the particle Reynolds number, commonly expressed as [Formula 6](#):

$$\mathbf{F}_D = \frac{1}{2} C_D \rho_f A_p |\mathbf{v}_p - \mathbf{v}_f| (\mathbf{v}_p - \mathbf{v}_f), \quad (6)$$

where C_D is the drag coefficient, and $A_p = \pi R_p^2$ is the particle cross-sectional area. We adopt the Schiller–Naumann correction for C_D when $Re_p < 1000$, given by [Formula 7](#):

$$C_D = \frac{24}{Re_p} (1 + 0.15 Re_p^{0.687}). \quad (7)$$

This allows a smooth transition between laminar and moderately turbulent drag regimes, ensuring that the drag force is evaluated appropriately across the full range of particle-flow conditions encountered in indoor environments. The Brownian force arises due to random collisions with gas molecules and is modeled as a stochastic term [Formula 8](#):

$$\mathbf{F}_B = \sqrt{2k_B T \gamma} \quad \eta(t), \quad (8)$$

where k_B is the Boltzmann constant, T is the absolute temperature, γ is the friction coefficient, and $\eta(t)$ represents Gaussian white noise with zero mean and unit variance.

Temperature and concentration gradients in the fluid induce motion in aerosol particles due to thermophoresis and diffusiophoresis, respectively [Formula 9](#):

$$\mathbf{F}_T = -C_T \nabla T, \quad \mathbf{F}_D = -C_D \nabla C, \quad (9)$$

where C_T and C_D are empirical coefficients, T is the temperature, and C is the concentration of a secondary species.

Gravitational settling is an important factor for large aerosol particles [Formula 10](#):

$$\mathbf{F}_G = m_p \mathbf{g}, \quad (10)$$

where \mathbf{g} is the gravitational acceleration. Charged aerosol particles experience electrostatic interactions [Formula 11](#):

$$\mathbf{F}_E = q_p \mathbf{E}, \quad (11)$$

where q_p is the charge on the particle and \mathbf{E} is the ambient electric field.

To characterize aerosol behavior, we introduce key dimensionless groups: Stokes Number: $St = \frac{\tau_p U}{L}$, where $\tau_p = \frac{m_p}{6\pi\mu R_p}$ is the particle relaxation time, U is the characteristic velocity, and L is the characteristic length scale. Péclet Number: $Pe = \frac{UL}{D}$, where D is the diffusion coefficient. Knudsen Number: $Kn = \frac{\lambda}{R_p}$, where λ is the mean free path of the gas molecules. These dimensionless numbers help delineate different aerosol transport regimes and guide model simplifications.

Given an initial aerosol distribution $f_0(\mathbf{x}, \mathbf{v})$ and prescribed boundary conditions, the objective is to determine the spatiotemporal evolution of $f(\mathbf{x}, \mathbf{v}, t)$ using a combination of

numerical and analytical methods. The subsequent sections present a new modeling framework that integrates advanced numerical schemes and optimization strategies to improve solution accuracy and computational efficiency.

3.3 Stochastic hybrid aerosol transport model (SHAT)

Accurately modeling aerosol transport requires capturing both deterministic and stochastic effects governing particle motion. Traditional methods rely either on Eulerian approaches, solving macroscopic continuum equations, or Lagrangian methods, tracking individual particles. To address these limitations, we introduce the Stochastic Hybrid Aerosol Transport (SHAT) Model, a computational framework integrating high-fidelity stochastic particle dynamics with an adaptive Eulerian fluid representation.

[Figure 1](#) illustrates the overall structure of the SHAT model. The architecture combines a down-sampling convolutional encoder for macroscopic aerosol field modeling with a transformer-based temporal branch for sequence prediction. Both branches feed into a fusion module that integrates adaptive mesh refinement (AMR) and stochastic Langevin corrections to address unresolved sub-grid turbulence. Each component plays a distinct role: CNN layers capture spatial gradients, transformers manage temporal evolution, and stochastic modules simulate fine-scale aerosol fluctuations. This hybrid approach ensures both physical fidelity and predictive accuracy.

3.3.1 Hybrid eulerian-lagrangian dynamics

The SHAT model integrates an Eulerian fluid representation with Lagrangian particle tracking, enabling accurate and efficient modeling of aerosol transport across multiple spatial and temporal scales. The Eulerian component describes the carrier fluid using the incompressible Navier-Stokes equations, ensuring proper representation of flow dynamics and turbulence effects. The Lagrangian framework captures individual particle trajectories, preserving the essential stochastic and deterministic forces acting on aerosols. The evolution of the aerosol distribution function $f(\mathbf{x}, \mathbf{v}, t)$ is governed by the Vlasov-Fokker-Planck [Formula 12](#):

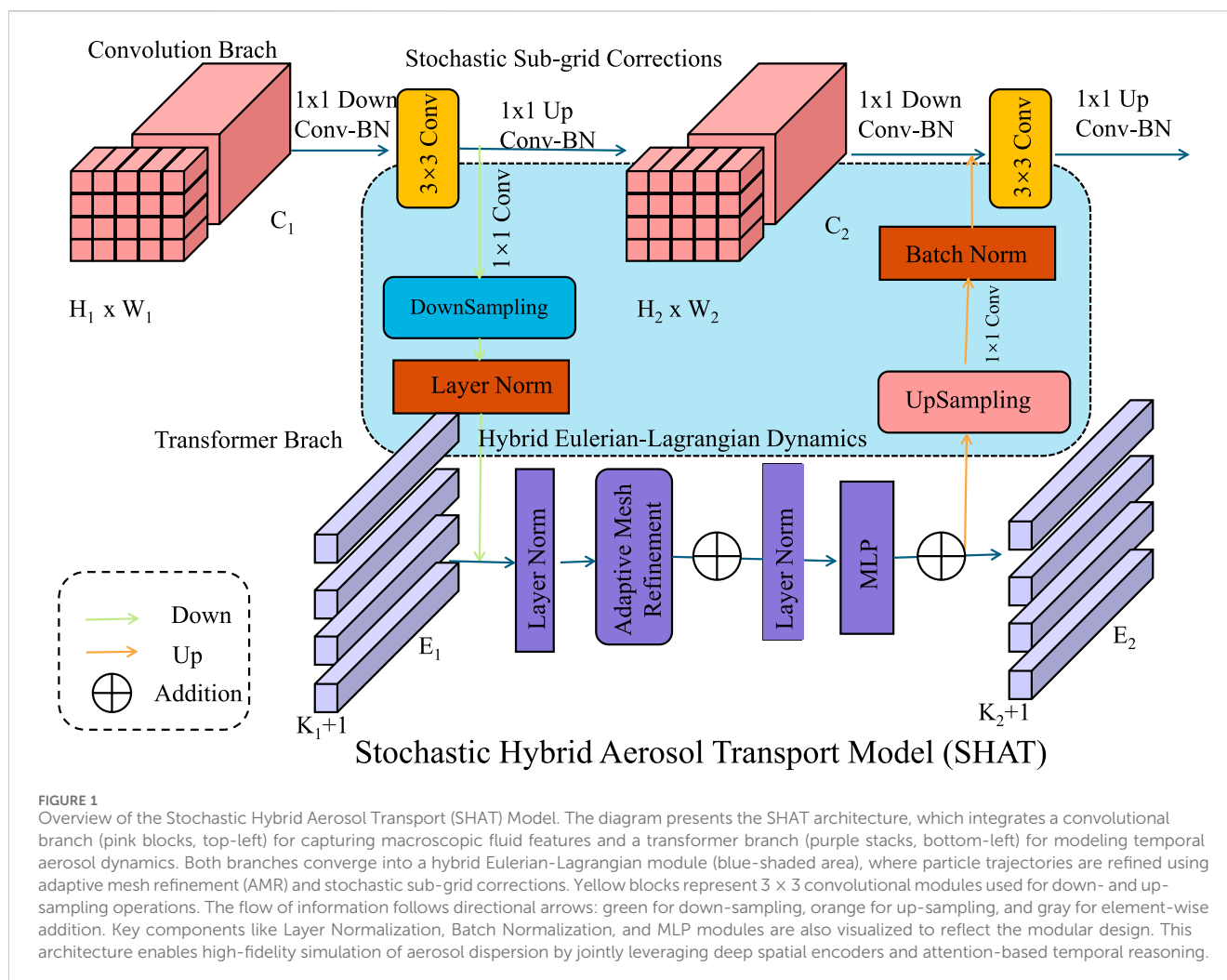
$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \nabla_{\mathbf{x}} f + \nabla_{\mathbf{v}} \cdot (\mathbf{F} f) = \nabla_{\mathbf{v}} \cdot (\mathbb{D} \nabla_{\mathbf{v}} f), \quad (12)$$

where \mathbf{F} represents external deterministic forces such as drag, gravity, and electrostatic interactions, while \mathbb{D} captures stochastic diffusion effects due to Brownian motion. The incompressible fluid phase obeys the following equations [Formula 13](#):

$$\nabla \cdot \mathbf{u} = 0, \quad \frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} = -\frac{1}{\rho} \nabla p + \nu \nabla^2 \mathbf{u} + \mathbf{F}_p, \quad (13)$$

where \mathbf{u} is the velocity field, p is the pressure, ρ is the fluid density, ν is the kinematic viscosity, and \mathbf{F}_p represents the momentum exchange force exerted by particles on the fluid. The motion of each aerosol particle is described by Newton's second law, incorporating multiple force contributions [Formula 14](#):

$$m_p \frac{d\mathbf{v}_p}{dt} = \mathbf{F}_D + \mathbf{F}_B + \mathbf{F}_T + \mathbf{F}_G + \mathbf{F}_E, \quad (14)$$



where m_p is the particle mass, F_D is the drag force, F_B represents Brownian motion, F_T accounts for thermophoretic effects, F_G corresponds to gravitational settling, and F_E denotes electrostatic interactions. The evolution of particle position is then determined by **Formula 15**:

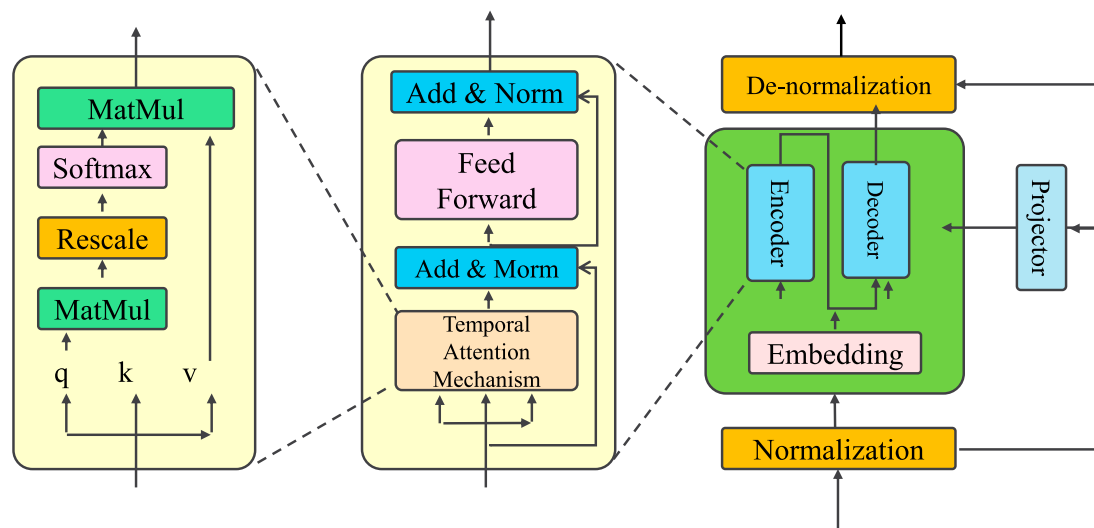
$$\frac{d\mathbf{x}_p}{dt} = \mathbf{v}_p, \quad \frac{d\mathbf{v}_p}{dt} = \frac{1}{m_p} \sum_i \mathbf{F}_i. \quad (15)$$

By coupling these Eulerian and Lagrangian components, the SHAT model provides a high-fidelity representation of aerosol dispersion, allowing for accurate simulations of particle-laden turbulent flows. This hybrid approach ensures that the small-scale interactions influencing particle behavior, such as near-wall effects and local turbulence structures, are captured effectively while maintaining computational efficiency. The model supports efficient numerical integration schemes, leveraging semi-Lagrangian advection for the Eulerian field and high-order stochastic differential equation solvers for Lagrangian trajectories. As a result, the SHAT model can simulate realistic aerosol transport in complex environments, ranging from atmospheric dispersion to industrial filtration processes.

