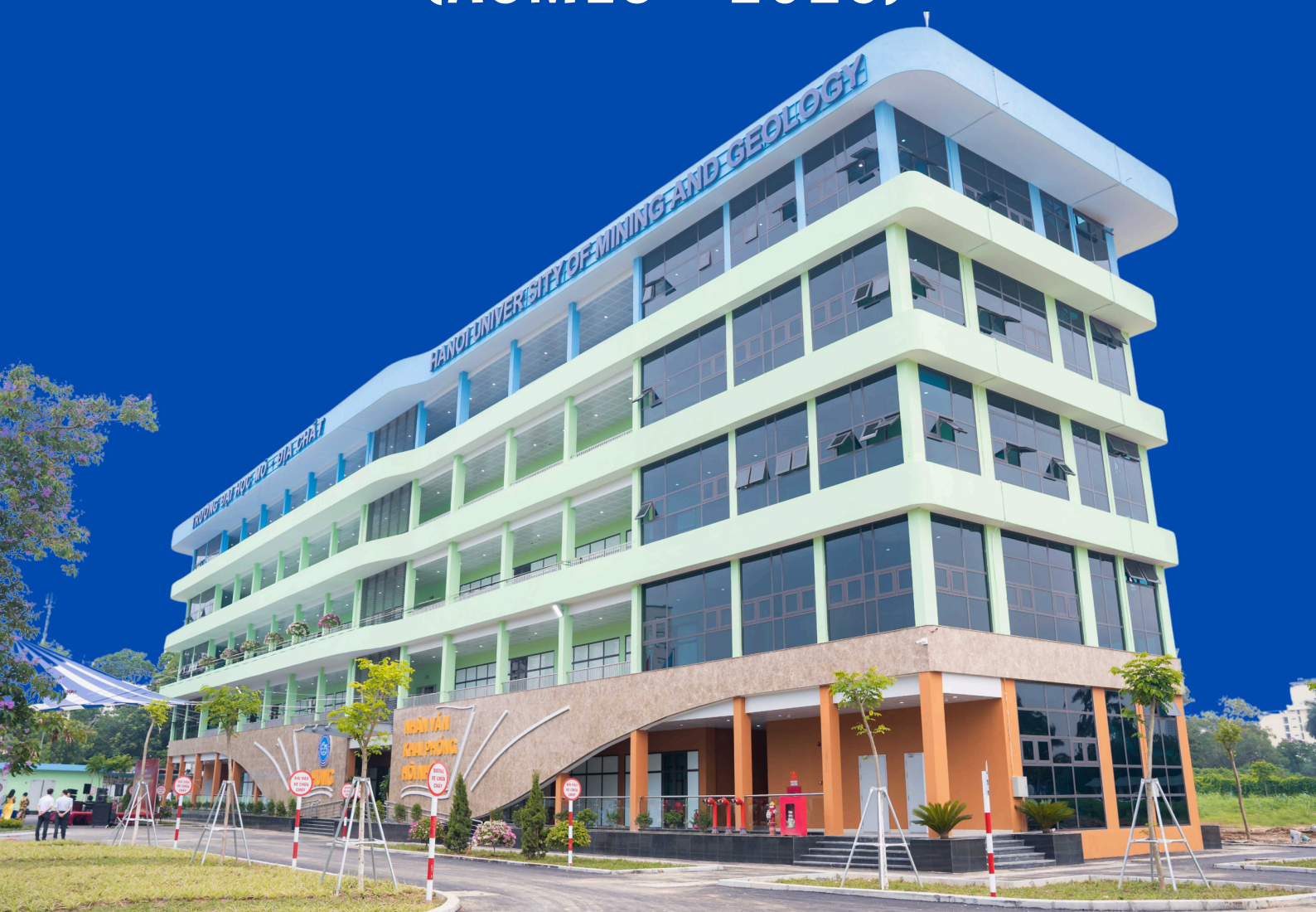


PROCEEDINGS OF THE INTERNATIONAL WORKSHOP ON ADVANCES IN
SURFACE MINING FOR ENVIRONMENTAL PROTECTION AND SUSTAINABLE
DEVELOPMENT 17 OCTOBER 2025, HA NOI, VIET NAM



ADVANCES IN SURFACE MINING FOR ENVIRONMENTAL PROTECTION AND SUSTAINABLE DEVELOPMENT-2025 (ASMES - 2025)



TRANSPORT PUBLISHING HOUSE, HA NOI, VIET NAM

PROCEEDINGS OF THE INTERNATIONAL WORKSHOP ON ADVANCES IN
SURFACE MINING FOR ENVIRONMENTAL PROTECTION AND SUSTAINABLE
DEVELOPMENT 17 OCTOBER 2025, HA NOI, VIET NAM

ADVANCES IN SURFACE MINING FOR ENVIRONMENTAL PROTECTION AND SUSTAINABLE DEVELOPMENT-2025 (ASMES - 2025)



Copyright @ 2025 Hanoi University of Mining and Geology, Viet Nam

All rights reserved. No part of this publication or the information contained herein may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, by photocopying, recording or otherwise, without written prior permission from the publisher.

Although all care is taken to ensure the integrity and quality of this publication and the information herein, no responsibility is assumed by the publishers nor the author for any damage to property or person as a result of operation or use of this publication and/or the information contained herein.

Published by: Transport Publishing House, Ha Noi, Viet Nam, 2025

ISBN: 978-604-76-3193-3

Printed in Viet Nam

MỤC LỤC

TT	TẬP THỂ TÁC GIẢ	TÊN BÀI BÁO	TRANG
1.	Bùi Xuân Nam, Trần Quang Hiếu, Nguyễn Anh Tuấn, Đỗ Ngọc Hoàn, Nguyễn Phú Vụ, Lê Văn Quyển	Định hướng đào tạo và nghiên cứu khoa học ngành Khai thác mỏ Lộ thiên: Vững bước hội nhập, vươn tầm kỷ nguyên mới	1-14
2.	Nguyen Anh Tuan, Pham Van Viet, Pham Van Hoa, Do Ngoc Hoan, Tran Dinh Bao,	Stone in Vietnam: Building stone, rock fill and armourstone in construction for Sustainable Development	15-23
3.	Xuan-Nam Bui, Carsten Drebenstedt, Nguyen Hoang, Nguyen Van Duc, Tran Quang Hieu, Nguyen Quoc Long, Le Ngoc Ninh	Smart Mining - Opportunities and Challenges for Vietnam's Mining Industry	24-41
4.	Lai Hong Thanh, Mai The Toan, Le Ngoc Ninh, Xuan-Nam Bui	Vietnam's 2024 Law on Geology and Minerals: Opportunities and Challenges for the Mining Industry	42-55
5.	Le Thi Thu Hoa, Nguyen Hoang, Tran Quang Hieu, Tran Dinh Bao, Le Qui Thao, Nguyen Truc Anh, Atrushkevich Victor, Xuan-Nam Bui	Some Technological Approaches, Optimization Strategies, and Sustainable Development Directions for Selective Mining in Vietnam's Open-Pit Coal Mines	56-66
6.	Le Qui Thao, Do Ngoc Le Dung, Nguyen Thi Thuc Anh, Do Ngoc Hoan, Xuan-Nam Bui	Some Technological Orientations for Safe and Environmentally Friendly Mining Technology of Coastal Placer Titanium Deposits in Vietnam	67-83
7.	Tran Quang Hieu, Nguyen Van Duc, Xuan-Nam Bui, Trinh Hai Son ² , Bui Hoang Bac	Sustainable Rare Earth Mining Technologies: Global Trends and a Tailored Model for Vietnam's Development	84-98
8.	Tran Quang Hieu, Bui Xuan Nam, Le Ngoc Ninh, Tran Van Dung, Do Duc Khoa	The management and utilization of industrial explosives in Vietnam: Achievements, opportunities, challenges, and future development orientations.	99-110
9.	Thang Dam Trong, Duc Ngo The	Experimental investigation on the influence of water saturation on the propagation velocity of blasting stress waves	111-123
10.	Nguyen Dinh An , Tran Dinh Bao	Controlling Harmful Impacts of Rock Blasting for Construction Site Preparation in Ha Giang Province, Vietnam	124-136
11.	Ngô Ngọc Thủy, Đàm Trọng Thắng	A Study on Mitigating Air Shock Wave Intensity in Surface Blasting for Rock Fragmentation Using Sand Cover Layers	137-148

12.	Van-Viet Pham, Pirat Jaroontattanapong, Nguyen Anh Tuan, Van -Hoa Pham	Prediction of an appropriated powder factor used for blasting under the consideration of blast direction and rock mass quality	149-163
13.	Vũ Tùng Lâm, Trần Quang Hiếu, Vũ Trọng Hiếu	Nghiên cứu phân tích, lựa chọn phương pháp hợp lý để xác định mức độ đập vỡ đất đá từ số liệu thực nghiệm	164-175
14.	Do Ngoc Hoan, Tran Quang Hieu, Tran Dinh Bao	Achievements and Challenges in Occupational Safety and Health Management in the Mineral Mining industry in Vietnam	176-189
15.	Lam Kieu Anh, Tran Quang Hieu, Do Ngoc Hoan	Assessment of Fire and Explosion Risks at Industrial Explosives Storage Facilities in Vietnam	190-209
16.	Do Van Trieu, Tran Dinh Bao, Dang Tien Anh, Dinh Ba Lam, Nguyen Van Hoang	Overview of Carbon Storage Technologies in Mine Waste Rock: Opportunities and Challenges for TKV Open-Pit Mines	210-223
17.	Đỗ Thị Lan Chi Vũ Văn Thú	Safety Culture In Enterprises: An Assessment And Intervention Framework Based On A Multi-Model Approach	224-233
18.	Xuan-Nam Bui, Thi-Tam Nguyen, Ngoc-Ninh Le, Duc-Tin Quach, Quang-Hieu Tran, Chang-Woo Lee, Van-Trieu Do, Van-Duc Nguyen	Decarbonizing Vietnam's Open-Pit Coal Mining: A Life Cycle Case Study of Emissions and Mitigation Pathways	234-248
19.	Tuan-Ngoc Le, Hoang Nguyen	Predicting PM2.5 Concentrations in an Open-Pit Copper Mine Using Advanced AI Techniques: Insights from the Sin Quyen Copper Mine, Vietnam	249-268
20.	Léopold Dupuy, Hoang Nguyen	Assessment and Prediction of Carbon Sequestration in Rehabilitated Open-Pit Mining Areas Using Remote Sensing and AI Techniques: A Case Study from Northern Vietnam	269-290
21.	Tran Dinh Bao, Vu Van Thu	Analysis, evaluation, and proposal of solutions to enhance the effectiveness of occupational safety and health practices in Vietnam in the current period	291-312
22.	Do Duc Thanh, Ngo Thanh Tuan	Impact of large capacity solar power sources when connected near high-voltage power system supplying electricity to mining areas	313-320
23.	Nguyen Hoang Long, Bui Xuan Nam, Tran Thi Hai Van, Dinh Bao Ngoc, Tran Truong Giang,	Development of a Management and Monitoring System for Production Activities in Open-Pit Quarries in Vietnam	321-328
24.	Nguyễn Mạnh Thường, Đàm Trọng Thắng, Tạ Quang Yên	Исследования, изготовление и испытание сборной плавучей рабочей платформы для подводного бурения	329-342

