

І. І. Склярів<sup>1</sup>, М.О. Геревич<sup>2</sup>

<sup>1</sup> Харківський національний автомобільно-дорожній університет, Харків, Україна

<sup>2</sup> Ужгородський національний університет, Ужгород, Україна

## МЕТОД ПРОГНОЗУВАННЯ ТЕХНІЧНОГО СТАНУ ТРАНСПОРТНИХ ЗАСОБІВ ІЗ ВИКОРИСТАННЯМ ТЕХНОЛОГІЙ ВЕЛИКИХ ДАНИХ

**Анотація. Актуальність.** Актуальність роботи зумовлена необхідністю підвищення ефективності сервісних технологій в автомобільній галузі в умовах зростання обсягів телематичних, діагностичних та експлуатаційних даних, що супроводжують функціонування сучасних транспортних засобів. Для сучасних автомобілів характерні складна структура технічних систем, наявність великої кількості взаємопов'язаних компонентів, різноманітність режимів експлуатації та підвищений ризик виникнення відмов, що ускладнює своєчасне оцінювання їх технічного стану. Існуючі підходи до прогнозування технічного стану автомобілів здебільшого не забезпечують комплексного врахування потокових сенсорних даних, історії технічного обслуговування та експлуатаційних параметрів, що ускладнює обґрунтоване прийняття сервісних рішень. Тому актуальною є розробка методу прогнозування технічного стану автомобілів на основі Big Data для підтримки процесу своєчасного виявлення ризикових станів і оптимізації технічного обслуговування. **Метою даної роботи** є розробка методу прогнозування технічного стану автомобілів на основі технологій Big Data шляхом інтеграції потокових сенсорних даних, історії технічного обслуговування та експлуатаційних параметрів для своєчасного виявлення ймовірних відмов, підвищення точності оцінювання технічного стану транспортних засобів і оптимізації сервісних рішень. **Об'єктом дослідження** процес прогнозування технічного стану автомобілів у системах сервісного обслуговування на основі аналізу великих обсягів різноманітних даних. **Предметом дослідження** методи, моделі та засоби прогнозування технічного стану автомобілів на основі Big Data шляхом інтеграції сенсорних, експлуатаційних і сервісних даних. **Результати.** У роботі розглянуто задачу прогнозування технічного стану автомобілів на основі Big Data з метою підвищення ефективності сервісних технологій в автомобільній галузі. Запропоновано метод, що передбачає інтеграцію експлуатаційних, сенсорних і сервісних даних у межах єдиного аналітичного контуру для оцінювання ризику виникнення відмов. Реалізацію методу виконано мовою Python у середовищі Google Colab із використанням відкритого набору даних SCANIA Component X. У процесі дослідження виконано підготовку даних, усунення витоку цільової змінної, побудову прогновної моделі, налаштування порога класифікації та формування сервісних рішень. Отримані результати підтвердили високу якість прогнозування та придатність запропонованого підходу до використання в задачах підтримки технічного обслуговування транспортних засобів.

**Ключові слова:** Big Data; прогнозування технічного стану; автомобільна галузь; машинне навчання; аналіз даних; технічне обслуговування; транспортні засоби; класифікація; бінарний класифікатор; Google Colab; Python.

### Вступ

Сучасний етап розвитку автомобільної галузі характеризується глибокою цифровою трансформацією, що охоплює не лише процеси проектування та виробництва транспортних засобів, а й системи їх технічного супроводу, сервісного обслуговування та підтримки після продажу. Умови зростання складності автомобільних систем, поширення електронних блоків керування, сенсорних мереж, телематичних платформ та підключених сервісів зумовлюють безперервне накопичення значних обсягів різноманітних даних про функціонування транспортних засобів. Такі дані формуються в процесі експлуатації автомобілів, технічного обслуговування, діагностики окремих вузлів і агрегатів, взаємодії з сервісною інфраструктурою та користувачами. У цьому контексті технології Big Data набувають особливого значення як інструмент інтеграції, зберігання, оброблення та інтелектуального аналізу великих масивів інформації з метою підвищення ефективності сервісних технологій в автомобільній галузі.

Традиційні підходи до оцінювання технічного стану автомобілів ґрунтуються переважно на регламентованих інтервалах технічного обслуговування, результатах періодичної діагностики або експертних оцінках фахівців. За таких умов прийняття рішень щодо обслуговування часто має реактивний характер,

тобто виконується після виявлення несправності або у межах заздалегідь встановлених часових чи інтервалів пробігу. Однак індивідуальні режими експлуатації транспортних засобів, відмінності в дорожніх, кліматичних і навантажувальних умовах, а також варіативність поведінки водія суттєво впливають на темпи зношування компонентів і ризику виникнення відмов. Унаслідок цього регламентні підходи не завжди забезпечують належну точність прогнозування технічного стану, що призводить або до передчасного виконання сервісних робіт, або до запізненого виявлення критичних змін у функціонуванні автомобіля.

Одним із перспективних напрямів розв'язання цієї проблеми є перехід до моделей обслуговування з використанням прогнозування, заснованих на аналізі фактичних експлуатаційних даних. Використання Big Data у поєднанні з методами інтелектуального аналізу даних і машинного навчання відкриває можливість виявлення прихованих закономірностей у поведінці технічних систем, ідентифікації процесів деградації та формування обґрунтованих прогнозів щодо майбутнього стану автомобіля. Особливою актуальності це набуває у сфері сервісних технологій, де своєчасне виявлення ознак потенційної відмови дозволяє оптимізувати графіки технічного обслуговування, знижувати експлуатаційні витрати, підвищувати надійність транспортних засобів і покращувати якість сервісу для кінцевого споживача.

