

Міністерство освіти і науки України
Львівський національний університет природокористування
Факультет механіки, енергетики та інформаційних технологій
Кафедра інформаційних технологій

КВАЛІФІКАЦІЙНА РОБОТА

другого (магістерського) рівня вищої освіти

на тему:

«Обґрунтування моделі машинного навчання для ідентифікації
пошкоджень дорожнього покриття»

Виконав: здобувач групи Іт-61
спеціальності 126
«Інформаційні системи та технології»

Сулятицький М.І.

Керівник:

Запорожцев С.Ю.

Рецензент:

Семерак В.М.

ЛЬВІВ-2024

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ЛЬВІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ПРИРОДОКОРИСТУВАННЯ
ФАКУЛЬТЕТ МЕХАНІКИ, ЕНЕРГЕТИКИ ТА ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ
КАФЕДРА ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ

Другий (магістерський) рівень вищої освіти
Спеціальність 126 – „Інформаційні системи та технології”

“ЗАТВЕРДЖУЮ”

Завідувач кафедри _____
д.т.н., проф. А.М. Тригуба
“ _____ ” _____ 202_ р.

ЗАВДАННЯ

на кваліфікаційну роботу здобувачу

Сулятицький Максим Іванович

1. Тема роботи: Обґрунтування моделі машинного навчання для ідентифікації пошкоджень дорожнього покриття

Керівник роботи Запорожцев Сергій Юрійович, к.т.н., доцент.

Затверджені наказом по університету від 12 вересня 2024 року № 616/к-с

2. Строк подання здобувачем роботи 05.12.2024 р.

3. Вихідні дані до роботи: Загальні відомості про методи оцінки якості та діагностики автомобільних доріг, варіанти підходів до машинного навчання та інструментарію для вирішення задачі

4. Зміст розрахунково-пояснювальної записки:

Вступ

1. Аналіз стану питання та постановка завдання. Сучасні тенденції та підходи до оцінки автомобільних доріг. Якість автомобільних доріг. Методи діагностики пошкоджень дорожнього полотна. Постановка задачі дослідження.

2. Обґрунтування та вибір інструментарію вирішення задачі. Аналіз динаміки автомобіля під впливом дорожніх дефектів. Застосування машинного навчання для моніторингу стану доріг. Вибір технічних засобів реалізації системи. Вибір мови програмування та середовища розробки.

3. Результати вирішення задачі. Отримання даних та їх первинна підготовка. Підготовка даних для машинного навчання. Навчання моделі та отримання результатів. Розробка загальної структури системи.

4. Охорона праці та безпека у надзвичайних ситуаціях.

5. Визначення ефективності інформаційної системи.

Висновки та пропозиції.

Список використаних джерел.

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень): _____
Мета роботи та аналіз предметної області. Сенсорні платформи, доступні для збору даних. Моделювання та вимірювання прискорень. Інструментарій та дані. Підготовка даних. Машинне навчання. Архітектура Big Data системи. Варіант архітектури потоку даних. Варіант SQL бази даних. Висновки

6. Консультанти з розділів:

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
1, 2, 3, 5	<i>Запорожцев С.Ю., доцент кафедри інформаційних технологій</i>		
4	<i>Городецький І.М., доцент кафедри управління проектами та безпеки виробництва</i>		

7. Дата видачі завдання 12 вересня 2024 р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів кваліфікаційної роботи	Строк виконання етапів роботи	Примітка
1	<i>Написання першого розділу</i>	12.09 - 25.09.24	
2	<i>Виконання другого розділу та аркушів ілюстраційного матеріалу до нього</i>	26.09 - 10.09.24	
3.	<i>Виконання третього розділу та аркушів ілюстраційного матеріалу до нього</i>	11.10 - 25.10.24	
4.	<i>Написання розділу «Охорона праці та безпека у надзвичайних ситуаціях»</i>	26.10 - 5.11.24	
5.	<i>Написання розділу «Визначення ефективності...»</i>	6.11 - 15.11.24	
6.	<i>Завершення оформлення розрахунково-пояснювальної записки та аркушів ілюстраційного матеріалу</i>	16.11 - 29.11.24	
7.	<i>Завершення роботи в цілому</i>	30.11 - 5.12.24	

Студент _____ Сулятицький М.І.
 (підпис)

Керівник роботи _____ Запорожцев С.Ю.
 (підпис)

РЕФЕРАТ

УДК 004.8 : 625.76

Обґрунтування моделі машинного навчання для ідентифікації пошкоджень дорожнього покриття

Сулятицький М.І. Кафедра ІТ – Дубляни, Львівський НУП, 2024.

Кваліфікаційна робота: 72 с. текст. част., 37 рис., 3 табл., 11 арк. ілюстраційного матеріалу, 35 джерел.

Об'єкт дослідження – автоматизація процесу діагностики стану дорожнього покриття.

Мета роботи – розробка моделі машинного навчання для оцінки стану дорожнього покриття за показниками, зібраними з множини мобільних пристроїв-акселерометрів.

Проведено аналіз предметної області, визначені фактори, які впливають на якість дороги та сформульовані задачі роботи. Розглянуті методи діагностики стану автомобільних доріг. Розглянуті математичні моделі, які використовуються в процесі моделювання вібрацій автомобіля під час руху. Визначено клас задачі, обрані мова програмування, середовище розробки, бібліотеки машинного навчання та джерела даних. Розроблена модель машинного навчання, проведено моделювання, отримані результати обчислень та зроблена їх оцінка. Розглянуті варіанти архітектурних рішень по системі, потокам даних та структурі бази даних. Описані питання охорони праці та визначення економічної ефективності отриманих рішень.

Ключові слова: діагностика пошкоджень дорожнього покриття, вимірювання прискорень, машинне навчання, задача кластеризації.

ABSTARCT

UDC 004.8 : 625.76

Rationale for a machine learning model for identifying road surface damage
Sulyatitsky M.I. Department of IT – Dublyany, Lviv NEU, 2024.

Qualification work: 72 p. text, 37 pict., 3 tabl., 11 p. illustrative material, 35 sources.

The object of the research is the automation of the process of diagnosing the condition of the road surface.

The purpose of the work is to develop a machine learning model for assessing the condition of the road surface according to indicators collected from a set of mobile accelerometer devices.

The analysis of the subject area was carried out, the factors that affect the quality of the road were identified and the tasks of the work were formulated. Methods for diagnosing the condition of highways are considered. The mathematical models used in the process of modeling the vibrations of the car while driving are considered. The class of the problem, the selected programming language, development environment, machine learning libraries and data sources are defined. A machine learning model has been developed, simulations have been carried out, the results of calculations have been obtained and their evaluation has been made. Options for architectural solutions for the system, data flows and database structure are considered. The issues of labor protection and determination of the economic efficiency of the solutions obtained are described.

Keywords: diagnostics of road surface damage, acceleration measurement, machine learning, clustering problem.

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

- ML – Machine Learning, машинне навчання;
- AI – Artificial Intelligence, штучний інтелект;
- GPS – Global Positioning System, глобальна система позиціонування;
- ДТП – дорожньо-транспортна пригода;
- FFT – Fast Fourier Transform, швидке перетворення Фур'є;
- CNN – Convolutional Neural Networks, згорткові нейронні мережі;
- IMU – Inertial Measurement Unit, інерціальні вимірювальні одиниці;
- SVM – Support Vector Machine, метод опорних векторів;
- RNN - Recurrent Neural Networks, рекурентні нейронні мережі;
- ОС - операційна система;
- CPU - Central Processing Unit, центральний процесор;
- GPU - Graphics Processing Unit, графічний процесор;
- TPU - Tensor Processing Unit, тензорний блок обробки;
- BDSA - Big Data Systems Architecture, архітектура систем великих даних;
- HDFS - розподілена файлова система Hadoop;
- SQL - Structured Query Language, мова структурованих запитів;
- API - Application Programming Interface, прикладний програмний інтерфейс;
- СУБД – система управління базами даних.

ЗМІСТ

ВСТУП	8
1. АНАЛІЗ СТАНУ ПИТАННЯ ТА ПОСТАНОВКА ЗАВДАННЯ	10
1.1. Сучасні тенденції та підходи до оцінки автомобільних доріг	10
1.2. Якість автомобільних доріг	14
1.3. Методи діагностики пошкоджень дорожнього полотна	20
Постановка задачі дослідження	23
2. ОБҐРУНТУВАННЯ ТА ВИБІР ІНСТРУМЕНТАРІЮ ВИРІШЕННЯ ЗАДАЧІ	24
2.1. Аналіз динаміки автомобіля під впливом дорожніх дефектів	24
2.2. Застосування машинного навчання для моніторингу стану доріг	28
2.3. Вибір технічних засобів реалізації системи	31
2.4. Вибір мови програмування та середовища розробки	35
Висновки до розділу	39
3. РЕЗУЛЬТАТИ ВИРІШЕННЯ ЗАДАЧІ	40
3.1. Отримання даних та їх первинна підготовка	40
3.2. Підготовка даних для машинного навчання	48
3.3. Навчання моделі та отримання результатів	52
3.4. Розробка загальної структури системи	57
Висновки до розділу	61
4. ОХОРОНА ПРАЦІ ТА БЕЗПЕКА У НАДЗВИЧАЙНИХ СИТУАЦІЯХ	62
5. ВИЗНАЧЕННЯ ЕФЕКТИВНОСТІ ІНФОРМАЦІЙНОЇ СИСТЕМИ	66
ВИСНОВКИ ТА ПРОПОЗИЦІЇ	69
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	70

ВСТУП

Дорожня інфраструктура є одним із найважливіших елементів економічного розвитку та соціального благополуччя сучасного суспільства. Стан дорожнього покриття безпосередньо впливає на безпеку дорожнього руху, ефективність транспортних перевезень та загальний рівень комфорту користувачів. Несправності та пошкодження дорожнього полотна, такі як тріщини, вибоїни, деформації, є причинами підвищення аварійності, зниження експлуатаційного ресурсу транспортних засобів, збільшення витрат палива та зростання витрат на ремонт доріг. Тому ідентифікація пошкоджень дорожнього покриття є одним із ключових завдань у забезпеченні сталого розвитку транспортної інфраструктури.

Актуальність теми роботи зумовлена необхідністю підвищення точності, швидкості та об'єктивності діагностики стану дорожнього полотна. Традиційні методи оцінки, які базуються на візуальних оглядах або ручних вимірюваннях, є трудомісткими, залежними від суб'єктивної оцінки фахівця та не завжди дозволяють своєчасно виявляти дефекти на великих ділянках доріг. У зв'язку з цим виникає потреба у впровадженні автоматизованих підходів до аналізу стану дорожнього покриття, які ґрунтуються на сучасних інформаційних технологіях.

Методи машинного навчання відкривають широкі перспективи для розв'язання цього завдання. Завдяки здатності аналізувати великі обсяги даних, виявляти приховані закономірності та адаптуватися до нових умов, алгоритми машинного навчання дозволяють створювати моделі, які автоматично ідентифікують пошкодження дорожнього покриття з високою точністю. Використання мобільних пристроїв, дронів, систем відеоспостереження та інших засобів збору даних у поєднанні з методами машинного навчання значно розширює можливості моніторингу доріг, роблячи цей процес ефективнішим і менш затратним.

Моделі машинного навчання для ідентифікації пошкоджень дорожнього покриття дозволяють не лише забезпечити якісне діагностування, але й сприяють ефективному плануванню ремонтних робіт, оптимізації витрат на обслуговування доріг та зниженню ризиків для учасників дорожнього руху. Крім того, такі моделі можуть бути інтегровані в системи керування дорожньою інфраструктурою та слугувати основою для створення інтелектуальних транспортних систем.

У контексті глобальних трендів цифровізації та автоматизації управлінських процесів, дослідження, спрямовані на вдосконалення методів оцінки стану дорожнього полотна, є особливо важливими. Зростання обсягів даних, доступних для аналізу, та розвиток обчислювальних потужностей створюють передумови для впровадження новітніх технологій у сферу дорожнього господарства. Це, у свою чергу, сприяє підвищенню конкурентоспроможності економіки та зростанню рівня життя населення.

Таким чином, обрана тема дослідження є актуальною, оскільки спрямована на вирішення важливого прикладного завдання – забезпечення ефективної діагностики пошкоджень дорожнього покриття шляхом використання сучасних технологій машинного навчання. Це дослідження має на меті зробити вагомий внесок у розвиток інноваційних підходів до управління дорожньою інфраструктурою, сприяючи підвищенню її якості, надійності та безпечності.

РОЗДІЛ 1

АНАЛІЗ СТАНУ ПИТАННЯ ТА ПОСТАНОВКА ЗАВДАННЯ

1.1. Сучасні тенденції та підходи до оцінки автомобільних доріг

Розвиток транспортної інфраструктури є основою економічного та соціального розвитку будь-якої країни. [1-3] В умовах сучасної економіки, де наголос робиться на інноваціях, автомобільні дороги стають критично важливими об'єктами, від стану яких залежить безпека, швидкість, комфорт перевезень та економічна ефективність транспорту. З огляду на це, оцінка ефективності автомобільних доріг виходить на перший план. Цей процес передбачає не лише аналіз технічних параметрів дорожнього полотна, але й облік факторів, пов'язаних з експлуатацією, довговічністю і відповідністю нормативним вимогам.

В останні роки спостерігається активне впровадження сучасних методів діагностики стану автомобільних доріг. [4, 5] Це зумовлено зростанням вимог до якості дорожнього полотна внаслідок інтенсифікації трафіку, збільшення навантажень на дорожнє покриття та необхідності економічно обґрунтованого управління інфраструктурою. Одним із ключових аспектів є оптимізація витрат на будівництво, ремонт та утримання доріг шляхом інтеграції технологій аналізу великих даних та моделей прогнозування. [6]

Проблеми оцінки ефективності автомобільних доріг виникають через використання застарілих стандартів та методів. Наприклад, часто застосовуються технічні показники, які не відображають реальних умов експлуатації. Крім того, для оцінки стану доріг необхідна точна діагностика таких параметрів, як рівність, зчеплення, міцність дорожнього полотна та інші характеристики, які безпосередньо впливають на безпеку дорожнього руху.

Одним із головних критеріїв ефективності автомобільних доріг є забезпечення високого рівня транспортно-експлуатаційних показників. [7, 8] Крім того, до таких критеріїв належать:

- рівність дорожнього покриття, що впливає на комфорт руху та витрати пального;
- коефіцієнт зчеплення коліс з дорогою, який визначає безпеку пересування в різних умовах;
- стійкість дорожнього полотна до деформацій, особливо в умовах підвищених температур або значних навантажень;
- надійність і довговічність конструкцій, що дозволяє мінімізувати частоту та обсяги ремонтних робіт.

Одним з основних методів є розрахунок експлуатаційного коефіцієнта, що враховує відношення максимальної фактичної швидкості до нормативної швидкості для певної категорії дороги. Цей коефіцієнт дозволяє комплексно оцінити стан дороги з точки зору її придатності для експлуатації.

Діагностика стану доріг [9, 10] є основою для забезпечення надійного функціонування дорожньої інфраструктури, безпеки руху та своєчасного планування ремонтних заходів. Основними завданнями діагностики є такі.

1. Організація моніторингу експлуатаційного стану доріг. Постійний контроль стану дорожнього покриття, елементів дорожньої інфраструктури та їх відповідність нормативним вимогам забезпечує виявлення проблемних ділянок на ранніх етапах, що дозволяє уникнути масштабних ремонтів і значних фінансових витрат.

2. Систематизація інформації про стан доріг. Збір, обробка та узагальнення даних з різних джерел, зокрема проектної та робочої документації, а також дорожніх баз даних, дозволяють створити єдину інформаційну платформу для аналізу поточного стану доріг і прогнозування їх змін у майбутньому.

3. Детальний аналіз проблемних ділянок. Проведення комплексної оцінки стану елементів дорожніх споруд у зонах із вираженою деградацією, таких як колійність, тріщини або зниження міцності дорожнього полотна. Виявлення конкретних причин дефектів сприяє розробці ефективних заходів з відновлення.

4. Використання сучасних приладів та методів діагностики. Впровадження інноваційних технологій, таких як лазерні сканери, ультразвукові трекери, георадарні системи, дрони та інші інструменти, дозволяє отримувати точні дані про стан дорожніх конструкцій. Це забезпечує підвищення якості діагностики та зменшення впливу людського фактора.

5. Оцінка експлуатаційного стану дорожнього покриття. Визначення техніко-експлуатаційних характеристик доріг, таких як міцність, рівність, коефіцієнт зчеплення та довговічність. Це дозволяє встановити рівень зношеності покриття, його відповідність транспортним навантаженням та погодним умовам.

6. Розрахунок залишкового ресурсу дорожніх конструкцій. Виконання аналізу залишкового ресурсу кожного елемента дорожнього одягу, включаючи ґрунти робочого шару та матеріали покриття. Це дає змогу визначити час, протягом якого дорога зможе функціонувати без суттєвих дефектів.

7. Оцінка впливу транспортного потоку. Аналіз інтенсивності руху, складу транспортного потоку та його динаміки дозволяє врахувати навантаження, яке впливає на стан дорожнього полотна, і спланувати заходи з його зміцнення.

8. Підготовка рекомендацій для ремонтних і відновлювальних робіт. Розробка на основі результатів діагностики конкретних технічних рішень для усунення виявлених дефектів, планування поточних і капітальних ремонтів.

