

Кількісна оцінка впливу погоди на надійність міського транспорту, використовуючи XGBoost та SHAP

Quantifying the Weather Impact on Urban Transport Reliability using XGBoost and SHAP

Юрій Мацелюх

Corresponding author: аспірант, e-mail: indeed.post@gmail.com, ORCID ID: 0000-0002-1721-7703

Василь Литвин

доктор технічних наук, професор, професор кафедри, e-mail: vasyly.v.lytvyn@lpnu.ua, ORCID ID: 0000-0002-9676-0180

Yurii Matseliukh

Corresponding author: PhD Student, e-mail: indeed.post@gmail.com, ORCID ID: 0000-0002-1721-7703

Vasyl Lytvyn

Dr.Sc., Professor, Professor Department, e-mail: vasyly.v.lytvyn@lpnu.ua, ORCID ID: 0000-0002-9676-0180

Національний університет "Львівська політехніка", м. Львів, Україна

Lviv Polytechnic National University, Lviv, Ukraine

Received: November 14, 2025 | Revised: December 06, 2025 | Accepted: December 31, 2025

UDC 004.85:656.4:551.5

DOI: <https://doi.org/10.33445/sds.2025.15.6.23>

Мета роботи. Розробка та перевірка методу кількісної оцінки впливу комплексу метеорологічних факторів на операційні збої громадського транспорту, розглядаючи їх як ризик для безпеки перевезень.

Метод дослідження. Дослідження базується на методах машинного навчання. Для прогнозування затримок використано високоточний ансамблевий алгоритм градієнтного бустингу XGBoost. Для глибокого аналізу факторів ризику та подолання проблеми «чорної скриньки» застосовано передовий інструмент інтерпретованого машинного навчання Shapley Additive Explanations, який здатний встановити внесок кожного фактору у конкретний прогноз.

Результати дослідження. Встановлено, що метеорологічні умови є значущим, нелінійним фактором ризику. Аналіз показав, що найбільшу загрозу для стабільності руху становлять температурні аномалії (різке зниження), значні опади та висока швидкість вітру. За допомогою запропонованого методу доведено, що вплив погоди є гетерогенним, тобто незначним у звичайних умовах та сильним тригером масштабних збоїв у години пік.

Практична цінність дослідження. Результати дослідження створюють наукову основу для розробки інформаційних систем підтримки прийняття рішень у сфері кризового управління. Запропонований метод дає змогу перейти від реагування на зміни, що вже відбулися, до здійснення попереджувальних заходів: завчасного інформування пасажирів, оптимізації ресурсів та підвищення стійкості транспортної системи до кліматичних загроз.

Тип статті. Емпірична.

Purpose. The purpose of the work is to develop and verify a method for quantitatively assessing the impact of a complex of meteorological factors on operational failures of public transport.

Method. The study is based on machine learning methods. The high-precision ensemble gradient boosting algorithm XGBoost was used to predict delays. For an in-depth analysis of risk factors and overcoming the «black box» problem, an advanced interpreted machine learning tool SHAP was used, which can establish the contribution of each factor to a specific forecast.

Findings. It was established that meteorological conditions are a significant, nonlinear risk factor. The analysis showed that the greatest threat to traffic stability is posed by temperature anomalies (a sharp decrease), significant precipitation and high wind speed. Using the proposed method, it was proven that the impact of weather is heterogeneous, i.e. insignificant under normal conditions and a strong trigger of large-scale failures during rush hours.

Practical implications. The results of the study create a scientific basis for the development of information systems to support decision-making in the field of crisis management. The proposed method makes it possible to move from responding to changes that have already occurred to implementing preventive measures: informing passengers in advance, optimizing resources and increasing the resilience of the transport system to climate threats.

Paper type. Empirical.

Ключові слова: надійність громадського транспорту; машинне навчання; інтерпретований штучний інтелект; управління ризиками; метеорологічні фактори; градієнтний бустинг; критична інфраструктура.

Key words: public transport reliability; machine learning; explainable AI; risk management; meteorological factors; gradient boosting; critical infrastructure.

Вступ

Стабільне функціонування системи громадського транспорту є однією з ключових складових концепції розумного міста та важливим фактором економічної активності (Banach and Długosz, 2025). Операційні затримки призводять до прямих економічних збитків (Комарницький та Бублик, 2008), зниження продуктивності праці та погіршення якості життя населення в містах (Guo et al., 2025). Одним із найбільш стохастичних, однак вагомих факторів, є погода. Проблема врахування впливу погодних умов на здійснення пасажирських перевезень у

громадському транспорту з кожним роком набуває щоразу більшої актуальності, особливо за умов зростання аномальних погодних явищ, таких як екстремальні зливи чи снігопади (Ma et al., 2025; Gu et al., 2025).

Незважаючи на загальне розуміння існування зв'язку погоди із затримками громадського транспорту, питання точної кількісної оцінки впливу комплексу метеорологічних факторів на відхилення від графіка руху залишається недостатньо вивченим. Сучасні підходи вимагають не лише констатації факту впливу, а й глибокого розуміння його нелінійного та гетерогенного характеру, що потребує залучення передових інструментів штучного інтелекту, серед яких машинне навчання, дерева рішень тощо (Arunachalam, 2025; Saki and Soori, 2025). Спираючись на роботи закордонних та вітчизняних вчених, які досліджували причини затримок громадського транспорту (Ройко та ін., 2018; Hanbarikarekani et al., 2018), потребує свого розвитку питання встановлення важливості впливу фактору погоди.

Сучасний штучний інтелект є комплексною галуззю науки, що займається створенням систем, здатних до розумної поведінки. В ієрархічній структурі штучного інтелекту прийнято виділяти як ключовий підрозділ машинне навчання, до складу якого входить глибоке навчання (Пасічник та ін., 2021; Maksoud et al., 2025). Машинне навчання поєднує в собі набір методів, що дозволяють комп'ютеру навчатися на основі даних, а не жорсткого програмування, тоді як глибоке навчання базується на багатопшарових нейронних мережах (Saki and Soori, 2025). Саме машинне навчання є ідеальним інструментом для вирішення завдань у сфері пасажирських перевезень. Транспортна система генерує величезні обсяги даних, залежності між якими є складними та нелінійними (Alonso-Fernandez et al., 2025). Методи машинного навчання дають змогу виявляти приховані патерни, будувати точні прогнозні моделі та оптимізувати операційні процеси, що є недосяжним для класичних статистичних підходів через складність та динамічність сучасних транспортних систем (Saki and Soori, 2025).

Метою роботи є розробка та перевірка методу кількісної оцінки впливу комплексу метеорологічних факторів на формування затримок громадського транспорту, що є джерелом ризику для безпеки перевезень у розумному місті. З огляду на підсилення можливостей моделі градієнтного бустингу інструментом інтерпретованого машинного навчання було сформувано наступні завдання:

1. Розробити та навчити модель градієнтного бустингу XGBoost для прогнозування затримок громадського транспорту на основі даних про виконання графіків зупинок та погодних даних.
2. Провести аналіз факторів впливу за допомогою методу інтерпретованого машинного навчання (SHAP) для ідентифікації та ранжування причин затримок пасажирських перевезень.
3. Кількісно оцінити та порівняти вплив операційних (часових, контекстуальних) та нелінійного впливу погодних факторів на надійність перевезень.

Наукова новизна полягає у розвитку методу оцінки впливу гетерогенних факторів на формування затримок громадського транспорту за допомогою поєднання методів градієнтного бустингу та інтерпретованого машинного навчання.