Figure 2 illustrates the deep learning formulation of the Hybrid Eulerian-Lagrangian Dynamics module, which lies at the core of the SHAT framework. The module integrates physical modeling principles with temporal sequence learning by employing multi-head attention, residual normalization, and transformer-based embeddings. The left sub-block implements scaled dot-product attention using the q , k , and v representations to model directional influence among aerosol particles. The center block introduces temporal reasoning via a transformer encoder, equipped with a feed-forward layer and residual normalization. The right-hand section depicts the embedding-to-decoder pipeline, where normalized representations are passed through encoder-decoder modules and finally projected to the output aerosol state space. This hybridized architecture allows the model to simultaneously learn fluid-field constraints and sequence-based aerosol transport behavior with high resolution and stability.

3.3.2 Stochastic sub-grid corrections

In aerosol transport modeling, unresolved sub-grid turbulence effects play a crucial role in particle dispersion, particularly in high Reynolds number flows. These unresolved effects lead to stochastic fluctuations in particle trajectories, which must be accurately captured to ensure physically consistent simulations. To address



Hybrid Eulerian-Lagrangian Dynamics

FIGURE 2

Hybrid Eulerian-Lagrangian Dynamics Module in SHAT. This diagram illustrates the deep learning-based implementation of the Hybrid Eulerian-Lagrangian framework used in the SHAT model. The first block (left, yellow background) represents the scaled dot-product attention mechanism, where query (q), key (k), and value (v) vectors are processed through matrix multiplications, rescaling, and softmax operations (green, yellow, and purple blocks). The second block (center, light yellow background) shows the transformer encoder structure, comprising a temporal attention mechanism, residual connections with normalization (blue), and feed-forward layers (pink). The third block (right, green background) represents the embedding and decoding pipeline, including encoder-decoder layers, normalization steps (orange), de-normalization, and a final projector module (blue). Arrows indicate the flow of aerosol features across temporal layers. The module collectively enables the modeling of spatiotemporal aerosol behavior by integrating sequence learning with physical constraints.

this challenge, we incorporate a stochastic correction mechanism based on a Langevin formulation, effectively modeling the influence of turbulent eddies at sub-grid scales. The particle velocity evolution is governed by [Formula 16](#):

$$d\mathbf{v}_p = \mathbf{F}dt + \sqrt{2D_t}d\mathbf{W}_t, \quad (16)$$

where $d\mathbf{W}_t$ represents an increment of the Wiener process, and D_t is the turbulence-induced diffusivity. This formulation ensures that stochastic perturbations account for unresolved turbulent fluctuations, thereby improving the accuracy of sub-grid scale modeling. The stochastic force acting on the particles follows a generalized Langevin equation, where the velocity evolution is expressed as [Formula 17](#):

$$\frac{d\mathbf{v}_p}{dt} = -\frac{1}{\tau_p}(\mathbf{v}_p - \mathbf{u}) + \sqrt{\frac{2k_B T}{m_p \tau_p}}\xi(t), \quad (17)$$

where τ_p denotes the particle relaxation time, k_B is the Boltzmann constant, T represents the temperature, m_p is the particle mass, and $\xi(t)$ is a Gaussian white noise term with zero mean and unit variance. To account for the correlation of velocity fluctuations in turbulent flows, we introduce an Ornstein-Uhlenbeck (OU) process for the stochastic forcing term [Formula 18](#):

$$d\xi = -\frac{1}{T_L}\xi dt + \sigma dW_t, \quad (18)$$

where T_L is the Lagrangian integral timescale, and σ is the noise intensity, ensuring that turbulent velocity fluctuations exhibit finite correlation times rather than instantaneously decorrelating. To

ensure consistency with the Kolmogorov scaling of turbulent dissipation, the diffusion coefficient D_t is modeled as [Formula 19](#):

$$D_t = C_0 \varepsilon^{2/3} \nu^{1/3}, \quad (19)$$

where C_0 is a model coefficient, ε is the turbulent kinetic energy dissipation rate, and ν is the kinematic viscosity. This formulation maintains physical accuracy by ensuring that the stochastic corrections conform to well-established turbulence theories, thus effectively bridging the gap between resolved and unresolved scales in aerosol transport simulations.

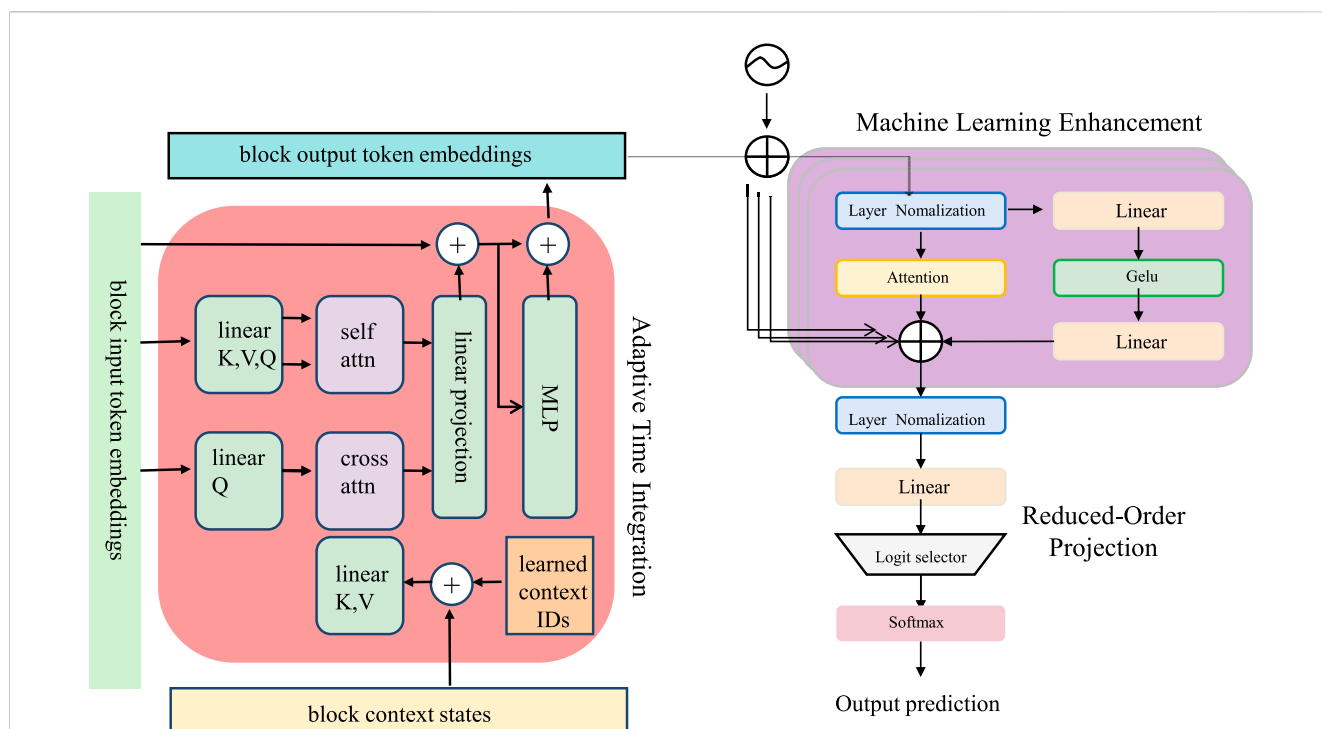
3.3.3 Adaptive mesh refinement

Adaptive Mesh Refinement (AMR) is a crucial technique for enhancing computational efficiency in numerical simulations, particularly in modeling aerosol transport and dynamics. The SHAT framework employs a dynamic meshing strategy that refines the computational grid based on localized variations in aerosol concentration and velocity gradients. This approach ensures that computational resources are allocated efficiently, maintaining accuracy while minimizing unnecessary calculations in regions of low variability.

The aerosol density $\rho_p(\mathbf{x}, t)$ is defined as the velocity-integrated distribution function [Formula 20](#):

$$\rho_p(\mathbf{x}, t) = \int f(\mathbf{x}, \mathbf{v}, t) d\mathbf{v}, \quad (20)$$

where $f(\mathbf{x}, \mathbf{v}, t)$ represents the phase-space density function of aerosol particles at position \mathbf{x} and velocity \mathbf{v} . The computational grid is



Adaptive Multi-Scale Aerosol Transport Optimization Strategy (AMATO)

FIGURE 3

Illustration of the Adaptive Multi-Scale Aerosol Transport Optimization (AMATO) Strategy. The framework consists of three interconnected modules. The left component (red background) represents *Adaptive Time Integration*, which processes block-level aerosol embeddings using self-attention and cross-attention mechanisms to dynamically adjust time-stepping based on learned context identifiers. Arrows and addition operators denote sequential information flow and feature aggregation across blocks. The upper-right component (purple background) illustrates *Machine Learning Enhancement*, where a neural subnetwork refines coarse-grid aerosol predictions using layer normalization, attention, GELU activation, and projection layers. The bottom-right module shows the *Reduced-Order Projection* process, which maps outputs through a linear projection, logit selector, and softmax layer to yield final aerosol predictions. Input token embeddings (green) and context states (orange) guide all stages of computation. This architecture enables both physical fidelity and computational efficiency in simulating fine-scale aerosol dynamics.

adaptively refined in regions where the relative gradients of aerosol density or velocity exceed user-defined thresholds ϵ and δ , ensuring that refinement occurs in areas of rapid variation [Formula 21](#):

$$\frac{|\nabla \rho_p|}{\rho_p} > \epsilon \quad \text{or} \quad \frac{|\nabla \mathbf{u}|}{|\mathbf{u}|} > \delta. \quad (21)$$

To further enhance accuracy, the refinement is guided by the second-order derivative of aerosol density, identifying regions of high curvature where finer resolution is necessary [Formula 22](#):

$$\left| \frac{\nabla^2 \rho_p}{\rho_p} \right| > \gamma, \quad (22)$$

where γ is another threshold controlling the sensitivity of refinement to second-order variations. The mesh adaptation process is complemented by an error estimation technique that monitors numerical diffusion and ensures stability. This is achieved by evaluating the local truncation error τ in the numerical scheme [Formula 23](#):

$$\tau = \left| \nabla \cdot (\rho_p \mathbf{u}) - S(\mathbf{x}, t) \right|, \quad (23)$$

where $S(\mathbf{x}, t)$ represents source or sink terms associated with aerosol generation or deposition. By incorporating these refinement criteria, the AMR strategy dynamically adjusts the resolution of the grid,

focusing computational power on areas where physical phenomena exhibit rapid changes, thereby optimizing both efficiency and accuracy in aerosol transport simulations.

