

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Західноукраїнський національний університет
Факультет комп'ютерних інформаційних технологій
Кафедра інформаційно-обчислювальних систем і управління

ШАЙНЮК Вадим Олександрович

**Метод прогнозування транспортних потоків за допомогою
Інтернету речей та машинного навчання / Method for
Predicting Traffic Flows Using the Internet of Things and
Machine Learning**

спеціальність: 122 - Комп'ютерні науки
освітньо-професійна програма - Комп'ютерні науки

Кваліфікаційна робота

Виконав студент групи КНм-21
В.О. Шайнюк

Науковий керівник:
к.т.н., професор В.В. Кочан

Кваліфікаційну роботу
допущено до захисту:
«___» _____ 20___ р.
В.о. завідувача кафедри
_____ Н.В. Дзюбановська

ТЕРНОПІЛЬ – 2025

Факультет комп'ютерних інформаційних технологій
Кафедра інформаційно-обчислювальних систем і управління
Освітній ступінь «магістр»
спеціальність: 122 – Комп'ютерні науки
освітньо-професійна програма – Комп'ютерні науки

ЗАТВЕРДЖУЮ
В.о. завідувача кафедри
Н.М. Васильків
« ____ » _____ 20__ р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ СТУДЕНТУ
ШАЙНЮКА Вадима Олександровича

(прізвище, ім'я, по батькові)

1. Тема кваліфікаційної роботи

Метод прогнозування транспортних потоків за допомогою Інтернету речей та машинного навчання / Method for Predicting Traffic Flows Using the Internet of Things and Machine Learning

керівник роботи к.т.н., професор В.В. Кочан

затверджені наказом по університету від 20 грудня 2024 року № 938.

2. Строк подання студентом закінченої кваліфікаційної роботи 1 грудня 2025 р.

3. Вихідні дані до кваліфікаційної роботи: завдання на кваліфікаційну роботу студента, наукові статті, технічна література.

4. Основні питання, які потрібно розробити

– проаналізувати предметну область керування транспортними потоками та інтелектуальних транспортних систем;

– виконати огляд сучасних підходів до прогнозування транспортних потоків, що базуються на використанні Інтернету речей, бездротових сенсорних мереж, технологій великих даних, методів машинного та глибинного навчання;

– сформулювати концепцію та архітектуру системи збору й обробки даних для прогнозування транспортних потоків, визначити джерела IoT-даних, структуру сховища даних та інформаційні потоки між компонентами;

– розробити методіку попередньої обробки та відбору ознак транспортних даних, що включає очищення, виявлення викидів, нормалізацію, а також застосування процедур вибору інформативних ознак з метою підвищення точності прогнозу та зменшення похибки;

– синтезувати та налаштувати моделі машинного навчання для прогнозування транспортних потоків на основі сформованого набору ознак, обґрунтувати вибір алгоритмів;

– провести експериментальні дослідження розробленого методу на реальному або відкритому наборі транспортних даних, виконати порівняльний аналіз якості прогнозування з базовими підходами.

5. Перелік графічного матеріалу у роботі

– архітектура системи прогнозування транспортного потоку на основі IoT та методів машинного навчання;

РЕЗЮМЕ

Кваліфікаційна робота на тему «Метод прогнозування транспортних потоків за допомогою Інтернету речей та машинного навчання» на здобуття освітнього ступеня «Магістр» зі спеціальності 122 «Комп'ютерні науки» освітньої програми «Комп'ютерні науки» написана обсягом в 70 сторінок і містить 4 ілюстрації, 2 таблиці, 1 додаток та 35 використаних джерел.

Метою кваліфікаційної роботи є розробка методу прогнозування транспортних потоків на основі даних Інтернету речей та машинного навчання, який забезпечує підвищення точності та стійкості прогнозів за рахунок інтеграції сенсорних даних, процедур попередньої обробки та відбору інформативних ознак.

Методи досліджень: аналіз і узагальнення наукових джерел, методи системного аналізу та структурного моделювання, апарат теорії ймовірностей і математичної статистики, методи попередньої обробки й інтелектуального аналізу даних, а також методи машинного навчання та експериментальне моделювання.

Результати дослідження: вдосконалено метод прогнозування транспортних потоків на основі даних Інтернету речей шляхом інтеграції спеціалізованої процедури попередньої обробки та відбору ознак на основі алгоритму рою частинок із моделями машинного навчання.

Результати роботи можуть застосовуватися для побудови та впровадження інтелектуальних систем прогнозування транспортних потоків у складі інтелектуальних транспортних систем та сервісів «розумного міста». Запропонований метод може бути використаний для адаптивного керування дорожнім рухом, оптимізації режимів роботи світлофорних об'єктів, планування маршрутів громадського транспорту та ін.

Ключові слова: ІНТЕЛЕКТУАЛЬНА ТРАНСПОРТНА СИСТЕМА; ПРОГНОЗУВАННЯ ТРАНСПОРТНИХ ПОТОКІВ; ІНТЕРНЕТ РЕЧЕЙ; РОЗУМНЕ МІСТО; МАШИННЕ НАВЧАННЯ; НЕЙРОННІ МЕРЕЖІ; АДАПТИВНЕ КЕРУВАННЯ РУХОМ.

ABSTRACT

Qualification work on the topic «Method for Predicting Traffic Flows Using the Internet of Things and Machine Learning» for Master's degree on speciality 122 «Computer Science» educational and professional program «Computer Science» is written on 70 pages and it contains 4 figures, 2 tables, 1 annex and 35 sources.

The purpose of this qualification work is to develop a method for traffic flow prediction based on Internet of Things data and machine learning, which improves the accuracy and robustness of forecasts by integrating sensor data, preprocessing procedures, and informative feature selection.

Research methods include the analysis and synthesis of scientific sources, methods of systems analysis and structural modelling, the apparatus of probability theory and mathematical statistics, methods of data preprocessing and intelligent data analysis, as well as machine learning methods and experimental modelling.

The research results consist in improving the method of traffic flow prediction based on Internet of Things data by integrating a specialized preprocessing and feature selection procedure based on the particle swarm optimization algorithm with machine learning models.

The results of this work can be used for the development and implementation of intelligent traffic flow prediction systems as part of intelligent transportation systems and smart city services. The proposed method can be applied for adaptive traffic control, optimization of traffic light operation modes, public transport route planning, etc.

Keywords: INTELLIGENT TRANSPORTATION SYSTEM; TRAFFIC FLOW PREDICTION; INTERNET OF THINGS; SMART CITY; MACHINE LEARNING; NEURAL NETWORKS; ADAPTIVE TRAFFIC CONTROL.

ЗМІСТ

Вступ.....	7
1 Аналіз предметної області та існуючих рішень.....	11
1.1 Опис предметної області.....	11
1.2 Аналіз існуючих рішень.....	14
1.3 Постановка задачі дослідження.....	20
Висновки до розділу 1.....	23
2 Розробка методу прогнозування транспортних потоків на основі IoT та машинного навчання.....	24
2.1 Концепція та архітектура системи збору й обробки даних.....	24
2.2 Попередня обробка та відбір ознак транспортних даних.....	29
2.3 Синтез і налаштування моделей машинного навчання для прогнозування транспортних потоків.....	31
2.4 Формалізація методу прогнозування транспортних потоків.....	34
Висновки до розділу 2.....	38
3 Програмна реалізація та експериментальні дослідження.....	39
3.1 Дослідження набору даних та постановка експериментів.....	39
3.2 Порівняльний аналіз результатів класифікаційних моделей.....	41
3.3 Порівняння із відомими підходами.....	44
Висновки до розділу 3.....	49
Висновки.....	50
Список використаних джерел.....	52
Додаток А Копії публікацій.....	56

ВСТУП

Актуальність роботи. Проблема транспортних заторів є одним із ключових викликів сучасних міст. Під затором зазвичай розуміють стан руху, за якого швидкість транспортних засобів помітно знижується, черги авто зростають, а фактичний час поїздки суттєво перевищує очікуваний. Починаючи з 2000-х років, завантаженість дорожньої мережі стабільно зростає внаслідок збільшення кількості приватних авто та нерівномірного розвитку інфраструктури й громадського транспорту [1]. Якщо якість і доступність громадських перевезень є низькими, мешканці віддають перевагу власному транспорту, що ще більше перевантажує мережу, посилює затори в години пік і збільшує втрати часу. У країнах, що розвиваються, ці процеси додатково ускладнюються відставанням розвитку дорожньої інфраструктури від темпів урбанізації та моторизації [1].

Затори мають виразну безпекову, економічну та екологічну складові. Вони підвищують імовірність дорожньо-транспортних пригод, призводять до пошкодження транспорту, травмування учасників руху й іноді до летальних наслідків. Водії та пасажирів зазнають втрат часу, збільшується споживання пального та викиди шкідливих речовин, погіршується якість життя в урбанізованих зонах. Тому управління транспортними потоками та зниження рівня заторів слід розглядати не лише як інженерне, а як комплексне соціально-економічне завдання.

Одним із базових інструментів підтримки рішень у цій сфері є прогнозування транспортних потоків. Його метою є оцінка майбутніх змін у використанні різних видів транспорту, інтенсивності руху на окремих ділянках та завантаженості маршрутів. Вихідним етапом виступає збір даних про поточний стан руху (інтенсивність, швидкість, затримки, тривалість поїздок) і супровідні параметри середовища. На основі цих даних формуються набори ознак, які можуть включати демографічні характеристики, структуру зайнятості, тарифи, вартість пального, показники економічної активності та метеорологічні умови [2].