Огляд літератури

Дослідження впливу зовнішніх факторів на транспортні системи є актуальним напрямом сучасної науки. Особлива увага приділяється стійкості міської інфраструктури до екстремальних погодних умов. Роботи (Gu et al., 2025; Ma et al., 2025) аналізують стійкість автобусних мереж до снігопадів та злив відповідно, застосовуючи теорію катастроф та системну динаміку для моделювання раптових збоїв. Дослідження Guo et al. (2025) пропонує керовану даними модель для оцінки та еволюційного аналізу стійкості транспортних агломерацій, тоді як (Kaunak et al., 2025) розробляють концепцію цифрового двійника міста

для аналізу наслідків повеней. Ці праці підкреслюють важливість кількісної оцінки впливу погоди, що є центральним завданням нашого дослідження.

Застосування штучного інтелекту та машинного навчання для аналізу транспортних систем є домінуючим трендом. Загальні огляди, представлені в роботах Saki and Soori (2025) та Maksoud et al. (2025), демонструють широкий спектр застосування машинного навчання від управління трафіком до керування автономними транспортними засобами. Автори (Zhang et al., 2025a) проводять глибокий аналіз застосування обробки природної мови та текстового майнінгу в транспортній галузі. Комплекс методологій, серед яких графові нейронні мережі для прогнозування попиту (Gong et al., 2025; Zhang et al., 2025b), порівняльний аналіз різних методів машинного навчання для виявлення маневрів вантажівок (Alonso-Fernandez et al., 2025), показують ефективність спеціалізованих моделей. У роботі (Masmoudi et al., 2025) використовуються методи машинного навчання для класифікації ризиків водіїв на основі психологічних та фізіологічних даних з симулятора.

Ключовими інструментами нашого дослідження є градієнтний бустинг XGBoost та інтерпретоване машинне навчання SHAP. Їх поєднання дає змогу не лише будувати високоточні прогнози, а й розуміти причини цих прогнозів. Дослідження Parsa et al. (2020) є яскравим прикладом застосування XGBoost та SHAP для виявлення ДТП у реальному часі на основі даних про трафік та погоду. Аналогічно, Zhang and Zou (2022) використовують цю комбінацію для прогнозування та аналізу причин морських аварій. У роботі Zhou et al. (2025) XGBoost (у складі Random Forest) використовується для оцінки залишкового пробігу електромобілів з урахуванням погоди та стилю водіння. Практичні аспекти та переваги SHAP для інтерпретації моделей XGBoost також детально описані в документації (Explain XGBoost, 2025). Ці роботи підтверджують, що обраний нами інструментарій є передовим для вирішення завдань причинно-наслідкового аналізу в складних транспортних системах.

Крім того, дослідження розширюють контекст безпеки та стійкості. У праці (Zhao et al., 2025) безпосередньо використовуються XGBoost та SHAP для аналізу стійкості транспортної поведінки під час екстремальних злив. Дослідження (Ha and Arimura, 2025) аналізує міську мобільність під час екстремальних погодних умов за допомогою просторової статистики, демонструючи, як несприятливі умови змінюють структуру транспортних потоків. Робота (Liu et al., 2025) присвячена оптимізації маршрутів перевезення небезпечних вантажів з урахуванням динамічної швидкості вітру, що є прикладом прямого впливу погодного фактора на логістичні рішення. Вчені у роботі (Theivadas та Ponnan, 2026) застосовують рекурентні нейронні мережі для прогнозування поведінки водія, що є важливим аспектом загальної безпеки на транспорті.

Управління транспортними потоками та мінімізація затримок є класичною інженерною задачею. Ю. Ройко та ін. (2018) визначають затримки на двосмугових вулицях біля зупинок громадського транспорту, Ю. Мацелюх та В. Литвин (2025) використовують метод кластеризації K-середніх для групування причин затримок пасажирських перевезень у розумному місті, а класифікацію ризиків на транспорті як основу економічної безпеки розглядає Рачинська (2016). Робота (Chukhray et al., 2019), однак, наголошує, що при впровадженні будь-яких нових технологій, перш за все, слід враховувати споживчі аспекти, у нашому випадку, потреби пасажирів, оскільки важливим є не лише технічна ефективність, а й соціально-економічна прийнятність запропонованих рішень.

Сучасні дослідження все частіше розглядають взаємодію різних погодних факторів та їхній комбінований вплив. Автори дослідження (Hanbarikarekani et al., 2018) пропонують мінімізувати середню затримку на перехрестях за допомогою керування швидкістю. У роботі (Arunachalam, 2025) розроблено оптимізовану нейронну мережу для прогнозування трафіку з урахуванням погоди. У дослідженні (Wang et al., 2025) автори аналізують, як поєднання погоди впливає на вибір між велосипедом та метро, що є важливим для розуміння мультимодальної

поведінки. У той же час, розробка складних просторово-часових моделей, як у дослідженні (Gao et al., 2025), дозволяє прогнозувати ризики ДТП з урахуванням географічних особливостей та динаміки трафіку. Це підтверджує думку, що дослідження впливу комплексу метеорологічних факторів на затримки громадського транспорту та глибокий аналіз причин їх виникнення є досі невирішеним питанням в управлінні транспортними системами, яке призводить до економічних збитків, спричинених затримками у роботі критичної інфраструктури (Kuzmin and Bublyk, 2016).

Матеріали та методи

Для досягнення мети дослідження було реалізовано послідовний, керований даними підхід, де методологія дослідження включає сім основних етапів, від збору даних до глибокого аналізу результатів моделювання, що схематично представлено на рис. 1.

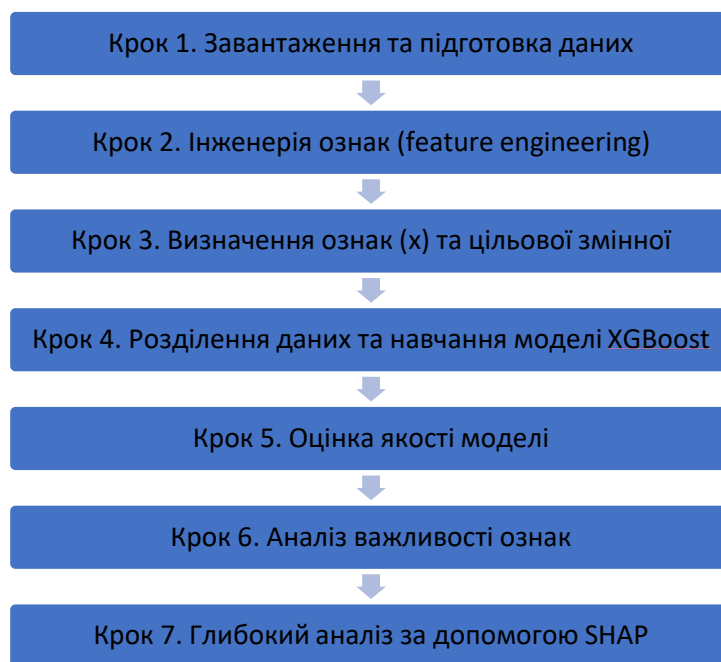


Рисунок 1 – Етапи розробки методу кількісної оцінки впливу метеорологічних факторів на затримки громадського транспорту

Примітка: розроблено з використанням стандартних методологій машинного навчання для аналізу затримок громадського транспорту із специфічною для даного дослідження глибокою інтерпретацією за допомогою SHAP

Джерело: розроблено автором

Ця структура дозволяє не тільки побудувати точну прогнозну модель, а й, що найважливіше, інтерпретувати її для отримання науково обґрунтованих висновків щодо причин затримок. Під час завантаження та підготовки даних на першому етапі дослідження було проведено збір та інтеграцію різнорідних даних з відкритих джерел, а саме було використано відкриті набори даних у форматі GTFS про рух громадського транспорту, що містили розклади руху, маршрути та фактичний час прибуття транспорту на зупинки, а також метеорологічні дані, отримані через API сервісу OpenWeather, що містили історичні погодні дані з погодинною деталізацією для конкретного географічного центру. Ці два набори даних були об'єднані за часом та географічним центром (географічними координатами) для створення єдиного датасету. Структура фінального набору даних, що використовувався для моделювання, представлена в табл. 1.