		шахтных скважин подводных взрывных работ в морских условиях в провинции Куангнинь	
25.	Nguyễn Văn Tính, Đàm Trọng Thắng, Lê Sỹ Thượng, Trần Quang Hiếu	Исследование расчета, проектирования и изготовления разумных малых зарядов взрывчатых веществ (ВВ) для экспериментальных работ по взрыву грунта и горных пород на миниатюрных моделях	343-358
26.	Нгуен В.Х, Дам Ч.ТХ, Нгуен В.И, Татаринов В.В, Чан В.Л	Экспериментальное исследование решений по снижению шума в модели электрического взрывного устройства	359-374
27.	Нгуен Мань Тхыонг, Дам Чонг Тханг, Та Куанг Йен	Исследования, изготовление и испытание сборной плавучей рабочей платформы для подводного бурения шахтных скважин подводных взрывных работ в морских условиях в провинции Куангнинь.	375-385
28.	Nguyễn Anh Tuấn, Phạm Văn Việt, Đỗ Ngọc Hoàn, Trần Quang Hiếu	Đề xuất bộ tiêu chí trong khai thác lộ thiên khoáng sản rắn đảm bảo hiệu quả, bền vững và tích ứng với biến đổi khí hậu ở Việt Nam	386-395
29.	Trần Quang Hiếu, Nguyễn Anh Tuấn, Đỗ Ngọc Hoàn, Bùi Xuân Nam, Lê Ngọc Ninh, Nguyễn Văn Đức, Lại Hồng Thanh	Tương lai ngành khai thác mỏ lộ thiên tại Việt Nam: Cơ hội, thách thức và các định hướng phát triển bền vững	396-412
30.	Lê Đức Phương, Đỗ Ngọc Hoàn	Phương pháp xác định biên giới kết thúc khai thác lộ thiên hợp lý cho các mỏ quặng đơn và đa khoáng thuộc TKV	413-422
31.	Nguyễn Anh Tuan, Pham Van Viet, Tran Trung Anh, Nguyen Van Tinh, Tran Dinh Bao, Pham Minh Tu, Noelie Escutenaire	Trung hòa chất lượng quặng nguyên khai: Áp dụng điều kiện khai thác khối 24-122, mỏ Tây Tân Rai, Tây Nguyên	423-431
32.	Nguyễn Anh Tuấn, Phạm Văn Việt, Trần Trung Anh, Trần Đình Bảo, Trần Đăng Hòa, Phạm Minh Tú, Noelie Escutenaire	Mô hình hàm lượng quặng bauxite: xây dựng bản đồ phân bố trữ lượng và hàm lượng quặng bauxite cho khối 24-122, mỏ Tây Tân Rai	432-440
33.	Khuong The Hung, Bui Hoang Bac, Vu The Thu	Sustainable development of apatite ore at the Coc mine, Lao Cai province based on a comprehensive assessment of geological conditions and mining status	441-447
34.	Lê Bá Phúc, Đỗ Văn Triều, Đặng Tiến Anh, Trần Đình Bảo	Nghiên cứu xây dựng phương pháp giám sát mô hình chất tải trên các ô tô vận chuyển đất đá tại các mỏ lộ thiên phục vụ chuyển đổi số	449-460
35.	Đỗ Ngọc Tước, Nhữ Việt Tuấn, Trịnh thị Phượng	Đánh giá tiềm năng khí hóa than ngầm bằng vi sinh tại bể than Đồng bằng Sông Hồng	461-470

36.	Đàm Quang Sang, Nguyễn Văn Tính, Hoàng Trung Hữu, Nguyễn Ngàn, Trần Văn Dũng, Trần Quang Hiếu	Phương pháp xác định các thông số của môi nổ dùng cho nổ mìn trên các mỏ lộ thiên ở Việt Nam	471-480
37.	Phạm Văn Hòa, Trần Văn Dũng, Hoàng Anh Phong, Trịnh Quang Trung, Lê Văn Quyển, Dương Thành Trung, Trần Đình Bảo, Trần Quang Hiếu, Nguyễn Anh Tuấn, Nguyễn Đình An, Phạm Văn Việt	Hệ thống quản lý công tác khoan nổ mìn – đáp ứng yêu cầu chuyển đổi số và phù hợp điều kiện thực tế cho các mỏ than lộ thiên vùng Quảng Ninh	481-491
38.	Đoàn Văn Thanh, Đỗ Văn Triều, Đinh Bá Lâm, Trần Đình Bảo	Nghiên cứu mô phỏng lan truyền sóng ứng suất nổ mìn trong điều kiện đất đá ngậm nước tại các mỏ than thuộc TKV	492-501
39.	Hoàng Mạnh Thắng, Đoàn Văn Thanh, Đinh Bá Lâm, Đặng Tiến Anh	Nghiên cứu một số phương pháp phá vỡ đất đá không sử dụng thuốc nổ cho các mỏ lộ thiên tại Việt Nam	502-511
40.	Lưu Văn Thực, Đỗ Văn Triều, Lê Bá Phúc, Lê Thị Thu Hoa	Nghiên cứu xây dựng cơ sở dữ liệu quan trắc dịch động và chương trình cảnh báo ổn định bờ mỏ, bãi thải bằng mô hình số kết hợp trí tuệ nhân tạo (AI) cho các mỏ lộ thiên thuộc TKV	512-521
41.	Đào Văn Chi, Vũ Thái Tiến Dũng, Phan Mai Linh	Kinh nghiệm quốc tế về lồng ghép bảo vệ môi trường trong các hoạt động khoáng sản và bài học cho Việt Nam	522-528
42.	Phạm Hải Anh, Nguyễn Anh Tuấn, Trịnh Văn Hiệp, Phạm Hồng Thái, Đặng Anh Đức	Tiềm năng ứng dụng công nghệ thu hồi, sử dụng và lưu trữ Carbon (CCUS) tại các nhà máy nhiệt điện thuộc TKV (Kỳ I)	529-539
43.	Nguyễn Anh Tuấn, Nguyễn Xuân Tú, Phạm Văn Việt, Đỗ Anh Sơn, Ngô Thanh Tuấn, Nguyễn Tiến Hưng.	Nghiên cứu đánh giá sự ảnh hưởng của moong khai thác lộ thiên đến các công trình hầm mỏ: Khai thác than mỡ khu Bắc Làng Cẩm, Thái Nguyên	540-557
44.	Nguyễn Đình An, Nhữ Văn Bách, Trần Đình Bảo	Ứng dụng trí tuệ nhân tạo lập hộ chiếu nổ mìn theo thời gian thực nhằm nâng cao hiệu quả và đảm bảo an toàn nổ mìn trong khai thác mỏ	558-565



ASMES - 2025

www.khaithacloathien.edu.vn



Predicting PM_{2.5} Concentrations in an Open-Pit Copper Mine Using Advanced AI Techniques: Insights from the Sin Quyen Copper Mine, Vietnam

Tuan-Ngoc Le¹, Nguyen Truc Anh², Hoang Nguyen^{3,4,*}, Xuan-Nam Bui^{5,6}

¹ Vinacomin – Minerals Holding Corporation, 193 Nguyen Huy Tuong Str., Thanh Xuan Trung ward, Thanh Xuan Dist., Hanoi, 100000, Vietnam

² Mining chemical industry holding corporation limited – Vinacomin, Hanoi 100000, Vietnam

³ Department of Surface Mining, Mining Faculty, Hanoi University of Mining and Geology, 18 Pho Vien, Dong Ngac Ward, Hanoi 100000, Vietnam

⁴ Innovations for Sustainable and Responsible Mining (ISRM) Research Group, Hanoi University of Mining and Geology, 18 Pho Vien, Dong Ngac Ward, Hanoi 100000, Vietnam

⁵ Laboratory of Sustainable Development in Natural Resources and Environment, Institute for Advanced Study in Technology, Ton Duc Thang University, Ho Chi Minh City, Vietnam

⁶ Faculty of Environment and Labour Safety, Ton Duc Thang University, Ho Chi Minh City, Vietnam

ABSTRACT

Accurate forecasting of PM_{2.5} concentrations in mining areas is vital for occupational health and environmental management. This study explores the application of advanced Artificial Intelligence (AI) models to predict PM_{2.5} levels at the Sin Quyen open-pit copper mine in northern Vietnam. A comprehensive dataset was constructed from historical PM_{2.5} measurements, meteorological variables (e.g., temperature, humidity, wind speed), and mine operational parameters. Several AI models—including Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), eXtreme Gradient Boosting (XGBoost), and Temporal Convolutional Networks (TCN)—were implemented and compared. Performance was evaluated using RMSE, MAE, and R² metrics. Among the tested models, TCN and LSTM demonstrated superior forecasting capabilities, capturing complex temporal dependencies in the dataset. These findings highlight the potential of AI-driven forecasting frameworks in enhancing air quality management and supporting sustainable mining practices.

Keywords: PM_{2.5} forecasting, Artificial Intelligence, Open-pit mining, Sin Quyen copper mine, LSTM, GRU

*Corresponding author.

E-mail address: nguyenhoang@humg.edu.vn

1. Introduction

Air pollution caused by fine particulate matter (PM_{2.5}) has become a major environmental and occupational health challenge, especially in open-pit mining regions (Rojano et al., 2018). PM_{2.5} particles, with aerodynamic diameters less than 2.5 micrometers, can penetrate deep into the human respiratory system, leading to severe cardiovascular and respiratory diseases (Chauhan et al., 2024). In open-pit mines, PM_{2.5} is generated from a combination of activities such as blasting, drilling, excavation, material handling, and vehicular movement on haul roads. Its concentration and dispersion are further influenced by meteorological factors, terrain morphology, and seasonal patterns.