Актуальність дослідження посилюється також тим, що на практиці доступ до повномасштабних реальних наборів даних автомобільної галузі часто є обмеженим через комерційну конфіденційність, неоднорідність форматів зберігання, неповноту спостережень і складність отримання тривалих часових рядів для великої кількості транспортних засобів. У зв'язку з цим обґрунтованим підходом є використання синтетичних датасетів, які відтворюють структурні, статистичні та причинно-логічні властивості реальних даних і дають змогу проводити відтворення експериментальні дослідження. Формування такого набору даних у контексті Big Data дозволяє моделювати значний обсяг записів, різноманітність джерел інформації, наявність шумів, пропусків, дисбалансу класів і залежностей між експлуатаційними режимами та технічними відмовами.

**Метою роботи** є розробка методу прогнозування технічного стану автомобілів на основі технологій Big Data шляхом інтеграції потокових сенсорних даних, історії технічного обслуговування та експлуатаційних параметрів для своєчасного виявлення ймовірних відмов, підвищення точності оцінювання технічного стану транспортних засобів і оптимізації сервісних рішень.

### Основна частина

Для розв'язання задачі прогнозування технічного стану автомобілів у межах розвитку сервісних технологій автомобільної галузі необхідно врахувати сучасні підходи до аналізу телематичних, діагностичних та експлуатаційних даних, визначити їхні функціональні можливості, переваги й обмеження, а також обґрунтувати доцільність розроблення власного методу на основі технологій Big Data. У зв'язку з цим доцільним є аналіз наукових досліджень, присвячених використанню методів машинного навчання, інтелектуального аналізу даних і прогнозного обслуговування у задачах оцінювання технічного стану транспортних засобів.

У статті [1] розглянуто актуальну науково-прикладну задачу прогнозування технічного стану автомобілів на основі технологій Big Data з метою підвищення ефективності сервісних технологій в автомобільній галузі. У роботі запропоновано метод прогнозування технічного стану автомобілів, який базується на інтеграції потокових даних із сенсорів, історії технічного обслуговування та експлуатаційних параметрів у єдиному аналітичному контурі. У роботі проаналізовано наукові публікації з позицій прикладних сценаріїв і типів методів машинного навчання, а також обговорюються відкриті проблеми, серед яких обмеженість доступних даних, залежність більшості підходів від розмічених вибірок, потреба в поєднанні кількох джерел даних і необхідність підвищення інтерпретації моделей.

У статті [2] автори розглядають задачу прогнозного технічного обслуговування в автомобільній галузі на основі даних бортових сенсорів і пропонують конвеєр PREPIPE для прогнозування стану засмічення кисневого датчика дизельного двигуна. У роботі використано часові ряди сигналів, зібраних з бортового блоку керування двигуном, а сам підхід

охоплює вибір інформативних сигналів, перетворення часових рядів у ознаки, відбір найважливіших характеристик, урахування історичних ознак і порівняння різних класифікаційних моделей, зокрема з архітектурами глибокого навчання. Автори показують, що якісна попередня обробка та інженерія ознак є критично важливими для досягнення високої точності прогнозування. Водночас дослідження зосереджене на окремому вузлі автомобіля, хоча було б логічним використовувати ширший підхід.

У роботі [3] автори розглядають сучасний стан застосування методів машинного навчання для прогнозного технічного обслуговування автомобілів і діагностики технічного стану транспортних засобів. У роботі підкреслено, що профілактика є важливим напрямком розвитку автомобільної галузі, оскільки дає змогу підвищити якість сервісу, своєчасно виявляти ризики відмов і зменшувати ймовірність критичних несправностей. Водночас наголошено, що в автомобільному секторі існують суттєві обмеження, пов'язані з недостатньою доступністю сенсорів, складністю отримання необхідних даних і обмеженістю окремих проектних рішень. У зв'язку з цим методи машинного навчання розглядаються як перспективний інструмент для аналізу навіть обмежених за обсягом даних і побудови моделей оцінювання технічного стану автомобілів.

У роботі [4] автори узагальнюють сучасні напрями застосування штучного інтелекту, машинного та глибокого навчання в транспортних системах. У роботі показано, що машинне навчання використовується для керування трафіком, автономного транспорту, інтелектуального паркування, оптимізації громадського транспорту, логістики, безпеки та моніторингу інфраструктури. Автори наголошують, що ці технології забезпечують прийняття рішень, дають змогу обробляти великі обсяги даних у реальному часі, прогнозувати тенденції та підвищувати стійкість транспортних систем. Водночас у статті окреслено ключові обмеження таких рішень, зокрема дефіцит і неоднорідність даних, недостатню узагальнюваність моделей, високі обчислювальні витрати, а також етичні й регуляторні проблеми, пов'язані з прозорістю, упередженістю та конфіденційністю даних.

У статті [5] автори здійснюють систематизований огляд застосування штучного інтелекту для діагностики несправностей транспортних засобів і прогнозного технічного обслуговування. У роботі підкреслено, що традиційні підходи, засновані на правилах і даних бортової діагностики, мають певні обмеження, оскільки переважно орієнтовані на виявлення вже наявних відмов, тоді як сучасні автомобілі потребують більш гнучких і прогностично орієнтованих засобів аналізу. Автори показують, що методи штучного інтелекту, зокрема машинного та глибокого навчання, дають змогу аналізувати дані з різних сенсорів, виявляти аномалії, прогнозувати деградацію технічного стану та підвищувати точність діагностики різних підсистем автомобіля, серед яких двигун, підвіска, шини, трансмісія та гальмівна система. Особливу увагу приділено підходам, заснованим на знаннях, аналізу даних з багатьох сенсорів, а також поєднанню різних

інтелектуальних методів для досягнення більш надійної й універсальної діагностичної підтримки. Водночас у роботі наголошено на проблемах гетерогенності даних, складності узагальнення моделей, залежності від якості вхідної інформації та потреби в комплексних рішеннях для діагностики транспортного засобу в цілому, а не лише його окремих вузлів.