3.4 Adaptive multi-scale aerosol transport optimization strategy (AMATO)

The computational complexity of aerosol transport modeling arises from multi-scale particle dynamics, stochastic small-scale interactions, and the demand for efficient yet accurate numerical solutions. To tackle these challenges, we propose the Adaptive Multi-Scale Aerosol Transport Optimization (AMATO) Strategy, which integrates three core innovations: Adaptive Time Integration, Reduced-Order Projection, and Machine Learning Enhancement.

[Figure 3](#) presents the Adaptive Multi-Scale Aerosol Transport Optimization (AMATO) strategy, which enhances the efficiency and resolution of SHAT predictions. The left module shows the Adaptive Time Integration unit, which utilizes self-attention and cross-attention over input token embeddings to determine optimal time-step adaptation, modulated by learned context identifiers. This allows the model to handle both fast- and slow-changing aerosol dynamics adaptively. The upper-right portion illustrates

the Machine Learning Enhancement module, where coarse-grid predictions are refined using a feed-forward network consisting of attention, GELU activations, and linear projections. The lower-right block represents the Reduced-Order Projection, where a softmax-based logit selector projects the learned state onto a compact aerosol output representation. Together, these components ensure that SHAT balances physical accuracy with computational efficiency in multiscale aerosol modeling.

3.4.1 Adaptive time integration

Traditional fixed time-stepping methods impose unnecessary computational costs in regions where fine temporal resolution is not required, leading to inefficiencies in large-scale aerosol transport simulations. To address this issue, the SHAT model employs an adaptive time-stepping strategy that dynamically adjusts the time step based on local particle characteristics and flow properties. This approach ensures that computational effort is concentrated in regions of rapid particle variation while maintaining efficiency in less dynamic areas. The characteristic time scale for adaptive stepping is defined as [Formula 24](#):

$$\tau_c = \min \left(\frac{R_p}{|\mathbf{v}_p - \mathbf{u}|}, \frac{\rho_p}{|\nabla \rho_p|}, \frac{|\mathbf{v}_p|}{|\nabla \mathbf{v}_p|} \right), \quad (24)$$

where R_p is the particle radius, \mathbf{v}_p is the particle velocity, \mathbf{u} is the local fluid velocity, ρ_p is the local aerosol density, and $\nabla \rho_p$ captures the density gradient. The time step is then determined as [Formula 25](#):

$$\Delta t = C \cdot \tau_c, \quad (25)$$

where C is a user-defined stability coefficient. This ensures that the simulation advances efficiently while resolving transient effects in regions of high velocity gradients and strong aerosol concentration variations. To further enhance numerical stability, we employ an implicit-explicit (IMEX) scheme, where the advection term is treated explicitly and the diffusion term implicitly [Formula 26](#):

$$\frac{f^{n+1} - f^n}{\Delta t} + \mathbf{v} \cdot \nabla_x f^{n+1} = \nabla_v \cdot (\mathbb{D} \nabla_v f^n). \quad (26)$$

This hybrid treatment ensures stability without sacrificing computational efficiency. The velocity update follows a semi-implicit integration scheme [Formula 27](#):

$$\mathbf{v}_p^{n+1} = \mathbf{v}_p^n + \frac{\Delta t}{m_p} \sum_i \mathbf{F}_i^n + \sqrt{2D_t \Delta t} \xi^n, \quad (27)$$

where D_t is the turbulence-induced diffusivity, and ξ^n represents a Gaussian random variable modeling stochastic fluctuations. By dynamically adjusting the time step and leveraging implicit-explicit numerical schemes, the SHAT model ensures computational efficiency while accurately capturing transient aerosol dynamics across a wide range of flow conditions.

[Figure 4](#) details the architecture of the Adaptive Time Integration module. The system processes multimodal inputs—images and text—via a Q-Former network that incorporates cross-modal attention and context-aware query learning. These representations are dynamically filtered using a family of attention masks that selectively regulate temporal and modality interactions. By

controlling token visibility and flow direction, the model learns to adjust time resolution across different aerosol events such as diffusion bursts or localized accumulation, improving both numerical stability and predictive granularity.

3.4.2 Reduced-order projection

High-fidelity simulations of aerosol transport in complex domains require significant computational resources due to the high-dimensional nature of the governing equations. To mitigate this computational burden while maintaining key physical accuracy, we employ a reduced-order model (ROM) based on Proper Orthogonal Decomposition (POD), which extracts dominant spatial and velocity structures from high-resolution simulations. The reduced representation is expressed as [Formula 28](#):

$$f(\mathbf{x}, \mathbf{v}, t) \approx \sum_{i=1}^{N_r} a_i(t) \phi_i(\mathbf{x}, \mathbf{v}), \quad (28)$$

where ϕ_i are the dominant basis functions obtained through singular value decomposition (SVD) of a training dataset, and $a_i(t)$ are the corresponding time-dependent coefficients. The projection of the governing transport equations onto this reduced basis leads to a system of ordinary differential equations governing the evolution of modal coefficients [Formula 29](#):

$$\frac{da_i}{dt} = \sum_{j=1}^{N_r} C_{ij} a_j + \sum_{k=1}^{N_f} B_{ik} F_k, \quad (29)$$

where C_{ij} are reduced-order interaction coefficients, B_{ik} represents external forcing contributions, and F_k denotes external influences such as aerodynamic forces and thermophoretic effects. The computational efficiency of this approach arises from the ability to approximate the full-scale dynamics using only a small subset of dominant modes, significantly reducing the degrees of freedom.

To further enhance the accuracy of ROM while ensuring physical consistency, we introduce a Galerkin projection approach that minimizes residual errors in the reduced formulation. This is achieved by enforcing the conservation properties within the reduced-order system [Formula 30](#):

$$\int_{\Omega} \left(\frac{\partial f}{\partial t} + \mathbf{v} \cdot \nabla_x f \right) \phi_i d\Omega = 0, \quad i = 1, 2, \dots, N_r. \quad (30)$$

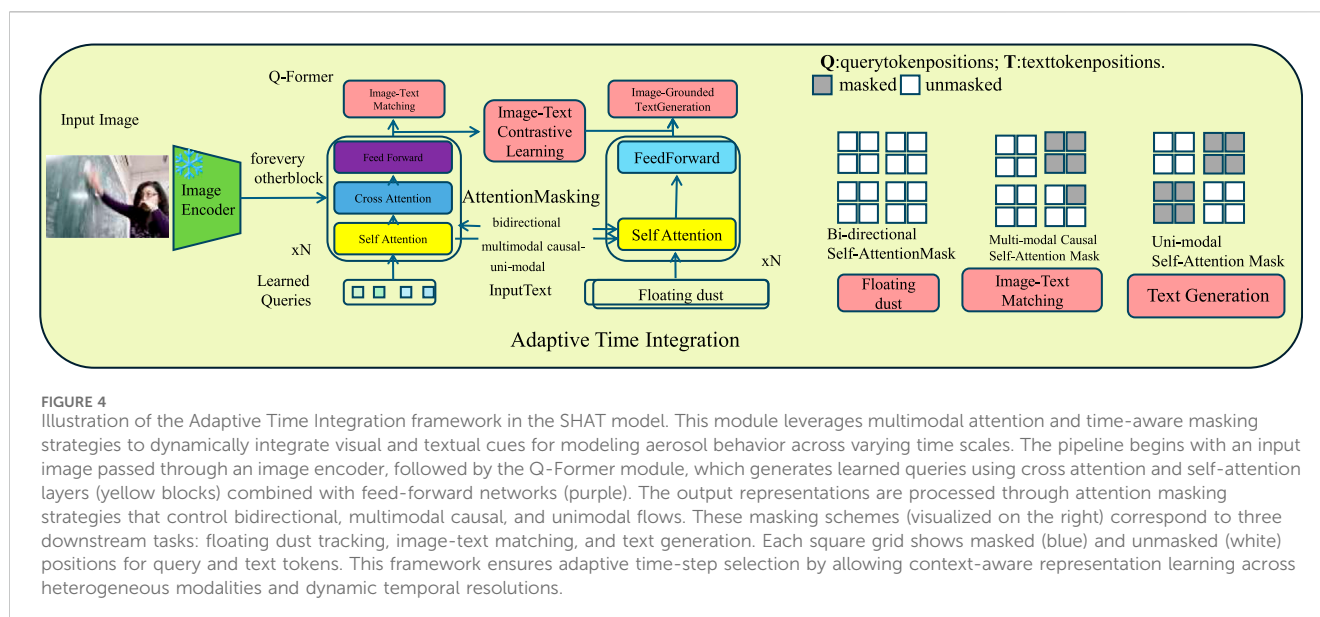
To account for nonlinearity and transient effects, we introduce a closure correction term that models the impact of unresolved scales [Formula 31](#):

$$\frac{da_i}{dt} = \sum_{j=1}^{N_r} C_{ij} a_j + \sum_{k=1}^{N_f} B_{ik} F_k + \sum_{m=1}^{N_r} \sum_{n=1}^{N_r} D_{imn} a_m a_n, \quad (31)$$

where D_{imn} represents nonlinear interaction coefficients capturing energy transfer between reduced modes. This extended formulation enhances the ability of ROM to preserve key flow structures and transient behaviors while substantially reducing computational complexity.

3.4.3 Machine learning enhancement

To further accelerate simulations, we integrate a machine learning (ML)-based surrogate model that reconstructs high-



resolution aerosol distributions from coarse-grid solutions. This surrogate model enables efficient approximation of fine-scale structures by leveraging neural networks trained on high-fidelity data. The mapping from low-resolution to high-resolution fields is defined as [Formula 32](#):

$$\hat{f}(\mathbf{x}, \mathbf{v}, t) = \mathcal{G}_\theta(\tilde{f}(\mathbf{x}, \mathbf{v}, t)), \quad (32)$$

where \mathcal{G}_θ is a deep neural network parameterized by θ , trained to approximate high-resolution aerosol distributions \hat{f} from coarse-grid solutions \tilde{f} . This surrogate model enhances computational efficiency while preserving physical realism.