Forecasting PM_{2.5} concentrations in mining environments is critical for enabling proactive air quality management, minimizing exposure risks for workers, and supporting regulatory compliance (Sanyal 2024; Saka et al., 2025). However, traditional statistical and deterministic modeling approaches—such as autoregressive integrated moving average (ARIMA), multiple linear regression (MLR), and Gaussian dispersion models—are often limited by their assumptions of linearity and stationarity (Talepour et al., 2024; Nourmohammad and Rashidi 2025). These models struggle to capture the nonlinear, dynamic, and multi-factorial nature of air quality data in the mining context.

In recent years, Artificial Intelligence (AI) and machine learning (ML) techniques have shown great potential in environmental modeling, especially for air pollution forecasting. Several studies have applied models such as Support Vector Regression (SVR), Artificial Neural Networks (ANN), and Decision Trees to predict PM concentrations in urban or industrial settings (Bai et al., 2018; Chang et al., 2020; Masood and Ahmad 2021). For example, Kleine Deters et al., (2017) demonstrated the ability of Boosted Trees (BTs) and Linear Support Vector Machines (L-SVM) to predict urban PM_{2.5} using meteorological and traffic data. Similarly, Marinov et al., (2022) applied deep learning techniques for air quality forecasting in large cities, showing improved accuracy over traditional methods. However, research focusing specifically on PM_{2.5} forecasting in open-pit mining environments remains limited, despite the unique emission profiles and operating conditions found in these areas.

Recent research has increasingly focused on applying modeling and AI-based approaches to air quality prediction in mining regions. Asif et al., (2018) developed the Mining Air Dispersion Model (MADM), an extension of Gaussian plume equations that explicitly incorporates deposition effects, and demonstrated its strong agreement with field measurements of PM_{2.5}, PM₁₀, NO₂, and heavy metals in an open-pit mine in Canada. More recently, Kamran et al., (2024) introduced a decision intelligence-driven framework that combined K-means clustering with random forests to predict air quality index (AQI) levels at the Haerwusu open-pit coal mine in China, achieving a remarkable classification accuracy of 97%. Complementary efforts by Wang et al., (2023) emphasized the integration of meteorological parameters and production intensity, showing that including operational factors such as rock and coal output substantially improved dust concentration forecasts in open-pit mines. Similarly, Chaulya et al., (2025) in India coupled air quality modeling with eco-friendly mitigation strategies, forecasting PM_{2.5}, PM₁₀, SO₂, and NO_x levels under expansion scenarios and evaluating the effectiveness of intelligent dust suppression systems.

Taken together, these studies highlight the growing recognition of AI and data-driven modeling as essential tools for mining-related environmental management. However, most

existing works emphasize either physical dispersion modeling (e.g., MADM) or single AI pipelines (e.g., RF-based AQI classification), while relatively few have systematically compared multiple machine learning and deep learning architectures. Moreover, the comprehensive integration of meteorological, operational, and topographic drivers into unified forecasting frameworks remains limited, particularly in open-pit mine contexts where spatiotemporal variability is high. This gap underscores the need for comparative, multi-model approaches that leverage both AI and domain-specific knowledge to improve the reliability of mine-site air quality forecasting.

This study addresses these gaps by conducting a comprehensive comparative analysis of four state-of-the-art AI models, including Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), eXtreme Gradient Boosting (XGBoost), Temporal Convolutional Networks (TCN). These models were chosen for their complementary strengths in capturing temporal dependencies (LSTM, GRU, TCN) and nonlinear variable interactions (XGBoost). By evaluating these architectures side-by-side on the same dataset, the study offers critical insights into their relative performance and suitability for mining-specific air quality forecasting.

A dataset was constructed using historical PM_{2.5} measurements from a real-world case study—the Sin Quyen open-pit copper mine in northern Vietnam—alongside meteorological variables (e.g., temperature, wind speed, wind direction, humidity). This multi-source integration enhances the representativeness and robustness of the modeling task.

The major contributions of this study are as follows:

1. Integration of diverse data sources (air quality and meteorology) to build a forecasting dataset tailored for open-pit mining environments.
2. Comparative evaluation of four advanced AI models (LSTM, GRU, XGBoost, and TCN) for PM_{2.5} forecasting, highlighting their strengths and limitations in handling temporal, nonlinear, and multivariate dependencies.
3. Demonstration of the suitability of TCN and LSTM models for capturing long-term dependencies in PM_{2.5} dynamics, with higher accuracy than tree-based models in this domain-specific application.
4. Provision of a practical AI-driven forecasting framework that can be integrated into real-time monitoring and early-warning systems for sustainable mine management.

This study contributes to the growing body of knowledge in AI-based environmental monitoring, especially within the underexplored context of open-pit mining. The findings offer not only methodological insights for future researchers but also actionable guidance for mine operators and policymakers aiming to improve air quality and protect occupational health.

2. Methodology

2.1. Long Short-Term Memory (LSTM)

Forecasting PM_{2.5} concentrations in open-pit mines is crucial for ensuring environmental safety and worker health. Due to the complex and dynamic nature of air quality in mining environments—impacted by blasting operations, haul road activities, meteorological factors, and terrain—traditional statistical models often fall short in accuracy.

Long Short-Term Memory (LSTM), a specialized form of Recurrent Neural Network (RNN), is well-suited for modeling such time-series problems thanks to its ability to learn long-term dependencies and nonlinear relationships (Abotaleb and Dutta 2024a).

LSTM is designed to overcome the vanishing gradient problem found in traditional RNNs (Hu et al., 2018). It introduces memory cells and gating mechanisms that allow the network to learn which information to remember, update, or forget over time.

Each LSTM cell contains:

- Forget gate:

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

This parameter aims to determine what portion of the previous cell state C_{t-1} to forget.

- Input gate:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (3)$$

This parameter controls which new information is stored in the cell.

- Cell state update:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (4)$$

- Output update:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = o_t * \tanh(C_t) \quad (6)$$

This step decides the output at current time step and what is passed to the next time step.

To forecast PM_{2.5} concentrations accurately in an open-pit mining context, the LSTM model typically uses multivariate time-series data, including: Historical PM_{2.5} concentrations; Meteorological factors: temperature, humidity, wind speed, wind direction, atmospheric pressure; Mining activities: blasting schedule, hauling operations, dust suppression activities; Topography and height of monitoring stations.

For forecasting air quality in open-pit mines, LSTM have many advantages, such as captures both short- and long-term dependencies in PM_{2.5} data (using time dependency feature); Models complex, nonlinear interactions between variables (using nonlinearity feature); Performs well with real-world noisy sensor data (using noise resistance feature); Can interpolate or learn despite missing intervals in time-series (using missing data tolerance feature); Learns lag effects from activities like blasting or rainfall on PM_{2.5} levels (using sequential impact feature).

In short, LSTM networks provide a powerful approach for forecasting PM_{2.5} in open-pit mining environments, where air quality is influenced by multiple dynamic and interrelated factors. Their ability to model temporal dependencies and adapt to real-world data makes them especially valuable for environmental monitoring, prediction, and early warning systems in the mining industry.

2.2. Gated Recurrent Unit (GRU)

Gated Recurrent Unit (GRU) is a streamlined variant of the Recurrent Neural Network (RNN), capable of modeling temporal dependencies while being computationally more efficient than LSTM (Abotaleb and Dutta 2024b), making it an excellent choice for real-time PM_{2.5} prediction.

GRU simplifies the LSTM architecture by combining the forget and input gates into a single update gate and using a reset gate to control memory (Zargar 2021). This reduces computational cost while maintaining comparable performance on many sequence learning tasks.

The key components of the GRU model are difference from the LSTM mode, and in each GRU cell contains the following:

- Update gate (z_t) helps to control how much of the past information needs to be passed along to the future.

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t] + b_z) \quad (7)$$

- Reset gate (r_t) helps to determine how much past information to forget.

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t] + b_r) \quad (8)$$

- Candidate activation ($\tilde{h}_t$) proposed memory content.

$$\tilde{h}_t = \tanh(W_h \cdot [r_t * h_{t-1}, x_t] + b_h) \quad (9)$$

- Final memory at time t (h_t) aims to interpolate between previous memory and new candidate.

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \quad (10)$$

The main advantages of the GRU model for forecasting air quality in open-pit mines are fewer gates than LSTM; faster training since it has simpler architecture; Suitable for edge devices or embedded systems in mining sites (using the efficiency feature); Maintains ability to capture complex time-series dependencies (using strong temporal learning feature); Ideal for deployment in near real-time systems (using the real-time forecasting feature); Learns from varied environmental and operational data (using the nonlinear patterns handling feature).

In conclusion, the GRU can offers an efficient and accurate solution for PM_{2.5} forecasting in open-pit mines. It strikes a balance between performance and computational cost, making it suitable for real-time applications. Its simplicity compared to LSTM enables easier tuning and faster deployment, especially in edge-computing environments like remote mines.

2.3. eXtreme Gradient Boosting (XGBoost)

Forecasting PM_{2.5} concentrations in open-pit mining environments is vital for minimizing health risks and complying with environmental standards. Unlike deep learning methods like LSTM or GRU, which are ideal for sequential data, XGBoost is a powerful machine learning algorithm known for its high accuracy, robustness, and scalability on tabular data (Shwartz-Ziv and Armon 2022).

XGBoost is particularly well-suited for modeling nonlinear relationships and complex interactions between environmental, operational, and meteorological variables affecting PM_{2.5}.

XGBoost stands for eXtreme Gradient Boosting, an optimized implementation of the gradient boosting framework (Trizoglou et al., 2021). It builds an ensemble of decision trees sequentially, where each new tree corrects the errors made by the previous ones.

The main concept of this model is using summing predictions from a series of decision trees to predict an output, using the Eq. (11):

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), \quad f_k \in F \quad (11)$$

where F is the space of regression trees, and f_k denotes the prediction function of the k^{th} tree.

In the XGBoost model, each tree is trained to minimize the regularized loss function using the Eq. (12):

$$L = \sum_i l(y_i, \hat{y}_i) + \sum_k \Omega(f_k) \quad (12)$$

where $\Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|^2$, with T = number of leaves and w = leaf weights.

The boosting stage is performed additively, using gradient descent to optimize the loss function.

