



APPLICATION OF POISSON AND NEGATIVE BINOMIAL MODELS IN EXPRESSWAY TRAFFIC ACCIDENT PREDICTION

Chu Tien Dung*, Le Xuan Quy

University of Transport and Communications, No 3 Cau Giay Street, Hanoi, Vietnam

ARTICLE INFO

TYPE: Research Article

Received: 12/03/2025

Revised: 23/07/2025

Accepted: 10/08/2025

Published online: 15/08/2025

<https://doi.org/10.47869/tcsj.76.6.1>

* *Corresponding author*

Email: dungchu@utc.edu.vn; Tel: 0912777219

Abstract. Traffic accidents on the expressway network remain a major challenge to road safety in Vietnam, requiring effective solutions to mitigate risks. This study uses data from 108 traffic accidents that occurred on expressways between January 2023 and February 2025 to analyze the impact of highway infrastructure and traffic characteristics on accident severity. The results show that vehicle type, accident type, and infrastructure conditions are significant factors: accidents involving coaches, head-on collisions, rollovers, and roads with a higher number of lanes tend to increase injury severity; whereas the presence of median barriers and emergency lanes significantly reduces the risk of fatalities and injuries. The study also highlights the importance of considering road design elements when developing safety policies for expressways. Through a comparative analysis of three regression models—Ordinary Least Squares (OLS), Poisson regression, and Negative Binomial (NB) regression—the NB model is identified as the most appropriate for predicting the severity of traffic accidents on expressways due to its ability to handle over-dispersed count data effectively.

Keywords: Traffic accident prediction, Poisson model, Negative Binomial model, Fatalities, Injuries, Expressway.

@ 2025 University of Transport and Communications



ỨNG DỤNG MÔ HÌNH POISSON VÀ MÔ HÌNH NHỊ THỨC ÂM TRONG DỰ BÁO TAI NẠN GIAO THÔNG TRÊN ĐƯỜNG CAO TỐC

Chu Tiến Dũng*, Lê Xuân Quý

Trường Đại học Giao thông vận tải, Số 3 Cầu Giấy, Hà Nội, Việt Nam

THÔNG TIN BÀI BÁO

CHUYÊN MỤC: Công trình khoa học

Ngày nhận bài: 12/03/2025

Ngày nhận bài sửa: 23/07/2025

Ngày chấp nhận đăng: 10/08/2025

Ngày xuất bản Online: 15/08/2025

<https://doi.org/10.47869/tcsj.76.6.1>

* Tác giả liên hệ

Email: dungchu@utc.edu.vn; Tel: 0912777219

Tóm tắt. Tai nạn giao thông trên hệ thống đường cao tốc tiếp tục là một thách thức lớn đối với an toàn giao thông tại Việt Nam, đòi hỏi các giải pháp hiệu quả nhằm giảm thiểu rủi ro. Nghiên cứu này sử dụng dữ liệu từ 108 vụ tai nạn xảy ra trên các tuyến cao tốc để phân tích tác động của các yếu tố kết cấu hạ tầng đường cao tốc và đặc điểm giao thông đến mức độ nghiêm trọng của tai nạn. Kết quả cho thấy loại phương tiện, dạng tai nạn và điều kiện kết cấu hạ tầng là những yếu tố có ảnh hưởng đáng kể: tai nạn liên quan đến xe khách, va chạm đối đầu, lật xe và số làn xe lớn có xu hướng làm tăng mức độ thương vong; trong khi sự hiện diện của dải phân cách và làn dừng khẩn cấp giúp giảm đáng kể nguy cơ tử vong và chấn thương. Qua phân tích so sánh ba mô hình hồi quy gồm: hồi quy tuyến tính (HQT), hồi quy Poisson và hồi quy nhị thức âm (NTÂ), mô hình nhị thức âm được xác định là phù hợp nhất trong việc dự báo mức độ nghiêm trọng của tai nạn giao thông trên đường cao tốc.

Từ khóa: Dự báo tai nạn giao thông, Mô hình Poisson, Mô hình nhị thức âm, Tử vong, Chấn thương, đường cao tốc.

@ 2025 Trường Đại học Giao thông vận tải

1. ĐẶT VẤN ĐỀ

Tai nạn giao thông đường bộ (TNGTĐB) vẫn là vấn đề nghiêm trọng toàn cầu. Theo WHO [1], năm 2021 có khoảng 1,19 triệu người tử vong do TNGTĐB, giảm 5% so với 2010. Tỷ lệ tử vong toàn cầu là 15/100.000 người, giảm 16% trong bối cảnh số phương tiện, dân số

và mạng lưới đường tăng mạnh. Tuy nhiên, mục tiêu giảm 50% tử vong trong “Thập kỷ Hành động vì An toàn Đường bộ 2011–2020” chưa đạt được, và khả năng hoàn thành mục tiêu tương tự theo các “Mục tiêu Phát triển Bền vững toàn cầu” vẫn còn xa.

Báo cáo WHO năm 2023 cho thấy TNGTDB là thách thức lớn về y tế và phát triển. Năm 2019, đây là nguyên nhân tử vong hàng đầu ở nhóm 5–29 tuổi, đứng thứ 12 ở mọi lứa tuổi. Gần 2/3 số ca tử vong thuộc nhóm lao động (18–59 tuổi). Nhóm đi xe máy/ba bánh chiếm 30% số ca tử vong, tiếp đến là ô tô (25%), người đi bộ (21%), xe đạp (5%) và xe tải/xe khách (19%). Khoảng 90% ca tử vong xảy ra ở các nước thu nhập thấp và trung bình, đặc biệt là Đông Nam Á (28%) và Tây Thái Bình Dương (25%).

Tại Đông Nam Á, Việt Nam có tỷ lệ tử vong do TNGT cao, chỉ sau Thái Lan. Theo UBATGT [2], 11 tháng đầu năm 2024 ghi nhận 21.691 vụ TNGT, 10.026 người tử vong và 16.103 người bị thương. So với cùng kỳ 2023: tăng 5,67% số vụ, giảm 7,89% ca tử vong, tăng 13,43% số người bị thương. Trên đường cao tốc, tuy chỉ chiếm khoảng 1% tổng số vụ nhưng hậu quả thường nghiêm trọng [3], đòi hỏi các giải pháp mạnh mẽ hơn.

Việt Nam bắt đầu xây dựng hệ thống cao tốc từ 1998. Đến năm 2025, tuyến Bắc – Nam phía Đông dài 2.063 km hoàn thành, nâng tổng chiều dài mạng lưới lên khoảng 3.000 km. Sau 2030, khi hoàn thiện tuyến phía Tây, hệ thống dự kiến đạt 8.400 km [4]. Một số tuyến có từ 4–6 làn, trang bị đầy đủ dải phân cách và làn dừng khẩn cấp, tốc độ tối đa 100–120 km/h. Tuy nhiên, do hạn chế ngân sách, nhiều tuyến phát triển theo giai đoạn, thiếu dải phân cách và làn dừng, tốc độ chỉ 80–90 km/h.

Ngoài yếu tố người điều khiển và thời tiết, nhiều TNGT trên cao tốc còn bắt nguồn từ hạn chế hạ tầng, nhất là ở các tuyến phân kỳ thiếu dải phân cách và làn dừng khẩn cấp. Điều này đặt ra yêu cầu cải thiện đồng bộ kết cấu hạ tầng để giảm nguy cơ tai nạn.

Tuy nhiên, nghiên cứu trong nước về yếu tố ảnh hưởng và mô hình dự báo TNGT trên cao tốc – đặc biệt nghiên cứu định lượng với mô hình thống kê – còn hạn chế. Do đó, nghiên cứu này nhằm xây dựng mô hình dự báo mức độ thương vong do TNGT và so sánh hiệu suất giữa các mô hình: Poisson, nhị thức âm (NTÂ), và hồi quy tuyến tính (HQT) – để xác định mô hình phù hợp nhất với dữ liệu đếm và tính chất quá phân tán của TNGT.