У статті [6] автори розглядають проблему формалізації процесу розробки моделей профілактичного технічного обслуговування в умовах великих обсягів даних, міждисциплінарної взаємодії та складної організації робіт. У роботі зазначено, що зростання доступності даних створює нові можливості для профілактичного обслуговування, однак сам процес побудови таких моделей залишається складним, трудомістким і потребує координації між фахівцями предметної області, аналітиками даних, ІТ-фахівцями та розробниками програмного забезпечення. Для подолання цих труднощів автори пропонують еталонну модель розробки рішень профілактичного обслуговування, побудовану на основі CRISP-DM і структуровану за чотирма фазами, що охоплюють повний життєвий цикл моделі: від опису компонента і сценарію використання до розгортання, моніторингу та повторного використання результатів. Особливу увагу приділено архітектурі роботи з даними, яка включає збір сенсорної інформації, попередню обробку, псевдонімізацію, централізоване зберігання в сховище даних, підготовку даних, моделювання, оцінювання та виконання моделі в хмарному або бортовому середовищі. Водночас стаття має переважно процесно-організаційний характер і зосереджується на еталонній схемі розробки моделей профілактичного обслуговування, а не на конкретному методі прогнозування технічного стану автомобіля.

У статті [7] автори здійснюють огляд підходів до моніторингу технічного стану транспортних засобів і діагностики несправностей для традиційних, електричних та автономних автомобілів. У роботі підкреслено, що автомобільна галузь досі не має уніфікованої структури, яка б поєднувала різні методи діагностики з урахуванням масштабованості, адаптивності та ефективності в різних типах транспортних засобів. Автори систематизують сучасні підходи обробки даних, акцентують увагу на ролі сучасних сенсорів, IoT, штучного інтелекту, багатосенсорної інтеграції та аналітики великих даних, а також пропонують концептуальну модель для раннього виявлення несправностей. У статті окремо наголошено, що методи штучного інтелекту та Big Data стають ключовими засобами для аналізу великих потоків сенсорних даних, виявлення аномалій і зниження ризику раптових відмов.

Варто зазначити, що в контексті обробки даних сенсорами велике значення мають інтелектуальні методи енергозбереження в сенсорних мережах. І також є доцільним розглянути дослідження за цим напрямком. В роботі [8] розглянуто застосування штучних нейронних мереж для підвищення енергоефективності безпроводних сенсорних мереж. Автор підкреслює, що обмеженість енергетичних ресурсів є однією з головних проблем таких мереж, а тому актуальним є використання інтелектуальних інструментів для зниження витрат на передавання даних, кластеризації вузлів,

прогнозування сенсорної інформації та оптимізації маршрутизації. У роботі показано, що нейронні мережі добре узгоджуються з природою сенсорних мереж завдяки можливості паралельної обробки, розподіленого зберігання даних, автоматизованої класифікації та роботи з шумними даними. Хоча стаття присвячена безпроводним сенсорним мережам, а не безпосередньо автомобільній галузі, її результати добре підтримують ідею використання сенсорних потоків, інтелектуальної обробки та прогнозних моделей у задачах оцінювання технічного стану автомобілів. Водночас робота має ширший інфраструктурний характер і не розглядає інтеграцію сервісної історії, експлуатаційних параметрів та автомобільних телематичних даних у єдиному контурі прогнозування.

Робота [9] присвячена підвищенню енергоефективності безпроводних сенсорних мереж в умовах обмежених ресурсів вузлів і необхідності тривалого автономного функціонування. У роботі підкреслено, що енергоспоживання таких мереж залежить не лише від апаратної конфігурації, а й від способу організації обміну даними, топології мережі, частоти передавання та впливу зовнішнього середовища. Хоча робота присвячена безпроводним сенсорним мережам, а не безпосередньо автомобільній галузі, вона підтверджує загальний принцип, за яким складні системи з великою кількістю сенсорів потребують інтелектуального аналізу даних, адаптивного керування та прогнозних механізмів для підвищення ефективності функціонування.

У статті [10] подано огляд сучасних підходів до діагностики несправностей і моніторингу технічного стану складних інженерних систем. Автори наголошують, що зі зростанням складності технічних об'єктів підвищуються вимоги до надійності, безпеки та своєчасного виявлення процесів деградації, а тому методи діагностики несправностей та моніторингу на основі даних набувають критичного значення для раннього виявлення дефектів, прогнозування залишкового ресурсу та планування технічного обслуговування. У роботі узагальнено результати шістнадцяти досліджень, присвячених застосуванню штучного інтелекту, глибокого навчання, цифрових двійників, аналізу вібраційних і струмових сигналів, моніторингу даних з сенсорів та методів роботи з малими й незбалансованими наборами даних. Окремо підкреслено, що сучасні рішення орієнтуються на раннє виявлення несправностей, підвищення точності діагностики, зменшення простоїв і підтримку прогнозного технічного обслуговування в умовах обмежених або шумних даних. Водночас стаття має широкий міжгалузевий характер і не зосереджується безпосередньо на автомобільній сервісній інфраструктурі чи інтеграції телематичних, сервісних та експлуатаційних даних автомобіля.