The XGBoost models also does not require sequential data and works well with historical data, weather data, mining operations, time-related features, and topographic data.

For these reasons, the XGBoost model has many advantages, like captures complex relationships between variables (using the nonlinearity feature); Handling outliers and missing data efficiently; Very fast training and inference; Provides interpretable feature contributions; Works well even with limited data compared to deep learning.

In short, XGBoost is a high-performance algorithm for PM_{2.5} forecasting in open-pit mines when sufficient historical data and contextual variables are available. Its speed, accuracy, and interpretability make it ideal for early-warning systems, operational planning, and regulatory compliance. While it doesn't inherently model sequence data like LSTM or GRU, its strong predictive power and feature importance analysis offer clear advantages in structured mining environments.

2.4. Temporal Convolutional Networks (TCN)

Temporal Convolutional Networks (TCN) offer a powerful alternative to recurrent models (e.g., LSTM, GRU) by leveraging dilated causal convolutions to capture long-range dependencies in time series (Gu et al., 2024). TCNs combine the strengths of convolutional neural networks (CNNs) and sequence modeling, making them well-suited for fast and accurate forecasting of PM_{2.5} in mining environments.

A TCN is a type of convolutional architecture designed specifically for sequence modeling. It ensures that predictions at time t are only influenced by inputs from time $\leq t$ (causality), and it uses dilated convolutions to cover a wide receptive field efficiently.

The core components of the TCN model include causal convolutions, dilated convolutions, and residual connections (Luo et al., 2021). Causal convolutions aim to ensure

no future information leaks into the past. Output at time t_t depends only on inputs from t_t and earlier. Dilated convolutions aim to introduce gaps in the convolutional filter to allow the network to cover longer time horizons without increasing model depth.

For example, dilation $d=2$ $d = 2$:

$$y_t = w_0x_t + w_1x_{t-2} + w_2x_{t-4} \quad (13)$$

Residual connections help in training deeper networks by allowing gradient flow and avoiding vanishing gradients.

The main advantages of the TCN model include the learning temporal patterns over long sequences efficiently using dilated convolutions, allowing faster training compared to sequential RNN models, avoiding vanishing/exploding gradient problems common in RNNs, adjustable to match specific forecast horizons, competitive or better performance than LSTM in many time-series benchmarks.

In summary, TCN provides a robust and scalable architecture for PM_{2.5} forecasting in open-pit mining environments. They combine the ability to model temporal dependencies with the efficiency of convolutional operations, making them an ideal solution for long-range, real-time environmental forecasting tasks. With proper tuning and feature engineering, TCNs can outperform or match traditional recurrent and tree-based models in accuracy and speed.

3. Dataset

3.1. Study site and data description

The study site is located within the Sin Quyen open-pit copper mine, situated in Lao Cai Province, northern Vietnam, approximately 300 km northwest of Hanoi. The mine lies within a mountainous region near the Vietnam–China border, characterized by a humid subtropical climate with distinct wet (May–September) and dry (October–April) seasons. The surrounding terrain is complex, with elevations ranging from 100 m to over 1,500 m above sea level, influencing local wind and atmospheric dispersion patterns.

A fixed environmental monitoring station was installed on the mine premises at an elevation of +102 meters relative to the mine pit base (Fig. 1). The location was strategically chosen to provide continuous measurements representative of ambient air quality in active mining areas; capture pollutant dispersion from multiple emission sources within the pit and along haul roads, and maintain accessibility for sensor maintenance and calibration.

The monitoring station is equipped with meteorological tower (mast height: ~10 m above ground level) for measuring temperature, humidity, pressure, wind speed, and wind direction, with air quality module housing the PM_{2.5} optical particle counter in a weatherproof enclosure with continuous power supply, and with data acquisition and transmission system for real-time data logging and remote retrieval.

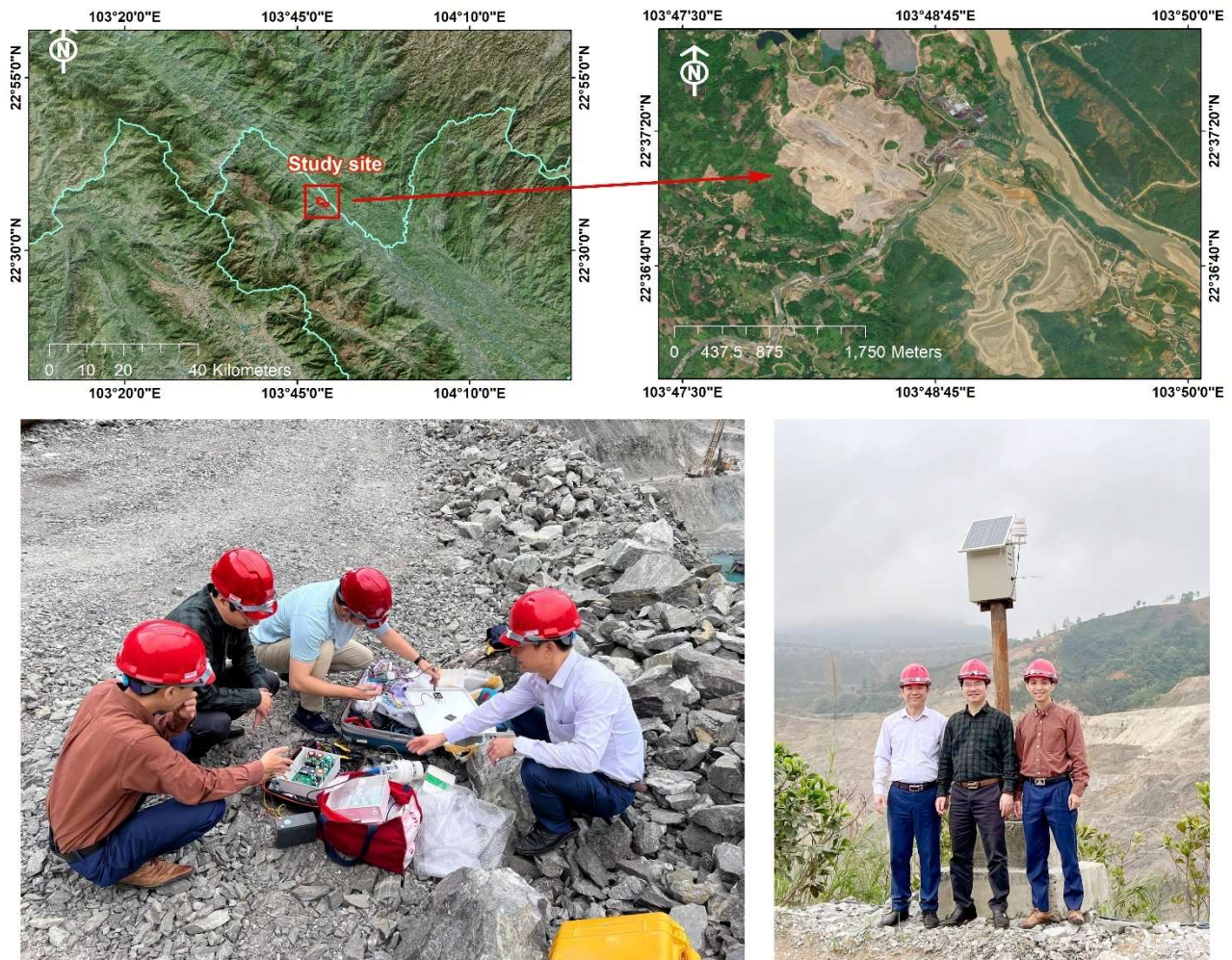


Figure 1. Location of the Sin Quyen monitoring station within the open-pit mine, Lao Cai Province, Vietnam

The raw dataset covers 102,410 consecutive records, spanning a continuous monitoring period at 5-minute sampling intervals. This high temporal resolution enables detection of rapid pollutant concentration spikes caused by short-term emission events (e.g., blasting, vehicle movement). It also provide fine-grained meteorological monitoring to study microclimate effects on pollutant dispersion.

For the purposes of $PM_{2.5}$ forecasting, the dataset was aggregated to hourly averages to reduce high-frequency noise caused by sensor fluctuations or short-lived micro-events and align with standard operational forecasting horizons used in regulatory and industrial air quality management.

Table 1. Summary of dataset variables

Category	Variable	Description	Unit	Sensor type / principle
Meteorological	temperature	Ambient air temperature	°C	Platinum resistance thermometer (PRT)
	humidity	Relative humidity of air	%	Capacitive humidity sensor

	pressure	Atmospheric pressure	hPa	Digital barometer
	windSpeed	Wind speed at monitoring height	m/s	Cup anemometer
	windDirection	Wind direction clockwise from north	degrees (°)	Wind vane
Air quality	dust25	Fine particulate matter $\leq 2.5 \mu\text{m}$	$\mu\text{g}/\text{m}^3$	Optical particle counter (light scattering)
Temporal index	timestamp	Observation date and time	—	System clock (synchronized)

3.2. Data preprocessing

3.2.1. Time alignment and resampling

The original dataset, recorded at 5-minute intervals, was reindexed and resampled to an hourly resolution to reduce high-frequency variability and to align with standard operational practices in air quality forecasting. This conversion ensured that each hourly timestamp contained synchronized measurements for all variables, allowing seamless integration of meteorological and pollutant data. The resampling was performed by computing the arithmetic mean of all observations falling within each hour. This approach preserved the diurnal and seasonal trends of interest while smoothing out transient micro-events and sensor noise that could otherwise hinder model generalization.