9. Підвищення якості нормативної бази. Проведення апробації нових методів і обладнання для діагностики, а також підготовка пропозицій щодо вдосконалення нормативно-правової бази у сфері оцінки стану доріг. Це дозволить адаптувати стандарти до сучасних вимог і технологій.

10. Організація ефективного управління дорожньою інфраструктурою. Використання результатів діагностики для стратегічного планування, оптимізації витрат на утримання доріг і підвищення рівня обслуговування учасників дорожнього руху.

Таким чином, завдання діагностики не обмежуються лише фіксацією поточного стану доріг. Вони охоплюють широкий спектр заходів, спрямованих

на підвищення ефективності управління дорожньою мережею, забезпечення безпеки руху та збереження інфраструктури в довгостроковій перспективі.

Сучасна оцінка ефективності доріг дедалі частіше базується на використанні методів моделювання. [11-14] Зокрема, застосовуються алгоритми машинного навчання для аналізу великих обсягів даних про стан доріг, трафік та погодні умови. Ці алгоритми дозволяють:

- прогнозувати зношення дорожнього полотна на основі історичних даних;
- оптимізувати планування ремонтних робіт залежно від інтенсивності використання дороги;
- визначати критичні ділянки інфраструктури, які потребують першочергової уваги.

Такі моделі враховують не лише технічні параметри, але й економічні аспекти, дозволяючи розробити довгострокові стратегії управління дорожньою мережею.

Сучасні підходи до оцінки доріг вимагають удосконалення нормативної бази. [15,16] Багато стандартів досі базуються на застарілих методах, які не враховують можливості сучасних технологій. Необхідно адаптувати нормативи до використання новітніх приладів для визначення рівності, зчеплення та інших характеристик дорожнього полотна.

Паралельно з цим впровадження технологій стандартизації експлуатаційних показників дозволить більш ефективно управляти якістю доріг. Наприклад, щорічна оцінка стану доріг за результатами моніторингу дозволяє отримувати актуальні дані про ефективність експлуатації та приймати своєчасні рішення про ремонт.

Перспективним є використання сенсорних мереж, вбудованих у дорожнє покриття, які в реальному часі можуть передавати інформацію про навантаження, зношеність і температурні зміни. Це дозволяє оперативно реагувати на критичні зміни в стані доріг. Крім того, розвиток технологій лазерного сканування та 3D-моделювання дає змогу створювати детальні

цифрові копії дорожньої інфраструктури, що значно полегшує її оцінку, планування ремонтів і реконструкції. [17-21]

Зростає актуальність впровадження автоматизованих систем управління дорожньою інфраструктурою, які інтегруються з транспортними системами, забезпечуючи ефективну взаємодію між транспортними потоками, станом доріг і погодними умовами. Особливу увагу приділяють розробці методів прогнозування залишкового ресурсу дорожніх конструкцій, що дозволяє планувати ремонтно-відновлювальні роботи заздалегідь, уникаючи аварійних ситуацій.

У перспективі значного розвитку набуває використання технологій «розумних доріг», які включають інтеграцію датчиків, зв'язку з транспортними засобами та енергозберігаючих рішень. Такі системи не тільки моніторять стан доріг, але й сприяють зменшенню витрат енергії, оптимізації транспортних потоків і підвищенню загальної безпеки дорожнього руху.

Іншим важливим напрямом є розвиток нормативної бази та стандартизація підходів до діагностики стану доріг, адаптація до сучасних технологій та європейських стандартів. Залучення міждисциплінарних досліджень, які охоплюють будівельні технології, матеріалознавство, інформаційні технології та економіку, дозволяє створювати ефективніші моделі управління дорожньою мережею.

Таким чином, розвиток діагностики стану доріг спрямований на підвищення точності, оперативності та ефективності процесів оцінки й управління, що сприятиме довговічності дорожньої інфраструктури, зменшенню витрат на її утримання та підвищенню безпеки дорожнього руху.

1.2. Якість автомобільних доріг

Якість доріг є критично важливим показником розвитку транспортної інфраструктури, що безпосередньо впливає на безпеку та комфорт транспортування. У цьому розділі ми розглянемо ключові аспекти визначення

поняття "якість автомобільних доріг", а також фактори, які на неї впливають. Окрім цього, будуть детально розглянуті методи дослідження та основні показники якості доріг, а також проблеми, з якими стикаються дослідники.

Якість доріг можна охарактеризувати як міру відповідності дорожнього покриття та інфраструктури вимогам безпеки, комфорту та ефективності транспортного руху. [22]

Основними аспектами якості доріг є кілька ключових характеристик, які визначають як безпеку, так і комфорт транспортування. Одним із важливих аспектів є стан дорожнього покриття. Якість покриття залежить від його рівності, цілісності та наявності дефектів, таких як тріщини, ями та вибоїни. Рівність покриття впливає на комфортність руху транспорту, оскільки нерівності можуть призвести до зниження швидкості і збільшення витрат пального, а також до підвищення ризику пошкодження автомобіля. Стан покриття також визначає наявність дефектів, які можуть бути результатом механічних пошкоджень або несприятливих погодних умов, що з часом призводять до зносу асфальту чи бетону.

Іншим важливим аспектом є шорсткість дорожнього покриття. Вона відіграє вирішальну роль у забезпеченні зчеплення між шинами автомобіля та дорогою. У разі недостатньої шорсткості знижується ефективність гальмування, особливо на мокрих або слизьких дорогах, що підвищує ризик ДТП. Надмірна шорсткість може призвести до підвищеного зносу шин і шуму під час руху, тому важливо забезпечити оптимальний рівень шорсткості для безпеки та комфорту водіїв.

Не менш важливим аспектом є інфраструктура дороги, до якої належать всі допоміжні елементи, що забезпечують безпеку та зручність руху. Це дорожні знаки, розмітка, освітлення, огорожі та відбійники. Ретельно продумана інфраструктура дозволяє знизити ймовірність аварій та покращити орієнтування водіїв, особливо у нічний час або в умовах обмеженої видимості. Крім того, добре організовані системи водовідведення, дренажу та дорожнього

покриття допомагають уникнути накопичення води на дорозі, що також підвищує безпеку транспортування.

Важливим фактором є безпека дорожнього руху, яка тісно пов'язана з якістю дороги. Безпечні дороги забезпечують не лише зручність, але й значно знижують рівень травматизму серед водіїв, пасажирів та пішоходів. Важливо враховувати параметри видимості дорожніх знаків, стан пішохідних переходів, наявність бар'єрів та розмітки. Система контролю за швидкістю, наявність камер спостереження та інших засобів зниження аварійності сприяють підвищенню безпеки на дорогах.

Якість доріг також визначається їх здатністю витримувати навантаження, яке створюють транспортні засоби. Це стосується не лише навантаження від автомобілів, але й від великих вантажівок і важких транспортних засобів. Важливим є врахування типів дорожнього покриття та міцності матеріалів, що використовуються при їх будівництві. Матеріали, які підбираються для будівництва дороги, мають бути стійкими до механічних впливів та погодних умов. Знос доріг може бути також пов'язаний з неякісними матеріалами чи неправильним проектуванням дорожнього покриття, що призводить до швидшого руйнування.

Також важливою є здатність дороги до адаптації до змінних умов навколишнього середовища. Погодні умови, такі як дощ, сніг, заморозки, можуть значно вплинути на стан дороги. Тому ефективна система водовідведення та належний температурний режим будівництва є важливими складовими якості дороги. Погіршення погодних умов може викликати замерзання води в тріщинах та інших дефектах покриття, що призводить до їх подальшого розширення.

Нарешті, один із важливих аспектів якості доріг — це їх здатність забезпечувати сталість і тривалий термін служби. Тривала експлуатація доріг без необхідності постійного ремонту є показником їх високої якості. Кожен з цих аспектів взаємопов'язаний з іншими і взаємно доповнює один одного, що

створює комплексне уявлення про загальний стан дороги та її здатність забезпечити безпеку і комфорт транспортування.

Можна говорити про такі фактори, що впливають на якість доріг:

- міцність і довговічність дороги значною мірою залежать від товщини асфальтобетонного покриття, якості матеріалів, наявності армувальних систем і водовідведення;

- високий рівень трафіку прискорює знос дорожнього покриття, що призводить до тріщин і вибоїн;

- погодні фактори, такі як температура, вологість і опади, також мають значний вплив на стан доріг, адже екстремальні умови можуть спричинити деформацію або руйнування покриття;

- безперервне обслуговування, включаючи очищення від сміття і снігу, а також ремонт пошкоджень, допомагає зберегти якість дороги в належному стані;

- врахування геологічних, гідрологічних умов та використання передових технологій будівництва також є дуже важливим для створення високоякісної дороги.

При дослідженні якості доріг виникають численні проблеми та виклики, що пов'язані з технічними, економічними, екологічними та організаційними аспектами. Однією з основних проблем є складність оцінки стану дорожнього покриття. Якість доріг не можна оцінити лише за зовнішніми ознаками, адже значна частина дефектів може бути непомітною на перший погляд, але впливати на експлуатаційні характеристики дороги. Наприклад, тріщини або дефекти в підземних шарах дорожнього покриття можуть призвести до серйозних руйнувань, які стають очевидними лише через певний час. Тому дослідження якості часто потребує застосування спеціалізованих методів, таких як дефектоскопія, георадарні дослідження, аналіз зображень та інші високотехнологічні інструменти.

Ще однією значущою проблемою є відсутність єдиного стандарту або методики для визначення якості доріг, який би враховував усі можливі

фактори, що впливають на стан дорожнього покриття. Різноманітність методів оцінки якості, відрізняються залежно від регіону, типу дороги, матеріалів покриття та умов експлуатації. Це ускладнює порівняння результатів досліджень, що проводяться в різних місцях або з використанням різних методик.

Існує також проблема з точністю і оперативністю збору даних про стан доріг. Зазвичай для таких досліджень необхідно використовувати різноманітні сенсори, датчики та спеціалізовані технології для збору точних даних про стан дороги, але це потребує значних фінансових і трудових витрат. Інколи дорожні служби не мають змоги проводити часті перевірки доріг через обмежений бюджет або через потребу в спеціалізованому обладнанні, що призводить до зволікань з виявленням дефектів.

Ще одним важливим викликом є зміни погодних умов і сезонні коливання, які значно впливають на стан доріг. Вода, сніг, мороз, спека — усі ці фактори можуть призвести до швидкого руйнування дорожнього покриття, що потребує своєчасного моніторингу та коригування планів ремонту. Врахування погодних умов є важливою частиною при плануванні ремонтних робіт, але часто стає проблемою через відсутність належного аналізу кліматичних змін і прогнозів.

Великим викликом є також вплив людського фактора на стан доріг. Наприклад, перевантаження автомобілями, а особливо важким транспортом, призводить до значного зносу покриття. Також важливою є проблема неправильного або недбалого використання технологій під час будівництва або ремонту доріг, коли використовуються неякісні матеріали або не дотримуються технологічні норми.

Крім того, питання фінансування залишається однією з основних проблем. Якісні дослідження вимагають значних інвестицій у технології, що не завжди можуть бути забезпечені з бюджету місцевих органів влади. Часто дорожні служби не мають необхідних коштів на регулярне проведення комплексних досліджень, що спричиняє затримки в діагностиці стану доріг і, як наслідок, учасників дорожнього руху стикаються з підвищеними ризиками аварійності.

Одним з викликів є і проблема актуальності зібраних даних. Існує потреба в тому, щоб зібрані дані були актуальними і відображали реальний стан дороги на момент їх збору. У випадку, якщо дослідження виконуються з затримкою або без належного контролю, це може призвести до невірних висновків і прийняття неефективних рішень щодо ремонту та утримання доріг.

Загалом, дослідження якості доріг стикається з численними складнощами, які потребують комплексного підходу і врахування різних факторів. Це вимагає постійного вдосконалення методик оцінки, залучення новітніх технологій, а також забезпечення належного фінансування та планування для підтримки високого рівня безпеки і комфортності дорожнього руху. Зведені показники по якості доріг можна дати в вигляді таблиці (приклад наведений в табл. 1.1).

Таблиця 1.1 - Показники якості автомобільних доріг

Фактори	Опис	Методи дослідження	Показники якості
Дорожнє покриття	Визначається станом асфальту або іншого матеріалу, наявністю тріщин, вибоїн, зносу. Важливе значення для комфорту та безпеки руху.	Візуальний огляд, профілювання, вимірювання висоти вибоїн	Глибина вибоїн, кількість тріщин, рівність покриття, коефіцієнт зносу.
Геометрія дороги	Визначається формою дороги, її шириною, поворотами, ухилами. Погіршення геометрії впливає на маневреність і безпеку.	Вимірювання геометричних параметрів, використання GPS і лазерних технологій	Ширина дороги, радіус поворотів, ухили, рівень безпеки на складних ділянках.
Стан розмітки	Оцінка видимості та чіткості дорожньої розмітки, що має важливе значення для безпеки руху та орієнтації водіїв.	Візуальний огляд, вимірювання ширини ліній, використання камер	Якість розмітки, яскравість фарб, чіткість ліній, наявність дефектів (розмиті, стерті лінії).
Швидкісний режим	Визначення допустимої швидкості руху для конкретної ділянки дороги, відповідність знаків швидкості. Невідповідність може призвести до збільшення аварійності.	Оцінка дорожніх знаків, аналіз аварійності, вимірювання швидкості	Частота порушень швидкості, середня швидкість на ділянці, відповідність швидкості до дорожніх умов.
Інфраструктура	Наявність додаткових елементів: освітлення, зупинки, пішохідні переходи, дорожні знаки. Їх стан значно впливає на безпеку руху.	Оцінка стану дорожніх елементів, інспекція	Якість освітлення, наявність пішохідних переходів, стан зупинок, дорожні знаки.

1.3. Методи діагностики пошкоджень дорожнього полотна

Методи діагностики стану дорожнього полотна мають велике значення для забезпечення безпеки дорожнього руху та підвищення якості дорожнього покриття. Проблеми, що виникають при оцінці стану доріг, можуть бути зумовлені як відсутністю достатньої точності оцінок, так і технічними складнощами при використанні тих чи інших методів. Для вирішення цих проблем важливо мати ефективні інструменти та методики, що дозволяють об'єктивно оцінювати стан дорожнього покриття, усувати дефекти та прогнозувати ймовірність аварійних ситуацій.

Дорожньо-транспортні пригоди (ДТП) часто зосереджуються на конкретних ділянках доріг, де є суттєві дефекти або невідповідність нормативам. Це вимагає від фахівців з безпеки руху своєчасного виявлення таких ділянок та визначення можливих причин аварій. У свою чергу, ефективність використання коштів на ремонт і підтримку доріг безпосередньо залежить від того, наскільки об'єктивно й точно оцінено стан доріг. Для цього застосовують різноманітні методи дослідження.

Методи діагностики стану дорожнього полотна можна поділити на три основні групи: інструментальні, візуальні та змішані (рис. 1.1). Кожна з цих груп має свої переваги та обмеження, що робить їх застосування в залежності від специфіки завдання і вимог. [23]



Рисунок 1.1 - Методи діагностики стану дорожнього покриття

Інструментальні методи дозволяють отримувати точні дані про стан дорожнього полотна за допомогою різних вимірювальних приладів. Вони можуть бути як прямими, так і непрямыми.

Прямі методи передбачають використання вимірювальних приладів, які дають точні показники, що безпосередньо відображають фізичні характеристики дорожнього покриття. Наприклад, однією з найбільш відомих інструментальних методик є застосування дорожніх рейок. Цей метод дозволяє визначати нерівності поверхні дорожнього покриття, а також вимірювати поздовжні та поперечні ухили. Дорожня рейка — це простий, але ефективний інструмент, що дозволяє з мінімальними затратами часу здійснювати вимірювання на місці. Водночас цей метод є досить трудомістким і вимагає присутності фахівця, що може бути обмеженням у випадку великих обсягів робіт.

Непрямі методи є більш автоматизованими та дозволяють обробляти більший обсяг даних. Одним із прикладів є використання поштовхомірів різних конструкцій. Ці прилади вимірюють коливання кузова автомобіля щодо підвіски при русі по дорожньому покриттю. Поштовхомір дозволяє за один день дослідити великі ділянки доріг (200-300 км), однак точність вимірювань може знижуватись через вплив факторів, таких як маса транспортного засобу, швидкість руху, температура навколишнього середовища та інші.

Візуальні методи є спрощеною формою оцінки стану доріг, але вони також мають свої переваги. Вони зазвичай застосовуються на першому етапі діагностики або для оперативного виявлення дефектів під час експлуатації доріг. Візуальний огляд може бути здійснений безпосередньо експертами або автоматизованими системами.

Перший спосіб візуальної оцінки передбачає огляд дороги спеціалістами, які рухаються на автомобілі зі швидкістю 10-20 км/год, фіксуючи всі дефекти та заносючи їх у журнал або електронні носії даних. За допомогою цього методу можна виявити пошкодження покриття, дефекти дорожніх споруд та інші

ознаки, що свідчать про погіршення стану дороги. Оцінка координат дефектів здійснюється через спідометр або за допомогою GPS-координат.

Другий спосіб передбачає використання відеокамер для зйомки ділянки дороги з автомобіля. Знятий матеріал переглядається і аналізується, після чого на екрані визначаються всі дефекти, які потребують ремонту чи ремонтування. Цей метод дозволяє зібрати великі обсяги даних для подальшого аналізу та допомагає значно знизити людський фактор в оцінці стану доріг.