Ngoài ra, nghiên cứu cũng phân tích nguyên nhân, đánh giá tác động của các yếu tố rủi ro và hậu quả trên mạng lưới cao tốc. Trên cơ sở đó, nhóm nghiên cứu đề xuất các giải pháp giảm thiểu tai nạn, thương vong và thiệt hại, đóng góp cơ sở khoa học cho quản lý và hoạch định chính sách an toàn giao thông đường cao tốc tại Việt Nam.

2. TỔNG QUAN

Tai nạn giao thông trên đường cao tốc là một vấn đề phức tạp, chịu ảnh hưởng bởi nhiều yếu tố như đặc điểm hình học tuyến đường, điều kiện vận hành, đặc tính phương tiện và hành vi người lái. Các nghiên cứu quốc tế đã chỉ ra rằng mức độ nghiêm trọng và tần suất tai nạn bị chi phối đáng kể bởi các yếu tố như tốc độ xe, điều kiện mặt đường và yếu tố hình học như bán kính đường cong nằm, chiều dài đường cong chuyển tiếp, chiều dài đoạn dốc, bán kính đường cong đứng lồi và lõm [5-7].

Qin và các cộng sự [5] đã nghiên cứu nguy cơ tai nạn trên các tuyến cao tốc vùng đồi núi, nơi địa hình phức tạp và điều kiện môi trường bị giới hạn bởi yếu tố tự nhiên, cho thấy rủi ro tai nạn có xu hướng gia tăng rõ rệt. Về đặc điểm hình học tuyến đường, các yếu tố như chiều dài đoạn dốc có tương quan dương với nguy cơ tai nạn. Các đoạn dốc dài, cả lên và xuống,

thường đi kèm với mức độ nguy hiểm cao hơn. Bán kính đường cong nằm nhỏ làm tăng lực ly tâm, dẫn đến nguy cơ trượt bánh, lật xe và hạn chế tầm nhìn. Đường cong chuyển tiếp ngắn gây khó khăn cho người lái trong việc điều chỉnh hướng di chuyển ở tốc độ cao, làm tăng khả năng mất ổn định ngang. Tương tự, bán kính nhỏ của các đường cong đứng lồi và lõm làm giảm tầm nhìn và ảnh hưởng xấu đến khả năng kiểm soát phương tiện khi lưu thông ở tốc độ lớn.

Về khía cạnh môi trường lái xe, nghiên cứu [5] cũng cho thấy tỷ lệ và tốc độ biến thiên của xe tải và xe con có tương quan dương với nguy cơ tai nạn, nhấn mạnh vai trò của lưu lượng và thành phần dòng xe trong việc ảnh hưởng đến an toàn giao thông. Đặc biệt, lưu lượng xe tải cao làm gia tăng rủi ro do bị che khuất tầm nhìn và hậu quả tai nạn thường nghiêm trọng hơn. Nguy cơ cũng tăng lên tại các khu vực như điểm dừng nghỉ và lối ra vào, nơi xảy ra sự pha trộn giữa các luồng giao thông và thay đổi tốc độ đột ngột, dẫn đến nhiều loạn dòng xe. Ngược lại, tỷ lệ thảm thực vật và đường có tương quan âm với nguy cơ tai nạn, gợi ý rằng môi trường đơn giản và tầm nhìn tốt giúp người lái giảm căng thẳng và ra quyết định kịp thời hơn.

Các nghiên cứu khác như của Kumar và các cộng sự [6] và Madushani và cộng sự [7] chỉ ra rằng các khu vực có đặc điểm địa hình phức tạp, điểm chuyển tiếp hình học (như đoạn vào – ra hầm, đoạn giao cắt) hoặc các đoạn đường có lưu lượng lớn là những vị trí có nguy cơ xảy ra tai nạn cao. Trong khi đó, Yang và các cộng sự [8] cũng như Bisht & Tiwari [9] nhấn mạnh tầm quan trọng của việc xác định các điểm đen tai nạn để có thể áp dụng các biện pháp can thiệp hiệu quả.

Về mặt phương pháp nghiên cứu, dữ liệu tai nạn giao thông thường có dạng đếm (số vụ tai nạn, số người tử vong, số người bị thương), nên các mô hình thống kê đếm như hồi quy Poisson và hồi quy NTÂ được sử dụng rộng rãi. Mô hình Poisson là cơ sở phổ biến nhưng nhạy cảm với hiện tượng phân tán quá mức – một đặc trưng phổ biến trong dữ liệu tai nạn [10]. Mô hình NTÂ khắc phục được hạn chế này và đã chứng minh hiệu quả trong nhiều nghiên cứu như của Zhang và các cộng sự [11], Naghawi và các cộng sự [12], Yan và các cộng sự [13], Ma và các cộng sự [14], và Ihueze & Onwurah [15]. Một số mô hình cải tiến như NTÂ có hiệu ứng ngẫu nhiên cũng được sử dụng nhằm xử lý biến thiên chưa quan sát được giữa các vị trí nghiên cứu.

Bảng 1. Tổng hợp một số nghiên cứu trên thế giới.

Tác giả	Quốc gia	Phương pháp	Yếu tố chính phân tích	Ghi chú
Qin và các cộng sự [5]	Trung Quốc	Mô hình học máy	Tỷ lệ xe tải, đoạn dốc dài, bán kính cong nhỏ	Dự báo mức độ rủi ro tai nạn
Kumar và các cộng sự [6]	Ấn Độ	Mô hình logit	Tốc độ, hình học tuyến, loại xe	Phân tích mức độ nghiêm trọng
Madushani và các cộng sự [7]	Sri Lanka	Mô hình học máy	Mặt đường, thời tiết, ánh sáng	Dự báo mức độ nghiêm trọng
Zhang và các cộng sự [13]	Trung Quốc	Poisson, NTÂ	Tốc độ, hình dạng tuyến	Phân tán quá mức được xử lý bằng NTÂ
Ma và các cộng sự [14]	Trung Quốc	NTÂ + Hiệu ứng ngẫu nhiên	Lưu lượng, độ dốc, chiều dài đoạn	So sánh NTÂ và NTÂ hiệu ứng ngẫu nhiên

Tại Việt Nam, các nghiên cứu áp dụng mô hình thống kê đếm để phân tích – dự báo tai nạn trên hệ thống đường cao tốc – vốn đang mở rộng mạnh mẽ – vẫn còn thiếu hụt. Đây là khoảng trống nghiên cứu mà bài báo này hướng tới, thông qua việc so sánh hiệu suất dự báo giữa các mô hình Poisson, NTÂ và HQT, trên cơ sở dữ liệu tai nạn thu thập được tại các

tuyến đường cao tốc ở Việt Nam.