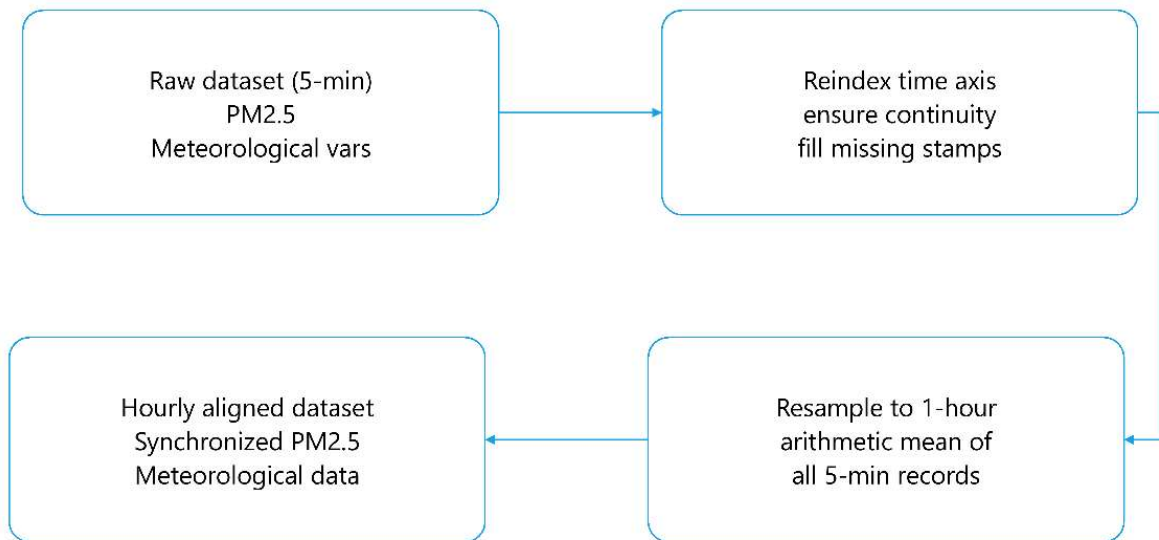


Figure 2. Flowchart of data preprocessing steps

3.2.2. Missing value assessment

After resampling, the dataset was systematically checked for missing entries across all variables. The inspection confirmed that the hourly dataset was complete, with no missing

values. This characteristic eliminates the need for data imputation and ensures that the models are trained solely on observed data, thus avoiding uncertainties introduced by synthetic estimations. The completeness of the dataset also reflects the reliability of the monitoring station's hardware and the stability of its data acquisition system.

3.2.3. Outlier review

Summary statistics and range checks were performed for each variable to assess the plausibility of the recorded values. While certain extreme $PM_{2.5}$ concentrations were detected—often corresponding to intense mining activities such as blasting, ore hauling, or strong wind-driven dust events—these data points were retained. The retention of extreme values was deemed essential, as they represent critical air quality exceedance events that the forecasting models must learn to identify and predict. Such peaks carry operational importance for real-time environmental management in open-pit mining contexts.

Table 2. Summary statistics of hourly dataset

Feature	count	mean	std	min
temperature	8535	17.151	7.239	2.601
humidity	8535	76.595	13.525	50.460
pressure	8535	100040.753	14.463	99997.179
windSpeed	8535	1.515	0.706	0.096
windDirection	8535	179.784	30.099	78.363
dust25	8535	56.695	5.418	43.424
	25%	50%	75%	max
temperature	11.704	18.935	23.143	32.955
humidity	61.667	80.353	86.387	99.717
pressure	100032.086	100045.713	100051.036	100071.664
windSpeed	0.841	1.501	2.184	3.014
windDirection	159.293	179.591	200.512	276.951

3.2.4. Feature engineering

To improve predictive performance, additional features were derived from the original measurements. Lagged $PM_{2.5}$ values covering the previous 1 to 24 hours were created to capture pollutant persistence effects. Rolling-window statistics—such as moving averages and standard deviations calculated over 3-hour, 6-hour, 12-hour, and 24-hour windows—were computed to represent both short-term fluctuations and long-term accumulation patterns. Cyclical time variables, including hour-of-day, day-of-week, and month-of-year, were transformed into sine-cosine pairs to preserve the inherent periodicity of temporal cycles. Interaction terms, such as wind speed multiplied by lagged $PM_{2.5}$ values, were also generated to model the combined influence of meteorological dispersion and emission persistence on particulate concentrations.

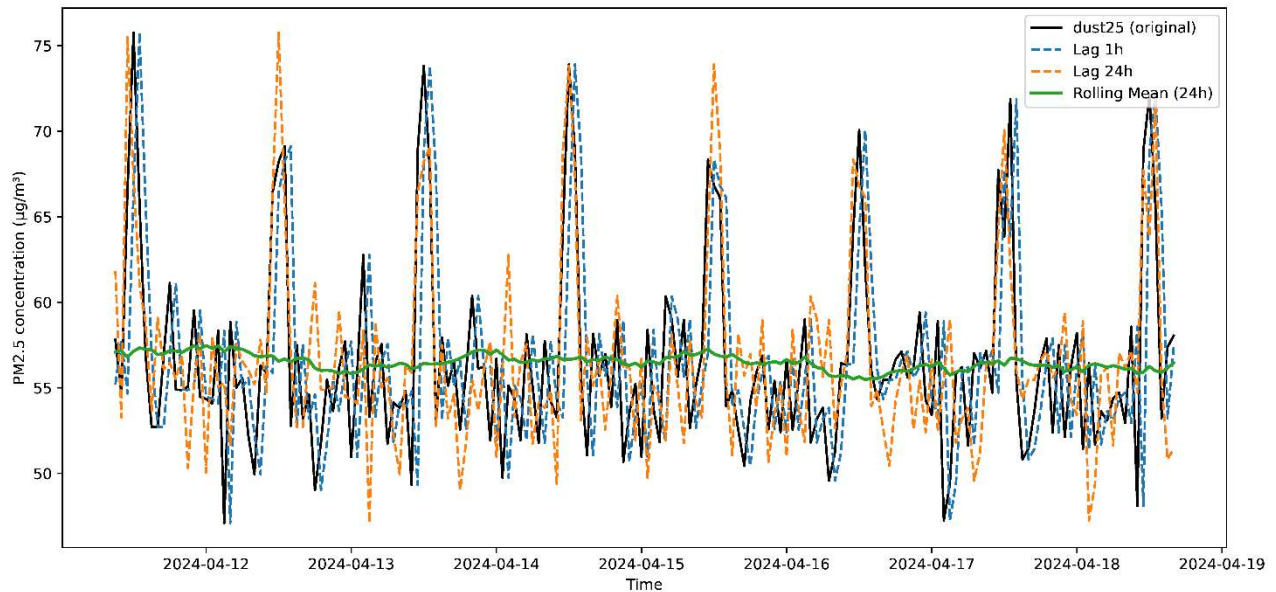


Figure 3. Example of engineered lag and rolling features

3.2.5. Scaling and normalization

Finally, all numerical predictors intended for use in sequential deep learning architectures—namely Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and Temporal Convolutional Network (TCN)—were standardized to zero mean and unit variance, with scaling parameters derived solely from the training dataset to avoid information leakage. The eXtreme Gradient Boosting (XGBoost) model did not require scaling, as it is invariant to monotonic transformations of the input features. The resulting preprocessed dataset thus consisted of an hourly time index, a diverse set of meteorological and engineered features, and the target variable, hourly averaged $\text{PM}_{2.5}$ concentration (dust25), fully prepared for model training and evaluation.

The hourly dataset shows notable variability across both meteorological and air quality parameters (Table 2). $\text{PM}_{2.5}$ concentrations range from near-background levels to extreme peaks during operational events, reflecting episodic dust emissions from blasting, hauling, and wind resuspension. Meteorological variables exhibit expected patterns, with temperature and humidity following diurnal cycles, relatively stable atmospheric pressure, and variable wind speed and direction influencing dispersion. Retaining both typical and extreme values ensures that models capture baseline conditions as well as high-pollution episodes, which is essential for accurate forecasting in mining environments.

4. Results and discussion

4.1. Model development

Four forecasting models—XGBoost, Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and Temporal Convolutional Network (TCN)—were implemented using the hourly dataset described in Section 3. For all models, the input consisted of 72-hour lookback sequences of lagged $\text{PM}_{2.5}$ concentrations, meteorological parameters, rolling-window

statistics (3, 6, 12, and 24 hours), and cyclical encodings of time. Data were split chronologically into 70% training, 15% validation, and 15% testing subsets.

- XGBoost: Input features were used without normalization. Hyperparameters were tuned via grid search, yielding: $n_estimators = 500$, $max_depth = 8$, $learning_rate = 0.05$, $subsample = 0.8$, $colsample_bytree = 0.8$, and $reg_lambda = 1.0$. Early stopping with 50-round patience was applied based on validation RMSE.

