

ĐÁNH GIÁ KHẢ NĂNG PHỤC HỒI GIAO THÔNG ĐÔ THỊ: TIẾP CẬN BẰNG HỌC MÁY DỰA TRÊN KỊCH BẢN

TS. NGUYỄN TRỌNG DŨNG*

PGS.TS. NGUYỄN THÙY ANH

ĐẶNG THỊ MINH

Nhóm nghiên cứu Ứng dụng công nghệ 4.0 trong Giao thông vận tải (I4T),
Trường Đại học Công nghệ Giao thông vận tải, 54 Triều Khúc, Thanh Liệt, Hà Nội, Việt Nam
*Email: anhnt@utt.edu.vn

Phản biện: TS. Cao Minh Quyền

TS. Hoàng Thị Hương Giang

TÓM TẮT

Nghiên cứu này đề xuất khung phương pháp luận nhằm định lượng khả năng phục hồi của mạng lưới giao thông đô thị trước các nhiễu loạn phi tuyến. Thông qua việc thiết lập môi trường Bản sao số, nghiên cứu mô phỏng kịch bản nhiễu loạn hỗn hợp nhằm khắc phục rào cản thiếu hụt dữ liệu lịch sử. Thuật toán tăng cường độ dốc cực hạn (XGBoost) được ứng dụng để kiểm thử sức chịu tải, phân tích động lực học lan truyền ùn tắc không-thời gian. Kết quả nghiên cứu đã toán học hóa thành công chỉ số phục hồi mạng lưới và định vị chính xác các ngưỡng suy giảm tới hạn. Khung tiếp cận này cung cấp nền tảng thực chứng vững chắc phục vụ công tác quy hoạch hạ tầng và thiết lập chiến lược điều tiết giao thông chủ động.

Từ khóa: Khả năng phục hồi giao thông đô thị, Học máy, Bản sao số, XGBoost, Ùn tắc không - thời gian.

ABSTRACT:

This study proposes a methodological framework to quantify the resilience of urban traffic networks against nonlinear disturbances. By establishing a Digital Twin environment, the research simulates mixed disturbance scenarios to overcome the barrier of historical data scarcity. The Extreme Gradient Boosting (XGBoost) algorithm is applied to conduct stress testing and analyze the spatiotemporal propagation dynamics of congestion. The research results have successfully mathematized the network resilience index and accurately pinpointed the critical degradation thresholds. This approach framework provides a solid empirical foundation for infrastructure planning and establishing proactive traffic regulation strategies.

Keywords: Urban traffic resilience, Machine learning, Digital Twin, XGBoost, Spatiotemporal congestion.

1. ĐẶT VẤN ĐỀ

Ùn tắc giao thông đô thị làm suy giảm năng lực thông hành và gia tăng chi phí vận hành mạng lưới. Sự phát triển của các hệ thống giao thông thông minh (ITS) dựa trên internet vạn vật và trí tuệ nhân tạo (AI) đã nâng cao đáng kể hiệu quả giám sát và dự báo lưu lượng [1, 2]. Tuy nhiên, phần lớn các kiến trúc dự báo hiện hành được tối ưu hóa cho điều kiện luồng giao thông ổn định, dẫn đến giảm độ tin cậy khi mạng lưới chịu tác động bởi các biến số nhiễu loạn ngẫu nhiên có biên độ lớn [3].

Hạn chế này đòi hỏi sự thay đổi phương pháp luận từ việc tối ưu hóa công suất thiết kế sang đánh giá khả năng phục hồi giao thông đô thị nhằm xác định các ngưỡng chịu đựng tới hạn. Thực tế cho thấy hơn 50% nguyên nhân gây ùn tắc

xuất phát từ các tác nhân ngoại sinh không lặp lại như thời tiết bất lợi hoặc sự kiện đô thị [4]. Việc xác định các ngưỡng tới hạn – mức độ nhiễu loạn tối đa mà hệ thống có thể chịu đựng trước khi giảm năng lực thông hành là thông số kỹ thuật cốt lõi để quy hoạch và quản lý hạ tầng dài hạn [5]. Mặc dù vậy, trong quy trình tiền xử lý dữ liệu học máy, các giá trị bất thường phản ánh các sự kiện này thường bị chuẩn hóa hoặc loại bỏ. Đồng thời, việc thiếu dữ liệu lịch sử về các kịch bản nhiễu loạn hỗn hợp cản trở đánh giá chính xác giới hạn suy giảm chức năng của mạng lưới.