3. THU THẬP VÀ PHÂN TÍCH THỐNG KÊ DỮ LIỆU

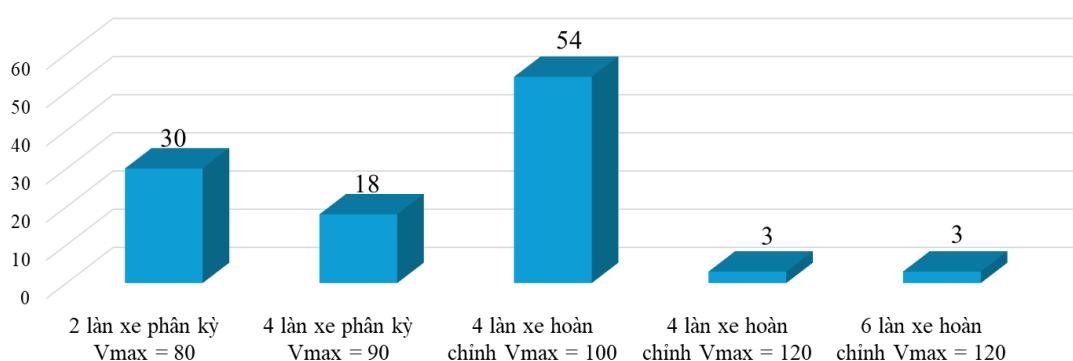
1. Thu thập dữ liệu

Nghiên cứu sử dụng bộ dữ liệu gồm 108 vụ tai nạn giao thông xảy ra trên các tuyến cao tốc ở Việt Nam từ 1/1/2023 đến 20/2/2025, thu thập thủ công từ các bản tin và báo mạng thông qua tính năng “tìm kiếm nâng cao” của Google Chrome với từ khóa “tai nạn giao thông trên cao tốc”, giới hạn trong lãnh thổ Việt Nam. Dữ liệu được đối chiếu từ nhiều nguồn để loại bỏ trùng lặp, thiếu chi tiết và được mã hóa, hệ thống hóa bằng Excel phục vụ phân tích định lượng. Mỗi vụ tai nạn ghi nhận các thông tin: tuyến đường, thời gian, số người tử vong và bị thương (theo mức độ), số phương tiện liên quan, nguyên nhân sơ bộ. Thông tin kỹ thuật tuyến cao tốc được bổ sung từ tài liệu thiết kế và cơ sở dữ liệu công khai, gồm số làn xe, chiều rộng, dải phân cách, làn dừng khẩn cấp và tốc độ tối đa. Trước khi phân tích, dữ liệu được xử lý sơ bộ để loại bỏ bản ghi thiếu thông tin, chuẩn hóa định dạng và đảm bảo tính nhất quán, sẵn sàng cho các mô hình thống kê trình bày ở phần tiếp theo. Các thông tin chính của dữ liệu được tóm tắt trong Bảng 2.

Bảng 2. Thống kê mô tả dữ liệu về tai nạn giao thông trên đường cao tốc.

STT	Giai đoạn đầu tư	Số làn	Bề rộng làn	Tốc độ tối đa	Làn dừng khẩn cấp	Dải phân cách	Số vụ	Chết	Bị thương nặng	Bị thương nhẹ
1	Phân kỳ	2	3,50	80	Không ^{*)}	Không	30	39	66	57
2	Phân kỳ	4	3,50	90	Không	Không	18	12	3	0
3	Hoàn chỉnh	4	3,75	100	Có	Có	54	12	60	39
4	Hoàn chỉnh	4	3,75	120	Có	Có	3	3	0	3
5	Hoàn chỉnh	6	3,75	120	Có	Có	3	6	12	18
Tổng							108	72	141	117

Ghi chú: ^{*)} Không có làn dừng khẩn cấp nhưng có 2x2m lề gia cố



Hình 1. Số vụ tai nạn theo loại cao tốc, số làn xe và tốc độ tối đa.

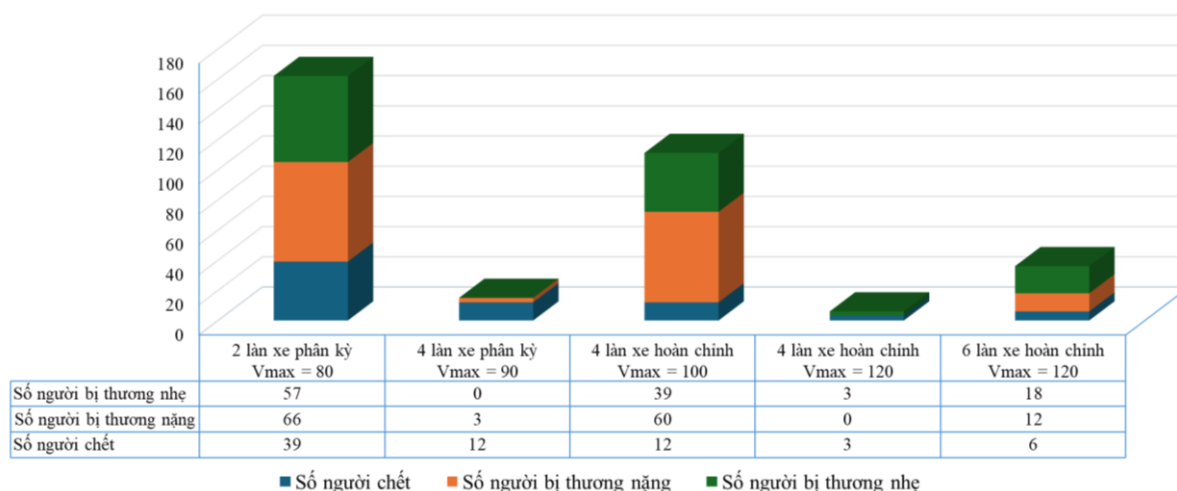
2. Yếu tố hạ tầng đường cao tốc và tai nạn giao thông

Dữ liệu tai nạn giao thông trên cao tốc cho thấy sự khác biệt rõ rệt theo giai đoạn đầu tư, số làn xe, giới hạn tốc độ và hạ tầng như làn dừng khẩn cấp, dải phân cách.

Trong giai đoạn phân kỳ, xảy ra 48 vụ tai nạn, chủ yếu trên tuyến 2 làn tốc độ 80 km/h (30 vụ, 39 người tử vong). Tuyến 4 làn tốc độ 90 km/h có 18 vụ (12 người tử vong). Giai đoạn hoàn chỉnh ghi nhận 60 vụ, tập trung trên tuyến 4 làn tốc độ 100 km/h (54 vụ, 12 người tử vong). Tuyến 4 làn và 6 làn tốc độ 120 km/h chỉ có 3 vụ mỗi tuyến, nhưng tuyến 6 làn gây hậu quả nghiêm trọng hơn (6 người tử vong, 12 người bị thương nặng) (Hình 1).

Xét theo số làn, tuyến 2 làn có tỷ lệ tai nạn cao nhất (27,8%), trong khi tuyến 6 làn thấp nhất (2,8%). Về tốc độ, các tuyến dưới 100 km/h có số vụ cao hơn, đặc biệt tuyến 2 làn tốc độ 80 km/h. Dù các tuyến tốc độ 120 km/h có ít tai nạn hơn, mức độ nghiêm trọng vẫn cao do va chạm ở tốc độ lớn.

Sự khác biệt về thương vong giữa các tuyến được thể hiện rõ trong Hình 2. Tuyến 2 làn có số thương vong cao nhất (39 tử vong, 66 bị thương nặng, 57 bị thương nhẹ), do mật độ xe cao, thiếu làn dừng và khả năng tránh va chạm thấp. Tuyến 4 làn giai đoạn phân kỳ (90 km/h) có thương vong thấp hơn nhiều, nhưng khi nâng cấp lên tốc độ 100 km/h lại ghi nhận thương vong tăng, cho thấy tốc độ này chưa phù hợp. Trong khi đó, các tuyến 4–6 làn tốc độ 120 km/h có thương vong thấp nhất, phản ánh hiệu quả của thiết kế hiện đại với làn dừng và dải phân cách đầy đủ.

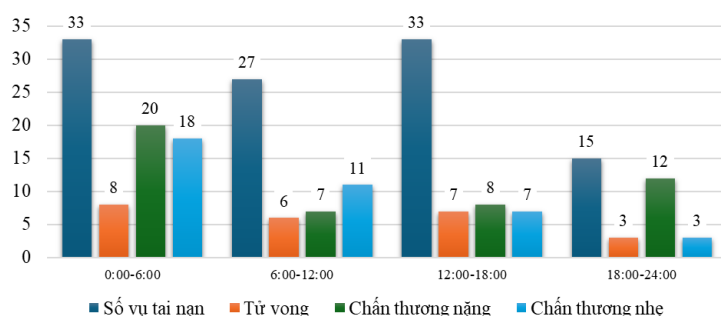


Hình 2. Số ca tử vong và chấn thương theo loại cao tốc, số làn xe và tốc độ tối đa.

Làn dừng khẩn cấp và dải phân cách đóng vai trò quan trọng trong giảm tai nạn. Các tuyến không có làn dừng, chỉ có lề gia cố 2x2m, ghi nhận 48 vụ tai nạn – cao hơn đáng kể so với các tuyến có hạ tầng đầy đủ (60 vụ dù tổng chiều dài lớn hơn). Tuyến 2 làn tốc độ 80 km/h, không có làn dừng hay dải phân cách cứng, có tỷ lệ thương vong cao nhất, cho thấy cần sớm nâng cấp hạ tầng. Ngược lại, các tuyến 4–6 làn hiện đại, tốc độ cao hơn lại an toàn hơn, khẳng định vai trò then chốt của thiết kế hạ tầng trong giảm thiểu thương vong.