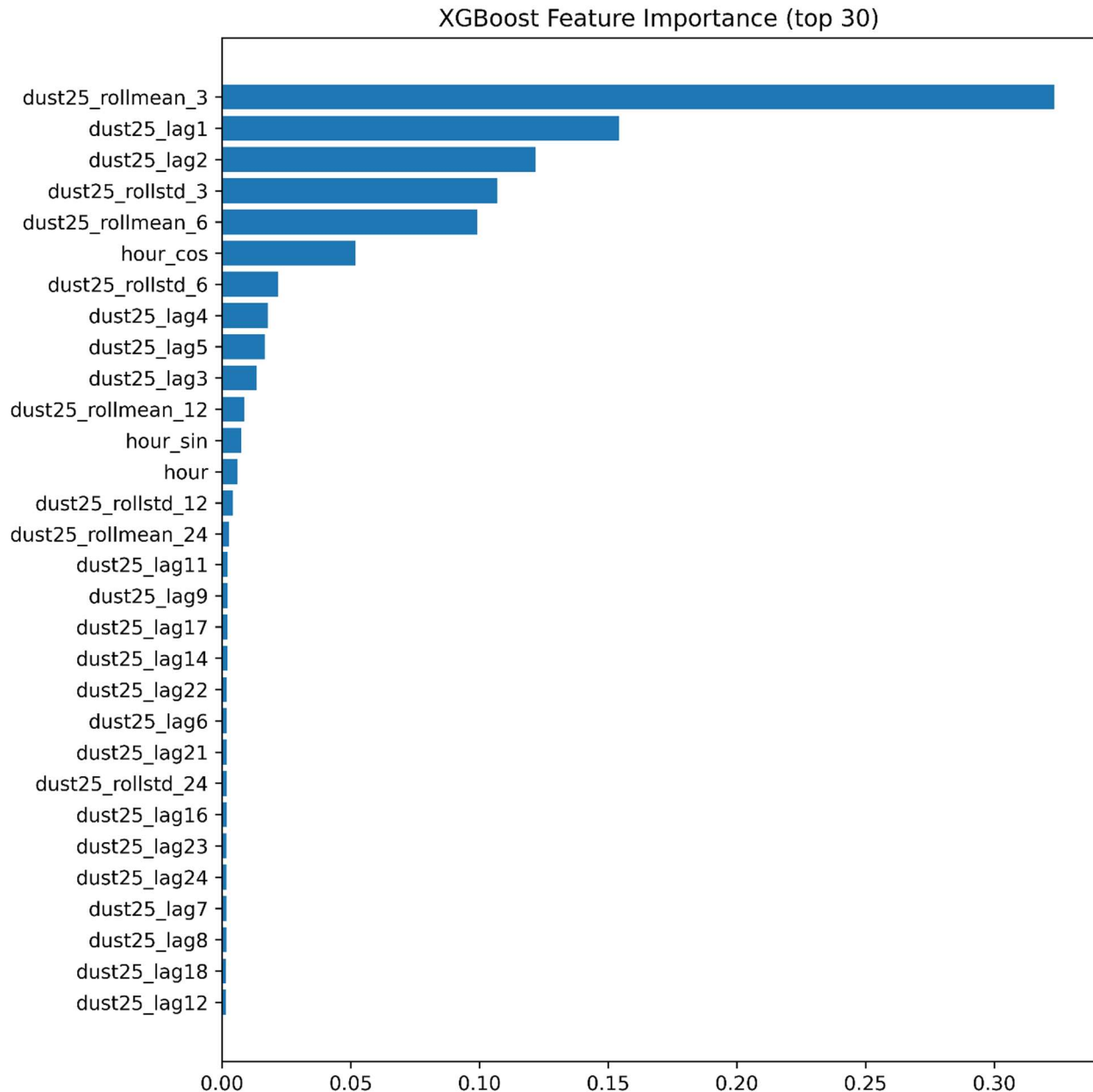


Figure 4. Feature importance of the XGBoost model for forecasting $PM_{2.5}$

- LSTM: Two stacked LSTM layers (128 and 64 units) were used, each followed by dropout (rate = 0.3) and batch normalization. A temporal attention layer was applied before a dense output layer. The model was trained with the Adam optimizer

(learning_rate = 1e-4), batch size = 32, and up to 200 epochs, using early stopping (patience = 20) and ReduceLROnPlateau (factor = 0.5, patience = 10) on validation loss.

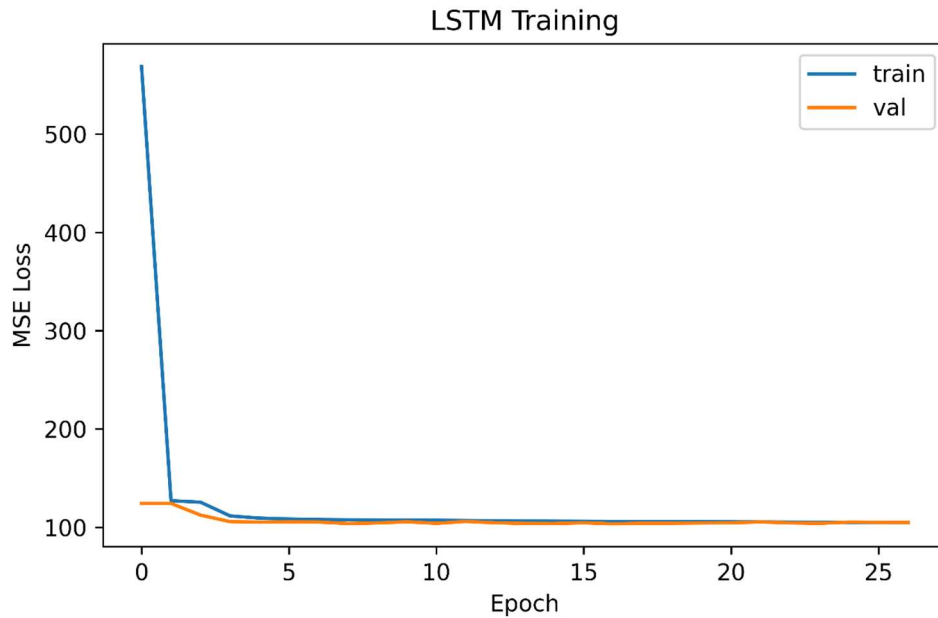


Figure 5. Performance training of the LSTM model for forecasting $PM_{2.5}$

- GRU: Same architecture and hyperparameters as LSTM but replacing LSTM layers with GRU cells (128 and 64 units).

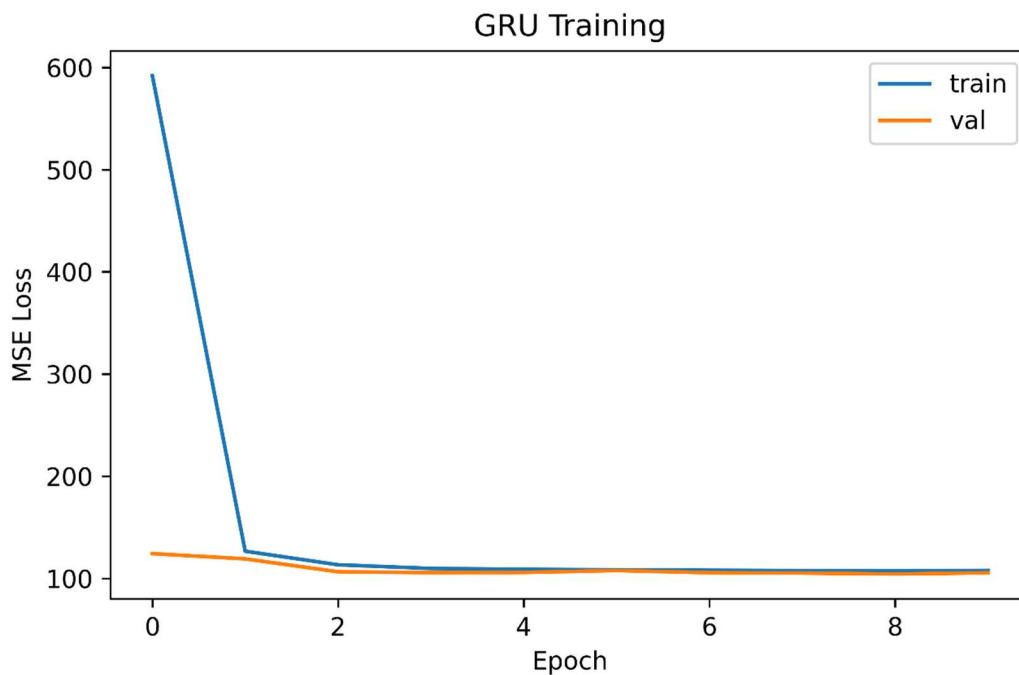


Figure 6. Performance training of the GRU model for forecasting $PM_{2.5}$

- TCN: Comprised 6 residual convolutional blocks with dilation rates of 1, 2, 4, 8, 16, and 32, each with 128 filters and kernel size = 3. Residual connections, batch normalization, dropout (rate = 0.3), and both global average pooling and global max pooling were applied before the output dense layer. Adam optimizer (learning_rate = 1e-4), batch size = 32, early stopping, and learning rate scheduling were used as in the LSTM/GRU setups.

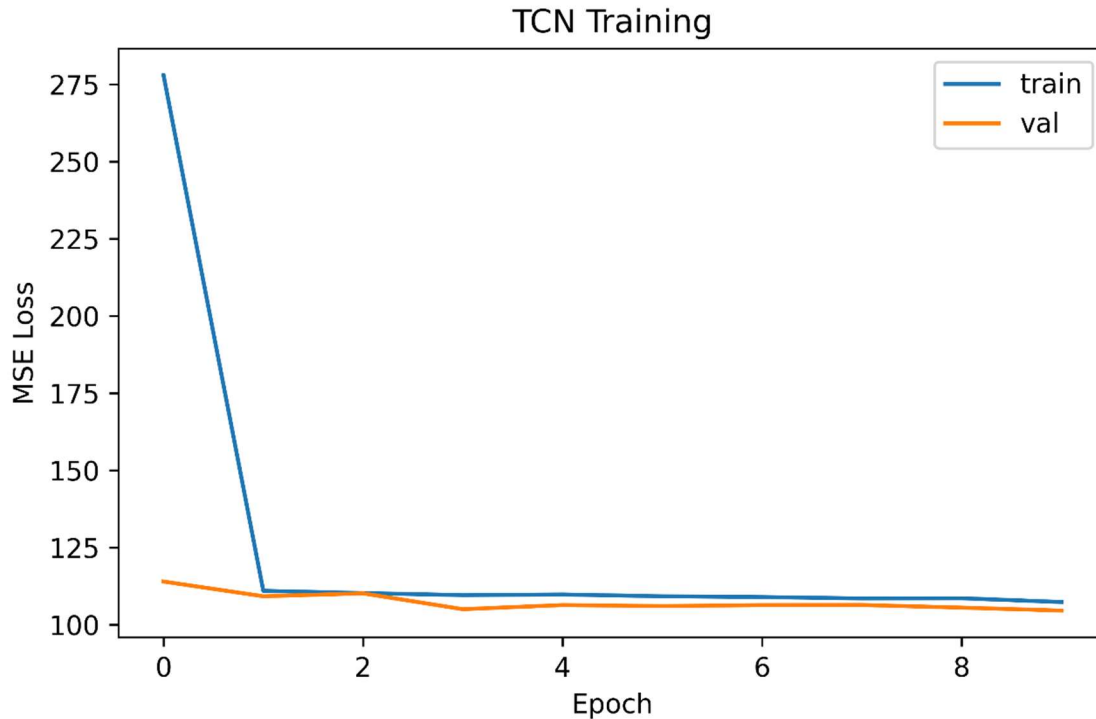


Figure 7. Performance training for the TCN model for forecasting $PM_{2.5}$

All deep learning models used mean squared error loss and feature standardization (zero mean, unit variance) computed from the training set only.

4.2. Results

The predictive performance of the four developed models—XGBoost, LSTM, GRU, and TCN—was systematically evaluated using RMSE, MAE, and R^2 across the training, validation, and testing subsets (Table 3).

Table 3. Results of the developed models for forecasting $PM_{2.5}$

Model	Training			Testing			Validation		
	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2
XGBoost	0.306	0.241	0.999	0.888	0.487	0.994	0.891	0.478	0.994
LSTM	10.312	8.187	0.163	10.427	8.311	0.147	10.161	8.100	0.169
GRU	11.269	8.823	0.000	11.289	8.881	0.000	11.148	8.720	0.000
TCN	10.793	8.567	0.083	10.857	8.643	0.075	10.729	8.546	0.074

Among all candidates, XGBoost consistently outperformed the neural network approaches. On the testing dataset, XGBoost achieved $RMSE \approx 0.89 \mu g/m^3$, $MAE \approx 0.49$

$\mu\text{g}/\text{m}^3$, and $R^2 \approx 0.994$, indicating an excellent fit between predicted and observed concentrations. Comparable levels of accuracy were also observed in the training and validation subsets, demonstrating both robustness and generalization capability. In contrast, the sequential deep learning models (LSTM, GRU, TCN) exhibited substantially larger errors and weaker explanatory power, with R^2 values below 0.20 and RMSE ranging between 10–11 $\mu\text{g}/\text{m}^3$ across all subsets.