Nhằm giải quyết vấn đề trên, nghiên cứu này đề xuất khung phương pháp định lượng khả năng phục hồi dựa trên môi trường bản sao số đối với mạng lưới giao thông Hà Nội. Môi trường mô phỏng được

sử dụng để tạo lập tập dữ liệu nhân tạo tích hợp đa biến số nhiễu loạn. Bằng việc ứng dụng thuật toán học máy XGBoost – một phương pháp có hiệu suất cao trong việc xử lý dữ liệu phi tuyến và nhiễu, nghiên cứu tiến hành các bài kiểm tra khả năng chịu đựng nhằm định vị các vùng không gian nhạy cảm và định lượng thời gian phục hồi của hệ thống dưới các điều kiện vận hành bất lợi [6].

2. PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Thiết kế môi trường bản sao số

Môi trường bản sao số trong nghiên cứu này được thiết lập như một thực thể số hóa động, phản ánh cấu trúc và quy luật vận hành của mạng lưới giao thông thực tế nhằm định lượng khả năng phục hồi dưới các tác nhân gây bất lợi. Quá trình thực thể hóa bắt đầu

bằng việc mô hình hóa hạ tầng dưới dạng đồ thị có hướng $G = (V, E)$, với V là tập hợp các nút giao và E là các phân đoạn đường bộ. Dữ liệu hình học được trích xuất từ OpenStreetMap thông qua thư viện OSMnx trong bán kính 15 km tính từ trung tâm Hà Nội. Mỗi cạnh trong đồ thị được gán các tham số vật lý như chiều dài, cấp hạng đường và tốc độ thiết kế để xác định năng lực thông hành cơ sở.

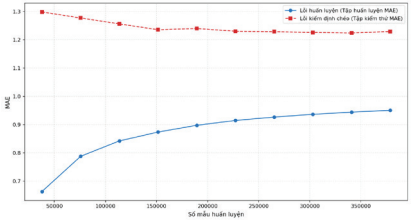
Để giám sát mạng lưới, nghiên cứu triển khai hệ thống cảm biến ảo tại các vị trí chiến lược, ghi nhận vận tốc trung bình, thời gian di chuyển và chỉ số ùn tắc theo chu kỳ 10 phút. Giải pháp này khắc phục các hạn chế về độ phủ và rủi ro mất mát dữ liệu của thiết bị vật lý, tạo ra chuỗi dữ liệu nhân tạo quy mô lớn để huấn luyện các thuật toán học máy.

Điểm then chốt của bản sao số là khả năng giả lập kịch bản nhiễu loạn hỗn hợp: sự tương tác phi tuyến giữa nhiễu loạn khí tượng và biến động nhu cầu đi lại đột biến. Cụ thể, lượng mưa được tích hợp qua hàm giảm năng lực thông hành, trong khi sự gia tăng nhu cầu được điều chỉnh bằng trọng số lưu lượng tại các khung giờ cao điểm. Cấu trúc này cho phép quan sát sự chuyển dịch trạng thái mạng lưới từ vận hành ổn định sang suy giảm chức năng, từ đó xác định ngưỡng tới hạn phục hồi của hệ thống dưới các tác động cộng hưởng.

2.2. Kiến trúc mô hình học máy

Kiến trúc mô hình dự báo trong nghiên cứu này được phát triển dựa trên thuật toán tăng cường XGBoost. Việc lựa chọn thuật toán này được căn cứ vào năng lực mô hình hóa các mối quan hệ phi tuyến tính phức tạp của dòng chảy giao thông, đồng thời kiểm soát chặt chẽ hiện tượng quá khớp thông qua việc tích hợp các thành phần điều chuẩn trực tiếp vào hàm mục tiêu. Không gian đặc trưng đầu vào được thiết lập dựa trên sự kết hợp đa chiều của ba nhóm biến số chỉ phối trạng thái vận hành mạng lưới: (1) nhóm đặc trưng hình học và chu kỳ thời gian giúp phân tích quy luật nhu cầu lặp lại; (2) nhóm

biến ngoại sinh (cường độ lượng mưa, sự kiện đặc biệt) nhằm cấu thành các kịch bản kiểm thử sức chịu đựng; và (3) nhóm đặc trưng tự hồi quy ghi nhận thời gian di chuyển trong quá khứ giúp nắm bắt động lực học lan truyền ùn tắc trên cấu trúc không - thời gian.