Таблиця 1 – Структура та опис вихідних даних для моделювання

Назва поля	Опис	Тип даних
trip_id	Унікальний ідентифікатор рейсу	Категоріальний
dt	Часова мітка у форматі Unix	Числовий
temp	Температура повітря, °C	Числовий
dew_point	Точка роси, °C	Числовий
feels_like	Температура, що відчувається, °C	Числовий
pressure	Атмосферний тиск, гПа	Числовий
humidity	Вологість, %	Числовий
wind_speed	Швидкість вітру, м/с	Числовий
wind_deg	Напрямок вітру, градуси	Числовий
rain_1h	Кількість опадів за останню годину, мм	Числовий
clouds_all	Хмарність, %	Числовий
weather_main	Загальний опис погоди	Категоріальний
weather_description	Детальний опис погоди	Категоріальний

Примітка: наведено основні поля; повний набір даних включав ідентифікатори зупинок та інші службові поля

Джерело: сформовано автором на основі відкритих даних GTFS та метеорологічного сервісу OpenWeather API

У результаті цього для дослідження було сформовано реляційну базу даних SQLite, що містить комплексну інформацію про рух громадського транспорту, приклад якої наведено на рис.2. Для забезпечення сфокусованого аналізу, з бази даних було вибірково завантажено записи, що стосуються виключно одного трамвайного маршруту Т03. Це дозволило уникнути шуму від змішування характеристик різних маршрутів.

	id	trip_id	dt	temp	dew_point	feels_like	temp_min	temp_max	pressure	humidity	wind_speed	wind_deg	rain_1h	clouds_all	weather_main	weather_description
1	174010	10639	1558360800	20.23	17.81	20.55	19.58	20.97	1007	86	3.66	163	8.29	100	Rain	heavy intensity rain
2	174011	10639	1558360800	20.23	17.81	20.55	19.58	20.97	1007	86	3.66	163	8.29	100	Rain	heavy intensity rain
3	174012	10639	1558360800	20.23	17.81	20.55	19.58	20.97	1007	86	3.66	163	8.29	100	Rain	heavy intensity rain
4	174013	10639	1558360800	20.23	17.81	20.55	19.58	20.97	1007	86	3.66	163	8.29	100	Rain	heavy intensity rain
5	174014	10639	1558360800	20.23	17.81	20.55	19.58	20.97	1007	86	3.66	163	8.29	100	Rain	heavy intensity rain
6	174015	10640	1558360800	20.23	17.81	20.55	19.58	20.97	1007	86	3.66	163	8.29	100	Rain	heavy intensity rain
7	174016	10640	1558360800	20.23	17.81	20.55	19.58	20.97	1007	86	3.66	163	8.29	100	Rain	heavy intensity rain
8	174017	10640	1558360800	20.23	17.81	20.55	19.58	20.97	1007	86	3.66	163	8.29	100	Rain	heavy intensity rain
9	174018	10640	1558360800	20.23	17.81	20.55	19.58	20.97	1007	86	3.66	163	8.29	100	Rain	heavy intensity rain
10	174019	10640	1558360800	20.23	17.81	20.55	19.58	20.97	1007	86	3.66	163	8.29	100	Rain	heavy intensity rain

Рисунок 2 – Етапи розробки методу кількісної оцінки впливу метеорологічних факторів на затримки громадського транспорту

Джерело: розроблено автором

Первинна підготовка даних включала об'єднання таблиць StopTimes, Trips та Routes для отримання повної інформації по кожній зупинці в рамках цільового маршруту. Також було перетворено текстові поля з часом (planned_arrival, actual_arrival) у формат datetime. Цільовою змінною у було визначено затримку прибуття транспорту на зупинку в секундах, що обчислювалася як різниця в секундах між фактичним та запланованим часом прибуття. При обробці пропущених значень для погодних показників, що фіксують опади (rain_1h, snow_1h), пропущені значення замінювалися на 0, припускаючи, що відсутність даних означає 0 опадів. На другому та третьому етапах було використано інженерію ознак для підготовки даних до моделювання. На основі підготовлених даних було сформовано набір предикторів (матрицю ознак X) та цільову змінну (y). Предиктори (ознаки x) були розділені на три групи: характеристики поїздки (stop_id – ідентифікатор зупинки; trip_id – ідентифікатор рейсу),

циклічні часові характеристики (hour_sin, hour_cos – синус та косинус години доби; dayofweek_sin, dayofweek_cos – синус та косинус дня тижня; month_sin, month_cos – синус та косинус місяця року), а також погодні характеристики, наведені в табл. 1 (temp, visibility, dew_point, feels_like, pressure, humidity, wind_speed, wind_deg, wind_gust, rain_1h, snow_1h, clouds_all, weather_id), де найбільш релевантними метеорологічними показниками були температура, вологість, швидкість вітру, опади. Цільовою змінною у була затримка в секундах (delay_seconds).

Показники, як година доби, день тижня та місяць року, що мають циклічну природу перетворювалися у неперервний двовимірний простір за допомогою тригонометричних функцій (1) та (2). Наприклад:

$$feature_sin = \sin\left(\frac{2\pi \cdot value}{max_value}\right), \quad (1)$$

$$feature_cos = \cos\left(\frac{2\pi \cdot value}{max_value}\right), \quad (2)$$

де $value$ – поточне значення ознаки;
 max_value – максимальне значення циклу;
 π – відношення довжини кола до його діаметра.

Таке кодування дозволяє моделі зрозуміти близькість кінцевих та початкових значень циклу.

На четвертому етапі для вирішення завдання прогнозування затримок було обрано алгоритм XGBoost (eXtreme Gradient Boosting), реалізований у вигляді XGBRegressor. Цей метод був обраний через його високу точність, швидкість та здатність моделювати складні нелінійні залежності. Процес моделювання моделі включав розділення даних і навчання. Для забезпечення надійного та неупередженого оцінювання моделі весь набір даних було розділено на три частини. Для об'єктивної оцінки якості моделі на даних, які вона не бачила під час навчання, спочатку було відділено тестову вибірку (20%), яка використовувалася лише один раз для фінальної оцінки якості повністю навченої моделі. Решту даних (80%) було поділено ще раз на тренувальну (64%) вибірку, яка використовувалася для безпосереднього навчання моделі, та валідаційну (16%) вибірку, яка використовувалася під час навчання для налаштування гіперпараметрів, зокрема для механізму ранньої зупинки (early stopping). Цей механізм дозволяє автоматично припинити навчання, якщо середня абсолютна помилка (MAE) на тестовій вибірці не покращується протягом 50 ітерацій, що є ефективним методом боротьби з перенавчанням.