Традиційно транспортні моделі спиралися на дані з окремих сенсорів – індукційних петель, фотоелектричних та п'єзоелектричних датчиків, установлених уздовж доріг [2]. Такі системи давали змогу формувати часові ряди інтенсивності та будувати короткострокові прогнози. З поширенням методів машинного навчання стало можливим враховувати складні нелінійні залежності, взаємодії між параметрами трафіку та інформацію з альтернативних маршрутів, що підвищило точність прогнозів.

Сучасний розвиток транспортної галузі пов'язаний із широким впровадженням Інтернету речей (IoT) та інтелектуальних транспортних систем (ITS). Масове розгортання датчиків у дорожній інфраструктурі та транспортних засобах забезпечує безпрецедентні обсяги даних, які оновлюються майже в реальному часі й охоплюють позицію, швидкість, тип транспортного засобу, стан покриття, погоду тощо. Поєднання IoT-інфраструктури з алгоритмами машинного навчання, на нашу думку, відкриває якісно нові можливості для побудови прогнозних моделей транспортних потоків.

Інтеграція даних з різнорідних джерел (сенсори, навігаційні сервіси, метеодані, соціально-економічні показники) в хмарних середовищах дозволяє перейти від «точкових» вимірювань до цілісної картини функціонування мережі. Моделі машинного навчання – регресія, дерева рішень, ансамблі, нейронні мережі – здатні відображати стохастичний і нелінійний характер транспортних процесів, працювати в режимі, близькому до реального часу, оцінювати ймовірність виникнення заторів та ефект від зміни режимів керування рухом. Водночас актуальними залишаються проблеми відбору інформативних ознак у великих IoT-датасетах, чутливості моделей до шумів і пропусків, узгодження різнорідних джерел та обмежень обчислювальних ресурсів.

Отже, постає науково-практичне завдання розробити метод прогнозування транспортних потоків на основі даних Інтернету речей та алгоритмів машинного навчання, який забезпечуватиме підвищену точність прогнозів завдяки інтеграції динамічних сенсорних даних, спеціалізованих процедур обробки й відбору ознак та адаптації моделі до змінних умов функціонування транспортної системи.

Мета і завдання дослідження. Метою роботи є розробка методу

прогнозування транспортних потоків на основі даних Інтернету речей та машинного навчання, який забезпечує підвищення точності та стійкості прогнозів за рахунок інтеграції сенсорних даних, процедур попередньої обробки та відбору інформативних ознак.

Для досягнення поставленої мети в роботі необхідно розв'язати такі основні завдання:

1. Проаналізувати предметну область керування транспортними потоками та інтелектуальних транспортних систем.

2. Виконати огляд сучасних підходів до прогнозування транспортних потоків, що базуються на використанні Інтернету речей, бездротових сенсорних мереж, технологій великих даних, методів машинного та глибинного навчання, а також виявити їх переваги та недоліки.

3. Сформуванати концепцію та архітектуру системи збору й обробки даних для прогнозування транспортних потоків, визначити джерела IoT-даних, структуру сховища даних та інформаційні потоки між компонентами.

4. Розробити методику попередньої обробки та відбору ознак транспортних даних, що включає очищення, виявлення викидів, нормалізацію, а також застосування процедур вибору інформативних ознак з метою підвищення точності прогнозу та зменшення похибки.

5. Синтезувати та налаштувати моделі машинного навчання для прогнозування транспортних потоків на основі сформованого набору ознак, обґрунтувати вибір алгоритмів.

6. Провести експериментальні дослідження розробленого методу на реальному або відкритому наборі транспортних даних, виконати порівняльний аналіз якості прогнозування з базовими підходами за допомогою відповідних метрик.

Об'єкт дослідження – процеси формування та динаміки транспортних потоків в умовах функціонування інтелектуальних транспортних систем на базі Інтернету речей.

Предмет дослідження – методи та моделі прогнозування транспортних потоків на основі даних, отриманих від IoT-пристроїв, зокрема архітектура

системи збору й обробки даних, процедури попередньої обробки та відбору ознак і алгоритми машинного навчання для прогнозування стану трафіку.

Методи дослідження. У роботі застосовано аналіз і узагальнення наукових джерел для визначення сучасного стану проблеми, методи системного аналізу та структурного моделювання для побудови архітектури IoT-орієнтованої системи, апарат теорії ймовірностей і математичної статистики для формалізації задачі прогнозування та оцінювання якості, методи попередньої обробки й інтелектуального аналізу даних, а також методи машинного навчання та експериментальне моделювання з порівняльним аналізом результатів.

Наукова новизна одержаних результатів полягає у вдосконаленні методу прогнозування транспортних потоків на основі даних Інтернету речей шляхом інтеграції спеціалізованої процедури попередньої обробки та відбору ознак на основі алгоритму рою частинок із моделями машинного навчання.

Практичне значення одержаних результатів полягає у можливості використання запропонованого методу прогнозування транспортних потоків як складової інтелектуальних транспортних систем та сервісів «розумного міста» для підвищення ефективності управління дорожнім рухом. Розроблена IoT-орієнтована архітектура та методика обробки даних дають змогу інтегрувати сенсорні мережі, бортові пристрої транспортних засобів і хмарну аналітику в єдину систему, здатну в режимі, наближеному до реального часу, оцінювати стан завантаженості мережі та формувати прогнози інтенсивності руху.

Публікації та апробація КР. Результати кваліфікаційної роботи апробовані та опубліковані у матеріалах (додаток А):

- II Всеукраїнської науково-практичної конференції «Інтелектуальні комп'ютерні системи та мережі», 25 листопада 2025 р., Тернопіль, Україна;
- 2nd International Scientific and Practical Conference «Progressive Approaches in Science and Engineering», November 26-28, 2025. Copenhagen, Denmark.

Кваліфікаційна робота складається із вступу, трьох розділів, висновків, списку використаних джерел та додатків.

1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ТА ІСНУЮЧИХ РІШЕНЬ

1.1 Опис предметної області

Проблема транспортних заторів належить до ключових викликів сучасних урбанізованих територій. Під затором зазвичай розуміють стан руху, коли швидкість транспортних засобів істотно знижується, формуються черги, а фактичний час поїздки значно перевищує нормативний [1]. Серед основних причин – надмірна кількість приватних автомобілів у поєднанні з недостатнім розвитком громадського транспорту. Якщо автобуси, трамваї та інші види транспорту працюють із перебоями, мають низький рівень сервісу або незручні маршрути, користувачі масово переходять на індивідуальний транспорт, що ще більше перевантажує вулично-дорожню мережу.

Особливо гостро це проявляється у країнах, що розвиваються, де темпи моторизації населення випереджають модернізацію інфраструктури. Обмежена кількість смуг, вузькі історичні вулиці, застарілі перехрестя, брак розв'язок та об'їзних шляхів формують «вузькі місця», де затори виникають хронічно [1]. Обмежений і нерівномірний парк громадського транспорту погіршує доступність перевезень і стимулює використання власного авто.

Наслідки заторів мають безпековий, економічний та екологічний вимір. За високої щільності потоку будь-який неузгоджений маневр може спровокувати аварійну ситуацію, дрібні зіткнення чи навіть летальні події [1]. Водії та пасажери втрачають значні обсяги часу, що трансформується у прямі та непрямі економічні збитки. Режим «stop-and-go» підвищує викиди забруднювальних речовин та рівень шуму, погіршуючи якість життя в містах.

На цьому тлі прогнозування транспортних потоків стає базовим інструментом підтримки рішень у транспортному плануванні та управлінні. Першим кроком є збір даних про поточний стан руху: інтенсивність, середню швидкість, завантаженість окремих ділянок, тривалість поїздок, а також супровідні параметри [2]. На основі цих даних формуються набори ознак, що можуть включати демографічні характеристики, рівень автомобілізації, тарифи, витрати на поїздки, показники економічної активності тощо.

У загальному вигляді мета транспортного прогнозування полягає у визначенні того, як і з якою інтенсивністю у майбутньому будуть використовуватися окремі види транспорту та елементи мережі [2]. Сучасні методи дають змогу враховувати вплив зовнішніх чинників (економічні, кліматичні, містобудівні зміни), працювати з потоками даних у реальному часі та будувати більш гнучкі, нелінійні моделі. Прогнози застосовуються для планування маршрутів і графіків громадського транспорту, проєктування та реконструкції магістралей, оцінки пропускнуої здатності перехресть, аналізу впливу інфраструктурних проєктів на довкілля й якість повітря [2].

Для опису динаміки транспортної системи дані про поточний трафік здебільшого поєднуються з демографічними, метеорологічними та економічними показниками [2]. Така інтеграція дозволяє отримати більш повну картину чинників, що формують потоки. Історично оцінювання обсягів руху ґрунтувалося на стаціонарних сенсорах (індукційні петлі, фотоелектричні та п'єзоелектричні датчики), розташованих на окремих ділянках доріг [2]. Отримані часові ряди інтенсивності використовувалися у класичних моделях. З розвитком методів машинного навчання стало можливим поєднувати поточні сенсорні дані з історичною інформацією по альтернативних маршрутах та будувати більш точні прогнози.