3. Khung giờ xảy ra tai nạn giao thông

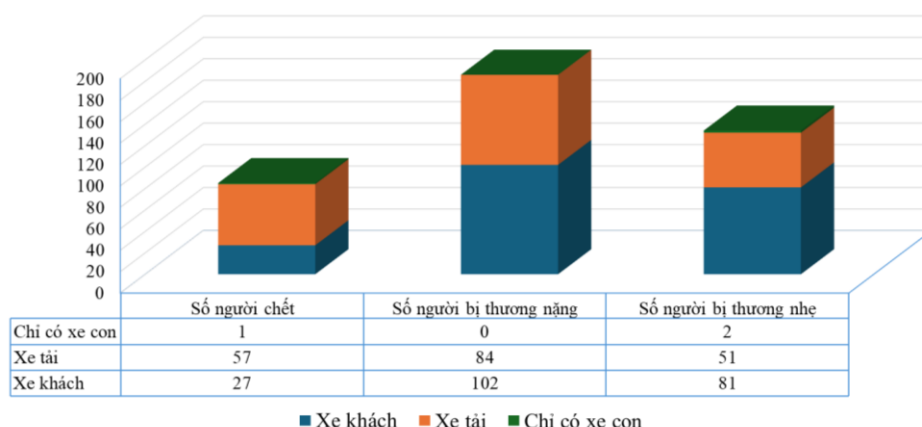
Mối quan hệ giữa khung giờ tai nạn và mức độ thương vong được thể hiện trong Hình 3. Tai nạn xảy ra nhiều nhất vào 0:00–6:00 và 12:00–18:00 (33 vụ), tiếp theo là 6:00–12:00 (27 vụ); ít nhất là 18:00–24:00 (15 vụ). Khung 0:00–6:00 cũng ghi nhận thương vong cao nhất (8 người tử vong, 20 bị thương nặng, 18 bị thương nhẹ), có thể do lái xe mệt mỏi, tốc độ cao, tầm nhìn kém và thời gian cứu hộ chậm. Dù ít vụ hơn, khung 18:00–24:00 vẫn có 12 ca bị thương nặng. Các khung giờ còn lại ghi nhận thương vong ở mức trung bình.



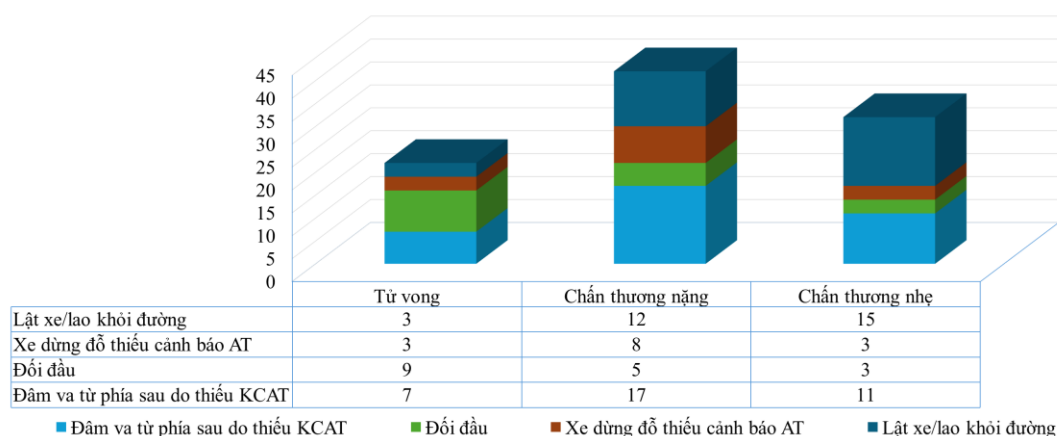
Hình 3. Mối quan hệ giữa số vụ tai nạn, mức độ thương vong và khung giờ xảy ra tai nạn.

3.1. Loại phương tiện, loại hình tai nạn và mức độ thương vong

Hình 4 cho thấy tai nạn liên quan đến xe khách có số ca chấn thương cao nhất (102 nặng, 81 nhẹ, 27 tử vong), còn xe tải có số tử vong cao nhất (57 người), cùng 84 chấn thương nặng và 51 nhẹ. Tai nạn xe con có thương vong thấp nhất (1 tử vong, 2 chấn thương nhẹ). Nguyên nhân là do xe khách chở đông người, ít thắt dây an toàn; xe tải lớn, va chạm mạnh, thường hoạt động ban đêm; còn xe con có hệ thống an toàn tốt, ít người trên xe.



Hình 4. Mức độ thương vong theo phương tiện.



Hình 5. Loại hình tai nạn và mức độ thương vong.

Về loại hình tai nạn (Hình 5), đâm từ phía sau phổ biến nhất (63 vụ, 62%), thường gây

nhiều chấn thương nhưng ít tử vong. Tai nạn đối đầu tuy chỉ 15 vụ nhưng nguy hiểm nhất (9 tử vong) do va chạm trực diện. Lật xe/lao khỏi đường và đâm vào xe dừng đỗ (đều 12 vụ) gây 3 ca tử vong mỗi loại, chủ yếu khi liên quan đến xe khách hoặc xe tải. Điều này cho thấy cần kiểm soát tốc độ, nâng cao ý thức tài xế và cải thiện kết cấu hạ tầng để giảm thiểu tai nạn.

4. LÝ THUYẾT VỀ CÁC MÔ HÌNH HỒI QUY

4.1. Hồi quy tuyến tính

Hồi quy tuyến tính là một kỹ thuật phổ biến dùng để ước lượng các tham số trong mô hình hồi quy tuyến tính, bằng cách giảm thiểu tổng bình phương sai lệch giữa giá trị quan sát và giá trị dự đoán từ mô hình. Mô hình HQT có dạng tổng quát [16]:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon \quad (1)$$

Y là biến phụ thuộc, $X_1, X_2, \dots, X_k$ là các biến độc lập, $\beta_1, \beta_2, \dots, \beta_k$ là các tham số chưa biết, ε là sai số ngẫu nhiên phân phối chuẩn với trung bình 0 và phương sai không đổi. Phương pháp HQT đơn giản, dễ áp dụng cho biến phụ thuộc liên tục, nhưng không phù hợp với dữ liệu phân phối không chuẩn, phương sai thay đổi hoặc dữ liệu đếm, khi đó mô hình Poisson và NTÂ được sử dụng thay thế.

4.2. Hồi quy Poisson

Phân phối Poisson mô hình hóa xác suất xảy ra y sự kiện (tử vong, chấn thương...) với công thức [17]:

$$\Pr(Y = y|\lambda) = \frac{e^{-\lambda} \lambda^y}{y!} \quad (y = 0, 1, 2, \dots) \quad (2)$$

Phân phối Poisson chỉ sử dụng một tham số duy nhất là λ , đại diện cho tỷ lệ xuất hiện trung bình của sự kiện hiếm trên đơn vị phơi nhiễm, có thể là thời gian, không gian, khoảng cách, diện tích, thể tích, dân số, v.v. Đặc điểm quan trọng của phân phối Poisson là kỳ vọng ($E[Y]$) và phương sai ($Var(Y)$) bằng nhau.

Trong mô hình hồi quy Poisson, giả định rằng kỳ vọng λ được xác định bởi một tập các biến giải thích (X_i):

$$\lambda_i = e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k} \quad (3)$$

Các hệ số $\beta_1, \beta_2, \dots, \beta_k$ là các tham số chưa biết, được ước lượng từ dữ liệu. Khi đó, mô hình hồi quy Poisson cơ bản cho từng quan sát i được viết là:

$$\Pr(Y = y_i|\lambda_i) = \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!} \quad (y_i = 0, 1, 2, \dots) \quad (4)$$

Hồi quy Poisson phù hợp khi phương sai số lần xảy ra sự kiện xấp xỉ bằng giá trị trung bình. Tuy nhiên, trong thực tế, dữ liệu thường có phương sai lớn hơn trung bình (quá phân tán), khiến mô hình Poisson không còn phù hợp và có thể dẫn đến sai lệch trong ước lượng tham số.