Проведений аналіз розглянутих праць засвідчив, що сучасні дослідження підтверджують високу ефективність підходів, заснованих на Big Data, машинному навчанні та інтелектуальному аналізі сенсорних даних, для задач діагностики несправностей і прогнозування технічного стану транспортних засобів. Водночас більшість наявних робіт або зосереджена на окремих вузлах автомобіля, або має оглядовий характер, не забезпечуючи цілісної інтеграції

потоків сенсорних даних, історії технічного обслуговування та експлуатаційних параметрів у межах єдиного методу. Це дає підстави стверджувати, що розроблення власного методу прогнозування технічного стану автомобілів на основі Big Data є обґрунтованим і актуальним.

Нехай множина автомобілів, для яких здійснюється моніторинг технічного стану, задається як

$$V = \{v_1, v_2, \dots, v_N\}, \quad (1)$$

де  $N$  – кількість автомобілів у вибірці. Для кожного автомобіля  $v_i$  у процесі експлуатації формується множина різнорідних даних, яка включає потокові сенсорні дані, історію технічного обслуговування та експлуатаційні параметри. Тоді інтегрований опис автомобіля можна подати у вигляді

$$D_i = \{S_i, H_i, E_i\}, \quad (2)$$

де  $S_i$  – потокові сенсорні дані,  $H_i$  – історія технічного обслуговування,  $E_i$  – експлуатаційні параметри. Потоків сенсорні дані для автомобіля  $v_i$  задаються часовим рядом

$$S_i(t) = (s_{i1}(t), s_{i2}(t), \dots, s_{im}(t)), \quad (3)$$

де  $m$  – кількість сенсорних параметрів,  $s_{ij}(t)$  – значення  $j$ -го сенсорного параметра в момент часу  $t$ . До таких параметрів можуть належати температура двигуна, тиск мастила, рівень вібрації, оберти двигуна, напруга акумулятора, швидкість, витрата пального та інші показники. Історія технічного обслуговування подається вектором

$$H_i = (h_{i1}, h_{i2}, \dots, h_{ip}), \quad (4)$$

де  $p$  – кількість ознак сервісної історії, серед яких можуть бути кількість попередніх ремонтів, тривалість інтервалу між сервісним обслуговуванням, кількість зафіксованих несправностей, дата останнього технічного обслуговування, типи виконаних сервісних робіт. Експлуатаційні параметри задаються як

$$E_i = (e_{i1}, e_{i2}, \dots, e_{iq}), \quad (5)$$

де  $q$  – кількість параметрів експлуатації, зокрема стиль водіння, тип дорожнього покриття, інтенсивність використання, кліматичні умови, середнє навантаження, режим руху. Оскільки сенсорні дані мають потіковий характер, для прогнозування технічного стану використовується механізм ковзного часового вікна. Для автомобіля  $v_i$  на інтервалі часу  $[t - \Delta t, t]$  формується підмножина даних

$$W_i(t) = \{S_i(\tau) \mid \tau \in [t - \Delta t, t]\}, \quad (6)$$

де  $\Delta t$  – довжина часового вікна. На основі часового вікна для кожного сенсорного параметра обчислюються статистичні та динамічні ознаки. Для  $j$  параметра вони можуть бути визначені таким чином:

- середнє значення

$$\mu_{ij}(t) = \sum_{k=1}^K s_{ij}(t_k) / K, \quad (7)$$

- стандартне відхилення

$$\sigma_{ij}(t) = \sqrt{\frac{1}{K} \sum_{k=1}^K (s_{ij}(t_k) - \mu_{ij}(t))^2}, \quad (8)$$

- лінійний тренд зміни параметра

$$\beta_{ij}(t) = \sum_{k=1}^K (t_k - \bar{t})(s_{ij}(t_k) - \bar{s}_{ij}) / \sum_{k=1}^K (t_k - \bar{t})^2, \quad (9)$$

де  $K$  – кількість спостережень у часовому вікні,  $\bar{t}$  – середнє значення часу у вікні,  $\bar{s}_{ij}$  – середнє значення  $j$ -го параметра у вікні.

Для виявлення аномальних режимів роботи визначається частота перевищення допустимого порогу:

$$a_{ij}(t) = \frac{1}{K} \sum_{k=1}^K I(s_{ij}(t_k) > s_j^{crit}), \quad (10)$$

де  $I()$  – індикаторна функція,  $s_j^{crit}$  – критичне значення  $j$ -го параметра. Після цього формується узагальнений вектор ознак технічного стану автомобіля:

$$X_i(t) = [F_i^{(S)}(t), H_i, E_i] \quad (11)$$

де  $F_i^{(S)}(t)$  – вектор ознак, сформований із сенсорних часових рядів у вікні  $[t - \Delta t, t]$ . Отже,  $X_i(t)$  є інтегрованим описом технічного стану автомобіля в момент часу  $t$ . Ціль методу полягає у прогнозуванні ймовірності виникнення відмови або потреби в технічному обслуговуванні у майбутньому часовому горизонті  $[t, t + \Delta p]$ , де  $\Delta p$  – горизонт прогнозування. Тоді цільова змінна задається як

$$y_i(t) = \begin{cases} 1, & \text{в } [t, t + \Delta p] \text{ відмова або потрібне ТО,} \\ 0, & \text{інакше.} \end{cases} \quad (12)$$