Запровадження інтелектуальних транспортних систем (Intelligent Transportation Systems, ITS) значно розширило склад доступних даних. До інформаційних потоків ITS належать не лише показники транспортного потоку, а й детальний опис дорожнього середовища: топологія мережі, властивості покриття, обмеження швидкості, режими роботи світлофорів, дані про громадський транспорт, парковки тощо [3, 4]. Водночас зберігається потреба в інтеграції додаткових факторів – складних погодних умов, ремонтних робіт, масових заходів.

Затор, з точки зору користувача, означає втрату часу та зниження комфортності поїздки, а з погляду органів влади – потребу підтримувати мережу у працездатному стані, оперативно реагувати на ДТП, зменшувати вплив транспорту на довкілля та підвищувати безпеку вразливих учасників руху [5].

Вузькі вулиці, дефіцит смуг, неефективні світлофорні плани та неузгоджені схеми руху формують локальні зони підвищеної аварійності [3, 5]. Повністю усунути затори практично неможливо, однак їхній негативний вплив можна суттєво зменшити за рахунок якісного прогнозування та інтелектуального керування потоками: оптимізації режимів світлофорів, перенаправлення транспорту на альтернативні маршрути, адаптації графіків громадського транспорту, завчасного виявлення небезпечних ситуацій.

У сучасних умовах дані перетворюються на ключовий ресурс транспортного управління. Розвиток IoT та IoE призвів до різкого зростання джерел інформації: окрім стаціонарних сенсорів, використовуються радарні системи, GPS-треки транспортних засобів, дані навігаційних сервісів і соціальних мереж [6]. Поєднання цих різномірних джерел дає змогу формувати багатовимірні моделі, що точніше відображають реальний стан мережі.

Центральну роль відіграють бездротові сенсорні мережі (Wireless Sensor Networks, WSN), які складаються з автономних вузлів із сенсорами, обчислювальними та комунікаційними модулями [7]. У транспортній сфері такі мережі розгортаються вздовж магістралей, на мостах, у тунелях, на транспортних вузлах та парковках. Вони забезпечують гнучке розгортання, масштабування та підключення до центральної платформи через шлюзи, точки доступу WLAN та стільникові модулі.

Паралельно розвиваються транспортні мережі зв'язку – VANET, MANET та Internet of Vehicles (IoV), де вузлами виступають транспортні засоби й інфраструктурні елементи [8]. Комунікації V2V і V2I забезпечують обмін даними про стан руху, дорожні умови, аварії. Для IoV пропонуються моделі на основі процесів прийняття рішень Маркова, біоінспіровані алгоритми та підходи теорії ігор, що розв'язують задачі розподілу ресурсів та забезпечення якості сервісу у хмарних обчислювальних середовищах [8].

Концепція Internet of Everything (IoE) розглядає у взаємодії чотири компоненти: людей, «речі» (пристрої), дані та процеси [9]. У транспортній сфері це означає:

– людей, які приймають рішення та взаємодіють із системою (оператори,

водії, пасажирів);

- пристрої, що вимірюють параметри руху та середовища;
- дані, які накопичуються та аналізуються;
- процеси, що регламентують верифікацію, доступ, маршрутизацію й аналітику.

Сюди ж належать онлайн-спільноти та краудсорсингові платформи, де користувачі повідомляють про затори, ДТП та перекриття доріг у режимі, близькому до реального часу. Таким чином, формується «спільнота сенсорів», де кожен користувач доповнює загальну картину транспортної ситуації.

У низці робіт описано підходи до прогнозування транспортних потоків, що інтегрують IoT-інфраструктуру, попередню обробку, відбір ознак і моделі машинного навчання [10]. Узагальнена схема передбачає: збір даних у реальному часі із сенсорів на дорогах і в транспортних засобах; доповнення їх історичними даними; збереження в хмарному сховищі; очищення та виявлення викидів; подальший відбір інформативних ознак, зокрема засобами Particle Swarm Optimization (PSO). Після цього будуються моделі, наприклад K-nearest neighbors (KNN), багатошарові перцептрони (MLP) та байєсівські мережі, які застосовуються для оцінки інтенсивності потоку чи класифікації станів трафіку [10]. Експериментальні результати (зокрема на UCI Traffic Dataset) свідчать, що комбінація PSO + KNN може забезпечувати високу точність прогнозів (порядку 96 %), низькі значення MSE і RMSE [10].

Отже, інтеграція даних Інтернету речей, процедур попередньої обробки, методів відбору ознак та алгоритмів машинного навчання створює методологічну основу для побудови ефективних систем прогнозування транспортних потоків.

1.2 Аналіз існуючих рішень

У наукових дослідженнях, присвячених прогнозуванню транспортних потоків, простежується кілька ключових напрямів: застосування технологій великих даних та інтелектуального аналізу, використання глибоких нейронних

мереж, розроблення гібридних статистико-евристичних моделей, залучення даних соціальних мереж, а також подальший розвиток класичних часових моделей. У цьому підрозділі узагальнено основні підходи, які є релевантними для побудови методу прогнозування транспортних потоків на основі Інтернету речей та машинного навчання.

Одним із напрямів є використання концепцій великих даних у міських інтелектуальних транспортних системах (ITS). Yachao Jia та співавтори запропонували мережеву модель, що спирається на методи data mining та big data для аналізу функціонування міських ITS [10]. Автори акцентують увагу на понятті «адаптивних транспортних даних», які, будучи збагаченими за рахунок великих обсягів та різноманітності джерел, істотно впливають на еволюцію транспортної інфраструктури. Серед ключових переваг використання технологій великих даних для транспорту виділяються: підвищення ефективності систем, можливість побудови прогнозів у режимі реального часу, а також розширення можливостей моніторингу стану мережі. Реалізація моделей моніторингу та прогнозування в реальному часі дозволяє одночасно збирати дані, оцінювати безпеку руху та збільшувати пропускну здатність транспортної системи [10].

Схожий акцент на інтелектуальному аналізі даних спостерігається у роботах Rahul Khokale та колег, де здійснено порівняльний аналіз методів data mining, що застосовуються для прогнозування дорожнього руху [11]. Особливу увагу приділено ролі бездротових сенсорних мереж (WSN) у «розумних містах». Сенсорні мережі розглядаються як основне джерело даних, а технології інтелектуального аналізу використовуються для підтримки рішень, зокрема під час керування світлофорами. Автори демонструють, як кластери та інші алгоритми можуть оптимізувати використання сенсорів у транспортних знаках, а також забезпечувати адаптацію режимів світлофорного регулювання до фактичного стану потоку [11].

Окрему групу робіт становлять дослідження, в яких застосовуються глибинні нейронні мережі для врахування просторово-часових залежностей у транспортних даних. Так, Yuheng Zhou та співавтори пропонують модель, що поєднує CNN і LSTM для прогнозування дорожнього трафіку [12]. У

запропонованому підході одна з компонент (LSTM) використовується для відображення просторових зв'язків, тоді як інша (CNN) – для опису часових закономірностей у даних. До вхідних ознак, крім базових параметрів руху, включаються метеорологічні характеристики за попередні періоди. Додатково застосовується відбір ознак за допомогою Random Forest (RF), що, за повідомленням авторів, дозволило суттєво підвищити точність моделі (на рівні близько 90 %) [12].

Гібридні підходи, які поєднують класичні математичні перетворення й сучасні алгоритми машинного навчання, також посідають помітне місце в літературі. Xianglong Luo та колеги пропонують модель прогнозування на основі поєднання Support Vector Regression (SVR) та методу Density Functional Theory (DFT) [13]. Історичні дані про рух проходять обробку з використанням DFT та порогових значень для виділення характерних компонент, після чого залишкові послідовності прогнозуються за допомогою SVR. Експериментальна перевірка, виконана на даних з автомобільних пунктів оплати проїзду провінції Цзянсу (Китай), продемонструвала перевагу такого підходу над традиційними моделями, особливо в умовах сезонних піків, характерних для періоду літніх відпусток [13].

У низці робіт увага зосереджена на класифікаційних методах машинного навчання, які використовують різні категорії ознак для характеристики маршрутів. Hashemi та співавтори застосували Random Forest (RF), Naïve Bayes Classifier (NBC), CN2 та Classification Trees (CT) для прогнозування стану трафіку, використовуючи множину категоріальних ідентифікаторів [14]. Запропонована система представляє характеристики маршрутів у вигляді набору правил типу «якщо–то», що полегшує інтерпретацію рішень. За результатами порівняння, RF і CT продемонстрували найкращу якість прогнозування, тоді як NBC виявився корисним для виявлення асоціацій між категоріями ознак, а CN2 – для аналізу «стилю» руху на різних маршрутах [14].

Важливий дослідницький вектор пов'язаний із залученням даних соціальних мереж до прогнозування транспортної ситуації. Kovuru Sridevi та колеги аналізують повідомлення із Twitter, пов'язані з дорожньою обстановкою,

та застосовують методи реального часу для оцінки рівня трафіку [15]. Дані збираються, класифікуються за географічним розташуванням, а текст повідомлень піддається аналізу тональності (позитивна, нейтральна, негативна). На основі цих ознак формується модель, яка оцінює транспортну ситуацію з використанням Random Forest. Запропонований підхід досягає точності близько 88 % і має важливу прикладну перевагу – можливість пропонувати альтернативні маршрути з меншим рівнем завантаженості, що безпосередньо підвищує зручність для пасажирів [15]. Цікавою є також виявлена кореляція між кількістю твітів з певного регіону та кількістю користувачів, зацікавлених у пересуванні цією територією, що додатково підтверджує інформативність соціальних даних.