На п'ятому етапі проводилася фінальна, неупереджена оцінка якості навченої моделі на тестовій вибірці, яку модель жодним чином не використовувала під час навчання чи налаштування. Для оцінки використовувалися стандартні метрики для задач регресії, а саме: середня абсолютна помилка (MAE) та коефіцієнт детермінації (R^2), що розраховувалися за формулами (3) та (4):

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad (3)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (4)$$

де n – кількість спостережень у тестовій вибірці;
 y_i – фактичне значення затримки;
 \hat{y}_i – прогнозоване значення затримки;
 \bar{y} – середнє значення фактичних затримок.

На шостому етапі аналізу важливості ознак спочатку проводився первинний аналіз факторів за допомогою вбудованого в XGBoost механізму `feature_importances_`. Цей атрибут розраховує важливість кожної ознаки, як правило, на основі того, наскільки часто та ефективно вона використовувалася для поділу в деревах рішень ансамблю. Це дозволило отримати базове ранжування факторів за їхнім внеском у модель.

На сьомому етапі глибокого аналізу за допомогою SHAP для глибокого та всебічного аналізу факторів впливу, що є ключовим завданням нашої роботи, було застосовано метод SHAP (SHapley Additive exPlanations). Цей підхід, на відміну від стандартної важливості ознак, дав змогу не лише визначити силу впливу фактора, а й зрозуміти його напрямок (збільшує чи зменшує затримку) та характер взаємодії з іншими факторами.

Аналіз проводився на репрезентативній вибірці з тестових даних розміром 50 000 записів для оптимізації обчислювального часу. Для пояснення моделі XGBoost використовувався спеціалізований `shap.TreeExplainer`. Пояснювальна адитивна модель SHAP має вигляд (5):

$$g(z') = \phi_0 + \sum_{j=1}^M \phi_j z'_j, \quad (5)$$

де	g	– пояснювальна модель;
	$z' \in \{0,1\}^M$	– вектор наявності ознак;
	M	– кількість ознак;
	ϕ_j	– SHAP-значення, що є внеском j -ої ознаки у прогноз;
	ϕ_0	– базове значення (середній прогноз моделі).

За допомогою SHAP були побудовані та проаналізовані діаграми залежностей безпосередньо для аналізу нелінійних ефектів окремих ознак та їх взаємодій, а саме, було досліджено, як вплив години доби на затримку змінюється залежно від інтенсивності дощу, чи як вплив дощу залежить від температури. Це дозволило кількісно оцінити та порівняти вплив операційних та погодних факторів, виявляючи складні приховані залежності.

Цей комплексний підхід дозволив не тільки створити точну прогнозну модель, але й глибоко проаналізувати та кількісно оцінити вплив кожного операційного та погодного фактора.

Результати та обговорення

У результаті моделювання затримок громадського транспорту при пасажирських перевезеннях навчена модель градієнтного бустингу XGBoost продемонструвала високу прогностичну здатність. При оцінці на тестовій вибірці, середня абсолютна помилка (MAE) склала 45,7 секунд, що вказує на високу точність прогнозування затримок в реальних умовах експлуатації. Первинний аналіз важливості факторів, проведений за допомогою вбудованого в XGBoost механізму `feature_importance`, дозволив проранжувати предиктори за їхнім сукупним внеском у точність моделі (Таблиця 2).

Аналіз даних таблиці 2 дав змогу зробити кілька ключових спостережень. По-перше, домінуючий вплив мають циклічні часові характеристики `dayofweek_cos`, `month_sin` та `month_cos`, які займають перші три місця з найбільшими значеннями важливості. Це свідчить про те, що модель в першу чергу орієнтується на передбачувані тижневі та сезонні патерни руху. По-друге, значний внесок роблять і погодні фактори, такі як швидкість вітру (`wind_speed`), температура, що відчувається (`feels_like`), тиск (`pressure`) та вологість (`humidity`). Це підтверджує вихідну гіпотезу дослідження. Водночас, викликає питання відносно низька важливість таких очевидних факторів, як дощ (`rain_1h`) та сніг (`snow_1h`). Цей парадокс вказує на те, що стандартний аналіз важливості не здатний розкрити складніші, нелінійні взаємозв'язки, для аналізу яких потрібні більш просунуті методи.

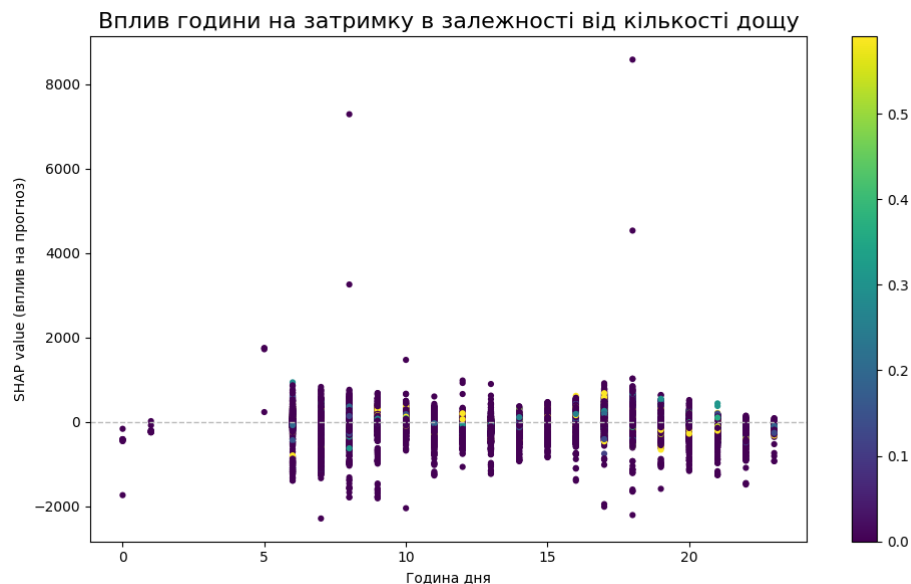
Таблиця 2 – Важливість ознак за результатами моделі XGBoost

Ознака	Важливість	Ознака	Важливість
dayofweek_cos	0.110005	temp	0.041122
month_sin	0.097793	visibility	0.040169
month_cos	0.094744	hour_cos	0.040126
wind_speed	0.057964	clouds_all	0.037653
dayofweek_sin	0.057848	trip_id	0.037299
feels_like	0.055244	dew_point	0.035978
pressure	0.052448	stop_id	0.03232
humidity	0.045525	wind_deg	0.031804
hour_sin	0.04474	weather_id	0.030555
rain_1h	0.020415	snow_1h	0.018976

Примітка: важливість є відносною величиною, що показує, як часто та ефективно ознака використовувалася для побудови дерев рішень в ансамблі

Джерело: розраховано автором

Для детального вивчення впливу кожного фактора та їхніх взаємодій було застосовано метод SHAP. На відміну від стандартної важливості, SHAP-значення показують не лише силу, а й напрямок впливу кожної ознаки на кожен окремий прогноз. Це дозволяє виявити приховані нелінійні залежності. На рисунку 3 представлено вплив години дня на затримку, де колір точок відображає інтенсивність дощу. Спостерігається чіткий добовий патерн: у ранкові години (6-9 ранку) та особливо у вечірні години пік (16-19 вечора) SHAP-значення є переважно позитивними, що вказує на збільшення затримок. Наявність навіть невеликої кількості дощу (зелені та жовті точки) суттєво зміщує SHAP-значення вгору, особливо у пікові години. Це означає, що дощ діє як мультиплікатор, значно посилюючи затримки саме тоді, коли система вже є навантаженою.