4.3. Hồi quy nhị thức âm

Phân phối Poisson có thể được tổng quát hóa bằng cách đưa vào nhiều gamma, trong đó biến gamma có kỳ vọng bằng 1 và tham số tỷ lệ là v . Sự kết hợp giữa phân phối Poisson và gamma (được gọi là phân phối nhị thức âm) có công thức xác suất [18]:

$$\Pr(Y = y_i | \lambda_i, \alpha) = \frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(y_i + 1)\Gamma(\alpha^{-1})} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \lambda_i} \right)^{\alpha^{-1}} \left(\frac{\lambda_i}{\alpha^{-1} + \lambda_i} \right)^{y_i} \quad (5)$$

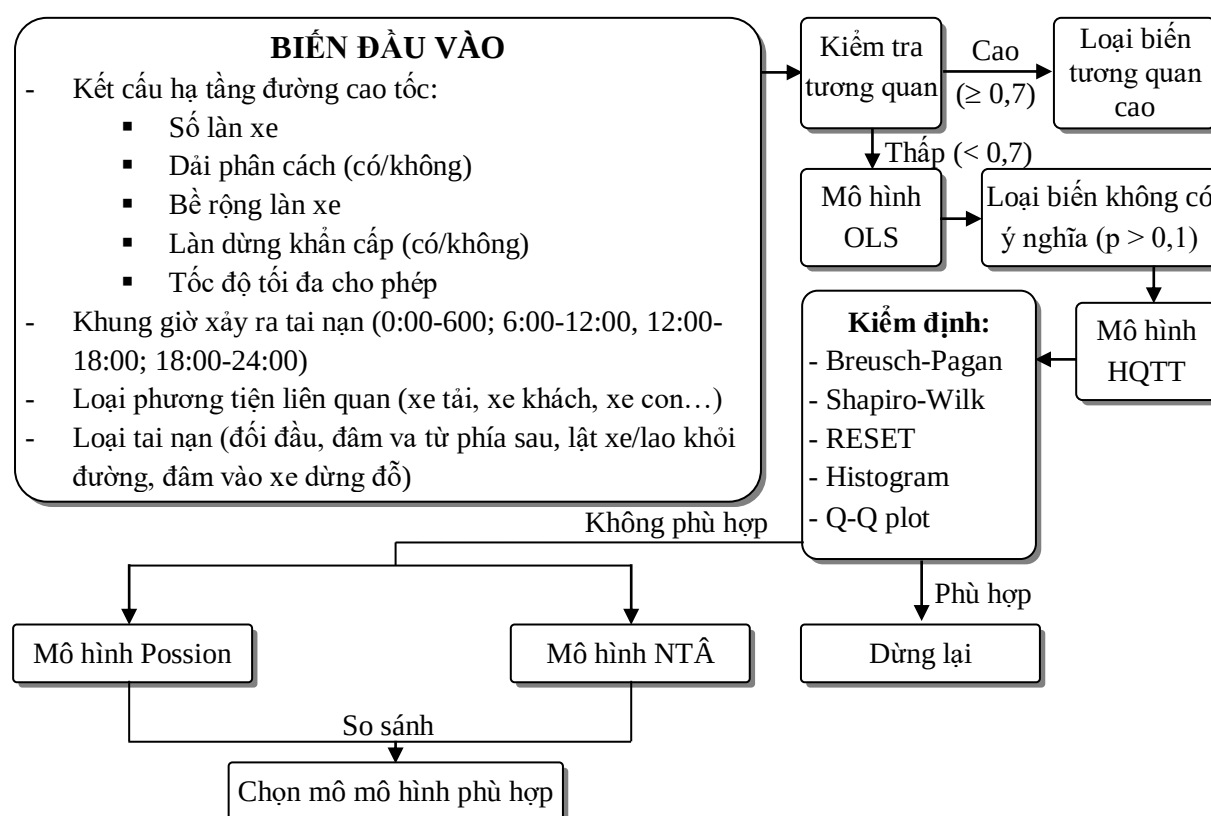
Trong đó: $\alpha = 1/v$: là tham số phân tán (dispersion parameter), tỷ lệ nghịch với mức độ biến thiên. Như vậy, hồi quy NTÂ là một mở rộng của hồi quy Poisson nhằm giải quyết vấn đề quá phân tán bằng cách bổ sung một tham số phân tán α , cho phép phương sai lớn hơn kỳ vọng theo công thức:

$$\text{Var}(Y) = \lambda + \alpha\lambda^2 = \lambda + \frac{\lambda^2}{v} \quad (6)$$

Khi $v \rightarrow \infty$, phương sai của mô hình NTÂ tiến dần về λ , khiến NTÂ trở thành mô hình Poisson, từ đó NTÂ trở nên linh hoạt hơn khi xử lý dữ liệu đếm có phương sai lớn hơn trung bình. Tuy nhiên, NTÂ phức tạp hơn Poisson và cần kiểm định để xác định tính cần thiết. Lựa chọn mô hình phụ thuộc vào tính chất dữ liệu: HQT phù hợp với biến liên tục; Poisson dùng khi phương sai bằng trung bình; còn nếu dữ liệu quá phân tán, NTÂ là lựa chọn tốt hơn. Các kiểm định như overdispersion và Vuong được dùng để chọn mô hình phù hợp.

5. XÂY DỰNG MÔ HÌNH VÀ KẾT QUẢ

5.1. Xây dựng mô hình



Hình 6. Trình tự xây dựng mô hình.

Theo quy định hiện hành của Việt Nam [19], cơ sở dữ liệu tai nạn chỉ ghi nhận các trường hợp tử vong hoặc bị thương, không phân biệt mức độ chấn thương. Do đó, bài báo chỉ

xây dựng mô hình dự báo tổng số ca tử vong hoặc chấn thương.

Bài báo sử dụng và so sánh ba mô hình HQT, Poisson và NTÁ để dự báo số ca tử vong và chấn thương trên đường cao tốc, đồng thời xác định các yếu tố ảnh hưởng (Hình 6). Trước tiên, các biến độc lập được kiểm tra tương quan và loại bỏ các biến có hệ số tương quan cao ($\geq 0,7$). Mô hình HQT được ước lượng và đánh giá bằng các kiểm định Breusch-Pagan, Shapiro-Wilk, RESET và phân tích đồ thị phần dư (histogram, Q-Q plot). Sau đó, các mô hình Poisson và NTÁ được ước lượng và so sánh thông qua kiểm định Vuong và kiểm định overdispersion để chọn mô hình phù hợp nhất.

5.2. Kết quả mô hình hồi quy tuyến tính

Mô hình HQT đã được ước lượng và các kết quả kiểm định của mô hình được thể hiện trên Bảng 3. Mô hình HQT có mức độ giải thích trung bình khá đối với số ca tử vong ($R^2 = 0,45$) và tốt hơn đối với số ca chấn thương ($R^2 = 0,58$). Hệ số R^2 hiệu chỉnh cũng cho thấy xu hướng tương tự, với giá trị cao hơn ở mô hình chấn thương (0,55) so với mô hình dự báo số ca tử vong (0,41). Điều này cho thấy mô hình có khả năng giải thích sự biến thiên của số ca chấn thương tốt hơn. Tuy nhiên, theo Ozili [20], đối với mô hình mà hầu hết các biến độc lập đều có ý nghĩa thống kê, thì giá trị R^2 có thể chấp nhận được là $R^2 \geq 0,5$. Theo tiêu chí này, chỉ có mô hình số ca chấn thương là có thể chấp nhận được. Kết quả kiểm định tổng thể bằng F-statistic đều cho thấy các biến độc lập có ảnh hưởng đáng kể đến biến phụ thuộc ($p < 0,001$). Tuy nhiên, mô hình HQT vi phạm một số giả định quan trọng. Kiểm định Breusch-Pagan ($p < 0,001$) chỉ ra hiện tượng phương sai không đồng nhất. Đồng thời, kiểm định Shapiro-Wilk ($p < 0,001$) cho thấy phần dư không tuân theo phân phối chuẩn, làm giảm tính chính xác của ước lượng. Ngoài ra, RESET test ($p < 0,001$) cũng cho thấy mô hình có vấn đề về tính tuyến tính. Những vi phạm này ảnh hưởng đến độ tin cậy của mô hình, do đó, cần cân nhắc sử dụng các mô hình phù hợp hơn như Poisson hoặc NTÁ để cải thiện độ chính xác và độ phù hợp.