Hình 1: Biểu đồ đường cong học tập

Quy trình huấn luyện và kiểm định (Hình 1) cho thấy sai số tuyệt đối trung bình (MAE) hội tụ ổn định khi quy mô mẫu tăng, khẳng định hệ thống đã học được các quy luật phân phối nội tại một cách khách quan và giải quyết vấn đề học vẹt. Về mặt định lượng (Bảng 1), mô hình XGBoost đạt hiệu suất dự báo tốt trên tập kiểm tra với hệ số $R^2 = 0.9852$, $RMSE = 3.1389$ và $MAE = 1.4274$. Kết quả này minh chứng năng lực tổng quát hóa đáp ứng đầy đủ tiêu chuẩn kỹ thuật để triển khai tính toán thời gian phục hồi và định lượng giới hạn sụp đổ của mạng lưới trong các kịch bản bất lợi [7].

Bảng 1: So sánh hiệu suất mô hình XGB trên 3 tập huấn luyện

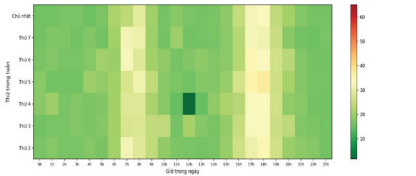
Tập dữ liệu	MAE	RMSE	R^2
Tập huấn luyện	0.9609	1.8584	0.9950
Tập kiểm tra	1.4274	3.1389	0.9852
Toàn bộ tập	1.0612	2.1975	0.9930

3. THIẾT LẬP THỰC NGHIỆM VÀ CHỈ SỐ PHỤC HỒI

Nhằm định lượng khả năng phục hồi của mạng lưới giao thông đô thị, nghiên cứu thiết lập khung kiểm tra sức chịu tải thông qua việc phân tích đối chiếu hai trạng thái vận hành mô phỏng như đã thiết lập tại mục 2.1: kịch bản cơ sở (đại diện cho trạng thái vận hành tiêu chuẩn) và kịch bản nhiễu loạn hỗn

hợp (tác động cộng hưởng giữa khí tượng bất lợi và sự kiện đô thị). Thông qua thuật toán học máy XGBoost, tập dữ liệu từ kịch bản này được phân tích để định lượng chính xác các ngưỡng đứt gãy mà tại đó hệ thống mất khả năng duy trì chức năng thông hành. Quy trình này cung cấp nền tảng toán học để tính toán biên độ suy giảm hiệu suất và chu kỳ thời gian phục hồi của cấu trúc giao thông dưới áp lực vận hành bất lợi.

Để thiết lập hệ quy chiếu khách quan về năng lực chống chịu của hạ tầng giao thông trước các tác nhân nhiễu loạn, nghiên cứu tiến hành toán học hóa khái niệm này thông qua việc xây dựng chỉ số phục hồi mạng lưới (NRI). Về mặt phương pháp luận, chỉ số NRI được cấu thành từ hai tham số động lực học cốt lõi: biên độ suy giảm hiệu suất vận hành cực đại và chu kỳ thời gian phục hồi.

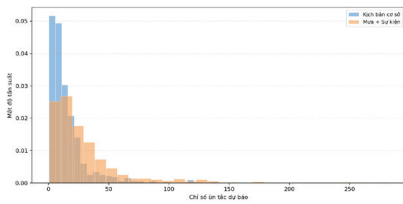


Hình 2: Bản đồ nhiệt thời gian

Để tìm hiểu chính xác các tham số này, nghiên cứu ứng dụng phương pháp phân tích không gian trạng thái thông qua bản đồ nhiệt thời gian (Hình 2) kết hợp với các tham số thống kê mô tả (Bảng 2). Dữ liệu thực chứng từ Bảng 2 chỉ ra tính phi đồng nhất và quy luật chu kỳ đặc trưng của luồng phương tiện đô thị, trong đó chỉ số ùn tắc trung bình duy trì ở ngưỡng cao vào các ngày làm việc trong tuần. Sự tập trung lưu lượng tại các khung giờ cao điểm (17h - 18h) đi kèm với độ lệch chuẩn lớn và giá trị cực đại gia tăng đột biến, cung cấp minh chứng cho sự tồn tại của các xung lực nhiễu động cục bộ có cường độ cao.

Khi tham chiếu các tham số thống kê này lên ma trận phân phối hai chiều tại Hình 2, sự biến thiên dải màu cung cấp cơ sở thực chứng liên tục để trích xuất chính xác thời điểm hệ thống bắt đầu suy giảm

chính xác biên độ rủi ro, từ đó thiết lập các ranh giới an toàn vững chắc phục vụ công tác quy hoạch và vận hành giao thông chủ động.