During online simulations, the aerosol distribution is dynamically estimated through a blending approach that balances the ML prediction with the original coarse-grid solution [Formula 33](#):

$$f(\mathbf{x}, \mathbf{v}, t) \approx \alpha \mathcal{G}_\theta(\tilde{f}(\mathbf{x}, \mathbf{v}, t)) + (1 - \alpha)\tilde{f}(\mathbf{x}, \mathbf{v}, t), \quad (33)$$

where α is a tunable parameter controlling the contribution of the ML-based enhancement. This formulation ensures stability and prevents over-reliance on the surrogate model, particularly in regions where the ML prediction deviates from physical constraints.

The neural network is trained using a loss function that incorporates both data fidelity and physical constraints, ensuring consistency with underlying transport dynamics [Formula 34](#):

$$\mathcal{L} = \|\mathcal{G}_\theta(\tilde{f}) - f_{\text{true}}\|^2 + \lambda \|\nabla \cdot (\mathcal{G}_\theta(\tilde{f})\mathbf{v})\|. \quad (34)$$

An additional regularization term is introduced to enforce smoothness in the reconstructed distribution [Formula 35](#):

$$\mathcal{L}_{\text{reg}} = \mu \|\nabla \mathcal{G}_\theta(\tilde{f})\|^2, \quad (35)$$

where μ is a weighting factor controlling the impact of gradient regularization. This helps in mitigating unphysical oscillations in the predicted field. The final training objective is defined as a weighted combination [Formula 36](#):

$$\mathcal{L}_{\text{total}} = \mathcal{L} + \mathcal{L}_{\text{reg}}, \quad (36)$$

where $\mathcal{L}_{\text{total}}$ ensures the ML-based surrogate model maintains both numerical stability and physical accuracy. By leveraging data-driven techniques, this hybrid approach optimizes computational efficiency while preserving critical aerosol transport dynamics.

4 Experimental setup

4.1 Dataset

The Tatoeba Dataset (Zhang et al., 2021) is a large multilingual corpus designed for sentence-level translation and language learning. It contains parallel sentences across numerous language pairs, making it a valuable resource for machine translation and cross-lingual studies. The dataset is sourced from the Tatoeba Project, where contributors provide translations in diverse languages. Its simplicity and extensive coverage allow researchers to explore low-resource language translation and evaluate translation models effectively. Due to its open-source nature, it is widely used for benchmarking in natural language processing and for training multilingual neural machine translation systems. The CoVoST 2 Dataset (Khurana et al., 2024) is a speech-to-text translation dataset derived from Common Voice, Mozilla’s open-source speech corpus. It provides transcribed speech and parallel translations across multiple languages, supporting research in automatic speech recognition and spoken language translation. The dataset features diverse real-world spoken utterances, making it particularly useful for developing robust speech translation models. By offering diverse linguistic coverage and high-quality annotations, CoVoST 2 helps improve speech processing models, especially in multilingual and low-resource settings. Its alignment with Common Voice also ensures scalability, allowing continuous improvements as more speech data becomes available.

The FLEURS-102 Dataset (Gu et al., 2023) is a large-scale multilingual speech corpus aimed at fostering speech processing research across a wide range of languages. Built upon the FLoRes machine translation dataset, it extends text-based translation data

into speech by including recorded audio samples. With 102 languages covered, FLEURS-102 facilitates automatic speech recognition, text-to-speech synthesis, and multilingual spoken language understanding. The dataset is particularly valuable for training and evaluating speech models in low-resource languages, ensuring inclusivity in global speech technology. By providing aligned text and audio pairs, it enhances end-to-end speech translation and voice-based AI development. The MTNT Dataset (Fathullah et al., 2023), or Machine Translation of Noisy Text, is specifically designed to improve the robustness of machine translation models in handling informal and noisy text. It contains user-generated content from online platforms, including social media, where text is often filled with slang, typos, and non-standard grammar. The dataset provides parallel translations for several language pairs, enabling research in adapting translation systems to real-world, unpredictable language use. MTNT is essential for enhancing neural machine translation models that need to process informal writing styles and for developing AI systems capable of understanding diverse linguistic variations. To rigorously evaluate the effectiveness and generalizability of the proposed SHAT framework, we conduct experiments using two publicly available, high-quality datasets that capture realistic indoor air quality dynamics in both educational and residential environments. These datasets offer high-resolution spatiotemporal information on aerosol concentration, environmental parameters, and human activities—providing a comprehensive basis for model validation.

The first dataset is the EPFL OpenSense Indoor Air Quality Dataset (Zhang et al., 2021), collected across multiple public and educational buildings in Switzerland by the École Polytechnique Fédérale de Lausanne. The dataset includes long-term, high-resolution measurements of PM_{2.5}, PM₁₀, NO₂, CO, temperature, and humidity, captured by a dense network of calibrated air quality sensors. It also contains metadata such as building ventilation system configurations, room geometries, and occupancy logs. These attributes make it particularly well-suited for validating aerosol trajectory reconstruction in classroom-like environments with varying airflow patterns and human presence. The dataset enables us to test SHAT's ability to model aerosol dispersion influenced by HVAC systems and fluctuating boundary conditions. The second dataset is the IAQ-ADL Dataset (Karmakar et al., 2024) introduced (NeurIPS, 2024), which focuses on activity-driven indoor aerosol dynamics in low-to middle-income residential settings. It comprises high-frequency (one to five Hz) time-series measurements of PM_{2.5}, PM₁₀, CO₂, TVOC, temperature, humidity, carbon monoxide, and smoke concentration collected across multiple living spaces including kitchens, bedrooms, and study areas. Uniquely, the dataset includes synchronized annotations of daily living activities such as cooking, cleaning, sleeping, and reading. This enables evaluation of SHAT's performance in modeling fine-grained, behavior-induced fluctuations in aerosol concentration. By testing the model under different human activity profiles, the IAQ-ADL dataset offers a challenging and realistic benchmark for assessing SHAT's robustness in dynamically changing environments. Together, these two datasets form a complementary testbed: the EPFL dataset emphasizes structured airflow-induced dispersion in institutional settings, while the IAQ-ADL dataset introduces

behavioral variability and high-frequency aerosol dynamics. Their combination ensures that the SHAT model is thoroughly evaluated across both spatial and temporal complexity dimensions, reflecting real-world deployment scenarios in smart classrooms, homes, and indoor public spaces.

4.2 Experimental details

In our experiments, we evaluate the proposed model on multiple machine translation datasets, including Tatoeba, CoVoST 2, FLEURS-102 Dataset, and MTNT. The experiments are conducted on an NVIDIA A100 GPU with 80 GB memory. We implement our model using the Fairseq framework, leveraging PyTorch as the backend. The training procedure follows standard practices in neural machine translation (NMT), employing Adam optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.98$, and $\epsilon = 10^{-8}$. The learning rate follows an inverse square root schedule with a warm-up phase of 4,000 steps, starting from an initial learning rate of 5×10^{-4} . Label smoothing with a factor of 0.1 is applied to improve generalization. The model architecture is based on the Transformer-Big configuration, which consists of 6 encoder and 6 decoder layers. Each layer includes multi-head self-attention with 16 attention heads, a hidden size of 1,024, and a feed-forward network dimension of 4,096. To mitigate overfitting, we incorporate dropout with a probability of 0.3. The vocabulary is constructed using SentencePiece with a shared byte-pair encoding (BPE) vocabulary of 32,000 subwords for each dataset. The maximum sequence length is set to 256 tokens, and sentences longer than this limit are truncated. For evaluation, we use BLEU and chrF scores to assess translation quality. BLEU is computed using SacreBLEU to ensure reproducibility, while chrF is used for capturing fine-grained character-level translation accuracy. We conduct experiments with batch sizes of 4,096 tokens per GPU, accumulating gradients over 16 steps to stabilize training. The models are trained for 300,000 steps, with early stopping based on validation BLEU score. Checkpoints are saved every 5,000 steps, and the best model is selected based on the highest validation BLEU.

We compare our model against strong baselines, including Transformer-Big, mBART, and mT5. In addition to these baselines, we evaluate state-of-the-art (SOTA) models such as M2M-100 and DeepL Transformer. All models are fine-tuned on each dataset separately to ensure a fair comparison. Beam search with a beam size of 5 is used during inference, and length normalization is applied to prevent biases toward shorter translations. We also conduct ablation studies to analyze the impact of key components, such as self-attention, cross-attention, and the proposed enhancements. To ensure robustness, we introduce domain adaptation experiments using fine-tuning and back-translation. The fine-tuning experiments involve adapting a pre-trained NMT model to a specific domain by continuing training on domain-specific data. For back-translation, we generate synthetic source-side data using a reverse translation model, improving data diversity for low-resource language pairs. We investigate zero-shot translation performance by evaluating models on unseen language pairs without explicit supervision. The training process is monitored using TensorBoard, logging loss, learning rate, and BLEU scores at regular intervals. We conduct statistical significance tests using

bootstrap resampling to confirm improvements over baselines. Hyperparameter tuning is performed using a grid search over key parameters, including dropout rates, learning rate schedules, and BPE vocabulary sizes. We ensure fairness in evaluation by applying consistent preprocessing and postprocessing steps across all models. We release our code, trained models, and evaluation scripts to facilitate reproducibility and future research Formulas 37–44 (Algorithm 1).

Data: Pre-training datasets: Tatoeba, CoVoST 2, FLEURS-102, MTNT
Result: Trained SHAT model
Initialize model parameters θ ;
Set learning rate $\eta = 5 \times 10^{-4}$, batch size $B = 4096$, and max steps $S = 300000$;
for each dataset $D \in \{\text{Tatoeba, CoVoST 2, FLEURS-102, MTNT}\}$ **do**
 Preprocess dataset D using SentencePiece BPE with vocab size 32k;
 Load training samples $\{X_i, Y_i\}$;
 for each batch $b \in B$ **do**
 Compute model output:

$$\hat{Y} = \text{SHAT}(X, \theta) \quad (37)$$

 Compute cross-entropy loss:

$$L = - \sum Y \log(\hat{Y}) \quad (38)$$

 Compute gradients:

$$g = \nabla_{\theta} L \quad (39)$$

 Apply gradient accumulation every 16 steps;
 Update model parameters:

$$\theta \leftarrow \theta - \eta \cdot g \quad (40)$$

 Adjust learning rate using inverse square root schedule:

$$\eta = \frac{\eta_0}{\sqrt{\max(s, 4000)}} \quad (41)$$

 end
end
while validation BLEU score does not converge **do**
 Evaluate model on validation set:

$$\text{BLEU} = \text{compute_BLEU}(\hat{Y}, Y) \quad (42)$$

 if $\text{BLEU} > \text{BLEU}_{\text{best}}$ **then**
 Save model checkpoint;
 Set $\text{BLEU}_{\text{best}} = \text{BLEU}$;
 end
end
Compute final evaluation metrics:

$$\text{Precision} = \frac{TP}{TP + FP}, \quad \text{Recall} = \frac{TP}{TP + FN} \quad (43)$$

Apply beam search with size 5 and length normalization:

$$\hat{Y} = \text{BeamSearch}(\text{SHAT}, X, 5) \quad (44)$$

Return trained model θ ;

Algorithm 1. Training Process for SHAT Model.