**Рисунок 3** – Вплив години на затримку в залежності від кількості дощу

Джерело: побудовано автором

На рисунку 4 показано аналогічну залежність, але з урахуванням вологості. Патерн є схожим: висока вологість (жовті точки) тісно корелює зі збільшенням затримок у пікові години. Оскільки висока вологість часто передуює або супроводжує опади, цей графік підтверджує

попередні висновки. Вплив вологості є найбільш вираженим у вечірні години, що свідчить про кумулятивний ефект втоми водіїв та погіршення видимості.

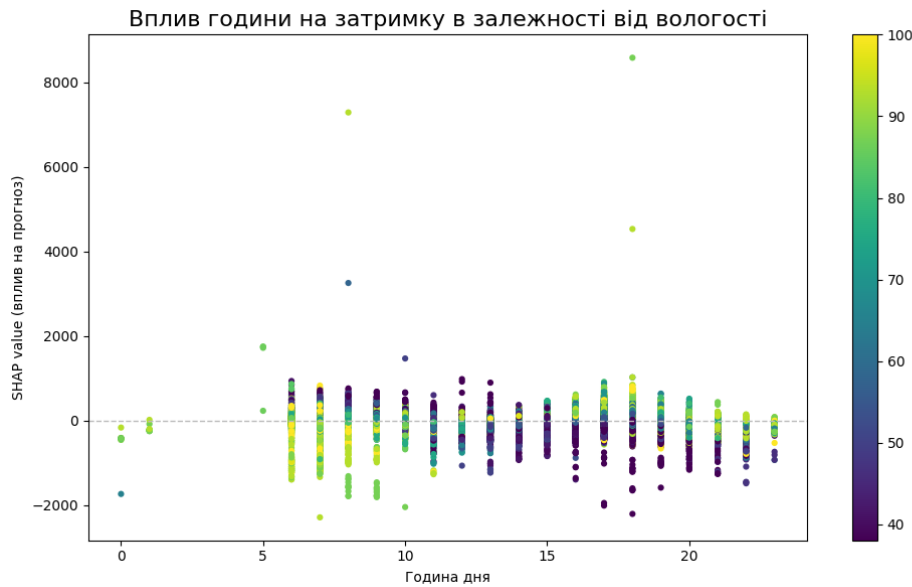


Рисунок 4 – Вплив години на затримку в залежності від вологості

Джерело: побудовано автором

На рисунку 5 аналізується взаємодія години дня та швидкості вітру. Хоча цей зв'язок менш виражений, ніж з опадами чи вологістю, простежується тенденція, де вища швидкість вітру (жовті точки) асоціюється з дещо більшими SHAP-значеннями, особливо в діапазоні від 7 до 18 години. Це може бути пов'язано як з обережнішим стилем водіння учасників руху, так і з можливим впливом сильного вітру на контактну мережу трамвая, що призводить до невеликих уповільнень.

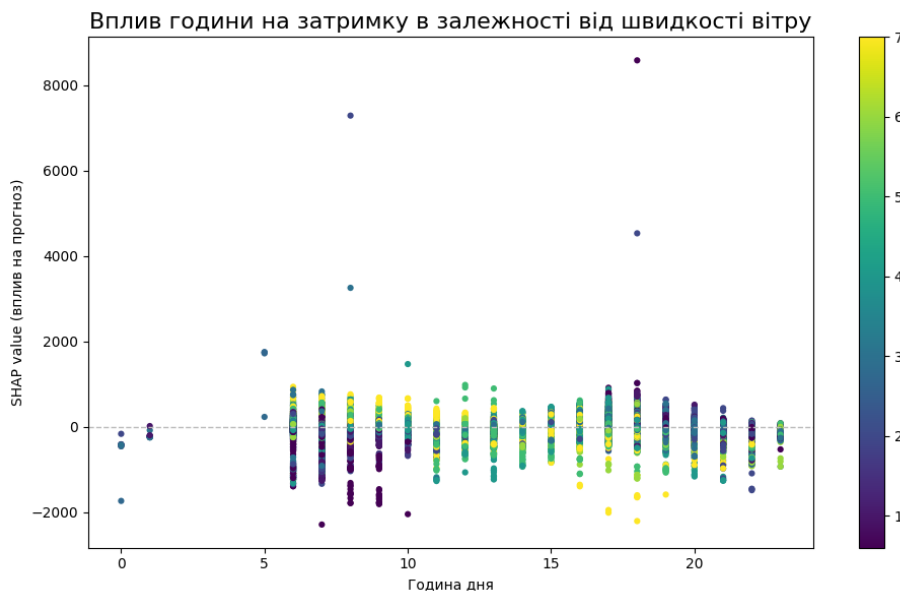


Рисунок 5 – Вплив години на затримку в залежності від швидкості вітру

Джерело: побудовано автором

Проведене дослідження кількісно підтвердило, що метеорологічні фактори мають значний, але складний та нелінійний вплив на затримки громадського транспорту. Ключовим результатом є виявлення того, що погода діє не як незалежний фактор, а як сильний модифікатор існуючих транспортних патернів, особливо годин пік. Цим пояснюється парадокс

низької формальної важливості ознак `rain_1h` та `snow_1h` у моделі XGBoost, оскільки дощ іде не постійно, його середній внесок протягом усього періоду є невисоким. Проте, як показав аналіз SHAP, його миттєвий вплив під час навантажених періодів є критичним.

Ці результати узгоджуються з висновками дослідників (Zhao et al., 2025), які також, використовуючи XGBoost та SHAP, вивчали стійкість транспортної поведінки під час екстремальних злив і виявили просторову та часову гетерогенність впливу. Наше дослідження доповнює їхню роботу, деталізуючи механізми взаємодії не лише дощу, а й вітру та вологості з добовими циклами трафіку. Крім того, отримані дані підтверджують загальні висновки про вразливість транспортних систем до погодних умов, зроблені в роботах вчених (Ma et al., 2025; Gu et al., 2025), надаючи конкретний кількісний інструментарій для оцінки цих ризиків.

З методологічної точки зору, застосування комбінації XGBoost та SHAP виявилось надзвичайно ефективним у розробленому методі кількісної оцінки впливу метеорологічних факторів на затримки громадського транспорту. Успішним виявився підхід, використаний у роботі (Parsa et al., 2020) для виявлення ДТП та дослідниками (Zhang & Zou, 2022) для аналізу морських інцидентів, який довів також свою придатність і для аналізу операційних затримок у нашому випадку. Це демонструє універсальність методу для широкого кола завдань у сфері транспортної безпеки та надійності.

Практична цінність отриманих результатів полягає у можливості створення дієвих систем управління рухом. Замість того, щоб реагувати на затори, що вже утворилися, транспортні служби можуть використовувати прогноз погоди як вхідні дані для прогнозної моделі, де прогноз невеликого дощу о 17:00 повинен генерувати значно вищий рівень тривоги та спонукати до превентивних заходів як виведення додаткових одиниць транспорту у час пік, ніж прогноз сильної зливи після 21:00. Це відповідає концепціям інтелектуальних транспортних систем, описаним у роботі (Saki & Soori, 2025), і може стати основою для реалізації цифрових двійників пасажирських перевезень громадським транспортом (Мацелюх та Литвин, 2025) та розумних міст (Каунак et al., 2025).