Змішані методи діагностики поєднують інструментальні та візуальні методи для отримання більш детальних даних. Наприклад, лазерний метод дозволяє автоматизовано здійснювати обстеження дороги, фіксуючи точні показники без втручання в рух транспортних засобів. Лазерна система сканує дорогу під кутом 360 градусів, збираючи хмари точок, що потім обробляються комп'ютерними програмами. Цей метод дозволяє здійснювати сканування доріг без перерви в русі транспорту та без необхідності великого часу на проведення обстежень.

Процес сканування передбачає два основних етапи: перший — це збір даних з допомогою лазерних датчиків, другий — це обробка отриманих точок в спеціалізованих програмних комплексах. В результаті обробки отриманих даних можна отримати тривимірну модель ділянки дороги, що дає можливість точно оцінити стан покриття та визначити дефекти, що потребують ремонту.

Також використовується метод відеоаналізу, при якому система відеокомп'ютерного сканування здійснює автоматичну обробку знятого матеріалу для визначення дефектів дорожнього покриття та їх класифікації. Такий метод дозволяє не тільки виявити дефекти, але й оцінити їх рівень небезпеки для безпеки дорожнього руху, що є важливим для планування ремонтних робіт та профілактичних заходів.

Таким чином, застосування сучасних методів діагностики стану дорожнього полотна дозволяє отримати об'єктивні дані про стан доріг, що сприяє ефективному управлінню дорожньою інфраструктурою та забезпечує підвищення безпеки дорожнього руху.

Постановка задачі дослідження

Підсумовуючи описане вище, можна визначити завдання, які треба вирішити в ході даної роботи:

- провести аналіз предметної області, визначити фактори, які впливають на якість дороги, а також зробити огляд методів та обмежень діагностики стану автомобільних доріг;

- розглянути математичні моделі, які використовуються в процесі моделювання вібрацій автомобіля під час руху та визначити, яким чином для обробки даних використовувати методи штучного інтелекту;

- визначити клас задачі та намітити шляхи її вирішення;

- обрати мову програмування, інструментарій та джерела даних;

- розробити модель машинного навчання, провести моделювання, отримати результати обчислень та зробити їх оцінку;

- розглянути варіанти архітектурних рішень по системі, потокам даних та структурі бази даних.

Виконання цих завдань вирішить загальну задачу, яка стоїть в даній роботі.

РОЗДІЛ 2

ОБҐРУНТУВАННЯ ТА ВИБІР ІНСТРУМЕНТАРІЮ ВИРІШЕННЯ ЗАДАЧІ

2.1. Аналіз динаміки автомобіля під впливом дорожніх дефектів

Стан дорожнього покриття суттєво впливає на безпеку руху, комфорт водія та збереження транспортного засобу. Дефекти, такі як тріщини, вибоїни, хвилі чи інші нерівності, створюють значні динамічні навантаження, які можуть викликати небезпечні коливання автомобіля, знижувати ефективність підвіски та підвищувати знос шин. Аналіз динамічної реакції автомобіля на такі дефекти дозволяє не тільки оцінити вплив дорожніх умов, але й розробляти методи їхнього моніторингу.

Автомобіль як об'єкт дослідження може бути поданий у вигляді спрощеної моделі з використанням масово-пружних систем. На рисунку 2.1 схематично наведена модель чверті автомобіля, яка вважається класичною. [24-26] Основна увага приділяється двом ключовим аспектам: вертикальним коливанням кузова та колеса автомобіля, а також передаванню вібрацій на водія.

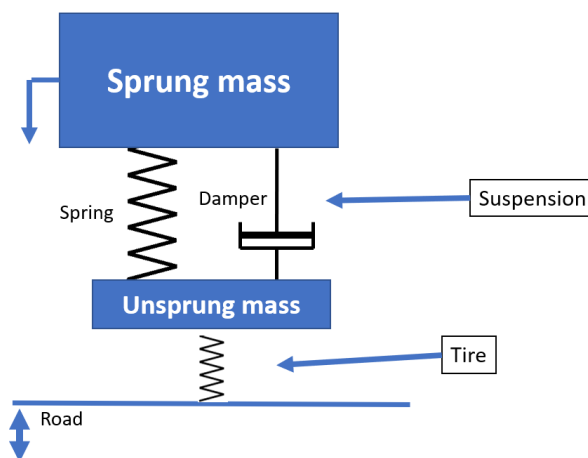


Рисунок 2.1 - Математична модель чверті автомобіля

Система "автомобіль-дорога" моделюється як двомасова система, яка описується такими параметрами:

- m_1 — маса кузова,
- m_2 — маса колеса (незалежно від підвіски),
- k — жорсткість підвіски,
- c — коефіцієнт демпфування,
- $z(t)$ — функція нерівностей дорожнього полотна.

Рух системи описується диференціальними рівняннями:

$$\begin{aligned} m_1 x_1'' + c(x_1' - x_2') + k(x_1 - x_2) &= 0, \\ m_2 x_2'' + c(x_2' - x_1') + k(x_2 - x_1) &= F_d, \end{aligned}$$

де x_1 і x_2 — зміщення кузова та колеса відповідно, а F_d — зовнішня сила, що моделює контакт із дорожньою нерівністю.

Наприклад, якщо нерівність дороги задається у вигляді синусоїдальної функції:

$$z(t) = A \sin(\omega t),$$

де A — амплітуда нерівності, ω — циклічна частота, тоді сила F_d може бути виражена як функція деформації та швидкості взаємодії колеса з поверхнею.

Розв'язання рівнянь руху для різних параметрів системи дозволяє дослідити вплив характеристик дорожнього полотна на динаміку автомобіля. Для цього використовують чисельні методи, наприклад, метод Рунге-Кутти, які забезпечують точне інтегрування рівнянь руху.

В рамках дослідження можливі сценарії з різними типами дефектів:

Одинична нерівність (яма):

$$z(t) = \begin{cases} A, & t_1 \leq t \leq t_2, \\ 0, & \text{інакше.} \end{cases}$$

Періодичні нерівності (хвилі):

$$z(t) = A \sin(\omega t).$$

Випадкові нерівності (шум):

$$z(t) = A \cdot \text{rand}(t),$$

де $\text{rand}(t)$ — випадкова функція.

Аналіз динаміки автомобіля може бути суттєво деталізований за рахунок урахування складніших форм нерівностей дорожнього полотна. Наприклад, реальні нерівності можуть бути представлені як трапециподібні профілі, які наближають дефекти у вигляді виступів або западин. В такому випадку формули будуть мати такий вигляд:

$$z(t) = \begin{cases} A \cdot \frac{t-t_1}{t_2-t_1}, & t_1 \leq t < t_2, \\ A \cdot \frac{t_3-t}{t_3-t_2}, & t_2 \leq t \leq t_3, \\ 0, & \text{інакше.} \end{cases}$$

Тут t_1 , t_2 , t_3 — точки початку, середини та кінця дефекту, A — амплітуда нерівності.

Випадкові нерівності можна моделювати як випадковий процес зі стандартним відхиленням σ та середнім значенням μ :

$$z(t) = \mu + \sigma \cdot \text{randn}(t),$$

де $\text{randn}(t)$ — функція, що генерує нормально розподілені випадкові значення.

Такі нерівності характерні для дорожнього полотна з тріщинами або іншими дефектами, що утворюють нерегулярний профіль. Для їх аналізу потрібен розширений підхід, наприклад, використання стохастичних диференціальних рівнянь:

$$m_1 x_1'' + c(x_1' - x_2') + k(x_1 - x_2) = -k_d z(t),$$

де k_d — коефіцієнт жорсткості контакту шини з дорожнім покриттям.

Крім того, у разі моделювання багатовісних транспортних засобів треба додатково враховувати зв'язок між осями.

Для оцінки впливу дефектів використовують також амплітудно-частотний аналіз. Частота природних коливань системи визначається як:

$$f_n = \frac{1}{2\pi} \sqrt{\frac{k}{m_1}}.$$

Якщо частота збурення ω наближається до f_n , в системі виникає резонанс, що супроводжується значним зростанням амплітуди коливань кузова. Це особливо небезпечно на високих швидкостях, коли енергія системи не встигає розсіюватися демпферами.

Динамічна реакція автомобіля також залежить від коефіцієнта демпфування ξ , який розраховується як:

$$\xi = \frac{c}{2\sqrt{km_1}}.$$

При малих значеннях ξ (недостатнє демпфування) автомобіль може зазнавати значних коливань, що негативно впливає на комфорт і безпеку.

Окрім визначення частоти природних коливань f_n , доцільно провести аналіз частотного спектру сигналів, що генеруються системою "автомобіль-дорога". Спектр сигналу нерівності можна отримати через швидке перетворення Фур'є (FFT):

$$Z(f) = \int_{-\infty}^{\infty} z(t)e^{-j2\pi ft} dt,$$

де $Z(f)$ — спектральна щільність потужності.

Для оцінки реакції автомобіля на вібрації використовується передавальна функція системи:

$$H(f) = \frac{X_1(f)}{Z(f)},$$

де $X_1(f)$ — спектр зміщення кузова. З її допомогою можна визначити резонансні частоти, на яких амплітуда зміщень досягає максимуму.

Крім того, для оцінки ефективності демпфування корисно врахувати логарифмічний декремент загасання:

$$\delta = \frac{1}{n} \ln \frac{x_1}{x_{n+1}},$$

де x_1 та x_{n+1} — амплітуди коливань на початку та після n -го циклу.

Як бачимо, дослідник при моделюванні поведінки автомобіля вимушений точно визначити його кількісні характеристики (різноманітні коефіцієнти, про які говорилось вище). Але в реальності немає можливості досліджувати кожен автомобіль в лабораторних умовах. Тому користуються збором та обробкою даних з великої кількості автомобілів, для чого застосовують методи машинного навчання.

2.2. Застосування машинного навчання для моніторингу стану доріг

Сучасні системи моніторингу активно використовують методи машинного навчання (ML), які дозволяють підвищити точність і ефективність виявлення дефектів. [27, 28] Розглянемо аспекти застосування машинного навчання для аналізу дорожнього покриття.

Одним із найпоширеніших підходів є аналіз дорожнього покриття за допомогою зображень, отриманих із камер, встановлених на автомобілях або безпілотниках. Методи обробки зображень у поєднанні з глибокими

нейронними мережами дозволяють ідентифікувати тріщини, ями, вибоїни та інші дефекти.

Основні кроки для цього полягають в наступному. Спочатку виконується попередня обробка зображень, яка включає фільтрацію шумів, нормалізацію освітлення та підвищення контрасту. Потім відбувається сегментація зображень, при цьому використовують алгоритми, такі як U-Net або Mask R-CNN, для виділення зон з дефектами. І нарешті, виконують етап класифікації, а саме застосування моделей, таких як ResNet або VGGNet, для визначення типу пошкодження.

Датчики вібрацій та інерціальні вимірювальні одиниці (IMU), встановлені на транспортних засобах, є ще одним джерелом даних для аналізу стану доріг. Машинне навчання дозволяє обробляти великі обсяги цих даних для прогнозування типу дефекту.

Наприклад, модель може використовувати сигнали прискорення для оцінки сили впливу нерівностей, а також частотний аналіз для виявлення характерних ознак пошкоджень через спектральні перетворення (наприклад, згадане вище перетворення Фур'є).

Математична модель сигналу вібрації може бути описана так:

$$x(t) = A \sin(2\pi ft + \phi),$$

де A — амплітуда, f — частота, ϕ — початкова фаза сигналу. Аналіз амплітуд і частотних компонентів допомагає визначити наявність та тип дефекту.

В залежності від типу даних та задачі, для моніторингу стану доріг використовують різні алгоритми машинного навчання.

Наприклад, алгоритм SVM (Support Vector Machine) добре працює на невеликих наборах даних, особливо коли класифікація дефектів має лінійний розподіл. Дерева рішень і Random Forest (Випадковий Ліс) ефективні для роботи з даними різної природи, включаючи зображення, сенсорні дані та сигнали вібрації. Глибокі нейронні мережі дозволяють виявляти складні

патерни в даних, зокрема для аналізу зображень та обробки великих масивів сенсорних даних.

У реальних умовах точність роботи ML-моделей залежить від багатьох факторів. Зокрема, зміна погодних умов, опади (дощ, сніг, туман) або недостатнє освітлення можуть погіршити якість зображень. Алгоритми з обробки зображень в такому випадку можуть застосовувати адаптивні методи, наприклад, зміну яскравості чи контрастності.

Сенсорні дані (вібрації) можуть мати значний шум, що вимагає попереднього фільтрування сигналів. Для цього можна використовувати, наприклад, фільтр Калмана для згладжування сенсорних сигналів або вейвлет-аналіз для виділення основних компонент.

Наприклад, для аналізу великих масивів даних доцільно використовувати модель

$$y = \sigma(WX + b),$$

де W — матриця ваг, X — вхідні дані, b — зміщення, σ — активаційна функція (наприклад, сигмоїда, ReLU або інша).

Глибокі нейронні мережі, такі як Convolutional Neural Networks (CNN), є оптимальними для обробки зображень дорожнього полотна. CNN автоматично виділяють ознаки дефектів (краї тріщин, форми вибоїн) через кілька рівнів згорток.

Математично згортковий шар описується так:

$$h_{i,j} = \sum_{m,n} x_{i+m,j+n} \cdot k_{m,n},$$

де $x_{i,j}$ — вхідні дані, $k_{m,n}$ — ядро згортки. Ця формула дозволяє знаходити локальні ознаки, наприклад, границі або текстурні зміни.

У випадках, коли необхідно працювати із часовими рядами (дані з акселерометрів чи IMU), доцільно використовувати рекурентні нейронні

мережі (RNN) або Long Short-Term Memory (LSTM) для аналізу часових залежностей.

Для впровадження ML-алгоритмів у дорожній моніторинг важливо враховувати такі аспекти, як: реальний час для забезпечення низької затримки обробки даних; енергоспоживання, що особливо важливе для автономних систем, таких як дрони; адаптацію, для чого моделі повинні враховувати різні погодні умови та типи доріг.

Незважаючи на високу ефективність моделей машинного навчання для ідентифікації дорожніх дефектів, існують певні виклики для їх впровадження. В першу чергу, це якість даних, а саме можлива недостатність міток у навчальних вибірках, що зазвичай може призводити до низької точності. Також проблемою є складність сучасних моделей, тому що великі нейронні мережі вимагають значних обчислювальних ресурсів.

Тим не менш, одним з перспективних напрямків може бути інтеграція даних із багатьох джерел (камери, сенсори, супутники) та розвиток самонавчальних моделей.

Остаточне рішення по вибору алгоритмів машинного навчання приймалось і описане в третьому розділі даної роботи після попереднього розвідувального аналізу тих даних, які були в наявності.

2.3. Вибір технічних засобів реалізації системи

Типовий метод моніторингу доріг покладається на систему збору даних, яка включає одновимірні дані часових рядів, двовимірні візуальні дані або тривимірні дані про глибину.

Таблиця 2.1 надає порівняльну оцінку датчиків наступного покоління, які розвинули весь процес збору даних для моніторингу стану доріг. [29] Зручне розміщення цих датчиків на землі та літальних платформах (тобто дронах) забезпечує їм перевагу з точки зору масштабованості та зручності використання.

Таблиця 2.1 - Порівняльна оцінка інтелектуальних датчиків

Характеристика	Камера	Лазер	Георадар	Тепловізор	Вібрація
Технології	2D зображення	3D побудова зображення за допомогою відображення	Радіохвилі для дослідження підземної поверхні; створює тривимірне зображення	Заснований на зміні температури навколишніх предметів за допомогою інфрачервоних хвиль	Показання акселерометрів, гіроскопів і GPS
Обробка	Складні алгоритми обробки зображень	Колекція 3D хмари точок	Потрібен збір глибинних зображень і даних моделювання	Збір теплових змін поверхні	Показання використовуються безпосередньо
Додаток реальному часу	Залежить від процесора	так	так	так	Не можна
Відчуття часу	При наближенні	При наближенні	При наближенні	При наближенні	Після події
Характер дефекту	Виходячи з форми та розміру	На основі 3D зображення	На основі 3D зображення	На основі теплових карт	Лише вздовж колісної траєкторії
Світлочутливість	Чутливий до рівня освітлення	Не чутливий до світлового впливу	Не чутливий до світлового впливу	Не чутливий до світла, але до температури поверхні	Жодного
Точність	Залежить від алгоритму	Високий	Високий	Високий	Багато помилок
Роздільна здатність	Від низького до високого	Висока	Залежить від частоти	Потребує вдосконалення	-
Час обробки	Швидкий збір і аналіз даних	Швидкий збір і аналіз даних	Збір даних автоматизований однак затримки при обробці	Швидкий збір і аналіз даних	Довгий, потрібна обробка даних
Вартість	Економна	Висока	Дуже дорого	Дуже дорого	Низька
Тип даних	2D, 3D	3D	3D	2D, 3D	1D

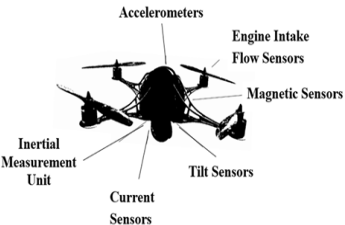

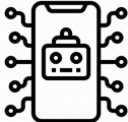
Недорогі датчики, такі як акселерометри, гіроскопи, магнітометри та GPS, часто використовуються для вимірювання руху, обертання, швидкості, орієнтації та розташування, які є одновимірними за своєю природою. Датчики, які зазвичай використовуються для отримання цифрових зображень, — це

високоякісні датчики RGB з високою роздільною здатністю, які полегшують отримання багатомасштабного низькорівневого та високорівневого виділення функцій. Однак ці зображення мають двовимірний характер і не можуть бути використані для характеристичного оцінювання ознак.