Паралельно з цим проводяться дослідження, у яких порівнюються різні алгоритми data mining на стандартних наборах даних. Poornam Rani та співавтори, використовуючи програмне середовище WEKA, проаналізували продуктивність J48, NBC, дерев рішень (DT) та Random Forest на транспортних даних [16]. Згідно з отриманими результатами, Random Forest не лише забезпечує вищу точність прогнозу, а й демонструє хорошу швидкодію навіть на великих масивах даних, що є важливим фактором для систем реального часу [16].

Окремий клас робіт стосується гібридних моделей, які інтегрують евристичні підходи зі статистичними методами та машинним навчанням. Chunjiao Dong та колеги вивчають можливості поєднання лінійних і нелінійних моделей для прогнозування транспортних потоків, зосереджуючись на Support Vector Regression як інструменті, що здатен врахувати як просторові, так і часові залежності [17]. Показано, що інтеграція різних моделей дозволяє підвищити точність прогнозів у середньому на 9,04 %, причому найбільший ефект спостерігається в умовах високої завантаженості мережі [17].

Xiaohue Yang та співавтори проводять порівняння статистичних моделей (ST, VAR, ARIMA) та моделей машинного навчання (RNN, SVM, MLP) в задачах мультигоризонтного прогнозування, урахування періодичних та пікових значень [18]. Автори пропонують гібридний підхід, у якому трафік розкладається на періодичну (регулярну) і випадкову компоненти. Періодичну складову описують тригонометричними моделями, тоді як залишкові компоненти апроксимуються

статистичними або машинними методами. Експерименти на даних автомагістралі в штаті Міннесота (Highway 394) із застосуванням метрик MAE, RMSE та MAPE показали, що урахування періодичної компоненти, особливо в години пік, суттєво підвищує точність прогнозів, а якість роботи моделі зростає зі збільшенням обсягу даних [18].

Особливий інтерес становлять роботи, де розглядається прогнозування швидкості руху як ключового показника транспортної ефективності. Luis Romo та колеги досліджують три моделі – EGB (Extreme Gradient Boosting), CNN та LSTMNN – на даних транспортної інфраструктури штату Каліфорнія [19]. Показано, що моделі машинного навчання здатні стабільно та з високою точністю прогнозувати швидкість руху в різних сценаріях, що робить їх придатними для інтеграції в системи підтримки прийняття рішень у ITS [19].

У дослідженні Xin Fu та співавторів запропоновано гібридну систему, що поєднує wavelet-перетворення з нейронними мережами [20]. Початкові дані розкладаються на високочастотну випадкову компоненту та низькочастотну трендову. Для моделювання високочастотної частини застосовується ARMA, тоді як для трендової компоненти використовується нейромережева модель. Подальша реконструкція сигналу дає змогу відновити просторово-часові залежності у даних і отримати точніший прогноз. Перевірка на даних міста Нінбо показала високу точність запропонованої системи та її здатність адаптуватися до складних умов руху [20].

Ще один приклад інтеграції штучних нейронних мереж та інтелектуальних агентів наведено в роботі Ioannis Loumiotis та співавторів [21]. Автори розробляють підхід на основі GRNN (General Regression Neural Network) для оцінювання швидкості транспортних засобів та керування режимами роботи світлофорів. Система дозволяє водіям отримувати більш точні прогнози стану трафіку, а керуючим органам – адаптувати сигнальні плани з метою зменшення заторів. Згідно з результатами, запропонований метод забезпечує меншу похибку порівняно з базовими моделями [21].

Нарешті, варто зазначити роботи, де застосовуються класичні часові моделі, зокрема ARIMA. Taghreed Alghamdi та колеги використовують ARIMA

для аналізу факторів, що суттєво впливають на рівень транспортних заторів, працюючи з часовими рядами, сформованими на основі коротких інтервалів спостереження та даних, які не підкоряються нормальному розподілу [22]. Автори демонструють, як за допомогою ARIMA можна оцінювати й прогнозувати несприятливі умови руху, що є корисним для підтримки рішень щодо управління трафіком. Модель тестували на даних із Каліфорнії, при цьому підготовка вибірки здійснювалася в середовищі R. Зроблено висновок про ефективність ARIMA для задач почасового прогнозування інтенсивності потоку [22].

У сучасних IoT-орієнтованих системах моніторингу, зокрема для «розумного міста» та транспортної інфраструктури, роботи В. Кочана [23-27] зосереджуються передусім на якісному сенсорному шарі та надійності вимірювань. Запропоновані рішення охоплюють інтелектуальні вимірювальні системи з компенсацією похибок, оптимізовані АЦП для вбудованих модулів, а також підходи до зменшення енергоспоживання автономних вузлів. Це дозволяє будувати сенсорні мережі, здатні тривалий час працювати в польових умовах, забезпечуючи стабільну точність і метрологічну надійність даних – ключову передумову для подальшого застосування методів машинного навчання.

Окремий напрям становлять рішення для IoT/ІоЕ-та smart-city-систем, де розглядаються архітектури розподілених сенсорних мереж, інтеграція різнорідних вузлів (стаціонарні сенсори, мобільні пристрої, дрони), а також моделі забезпечення надійності й відмовостійкості (multi-version підхід, резервування каналів). У контексті прогнозування транспортних потоків ці підходи можна розглядати як відомі рішення для нижнього рівня системи: вони задають вимоги до точності, енергоефективності та структури IoT-інфраструктури, на базі якої вже будуються аналітичні моделі та алгоритми прогнозування. Таким чином, роботи В. Кочана формують міцний інженерний фундамент для проєктування інтелектуальних транспортних систем, орієнтованих на довготривалий, надійний та масштабований збір даних.

Узагальнюючи, можна зазначити, що сучасні підходи до прогнозування транспортних потоків:

- активно інтегрують IoT/ІоЕ-інфраструктуру та дані з різномірних джерел (сенсори, соціальні мережі, навігаційні платформи);
- використовують як класичні статистичні моделі, так і методи машинного та глибинного навчання, а також їхні гібриди;
- приділяють значну увагу відбору ознак, обробці часових і просторових залежностей та роботі з великими масивами даних у режимі, близькому до реального часу.

Ці напрацювання створюють теоретичне та практичне підґрунтя для розроблення у подальших розділах роботи методу прогнозування транспортних потоків на основі Інтернету речей та машинного навчання з використанням процедур відбору ознак.

1.3 Постановка задачі дослідження

Попри значний прогрес у розвитку інтелектуальних транспортних систем (ITS), наявні рішення мають низку обмежень, які суттєво звужують їхню ефективність у реальних умовах. Ці проблеми пов'язані як із вибором технічної інфраструктури збору даних, так і з обробкою великих масивів інформації та особливостями алгоритмічних підходів.

Однією з базових технологій, на які спираються адаптивні транспортні системи, є бездротові сенсорні мережі (WSN). У більшості існуючих рішень WSN використовуються для вимірювання щільності трафіку, а дані від сенсорів додатково групуються за допомогою кластеризаційних алгоритмів. Однак побудова такої інфраструктури є одночасно критично важливою і вкрай дорогою. Встановлення та обслуговування великої кількості сенсорних вузлів, забезпечення живлення, захист від зовнішніх впливів і підтримка стабільного зв'язку потребують значних фінансових та організаційних ресурсів. До того ж робота таких систем є чутливою до інтенсивності прибуття та вибуття транспортних засобів: різкі зміни потоків можуть призводити до нерівномірного завантаження сенсорної мережі та ускладнювати інтерпретацію результатів.

Інша важлива група проблем стосується управління даними, що

генеруються пристроями ІоЕ/ІоТ у транспортній системі. Сучасні датчики, навігаційні системи, бортові пристрої, мобільні застосунки користувачів формують гігантські обсяги інформації. Для таких потоків потрібна ефективна, масштабована система збирання, зберігання, попередньої обробки та аналітики. Водночас від цих систем очікують швидкого отримання результатів – як у вигляді прогнозів, так і у вигляді рекомендацій для керування рухом. Якщо обробка даних є надто повільною або недостатньо оптимізованою, прогностичні моделі втрачають актуальність, а рішення не встигають впливати на реальний стан трафіку.

Додатковим аспектом є залежність від людського чинника. Там, де рішення про зміну режимів роботи інфраструктури, перенаправлення потоків або інформування користувачів приймаються вручну, неминуче виникають затримки. Для сучасних систем, які працюють у динамічному середовищі, навіть невеликі часові втрати можуть зменшувати ефективність керування. Це підкреслює потребу у створенні таких ІТS, які здатні самостійно приймати рішення на основі моделей машинного навчання, мінімізуючи втручання людини на оперативному рівні.

Окремо варто зазначити проблему неточності маршрутних рекомендацій у популярних навігаційних сервісах під час аварій, дорожніх ремонтів або тимчасових перекриттів. Навіть за наявності великого масиву історичних даних і оновлень від користувачів, сервіси на кшталт Google Maps не завжди встигають коректно врахувати усі зміни в інфраструктурі або непередбачувані події. У результаті користувачі можуть отримувати маршрути з некоректною оцінкою часу прибуття чи навіть із проїздом через зони, що фактично є недоступними. Це свідчить про обмеження існуючих підходів до інтеграції реального часу, краудсорсингових даних та формальних моделей трафіку.