Слід зазначити, що дане дослідження було обмежене аналізом одного трамвайного маршруту. Для більшого узагальнення висновків у майбутньому слід розширити аналіз на всю транспортну мережу міста. Крім того, до моделі доцільно також включити додаткові фактори як проведення масових заходів, дорожніх робіт та інформацію про ДТП. Подальший розвиток методу може полягати в його інтеграції в систему прийняття рішень у реальному часі для диспетчерських служб.

Висновки

Проведене дослідження доводить, що поєднання методів градієнтного бустингу та інтерпретованого машинного навчання є ефективним інструментом для сценарного аналізу ризиків та підвищення операційної безпеки критичної інфраструктури. Встановлено, що метеорологічні фактори є значущою, але складною нелінійною загрозою. Інструмент SHAP дозволяють декомпонувати цей ризик, кількісно оцінити вплив окремих параметрів та їх комбінацій, що є неможливим при використанні класичних методів.

У ході дослідження було успішно розроблено та апробовано метод кількісної оцінки впливу метеорологічних факторів на формування затримок у роботі громадського транспорту, що повністю відповідає поставленій меті. Застосування ансамблевого методу градієнтного бустингу дозволило створити високоточну прогнозу модель, яка здатна передбачати затримки з середньою абсолютною помилкою близько 45.7 секунд, що є високим показником для динамічної транспортної системи.

Ключовим результатом дослідження стало не лише підтвердження факту впливу погоди, а й глибоке розкриття його нелінійного та гетерогенного характеру. На відміну від стандартних методів аналізу важливості ознак, які недооцінювали роль опадів та снігу,

застосування інтерпретованого машинного навчання дозволило встановити, що метеорологічні фактори діють як мультиплікатори ризику. Встановлено, що інтенсивність дощу та висока вологість мають найбільш виражений синергетичний ефект з ранковими та вечірніми годинами пік, значно посилюючи затримки саме тоді, коли транспортна система є найбільш навантаженою. Вплив швидкості вітру виявився менш значущим, проте також стабільно призводив до збільшення тривалості поїздки.

Наукова новизна роботи полягає у розвитку методу оцінки впливу гетерогенних факторів, який, завдяки поєднанню прогностичної потужності XGBoost та інтерпретаційних можливостей SHAP, дозволяє перейти від простої констатації кореляцій до глибокого розуміння причинно-наслідкових зв'язків у системі «погода-трафік». Цей підхід дає змогу кількісно оцінити, як саме комбінація операційних (час доби, день тижня) та погодних умов формує ризик операційного збою.

Практична цінність отриманих результатів полягає у створенні наукового підґрунтя для розробки дієвих систем управління транспортними потоками. Замість того, щоб реагувати на затори, що вже виникли, транспортні служби можуть використовувати метеорологічні прогнози як вхідні дані для моделі, щоб завчасно оцінювати ризики та вживати превентивних заходів. Це включає тимчасову зміну розкладу руху, виведення на лінію додаткових транспортних засобів перед очікуваними піковими навантаженнями під час негоди та своєчасне інформування пасажирів, що в кінцевому підсумку сприятиме підвищенню надійності та безпеки перевезень у концепції розумного міста.

Перспективи подальших досліджень полягають у масштабуванні запропонованого підходу на всю транспортну мережу міста, а не лише на один маршрут, для перевірки отриманих висновків. Доцільним є включення до моделі додаткових динамічних факторів, таких як дані про дорожньо-транспортні пригоди, проведення дорожніх робіт та масові заходи, що дозволить створити ще більш комплексну та точну модель. Кінцевою метою може стати інтеграція розробленого методу в автоматизовану систему підтримки прийняття рішень для диспетчерських центрів або його використання як одного з модулів у концепції цифрового двійника міста.

Фінансування

Це дослідження не отримало конкретної фінансової підтримки.

Конкуруючі інтереси

Автори заявляють, що у них немає конкуруючих інтересів.

Список використаних джерел

- [1] Комарницький, І. М., та Бублик, М. І. (2008). Оцінка техногенних збитків та аналіз підходів до їхнього розрахунку у глобальному та регіональному аспектах. *Вісник НУ «Львівська політехніка»*. «Проблеми економіки та управління», 628, 134-144.
- [2] Мацелюх, Ю. Р., та Литвин, В. В. (2025). Метод К-середніх для аналізу даних з організації пасажирських перевезень у розумному місті. *Сучасний стан наукових досліджень та технологій в промисловості*, 1(31), 83–101. <https://doi.org/10.30837/2522-9818.2025.1.083>
- [3] Мацелюх, Ю. Р., та Литвин, В. В. (2025). Концепція цифрового двійника пасажирських перевезень громадським транспортом у розумному місті. *Таврійський науковий вісник. Серія: Технічні науки*, 1(4), 425-436. <https://doi.org/10.32782/tnv-tech.2025.4.1.40>

- [4] Пасічник, А. М., Пасічник, В. А., та Кущенко, Є. С. (2021). Застосування системного підходу в методології розв'язання інженерних задач. *Транспортні системи та технології перевезень*, 21, 23-32. <https://doi.org/10.15802/tstt2021/237635>
- [5] Рачинська, А. В. (2016). Класифікація ризиків на залізничному транспорті як основа формування системи економічної безпеки його функціонування. *Економіка та суспільство*, 6, 81-87.
- [6] Ройко, Ю., Бура, Р., та Максимюк, С. (2018). Визначення затримок на двосмугових вулицях у межах зупинок громадського транспорту. *Сучасні технології в машинобудуванні та транспорті*, 2(11), 24-28.
- [7] Alonso-Fernandez, F., Butt, T. H., and Tiwari, P. (2025). Overtake detection in trucks using CAN bus signals: A comparative study of machine learning methods. *Expert Systems with Applications*, 297, 129339. <https://doi.org/10.1016/j.eswa.2025.129339>
- [8] Arunachalam, S. (2025). Long-term and short-term traffic flow prediction with different weather conditions using optimized wheel-graph attention based neural network. *Future Generation Computer Systems*, 176, 108198. <https://doi.org/10.1016/j.future.2025.108198>
- [9] Banach, M., and Długosz, R. (2025). Solutions for planning smart hybrid public transportation system based on Google Maps and Voronoi diagrams. *Journal of Computational and Applied Mathematics*, 472, 116775. <https://doi.org/10.1016/j.cam.2025.116775>
- [10] Chukhray, N., Shakhovska, N., Mrykhina, O., et al (2019). Consumer aspects in assessing the suitability of technologies for the transfer. In *2019 IEEE 14th International Conference on Computer Sciences and Information Technologies (CSIT)* (pp. 142–147). <https://doi.org/10.1109/stc-csit.2019.8929879>
- [11] Explain XGBoost Predictions with SHAP. (2025). *XGBoosting.com*. Retrieved from <https://xgboosting.com/explain-xgboost-predictions-with-shap/>
- [12] Gao, P., Shuai, B., Zhang, R., and Wang, B. (2025). STCAGNN-RNKDE: a traffic accident prediction model for spatiotemporal combinatorial attention graph neural networks using Ripley's K and network kernel density estimation. *Reliability Engineering and System Safety*, 265, 111593. <https://doi.org/10.1016/j.ress.2025.111593>
- [13] Gong, Y., Yu, P., Zhang, X., Zhang, X., Nie, X., and Sun, H. (2025). Exploiting dynamic spatio-temporal correlations for origin-destination demand prediction. *Expert Systems With Applications*, 299, 130095. <https://doi.org/10.1016/j.eswa.2025.130095>
- [14] Gu, S., Miao, A., Yan, D., Hu, B., Cheng, G., and Xu, Z. (2025). Resilience Assessment of Urban Conventional Bus Networks in Cold Regions under Snowfall Weather. *Reliability Engineering and System Safety*. <https://doi.org/10.1016/j.ress.2025.111824>
- [15] Guo, Z., Zhou, X., Lu, C., Li, S., Zhu, Y., Liu, Y., and Li, Z. (2025). Urban agglomeration transportation resilience: evaluation and evolution analysis using a data-driven model. *Environmental and Sustainability Indicators*, 26, 100714. <https://doi.org/10.1016/j.indic.2025.100714>
- [16] Ha, T. V., and Arimura, M. (2025). Mapping urban mobility during extreme weather events using mobile spatial statistics data. *Transportation Research Interdisciplinary Perspectives*, 33, 101609. <https://doi.org/10.1016/j.trip.2025.101609>
- [17] Hanbarikarekani, M., Qu, X., Zeibots, M., and Qi, W. (2018). Minimizing the average delay at intersections via presignals and speed control. *Journal of Advanced Transportation*, 2018, 4121582. <https://doi.org/10.1155/2018/4121582>
- [18] Kaynak, S., Kaynak, B., Mermer, O., and Demir, I. (2025). City-scale digital twin framework for flood impact analysis: Integrating urban infrastructure and real-time data analytics. *Urban Climate*, 64, 102640. <https://doi.org/10.1016/j.uclim.2025.102640>
- [19] Kuzmin, O., and Bublyk, M. (2016). Economic evaluation and government regulation of technogenic (man-made) damage in the national economy. In *2016 XIth International*