Глибина аналізу таких зображень дозволяє використовувати багаторазові параметричні атрибути, що полегшують дослідження внутрішніх характеристик і можуть бути отримані за допомогою тепловізійних датчиків, LiDAR, лазерних датчиків і георадара тощо.

Таблиця 2.2 ілюструє переваги та обмеження різних платформ, на яких зазвичай розгортаються датчики. [29]

Таблиця 2.2 - Сенсорні платформи, доступні для збору даних

Сенсорна платформа	Умовне зображення	Переваги	Обмеження
Безпілотні літальні апарати		<ul style="list-style-type: none"> - Великий кут зору. - Висока роздільна здатність. - Детальні дані. - Простота розгортання та доступність у небезпечних зонах. - Гнучкість для швидкої перевірки. 	<ul style="list-style-type: none"> - Обмеження корисного навантаження та пам'яті. - Законодавчі обмеження.
Спеціальні наземні транспортні засоби		<ul style="list-style-type: none"> - Наявність тривалого маршруту. - Масив датчиків. - Зображення високої роздільної здатності. - Можливість працювати в дуже щільній місцевості. 	<ul style="list-style-type: none"> - Малий кут огляду. - Менш економічно ефективний. - Висока залежність від робочої сили.
Смартфони		<ul style="list-style-type: none"> - Легка технологія. - Економічно вигідний. 	<ul style="list-style-type: none"> - Зображення з низькою роздільною здатністю. - Обмежено даними RGB.

Інтелектуальні методи вимірювання надають інформацію високої роздільної здатності. Швидше встановлення та гнучкість у розгортанні безконтактних датчиків на різних мобільних платформах збору даних

полегшують зручні способи перевірки доріг, долаючи обмеження простору та навколишнього середовища. Кожна платформа має свої переваги та обмеження та доповнює зручність використання одна одної.

Схематичне зображення датчиків і платформ збору даних показано на рис.2.1. Слід зазначити, що даний перелік не є остаточним.

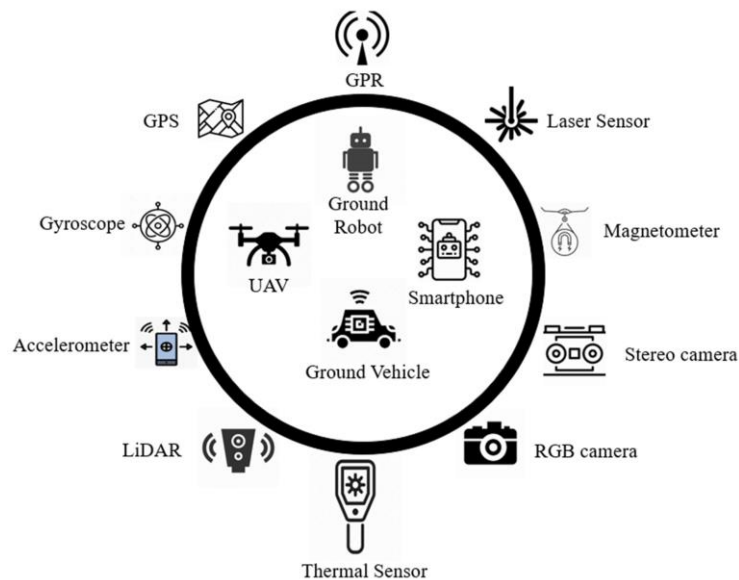


Рисунок 2.1 - Схематичне зображення датчиків наступного покоління та їх платформ

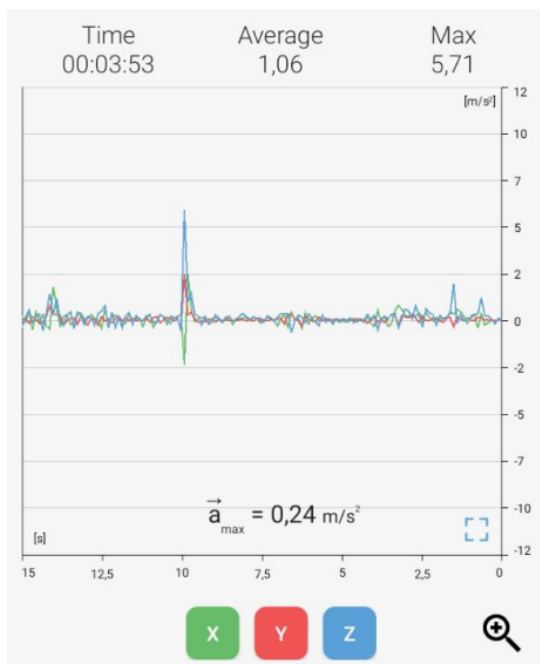
Основний недолік перспективних датчиків - їх висока ціна. Крім того, додатково треба врахувати ще вартість впровадження, масштабування та підтримки. Тому швидкість розгортання сучасних систем напряму залежить від необхідного фінансування.

В даній роботі було прийнято рішення використовувати прості датчики вимірювання прискорення, які є практично в кожному смартфоні.

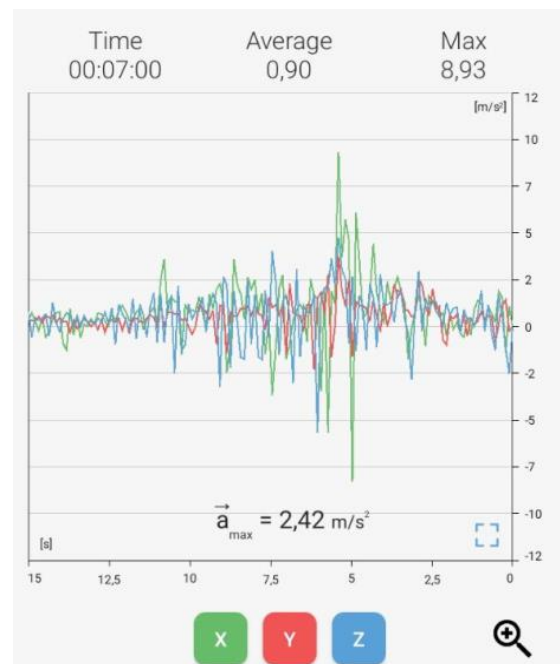
В якості програмних засобів вимірювання прискорень можуть бути застосовані практично будь-які рішення, які мають відповідні засоби обміну інформацією з іншими додатками. Прикладом такого рішення є додаток Vibrometer на смартфоні з ОС Android. [30]

Наведемо приклади роботи додатка. На рис.2.2(а) бачимо зафіксовані коливання при проїзді одинарного дефекту на трасі. Коливання, які показані на

рис.2.2(б) демонструють приклад прискорень корпусу автомобіля при проїзді залізничного переїзду.



а)



б)

Рисунок 2.2 - Приклади роботи додатку Accelerometer

Як бачимо, форма коливань досить складна, тому при обробці можна обмежитись обчисленням тільки амплітуди максимальних коливань в певний проміжок часу, наприклад, протягом кожної секунди.

2.4. Вибір мови програмування та середовища розробки

У сфері створення систем моніторингу стану доріг правильний вибір мови програмування та середовища розробки має важливе значення. Кожна мова програмування має свої переваги й обмеження, і їх варто враховувати під час реалізації проекту такого масштабу. [31]

Python є одним із найпопулярніших варіантів для розробки алгоритмів машинного навчання. Простий та зрозумілий синтаксис цієї мови дозволяє зосередитись на вирішенні прикладних задач, не витрачаючи надмірно багато

часу на опанування інструментів. Python підтримує широкий спектр бібліотек, таких як TensorFlow, PyTorch, scikit-learn, що надають потужні можливості для розробки, навчання й оцінки моделей. Ці бібліотеки є не лише добре документованими, але й активно підтримуються спільнотою, що знижує ризики у випадку виникнення труднощів. Однак Python має й певні недоліки. Наприклад, він поступається C++ за швидкістю виконання, що може бути критичним для обчислювально інтенсивних задач. Водночас цей недолік можна частково компенсувати використанням апаратного прискорення.

Порівняння Python із іншими мовами програмування демонструє його очевидні переваги в контексті розробки систем машинного навчання. Наприклад, Java і C++ надають вищу продуктивність завдяки компіляції в машинний код, але вимагають більше часу та ресурсів для написання та підтримки коду. На відміну від них, Python пропонує значно ширший спектр бібліотек, які суттєво зменшують час розробки. Мови, такі як R, мають сильну орієнтацію на статистичний аналіз, але вони не настільки гнучкі для побудови складних систем, як Python. Різниця стає особливо помітною при розгляді доступу до інструментів глибокого навчання, де Python займає лідируючі позиції завдяки бібліотекам, як-от TensorFlow і PyTorch.

Інші мови, такі як R, C++ чи Java, також можна розглядати для таких задач. Як було сказано вище, R вирізняється своєю спеціалізацією на статистичному аналізі та візуалізації, однак у питаннях створення комплексних систем машинного навчання Python значно перевершує його. C++ і Java хоч і забезпечують високу продуктивність і контроль над ресурсами, проте їхній складний синтаксис і більша тривалість розробки роблять ці мови менш привабливими в контексті гнучких та швидких змін у проєктах.

Обрання Python як основної мови для цього проєкту є обґрунтованим не лише через її технічні переваги, але й через її інтеграцію з багатьма середовищами розробки. Одним з таких середовищ є Google Colab.

Серед інших численних варіантів інструментів Google Colab вирізняється своєю доступністю й функціональністю. Це безкоштовна платформа, яка

дозволяє працювати з Python у браузері, уникаючи витрат часу на локальне налаштування середовища. Завдяки можливості підключення GPU і TPU, обчислення можуть виконуватись швидше, що є критично важливим для великих моделей машинного навчання. Інша суттєва перевага Google Colab — можливість співпраці. Розробники можуть спільно редагувати код у реальному часі, що значно полегшує командну роботу.

На відміну від Google Colab, локальні середовища, такі як Jupyter Notebook чи PyCharm, мають свої особливості. Jupyter Notebook забезпечує інтерактивність роботи та ідеально підходить для досліджень і навчання, проте не має вбудованого апаратного прискорення. PyCharm, хоч і пропонує розширений функціонал, вимагає більше ресурсів і часто використовується для проєктів, що потребують комплексного підходу до налагодження.

На наш погляд Google Colab є ідеальним вибором для цього проєкту через свою адаптованість до задач машинного навчання. Її інтеграція з бібліотеками Python, доступ до обчислювальних ресурсів і простота використання забезпечують оптимальні умови для створення систем моніторингу стану доріг. Python і Google Colab у поєднанні стають інструментами, які дозволяють ефективно вирішувати складні задачі й підтримувати сучасні стандарти розробки програмного забезпечення.

Треба додати, що для систем моніторингу стану доріг, заснованих на машинному навчанні, вибір платформи є одним з ключових аспектів, що впливає на ефективність, масштабованість і зручність розробки.

Окрім Google Colab, існують кілька конкурентних платформ, кожна з яких має свої особливості. Наприклад, Microsoft Azure Machine Learning пропонує інтеграцію з екосистемою Microsoft, що зручно для користувачів Windows і корпоративних клієнтів, які вже використовують Azure для інших задач. Однак, використання Azure може бути дорожчим, і воно вимагає більш складного налаштування. Amazon SageMaker є ще однією потужною платформою, яка забезпечує масштабованість та інтеграцію з іншими сервісами AWS. Незважаючи на високі можливості, SageMaker часто критикують за складність

налаштування та значні фінансові витрати. IBM Watson Studio фокусується на корпоративних клієнтах і забезпечує високий рівень персоналізації, проте поступається конкурентам за популярністю серед окремих розробників і невеликих команд. У цьому контексті Google Colab вирізняється простотою доступу, мінімальними вимогами до попередньої підготовки та відсутністю плати для базового використання.

Технічна інтеграція Python із платформами машинного навчання є безперечною перевагою. Google Colab забезпечує зручне середовище для роботи з Python, яке підтримує використання зовнішніх бібліотек і API. Наприклад, через бібліотеку `gspread` розробники можуть інтегруватися з Google Sheets, а з `pandas` — легко обробляти великі набори даних. Для роботи з графічними процесорами Colab пропонує готові драйвери для використання NVIDIA CUDA, що дозволяє швидко масштабувати ресурси для обчислень. Крім того, платформа підтримує прямий доступ до Google Drive, що спрощує завантаження й збереження результатів роботи. На відміну від локальних середовищ, Google Colab автоматично зберігає стан роботи в хмарі, забезпечуючи безпеку даних і зручність у командних проектах.

Оптимізація обчислювальних ресурсів на платформі також одна з беззаперечних переваг для ефективного виконання задач. У Google Colab можна обирати між CPU, GPU і TPU, залежно від потреб конкретного етапу проекту. Наприклад, попередню обробку даних краще виконувати на CPU, тоді як для навчання моделей доцільніше використовувати GPU або TPU. Обмеження Google Colab у вигляді лімітів на час використання та доступ до ресурсів у безкоштовній версії можна обійти шляхом підключення до платних підписок. Для ефективного використання пам'яті можна застосовувати стратегії батчінгу (розподілу даних на невеликі партії), а також зберігати дані у стислому форматі для передачі між етапами обробки. Використання оптимізованих бібліотек, таких як NumPy чи CuPy, також дозволяє значно скоротити час обчислень.

Таким чином, у виборі мови програмування та середовища розробки для задач моніторингу стану доріг Python і Google Colab формують унікальний симбіоз, що поєднує потужність із доступністю, що є ключовим для ефективного впровадження сучасних рішень.

Висновки до розділу

В результаті виконання даного розділу вирішені такі питання:

- розглянуті деякі математичні моделі коливань автомобіля, визначено, що для них потрібно точно знати параметри системи, що в практичних умовах нереально, тому обраний шлях моделювання за допомогою машинного навчання;

- зроблений огляд методів машинного навчання для моніторингу стану доріг, з'ясовано, що остаточний вибір методів залежить від наявних даних;

- розглянуті варіанти технічних засобів при реалізації системи, вибір зроблено на користь мобільних пристроїв з вбудованими акселерометрами;

- обрані мова програмування (Python), середовище розробки (Google Colab) та стандартні бібліотеки машинного навчання та візуалізації.

РОЗДІЛ 3

РЕЗУЛЬТАТИ ВИРІШЕННЯ ЗАДАЧІ

3.1. Отримання даних та їх первинна підготовка

Перш за все, треба визначитись, яким чином отримувати дані для навчання моделі. В реальних умовах при розповсюдженні створеного додатка та встановленні його на багатьох пристроях можна буде користуватись тим потоком даних, який автоматично створить ця сукупність пристроїв. Інший шлях - встановити аналог додатка і спробувати на автомобілі (або декількох різних машинах) проїхати по різних маршрутах та зафіксувати реальні коливання корпусу автомобіля. Цей шлях потребує певних витрат ресурсів, часу та фінансів. Тому, поки немає ані потоку даних, ані ресурсів для “польових” досліджень, доводиться користуватись тим, що є в інтернеті.

Як відомо, абсолютна більшість комерційних проєктів в інтернеті суворо охороняють свої напрацювання, особливо дані, отримані при роботі своїх сервісів. Тому за відсутністю таких даних у вільному доступі було прийнято рішення звернутись до інших відкритих джерел. Одним з таких джерел є база геодезичних даних з вільним доступом OpenStreetMap. [32] Цей сервіс збирає геодані від зареєстрованих користувачів та вільно надає їх зацікавленим сервісам та приватним особам.

Для задач даної роботи були обрані геодані міста Львів, які задані такими приблизними координатами земної поверхні: від 49,83 до 49,85 градусів північної широти та від 24,01 до 24,04 градусів східної довготи. Це приблизний центр Львова.

Веб-сайт OpenStreetMap та його наявні «дзеркальні» джерела надають дані, які слідують деревовидній структурі та завантажуються у форматі OSM, який має вигляд, аналогічний популярному формату XML.

Спочатку підключаємо основні бібліотеки, необхідні для обробки та візуалізації даних (приклад відповідного фрагмента Python-коду показано на рис. 3.1).

```
▶ %matplotlib inline
import math
import numpy as np
import pandas as pd
from datetime import timedelta
import datetime as dt
import matplotlib.pyplot as plt
plt.rcParams['figure.figsize'] = [16, 10]
import seaborn as sns
import random
```

Рисунок 3.1 - Імпорт бібліотек

Щоб спростити роботу з даними, файл з OpenStreeledMap було збережено на Google Drive. Завдяки інтеграції з Google Colab можна уникнути повторного завантаження файлу щоразу, використовуючи прямий доступ до сервісу (рис. 3.2). Під час виконання програми з'явиться запит на надання дозволу, який необхідно підтвердити.

```
▶ from google.colab import drive
drive.mount('/gdrive')
%cd /gdrive
```

Рисунок 3.2 - Організація прямого доступу до Google Drive у Google Colab

Використовуючи доступ до Google Drive і знаючи шлях до файлу, завантажуюмо дані, створюючи початковий датафрейм (рис. 3.3).