Задача прогнозування формалізується як задача бінарної класифікації, у якій необхідно побудувати відображення

$$f : X_i(t) \rightarrow \hat{y}_i(t) \quad (13)$$

де  $\hat{y}_i(t)$  – прогнозований клас технічного стану. Для оцінювання ризику відмови використовується ймовірнісна модель

$$P_i(t) = P(y_i(t) = 1 \mid X_i(t)), \quad (14)$$

У найпростішому випадку така модель може бути реалізована логістичною функцією:

$$P_i(t) = \frac{1}{1 + \exp\left(-\left(w_0 + \sum_{r=1}^M w_r x_{ir}(t)\right)\right)}, \quad (15)$$

де  $M$  – розмірність вектору ознак,  $x_{ir}(t)$  –  $r$ -та ознака для автомобіля,  $v_i$ ,  $w_r$  – вагові коефіцієнти моделі. Правило прийняття рішення щодо технічного стану має такий вигляд:

$$\hat{y}_i(t) = \begin{cases} 1, & \text{якщо } P_i(t) \geq \theta, \\ 0, & \text{якщо } P_i(t) < \theta. \end{cases} \quad (16)$$

де  $\theta$  – порогове значення, що визначається експериментально.

Для підвищення чутливості методу до процесів деградації вводиться інтегральний індекс технічного ризику, який враховує вплив сенсорних, сервісних та експлуатаційних факторів:

$$R_i(t) = \alpha R_i^{(S)}(t) + \beta R_i^{(H)} + \gamma R_i^{(E)}, \quad (17)$$

де  $R_i^{(S)}(t)$  – ризик, оцінений за сенсорними даними,  $R_i^{(H)}$  – ризик на основі історії технічного обслуговування,  $R_i^{(E)}$  – ризик, зумовлений умовами експлуатації,  $\alpha, \beta, \gamma$  – вагові коефіцієнти, для яких виконується така умова:

$$\alpha + \beta + \gamma = 1. \quad (18)$$

Сенсорна складова ризику може бути визначена як зважена сума нормованих ознак:

$$R_i^{(S)}(t) = \sum_{j=1}^m \lambda_j z_{ij}(t), \quad (19)$$

де  $z_{ij}(t)$  – нормоване значення  $j$ -ї сенсорної ознаки,

$\lambda_j$  – коефіцієнт важливості ознаки,  $\sum_{j=1}^m \lambda_j = 1$ .

Сервісна складова ризику може бути задана, наприклад, через кількість попередніх відмов і тривалість після останнього технічного обслуговування:

$$R_i^{(H)} = \eta_1 \frac{n_i^{fail}}{n_{\max}^{fail}} + \eta_2 \frac{d_i^{serv}}{d_{\max}^{serv}}, \quad (20)$$

де  $n_i^{fail}$  – кількість попередніх відмов,  $d_i^{serv}$  – пробіг або час після останнього технічного обслуговування,  $\eta_1 + \eta_2 = 1$ , а експлуатаційна складова ризику визначається як

$$R_i^{(E)} = \sum_{l=1}^q \rho_l e_{il}^*, \quad (21)$$

де  $e_{il}^*$  – нормовані експлуатаційні параметри,  $\rho_l$  – їх вагові коефіцієнти,  $\sum_{l=1}^q \rho_l = 1$ . Тоді кінцеве правило сервісного рішення можна подати як

$$u_i(t) = \begin{cases} u_1, & \text{якщо } R_i(t) < \delta_1, \\ u_2, & \text{якщо } \delta_1 \leq R_i(t) < \delta_2, \\ u_3, & \text{якщо } R_i(t) \geq \delta_2, \end{cases} \quad (22)$$

де  $u_1$  – продовження експлуатації без втручання,  $u_2$  – призначення додаткової діагностики,  $u_3$  – рекомендація щодо термінового технічного обслуговування;  $\delta_1, \delta_2$  – пороги прийняття рішення.

Отже, запропонований метод включає послідовність етапів: збір та інтеграцію різномірних даних, формування часових вікон, обчислення статистичних і динамічних ознак, побудову прогнозу моделі для визначення ймовірності відмови, обчислення інтегрального індексу ризику та формування сервісного рішення.

На відміну від підходів, що спираються лише на окремі сенсорні сигнали або на вузькоспеціалізовані дані, запропонований метод враховує багатофакторну природу технічного стану автомобіля та дає змогу підвищити обґрунтованість прогнозування в умовах Big Data.

Для перевірки працездатності запропонованого методу використано середовище Google Colab і мову Python. Реалізацію виконано із застосуванням бібліотек NumPy, Pandas, Scikit-learn та Matplotlib, що забезпечило повний цикл оброблення даних, формування ознак, навчання моделі та оцінювання результатів.

Для експериментальної перевірки запропонованого методу було обрано відкритий набір даних SCANIA Component X Dataset [11], оскільки він найбільшою мірою відповідає предметній області дослідження та містить реальні багатовимірні дані про технічний стан транспортних засобів, їх експлуатаційні характеристики та сервісно-орієнтовані мітки.

Для формування робочого масиву даних було використано шість файлів набору:

- test\_operational\_readouts.csv,
- test\_specifications.csv,
- train\_tte.csv,
- validation\_operational\_readouts.csv,
- validation\_specifications.csv,
- validation\_labels.csv.