- Scientific and Technical Conference Computer Sciences and Information Technologies (CSIT).*
- [20] Liu, L., Wang, R., Tian, T., and Li, S. (2025). Hazardous materials transportation route optimization considering dynamic wind speeds and risk equity. *Journal of Loss Prevention in the Process Industries*, 98, 105748. <https://doi.org/10.1016/j.ilp.2025.105748>
- [21] Ma, F., Zhang, Y., Liu, Q., Guo, Y., Yang, Z., and Zhang, J. (2025). Modeling urban transportation safety resilience under extreme rainstorms: A catastrophe theory approach. *Reliability Engineering and System Safety*, 263, 111301. <https://doi.org/10.1016/j.ress.2025.111301>
- [22] Maksoud, N., AlJassmi, H., Ali, L., and Masoud, A. R. (2025). Applications of large language models and generative AI in transportation: A systematic review and bibliometric analysis. *Transportation Research Interdisciplinary Perspectives*, 34, 101699. <https://doi.org/10.1016/j.trip.2025.101699>
- [23] Masmoudi, M., Shakrouf, Y., Omar, O. H., Shikhli, A., Abdalla, F., Alketbi, W., ... and Siam, A. I. (2025). Driver risk classification for transportation safety: A machine learning approach using psychological, physiological, and demographic factors with driving simulator. *Engineering Applications of Artificial Intelligence*, 162, 112585. <https://doi.org/10.1016/j.engappai.2025.112585>
- [24] Parsa, A. B., Movahedi, A., Taghipour, H., Derrible, S., and Mohammadian, A. (K.). (2020). Toward safer highways, application of XGBoost and SHAP for real-time accident detection and feature analysis. *Accident Analysis and Prevention*, 136, 105405. <https://doi.org/10.1016/j.aap.2019.105405>
- [25] Saki, S., and Soori, M. (2025). Artificial intelligence, machine learning and deep learning in advanced transportation systems, a review. *Multimodal Transportation*, 5, 100242. <https://doi.org/10.1016/j.multra.2025.100242>
- [26] Theivadas, J. R., and Ponnann, D. S. (2026). DrivNet: Adaptive driving behavior prediction for improving the safety of multimodal transportation in autonomous vehicle systems. *Multimodal Transportation*, 5, 100266. <https://doi.org/10.1016/j.multra.2025.100266>
- [27] Wang, X., Pan, Z., Liu, P., Wang, X., Fan, Z., Liu, N., and Wu, J. (2025). The interaction of weather factors on multimodal public transportation: An empirical research on the combination of bike-sharing and metro. *Cities*, 169, 106537. <https://doi.org/10.1016/j.cities.2025.106537>
- [28] Zhang, C., Zou, X., and Lin, C. (2022). Fusing XGBoost and SHAP Models for Maritime Accident Prediction and Causality Interpretability Analysis. *Journal of Marine Science and Engineering*, 10(8), 1154. <https://doi.org/10.3390/jmse10081154>
- [29] Zhang, X., Gao, R., Xiao, Z., Wang, K., Liu, T., Liang, M., and Zhang, J. (2025a). Natural language processing and text mining in transportation: Current status, challenges, and future roadmap. *Expert Systems with Applications*, 296, 129050. <https://doi.org/10.1016/j.eswa.2025.129050>
- [30] Zhang, Z., Yang, J., Chen, X., and Wu, Y. (2025b). Graph neural networks with flow conservation constraints for real-time origin-destination matrix completion. *Pattern Recognition*, 170, 112046. <https://doi.org/10.1016/j.patcog.2025.112046>
- [31] Zhao, N., Zhang, Y., Lu, Y., and Zhai, G. (2025). Spatial heterogeneity and multimodal complementarity in travel behavior resilience under extreme rainfall: A machine learning approach. *Journal of Transport Geography*, 130, 104472. <https://doi.org/10.1016/j.jtrangeo.2025.104472>
- [32] Zhou, L., Wang, Z., Mao, Z., Wang, Q., Zhang, D., Zhang, Z., and Chen, Z. (2025). Data-driven remaining driving range estimation and analysis framework for electric vehicles under real-world conditions. *Applied Energy*, 402, 126881. <https://doi.org/10.1016/j.apenergy.2025.126881>