При аналізі даних помічаємо значну кількість пропусків та несуттєвої інформації. Для подальшої роботи залишаємо лише часові позначки, ідентифікатори користувачів та географічні координати (широту 'lat' та довготу 'lon'). Загальну інформацію про первинний датасет та очищений набір даних наведено, відповідно, на рис. 3.4 та рис. 3.5.

```
df = pd.read_xml('/gdrive/MyDrive/ЛНУП/24-25/Дипл/мaп')
df
```

	note	osm_base	minlat	minlon	maxlat	maxlon	id	lat	lon	version	timestamp	changeset
0	The data included in this document is from www...	None	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	None	NaN
1	None	2024-11-24T20:18:29Z	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	None	NaN
2	None	None	49.83477	24.0177	49.84759	24.04328	NaN	NaN	NaN	NaN	None	NaN
3	None	None	NaN	NaN	NaN	NaN	26150791.0	49.841952	24.031592	133.0	2024-11-01T20:40:17Z	158636113.0
4	None	None	NaN	NaN	NaN	NaN	30508730.0	49.843264	24.022851	24.0	2024-06-13T13:38:15Z	152634694.0
...
40209	None	None	NaN	NaN	NaN	NaN	18242157.0	NaN	NaN	1.0	2024-11-02T16:39:19Z	158667052.0
40210	None	None	NaN	NaN	NaN	NaN	18242158.0	NaN	NaN	1.0	2024-11-02T16:39:19Z	158667052.0
40211	None	None	NaN	NaN	NaN	NaN	18242176.0	NaN	NaN	1.0	2024-11-02T16:51:40Z	158667538.0
40212	None	None	NaN	NaN	NaN	NaN	18242177.0	NaN	NaN	1.0	2024-11-02T16:51:40Z	158667538.0
40213	None	None	NaN	NaN	NaN	NaN	18310906.0	NaN	NaN	1.0	2024-11-21T00:13:34Z	159389759.0

40214 rows × 17 columns

Рисунок 3.3 - Завантаження даних

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 40214 entries, 0 to 40213
Data columns (total 17 columns):
 #   Column      Non-Null Count  Dtype
---  -
 0   note        1 non-null     object
 1   osm_base    1 non-null     object
 2   minlat      1 non-null     float64
 3   minlon      1 non-null     float64
 4   maxlat      1 non-null     float64
 5   maxlon      1 non-null     float64
 6   id          40211 non-null float64
 7   lat         33563 non-null float64
 8   lon         33563 non-null float64
 9   version     40211 non-null float64
10  timestamp   40211 non-null object
11  changeset   40211 non-null float64
12  uid         40211 non-null float64
13  user        40211 non-null object
14  tag         0 non-null     float64
15  nd          0 non-null     float64
16  member      0 non-null     float64
dtypes: float64(13), object(4)
memory usage: 5.2+ MB
```

Рисунок 3.4 - Загальна інформація про первинний датасет

```

df_lat_lon = df[['timestamp', 'user', 'lat', 'lon']]
df_lat_lon.info()

```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 40214 entries, 0 to 40213
Data columns (total 4 columns):
 #   Column      Non-Null Count  Dtype
---  ---
 0   timestamp  40211 non-null   object
 1   user        40211 non-null   object
 2   lat         33563 non-null   float64
 3   lon         33563 non-null   float64
dtypes: float64(2), object(2)
memory usage: 1.2+ MB

```

Рисунок 3.5 - Виділення ключових параметрів

Варто зазначити, що геоданих у наборі близько 33,5 тисяч точок, тоді як інші записи не є актуальними. Для оптимізації скорочуємо обсяг даних, водночас впорядковуючи їх (рис. 3.6).

```

df_lat_lon = df_lat_lon.dropna().sort_values(
    by=['timestamp', 'user', 'lat', 'lon']).reset_index(drop=True)
df_lat_lon

```

	timestamp	user	lat	lon
0	2008-05-03T12:05:25Z	zoidberg	49.845900	24.020238
1	2009-09-18T14:51:03Z	Badran	49.845133	24.039540
2	2009-10-07T15:33:04Z	datalogg	49.847311	24.039396
3	2009-10-07T15:33:04Z	datalogg	49.847514	24.039620
4	2009-10-07T15:33:04Z	datalogg	49.847546	24.040603
...
33558	2024-11-24T06:47:41Z	Oppair	49.841686	24.026894
33559	2024-11-24T06:47:41Z	Oppair	49.841686	24.023159
33560	2024-11-24T06:47:41Z	Oppair	49.841917	24.022787
33561	2024-11-24T06:47:41Z	Oppair	49.841924	24.023667
33562	2024-11-24T06:47:41Z	Oppair	49.841950	24.023814

33563 rows x 4 columns

Рисунок 3.6 - Виділення геоданих та їх упорядкування

Встановлюємо координати для центру міста та прилеглих районів (код показано на рис. 3.7).

```

min_lat, max_lat = 49.83477, 49.84759
min_lon, max_lon = 24.01770, 24.04328
df_lat_lon_1 = df_lat_lon[(df_lat_lon['lat']>min_lat) & (df_lat_lon['lat']<max_lat) &
(df_lat_lon['lon']>min_lon) & (df_lat_lon['lon']<max_lon)]
df_lat_lon_1

```

	timestamp	user	lat	lon
0	2008-05-03T12:05:25Z	zoidberg	49.845900	24.020238
1	2009-09-18T14:51:03Z	Badran	49.845133	24.039540
2	2009-10-07T15:33:04Z	datalogg	49.847311	24.039396
3	2009-10-07T15:33:04Z	datalogg	49.847514	24.039620
4	2009-10-07T15:33:04Z	datalogg	49.847546	24.040603
...
33558	2024-11-24T06:47:41Z	Oppair	49.841686	24.026894
33559	2024-11-24T06:47:41Z	Oppair	49.841686	24.023159
33560	2024-11-24T06:47:41Z	Oppair	49.841917	24.022787
33561	2024-11-24T06:47:41Z	Oppair	49.841924	24.023667
33562	2024-11-24T06:47:41Z	Oppair	49.841950	24.023814

30818 rows × 4 columns

Рисунок 3.7 - Датасет центра Львова

Для аналізу візуалізуємо отримані дані. Фрагмент коду, який використовується для побудови графіку, представлено на рис. 3.8, а результат первинної візуалізації – на рис. 3.9.

```

N = df_lat_lon_1.shape[0]

city_lat_border = (min_lat, max_lat)
city_long_border = (min_lon, max_lon)
plt.figure(figsize=(7,6))
plt.scatter(df_lat_lon_1['lon'][:N], df_lat_lon_1['lat'][:N], color='blue',
            s=1, alpha=0.5)
plt.ylabel('latitude')
plt.xlabel('longitude')
plt.ylim(city_lat_border)
plt.xlim(city_long_border)
plt.show()

```

Рисунок 3.8 - Код для створення первинної візуалізації

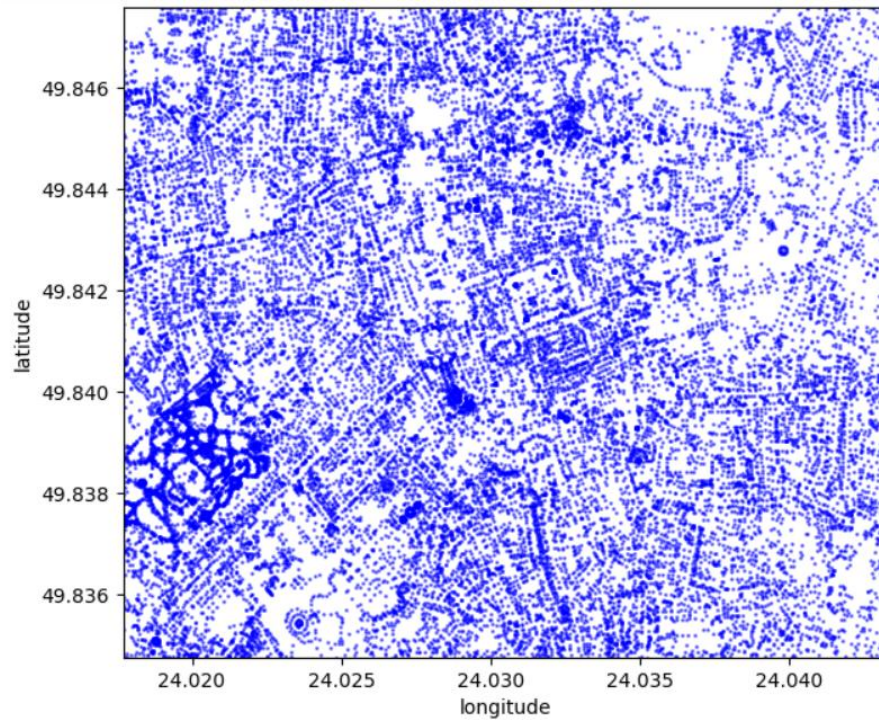


Рисунок 3.9 - Результат первинної візуалізації

Для перевірки можна зіставити побудоване зображення з картою цієї ж місцевості з сервісу OpenStreetMap (рис. 3.10). Хоча безпосереднє порівняння набору точок із векторними об'єктами є складним, на обох картах легко помітити подібність у розташуванні окремих вулиць і ключових елементів.

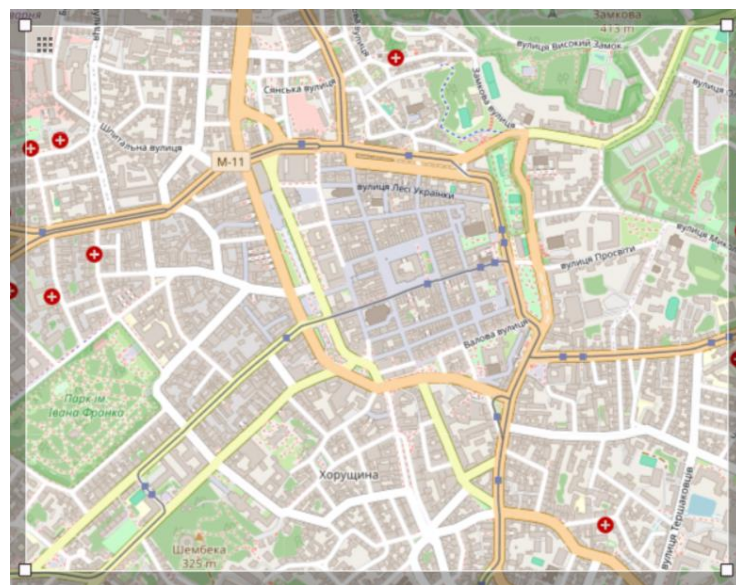
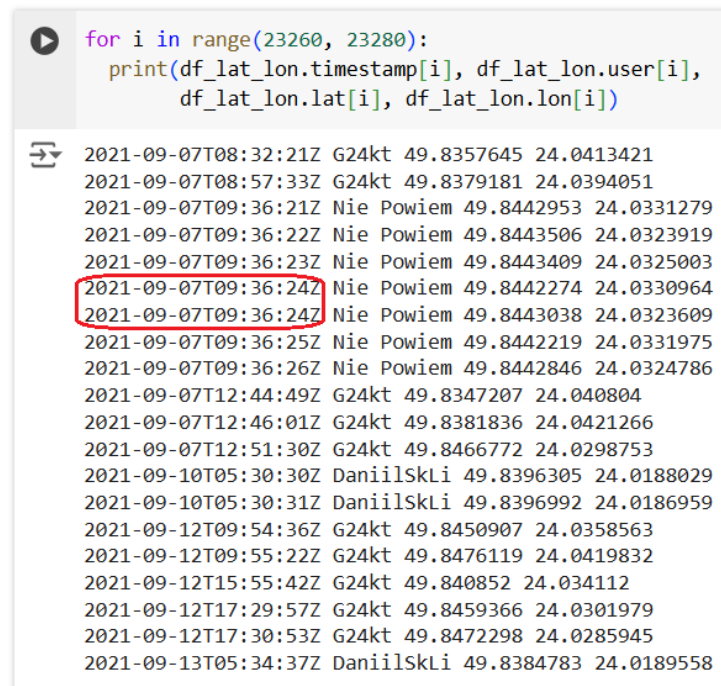


Рисунок 3.10 - Карта відповідного району в сервісі OpenStreetMap

Під час подальшого аналізу виявлено повторення часових позначок для одних і тих самих користувачів (рис. 3.11).

```
▶ for i in range(23260, 23280):  
    print(df_lat_lon.timestamp[i], df_lat_lon.user[i],  
          df_lat_lon.lat[i], df_lat_lon.lon[i])
```



```
↵ 2021-09-07T08:32:21Z G24kt 49.8357645 24.0413421  
2021-09-07T08:57:33Z G24kt 49.8379181 24.0394051  
2021-09-07T09:36:21Z Nie Powiem 49.8442953 24.0331279  
2021-09-07T09:36:22Z Nie Powiem 49.8443506 24.0323919  
2021-09-07T09:36:23Z Nie Powiem 49.8443409 24.0325003  
2021-09-07T09:36:24Z Nie Powiem 49.8442274 24.0330964  
2021-09-07T09:36:24Z Nie Powiem 49.8443038 24.0323609  
2021-09-07T09:36:25Z Nie Powiem 49.8442219 24.0331975  
2021-09-07T09:36:26Z Nie Powiem 49.8442846 24.0324786  
2021-09-07T12:44:49Z G24kt 49.8347207 24.040804  
2021-09-07T12:46:01Z G24kt 49.8381836 24.0421266  
2021-09-07T12:51:30Z G24kt 49.8466772 24.0298753  
2021-09-10T05:30:30Z DaniilSkLi 49.8396305 24.0188029  
2021-09-10T05:30:31Z DaniilSkLi 49.8396992 24.0186959  
2021-09-12T09:54:36Z G24kt 49.8450907 24.0358563  
2021-09-12T09:55:22Z G24kt 49.8476119 24.0419832  
2021-09-12T15:55:42Z G24kt 49.840852 24.034112  
2021-09-12T17:29:57Z G24kt 49.8459366 24.0301979  
2021-09-12T17:30:53Z G24kt 49.8472298 24.0285945  
2021-09-13T05:34:37Z DaniilSkLi 49.8384783 24.0189558
```

Рисунок 3.11 - Приклад дублювання часових позначок

Це пояснюється тим, що геодані можуть надходити кілька разів за секунду. У такому форматі розрахунок швидкості руху користувачів стає проблематичним. Для вирішення цієї проблеми усереднюємо координати для записів з однаковими часовими позначками. Такий підхід дозволить точно визначати швидкість руху.

Приклад виконання розрахунків швидкості наведено на рис. 3.12. Крім того, для уникнення повторів проводимо агрегацію даних, результат якої, разом із кодом, показано на рис. 3.13.

Перевіряємо загальну інформацію про отриманий датасет. Результати перевірки наведено на рис.3.14.

```

u = 'Nie Powiem'
time1 = '2021-09-07T09:36:23Z'
time2 = '2021-09-07T09:36:24Z'
lat1 = df_lat_lon[(df_lat_lon.user == u) &
                  (df_lat_lon.timestamp == time1)].lat.mean()
lat2 = df_lat_lon[(df_lat_lon.user == u) &
                  (df_lat_lon.timestamp == time2)].lat.mean()
lon1 = df_lat_lon[(df_lat_lon.user == u) &
                  (df_lat_lon.timestamp == time1)].lon.mean()
lon2 = df_lat_lon[(df_lat_lon.user == u) &
                  (df_lat_lon.timestamp == time2)].lon.mean()
diff_lat, diff_lon = abs(lat2-lat1)*scale_lat, abs(lon2-lon1)*scale_lon
velocity = (diff_lat**2 + diff_lon**2) ** 0.5 * 3.6
print ("Середня широта в першу секунду, град:", lat1)
print ("Середня широта в другу секунду, град:", lat2)
print ("Середня довгота в першу секунду, град:", lon1)
print ("Середня довгота в другу секунду, град:", lon2)
print ("Відстань по широті, м:", diff_lat)
print ("Відстань по довготі, м:", diff_lon)
print ("Середня швидкість, км/год:", round(velocity, 1))

```


 Середня широта в першу секунду, град: 49.8443409
 Середня широта в другу секунду, град: 49.8442656
 Середня довгота в першу секунду, град: 24.0325003
 Середня довгота в другу секунду, град: 24.032728650000003
 Відстань по широті, м: 8.28299999877916
 Відстань по довготі, м: 15.984500000278956
 Середня швидкість, км/год: 64.8

Рисунок 3.12 - Розрахунок швидкості руху

```

df_lat_lon_agg = df_lat_lon_1.groupby(df_lat_lon_1.columns.
                                     drop(['lat', 'lon']).tolist(), as_index=False).agg(np.mean)
df_lat_lon_agg

```

	timestamp	user	lat	lon
0	2008-05-03T12:05:25Z	zoidberg	49.845900	24.020238
1	2009-09-18T14:51:03Z	Badran	49.845133	24.039540
2	2009-10-07T15:33:04Z	datalogg	49.847503	24.040471
3	2009-10-07T15:33:07Z	datalogg	49.846906	24.035732
4	2009-10-07T15:33:09Z	datalogg	49.847237	24.038930
...
3822	2024-11-20T15:51:54Z	G24kt	49.841398	24.033839
3823	2024-11-21T00:12:44Z	hvictor	49.840388	24.024726
3824	2024-11-21T00:16:49Z	hvictor	49.839921	24.024177
3825	2024-11-21T07:08:28Z	*Martin*	49.842576	24.033601
3826	2024-11-24T06:47:41Z	Oppair	49.841833	24.024064

3827 rows x 4 columns

Рисунок 3.13 - Код для агрегації та його результат

```
df_lat_lon_agg.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3827 entries, 0 to 3826
Data columns (total 4 columns):
#   Column      Non-Null Count  Dtype
---  -
0   timestamp  3827 non-null   object
1   user        3827 non-null   object
2   lat         3827 non-null   float64
3   lon         3827 non-null   float64
dtypes: float64(2), object(2)
memory usage: 119.7+ KB
```

Рисунок 3.14 - Загальна інформація про підготовлений датасет

На цьому етап попередньої підготовки даних можна вважати закінченим. Отриманий датасет має майже 4 тис. об'єктів та готовий для подальших дій.