На першому етапі було проаналізовано структуру таблиць і склад їхніх атрибутів, після чого виконано поєднання файлів за спільними ідентифікаційними ознаками. Операційні дані, що містили показники функціонування компонентів у часовій динаміці, були інтегровані зі параметрами специфікації транспортних засобів, а також із цільовими мітками, що характеризують технічний стан. У результаті злиття було сформовано єдиний узгоджений датасет, придатний для подальшого машинного аналізу. Після об'єднання даних було виконано їх первинну перевірку, в межах якої оцінено розмір отриманої вибірки, структуру ознак та наявність пропущених значень. Сформований масив виявився неоднорідним за повнотою заповнення, що відповідає природі реальних технічних даних. Подальша підготовка полягала у виділенні рядків, для яких наявна цільова мітка, а також у приведенні задачі до бінарної постановки: нормальний технічний стан було позначено як клас 0, а наявність ризику відмови – як клас 1. Такий підхід дозволив узгодити експеримент із розробленим методом прогнозування технічного стану в термінах своєчасного виявлення небезпечних станів. Для забезпечення коректності обчислювального експерименту зі складу ознак було вилучено змінні, що могли прямо або опосередковано містити інформацію про цільовий клас і призводити до витоку цільової змінної.

Окремо було виконано фільтрацію ознак із надмірною часткою пропусків, після чого набір було поділено на числові та категоріальні параметри. Це створило основу для подальшої побудови прогнозу моделі та дало змогу перейти до етапу оцінювання якості методу на очищених і структурованих даних.

Основні результати експерименту свідчать, що після усунення витoku цільової змінної та оптимізації порога класифікації запропонований метод забезпечив високу якість прогнозування технічного стану автомобілів. Матриця помилок на рис. 1 показує, що модель правильно ідентифікувала 38202 випадки нормального технічного стану та 925 випадків ризику відмови.



Рис. 1. Матриця помилок

Водночас зафіксовано 29 хибнопозитивних рішень, коли справний стан було віднесено до ризикового, і 90 хибнонегативних рішень, коли ризиковий стан залишився невиявленим. Такі результати дають підстави стверджувати, що модель характеризується дуже високою точністю розпізнавання нормального стану та достатньо високою здатністю виявляти ризикові режими. Особливо важливим є те, що кількість хибнопозитивних рішень залишається незначною, що знижує ймовірність необґрунтованого сервісного втручання.

Отримані значення  $ROC-AUC = 0.9962$  та  $AP = 0.9674$  на рис. 2 та рис. 3 відповідно підтверджують високу роздільну здатність моделі та її здатність відокремлювати ризикові стани від нормальних навіть в умовах істотного дисбалансу класів.

Це свідчить про те, що запропонований метод є не лише придатним для класифікації, а й ефективним для ранжування об'єктів за рівнем ризику, що має особливе значення для задач сервісної аналітики.

Аналіз графіка залежності метрик від порога, який представлений на рис. 4 показав, що використання стандартного порога не є оптимальним для даної задачі. Підібране значення порога на рівні 0.15 забезпечило більш збалансоване співвідношення між точністю та повнотою виявлення ризикових станів.

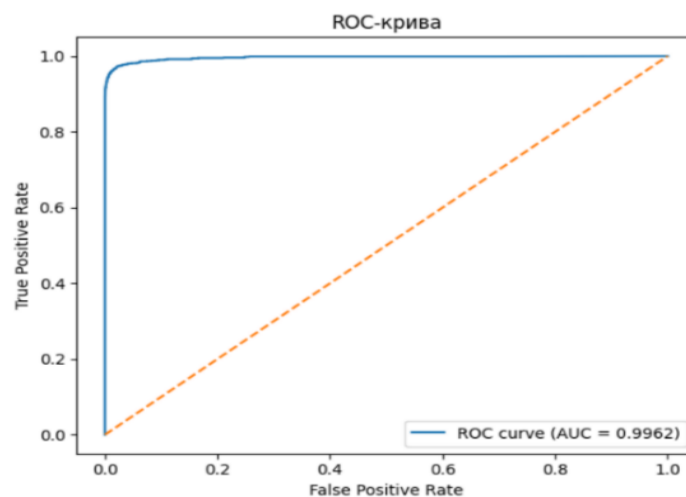


Рис. 2. ROC крива

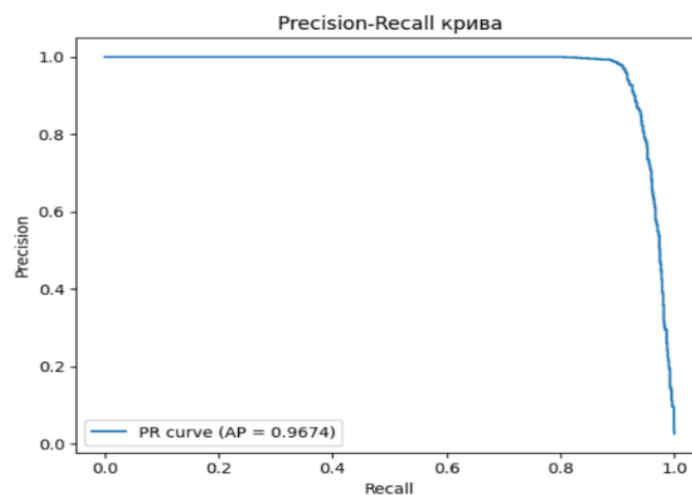


Рис. 3. Precision-Recall крива

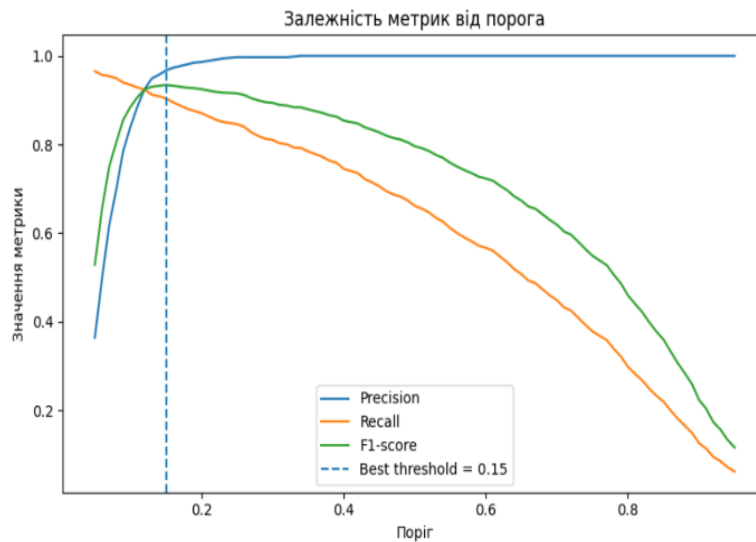


Рис. 4. Залежність метрик від порога

Саме завдяки цьому вдалося зменшити кількість пропущених відмовних випадків і підвищити практичну цінність моделі для прийняття сервісних рішень.