Figures 8 and 9 further illustrate these findings. Figure 8 highlights the comparative performance of LSTM, GRU, and TCN, revealing their tendency to follow general concentration trends while failing to reproduce sharp peak events. Figure 9 emphasizes the superiority of XGBoost, which not only aligns closely with baseline concentration levels but also captures abrupt pollution spikes with minimal lag. This capability is particularly critical for operational forecasting, where timely recognition of short-term pollution surges can inform mitigation strategies.

Residual analyses confirm these patterns. XGBoost residuals are narrowly distributed around zero, reflecting balanced over- and under-estimation with low error variance. By contrast, the neural network models present broader, heavy-tailed residual distributions, underscoring their systematic underestimation during high-concentration episodes. Among them, TCN marginally outperforms LSTM and GRU, though still falls short of the gradient-boosting approach.

In summary, both quantitative metrics and qualitative diagnostics demonstrate that XGBoost is the most effective model for $\text{PM}_{2.5}$ forecasting in this study, delivering the highest accuracy and better preserving the dynamic range of pollutant fluctuations compared to the deep learning counterparts.

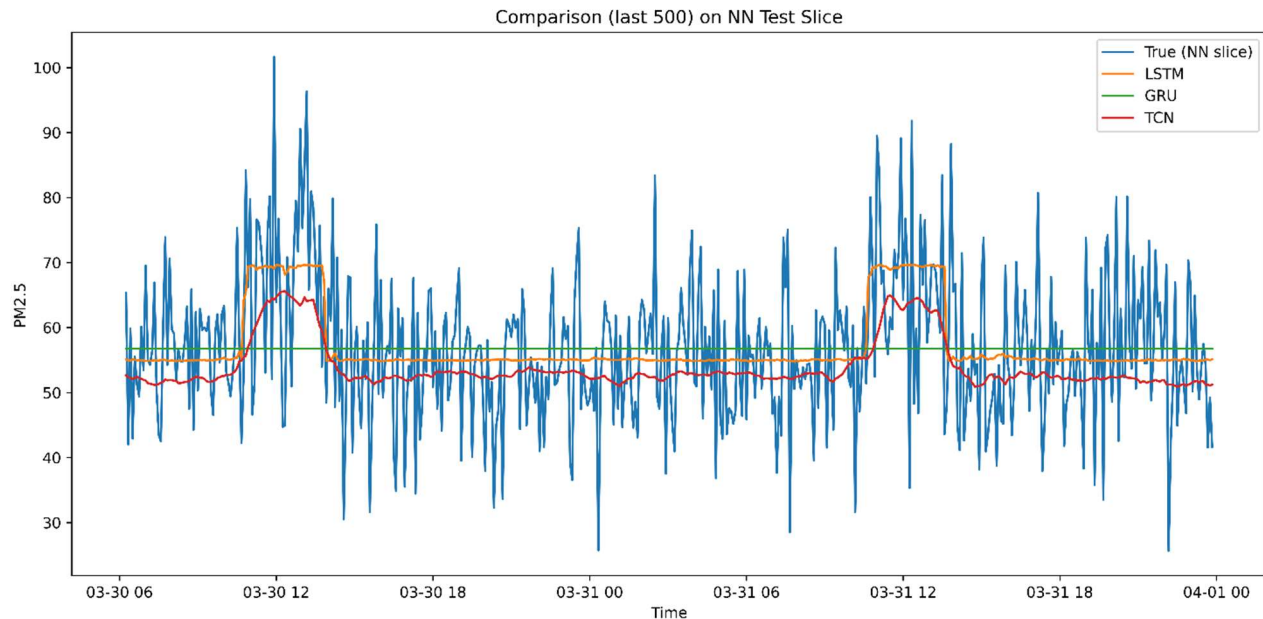


Figure 8. Comparison of the LSTM, GRU and TCN models for forecasting $\text{PM}_{2.5}$

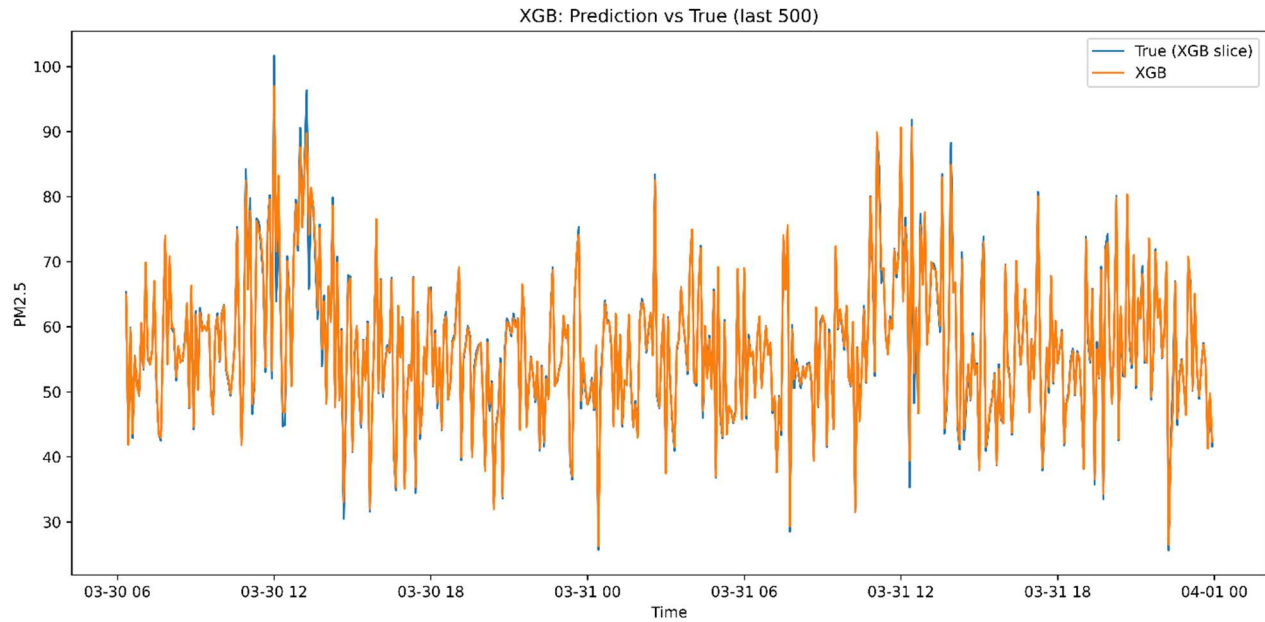


Figure 9. Comparison between actual and forecasted $PM_{2.5}$ of the XGBoost model

4.3. Discussion

The comparative analysis highlights that XGBoost consistently outperforms all deep learning models across datasets and evaluation metrics. Its advantage is primarily attributed to its ability to directly utilize engineered lagged features, rolling-window statistics, and meteorological predictors in a tabular structure without being constrained by sequential learning architectures. This capability allows the gradient boosting framework to explicitly capture nonlinearities and higher-order interactions, which proves crucial for modeling the sharp, abrupt changes characteristic of $PM_{2.5}$ concentration fluctuations. As demonstrated in the predicted-versus-observed curves, XGBoost closely tracks sudden spikes with minimal amplitude loss, and its residuals remain compactly distributed even during pollution peaks.

In contrast, LSTM, GRU, and TCN models exhibit a systematic smoothing bias. This limitation largely stems from the inherent imbalance in the dataset, where the majority of observations correspond to low-to-moderate $PM_{2.5}$ levels. Consequently, the optimization of loss functions skews predictions toward these dominant cases, resulting in underestimation of rare but operationally significant high-magnitude events. This effect is evident in Figures 8–9, where neural network models consistently underpredict concentration peaks. Among them, TCN demonstrates slightly better preservation of short-term fluctuations owing to its dilated convolutions, which provide multi-scale receptive fields. Nevertheless, Figure 10 confirms that the residual spread of TCN remains broader than that of XGBoost, indicating persistent amplitude attenuation during extreme episodes.