References

- [1] Komarnytsky, I.M., Bublyk, M.I. (2008). Assessment of man-made damages and analysis of approaches to their calculation in global and regional aspects. *Bulletin of Lviv Polytechnic National University. Series: Economics and Management Issues*, 628, 134–144.
- [2] Matseliukh Y., & Lytvyn, V. (2025). Application of the K-means clustering method in the organisation of passenger transport in a smart city. *Innovative Technologies and Scientific Solutions for Industries*, 1 (31), 83–101. <https://doi.org/10.30837/2522-9818.2025.1.083>
- [3] Matseliukh Y., & Lytvyn, V. (2025). Concept of a digital twin of passenger transportation by public transport in a smart city. *Taurida Scientific Herald. Series: Technical Sciences*, 1(4), 425-436. <https://doi.org/10.32782/tnv-tech.2025.4.1.40>
- [4] Pasichnyk, A., Pasichnyk, V., & Kushchenko, Y. (2021). Application of system approach in methodology of solving engineering problems. *Transport Systems and Transportation Technologies*, (21), 23. <https://doi.org/10.15802/tstt2021/237635>
- [5] Rachynska, A.V. Classification of risks of the railways as a basis for formation of economic security for its functioning. *Economy and Society*, 6, 81-87.
- [6] Royko Yu., Bura R., Maksymiuk S. Determination of delays on two-lane streets within public transport stops. *Advances In Mechanical Engineering and Transport*, 2(11), 24-28.
- [7] Alonso-Fernandez, F., Butt, T. H., and Tiwari, P. (2025). Overtake detection in trucks using CAN bus signals: A comparative study of machine learning methods. *Expert Systems with Applications*, 297, 129339. <https://doi.org/10.1016/j.eswa.2025.129339>
- [8] Arunachalam, S. (2025). Long-term and short-term traffic flow prediction with different weather conditions using optimized wheel-graph attention based neural network. *Future Generation Computer Systems*, 176, 108198. <https://doi.org/10.1016/j.future.2025.108198>
- [9] Banach, M., and Długosz, R. (2025). Solutions for planning smart hybrid public transportation system based on Google Maps and Voronoi diagrams. *Journal of Computational and Applied Mathematics*, 472, 116775. <https://doi.org/10.1016/j.cam.2025.116775>
- [10] Chukhray, N., Shakhovska, N., Mrykhina, O., et al (2019). Consumer aspects in assessing the suitability of technologies for the transfer. In *2019 IEEE 14th International Conference on Computer Sciences and Information Technologies (CSIT)* (pp. 142–147). <https://doi.org/10.1109/stc-csit.2019.8929879>
- [11] Explain XGBoost Predictions with SHAP. (2025). *XGBoosting.com*. Retrieved from <https://xgboosting.com/explain-xgboost-predictions-with-shap/>
- [12] Gao, P., Shuai, B., Zhang, R., and Wang, B. (2025). STCAGNN-RNKDE: a traffic accident prediction model for spatiotemporal combinatorial attention graph neural networks using Ripley's K and network kernel density estimation. *Reliability Engineering and System Safety*, 265, 111593. <https://doi.org/10.1016/j.res.2025.111593>
- [13] Gong, Y., Yu, P., Zhang, X., Zhang, X., Nie, X., and Sun, H. (2025). Exploiting dynamic spatio-temporal correlations for origin-destination demand prediction. *Expert Systems With Applications*, 299, 130095. <https://doi.org/10.1016/j.eswa.2025.130095>
- [14] Gu, S., Miao, A., Yan, D., Hu, B., Cheng, G., and Xu, Z. (2025). Resilience Assessment of Urban Conventional Bus Networks in Cold Regions under Snowfall Weather. *Reliability Engineering and System Safety*. <https://doi.org/10.1016/j.res.2025.111824>
- [15] Guo, Z., Zhou, X., Lu, C., Li, S., Zhu, Y., Liu, Y., and Li, Z. (2025). Urban agglomeration transportation resilience: evaluation and evolution analysis using a data-driven model. *Environmental and Sustainability Indicators*, 26, 100714. <https://doi.org/10.1016/j.indic.2025.100714>
- [16] Ha, T. V., and Arimura, M. (2025). Mapping urban mobility during extreme weather events using mobile spatial statistics data. *Transportation Research Interdisciplinary Perspectives*, 33, 101609. <https://doi.org/10.1016/j.trip.2025.101609>
- [17] Hanbarikarekani, M., Qu, X., Zeibots, M., and Qi, W. (2018). Minimizing the average delay at intersections via presignals and speed control. *Journal of Advanced Transportation*, 2018, 4121582. <https://doi.org/10.1155/2018/4121582>

- [18] Kaynak, S., Kaynak, B., Mermer, O., and Demir, I. (2025). City-scale digital twin framework for flood impact analysis: Integrating urban infrastructure and real-time data analytics. *Urban Climate*, 64, 102640. <https://doi.org/10.1016/j.uclim.2025.102640>
- [19] Kuzmin, O., and Bublyk, M. (2016). Economic evaluation and government regulation of technogenic (man-made) damage in the national economy. In *2016 XIth International Scientific and Technical Conference Computer Sciences and Information Technologies (CSIT)*.
- [20] Liu, L., Wang, R., Tian, T., and Li, S. (2025). Hazardous materials transportation route optimization considering dynamic wind speeds and risk equity. *Journal of Loss Prevention in the Process Industries*, 98, 105748. <https://doi.org/10.1016/j.jlp.2025.105748>
- [21] Ma, F., Zhang, Y., Liu, Q., Guo, Y., Yang, Z., and Zhang, J. (2025). Modeling urban transportation safety resilience under extreme rainstorms: A catastrophe theory approach. *Reliability Engineering and System Safety*, 263, 111301. <https://doi.org/10.1016/j.res.2025.111301>
- [22] Maksoud, N., AlJassmi, H., Ali, L., and Masoud, A. R. (2025). Applications of large language models and generative AI in transportation: A systematic review and bibliometric analysis. *Transportation Research Interdisciplinary Perspectives*, 34, 101699. <https://doi.org/10.1016/j.trip.2025.101699>
- [23] Masmoudi, M., Shakrouf, Y., Omar, O. H., Shikhli, A., Abdalla, F., Alketbi, W., ... and Siam, A. I. (2025). Driver risk classification for transportation safety: A machine learning approach using psychological, physiological, and demographic factors with driving simulator. *Engineering Applications of Artificial Intelligence*, 162, 112585. <https://doi.org/10.1016/j.engappai.2025.112585>
- [24] Parsa, A. B., Movahedi, A., Taghipour, H., Derrible, S., and Mohammadian, A. (K.). (2020). Toward safer highways, application of XGBoost and SHAP for real-time accident detection and feature analysis. *Accident Analysis and Prevention*, 136, 105405. <https://doi.org/10.1016/j.aap.2019.105405>
- [25] Saki, S., and Soori, M. (2025). Artificial intelligence, machine learning and deep learning in advanced transportation systems, a review. *Multimodal Transportation*, 5, 100242. <https://doi.org/10.1016/j.multra.2025.100242>
- [26] Theivadas, J. R., and Ponnann, D. S. (2026). DrivNet: Adaptive driving behavior prediction for improving the safety of multimodal transportation in autonomous vehicle systems. *Multimodal Transportation*, 5, 100266. <https://doi.org/10.1016/j.multra.2025.100266>
- [27] Wang, X., Pan, Z., Liu, P., Wang, X., Fan, Z., Liu, N., and Wu, J. (2025). The interaction of weather factors on multimodal public transportation: An empirical research on the combination of bike-sharing and metro. *Cities*, 169, 106537. <https://doi.org/10.1016/j.cities.2025.106537>
- [28] Zhang, C., Zou, X., and Lin, C. (2022). Fusing XGBoost and SHAP Models for Maritime Accident Prediction and Causality Interpretability Analysis. *Journal of Marine Science and Engineering*, 10(8), 1154. <https://doi.org/10.3390/jmse10081154>
- [29] Zhang, X., Gao, R., Xiao, Z., Wang, K., Liu, T., Liang, M., and Zhang, J. (2025a). Natural language processing and text mining in transportation: Current status, challenges, and future roadmap. *Expert Systems with Applications*, 296, 129050. <https://doi.org/10.1016/j.eswa.2025.129050>
- [30] Zhang, Z., Yang, J., Chen, X., and Wu, Y. (2025b). Graph neural networks with flow conservation constraints for real-time origin-destination matrix completion. *Pattern Recognition*, 170, 112046. <https://doi.org/10.1016/j.patcog.2025.112046>
- [31] Zhao, N., Zhang, Y., Lu, Y., and Zhai, G. (2025). Spatial heterogeneity and multimodal complementarity in travel behavior resilience under extreme rainfall: A machine learning approach. *Journal of Transport Geography*, 130, 104472. <https://doi.org/10.1016/j.jtrangeo.2025.104472>
- [32] Zhou, L., Wang, Z., Mao, Z., Wang, Q., Zhang, D., Zhang, Z., and Chen, Z. (2025). Data-driven remaining driving range estimation and analysis framework for electric vehicles under real-world conditions. *Applied Energy*, 402, 126881. <https://doi.org/10.1016/j.apenergy.2025.126881>