3.2. Підготовка даних для машинного навчання

Для визначення умовних дефектів на окремих ділянках, ґрунтуючись на результатах аналізу [33], було прийнято таку класифікацію можливих прискорень:

- до 1 м/с^2 – незначні дефекти;
- близько 2 м/с^2 – помітні пошкодження;
- 3 м/с^2 і більше – значні дефекти, які можуть потребувати термінового ремонту.

Обираємо 10 випадкових точок і присвоюємо їм умовні значення дефектів (рис. 3.15). Ці дані будуть в якості “генератора” сигналів.

Виконуємо візуалізацію аналогічно до первинної, щоб оцінити розташування умовних дефектних зон (код наведено на рис. 3.16, а результат – на рис. 3.17).

При аналізі точок, розташованих поблизу дефектів, було виявлено незбалансованість у датасеті. Це може бути зумовлено відмінностями у щільності транспортних потоків. Для виправлення ситуації вирішено збалансувати дані. Умовні значення прискорень встановлюємо випадковими,

але з урахуванням середнього значення, характерного для заданої дефектної зони.

```
df_lat_lon_agg_defekt = df_lat_lon_agg.sample(n=10, random_state=4)[['lat', 'lon']]
df_lat_lon_agg_defekt['accel'] = 1
df_lat_lon_agg_defekt.accel[1::3] = 2
df_lat_lon_agg_defekt.accel[2::3] = 3
df_lat_lon_agg_defekt
```

	lat	lon	accel
1460	49.843015	24.029010	1
1520	49.836078	24.017897	2
331	49.843676	24.019297	3
1845	49.838874	24.027777	1
2849	49.843495	24.034450	2
3260	49.846123	24.027176	3
2001	49.840379	24.031810	1
2365	49.843358	24.039285	2
640	49.843440	24.042625	3
2120	49.840440	24.019153	1

Рисунок 3.15 - Координати умовних дефектних ділянок

```
N = df_lat_lon_1.shape[0]
city_lat_border = (min_lat, max_lat)
city_long_border = (min_lon, max_lon)
plt.figure(figsize=(7,6))
plt.scatter(df_lat_lon_1['lon'][:N], df_lat_lon_1['lat'][:N], color='blue',
            s=1, alpha=0.5)
plt.scatter(df_lat_lon_agg_defekt.loc[df_lat_lon_agg_defekt.accel==1].lon,
            df_lat_lon_agg_defekt.loc[df_lat_lon_agg_defekt.accel==1].lat,
            color='yellow', s=100, alpha=1)
plt.scatter(df_lat_lon_agg_defekt.loc[df_lat_lon_agg_defekt.accel==2]['lon'],
            df_lat_lon_agg_defekt.loc[df_lat_lon_agg_defekt.accel==2]['lat'],
            color='orange', s=100, alpha=1)
plt.scatter(df_lat_lon_agg_defekt.loc[df_lat_lon_agg_defekt.accel==3]['lon'],
            df_lat_lon_agg_defekt.loc[df_lat_lon_agg_defekt.accel==3]['lat'],
            color='red', s=100, alpha=1)
plt.ylabel('latitude')
plt.xlabel('longitude')
plt.ylim(city_lat_border)
plt.xlim(city_long_border)
plt.show()
```

Рисунок 3.16 - Код візуалізації дефектних ділянок

В реальних умовах є сенс також враховувати щільність транспортного потоку, щоб балансувати дані.

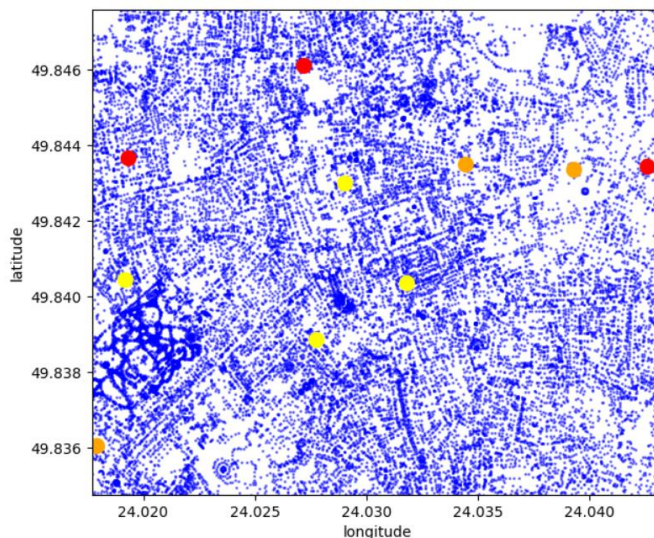


Рисунок 3.17 - Візуалізація умовних дефектних ділянок

Код, що реалізує балансування даних, разом із результатами його виконання, представлений на рис. 3.18. Для підвищення варіативності додаємо ще "шум" у вигляді випадкових координат та коливань у кількох додаткових точках (рис. 3.19).

```

accr = 0.0005
quant = 15
df_lat_lon_accel = pd.DataFrame(columns=list(df_lat_lon_agg_defekt.axes[1]))
for i in df_lat_lon_agg_defekt.index:
    add = df_lat_lon[(abs(df_lat_lon['lat'])-df_lat_lon_agg_defekt.loc[i].lat)<=accr) &
                (abs(df_lat_lon['lon'])-df_lat_lon_agg_defekt.loc[i].lon)<=accr)][['lat', 'lon']]
    add['accel'] = 0
    if (add.shape[0]>quant*1.3):
        add = add.sample(random.randrange(int(quant*0.7),int(quant*1.3)))
    elif (add.shape[0]<quant*0.7):
        k = quant // add.shape[0]
        a = add
        for j in range(k):
            add = pd.concat([add, a])
    print(add.shape[0])
    for j in add.index:
        add.accel[j] = df_lat_lon_agg_defekt.accel[i] + random.uniform(-0.8, 0.8)
df_lat_lon_accel = pd.concat([df_lat_lon_accel, add])
df_lat_lon_accel

```

	lat	lon	accel
19133	49.843339	24.029476	0.605509
33514	49.842833	24.028893	1.036758
28178	49.842922	24.029001	1.299952
9772	49.843072	24.029470	0.238235
13409	49.842964	24.028701	1.655061
...
2133	49.840896	24.018862	1.323514
15546	49.839965	24.018999	0.700496
15559	49.840055	24.018904	1.099370
13140	49.840884	24.019327	0.254053
2554	49.840659	24.019460	1.054523

143 rows x 3 columns

Рисунок 3.18 - Збалансовані дані

```
df_lat_lon_noise = pd.DataFrame(columns=list(df_lat_lon_agg_defekt.axes[1]))
for i in range(10):
    df_lat_lon_noise.loc[len(df_lat_lon_noise)] = [random.uniform(min_lat, max_lat),
                                                    random.uniform(min_lon, max_lon), random.uniform(0.5, 3.5)]
df_lat_lon_accel = pd.concat([df_lat_lon_accel, df_lat_lon_noise])
df_lat_lon_accel
```

	lat	lon	accel
19133	49.843339	24.029476	0.605509
33514	49.842833	24.028893	1.036758
28178	49.842922	24.029001	1.299952
9772	49.843072	24.029470	0.238235
13409	49.842964	24.028701	1.655061
...
5	49.845281	24.026986	3.257157
6	49.846470	24.024354	1.324313
7	49.839453	24.025503	1.426592
8	49.840148	24.035113	0.982677
9	49.840964	24.040309	1.829515

153 rows x 3 columns

Рисунок 3.19 - Підготовлений набір даних для машинного навчання

Результатом наших дій став готовий набір даних, придатний для використання у задачах машинного навчання. Для наочності виконаємо його візуалізацію на карті міста (рис. 3.20).

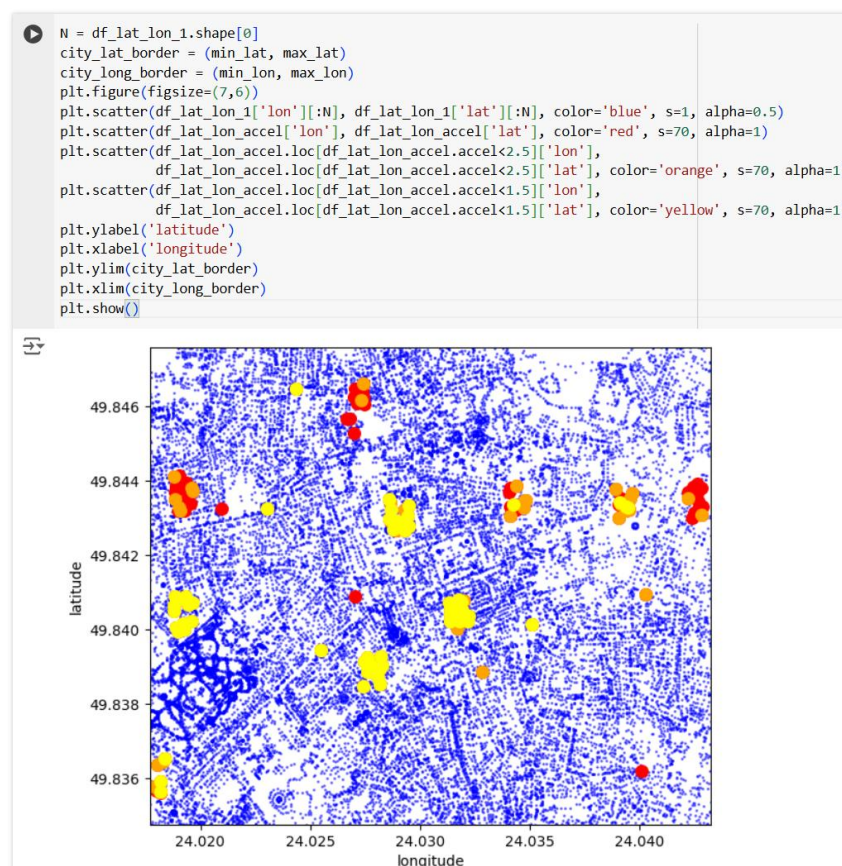


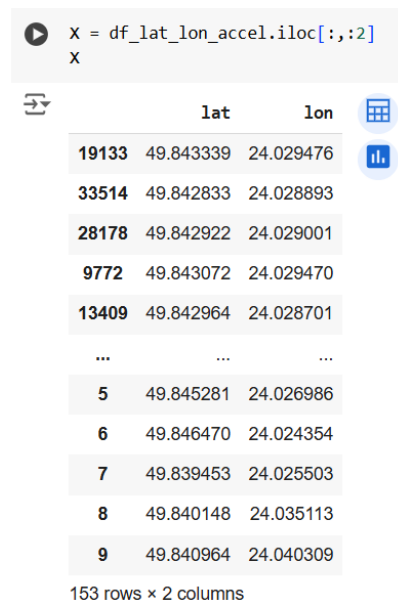
Рисунок 3.20 - Візуалізація набору даних для машинного навчання

Наступний крок - безпосередньо налаштування моделі машинного навчання, отримання та оцінка результатів.

3.3. Навчання моделі та отримання результатів

Для вирішення задачі машинного навчання була обрана модель кластеризації. Це зумовлено тим, що в реальних умовах ми маємо лише дані про прискорення (або коливання) автомобілів, прив'язані до географічних координат, і ніяких додаткових параметрів. Таким чином, задача полягає у виявленні проблемних ділянок дороги за допомогою алгоритму "без вчителя", що шукає "скупчення" сигналів про дефекти.

Спершу з навчального датасету обираються лише геодані (рис. 3.21).



```
X = df_lat_lon_accel.iloc[:, :2]
X
```

	lat	lon
19133	49.843339	24.029476
33514	49.842833	24.028893
28178	49.842922	24.029001
9772	49.843072	24.029470
13409	49.842964	24.028701
...
5	49.845281	24.026986
6	49.846470	24.024354
7	49.839453	24.025503
8	49.840148	24.035113
9	49.840964	24.040309

153 rows x 2 columns

Рисунок 3.21 - Навчальна вибірка для кластеризації

Для пошуку оптимальної кількості кластерів навчання моделі проводиться методом К-середніх із різною кількістю кластерів (від 5 до 20). Результати оцінюються за показником інерції (рис. 3.22).

```

from sklearn.cluster import KMeans
inertia = []
for k in range(5,20):
    km = KMeans(n_clusters=k, n_init='auto', random_state=1).fit(X)
    inertia.append(np.sqrt(km.inertia_))

```

Рисунок 3.22 - Навчання моделі з різною кількістю кластерів

Отримані значення інерції візуалізуються для аналізу "зламів" кривої, які вказують на потенційно оптимальну кількість кластерів (рис. 3.23).

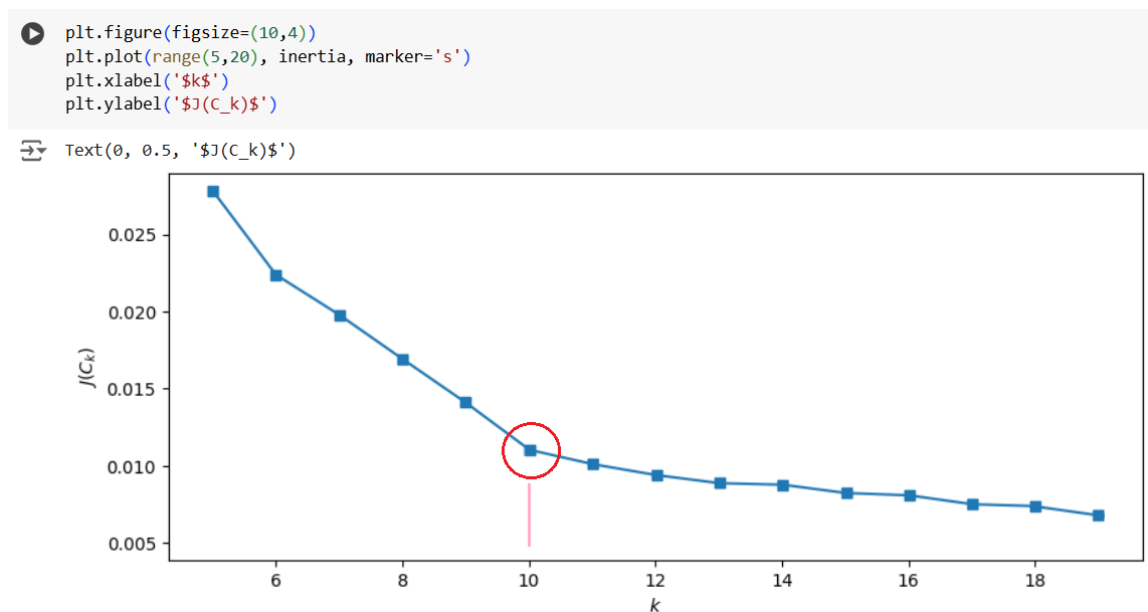


Рисунок 3.23 - Візуалізація показника інерції

Для точнішого визначення оптимальної кількості кластерів використовується метод розрахунку вуглів (рис. 3.24) між точками кривої інерції. Бачимо, що оптимальна кількість кластерів дорівнює 10.