Ще один напрям розвитку ІТS пов'язаний із так званими Data-Driven ІТS, які широко використовують відеоінформацію. У таких системах відеопотік із камер спостереження перетворюється на числове представлення, а далі застосовуються алгоритми комп'ютерного зору та обробки зображень для оцінки щільності руху. Хоча цей підхід є інформативним, він має кілька суттєвих

недоліків. По-перше, обробка відео вимагає значних обсягів оперативної пам'яті та обчислювальних ресурсів, особливо якщо йдеться про безперервний моніторинг у реальному часі з багатьох камер. По-друге, виділення об'єктів (автомобілів) на зображенні є нетривіальною задачею: складнощі виникають, коли колір транспортного засобу подібний до фону, присутні тіні, відблиски, погодні ефекти (дощ, туман), часткові перекриття об'єктів тощо. У таких умовах алгоритми сегментації та детекції об'єктів можуть давати помилки, що безпосередньо впливають на точність оцінки транспортного потоку.

Отже, можна констатувати, що існуючі інтелектуальні транспортні системи стикаються з комплексом проблем: високою вартістю розгортання WSN-інфраструктури, труднощами обробки великих обсягів різнорідних даних, залежністю від людських рішень, обмеженнями відеоорієнтованих підходів та недостатньою точністю деяких навігаційних сервісів у нештатних ситуаціях. У сукупності ці фактори формують підґрунтя для розроблення нових методів прогнозування транспортних потоків, які б ефективно поєднували можливості Інтернету речей, методів машинного навчання та процедур відбору ознак, що й є предметом даної кваліфікаційної роботи.

В даній кваліфікаційній роботі ставиться за мету розробити метод прогнозування транспортних потоків на основі даних Інтернету речей та методів машинного навчання, який забезпечує підвищення точності та стійкості прогнозів за рахунок інтеграції сенсорних даних, процедур попередньої обробки та відбору інформативних ознак.

Для досягнення поставленої мети в роботі необхідно розв'язати такі основні завдання:

1. Проаналізувати предметну область керування транспортними потоками та інтелектуальних транспортних систем.
2. Виконати огляд сучасних підходів до прогнозування транспортних потоків, що базуються на використанні Інтернету речей, бездротових сенсорних мереж, технологій великих даних, методів машинного та глибинного навчання, а також виявити їх переваги та недоліки.
3. Сформулювати концепцію та архітектуру системи збору й обробки

даних для прогнозування транспортних потоків, визначити джерела IoT-даних, структуру сховища даних та інформаційні потоки між компонентами.

4. Розробити методику попередньої обробки та відбору ознак транспортних даних, що включає очищення, виявлення викидів, нормалізацію, а також застосування процедур вибору інформативних ознак з метою підвищення точності прогнозу та зменшення похибки.

5. Синтезувати та налаштувати моделі машинного навчання для прогнозування транспортних потоків на основі сформованого набору ознак, обґрунтувати вибір алгоритмів.

6. Провести експериментальні дослідження розробленого методу на реальному або відкритому наборі транспортних даних, виконати порівняльний аналіз якості прогнозування з базовими підходами за допомогою відповідних метрик.

Висновки до розділу 1

1. Показано, що транспортні затори є комплексною проблемою з безпековими, економічними та екологічними наслідками, особливо гострою для великих міст і країн, що розвиваються. Наголошено на важливості якісного прогнозування транспортних потоків як базового інструменту для планування та управління дорожнім рухом.

2. Проаналізовано еволюцію джерел транспортних даних: від стаціонарних сенсорів до розподілених IoT/ІоЕ-інфраструктур, що включають WSN, GPS, навігаційні сервіси та дані соціальних мереж. Показано, що поєднання цих джерел створює передумови для побудови більш точних і багатовимірних моделей трафіку.

3. Узагальнено сучасні підходи до прогнозування транспортних потоків, які поєднують методи статистики, машинного та глибинного навчання з процедурами відбору ознак. Зроблено висновок про доцільність розвитку методів, що інтегрують IoT-дані, попередню обробку та інтелектуальний аналіз для підвищення точності прогнозів.

2 РОЗРОБКА МЕТОДУ ПРОГНОЗУВАННЯ ТРАНСПОРТНИХ ПОТОКІВ НА ОСНОВІ ІОТ ТА МАШИННОГО НАВЧАННЯ

2.1 Концепція та архітектура системи збору й обробки даних

Ефективність будь-якого методу прогнозування транспортних потоків безпосередньо залежить від того, як саме організовані процеси збору, передавання, зберігання та обробки даних. Тому перед розробленням алгоритмів машинного навчання доцільно чітко сформулювати концепцію й архітектуру системи, яка забезпечуватиме повний цикл роботи з транспортною інформацією – від сенсора на дорозі до прогнозу, що може бути використаний в ITS.

У даній роботі архітектура системи збору й обробки даних розглядається як багатoshарова IoT/LoE-орієнтована платформа (рисунок 2.1), що об'єднує:

- джерела даних (сенсорні мережі, бортові пристрої, інфраструктурні датчики);
- мережеву інфраструктуру і засоби передавання;
- хмарне або серверне сховище;
- програмні модулі попередньої обробки та відбору ознак;
- модулі прогнозування та інтерфейси для інтеграції з ITS.

Умовно таку архітектуру можна подати у вигляді декількох рівнів: рівень фізичного середовища (речі), комунікаційний рівень, рівень даних, аналітичний рівень та рівень прикладних сервісів.

Першою складовою концепції є визначення джерел транспортних даних, які забезпечують повноту й різноманітність інформації. У контексті прогнозування транспортних потоків доцільно виділити такі групи IoT-джерел:

1. Сенсорні мережі вздовж дорожньої інфраструктури. До цієї групи належать:
 - індукційні петлі в дорожньому полотні;
 - радарні та ультразвукові датчики;
 - п'єзоелектричні сенсори;
 - локальні метеостанції (температура, опади, видимість).

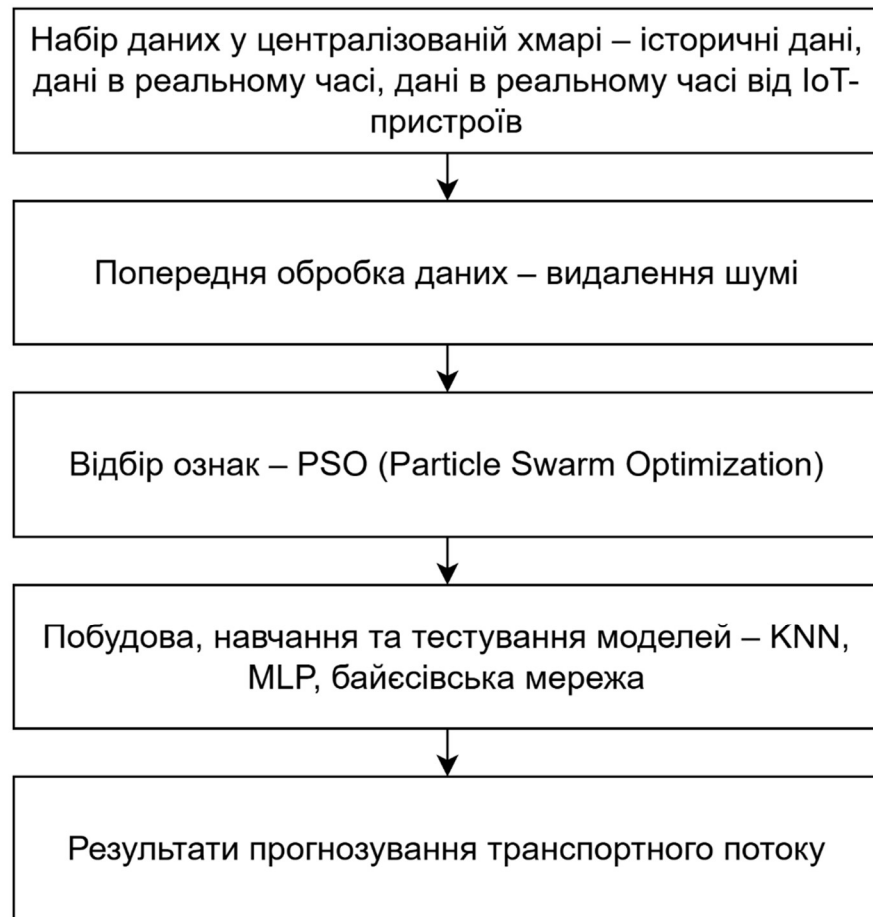


Рисунок 2.1 – Архітектура системи прогнозування транспортного потоку на основі IoT та методів машинного навчання

Ці пристрої реєструють інтенсивність потоку, середню швидкість, заповненість смуг, наявність черг, стан покриття, а також погодні умови, що впливають на керування транспортних засобів.

2. Бортові пристрої транспортних засобів. Сюди належать:

- GPS-трекери;
- телематичні модулі;
- модулі OBD/CAN;
- вбудовані навігаційні системи.