Bảng 3. Kết quả kiểm định của mô hình HQT.

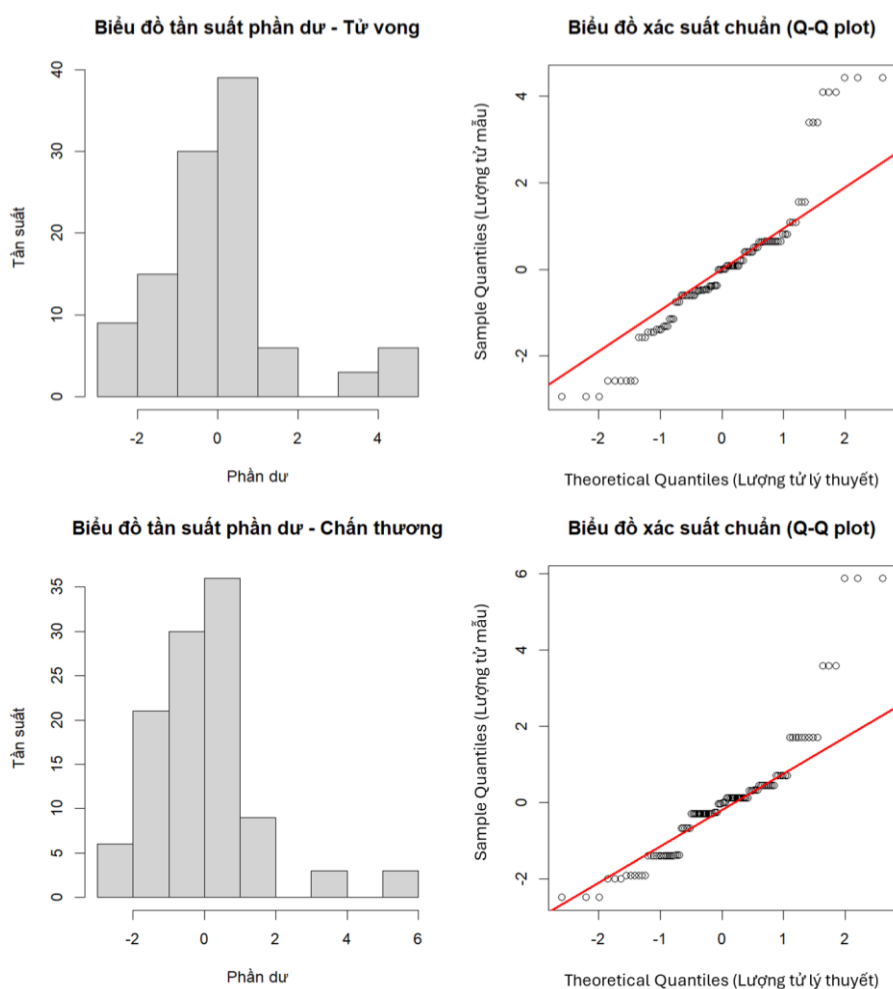
Chỉ số	HQT – Số ca tử vong	HQT – Số ca chấn thương
Hệ số R^2	0,45	0,58
Hệ số R^2 hiệu chỉnh	0,41	0,55
F-statistic (p-value)	10,22 ($p < 0,001$)	19,39 ($p < 0,001$)
Kiểm định Breusch-Pagan (p-value)	35,99 ($p < 0,001$)	49,94 ($p < 0,001$)
Kiểm định Shapiro-Wilk (p-value)	0,90 ($p < 0,001$)	0,85 ($p < 0,001$)
Kiểm định RESET (p-value)	24,064 ($p < 0,001$)	76,39 ($p < 0,001$)

Hai biểu đồ tần suất phần dư và Q-Q plot (Hình 7) cho thấy phần dư của mô hình không tuân theo phân phối chuẩn. Biểu đồ tần suất phần dư có xu hướng lệch, trong khi Q-Q plot thể hiện sự lệch khỏi đường chuẩn ở hai đầu, đặc biệt với số ca tử vong và chấn thương. Điều này củng cố kết quả kiểm định Shapiro-Wilk, cho thấy giả định về phân phối chuẩn của phần dư bị vi phạm, ảnh hưởng đến độ tin cậy của suy luận thống kê từ mô hình HQT.

5.3. Kết quả mô hình hồi quy Poisson và NTÁ

Kết quả kiểm định mô hình Poisson và NTÁ trên Bảng 4 cho thấy NTÁ có Log-Likelihood cao hơn và AIC thấp hơn trong cả hai mô hình dự báo số ca tử vong và chấn thương, cho thấy khả năng phù hợp tốt hơn. Tuy nhiên, các chỉ số đánh giá độ phù hợp như

McFadden R^2 , r^2_{ML} và r^2_{CU} của Poisson đều cao hơn so với NTÂ, đặc biệt là McFadden R^2 (0,37 so với 0,19 đối với mô hình dự báo số ca tử vong và 0,44 so với 0,28 đối với mô hình dự báo số ca chấn thương), cho thấy mô hình Poisson có thể giải thích biến phụ thuộc tốt hơn.



Hình 7. Biểu đồ tần suất phần dư và Q-Q.

Lưu ý rằng, theo Louviere và cộng sự [21], chỉ số McFadden R^2 hiếm khi đạt giá trị gần 1 khi sử dụng dữ liệu thực nghiệm. Để có giá trị bằng 1 (tức là mô hình khớp hoàn hảo), giá trị likelihood phải bằng 1, và do đó log-likelihood sẽ bằng 0. Trên thực tế, các giá trị này thường thấp hơn nhiều so với chỉ số R^2 trong hồi quy tuyến tính. Theo quy ước chung, các giá trị từ 0,2 đến 0,4 có thể được xem là biểu hiện của một mô hình có độ phù hợp tốt. Theo tiêu chí này, mô hình NTÂ dự báo số ca tử vong có phần bất lợi hơn so với mô hình Poisson do hệ số McFadden R^2 chỉ đạt 0,19.

Kiểm định Likelihood Ratio Test (LRT) giữa hai mô hình cho thấy với mô hình dự báo số ca tử vong, giá trị p-value = 0,007 ($< 0,05$) cho thấy NTÂ phù hợp hơn, trong khi với mô hình dự báo số ca chấn thương, p-value = 0,9605 ($> 0,05$) không cho thấy sự khác biệt đáng kể, đồng nghĩa với việc Poisson có thể vẫn được sử dụng. Kiểm định overdispersion cho thấy không có dấu hiệu phân tán mạnh (p-value $> 0,05$ trong cả hai trường hợp), điều này cho thấy mô hình Poisson vẫn có thể áp dụng mà không gây sai lệch đáng kể. Nhìn chung, nếu dữ liệu có overdispersion, mô hình NTÂ sẽ là lựa chọn hợp lý. Tuy nhiên, trong trường hợp không có overdispersion rõ ràng, mô hình Poisson vẫn có thể được ưu tiên nhờ chỉ số R^2 cao hơn và sự

chênh lệch AIC không quá lớn.

Bảng 4. Kết quả kiểm định của mô hình Poisson và NTÂ.