Hình 5: Biểu đồ so sánh kịch bản

4.3. Sự không đồng nhất về không gian

Đánh giá khả năng phục hồi của mạng lưới giao thông cần phân tích sự không đồng nhất về không gian, do năng lực chống chịu của hạ tầng phân bố không đồng đều trên toàn bộ phạm vi đô thị. Dữ liệu tại Bảng 4 phân tích tính nhạy cảm của mạng lưới theo 4 vùng địa lý tính từ trung tâm thành phố. Kết quả thống kê chỉ ra sự phân hóa rõ rệt về mức độ bất ổn định của dòng phương tiện: vùng ngoại vi ghi nhận chỉ số ùn tắc trung bình và độ lệch chuẩn cao nhất cùng khoảng tứ phân vị rộng. Điều này khẳng định đây là khu vực có năng lực phục hồi thấp nhất, rất dễ bị suy giảm chức năng khi chịu các tác động cộng hưởng.

Tương tự, vùng trung tâm cũng thể hiện mức độ biến động lớn, phản ánh tình trạng quá tải thường xuyên do mật độ giao cắt cao. Đáng chú ý, mặc dù vùng cận ngoại có chỉ số ùn tắc trung bình tương đối thấp, khu vực này lại chứa các điểm dữ liệu có giá trị ùn tắc cực đại cao nhất trên mạng lưới. Sự đối lập này cho thấy sự tồn tại của các nút thắt cục bộ có tính chất bất lợi, nơi khả năng lưu thông có thể bị gián đoạn hoàn toàn khi tiệm cận các ngưỡng đứt gãy.

Phân tích trực quan thông qua biểu đồ hộp tại Hình 6 xác nhận cấu trúc phi tuyến của các vùng nhạy cảm. Sự phân tán rộng của các điểm dữ liệu bất thường tại vùng ngoại vi và vùng trung tâm chỉ ra rằng đây là các khu vực có giới hạn chịu đựng yếu. Trong các kịch bản kiểm thử, những vùng này sẽ đóng vai trò là

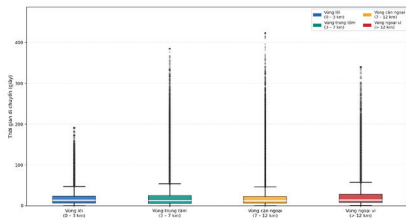
Bảng 3: Thống kê so sánh hai kịch bản

Kịch bản	Trung bình	Độ lệch chuẩn	Min	25%	Trung vị	75%	Max
Kịch bản cơ sở	15.6526	18.3364	0.4070	5.4238	10.1128	18.9177	157.1871
Mưa + Sự kiện	30.7081	35.8385	0.8356	10.8208	20.2405	36.3128	281.9753

Bảng 4: Thống kê ùn tắc theo vùng

Vùng	Trung bình	Độ lệch chuẩn	Min	Q1 (25%)	Trung vị	Q3 (75%)	Max	IQR (biên độ)
Vùng lõi (0 - 3 km)	18.5270	18.9639	0.4263	6.8780	13.5695	23.1996	191.3686	16.3216
Vùng trung tâm (3 - 7 km)	21.6534	28.0645	0.3876	5.9144	12.8896	25.3351	385.1975	19.4207
Vùng cận ngoại (7 - 12 km)	18.9269	24.1735	0.4914	6.7239	12.3629	22.6560	423.5278	15.9321
Vùng ngoại vi (> 12 km)	24.7918	30.1857	0.9915	8.3732	14.1417	28.1198	340.0147	19.7466

tác nhân gây ra sự suy giảm chức năng, làm chậm chu kỳ phục hồi của toàn bộ hệ thống. Việc định vị chính xác cấu trúc không gian của các khu vực này cung cấp cơ sở để thiết lập mức độ ưu tiên phân bổ nguồn lực và triển khai các chiến lược kiểm soát giao thông cục bộ.