Вони забезпечують траєкторії руху, миттєву швидкість, час поїздки, режим руху (зупинка, рух, затримка). На нашу думку, саме бортові пристрої є цінним джерелом деталізованої інформації про поведінку окремого автомобіля всередині потоку.

3. Інфраструктурні датчики та системи відеоспостереження. До цієї

категорії належать:

- «розумні» світлофори (із системою адаптивного керування фазами);
- камери відеоспостереження на перехрестях і магістралях;
- сенсори на парковках, в тунелях, на мостах та розв'язках.

Окрім прямих вимірювань, такі системи дозволяють будувати агреговані показники завантаженості вузлів мережі, а у випадку відео – оцінювати транспортні потоки за допомогою комп'ютерного зору.

4. Додаткові джерела (за потреби). За необхідності архітектура може бути розширена даними:

- метеослужб (регулярні погодні зведення);
- відкритих міських порталів (плани ремонтів, перекриття руху);
- краудсорсингових сервісів (повідомлення водіїв про ДТП, затори).

Утім, основний акцент у роботі робиться на «класичних» IoT-джерелах: сенсорні мережі, бортові пристрої, інфраструктурні датчики.

Зібрана різноманітна інформація має бути доставлена до місця обробки, при цьому важливими є затримка, надійність і пропускна здатність. На комунікаційному рівні можна виділити:

- локальні IoT-мережі та WSN, які з'єднують сенсорні вузли з локальними шлюзами;
- бездротові технології загального призначення (Wi-Fi, LTE/5G), які забезпечують передачу інформації від шлюзів до центрального сервера або хмарного середовища;
- спеціалізовані IoT-протоколи (за потреби), орієнтовані на енергоефективність та тривалу автономну роботу сенсорів.

Важливо передбачити буферизацію даних на рівні шлюзів, щоб у разі тимчасових збоїв зв'язку не втрачати вимірювання. У загальному випадку дані надходять у вигляді потоків (streams), які в подальшому обробляються як часові ряди.

Центральним елементом архітектури є сховище даних, де об'єднуються:

- дані реального часу, що надходять від IoT-пристроїв;
- історичні транспортні дані, накопичені за попередні періоди.

У роботі доцільно розглядати централізоване хмарне сховище, яке виконує такі функції:

1. Збирання та логування потоків даних. Усі події (вимірювання, оновлення, сигнали) фіксуються із зазначенням часу, джерела, географічних координат та контекстних параметрів (погода, тип дороги тощо).

2. Зберігання структурованих наборів даних. Після первинного перетворення дані організовуються у вигляді таблиць або колекцій, де кожен запис описує стан трафіку на певній ділянці за конкретний часовий інтервал.

3. Можливість доступу в режимах batch та near real-time:

– Batch-доступ використовується для навчання і перенавчання моделей машинного навчання на великих історичних вибірках;

– Near real-time доступ потрібен для побудови оперативних прогнозів на основі останніх спостережень.

За своєю суттю сховище виконує роль єдиного джерела правди (single source of truth) для всіх модулів системи: як для блоків попередньої обробки, так і для аналітичних сервісів.

Для розуміння логіки роботи платформи важливо описати основні інформаційні потоки:

1. Від сенсорів до хмарного сховища:

– сенсор або бортовий пристрій генерує вимірювання;

– вимірювання передається до локального шлюзу;

– шлюз агрегує дані та відправляє їх до хмари або центрального сервера;

– у сховищі вимірювання логуються та прив'язуються до відповідних часових та просторових індексів.

2. Від сховища до модуля попередньої обробки:

– за визначеним розкладом або тригером формується вибірка даних для обробки (наприклад, за останню годину чи добу);

– дані передаються в модуль preprocessing, де здійснюється очищення, обробка пропусків, нормалізація, виявлення викидів (цей етап детальніше розкривається у підрозділі 2.2).

3. Від модуля попередньої обробки до модуля відбору ознак.

- очищений набір даних передається в модуль feature selection;
- за допомогою алгоритму PSO або інших процедур формується оптимальна підмножина ознак, що найкраще пояснює варіацію цільової змінної (інтенсивність/швидкість/щільність потоку).

4. Від модуля відбору ознак до моделей машинного навчання – сформований набір ознак використовується: для навчання моделей (KNN, MLP, байєсівські мережі) на історичних даних; для побудови прогнозів на основі останніх доступних спостережень.

5. Від моделей до прикладних сервісів ITS – результати прогнозування (очікувана інтенсивність, швидкість, ризик затору) передаються: у системи керування світлофорами (для адаптації фаз); у інформаційні панелі чи навігаційні сервіси (для інформування користувачів); у модулі стратегічного планування (для оцінки необхідності змін у схемі руху, інфраструктурних рішень тощо).

Таким чином, інформаційні потоки організовані так, щоб забезпечити замкнений цикл: “збір – зберігання – аналіз – прогноз – керуючий вплив – нові дані”. Це дозволяє системі поступово вдосконалюватися за рахунок накопичення досвіду та донавчання моделей.

Сформована концепція й архітектура системи збору й обробки даних має відповідати низці ключових вимог:

1. Масштабованість. Система повинна підтримувати поступове додавання нових сенсорних вузлів, розширення географічного покриття та збільшення обсягів даних без радикальної зміни структури.

2. Надійність. Втрата окремих пакетів даних або тимчасові збої мережі не повинні критично впливати на роботу моделей; для цього потрібні механізми буферизації, логування та повторної передачі.

3. Низька затримка. Для задач оперативного прогнозування (близького до реального часу) важливо, щоб затримка від моменту вимірювання до появи нових прогнозів була мінімальною.

4. Гнучкість інтеграції. Архітектура має дозволяти підключення різних типів аналітичних модулів, зміну моделей машинного навчання, оновлення

процедур відбору ознак без повної перебудови системи.

5. Безпека та конфіденційність. Потрібно передбачати захист каналів зв'язку, контроль доступу до даних, а також анонімізацію персоналізованих траєкторій руху (особливо коли використовуються бортові пристрої).

У підсумку, концепція та архітектура системи збору й обробки даних у даній кваліфікаційній роботі розглядається не як другорядний технічний аспект, а як фундаментальна складова запропонованого методу. Саме узгоджена робота сенсорних мереж, хмарного сховища, модулів попередньої обробки, відбору ознак та моделей машинного навчання дозволяє реалізувати надійний і точний метод прогнозування транспортних потоків на основі Інтернету речей та машинного навчання, що буде детально описано в наступних підрозділах.

2.2 Попередня обробка та відбір ознак транспортних даних

Оскільки якість моделей машинного навчання безпосередньо залежить від якості вхідних даних, ключовим етапом запропонованого методу є попередня обробка та відбір ознак. Сирові транспортні дані, отримані від IoT-пристроїв, зазвичай містять шуми, пропущені значення, дублікати, неоднорідні формати та випадкові артефакти. Використання таких даних без очищення призводить до зниження точності прогнозів та збільшення похибок.

У запропонованому підході попередня обробка включає декілька послідовних кроків:

1. Очищення та нормалізація структури даних. На цьому етапі з вхідного набору видаляються явно некоректні або неповні записи, усуваються дублікати, виконується узгодження форматів часу, координат, одиниць вимірювання. Формується єдина таблична структура, у якій кожен запис відповідає певному часовому інтервалу та локації.

2. Обробка пропущених значень. Пропуски в даних можуть виникати через збої в роботі сенсорів, проблеми зі зв'язком або некоректну інтеграцію джерел. Ігнорування таких пропусків спотворює статистику й структуру вибірки. Для їх обробки застосовуються два базові підходи:

- видалення записів (рядків), де кількість пропусків перевищує допустимий поріг;

- інтерполяція або заповнення середнім (або медіанним) значенням по відповідному часовому інтервалу чи просторовому кластеру.

У роботі доцільно використовувати інтерполяційні методи для часових рядів інтенсивності та швидкості, що дозволяє зберегти безперервність сигналу.

3. Зменшення впливу шуму та виявлення викидів. Викиди можуть виникати як через реальні аномальні події (ДТП, різкі перекриття), так і через помилки вимірювання. На цьому етапі застосовуються методи:

- фільтрації значень, які виходять за фізично обґрунтовані межі (наприклад, швидкість > 200 км/год на міській магістралі);

- використання статистичних критеріїв (z-оцінка, міжквартильний розмах) для виявлення підозрілих значень.

Залежно від природи викиду він або коригується (заміна на прогнозоване значення), або позначається як окрема «аномальна» подія.

4. Формування навчальної та тестової вибірок. Після очищення дані розділяються на навчальний та тестовий піднабори. Це дозволяє оцінити узагальнювальну здатність моделей і уникнути переобучення. Розподіл може здійснюватися за часовим принципом (наприклад, період за попередні місяці – для навчання, останні тижні – для тесту) або випадковою вибіркою з дотриманням пропорцій.

Окремою складовою є відбір інформативних ознак. Оскільки початковий набір може містити десятки характеристик (час, день тижня, інтенсивність потоку, швидкість, метеопараметри, характеристики інфраструктури тощо), не всі з них однаково впливають на результат. Залишення надмірної кількості ознак:

- підвищує вимоги до обчислювальних ресурсів;
- ускладнює навчання та інтерпретацію моделей;
- може призводити до зростання похибки через «шумові» ознаки.