Додатково встановлено, що найбільший внесок у формування прогнозу мають ознаки груп 459, 158, 837, 167, 309 і 397, що представлені на рис.5, які можна розглядати як найбільш інформативні параметри технічного стану в межах досліджуваного набору даних.

Це підтверджує, що після очищення набору ознак модель спирається на реальні експлуатаційні характеристики, а не на службові або цільові змінні.

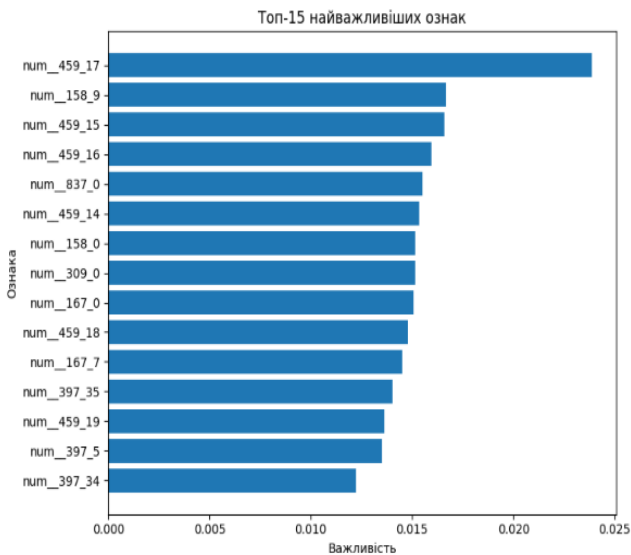


Рис. 5. Важливі ознаки

Таким чином, результати експерименту підтверджують працездатність запропонованого методу та його придатність для задач прогнозування технічного стану автомобілів на основі великих даних. Практичне значення отриманих результатів полягає у можливості своєчасного виявлення ризикових станів, зниження ймовірності раптових відмов і підвищення обґрунтованості рішень щодо технічного обслуговування.

## Висновки

Розроблено метод прогнозування технічного стану автомобілів, який базується на інтеграції сенсорних, експлуатаційних та сервісних даних і дозволяє формувати оцінку ризику відмови з подальшим перетворенням її на сервісне рішення.

Метод формалізовано математично, описано його основні етапи та реалізовано засобами Python у середовищі Google Colab. В якості датасету використано SCANIA Component X, на базі якого сформовано єдиний робочий набір даних для перевірки працездатності підходу. У процесі моделювання виконано підготовку даних, усунення витоку цільової змінної, відбір інформативних ознак, побудову моделі для прогнозу та налаштування порога класифікації. Отримані результати показали високу якість прогнозування: модель забезпечила надійне відокремлення нормальних станів від ризикових, а також продемонструвала високу придатність до ранжування об'єктів за рівнем ризику. Аналіз матриці помилок, ROC- та PR-характеристик підтвердив, що запропонований метод може бути використаний для підтримки рішень щодо продовження експлуатації, додаткової діагностики або термінового технічного обслуговування.

Практичне значення роботи полягає в тому, що запропонований підхід може бути основою для побудови інтелектуальних сервісних систем в автомобільній галузі, орієнтованих на своєчасне виявлення ризикових технічних станів, зниження ймовірності раптових відмов і підвищення обґрунтованості технічного обслуговування.

**Конфлікт інтересів.** Автори декларують, що не мають конфлікту інтересів стосовно даного дослідження, в тому числі фінансового, особистісного характеру, авторства чи іншого характеру, що міг би вплинути на дослідження та його результати, представлені в даній статті.

**Використання засобів штучного інтелекту.** Автори підтверджують, що не використовували технології штучного інтелекту при створенні представленої роботи.