На наступному етапі модель навчається для 10 кластерів, після чого визначається, до якого кластеру належить кожна точка (рис. 3.25 і 3.26).

```

▶ angles = []
for i in range(13):
    x1, y1 = i, inertia[i]
    x2, y2 = i+1, inertia[i+1]
    x3, y3 = i+2, inertia[i+2]
    a = math.sqrt((x2 - x1)*(x2 - x1) + (y2 - y1)*(y2 - y1))
    b = math.sqrt((x2 - x3)*(x2 - x3) + (y2 - y3)*(y2 - y3))
    c = math.sqrt((x3 - x1)*(x3 - x1) + (y3 - y1)*(y3 - y1))
    angles.append(math.acos((a*a + b*b - c*c) / (2*a*b)))
print('Min angle in radians = %7.5f' % min(angles))
print('Clusters =', angles.index(min(angles)) + 10)

```

↗ Min angle in radians = 3.13874
 Clusters = 10

Рисунок 3.24 - Пошук оптимальної кількості кластерів

```

▶ km = KMeans(n_clusters=10, n_init='auto', random_state=2).fit(X)
km.cluster_centers_

```

↗ array([[49.84317794, 24.0344756],
 [49.84358735, 24.01943495],
 [49.84285893, 24.03942852],
 [49.84614934, 24.02694614],
 [49.83593758, 24.01803335],
 [49.840377 , 24.03182789],
 [49.84300196, 24.02899227],
 [49.83913444, 24.02760822],
 [49.84346469, 24.04260579],
 [49.84048376, 24.01917939]])

Рисунок 3.25 - Навчання моделі для 10 кластерів

```

▶ df_cluster = df_lat_lon_accel
df_cluster['cluster'] = kmp
df_cluster

```

	lat	lon	accel	cluster
19133	49.843339	24.029476	0.605509	6
33514	49.842833	24.028893	1.036758	6
28178	49.842922	24.029001	1.299952	6
9772	49.843072	24.029470	0.238235	6
13409	49.842964	24.028701	1.655061	6
...
5	49.845281	24.026986	3.257157	3
6	49.846470	24.024354	1.324313	3
7	49.839453	24.025503	1.426592	7
8	49.840148	24.035113	0.982677	0
9	49.840964	24.040309	1.829515	2

153 rows × 4 columns

Рисунок 3.26 - Розподіл точок по кластерам

Дані групуються за кластерами, обчислюються середні значення прискорень для кожного кластеру (рис. 3.27), і додаються координати центрів (рис. 3.28), а це означає, що фактично ми отримали результати ідентифікації.

```
df_cluster_accel = df_cluster.groupby(df_cluster.columns.drop(['lat', 'lon', 'accel']).tolist() as_index=False).agg(np.mean).drop(['lat', 'lon'], axis=1)
df_cluster_accel
```

	cluster	accel
0	0	2.003092
1	1	2.781978
2	2	1.983813
3	3	2.808073
4	4	1.932265
5	5	0.992996
6	6	1.084358
7	7	1.030300
8	8	2.988841
9	9	0.802865

Рисунок 3.27 - Групування даних по кластерах

```
df_ident_clusters = df_cluster_accel
df_ident_clusters['lat'] = km.cluster_centers[:, 0]
df_ident_clusters['lon'] = km.cluster_centers[:, 1]
df_ident_clusters
```

	cluster	accel	lat	lon
0	0	2.003092	49.843178	24.034476
1	1	2.781978	49.843587	24.019435
2	2	1.983813	49.842859	24.039429
3	3	2.808073	49.846149	24.026946
4	4	1.932265	49.835938	24.018033
5	5	0.992996	49.840377	24.031828
6	6	1.084358	49.843002	24.028992
7	7	1.030300	49.839134	24.027608
8	8	2.988841	49.843465	24.042606
9	9	0.802865	49.840484	24.019179

Рисунок 3.28 - Результати ідентифікації

Порівняння координат центрів із даними умовних дефектів дозволяє оцінити похибки у визначенні відстаней та прискорень (рис. 3.29).

```

diffs = pd.DataFrame(columns=['dist','accel'])
for i in df_lat_lon_agg_defekt.index:
    min_dist = 1000
    j_cluster = 0
    for j in df_ident_clusters.index:
        dist = math.sqrt((df_lat_lon_agg_defekt.lat[i]-df_ident_clusters.lat[j])**2 +
                        (df_lat_lon_agg_defekt.lon[i]-df_ident_clusters.lon[j])**2)
        if dist < min_dist:
            min_dist = dist
            j_cluster = j
    diffs.loc[len(diffs)] = [min_dist*scale_dist, abs(df_ident_clusters.accel[j_cluster]-
                                                    df_lat_lon_agg_defekt.accel[i])]
diffs

```

	dist	accel
0	0.440170	0.084358
1	3.914453	0.067735
2	3.279505	0.218022
3	6.206918	0.030300
4	6.361838	0.003092
5	4.627355	0.191927
6	0.360023	0.007004
7	10.385919	0.016187
8	0.625565	0.011159
9	1.022081	0.197135

Рисунок 3.29 - Похибки у відстанях та прискореннях

Чисельна оцінка похибок наведена у вигляді мінімальних, максимальних і середніх значень (рис. 3.30).

```

print('Errors of dist (m): Max =', round(diffs.dist.max(), 1),
      '; Min =', round(diffs.dist.min(), 1),
      '; Mean =', round(diffs.dist.mean(), 1))
print('Errors of accel (m/s**2): Max =', round(diffs.accel.max(), 2),
      '; Min =', round(diffs.accel.min(), 2),
      '; Mean =', round(diffs.accel.mean(), 2))
print('Accuracy of identification (%)', abs(round((1-diffs.accel.mean())*100, 2)))

```

```

Errors of dist (m): Max = 10.4 ; Min = 0.4 ; Mean = 3.7
Errors of accel (m/s**2): Max = 0.22 ; Min = 0.0 ; Mean = 0.08
Accuracy of identification (%): 91.73

```

Рисунок 3.30 - Чисельна оцінка похибок

Аналіз результатів показує, що середня похибка у визначенні координат дефектних ділянок становить близько 4 метрів, що є прийнятним для багатьох

практичних застосувань. Точність ідентифікації класу дефекту досягає майже 92%, що свідчить про високу ефективність обраної моделі кластеризації.

Такі результати демонструють потенціал застосування підходу для виявлення та класифікації дефектів дорожнього покриття в реальних умовах.

3.4. Розробка загальної структури системи

Поява пристроїв IoT, а саме датчиків, і їх здатність з'єднуватися один з одним для створення інтелектуальної мережі для постійного обміну інформацією породила нову концепцію роботи з даними під назвою Big Data. Мова йде про технологію Big Data Systems Architecture (BDSA). Цей підхід вимагає прийняття належних технологій для задоволення потреб у зборі, зберіганні, обробці та аналізі даних. [34]

Таким чином, якщо для моніторингу дорожніх покриттів пропонується BDSA-підхід, то можна позначити основні шари та компоненти системи. Приклад такої архітектури наведений на рис. 3.31.

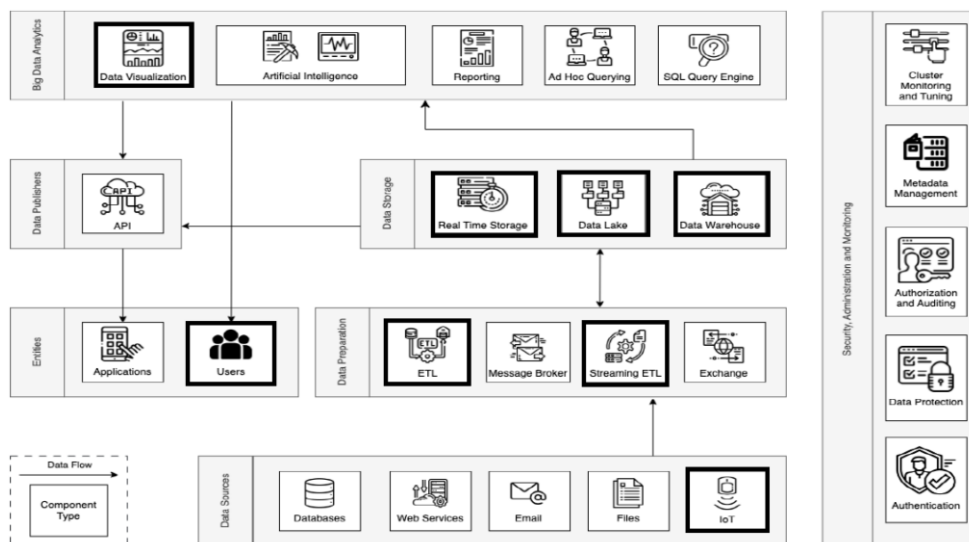


Рисунок 3.31 - Архітектура BDSA-системи для моніторингу доріг

BDSA — це багаторівнева архітектура, яка забезпечує логічне поділ завдань, де кожен рівень виконує конкретну роль і взаємодіє з шарами над і під

ним. Це допомагає побудувати масштабовану, надійну та потужну архітектуру, здатну впоратися з вимогами обробки та зберігання великих даних.

Перший рівень BDSA – це джерела даних, які можуть включати дорожні датчики, камери моніторингу та дані системи керування дорожнім рухом, серед іншого. Ці джерела даних відповідають за збір відповідної інформації про дорожній рух і дороги. Різноманітність джерел даних означає, що дані можуть бути зібрані в різних форматах, а саме структурованих, неструктурованих або напівструктурованих даних.

Другий рівень — це обробка даних, де необроблені дані аналізуються та перетворюються на більш корисну та значущу інформацію. На цьому етапі можна застосувати такі методи, як очищення, стандартизація та збагачення даних, а також алгоритми машинного навчання для вдосконаленої обробки та перетворення даних.

Після обробки дані спрямовуються на рівень зберігання, де вони зберігаються структурованим чином і стають доступними для майбутніх запитів. Цей рівень може передбачати використання баз даних або розподілених систем зберігання, таких як розподілена файлова система Hadoop (HDFS), Google Big Table або S3 Amazon у контексті великих даних. Основним завданням на цьому рівні є ефективне зберігання величезних обсягів даних таким чином, щоб забезпечити швидкий пошук і аналіз.

Після того, як дані зберігаються, рівень Big Data Analytics включає аналіз цих даних. Зазвичай це включає візуалізацію даних, штучний інтелект для моделювання прогнозів щодо доріг здоров'я, звітність, спеціальні запити (ідеально підходить для дослідницького інтерактивного аналізу даних) і систему запитів на мові структурованих запитів (SQL), що використовується для планових звітів і обробки транзакцій. Мета тут полягає в тому, щоб отримати інформацію, тенденції та закономірності зі збережених даних, представлених у вигляді графіків, таблиць або інших форм візуального представлення. Це дозволяє зовнішнім додаткам або користувачам, таким як

дорожні інженери або органи дорожнього руху, легко отримувати доступ і інтерпретувати інформацію, отриману від моніторингу доріг.

П'ятий рівень – це Data Publisher, який відповідає за публікацію даних і результатів аналізу в зовнішніх системах, тобто програмах. Це можуть бути бази даних, інші системи великих даних або інтерфейси прикладних програм (API). Цей рівень забезпечує широку доступність результатів системи обробки великих даних для використання іншими компонентами ширшої технологічної екосистеми.

Сутності рівня представляють споживачів інформації, отриманої з рівня аналітики, і можуть бути людьми або машинами. Люди, наприклад дорожні інженери чи органи дорожнього руху, шукають важливу інформацію на інформаційних панелях або звітах. Машини можуть використовувати ці дані для прийняття рішень або автоматизації.

Нарешті, рівень безпеки, адміністрування та моніторингу перетинає всі попередні рівні. Це передбачає забезпечення безпеки даних і самої системи великих даних, адміністрування системи (наприклад, керування доступом користувачів, розподіл ресурсів) і моніторинг системи (наприклад, відстеження продуктивності, ідентифікація та обробка збоїв). Ці аспекти мають вирішальне значення для загальної цілісності, надійності та продуктивності системи великих даних.

Однією з вимог до системи є можливість відстежувати трафік у реальному часі, що називається обробкою в реальному часі. Він відповідає за збір і обробку даних на високій швидкості, а також робить їх доступними в інструменті візуалізації, тобто на інформаційній панелі. На додаток до панелі приладів, він також може показувати таблицю, що містить журнали тривожних подій для зареєстрованих транспортних засобів з дуже низькою швидкістю, надмірною швидкістю або надмірною вагою, а також низькими або високими температурами. Іншою системною вимогою є підтримка пакетної обробки; він відповідає за збір і перетворення даних, але етапи зберігання є складнішими та застосовуються до більшого обсягу історичних даних (рис.3.32).

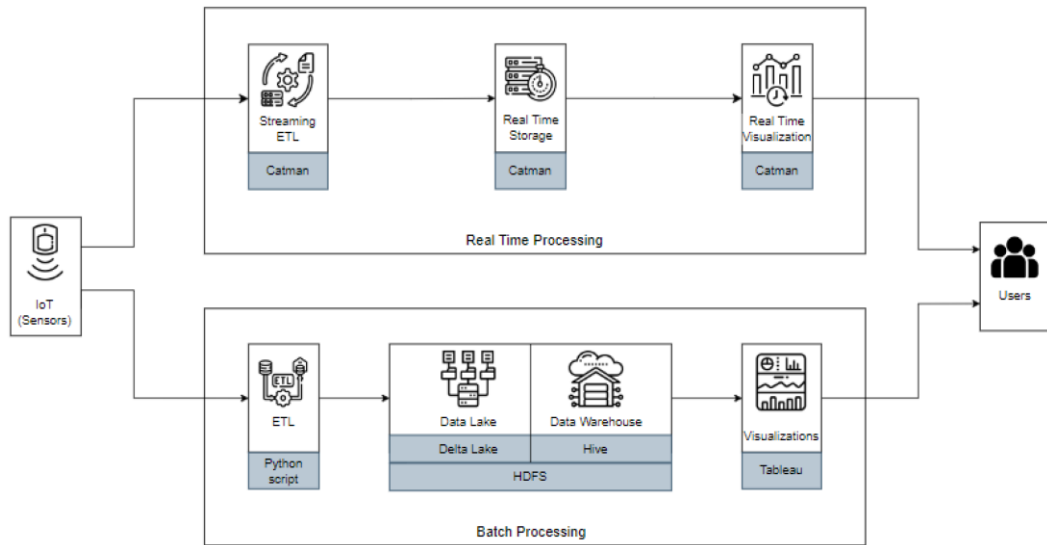


Рисунок 3.32 - Один з варіантів архітектури потоку даних

Якщо в якості датчиків використовуються не тільки акселерометри, то можна формувати різноманітні бази даних відповідних напрямків даних (прискорень, температури повітря, вантажності автівок і багато іншого). Інформація з датчиків попередньо обробляється та агрегується, щоб високоструктуровані дані. Такий підхід дає змогу відповідати на питання, що стосуються моніторингу доріг, а також використовувати різноманітні алгоритми для бізнес-процесів. Варіант такої бази даних наведено на рис. 3.33).

Таким чином, використання BDSA-підходу дозволяє обробляти великі обсяги даних із різних джерел, таких як датчики, відеокамери та мобільні пристрої, забезпечуючи гнучкість у масштабуванні з додаванням нових джерел даних. Архітектура дозволяє об'єднувати структуровані й неструктуровані дані (наприклад, геодані, зображення, звіти) в єдиній системі для отримання більш комплексного аналізу. Використання BDSA дозволяє застосовувати алгоритми машинного навчання та штучного інтелекту для прогнозування стану доріг, пріоритизації ремонтів і оптимізації витрат. Система може обробляти дані в розподіленому середовищі, забезпечуючи стійкість до збоїв і безперервну роботу навіть за умов високого навантаження.

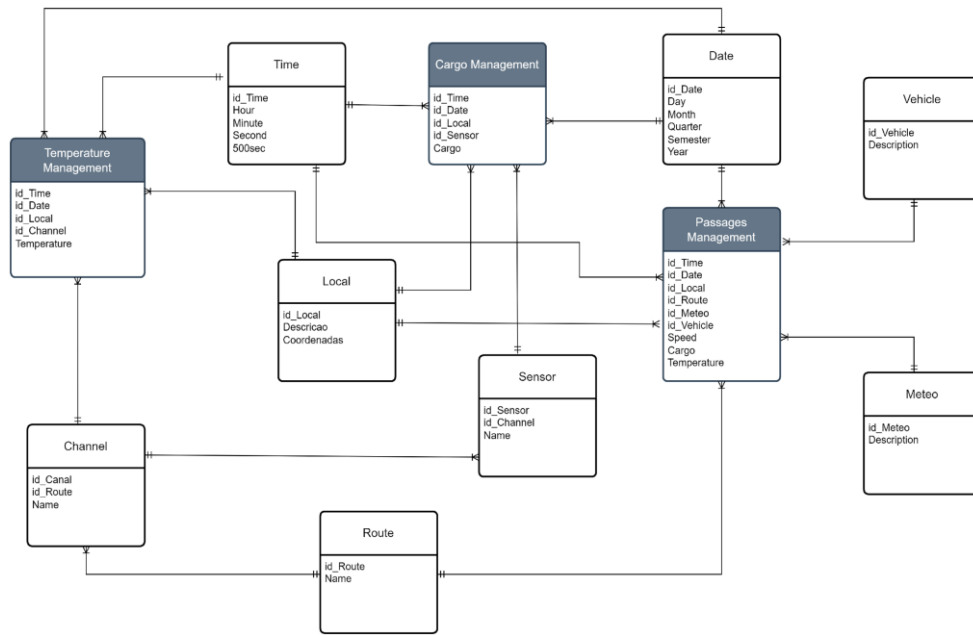


Рисунок 3.33 - Варіант SQL бази даних

Модульна структура BDSA дозволяє легко адаптувати систему до змін у вимогах або технологіях, наприклад, інтегрувати нові датчики чи алгоритми обробки. Також впровадження BDSA-підходу робить системи моніторингу доріг не лише ефективнішими, а й більш перспективними для майбутніх удосконалень.

Висновки до розділу

Таким чином, в даному розділі були виконані наступні завдання:

- обрано джерело даних, проведено завантаження та попередня обробка;
- визначено, що задача відноситься до задач кластеризації;
- здійснено пошук оптимальної кількості кластерів по показнику інерції;
- отримані результати ідентифікації дефектних ділянок, здійснена перевірка на похибки та їх чисельна оцінка;
- запропоновані варіанти структури системи та архітектури потоків даних.

РОЗДІЛ 4

ОХОРОНА ПРАЦІ ТА БЕЗПЕКА У НАДЗВИЧАЙНИХ СИТУАЦІЯХ

Охорона праці та дотримання техніки безпеки є невід'ємними складовими роботи з автоматизованими системами, що базуються на моделях машинного навчання для ідентифікації пошкоджень дорожнього покриття. Безпека працівників, які займаються розробкою, впровадженням і експлуатацією таких систем, вимагає врахування специфіки роботи із цифровим обладнанням, програмними засобами, а також потенційних ризиків у польових умовах.

Під час роботи з обладнанням для збору даних, наприклад, мобільними пристроями, дронами чи автомобілями, оснащеними датчиками та камерами, важливо дотримуватися правил техніки безпеки. Працівники повинні носити спеціальний одяг із світловідбиваючими елементами, щоб бути помітними на дорозі, особливо в умовах недостатньої видимості. Перед початком роботи слід провести інструктаж із безпеки, зокрема щодо дій у разі виникнення надзвичайних ситуацій, таких як аварії чи погіршення погодних умов.