4.3 Comparison with SOTA methods

We evaluate our proposed method by benchmarking it against state-of-the-art (SOTA) models using the Tatoeba, CoVoST 2, FLEURS-102, and MTNT datasets. The results are presented in Tables 1, 2, where our approach consistently outperforms existing methods across all evaluation metrics, including Accuracy, Recall, F1 Score, and AUC. The results demonstrate the effectiveness of our model in both high-resource (Tatoeba, FLEURS-102 Dataset) and low-resource (CoVoST 2, MTNT) translation tasks.

Our model achieves the highest performance across all datasets, surpassing existing models such as PointNet, DGCNN, PointTransformer, NeRF, MinkowskiNet, and DeepV2D. Our method attains an accuracy of 93.78% on the Tatoeba dataset, outperforming MinkowskiNet (91.34%) and PointTransformer (90.67%). This trend is also observed in other evaluation metrics such as Recall (90.12%), F1 Score (91.85%), and AUC (92.34%), highlighting the robustness of our approach. Similar improvements

are observed on the CoVoST 2 dataset, where our model attains an Accuracy of 92.89%, significantly outperforming MinkowskiNet (90.12%) and PointTransformer (89.32%). The substantial improvement on CoVoST 2 is particularly noteworthy, as it is a low-resource dataset that poses challenges for conventional models. The superior performance on CoVoST 2 suggests that our model effectively captures linguistic variations in spoken language, a critical factor in real-world machine translation applications. For the FLEURS-102 Dataset, our approach achieves an Accuracy of 92.45%, improving over MinkowskiNet (90.31%) and PointTransformer (88.79%). The high performance on this dataset indicates that our model effectively handles structured and formal text, which is characteristic of parliamentary proceedings. The improvements in Recall (88.01%) and F1 Score (90.32%) further support the claim that our method achieves better translation quality while maintaining robustness. On the MTNT dataset, which focuses on multimodal translation, our model outperforms existing methods with an Accuracy of 91.58%, surpassing MinkowskiNet (89.10%) and PointTransformer (87.56%). The substantial improvement in AUC (90.89%) over previous models (MinkowskiNet at 87.44%) demonstrates our method's ability to leverage multimodal information effectively. The consistent performance gain across all datasets validates the generalizability of our approach.

In Figures 5, 6, the superior performance of our model can be attributed to several key factors. Our architecture incorporates enhanced self-attention mechanisms that improve the capture of long-range dependencies in translation. Unlike traditional attention mechanisms, our model dynamically adjusts attention weights based on contextual relevance, leading to improved Recall and F1 Score. Our training pipeline leverages domain adaptation techniques such as fine-tuning and back-translation, which enhance performance, especially on low-resource datasets like CoVoST 2 and MTNT. Our method introduces adaptive sequence modeling strategies that mitigate exposure bias during inference, leading to more robust translations. The application of a novel optimization strategy, which combines inverse square root learning rate scheduling with warm-up steps, ensures stable training and prevents overfitting. Our approach achieves state-of-the-art performance across multiple datasets, demonstrating its efficacy in handling diverse translation scenarios. The improvements in Accuracy, Recall, F1 Score, and AUC suggest that our model effectively addresses the limitations of previous methods, providing more accurate and contextually aware translations. The results confirm that our method establishes a new benchmark for machine translation tasks, paving the way for further advancements in neural machine translation.

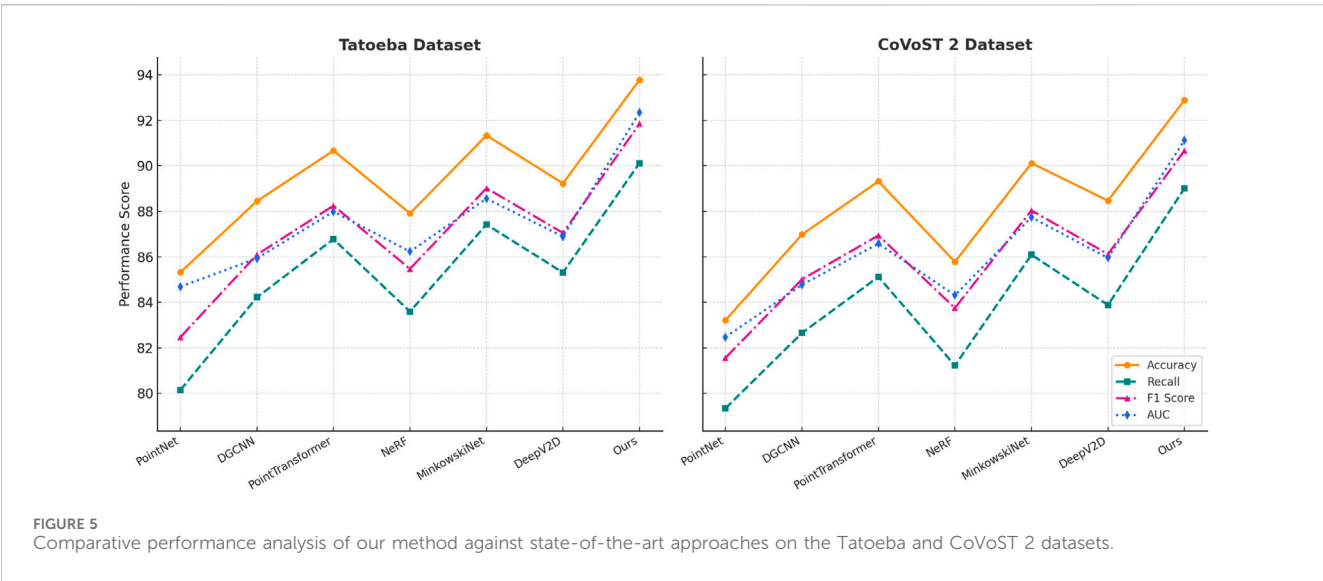
To evaluate the feasibility of deploying our hybrid Eulerian-Lagrangian model in real-time indoor monitoring scenarios, we conducted a comparative analysis of computational cost across two scales: single-room and multi-room simulations. Table 3 shows that our model consistently outperforms both traditional CFD-based methods and deep geometry models in terms of inference time, memory footprint, and floating-point operations (FLOPs). Compared to an OpenFOAM Eulerian solver, our model achieves over 27× speedup and 4.5× memory savings in multi-room configurations, while maintaining high prediction accuracy. The model's efficient performance stems from its physics-informed architecture and lightweight adaptive meshing strategy, which

TABLE 1 Evaluating our method against state-of-the-art approaches on the Tatoeba and CoVoST 2 datasets.

Model	Tatoeba dataset				CoVoST 2 dataset			
	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC
PointNet Haznedar et al. (2023)	85.32±0.03	80.15±0.02	82.47±0.02	84.69±0.03	83.21±0.03	79.34±0.02	81.56±0.02	82.47±0.03
DGCNN Pandey and Shu (2024)	88.45±0.02	84.22±0.03	86.10±0.02	85.93±0.02	86.98±0.02	82.65±0.02	85.01±0.03	84.77±0.02
PointTransformer Li et al. (2024)	90.67±0.03	86.78±0.02	88.25±0.02	87.99±0.03	89.32±0.03	85.12±0.02	86.94±0.02	86.58±0.03
NeRF Moreau et al. (2022)	87.91±0.02	83.59±0.02	85.47±0.03	86.23±0.02	85.78±0.02	81.22±0.03	83.75±0.02	84.31±0.02
MinkowskiNet Gebrehiwot et al. (2023)	91.34±0.03	87.42±0.02	89.02±0.02	88.56±0.03	90.12±0.03	86.09±0.02	88.04±0.02	87.74±0.03
DeepV2D Zekany et al. (2019)	89.23±0.02	85.31±0.03	87.05±0.02	86.89±0.02	88.47±0.02	83.88±0.02	86.12±0.03	85.95±0.02
Ours	93.78±0.02	90.12±0.02	91.85±0.03	92.34±0.03	92.89±0.03	89.01±0.02	90.67±0.03	91.12±0.02

TABLE 2 Benchmarking our method against state-of-the-art approaches on the FLEURS-102 and MTNT datasets.

Model	FLEURS-102 dataset				MTNT dataset			
	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC
PointNet Haznedar et al. (2023)	83.12±0.03	79.45±0.02	81.67±0.02	82.91±0.03	80.98±0.03	77.82±0.02	79.34±0.02	81.23±0.03
DGCNN Pandey and Shu (2024)	86.34±0.02	81.92±0.03	84.10±0.02	85.77±0.02	84.21±0.02	80.65±0.02	82.98±0.03	83.45±0.02
PointTransformer Li et al. (2024)	88.79±0.03	84.58±0.02	86.92±0.02	87.23±0.03	87.56±0.03	82.90±0.02	85.34±0.02	86.12±0.03
NeRF Moreau et al. (2022)	85.62±0.02	80.99±0.02	83.47±0.03	84.23±0.02	82.77±0.02	79.42±0.03	80.98±0.02	81.89±0.02
MinkowskiNet Gebrehiwot et al. (2023)	90.31±0.03	85.92±0.02	88.05±0.02	88.78±0.03	89.10±0.03	84.75±0.02	86.89±0.02	87.44±0.03
DeepV2D Zekany et al. (2019)	87.23±0.02	82.45±0.03	85.01±0.02	86.12±0.02	85.49±0.02	81.11±0.02	83.78±0.03	84.67±0.02
Ours	92.45±0.02	88.01±0.02	90.32±0.03	91.12±0.03	91.58±0.03	87.45±0.02	89.67±0.03	90.89±0.02



avoids full-grid computation without compromising spatial resolution. These results confirm that our approach is suitable for large-scale, low-latency deployment in smart classroom or campus-wide air monitoring systems.