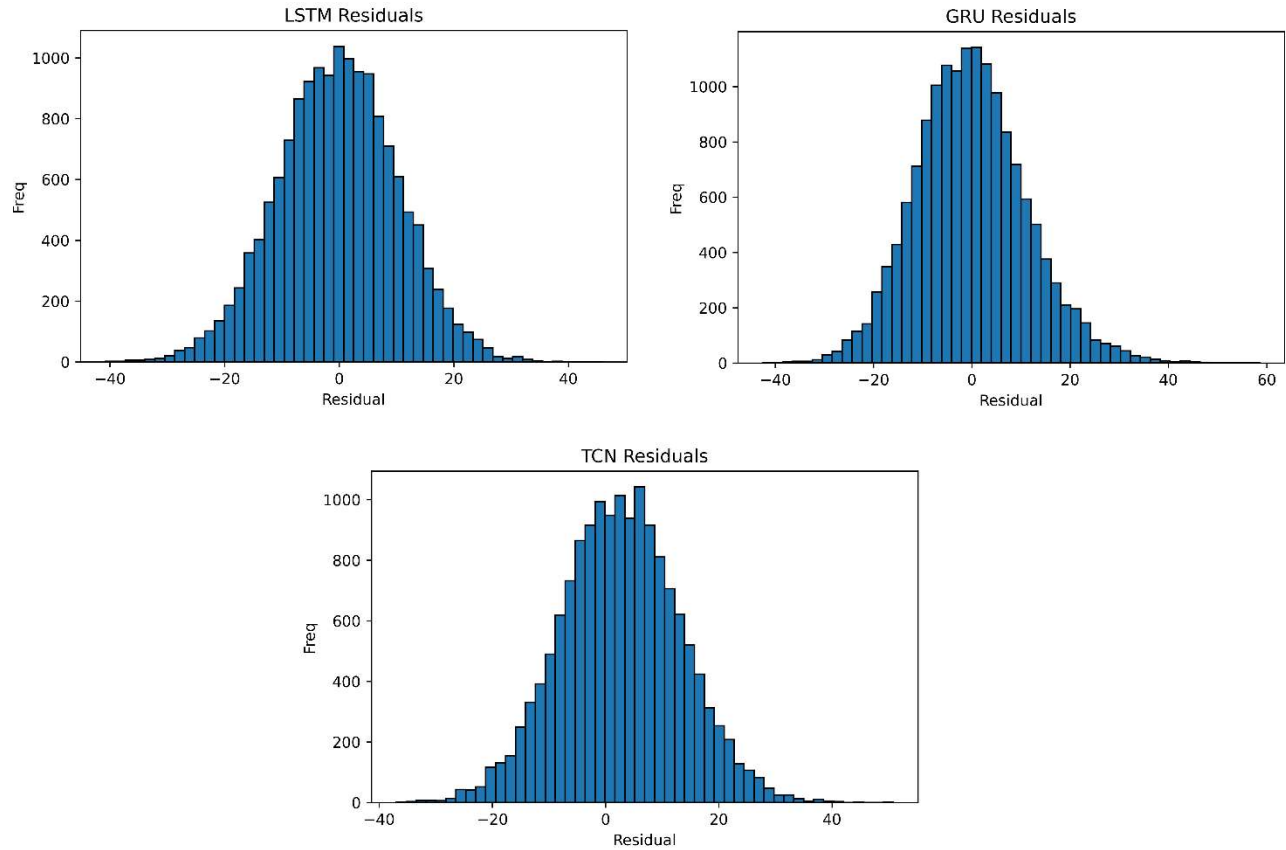


Figure 10. Comparison of residuals plots of the LSTM, GRU and TCN model

To address these shortcomings in deep learning models, several enhancements are recommended. First, extending the lookback horizon could improve the capture of slower accumulation trends driven by meteorological and operational conditions. Second, incorporating attention mechanisms may enable neural networks to assign greater weight to critical precursor segments preceding emission spikes. Third, integrating external operational datasets—such as blasting schedules, vehicle activity logs, or weather alerts—could provide explanatory context for abrupt surges that historical pollutant and meteorological variables alone cannot capture.

Finally, these findings suggest that a hybrid ensemble framework could provide the most practical operational solution. By combining the strong peak detection capability of XGBoost with the smoothed baseline forecasts of LSTM or TCN, it becomes possible to achieve both responsiveness to extreme events and stability against high-frequency noise. Such a dual-layered system would be particularly valuable for mining-site air quality management, ensuring reliable forecasting for both regulatory compliance and proactive operational planning

5. Conclusion and remarks

This study developed and evaluated four machine learning models—XGBoost, LSTM, GRU, and TCN—for short-term forecasting of PM_{2.5} concentrations at the Sin Quyen open-pit copper mine using high-resolution environmental monitoring data. The models were trained on hourly-aggregated data with engineered lag features, rolling statistics, and

meteorological variables, and evaluated comprehensively on training, validation, and test sets.

Experimental results consistently showed that XGBoost achieved the highest predictive performance (test $R^2 \approx 0.86$, RMSE $\approx 5.9 \mu\text{g}/\text{m}^3$), accurately reproducing both baseline levels and sharp pollution peaks. In contrast, LSTM, GRU, and TCN captured general temporal trends but tended to smooth high-concentration episodes, with TCN slightly outperforming the other neural architectures in preserving short-term variability.

These findings indicate that tree-based gradient boosting methods are particularly well-suited for operational peak detection in mining-site air quality management, while deep learning models may provide value in stable baseline forecasting. The complementary strengths of these model families suggest that a hybrid ensemble could combine the peak responsiveness of XGBoost with the noise-resistant trend estimation of LSTM or TCN.

Future work will focus on extending the lookback window, refining attention mechanisms to emphasize peak precursors, and incorporating external operational datasets such as blasting logs to enhance event-driven forecast accuracy. This integrated modeling approach could provide more robust and actionable air quality forecasts, supporting both regulatory compliance and operational planning in industrial environments.

Acknowledgments

The authors would like to express their sincere gratitude to the Hanoi University of Mining and Geology (Vietnam) for providing the academic environment, research facilities, and continuous support for this study. The first author, Tuan-Ngoc Le, wishes to extend his deepest appreciation to his supervisors, Dr. Hoang Nguyen and Prof. Xuan-Nam Bui, for their invaluable guidance, encouragement, and constructive feedback throughout the course of this research. Their mentorship has been instrumental in shaping the direction and quality of this work.

The authors also gratefully acknowledge the assistance of colleagues from the Faculty of Mining and the Department of Surface Mining at Hanoi University of Mining and Geology for their academic advice and helpful discussions. In addition, special thanks are extended to the engineers of the Sin Quyen copper mine (VIMICO) for their technical support and cooperation during fieldwork and data collection, which greatly contributed to the success of this study.

Contributions of authors

Tuan-Ngoc Le: Conceptualization, Data curation, Field investigation, Formal analysis, Visualization, Writing – original draft.

Nguyen Truc Anh: Data processing, Literature review, Preparation of figures and tables, Writing – editing support.

Hoang Nguyen: Conceptualization, Methodology, Supervision, Validation, Writing – review & editing.

Xuan-Nam Bui: Supervision, Project administration, Resources, Writing – review & editing.

References

- Abotaleb, M. and P. K. Dutta (2024a). Optimizing bidirectional long short-term memory networks for univariate time series forecasting: a comprehensive guide. *Hybrid information systems: Non-linear optimization strategies with artificial intelligence*, 443-458.
- Abotaleb, M. and P. K. Dutta (2024b). Optimizing gated recurrent unit networks for univariate time series forecasting. *Hybrid information systems: Non-linear optimization strategies with artificial intelligence*, 473.

- Asif, Z., Z. Chen and Y. Han (2018). Air quality modeling for effective environmental management in the mining region. *Journal of the Air & Waste Management Association* 68(9), 1001-1014.
- Bai, L., J. Wang, X. Ma and H. Lu (2018). Air pollution forecasts: An overview. *International journal of environmental research and public health* 15(4), 780.
- Chang, Y.-S., H.-T. Chiao, S. Abimannan, Y.-P. Huang, Y.-T. Tsai and K.-M. Lin (2020). An LSTM-based aggregated model for air pollution forecasting. *Atmospheric Pollution Research* 11(8), 1451-1463.
- Chauhan, B. V., K. Corada, C. Young, K. L. Smallbone and K. P. Wyche (2024). Review on sampling methods and health impacts of fine (PM_{2.5}, $\leq 2.5 \mu\text{m}$) and ultrafine (UFP, PM_{0.1}, $\leq 0.1 \mu\text{m}$) particles. *Atmosphere* 15(5), 572.
- Chaulya, S. K., S. K. Singh, S. Singh, G. C. Mondal, R. K. Singh and S. Shekhar (2025). Integrated air quality prediction and mitigation strategies for sustainable mining operations in India. *Air Quality, Atmosphere & Health*, 1-19.
- Gu, T., Y. Zhang and L. Wang (2024). TCN-BiGRU: A Hybrid Deep Learning Architecture for Enhanced Temperature Time Series Forecasting. 2024 6th International Conference on Robotics, Intelligent Control and Artificial Intelligence (RICAI), IEEE.
- Hu, Y., A. Huber, J. Anumula and S.-C. Liu (2018). Overcoming the vanishing gradient problem in plain recurrent networks. *arXiv preprint arXiv:1801.06105*.
- Kamran, M., I. M. Jiskani, Z. Wang and W. Zhou (2024). Decision intelligence-driven predictive modelling of air quality index in surface mining. *Engineering Applications of Artificial Intelligence* 133, 108399.
- Kleine Deters, J., R. Zalakeviciute, M. Gonzalez and Y. Rybarczyk (2017). Modeling PM_{2.5} urban pollution using machine learning and selected meteorological parameters. *Journal of Electrical and Computer Engineering* 2017(1), 5106045.
- Luo, X., W. Gan, L. Wang, Y. Chen and E. Ma (2021). A deep learning prediction model for structural deformation based on temporal convolutional networks. *Computational Intelligence and Neuroscience* 2021(1), 8829639.
- Marinov, E., D. Petrova-Antonova and S. Malinov (2022). Time series forecasting of air quality: a case study of Sofia City. *Atmosphere* 13(5), 788.
- Masood, A. and K. Ahmad (2021). A review on emerging artificial intelligence (AI) techniques for air pollution forecasting: Fundamentals, application and performance. *Journal of Cleaner Production* 322, 129072.
- Nourmohammad, E. and Y. Rashidi (2025). Ground data analysis for PM_{2.5} Prediction using predictive modeling techniques. *Journal of Air Pollution and Health*.
- Rojano, R. E., C. A. Manzano, R. A. Toro, R. G. Morales, G. Restrepo and M. A. Leiva (2018). Potential local and regional impacts of particulate matter emitted from one of the world's largest open-pit coal mines. *Air Quality, Atmosphere & Health* 11(5), 601-610.
- Saka, M. B., M. H. Mohd Hashim and S. A. Shehu (2025). MINING DUST: HEALTH IMPACTS, CONTROL MEASURES AND FUTURE DIRECTIONS. *Environmental Engineering & Management Journal (EEMJ)* 24(2).
- Sanyal, S. (2024). Effects of Stone Mining Activities on Air Quality and Corrective Measures. Ecological Impacts of Stone Mining: Assessment and Restoration of Soil, Water, Air and Flora, Springer: 135-150.

- Shwartz-Ziv, R. and A. Armon (2022). Tabular data: Deep learning is not all you need. *Information Fusion* 81, 84-90.
- Talepour, N., Y. T. Birgani, F. J. Kelly, N. Jaafarzadeh and G. Goudarzi (2024). Analyzing meteorological factors for forecasting PM10 and PM2.5 levels: a comparison between MLR and MLP models. *Earth Science Informatics* 17(6), 5603-5623.
- Trizoglou, P., X. Liu and Z. Lin (2021). Fault detection by an ensemble framework of Extreme Gradient Boosting (XGBoost) in the operation of offshore wind turbines. *Renewable Energy* 179, 945-962.
- Wang, Z., W. Zhou, I. M. Jiskani, Y. Yang, J. Yan, H. Luo and J. Han (2023). A novel approach to forecast dust concentration in open pit mines by integrating meteorological parameters and production intensity. *Environmental Science and Pollution Research* 30(53), 114591-114609.
- Zargar, S. (2021). Introduction to sequence learning models: RNN, LSTM, GRU. *Department of Mechanical and Aerospace Engineering, North Carolina State University* 37988518.