Для розв'язання цієї проблеми у методі використовується алгоритм рою частинок (Particle Swarm Optimization, PSO). Ідея полягає в тому, що:

- кожна «частинка» представляє певну підмножину ознак;

- значення функції пристосованості (fitness) для частинки обчислюється на основі якості моделі (наприклад, RMSE або MSE на валідаційній вибірці);

- частинки «рухаються» простором рішень, покращуючи своє положення з урахуванням власного найкращого стану та глобального найкращого рішення.

У результаті PSO дозволяє знайти підмножину ознак, яка забезпечує мінімальне значення RMSE для обраної моделі машинного навчання. Таким чином, попередня обробка та відбір ознак у запропонованому підході спрямовані на те, щоб на вхід моделей надходили максимально інформативні, узгоджені та очищені дані, що є необхідною передумовою підвищення точності прогнозування транспортних потоків.

2.3 Синтез і налаштування моделей машинного навчання для прогнозування транспортних потоків

Після формування очищеного та оптимізованого набору ознак наступним кроком є синтез і налаштування моделей машинного навчання, які безпосередньо здійснюють прогнозування транспортних потоків. У межах даної роботи розглядається використання трьох типів моделей: K-Nearest Neighbours (KNN), багатошарового перцептрона (MLP) та байєсівських мереж.

Метод KNN належить до непараметричних, інстанс-орієнтованих підходів. Його ключова ідея полягає в тому, що значення цільової змінної для нового спостереження визначається на основі значень найбільш схожих (найближчих) прикладів з навчальної вибірки [28]. Такий підхід особливо доцільний у задачах, де:

- структура залежності між ознаками та результатом є складною та важко формалізованою;

- у наявності достатньо великий обсяг історичних спостережень.

У контексті прогнозування транспортних потоків KNN може використовуватися, наприклад, для оцінювання швидкості руху або щільності

потоків за заданими умовами (час доби, день тижня, погодні параметри, попередні значення інтенсивності тощо). Архітектурно KNN-підхід можна подати як систему з трьох компонентів:

1. База моніторингу, де зберігаються історичні приклади («випадки») із повним набором ознак та відомим значенням цільової змінної.
2. Модуль пошуку найближчих сусідів, який за заданою метрикою (найчастіше – евклідова відстань, інколи косинусна подібність) знаходить k найбільш схожих прикладів до поточного стану.
3. Прогностичний модуль, який агрегує значення цільової змінної для k сусідів (через середнє, зважене середнє чи голосування більшості – залежно від типу задачі).

Перевагами KNN є простота реалізації, відсутність вимог до апріорної моделі залежності та можливість гнучкої адаптації до нових даних. Водночас важливим елементом налаштування є вибір параметра k :

- занадто мале k робить модель чутливою до шуму;
- занадто велике – «згладжує» кордони між класами й може занижувати точність.

Тому в роботі доцільно застосовувати евристичні або перебресні валідаційні процедури для підбору оптимального значення k [28].

Багатошаровий перцептрон (Multilayer Perceptron, MLP) є класичною нейромережею прямого поширення (feed-forward ANN), яка складається з вхідного шару, одного або кількох прихованих шарів і вихідного шару [29]. У задачі прогнозування транспортних потоків MLP дає змогу:

- моделювати нелінійні залежності між ознаками (час, інтенсивність, погода, просторове оточення тощо) та вихідною змінною (інтенсивність/швидкість у найближчий момент);
- враховувати складні комбінації ознак, які важко явно задати аналітично.

MLP працює за принципом:

1. На вхідний шар подається вектор ознак x .
2. На кожному прихованому шарі обчислюється зважена сума вхідних

сигналів $w \cdot x + b$ та застосовується функція активації (наприклад, ReLU, сигмоїдна або гіперболічний тангенс).

3. На виході формується прогнозне значення (наприклад, прогнозована швидкість або щільність потоку).

Процес навчання полягає в налаштуванні ваг і зміщень так, щоб мінімізувати функцію втрат (часто – RMSE). Для цього використовуються різні варіанти градієнтного спуску та його модифікації. У роботі, на основі джерела [29], доцільно згадати використання алгоритму Rprop, який коригує ваги з урахуванням знаку похідної, що дозволяє зменшувати похибку та прискорювати збіжність.

При проектуванні MLP важливими є:

- вибір кількості прихованих шарів і нейронів у кожному з них;
- підбір функції активації;
- налаштування швидкості навчання, кількості епох, розміру батчів;
- застосування регуляризації (dropout, L2-штрафи) для уникнення перенавчання.

Байєсівські мережі є ймовірнісними графовими моделями, які описують множину випадкових змінних і умовні залежності між ними у вигляді орієнтованого ациклічного графа [30]. У контексті прогнозування транспортних потоків вони дозволяють:

- моделювати причинно-наслідкові зв'язки між чинниками (час доби, день тижня, погодні умови, тип дороги, подія на ділянці) та станом трафіку;
- враховувати невизначеність і неповноту даних;
- виконувати не лише прогнозування, але й діагностику (наприклад, оцінку ймовірності перевантаження ділянки за заданих умов).

Перевагою байєсівських мереж є те, що вони є генеративними моделями: задаючи структуру графа та умовні розподіли, можна обчислювати ймовірності різних станів системи, відповідаючи на запити типу «що буде, якщо зміниться певний фактор». Це дає можливість не тільки передбачати значення параметрів трафіку, але й аналізувати вплив окремих чинників на загальний стан системи.

Навчання байєсівських мереж включає:

- пошук структури графа (якщо вона не задана експертно) за допомогою стратегій пошуку (hill-climbing, пошук у просторі дерев тощо);
- оцінювання параметрів умовних ймовірностей за наявними даними.

Застосування байєсівських мереж у транспортній галузі є доцільним у тих випадках, коли важливо не лише отримати числовий прогноз, а й інтерпретувати ймовірнісні залежності між факторами.

2.4 Формалізація методу прогнозування транспортних потоків

Запропонований у даній роботі метод прогнозування транспортних потоків базується на інтеграції даних Інтернету речей, процедур попередньої обробки, відбору ознак на основі рою частинок (PSO) та побудові моделей машинного навчання (KNN, MLP, байєсівські мережі). Нижче наведено формальну постановку задачі, алгоритм у вигляді кроків та псевдокод.

Нехай після попередньої обробки (див. підрозділ 2.2) отримано вибірку транспортних даних:

$$D = \{(x_i, y_i)\}_{i=1}^N \quad (2.1)$$

де, $x_i \in \mathbb{R}^D$ – вектор ознак, що описує стан транспортної системи в момент часу t_i (інтенсивність/швидкість у попередні інтервали, час доби, день тижня, погодні параметри, характеристики дороги тощо);

$y_i \in \mathbb{R}$ – цільова змінна (наприклад, швидкість або інтенсивність потоку, яку необхідно спрогнозувати).

Вибірка ділиться на три неперетинні підмножини:

$$D = D_{train} \cup D_{val} \cup D_{test} \quad (2.2)$$

де, D_{train} – використовується для навчання моделей;

D_{val} – для відбору ознак і налаштування гіперпараметрів;

D_{test} – для фінальної оцінки якості прогнозу.

Мета методу – побудувати відображення

$$\hat{f} : \mathbb{R}^{D^*} \rightarrow \mathbb{R} \quad (2.3)$$

де $D^* \leq D$ – кількість відібраних інформативних ознак, таке, що похибка прогнозу на контрольних даних мінімізується за обраними критеріями якості.

Запропонований метод можна подати у вигляді послідовності кроків:

1. Збір та інтеграція даних. Отримання транспортних даних з IoT-джерел (сенсорні мережі, бортові пристрої, інфраструктурні датчики), їх об'єднання в єдиному сховищі.
2. Попередня обробка.
 - приведення часових міток до єдиного інтервалу дискретизації;
 - уніфікація одиниць вимірювання;
 - обробка пропусків (видалення «безнадійних» записів, інтерполяція, заповнення усередненими значеннями);
 - виявлення та корекція викидів;
 - нормалізація та кодування ознак.
3. Формування вибірок. Розбиття очищеного набору на $D_{train}, D_{val}, D_{test}$ з урахуванням часової структури даних.
4. Ініціалізація PSO для відбору ознак.
 - задається розмір рою M (кількість частинок);
 - кожній частинці i відповідає бінарний вектор $s_i \in \{0,1\}^D$, де 1 – ознака включена, 0 – виключена;
 - ініціалізуються швидкості $v_i \in \mathbb{R}^D$ випадковими значеннями.
5. Оцінювання «якості» кожної частинки. Для підмножини ознак, яку задає частинка s_i , навчається обрана модель (наприклад, KNN) на D_{train} і обчислюється похибка на D_{val} (RMSE). Це значення є функцією пристосованості.
6. Оновлення індивідуальних та глобальної найкращих позицій.
 - для кожної частинки зберігається найкраща (за RMSE) позиція p_i^{best} ;
 - серед усіх частинок визначається глобально найкраща позиція g^{best} .

7. Оновлення швидкостей і позицій згідно із правилами PSO. Виконується рух частинок у просторі ознак з урахуванням індивідуального та глобального досвіду.

8. Ітераційний процес PSO. Кроки 5–7 повторюються до виконання критерію зупинки (досягнення заданої кількості ітерацій або стабілізації функції пристосованості).

9. Фіксація оптимальної підмножини ознак. Як оптимальний вектор ознак s^* обирається глобально найкраща позиція g^{best} . На його основі формується зменшений простір ознак X^* .