Робота з технічним обладнанням для збору даних потребує уважності та ретельного огляду пристроїв перед початком експлуатації. Усі елементи обладнання, включно з камерами, лазерними сканерами, датчиками та акумуляторами, мають бути перевірені на справність. Важливо уникати контакту обладнання з вологою, пилом чи іншими шкідливими факторами, які можуть спричинити його поломку. При підготовці техніки до роботи необхідно використовувати тільки сертифіковані кабелі та зарядні пристрої, щоб запобігти короткому замиканню або перегріванню компонентів.

У польових умовах особливу увагу слід приділити розташуванню обладнання. Якщо для збору даних використовується автомобіль, оснащений системою збору інформації, водій і оператор повинні дотримуватися правил дорожнього руху. Переміщення транспортного засобу під час збору даних має здійснюватися на безпечній швидкості, щоб уникнути аварійних ситуацій. При

зупинці автомобіля необхідно встановити сигнальні знаки та увімкнути аварійну сигналізацію.

Під час використання дронів необхідно дотримуватися правил, встановлених для безпілотних літальних апаратів. Оператор дрона повинен пройти відповідне навчання і володіти необхідними сертифікатами. Політ дрона повинен виконуватися лише у встановлених зонах, без наближення до ліній електропередач, дерев чи інших потенційно небезпечних об'єктів. Перед кожним запуском необхідно перевірити заряд акумуляторів дрона та справність його систем навігації.

При роботі з комп'ютерними системами для обробки даних і тренування моделей машинного навчання важливо дотримуватися стандартів ергономіки. Робоче місце оператора повинно бути обладнане таким чином, щоб запобігти тривалому перебуванню в статичній позі. Монітор слід розташовувати на рівні очей, а освітлення в приміщенні має бути достатнім і не спричиняти відблисків на екрані. Рекомендується дотримуватися режиму праці й відпочинку: після кожної години роботи робити коротку перерву.

Під час роботи із серверним обладнанням чи комп'ютерами, які обробляють великі обсяги даних, необхідно забезпечити їх правильне охолодження. Перегрів систем може спричинити збій у роботі обладнання або навіть пожежу. У приміщенні, де розташовані сервери, слід дотримуватися температурного режиму, встановленого виробником обладнання, а також мати справні засоби пожежогасіння.

Особливу увагу необхідно приділяти безпеці при роботі з великими обсягами даних. Усі працівники повинні дотримуватися норм кібербезпеки: використовувати надійні паролі, уникати підключення до незахищених мереж і не надавати доступ до даних стороннім особам. У разі виявлення витоку інформації слід негайно повідомити відповідальних осіб і вжити заходів для мінімізації ризиків.

Робота з системами машинного навчання передбачає взаємодію з потужними обчислювальними ресурсами, що може викликати підвищене

електромагнітне випромінювання. Для захисту працівників необхідно дотримуватися встановлених норм перебування в зоні дії випромінювання. У приміщеннях, де встановлені такі системи, повинна бути організована відповідна вентиляція.

Усі працівники, які залучені до роботи з системами ідентифікації пошкоджень дорожнього покриття, повинні регулярно проходити медичні огляди та навчання з охорони праці. Це дозволяє знизити ризик виникнення професійних захворювань і підвищити рівень безпеки під час виконання робіт.

В разі виникнення надзвичайної ситуації працівники, які обслуговують систему, повинні діяти відповідно до визначеного порядку, що забезпечує їхню безпеку та мінімізує можливі втрати обладнання або даних.

Якщо під час роботи в польових умовах виникає небезпека для життя або здоров'я працівників (наприклад, аварія на дорозі, несприятливі погодні умови, загроза зіткнення чи інші події), необхідно негайно припинити всі роботи. Персонал повинен покинути небезпечну зону, дотримуючись правил безпеки, і знайти укриття. У разі аварії слід повідомити відповідні служби (поліцію, швидку допомогу) та забезпечити захист місця події до їх прибуття.

У разі виявлення несправностей технічного обладнання, таких як дронів, камер чи датчиків, необхідно зупинити їхню експлуатацію. Працівники повинні відключити несправні пристрої від джерела живлення, забезпечити їх зберігання у безпечному місці та повідомити про інцидент відповідальну особу чи службу технічного обслуговування. У разі можливості усунення несправності на місці, необхідно дотримуватися інструкцій з ремонту та використовувати лише сертифіковані інструменти.

При загрозі втрати зібраних даних або збою в роботі програмного забезпечення важливо оперативно створити резервну копію всіх наявних даних. Якщо система виявляє аномалії у функціонуванні (наприклад, некоректне зчитування даних або помилки в обробці), слід зупинити обробку і повідомити команду розробників. Роботу із системою можна продовжувати тільки після її детальної перевірки та усунення проблем.

Якщо в приміщенні з комп'ютерним обладнанням чи серверами виникає загроза пожежі, працівники повинні негайно відключити електроживлення та скористатися засобами пожежогасіння, якщо це можливо без загрози для життя. У разі значного поширення вогню слід негайно покинути приміщення, повідомити рятувальну службу та дотримуватися інструкцій з евакуації.

У разі надзвичайних погодних умов, таких як сильні дощі, снігопади або урагани, що можуть вплинути на роботу обладнання, необхідно припинити роботу, захистити технічні пристрої від впливу навколишнього середовища та перенести їх до безпечного місця. Особливо це стосується обладнання, яке використовується на відкритих територіях.

Після завершення надзвичайної ситуації всі пристрої, які використовувалися, мають пройти перевірку на справність. Зібрані дані слід проаналізувати на предмет втрат або пошкоджень. Відповідальні працівники повинні скласти звіт про надзвичайну ситуацію, описати вжиті заходи та надати рекомендації для уникнення подібних інцидентів у майбутньому.

Дотримання цього порядку дій забезпечує ефективне реагування на надзвичайні ситуації, зменшує ризики для життя та здоров'я працівників, а також сприяє збереженню обладнання та даних.

РОЗДІЛ 5

ВИЗНАЧЕННЯ ЕФЕКТИВНОСТІ ІНФОРМАЦІЙНОЇ СИСТЕМИ

Для оцінки економічної ефективності впровадження системи машинного навчання для ідентифікації пошкоджень дорожнього покриття необхідно врахувати загальні витрати на створення та експлуатацію системи, а також очікувану економію завдяки її використанню.

Першим кроком визначають витрати. Загальні витрати на впровадження системи складаються з:

- вартості розробки програмного забезпечення $C_{\text{розр}}$;
- витрат на технічне обладнання $C_{\text{техн}}$;
- витрат на навчання персоналу $C_{\text{навч}}$;
- експлуатаційних витрат $C_{\text{експл}}$.

Сумарні витрати визначаються за формулою:

$$C_{\text{заг}} = C_{\text{розр}} + C_{\text{техн}} + C_{\text{навч}} + C_{\text{експл}} .$$

Наприклад, якщо:

- $C_{\text{розр}} = 1\,000\,000$ грн,
- $C_{\text{техн}} = 500\,000$ грн,
- $C_{\text{навч}} = 100\,000$ грн,
- $C_{\text{експл}} = 300\,000$ грн,

тоді:

$$C_{\text{заг}} = 1\,000\,000 + 500\,000 + 100\,000 + 300\,000 = 1\,900\,000 \text{ грн.}$$

На другому кроці проводять оцінку можливої економії. Нехай економія завдяки впровадженню системи включає:

- зменшення витрат на ремонт доріг $E_{\text{рем}}$;
- зниження аварійності $E_{\text{авар}}$;

- зменшення соціальних витрат $E_{\text{соц}}$.

Загальна економія визначається як:

$$E_{\text{заг}} = E_{\text{рем}} + E_{\text{авар}} + E_{\text{соц}} .$$

Припустимо, що:

- $E_{\text{рем}} = 1\,200\,000$ грн,

- $E_{\text{авар}} = 800\,000$ грн,

- $E_{\text{соц}} = 500\,000$ грн,

тоді:

$$E_{\text{заг}} = 1\,200\,000 + 800\,000 + 500\,000 = 2\,500\,000 \text{ грн.}$$

Третій крок - розрахунок коефіцієнта економічної ефективності. Даний коефіцієнт економічної ефективності системи визначається як:

$$K_{\text{еф}} = \frac{E_{\text{заг}}}{C_{\text{заг}}} .$$

Підставляючи значення:

$$K_{\text{еф}} = \frac{2\,500\,000}{1\,900\,000} \approx 1,32.$$

Це означає, що кожна вкладена гривня приносить 1,32 грн економії.

Наступний крок - розрахунок терміну окупності. Як відомо, термін окупності інвестицій у систему зазвичай розраховується за формулою:

$$T_{\text{окуп}} = \frac{C_{\text{заг}}}{E_{\text{заг}}} .$$

Для нашого прикладу:

$$T_{\text{окуп}} = \frac{1\,900\,000}{2\,500\,000} \approx 0,76 \text{ року.}$$

Тобто система окупиться менше ніж за рік.

І нарешті, можна оцінити соціально-економічну ефективність. Для врахування непрямой вигоди використовується показник соціально-економічної ефективності:

$$K_{\text{соц-еф}} = \frac{E_{\text{зар}} + E_{\text{соц}}}{C_{\text{зар}}}.$$

У нашому випадку:

$$K_{\text{соц-еф}} = \frac{2\,500\,000 + 500\,000}{1\,900\,000} \approx 1,58.$$

Це підкреслює значний вплив системи на підвищення безпеки та ефективності використання ресурсів.

Таким чином, розрахунки демонструють високу економічну та соціальну ефективність впровадження системи. Завдяки цьому досягається суттєва економія коштів, скорочуються соціальні витрати, пов'язані з аварійністю, і забезпечується окупність системи у короткі терміни.

ВИСНОВКИ ТА ПРОПОЗИЦІЇ

В ході виконання кваліфікаційної роботи була розроблена модель машинного навчання для ідентифікації пошкоджень дорожнього покриття. В ході виконання роботи вирішені такі задачі:

1. Проведено аналіз предметної області, визначені фактори, які впливають на якість дороги та сформульовані задачі роботи.

2. Розглянуті методи та обмеження діагностики стану автомобільних доріг.

3. Розглянуті деякі математичні моделі, які використовуються в процесі моделювання вібрацій автомобіля під час руху та визначено, що для обробки даних слід використовувати методи штучного інтелекту.

4. Визначено клас задачі (задача кластеризації) та намічені шляхи її вирішення.

5. Обрані мова програмування, середовище розробки, бібліотеки машинного навчання та джерела даних.

6. Розроблена модель машинного навчання, проведено моделювання, отримані результати обчислень та зроблена їх оцінка.

7. Розглянуті варіанти архітектурних рішень по системі, потокам даних та структурі бази даних.

8. Описані питання охорони праці та визначення економічної ефективності отриманих рішень.

9. По результатам роботи опубліковані тези доповіді на студентській науковій конференції [35].

Подальший розвиток досліджень може бути направлений на використання інших методів машинного навчання для ідентифікації пошкоджень дорожнього покриття.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Шиба О.А. Вплив розвитку транспортної інфраструктури на економічне зростання країн-членів Європейського союзу. URL: https://www.lnu.edu.ua/wp-content/uploads/2017/05/aref_shyba.pdf
2. Розвиток транспортної інфраструктури як фактор підвищення глобальної конкурентоспроможності. URL: <https://doi.org/10.32782/2524-0072/2024-60-108>
3. Гринь А., Толок П. Вплив транспортної інфраструктури на економічний розвиток регіонів, оцінка ефективності управління дорожнім господарством. URL: <https://doi.org/10.31891/2307-5740-2024-334-13>
4. Мороз Т. Аналіз методів діагностування якісних показників стану автомобільних доріг. URL: <http://dx.doi.org/10.33744/0365-8171-2022-112-093-103>
5. Інтерактивний моніторинг автомобільних доріг : монографія / В. О. Алексієв, О. П. Алексієв, А. А. Видмиш, В. О. Хабаров. – Вінниця : ВНТУ, 2012. – 144 с.
6. Технічні правила ремонту та утримання автомобільних доріг загального користування України. URL: <https://ips.ligazakon.net/document/FIN4934>
7. Транспортно-експлуатаційні якості автомобільних доріг та міських вулиць : навчальний посібник / Кашканов А. А., Кашканов В. А., Кужель В. П. – Вінниця : ВНТУ, 2018. – 113 с.
8. Пивоварова К.Б. Інтерактивна оцінка та діагностика стану автомобільних доріг. URL: <https://dspace.khadi.kharkov.ua/items/f7fd5a4a-0ff3-4397-8c8f-656ce5040e8f>
9. Положення про Державну службу автомобільних доріг України. URL: <https://www.kmu.gov.ua/npas/75384202>
10. Звіт Державного агентства автомобільних доріг України 2022. URL: <https://restoration.gov.ua/4489/zvity/54474/54475.pdf>
11. Соколова Н.М. Застосування методів аналітичного ієрархічного процесу у кваліметричній моделі оцінювання вартості автомобільних доріг. URL: <https://nauka.com.ua/index.php/ee/article/view/441/444>

12. Білоцерковець С.А. Інтелектуальна технологія моніторингу функціонального стану доріг та узбіч. URL: https://essuir.sumdu.edu.ua/bitstream-download/123456789/86722/1/Bilotserkovets_mag_rob.pdf
13. Чебанова О.П., Волохов В.А. Використання технологій машинного навчання для оптимізації логістики. URL: <https://doi.org/10.18664/btie.83.300406>
14. Аналіз методів, які використовуються при обґрунтуванні реконструкції автомобільних доріг. URL: <https://nidi.org.ua/ua/analiz-metodiv-yaki-vikoristovuyutysya-pri-obruntuvanni-rekonstrukciyi-avtomobilynih-dorig>
15. Щодо удосконалення механізмів розвитку дорожньої галузі в контексті реалізації реформи державного управління автомобільними дорогами. URL: <https://www.niss.gov.ua/doslidzhennya/ekonomika/schodo-udoskonalennya-mekhanizmiv-rozvitku-dorozhnoi-galuzi-v-konteksti>
16. Потеєва М.А. Удосконалення механізму державного регулювання транспортної інфраструктури України. URL: <http://dspace.nbuu.gov.ua/bitstream/handle/123456789/12027/17-Poteeva.pdf>
17. Лагода А.В. Аналіз принципів використання технології Інтернету речей в керуванні дорожнім рухом в рамках концепції “Smart City”. URL: <https://ela.kpi.ua/server/api/core/bitstreams/a1af55f7-74f0-46bd-ae85-224837d64cf8/content>
18. Інтелектуалізація електронних систем та програмного забезпечення сучасних автомобілів. URL: <https://doi.org/10.31649/2413-4503-2022-16-2-88-95>
19. Аналіз функціональних можливостей сенсорних дисплеїв в інформаційних системах транспортних засобів. URL: <https://doi.org/10.30977/VEIT.2020.17.0.42>
20. Інформаційні системи і технології в транспортній логістиці: навч. посібник. URL: https://document.kdu.edu.ua/monogr/2022_128.pdf
21. Проектування бездротових мереж для рухомих об’єктів. URL: https://repo.btu.kharkov.ua/bitstream/123456789/11969/1/22_196_2018.pdf
22. Основи експлуатації автомобільних доріг і аеродромів : навч. посіб. / В. С. Степура, А. О. Белятинський, Н. В. Кужель. — К. : НАУ, 2013. — 204 с.

23. Класифікація методів діагностики стану дорожнього полотна. URL: <https://api.dspace.khadi.kharkov.ua/server/api/core/bitstreams/83a2f4ea-798e-44ea-b12f-9a27a024f0e3/content>
24. Підвіска автомобіля. URL: https://uk.wikipedia.org/wiki/Підвіска_автомобіля
25. Динаміка й оптимізація машин / В.С. Ловейкін, Ю.О. Ромасевич, Р.А. Кульпін. – К.: ЦП «Компринт», 2018. - 310 с.
26. Introduction to Vehicle dynamics and suspension systems. URL: <https://www.school-mechademic.com/blog/vehicle-dynamics>
27. Нова інформаційна система для класифікації стану дорожнього покриття за допомогою методів машинного навчання та паралельних обчислень. URL: <https://doi.org/10.31891/csit-2023-1-7>
28. Моніторинг стану дорожнього покриття через мобільний застосунок. URL: <https://dspace.khadi.kharkov.ua/items/d4ff55f0-bd51-4887-b5e3-95b5a7eed2cf>
29. Road Condition Monitoring Using Smart Sensing and Artificial Intelligence: A Review. URL: <https://doi.org/10.3390/s22083044>
30. Vibration Meter - Apps on Google Play. URL: <https://play.google.com/store/apps/details?id=com.exatools.vibrometer>
31. Найкращі мови програмування для штучного інтелекту. URL: <https://pnn.com.ua/ua/blog/detail/top-programming-languages-for-artificial-intelligence>
32. OpenStreetMap. URL: <https://www.openstreetmap.org/>
33. Конспект лекцій з дисципліни «Робочі процеси автомобілів» для студентів спеціальності 133 «Галузеве машинобудування». Системи управління і підвіска. URL: <https://eir.zp.edu.ua/server/api/core/bitstreams/6d7ddddb-a74b-46bd-8db1-ec7fd9ed4732/content>
34. A Big Data System Architecture to Support the Monitoring of Paved Roads. URL: <https://doi.org/10.3390/infrastructures8120167>
35. Сулятицький М. Особливості моделей машинного навчання для ідентифікації пошкоджень дорожнього покриття. URL: https://lnup.edu.ua/attachments/article/8156/СТУД.ФОРУМ_2024.pdf#page=370