Mô hình	Số ca tử vong		Số ca chấn thương	
Chỉ số	Poisson	NTÂ	Poisson	NTÂ
Log-Likelihood	-137,61	-133,98	-127,26	-126,27
AIC	293,23	287,96	270,53	268,53
McFadden R²	0,37	0,19	0,44	0,28
r²ML	0,77	0,43	0,84	0,61
r²CU	0,79	0,46	0,85	0,63
Likelihood Ratio Test (LRT) giữa Poisson và NTÂ:				
χ² (p-value)	7,26 (p-value = 0,007)		0,0025 (p-value = 0,9605)	
Kiểm định Overdispersion cho NTÂ:				
Tỷ lệ phân tán (p-value)	0,99 (p-value = 0,504)		1,11 (p-value = 0,2173)	

Kết quả phân tích mô hình hồi quy Poisson và NTÂ được thể hiện trong Bảng 4. Nhìn chung, hệ số ước lượng không có nhiều khác biệt giữa hai mô hình cho cả số ca tử vong và chấn thương. Kết quả cho thấy nhiều yếu tố ảnh hưởng đáng kể đến mức độ nghiêm trọng của tai nạn giao thông. Hằng số có giá trị âm lớn (-3,00*** đến -3,39*** cho tử vong, -2,27*** cho chấn thương), cho thấy xác suất xảy ra tai nạn nghiêm trọng trong điều kiện cơ sở là thấp.

Bảng 5. Kết quả ước lượng của mô hình Poisson và NTÂ.

Mô hình	Số ca tử vong		Số ca chấn thương	
	Poisson	NTÂ	Poisson	NTÂ
Hằng số	-3,00***	-3,39***	-2,27***	-2,27***
Tai nạn xảy ra vào 6:00-12:00: 1-đúng, 0-sai	-0,83**	-1,01**	-	-
Tai nạn có liên quan xe khách: 1-đúng, 0-sai	1,47***	1,47***	1,54***	1,54***
Tai nạn có liên quan xe tải: 1-đúng, 0-sai	0,98***	1,14**	0,53*	0,53*
Tai nạn đối đầu: 1-đúng, 0-sai	0,97***	0,98**	0,96***	0,96***
Tai nạn lật xe/lao khỏi đường: 1-đúng, 0-sai	1,36***	1,39***	1,27***	1,27***
Số làn xe/chiều	2,20***	2,43***	1,71***	1,71***
Dải phân cách: 1 có, 0 không có	-4,3***	-4,41***	-3,59***	-3,59***
Làn dừng khẩn cấp: 1 có, 0 không có	-1,43***	-1,44***	-1,13***	-1,13***
Ghi chú: *** p-value < 0,01; ** p-value < 0,05; * p-value < 0,1				

Tai nạn xảy ra trong khoảng 6:00-12:00 có hệ số âm (-0,83** đến -1,01**) trong mô hình dự báo số ca tử vong, cho thấy khoảng thời gian này có thể liên quan đến mức độ nghiêm trọng thấp hơn. Tuy nhiên, biến này không có ý nghĩa đối với mô hình dự báo số ca chấn thương (và đã bị loại bỏ). Có thể các khung giờ khác (chẳng hạn như 0:00 – 6:00) có mức độ thương vong cao hơn do tốc độ di chuyển cao và nguy cơ tài xế buồn ngủ.

Loại phương tiện liên quan đến tai nạn có ảnh hưởng đáng kể. Tai nạn liên quan đến xe khách làm tăng nguy cơ tử vong (1,47***) và chấn thương (1,54***), trong khi xe tải cũng có tác động tăng nhưng ở mức thấp hơn (0,98*** đến 1,14** cho mô hình dự báo số ca tử vong,

0,53* cho mô hình dự báo số ca chấn thương). Điều này hàm ý rằng xe khách thường chở số lượng hành khách lớn, và hiện nay nhiều xe khách không yêu cầu hành khách thắt dây an toàn, dẫn đến hậu quả thương vong nghiêm trọng hơn khi xảy ra tai nạn.

Về nguyên nhân tai nạn, tai nạn đối đầu làm tăng nguy cơ tử vong (0,97*** đến 0,98**) và chấn thương (0,96***), trong khi tai nạn do lật xe hoặc lao khỏi đường có tác động mạnh hơn (1,36*** đến 1,39*** cho mô hình dự báo số ca tử vong, 1,27*** cho mô hình dự báo số ca chấn thương).

Kết cấu hạ tầng đường cao tốc ảnh hưởng đáng kể đến mức độ nghiêm trọng của tai nạn. Số làn xe trên mỗi chiều có hệ số dương lớn (2,2*** đến 2,43*** cho tử vong, 1,71*** cho chấn thương), cho thấy số làn xe càng nhiều, mức độ nghiêm trọng của tai nạn càng cao. Cần lưu ý rằng đường có nhiều làn xe thường có tốc độ khai thác cao (100 hoặc 120 km/h). Do vậy, mặc dù tăng số làn xe có thể giúp giảm số vụ tai nạn giao thông, nhưng nếu tai nạn xảy ra, hậu quả thương vong có thể nghiêm trọng hơn. Ngược lại, sự hiện diện của dải phân cách giúp giảm đáng kể nguy cơ tử vong (-4,3*** đến -4,41***) và chấn thương (-3,59***). Điều này hàm ý rằng các đường cao tốc chỉ có hai làn xe và không có dải phân cách giữa sẽ có nguy cơ thương vong rất cao khi xảy ra tai nạn.

Đáng chú ý, làn dừng khẩn cấp có hệ số âm (-1,43*** đến -1,44*** cho số ca tử vong, -1,13*** cho số ca chấn thương), cho thấy sự hiện diện của làn này giúp giảm mức độ nghiêm trọng của tai nạn. Điều này có thể là do làn dừng khẩn cấp cung cấp không gian an toàn cho phương tiện gặp sự cố, tránh việc dừng trực tiếp trên làn xe chạy của cao tốc.

Nhìn chung, các yếu tố như xe khách, xe tải, số làn xe lớn, và nguyên nhân tai nạn nghiêm trọng như đối đầu hoặc lật xe đều làm tăng mức độ nghiêm trọng của tai nạn. Trong khi đó, dải phân cách và làn dừng khẩn cấp đóng vai trò quan trọng trong việc giảm số ca tử vong và chấn thương. Kết quả này nhấn mạnh tầm quan trọng của việc quy hoạch kết cấu hạ tầng đường cao tốc hợp lý nhằm giảm thiểu hậu quả tai nạn.

6. KẾT LUẬN

Tai nạn giao thông trên hệ thống đường cao tốc tiếp tục là một thách thức lớn đối với an toàn giao thông tại Việt Nam. Dù đã có nhiều nỗ lực cải thiện, số vụ tai nạn và mức độ nghiêm trọng của chúng vẫn đặt ra yêu cầu cấp thiết về các giải pháp hiệu quả nhằm giảm thiểu rủi ro.

Nghiên cứu này đã phân tích dữ liệu từ 108 vụ tai nạn trên các tuyến cao tốc tại Việt Nam để đánh giá ảnh hưởng của các yếu tố kết cấu hạ tầng đường cao tốc và đặc điểm giao thông đến mức độ nghiêm trọng của tai nạn. Kết quả cho thấy, ngoài yếu tố con người và điều kiện thời tiết, những hạn chế về thiết kế kết cấu hạ tầng như thiếu dải phân cách giữa, không có làn dừng khẩn cấp, và tốc độ khai thác không phù hợp có tác động đáng kể đến mức độ thương vong.

Các yếu tố kết cấu này ảnh hưởng đến cơ hội xử lý và giảm nhẹ va chạm khi có sự cố xảy ra. Cụ thể, thiếu dải phân cách giữa làm tăng khả năng xảy ra tai nạn đối đầu, vốn là loại tai nạn có mức độ nghiêm trọng cao. Không có làn dừng khẩn cấp khiến các phương tiện gặp sự cố không có không gian an toàn để dừng, dẫn đến va chạm dây chuyền hoặc cản trở lưu thông. Trong điều kiện vận hành tại Việt Nam, nơi ý thức chấp hành luật giao thông chưa cao và còn tồn tại tình trạng lấn làn, dừng đỗ sai quy định, thì các yếu tố hạ tầng này càng trở nên quan trọng.