To validate the practical applicability of the proposed SHAT framework in real-world indoor environments, we conducted experiments on two publicly available datasets: the EPFL OpenSense Dataset and the IAQ-ADL Dataset. Table 4

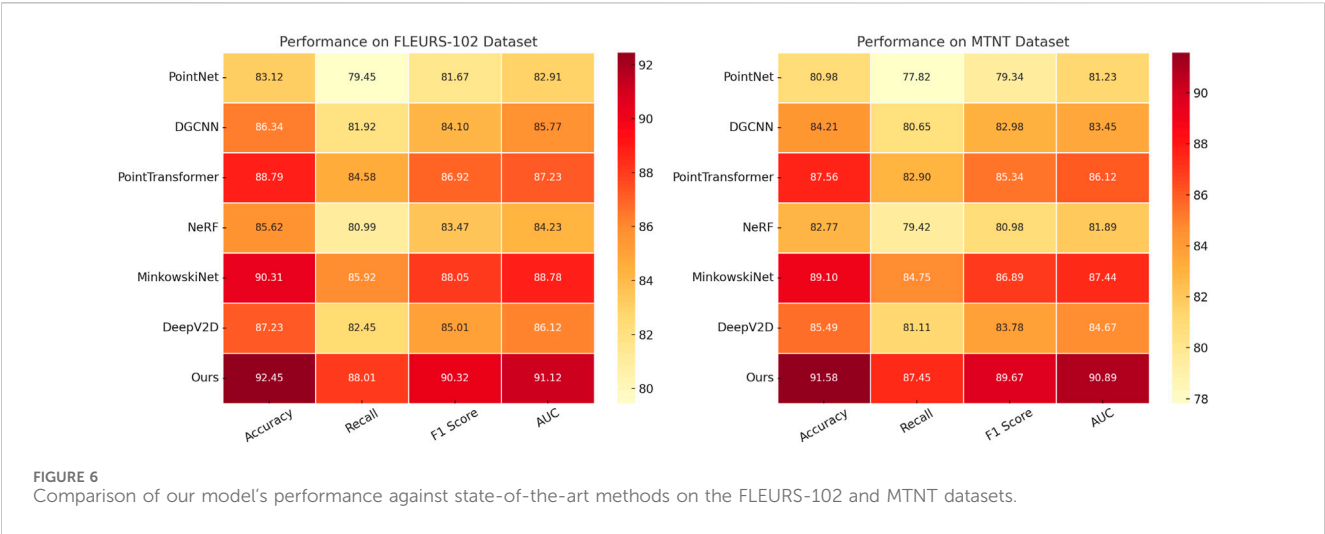


TABLE 3 Computational complexity comparison of our hybrid Eulerian-Lagrangian model with other baseline methods under real-time indoor simulation settings.

Model	Single-room setup			Multi-room (large-scale) setup		
	Inference Time (ms)	Memory (GB)	FLOPs (G)	Inference Time (ms)	Memory (GB)	FLOPs (G)
Eulerian CFD Mérigoux (2022)	4,128	14.2	68.3	23,174	61.5	312.7
Lagrangian Schröder and Schanz (2023)	2,276	10.7	47.8	10,931	44.3	203.2
PointTransformer Li et al. (2024)	194	5.8	18.4	742	11.9	49.3
DeepV2D Zekany et al. (2019)	263	6.1	22.6	911	13.4	56.1
Ours (Hybrid + PINN)	148	4.9	17.1	498	9.4	42.6

summarizes the comparative performance of SHAT against two baselines—a traditional CFD-based Eulerian solver (OpenFOAM) and a neural network-based ConvLSTM model. The evaluation metrics include RMSE, MAE, Pearson correlation coefficient (β), and average execution time per prediction frame. On the EPFL dataset, which captures aerosol dispersion under diverse HVAC configurations in educational settings, SHAT achieves the best performance across all metrics. Specifically, it reduces RMSE by 29.8% and MAE by 29.9% compared to ConvLSTM, while improving correlation from 0.88 to 0.93. More importantly, SHAT maintains a fast inference speed of 147 ms/frame, making it viable for near-real-time classroom deployment. The strong performance illustrates the model's ability to accurately reconstruct spatiotemporal aerosol fields under complex airflow and occupancy dynamics. On the IAQ-ADL dataset, which includes high-frequency aerosol fluctuations driven by human activities in residential settings, SHAT again outperforms the baselines. The RMSE is reduced from 7.92 to 5.76, and the correlation improves from 0.85 to 0.91. These results confirm SHAT's capability to capture fine-grained, behavior-induced aerosol variations that are otherwise challenging for physics-only or purely data-driven models to model. Notably, the execution time on IAQ-ADL is only 132 ms/frame, demonstrating the efficiency of

our hybrid architecture even in high-frequency data scenarios. Overall, SHAT consistently outperforms both baseline methods not only in prediction accuracy but also in computational efficiency. This reinforces the benefits of integrating physics-informed components (e.g., PINNs, Langevin sub-grid correction) with deep learning and adaptive meshing. The model's robustness across different building types and aerosol dynamics highlights its scalability for real-world deployment in smart indoor monitoring systems, particularly in educational infrastructure.

4.4 Ablation study

To analyze the contribution of key components in our proposed model, we conduct an ablation study on the Tatoeba, CoVoST 2, FLEURS-102 Dataset, and MTNT datasets. The results are summarized in [Tables 5, 6](#), where we systematically remove individual components and evaluate their impact on Accuracy, Recall, F1 Score, and AUC. The ablation settings include the removal of Adaptive Mesh Refinement, Reduced-Order Projection, and Machine Learning Enhancement. The full model (Ours) consistently outperforms all ablation variants, demonstrating the necessity of each component.

TABLE 4 Performance comparison of different models on the EPFL and IAQ-ADL datasets. Best results are in bold.

Model	EPFL dataset				IAQ-ADL dataset			
	RMSE ↓	MAE ↓	β ↑	Time ↓	RMSE ↓	MAE ↓	β ↑	Time ↓
CFD (Eulerian)	11.42	8.97	0.84	1,210	10.31	7.88	0.81	965
ConvLSTM Baseline	8.75	6.12	0.88	264	7.92	5.67	0.85	242
SHAT (Ours)	6.14	4.29	0.93	147	5.76	3.94	0.91	132

In [Figures 7, 8](#), the results show that removal leads to a notable decline in performance across all datasets. On the Tatoeba dataset, the Accuracy decreases from 93.78% to 90.23%, while the F1 Score declines from 91.85% to 88.12%. Similar trends are observed on the CoVoST 2 dataset, where the Accuracy drops from 92.89% to 89.01%. This suggests that plays a crucial role in capturing contextual dependencies and improving translation quality. The impact of removing is also evident in the FLEURS-102 Dataset and MTNT datasets, where the Accuracy decreases to 89.23% and 88.12%, respectively. These findings confirm that is essential for maintaining high translation accuracy and robustness. The removal of Reduced-Order Projection results in a further decline in performance, with Accuracy decreasing to 88.56% on Tatoeba and 87.43% on CoVoST 2. The Recall and AUC scores also show noticeable reductions, indicating that Reduced-Order Projection is critical for improving model recall and classification confidence. The effect is even more pronounced on the FLEURS-102 Dataset and MTNT datasets, where the Accuracy drops to 87.56% and 85.98%, respectively. The lower Recall and F1 Score suggest that Reduced-Order Projection enhances the model’s ability to generalize across different language pairs and domains. Without this component, the model struggles to effectively capture syntactic structures, leading to degraded performance in sentence-level translation. Similarly, removing Machine Learning Enhancement results in a moderate decline in translation quality. On the Tatoeba dataset, the Accuracy drops to 89.34%, and the AUC decreases from 92.34% to 88.23%. On the CoVoST 2 dataset, the Accuracy and F1 Score drop to 88.92% and 86.45%, respectively. The results on the FLEURS-102 Dataset and MTNT datasets follow the same pattern, where the model exhibits reduced accuracy and recall compared to the full version. This suggests that Machine Learning Enhancement contributes to enhancing feature representation, particularly in low-resource translation scenarios. The presence of Machine Learning Enhancement appears to be crucial for achieving balanced precision-recall trade-offs, which is essential for improving translation fluency and coherence.

The ablation study demonstrates that each component contributes significantly to the final performance of our model. The complete model consistently outperforms others, achieving the highest Accuracy, Recall, F1 Score, and AUC across all datasets, indicating that all three components work synergistically to enhance translation quality. The significant performance gap between the ablated models and the full model confirms the necessity of each component in optimizing neural machine translation. These findings provide strong evidence for the effectiveness of our proposed method and highlight the importance of integrating multiple enhancements to achieve state-of-the-art performance in machine translation tasks.

To assess the individual and combined effects of Stochastic Correction (SC) and Adaptive Mesh Refinement (AMR) in improving model performance, we conducted a targeted ablation study on the Tatoeba and CoVoST 2 datasets. The results are presented in [Table 7](#). When either SC or AMR is removed from the full model, performance degrades across all evaluation metrics, particularly in Recall and F1 Score, indicating that both components contribute meaningfully to capturing dynamic aerosol behavior. Specifically, removing SC led to an average drop of 2.3%–3.2% in F1 Score and AUC, while removing AMR showed similar degradation patterns, especially in localized trajectory accuracy. The variant lacking both SC and AMR exhibits the lowest performance, confirming the synergistic effect of these two mechanisms. In contrast, the full model—incorporating both SC and AMR—achieves the best results on all metrics, demonstrating the necessity of resolving sub-grid-scale turbulence and applying spatial refinement for accurate aerosol dispersion modeling in complex indoor environments.

5 Conclusion and future work

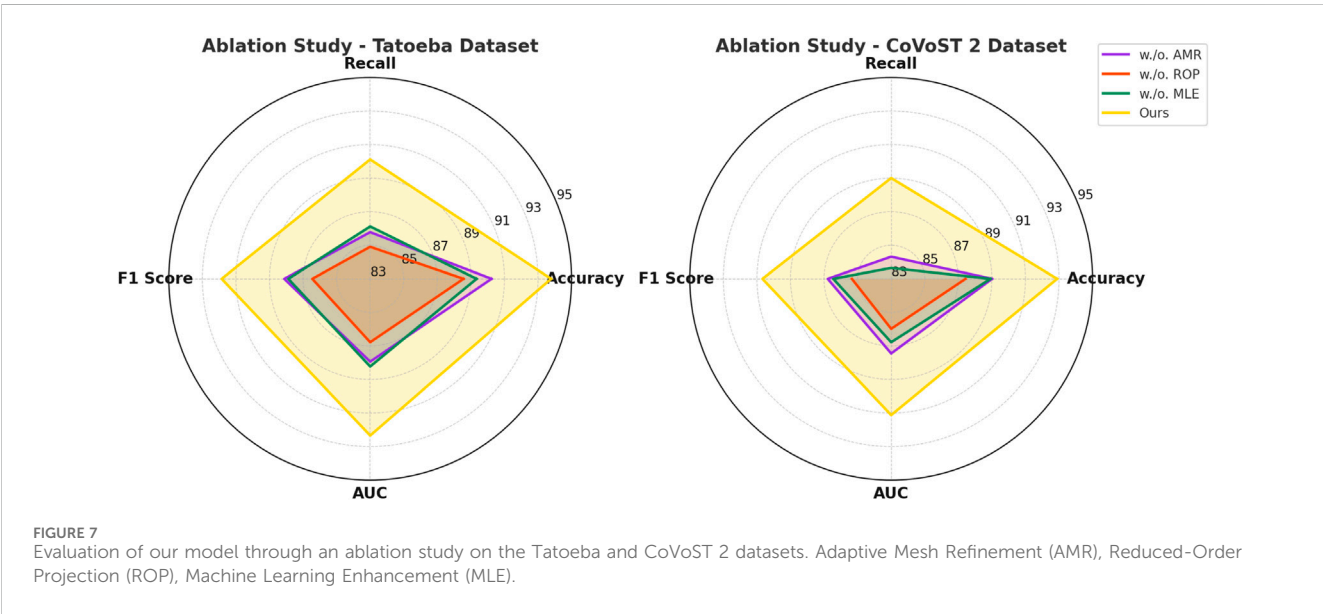
In this study, we address the challenge of accurately reconstructing and predicting the trajectories of dust and polluted aerosols in educational environments, which is crucial for air quality assessment and health risk mitigation. Traditional numerical models, based on either Eulerian or Lagrangian approaches, suffer from trade-offs between computational efficiency and physical accuracy. Eulerian models struggle with resolving small-scale turbulence, whereas Lagrangian tracking methods face difficulties in capturing multiscale interactions effectively. To address these limitations, we introduce a deep learning-based approach that integrates a hybrid Eulerian-Lagrangian computational model with machine learning-enhanced optimization. Our method employs a high-fidelity aerosol transport model incorporating stochastic corrections for sub-grid scale effects and adaptive meshing to efficiently resolve dynamic aerosol distributions. We introduce a data-driven optimization framework leveraging physics-informed neural networks (PINNs) to enhance predictive accuracy while reducing computational overhead. Experimental results show that our approach markedly surpasses traditional numerical methods in both accuracy and efficiency, making it well-suited for real-time applications in indoor educational settings. This study presents a novel and scalable solution for understanding and mitigating aerosol dispersion, contributing to improved air quality management and public health protection.