## СПИСОК ЛІТЕРАТУРИ

1. Andreas Theissler, Judith Pérez-Velázquez, Marcel Kettelgerdes, Gordon Elger. Predictive maintenance enabled by machine learning: Use cases and challenges in the automotive industry. *Reliability Engineering & System Safety*. Volume 215. 2021. 23 p. <https://doi.org/10.1016/j.ress.2021.107864>.
2. Danilo Giordano, Flavio Giobergia, Eliana Pastor, Antonio La Macchia, Tania Cerquitelli, Elena Baralis, Marco Mellia, Davide Tricarico. Data-driven strategies for predictive maintenance: Lesson learned from an automotive use case. *Computers in Industry*. Volume 134. 2022. <https://doi.org/10.1016/j.compind.2021.103554>.
3. M. Jain, D. Vasdev, K. Pal, V. Sharma. Systematic literature review on predictive maintenance of vehicles and diagnosis of vehicle's health using machine learning techniques. *Computational Intelligence*, 38. 2022. p. 1990–2008. <https://doi.org/10.1111/coin.12553>.
4. Siavash Saki, Mohsen Soori. Artificial intelligence, machine learning and deep learning in advanced transportation systems, a review. *Multimodal Transportation*. Volume 5, Issue 1. 2026. 25 p. <https://doi.org/10.1016/j.multra.2025.100242>.
5. Md Naeem Hossain, Md Mustafizur Rahman, Devarajan Ramasamy. Artificial Intelligence-Driven Vehicle Fault Diagnosis to Revolutionize Automotive Maintenance: A Review. *CMES - Computer Modeling in Engineering and Sciences*. Volume 141, Issue 2. 2024. P. 951-996. <https://doi.org/10.32604/cmcs.2024.056022>.
6. Sielaff Lennard, Lucke Dominik, Wolf Yannic. A Reference Model for Predictive Maintenance Model Development. *Procedia CIRP*. Volume 130. 2024. P. 1537-1542. <https://doi.org/10.1016/j.procir.2024.10.279>.
7. Md Naeem Hossain, Md Mustafizur Rahman, Devarajan Ramasamy. Advances in intelligent vehicular health monitoring and fault diagnosis: Techniques, technologies, and future directions. *Measurement*. Volume 253, Part B. 2025. <https://doi.org/10.1016/j.measurement.2025.117618>.
8. Дяченко В.О. Інтелектуальні підходи енергозбереження у безпроводних сенсорних комп'ютерних мережах. *Системи управління, навігації та зв'язку*, т. 4 (62), 2020. P. 114-118. <https://doi.org/10.26906/SUNZ.2020.4.114>.
9. Harmash V., Diachenko V., Mikhal O., Znaidiuk V. Energy-Saving Method In Wireless Sensor Networks // *Control, Navigation and Communication Systems*, v.2 (80), 2025. P.54–58. <https://doi.org/10.26906/SUNZ.2025.2.054>.
10. Li Y, Wang T, Noman K, Li B. Advanced Fault Diagnosis and Health Monitoring Techniques for Complex Engineering Systems: 2nd Edition. *Sensors*. 2025; 25(22):7054. <https://doi.org/10.3390/s25227054>.
11. SCANIA Component X Dataset: A Real-World Multivariate Time Series Dataset for Predictive Maintenance. <https://doi.org/10.58141/1w9m-yz81>.

Received (Надійшла) 09.01.2026

Accepted for publication (Прийнята до друку) 25.03.2026

Publication date (Дата публікації) 22.05.2026

## ВІДОМОСТІ ПРО АВТОРІВ/ ABOUT THE AUTHORS

**Склярів Ілля Ігорович** – аспірант кафедри комп'ютерних наук і інформаційних систем, Харківський національний автомобільно-дорожній університет, Харків, Україна;

**Illia Skliarov** – PhD student, Department of Computer Science and Information Systems, Kharkiv National Automobile and Highway University, Kharkiv, Ukraine;

e-mail: [is.ilyasklyarov@gmail.com](mailto:is.ilyasklyarov@gmail.com); ORCID Author ID: <http://orcid.org/0009-0001-9116-5315>.

**Геревич Михайло Олександрович** – доктор філософії, доцент кафедри теорії та історії держави і права, Ужгородський національний університет, Ужгород, Україна;

**Mykhailo Herevych** – PhD, Associate Professor of Department of Theory and History of State and Law, Uzhhorod, Ukraine; e-mail: [mykhailo.herevych@uzhnu.edu.ua](mailto:mykhailo.herevych@uzhnu.edu.ua); ORCID Author ID: <https://orcid.org/0000-0002-0842-2828>.

**Method for predicting the technical condition of vehicles based on Big Data**

Illia Skliarov, Mykhailo Herevych

**Abstract. Relevance.** The relevance of the study is determined by the need to improve the efficiency of service technologies in the automotive industry under conditions of growing volumes of telematics, diagnostic, and operational data accompanying the functioning of modern vehicles. Modern automobiles are characterized by a complex structure of technical systems, the presence of a large number of interconnected components, the heterogeneity of operating modes, and an increased risk of failures, which complicates the timely assessment of their technical condition. Existing approaches to forecasting the technical condition of automobiles mostly do not ensure the comprehensive consideration of streaming sensor data, maintenance history, and operational parameters, which complicates the substantiated decision-making process regarding service actions. Therefore, the development of a method for forecasting the technical condition of automobiles based on Big Data is relevant for supporting the timely detection of risky conditions and optimizing maintenance. **Purpose of the article** is to develop a method for forecasting the technical condition of automobiles based on Big Data technologies through the integration of streaming sensor data, maintenance history, and operational parameters for the timely detection of probable failures, improvement of the accuracy of assessing the technical condition of vehicles, and optimization of service decisions. **Object of research** is the process of forecasting the technical condition of automobiles in service maintenance systems based on the analysis of large volumes of heterogeneous data. **Subject of research** is methods, models, and tools for forecasting the technical condition of automobiles based on Big Data through the integration of sensor, operational, and service data. **Research results.** The study addresses the problem of forecasting the technical condition of automobiles based on Big Data in order to improve the efficiency of service technologies in the automotive industry. A method is proposed that involves the integration of operational, sensor, and service data within a unified analytical framework for assessing the risk of failures. The method was implemented in Python in the Google Colab environment using the open SCANIA Component X dataset. In the course of the study, data preparation, elimination of target variable leakage, construction of a predictive model, adjustment of the classification threshold, and formation of service decisions were carried out. The obtained results confirmed the high quality of forecasting and the suitability of the proposed approach for use in tasks supporting vehicle maintenance.

**Keywords:** Big Data, technical condition forecasting, automotive industry, machine learning, data analysis, maintenance, vehicles, classification, binary classifier, Google Colab, Python.