So sánh giữa các mô hình hồi quy HQT, Poisson và NTÂ cho thấy mô hình NTÂ có độ phù hợp cao nhất trong việc dự báo số ca tử vong và chấn thương. Kết quả phân tích hồi quy chỉ ra rằng loại phương tiện là yếu tố quan trọng, với xe khách làm tăng nguy cơ tử vong và chấn thương cao hơn so với xe tải, có thể do số lượng hành khách đông và thiết kế an toàn hạn chế. Các loại tai nạn nghiêm trọng như đầu đối đầu hoặc lật xe cũng làm gia tăng mức độ thương vong. Về hạ tầng, số làn xe lớn có thể làm tăng tốc độ khai thác, kéo theo hậu quả nặng nề hơn khi xảy ra tai nạn, trong khi dải phân cách và làn dừng khẩn cấp góp phần giảm mức độ thương vong thông qua cải thiện khả năng xử lý và phản ứng khi có sự cố.

Tuy nhiên, nghiên cứu này vẫn còn một số hạn chế, đặc biệt là do quy mô dữ liệu chưa đủ lớn, chỉ bao gồm 108 vụ tai nạn, có thể ảnh hưởng đến tính chính xác của các mô hình dự báo. Do đó, nghiên cứu trong tương lai cần mở rộng quy mô dữ liệu để nâng cao độ tin cậy của kết quả phân tích. Ngoài ra, các phương pháp tiên tiến như học máy, bao gồm mạng nơ-ron nhân tạo (ANN), rừng ngẫu nhiên (Random Forest) và thuật toán tăng cường độ dốc cực đại XGBoost, có thể được ứng dụng để cải thiện khả năng dự báo và cung cấp cái nhìn sâu hơn về các yếu tố ảnh hưởng đến mức độ nghiêm trọng của tai nạn.

Dựa trên những kết quả đạt được, một số giải pháp đề xuất bao gồm: nâng cấp các tuyến cao tốc chưa phân kỳ đầy đủ, bổ sung dải phân cách và làn dừng khẩn cấp, kiểm soát tốc độ hiệu quả hơn, đồng thời nâng cao ý thức người tham gia giao thông thông qua các chương trình đào tạo và tuyên truyền phù hợp với điều kiện vận hành tại Việt Nam. Những biện pháp này không chỉ giúp giảm số vụ tai nạn mà còn góp phần hạn chế mức độ thiệt hại về người và tài sản. Để nâng cao an toàn giao thông một cách bền vững, cần một chiến lược tổng thể kết hợp giữa quy hoạch kết cấu hạ tầng hợp lý, chính sách kiểm soát hiệu quả và ứng dụng các mô hình phân tích hiện đại nhằm hỗ trợ ra quyết định và quản lý rủi ro trên các tuyến cao tốc.

LỜI CẢM ƠN

Nghiên cứu này được tài trợ bởi Bộ Giáo dục và Đào tạo (MoET) trong Đề tài mã số B2025-GHA-01.

TÀI LIỆU THAM KHẢO

- [1]. World Health Organization, Global status report on road safety 2023. <https://iris.who.int/bitstream/handle/10665/374868/9789240086456-eng.pdf>, 2023 (truy cập ngày 15/03/2025).
- [2]. Ủy ban An toàn Giao thông Quốc gia, Số liệu tai nạn giao thông năm 2023. <https://baolamdong.vn/xa-hoi/202401/nam-2023-toan-quoc-xay-ra-22067-vu-tai-nan-giao-thong-lam-chet-11628-nguoi-0331026/>, 2023 (truy cập ngày 15/03/2025).
- [3]. Báo Nhân Dân: Nỗ lực giảm thiểu các vụ tai nạn giao thông trên đường cao tốc, <https://nhandan.vn/no-luc-giam-thieu-cac-vu-tai-nan-giao-thong-tren-duong-cao-toc-post821957.html>, 2024 (truy cập ngày 31/05/2025).
- [4]. VnExpress, Quy hoạch mạng lưới cao tốc Việt Nam. <https://vnexpress.net/quy-hoach-mang-luoi-cao-toc-viet-nam-4855526.html>, 2025 (truy cập ngày 15/03/2025).
- [5]. P. Qin, J. He, C. Zhang, X. Yan, C. Wang, Y. Ye, Z. Fang, Crash risk prediction and analysis from the perspective of alignment and environment features: A study on an expressway in a hilly area, Traffic Injury Prevention, (2025) 1-11. <https://doi.org/10.1080/15389588.2025.2459297>.
- [6]. P. Kumar, J.K. Jain, G. Singh, Analysing crash severity on expressways in India: Statistical and machine learning models, Proceedings of the Institution of Civil Engineers-Transport, (2024) 1-14. <https://doi.org/10.1680/jtran.24.00071>.

- [7]. J.S. Madushani, R.K. Sandamal, D.P.P. Meddage, H.R. Pasindu, P.A. Gomes, Evaluating expressway traffic crash severity using logistic regression and explainable supervised machine learning classifiers, *Transportation Engineering*, 13 (2023) 100190. <https://doi.org/10.1016/j.treng.2023.100190>.
- [8]. Y. Yang, T. Zheng, Y. Zhang, Y. Jiang, Y. Hu, Prediction and analysis of expressway tunnel crashes based on modified convolutional neural network and Shapley additive explanations, *Proceedings of the Institution of Civil Engineers-Transport*, (2025). <https://doi.org/10.1680/jtran.24.00129>.
- [9]. L.S. Bisht, G. Tiwari, Identification of road traffic crash hotspots on an intercity expressway in India using geospatial techniques, *IATSS Research*, 47 (2023) 349-356. <https://doi.org/10.1016/j.iatssr.2023.07.003>.
- [10]. Q. Hou, A.P. Tarko, X. Meng, Investigating factors of crash frequency with random effects and random parameters models: New insights from a Chinese freeway study, *Accident Analysis & Prevention*, 120 (2018) 1–12. <https://doi.org/10.1016/j.aap.2018.07.010>.
- [11]. H. Zhang, S. Li, C. Wu, Q. Zhang, Y. Wang, Predicting crash frequency for urban expressways considering collision types using real-time traffic data, *Journal of Advanced Transportation*, 1 (2020) 8523818. <https://doi.org/10.1155/2020/8523818>.
- [12]. H. Naghawi, Negative binomial regression model for road crash severity prediction, *Modern Applied Science*, (2018) 38-46. <https://doi.org/10.5539/mas.v12n4p38>
- [13]. Y. Yan, Y. Zhang, X. Yang, J. Hu, J. Tang, Z. Guo, Crash prediction based on random effect negative binomial model considering data heterogeneity, *Physica A: Statistical Mechanics and Its Applications*, 547 (2020) 123858. <https://doi.org/10.1016/j.physa.2019.123858>.
- [14]. Z. Ma, H. Zhang, I. Steven, J. Chien, J. Wang, C. Dong, Predicting expressway crash frequency using a random effect negative binomial model: A case study in China, *Accident Analysis & Prevention*, 98 (2017) 214-222. <https://doi.org/10.1016/j.aap.2016.10.012>.
- [15]. C.C. Ihueze, U.O. Onwurah, Road traffic accidents prediction modeling: An analysis of Anambra State, Nigeria, *Accident Analysis & Prevention*, 112 (2018) 21-29. <https://doi.org/10.1016/j.aap.2017.12.016>.
- [16]. NCSS Statistical Software. Chapter 300: Linear Regression and Correlation. https://www.ncss.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Linear_Regression_and_Correlation.pdf (truy cập ngày 01/06/2025).
- [17]. NCSS Statistical Software. Chapter 325: Poisson Regression. https://www.ncss.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Poisson_Regression.pdf (truy cập ngày 01/06/2025).
- [18]. Chapter 326 Negative Binomial Regression. https://www.ncss.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Negative_Binomial_Regression.pdf. (truy cập ngày 01/06/2025).
- [19]. Thông tư 41/2024/TT-BGTVT: Quy định về quản lý, vận hành, khai thác và bảo trì kết cấu hạ tầng đường bộ, 2024.
- [20]. P. K. Ozili. The acceptable R-square in empirical modelling for social science research. In *Social research methodology and publishing results: A guide to non-native English speakers*. IGI global, (2023) 134-143.
- [21]. J. Louviere, D. Hensher, J. Swait. *Stated choice methods*. Cambridge University Press, 2000.