Hình 6: Biểu đồ boxplot theo vùng

4.4. Động lực học lan truyền

Đánh giá động lực học lan truyền ùn tắc là nền tảng phân tích thiết yếu nhằm giải mã cơ chế suy giảm khả năng phục hồi của cấu trúc mạng lưới trên trục không - thời gian. Dữ liệu thực chứng tại Bảng 5 cung cấp tiền đề phân tích chi tiết về sự biến thiên của chỉ số ùn tắc trong chu kỳ 150 phút, với thời điểm hệ thống chịu tải lượng phương tiện cực đại T_0 làm hệ quy chiếu trung tâm. Tại pha tích lũy áp lực kéo dài từ $T_0 - 60'$ đến $T_0 - 30'$, chỉ số ùn tắc trung bình duy trì ở ngưỡng cao, minh chứng cho sự hình thành và gia tăng mật độ của các điểm nghẽn cục bộ. Khi tiến đến thời điểm T_0 , mạng lưới chậm ngưỡng bão hòa với giá

trị ùn tắc trung bình đạt mức cực đại là 43.4089. Đáng chú ý, quá trình hệ thống tự phân tán áp lực sau đó diễn ra theo một cấu trúc hoàn toàn phi tuyến. Cụ thể, thay vì suy giảm đồng đều, chỉ số này dao động phức tạp và gia tăng trở lại mức 43.4061 tại $T_0 + 60'$ trước khi thực sự phân tán xuống mức 24.7287 tại $T_0 + 90'$. Sự trì hoãn này cung cấp luận cứ toán học chứng minh cho sự tồn tại của các sóng xung kích giao thông và hiện tượng trào ngược hàng chờ, trực tiếp làm gián đoạn và kéo dài chu kỳ phục hồi của mạng lưới.

Cơ chế hình thành hiện tượng suy giảm phi tuyến này được thể hiện thông qua chuỗi biểu đồ diễn biến lan truyền tại Hình 7. Các phân tích trực quan xác nhận mối tương quan không gian chặt chẽ, trong đó các cụm quá tải cục bộ không bị cô lập mà liên tục khuếch tán dọc theo các liên kết của đồ thị mạng lưới ra các phân vùng lân cận. Sự mở rộng cấu trúc không gian này biến các điểm nghẽn cục bộ thành các tác nhân trực tiếp kích hoạt hiện tượng suy giảm chức năng toàn cục. Sự đồng thuận giữa các tham số biến thiên thời gian và cấu trúc mở rộng không gian khẳng định rằng năng lực phục hồi của mạng lưới không phải là một tham số tĩnh, mà thực chất là một quá trình động lực học phức tạp phụ thuộc chặt chẽ vào cấu trúc hình học hạ tầng và tốc độ

khuếch tán nhiễu loạn dưới các tác động cộng hưởng.

5. KẾT LUẬN

Nghiên cứu này đã đề xuất một khung phương pháp luận tiên tiến nhằm đánh giá định lượng khả năng phục hồi của mạng lưới giao thông đô thị dưới các tác động cộng hưởng. Thông qua việc kiến tạo môi trường bản sao số độ trung thực cao cho khu vực hạ tầng lõi, nghiên cứu đã giải quyết vấn đề khan hiếm dữ liệu trong các kịch bản bất lợi. Sự tích hợp của thuật toán XGBoost đã chứng minh ưu thế trong việc giải mã các động lực học phi tuyến của dòng giao thông dưới kịch bản nhiễu loạn hỗn hợp, bao gồm suy giảm năng lực do biến động khí tượng và gia tăng nhu cầu vận tải đột biến.

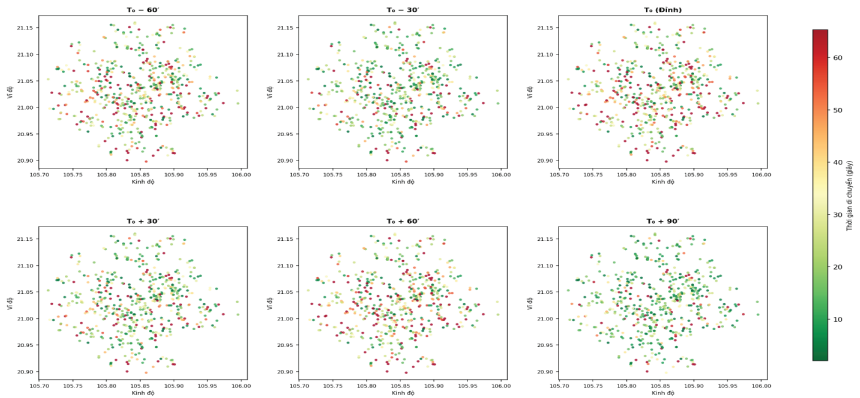
Kết quả thực chứng chỉ ra rằng, sức chịu tải của hệ thống không mang tính tuyến tính mà tồn tại các ngưỡng đứt gãy. Sự tiệm cận các giới hạn này kích hoạt hiệu ứng lan truyền ùn tắc không - thời gian, làm suy giảm cục bộ năng lực thông hành và gia tăng theo cấp số nhân thời gian phục hồi.