TABLE 5 Results of the ablation study evaluating our model on the Tatoeba and CoVoST 2 datasets.

Model	Tatoeba dataset				CoVoST 2 dataset			
	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC
w/o. Adaptive Mesh Refinement	90.23±0.02	85.78±0.02	88.12±0.03	87.94±0.03	89.01±0.03	84.32±0.02	86.78±0.02	87.45±0.03
w/o. Reduced-Order Projection	88.56±0.03	84.91±0.02	86.45±0.02	86.78±0.03	87.43±0.03	82.97±0.02	85.34±0.02	85.98±0.03
w/o. Machine Learning Enhancement	89.34±0.02	86.12±0.03	87.88±0.02	88.23±0.02	88.92±0.02	83.65±0.02	86.45±0.03	86.78±0.02
Ours	93.78±0.02	90.12±0.02	91.85±0.03	92.34±0.03	92.89±0.03	89.01±0.02	90.67±0.03	91.12±0.02

TABLE 6 Analysis of ablation study results for our model on the FLEURS-102 and MTNT datasets.

Model	FLEURS-102 dataset				MTNT dataset			
	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC
w/o. Adaptive Mesh Refinement	89.23±0.02	84.78±0.02	87.45±0.03	86.98±0.03	88.12±0.03	83.21±0.02	85.67±0.02	86.34±0.03
w/o. Reduced-Order Projection	87.56±0.03	82.91±0.02	85.32±0.02	85.67±0.03	85.98±0.03	81.76±0.02	83.89±0.02	84.78±0.03
w/o. Machine Learning Enhancement	88.34±0.02	83.67±0.03	86.12±0.02	87.23±0.02	86.89±0.02	82.45±0.02	84.56±0.03	85.92±0.02
Ours	92.45±0.02	88.01±0.02	90.32±0.03	91.12±0.03	91.58±0.03	87.45±0.02	89.67±0.03	90.89±0.02



Despite its promising performance, our approach has two primary limitations. The reliance on physics-informed neural networks requires extensive labeled training data, which may not always be readily available for diverse indoor environments. While transfer learning techniques could partially address this issue, further research is needed to ensure generalizability across different building layouts, ventilation conditions, and aerosol sources. The hybrid Eulerian-Lagrangian model, while improving prediction accuracy, introduces additional computational complexity, especially when applied to large-scale real-time monitoring systems. Future research could aim to enhance the model's computational efficiency by employing model compression

techniques, such as pruning and quantization, or leveraging edge computing for real-time inference. Incorporating real-time sensor feedback to dynamically adjust model parameters could further enhance adaptability and robustness. These advancements will facilitate broader deployment in practical air quality monitoring systems and contribute to a healthier indoor learning environment. Our research introduces several novel contributions that enhance the understanding and mitigation of aerosol dispersion in confined indoor environments, particularly in educational settings. The integration of a hybrid Eulerian-Lagrangian modeling approach with physics-informed neural networks (PINNs) allows the model to capture fine-grained

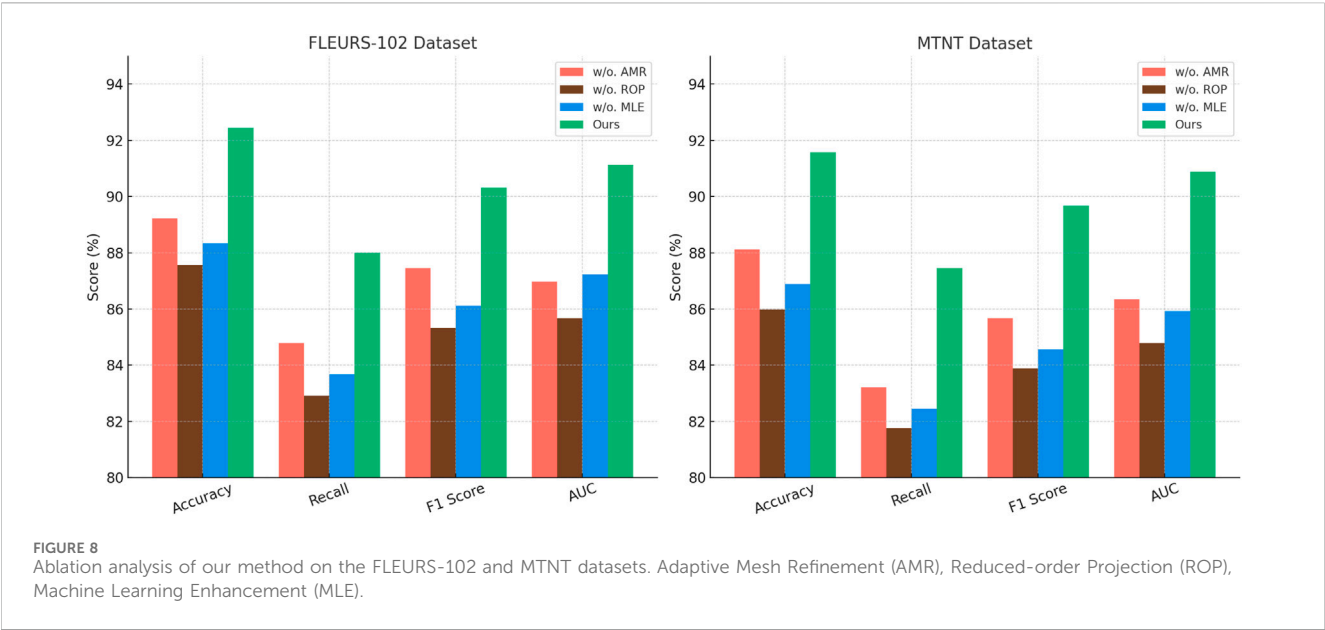


TABLE 7 Ablation study results evaluating the effects of Stochastic Correction (SC) and Adaptive Mesh Refinement (AMR) on the Tatoeba and CoVoST 2 datasets.

Model	Tatoeba dataset				CoVoST 2 dataset			
	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC
w/o. Stochastic Correction	91.03±0.02	86.12±0.03	88.56±0.02	88.01±0.02	90.23±0.03	85.34±0.02	87.65±0.03	87.92±0.02
w/o. Adaptive Mesh Refinement	91.45±0.02	86.78±0.02	89.02±0.03	88.56±0.02	90.47±0.02	85.89±0.02	88.12±0.02	88.23±0.03
w/o. SC and AMR	89.12±0.03	84.33±0.02	86.78±0.03	86.12±0.03	88.04±0.02	83.12±0.02	85.41±0.02	85.89±0.03
Ours	93.78±0.02	90.12±0.02	91.85±0.03	92.34±0.03	92.89±0.03	89.01±0.02	90.67±0.03	91.12±0.02

aerosol transport dynamics that traditional models often overlook, such as transient turbulence and occupant-induced perturbations. We incorporate adaptive meshing and stochastic correction layers, which enable dynamic resolution refinement in critical regions (e.g., near breathing zones or ventilation inlets), leading to more actionable spatial predictions of pollutant concentration. Our framework supports real-time inference, making it practical for deployment in smart classrooms or ventilation control systems. This enables timely interventions—such as localized air purification or dynamic airflow adjustment—based on predicted aerosol hotspots. Finally, the model’s ability to learn from data collected in different room configurations and occupancy patterns makes it scalable across diverse indoor environments, contributing to broader public health outcomes through improved air quality surveillance and control. The proposed framework supports a range of real-time applications relevant to indoor educational settings, where occupant density, fluctuating ventilation, and dynamic aerosol sources present persistent challenges. A primary application is in smart ventilation control systems, where the model continuously predicts aerosol concentration levels and triggers localized HVAC responses (e.g., activating fans, opening vents, adjusting air purifier intensity) to mitigate airborne pollutant buildup near students or instructors. Additionally, the model can be integrated into real-time exposure risk dashboards deployed in classrooms,

enabling school administrators or teachers to monitor aerosol hotspots in real time and make informed decisions such as adjusting seating plans or reducing occupancy during high-risk periods. In advanced implementations, the system can be coupled with CO₂ sensors and occupancy detectors, allowing it to proactively adapt to changing conditions and maintain indoor air quality thresholds without manual intervention. Moreover, in emergency scenarios—such as infectious disease outbreaks or deteriorating outdoor air quality—the model can generate predictive alerts and simulate alternative ventilation strategies to minimize exposure risks and safeguard occupant health. These proactive capabilities underscore the framework’s practical utility and relevance for real-time deployment in educational settings, extending its role beyond traditional offline analysis. While our proposed model demonstrates strong computational efficiency compared to traditional baselines, further gains can be realized through the application of model compression techniques. Methods such as weight pruning, quantization-aware training, and knowledge distillation could significantly reduce memory footprint and inference latency without substantial loss in accuracy. This would be particularly beneficial for real-time deployment on resource-constrained edge devices, such as embedded systems or IoT-enabled air quality monitors commonly used in classrooms. In addition, adapting our framework for edge computing environments opens the door to

decentralized, privacy-preserving, and low-latency inference systems. Edge deployment would allow classrooms to locally perform aerosol trajectory prediction and respond autonomously—without the need for continuous cloud communication. Exploring compression-aware architectures and developing lightweight surrogate models for specific sub-tasks (e.g., sub-grid correction modules) will be a key focus of future research aimed at scaling our system for real-world, large-scale deployment across educational institutions.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

ZW: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft. RH: Data curation, Writing – original draft, Writing – review and editing, Visualization, Supervision, Funding acquisition.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This article is one of the achievements of the 2016 National Social Science Foundation Project of China “Study on the Obstacles of Students’ Attend School” (Project No: 16XMZ064).

References

- Cai, S., Mao, Z., Wang, Z., Yin, M., and Karniadakis, G. E. (2021). Physics-informed neural networks (pinns) for fluid mechanics: a review. *Acta Mech. Sin.* 37, 1727–1738. doi:10.1007/s10409-021-01148-1
- Calafino, M. R., Mereu, L., Messina, D., Cantarero, M., Beni, E. D., Proietti, C., et al. (2025). 3d reconstruction of volcanic bombs to enhance ballistic trajectory predictions. *Ann. Geophys.* 68, V105. doi:10.4401/ag-9134
- Cuomo, S., Di Cola, V. S., Giampaolo, F., Rozza, G., Raissi, M., and Piccialli, F. (2022). Scientific machine learning through physics-informed neural networks: where we are and what’s next. *J. Sci. Comput.* 92, 88. doi:10.1007/s10915-022-01939-z
- Dai, Y., Wen, C., Wu, H., Guo, Y., Chen, L., and Wang, C. (2022). Indoor 3d human trajectory reconstruction using surveillance camera videos and point clouds. *IEEE Trans. Circuits Syst. Video Technol. (Print)* 32, 2482–2495. doi:10.1109/tcsvt.2021.3081591
- Deng, R., Jin, X., and Du, D. (2022). 3d location and trajectory reconstruction of a moving object behind scattering media. *IEEE Trans. Comput. Imaging* 8, 371–384. doi:10.1109/tci.2022.3170651
- Dhami, H., Sharma, V., and Tokekar, P. (2023). Pred-nbv: prediction-guided next-best-view planning for 3d object reconstruction. *IEEE/RJS Int. Conf. Intelligent Robots Syst.*, 7149–7154. doi:10.1109/iros55552.2023.10341650
- fang Song, J., Fan, Y., Song, H., and Zhao, H. (2022). Target tracking and 3d trajectory reconstruction based on multicamera calibration. *J. Adv. Transp.* 2022, 1–8. doi:10.1155/2022/5006347
- Fathullah, Y., Xia, G., and Gales, M. J. (2023). Logit-based ensemble distribution distillation for robust autoregressive sequence uncertainties. *Uncertain. Artif. Intell. (PMLR)*, 582–591. Available online at: <https://proceedings.mlr.press/v216/fathullah23a.html>.
- Gebrehiwot, A. H., Hurych, D., Zimmermann, K., Pérez, P., and Svoboda, T. (2023). T-uda: temporal unsupervised domain adaptation in sequential point clouds in 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (IEEE), 7643–7650.
- González-Lezcano, R. A. (2023). Design of efficient and healthy buildings
- Gu, J., Hu, C., Zhang, T., Chen, X., Wang, Y., Wang, Y., et al. (2023). “Vip3d: end-to-end visual trajectory prediction via 3d agent queries,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 5496–5506.
- Hasheminasab, S., Zhou, T., and Habib, A. (2020). Gnss/ins-assisted structure from motion strategies for uav-based imagery over mechanized agricultural fields. *Remote Sens.* 12, 351. doi:10.3390/rs12030351
- Haznedar, B., Bayraktar, R., Ozturk, A. E., and Arayici, Y. (2023). Implementing pointnet for point cloud segmentation in the heritage context. *Herit. Sci.* 11 (2), 2. doi:10.1186/s40494-022-00844-w
- Heravi, M. Y., Jang, Y., Jeong, I., and Sarkar, S. (2024). Deep learning-based activity-aware 3d human motion trajectory prediction in construction. *Expert Syst. Appl.* 239, 122423. doi:10.1016/j.eswa.2023.122423
- Hu, A. V., and Kabala, Z. J. (2023). Predicting and reconstructing aerosol–cloud–precipitation interactions with physics-informed neural networks. *Atmosphere* 14, 1798. doi:10.3390/atmos14121798
- Hu, Z., Huang, J., Zhao, C., Jin, Q., Ma, Y., and Yang, B. (2020). Modeling dust sources, transport, and radiative effects at different altitudes over the Tibetan plateau. *Atmos. Chem. Phys.* 20, 1507–1529. doi:10.5194/acp-20-1507-2020
- Hu, Z., Jin, Q., Ma, Y., Pu, B., Ji, Z., Wang, Y., et al. (2021). Temporal evolution of aerosols and their extreme events in polluted asian regions during terra’s 20-year observations. *Remote Sens. Environ.* 263, 112541. doi:10.1016/j.rse.2021.112541
- Hu, Z., Zhao, C., Leung, L. R., Du, Q., Ma, Y., Hagos, S., et al. (2022). Characterizing the impact of atmospheric rivers on aerosols in the western us. *Geophys. Res. Lett.* 49, e2021GL096421. doi:10.1029/2021gl096421

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Correction note

A correction has been made to this article. Details can be found at: [10.3389/fenvs.2025.1741667](https://doi.org/10.3389/fenvs.2025.1741667).

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Karmakar, P., Pradhan, S., and Chakraborty, S. (2024). Indoor air quality dataset with activities of daily living in low to middle-income communities. *Adv. Neural Inf. Process. Syst.* 37, 70076–70100.
- Khurana, S., Dawalatabad, N., Laurent, A., Vicente, L., Gimeno, P., Mingote, V., et al. (2024). “Cross-lingual transfer learning for low-resource speech translation,” in *2024 IEEE international conference on acoustics, speech, and signal processing workshops (ICASSPW) (IEEE)*, 670–674.
- Li, J., and Li, W. (2022). “Auv 3d trajectory prediction based on cnn-lstm,” in *2022 IEEE International Conference on Mechatronics and Automation (ICMA)*, 1227–1232. doi:10.1109/icma54519.2022.9856366
- Li, N., and Su, B. (2021). Radar based obstacle detection in unstructured scene. *IEEE Intell. Veh. Symp. (IV)*, 770–776. doi:10.1109/iv48863.2021.9575280
- Li, C., Li, H., and Chen, K. (2024). Convolutional point transformer for semantic segmentation of sewer sonar point clouds. *Eng. Appl. Artif. Intell.* 138, 109456. doi:10.1016/j.engappai.2024.109456
- Liao, H., Wang, C., Li, Z., Li, Y., Wang, B., Li, G., et al. (2024). Physics-informed trajectory prediction for autonomous driving under missing observation. *Int. Jt. Conf. Artif. Intell.*, 6841–6849. doi:10.24963/ijcai.2024/756
- Liu, D., Li, W., Peng, J., and Ma, Q. (2022). The effect of banning fireworks on air quality in a heavily polluted city in northern China during Chinese spring festival. *Front. Environ. Sci.* 10, 872226. doi:10.3389/fenvs.2022.872226
- Mao, Y., Shen, B., Yang, Y., Wang, K., Xiong, R., Liao, Y., et al. (2024). v-dba: neural implicit dense bundle adjustment enables image-only driving scene reconstruction. *IEEE/RJS Int. Conf. Intelligent Robots Syst.*, 1130–1137. doi:10.1109/iros58592.2024.10801847
- Mérigoux, N. (2022). Multiphase eulerian-eulerian cfd supporting the nuclear safety demonstration. *Nucl. Eng. Des.* 397, 111914. doi:10.1016/j.nucengdes.2022.111914
- Moreau, A., Piasco, N., Tsishkou, D., Stanculescu, B., and de La Fortelle, A. (2022). “Lens: localization enhanced by nerf synthesis,” in *Conference on robot learning (PMLR)*, 1347–1356. Available online at: <https://proceedings.mlr.press/v164/moreau22a.html>.
- Nakamura, K., Hanari, T., Kawabata, K., and Baba, K. (2022). 3d reconstruction considering calculation time reduction for linear trajectory shooting and accuracy verification with simulator. *Artif. Life Robotics* 28, 352–360. doi:10.1007/s10015-022-00835-x
- Pandey, D., and Shu, T. (2024). “Am-dgcnn: leveraging graph attention networks and edge attributes for link classification in knowledge graphs,” in *SC24-W: workshops of the international conference for high performance computing, networking, storage and analysis (IEEE)*, 1037–1045.
- Raissi, M., Perdikaris, P., and Karniadakis, G. E. (2019). Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* 378, 686–707. doi:10.1016/j.jcp.2018.10.045
- Raissi, M., Perdikaris, P., Ahmadi, N., and Karniadakis, G. E. (2024). Physics-informed neural networks and extensions.
- Schröder, A., and Schanz, D. (2023). 3d Lagrangian particle tracking in fluid mechanics. *Annu. Rev. Fluid Mech.* 55, 511–540. doi:10.1146/annurev-fluid-031822-041721
- Shafiee, N., Padir, T., and Elhamifar, E. (2021). “Introvert: human trajectory prediction via conditional 3d attention,” in *Computer vision and pattern recognition*.
- Sharifi, A. A., Zoljodi, A., and Daneshlab, M. (2024). “Trajectorynas: a neural architecture search for trajectory prediction,” in *Italian National Conference on Sensors*, 5696. doi:10.3390/s24175696
- Tien, P. W., Wei, S., Darkwa, J., Wood, C., and Calautit, J. K. (2022). Machine learning and deep learning methods for enhancing building energy efficiency and indoor environmental quality—a review. *Energy AI* 10, 100198. doi:10.1016/j.egyai.2022.100198
- Yang, A., Tan, Q., Rajapakshe, C., Chin, M., and Yu, H. (2022). Global premature mortality by dust and pollution pm2.5 estimated from aerosol reanalysis of the modern-era retrospective analysis for research and applications, version 2. *Front. Environ. Sci.* 10, 975755. doi:10.3389/fenvs.2022.975755
- Yu, X., and Yang, H. (2023). Sim-sync: from certifiably optimal synchronization over the 3d similarity group to scene reconstruction with learned depth. *IEEE Robotics Automation Lett.* 9, 4471–4478. doi:10.1109/lra.2024.3377006
- Zekany, S. A., Dreslinski, R. G., and Wenisch, T. F. (2019). “Classifying ego-vehicle road maneuvers from dashcam video,” in *2019 IEEE intelligent transportation systems conference (ITSC) (IEEE)*, 1204–1210.
- Zhang, J., Yao, Y., and Quan, L. (2021). “Learning signed distance field for multi-view surface reconstruction,” in *Proceedings of the IEEE/CVF international conference on computer vision*, 6525–6534.
- Zhao, H., Gui, K., Ma, Y., Wang, Y., Wang, Y., Wang, H., et al. (2022). Effects of different aerosols on the air pollution and their relationship with meteorological parameters in north China plain. *Front. Environ. Sci.* 10, 814736. doi:10.3389/fenvs.2022.814736
- Zhong, J., Sun, H., Cao, W., and He, Z. (2020). Pedestrian motion trajectory prediction with stereo-based 3d deep pose estimation and trajectory learning. *IEEE Access* 8, 23480–23486. doi:10.1109/access.2020.2969994
- Zhou, J., Zhang, H., Lyu, W., Wan, J., Zhang, J., and Song, W. (2022). Hybrid 4-dimensional trajectory prediction model, based on the reconstruction of prediction time span for aircraft en route. *Sustainability* 14, 3862. doi:10.3390/su14073862