Về mặt thực tiễn, nghiên cứu tạo tiền đề khoa học để dịch chuyển mô hình quản lý giao thông từ đáp ứng thụ động sang đánh giá ranh giới phục hồi chủ động. Các tham số về ngưỡng đứt gãy cần được đưa trực tiếp vào cấu trúc của hệ thống ITS để tối ưu hóa chiến lược điều tiết lưu lượng động và phân bổ nguồn lực khẩn cấp. Trong tương lai, việc mở rộng quy mô đô thị không gian và tích hợp các luồng dữ liệu đa phương thức theo thời gian thực sẽ tiếp tục hoàn thiện khung đánh giá, định hướng xây dựng hạ tầng đô thị thông minh và bền vững.■

Lời cảm ơn: Nghiên cứu này là một phần nội dung nằm trong đề tài trọng điểm cấp trường năm học 2024 - 2025: “Xây dựng mô hình AI: Dự đoán ùn tắc giao thông dựa trên lịch sử và thời gian thực”, mã số ĐTTĐ-21, Chủ nhiệm đề tài: Đặng Thị Minh. Nhóm tác giả chân

Bảng 5: Thống kê ùn tắc theo khung thời gian quanh đỉnh

Khung thời gian	Lệch (phút)	Trung bình	Trung vị	Độ lệch chuẩn	Max
$T_0 - 60'$	-60	39.3532	25.5431	46.5448	423.5278
$T_0 - 30'$	-30	30.2429	18.7649	34.8858	271.6802
T_0 (Đỉnh)	0	43.4089	27.7843	50.2567	423.5278
$T_0 + 30'$	30	32.3626	20.8571	38.1939	311.3019
$T_0 + 60'$	60	43.4061	27.7843	50.2567	423.5278
$T_0 + 90'$	90	24.7287	15.8023	29.0974	236.3719



Hình 7: Biểu đồ diễn biến lan truyền

thành cảm ơn sự hỗ trợ về kinh phí của Trường Đại học Công nghệ

 Giao thông vận tải để nhóm tác giả thực hiện các nội dung trong đề tài.

TÀI LIỆU THAM KHẢO

[1] S. Afandizadeh, S. Abdolahi, và H. Mirzahassein, “Deep Learning Algorithms for Traffic Forecasting: A Comprehensive Review and Comparison with Classical Ones”, Journal of Advanced Transportation, vol 2024, số p.h 1, tr 9981657, 2024, doi: 10.1155/2024/9981657.

[2] K. Saini và S. Sharma, “Smart Road Traffic Monitoring: Unveiling the Synergy of IoT and AI for Enhanced Urban Mobility”, ACM Comput. Surv., vol 57, số p.h 11, tr 276:1-276:45, Tháng Sáu 2025, doi: 10.1145/3729217.

[3] A. Mystakidis, P. Koukaras, và C. Tjortjis, “Advances in Traffic Congestion Prediction: An Overview of Emerging Techniques and Methods”, Smart Cities, vol 8, số p.h 1, tháng 2 2025, doi: 10.3390/smartcities8010025.

[4] M. Shaygan, C. Meese, W. Li, X. (George) Zhao, và M. Nejad, “Traffic prediction using artificial intelligence: Review of recent advances and emerging opportunities”, Transportation Research Part C: Emerging Technologies, vol 145, tr 103921, tháng 12 2022, doi: 10.1016/j.trc.2022.103921.

[5] A.-L. Toba, S. Kulkarni, W. Khallouli, và T. Pennington, “Long-Term Traffic Prediction Using Deep Learning Long Short-Term Memory”, Smart Cities, vol 8, số p.h 4, tháng 7 2025, doi: 10.3390/smartcities8040126.

[6] C. Bentéjac, A. Csörgő, và G. Martínez-Muñoz, A Comparative Analysis of XGBoost. 2019. doi: 10.48550/arXiv.1911.01914.

[7] J. Friedman, “Greedy Function Approximation: A Gradient Boosting Machine”, The Annals of Statistics, vol 29, tháng 11 2000, doi: 10.1214/aos/1013203451.

Ngày nhận bài: 08/5/2026; Ngày chấp nhận đăng: 29/5/2026