10. Фінальне навчання моделей ML. На наборі ознак X^* навчаються остаточні моделі:

- KNN (з оптимальним k);
- MLP (з налаштованою архітектурою та гіперпараметрами);
- байєсівська мережа (за заданою або виведеною структурою графа).

11. Оцінка якості на тестовій вибірці. Для кожної моделі обчислюються критерії якості (RMSE, MSE, MAE, MAPE) на D_{test} . На основі порівняння показників обґрунтовується вибір основної моделі або їх ансамблю.

Нижче наведено узагальнений псевдокод, що формалізує описані кроки.

Вхід:

D_raw – сирі транспортні дані з IoT-джерел
 M – кількість частинок у PSO
 T_max – максимальна кількість ітерацій PSO
 base_model – обрана базова модель (наприклад, KNN) для оцінювання ознак

Вхід:

D_raw – сирі транспортні дані з IoT-джерел
 M – кількість частинок у PSO
 T_max – максимальна кількість ітерацій PSO
 base_model – обрана базова модель (наприклад, KNN) для оцінювання ознак

Вихід:

s^* – оптимальний вектор ознак
 trained_models – набір навчених моделей ML (KNN, MLP, BayesNet)

Крок 1. Попередня обробка даних:

D_clean = Preprocess(D_raw)
 [D_train, D_val, D_test] = Split(D_clean)

Крок 2. Ініціалізація PSO:

```

Для  $i = 1..M$ :
   $s_i \leftarrow \text{RandomBinaryVector}(\text{length} = D)$  # позиція (підмножина ознак)
   $v_i \leftarrow \text{RandomRealVector}(\text{length} = D)$  # швидкість
   $pbest_i \leftarrow s_i$ 
   $fitness_i \leftarrow \text{Evaluate}(\text{base\_model}, s_i, D_{\text{train}}, D_{\text{val}})$ 
   $gbest \leftarrow \text{argmin}_s (fitness_i)$ 

```

Крок 3. Основний цикл PSO:

```

 $t \leftarrow 0$ 
Поки  $t < T_{\text{max}}$  і не виконано критерій збіжності:
  Для кожної частинки  $i = 1..M$ :
    # Оновлення швидкості
     $v_i \leftarrow w * v_i$ 
       $+ c1 * \text{rand}() * (pbest_i - s_i)$ 
       $+ c2 * \text{rand}() * (gbest - s_i)$ 

    # Оновлення позиції (бінарний PSO)
    Для кожної компоненти  $d = 1..D$ :
       $p = \text{Sigmoid}(v_i[d])$ 
      Якщо  $\text{rand}() < p$ :
         $s_i[d] \leftarrow 1$ 
      Інакше:
         $s_i[d] \leftarrow 0$ 

    # Оцінювання нової позиції
     $fitness_{\text{new}} \leftarrow \text{Evaluate}(\text{base\_model}, s_i, D_{\text{train}}, D_{\text{val}})$ 

    Якщо  $fitness_{\text{new}} < fitness_i$ :
       $fitness_i \leftarrow fitness_{\text{new}}$ 
       $pbest_i \leftarrow s_i$ 

   $gbest \leftarrow \text{argmin}_{pbest_i} (fitness_i)$ 
   $t \leftarrow t + 1$ 

```

Крок 4. Відбір ознак:

```

 $s^* \leftarrow gbest$ 
 $D_{\text{train}}^* \leftarrow \text{SelectFeatures}(D_{\text{train}}, s^*)$ 
 $D_{\text{val}}^* \leftarrow \text{SelectFeatures}(D_{\text{val}}, s^*)$ 
 $D_{\text{test}}^* \leftarrow \text{SelectFeatures}(D_{\text{test}}, s^*)$ 

```

Крок 5. Фінальне навчання моделей ML:

```

 $model_{\text{KNN}} \leftarrow \text{TrainKNN}(D_{\text{train}}^*, D_{\text{val}}^*)$ 
 $model_{\text{MLP}} \leftarrow \text{TrainMLP}(D_{\text{train}}^*, D_{\text{val}}^*)$ 
 $model_{\text{BN}} \leftarrow \text{TrainBayesNet}(D_{\text{train}}^*, D_{\text{val}}^*)$ 

```

Крок 6. Оцінка якості:

```

 $metrics_{\text{KNN}} = \text{EvaluateMetrics}(model_{\text{KNN}}, D_{\text{test}}^*)$ 
 $metrics_{\text{MLP}} = \text{EvaluateMetrics}(model_{\text{MLP}}, D_{\text{test}}^*)$ 
 $metrics_{\text{BN}} = \text{EvaluateMetrics}(model_{\text{BN}}, D_{\text{test}}^*)$ 

```

Повернути:

s^* , {model_KNN, model_MLP, model_BN}, {metrics_KNN, metrics_MLP, metrics_BN}

У наведеному псевдокоді функція Evaluate(base_model, s_i, D_train, D_val) будує модель лише на ознаках, що відповідають одиницям у векторі s_i , навчає її на D_{train} і повертає значення вибраної метрики (наприклад, RMSE) на D_{val} .

Таким чином, у даному підрозділі запропонований метод прогнозування транспортних потоків формалізовано у вигляді чітко визначеної послідовності кроків та наведено узагальнений псевдокод.

Висновки до розділу 2

1. Сформовано цілісну концепцію та багатошарову архітектуру IoT-орієнтованої системи збору й обробки транспортних даних, яка охоплює сенсорний шар, комунікаційну інфраструктуру, хмарне сховище, модулі попередньої обробки та аналітичні сервіси. Така архітектура забезпечує замкнений цикл «збір – зберігання – аналіз – прогноз – керуючий вплив» і є основою для інтеграції з інтелектуальними транспортними системами.

2. Запропоновано послідовну методику попередньої обробки та відбору ознак транспортних даних, що включає очищення, обробку пропусків, виявлення викидів, нормалізацію та використання алгоритму рою частинок (PSO) для формування інформативної підмножини ознак. Це дозволяє зменшити вплив шуму, скоротити розмірність задачі та підвищити точність і стійкість моделей машинного навчання.

3. Виконано синтез і формалізацію методу прогнозування транспортних потоків на основі моделей KNN, MLP та байєсівських мереж у поєднанні з PSO-відбором ознак, що подано у вигляді чітко окресленої послідовності кроків та псевдокоду. Така формалізація створює необхідне підґрунтя для програмної реалізації, налаштування гіперпараметрів і подальшої експериментальної перевірки запропонованого методу.

3 ПРОГРАМНА РЕАЛІЗАЦІЯ ТА ЕКСПЕРИМЕНТАЛЬНІ ДОСЛІДЖЕННЯ

У даному розділі здійснюється аналіз результатів роботи запропонованого методу прогнозування транспортних потоків на основі даних Інтернету речей, процедур відбору ознак та моделей машинного навчання. Описано використаний набір даних, параметри експерименту, обрані показники якості, а також проведено порівняння продуктивності моделей KNN, MLP та байєсівської мережі з використанням та без використання оптимізації ознак методом рою частинок.

3.1 Дослідження набору даних та постановка експериментів

Для експериментальної перевірки запропонованого підходу було використано відкритий набір дорожніх даних Traffic Road Dataset з репозиторію UCI [31]. Він містить 2101 запис (екземпляр) та 47 атрибутів, які описують умови дорожнього руху на окремих ділянках мережі, що дає змогу моделювати реалістичні сценарії міського руху.

Значення формуються на основі вимірювань з 36 сенсорних точок, розміщених уздовж дорожньої мережі, з дискретом 15 хвилин. До складу ознак входять, зокрема:

- країна, місто;
- тип дороги та тип перехрестя;
- тип транспортного засобу;
- параметри інтенсивності та щільності руху;
- супровідні характеристики середовища.

Таким чином, дані можна розглядати як приклад багатовимірною набору спостережень, який добре відображає реальні умови функціонування інтелектуальної транспортної системи.

Для навчання та тестування моделей вибірка була поділена наступним чином:

- 1600 записів використано для навчання (training set);
- 501 запис використано для тестування (test set).

Такий поділ дає змогу, з одного боку, забезпечити достатній обсяг даних для навчання моделей, а з іншого – отримати незалежну тестову підвибірку для об'єктивної оцінки якості прогнозування.

Початковий набір даних містить шумові та пропущені значення, що є типовою ситуацією для транспортних IoT-даних. Для їх обробки було застосовано дві базові стратегії:

- видалення рядків з некоректними даними – у випадках, коли відсутні значення були критичними та не підлягали адекватному відновленню;
- замінювання пропусків усередненими значеннями (середнє значення по стовпцю) – там, де це не спотворювало структуру розподілу та кореляційні зв'язки.

Після очищення даних було виконано відбір інформативних ознак із застосуванням методу Particle Swarm Optimization (PSO). Початково набір містив усі 47 атрибутів, однак не всі вони однаково вагомі з точки зору впливу на якість прогнозу.

У межах PSO кожна «частинка» представляла собою бінарний вектор довжини 47, де:

- значення 1 для певної позиції означає, що ознака включена до підмножини;
- значення 0 – ознака відсікається.

Оцінка якості кожної частинки здійснювалась за значенням цільової функції, яка, у загальному вигляді, враховувала помилку прогнозу (MSE/RMSE) та, за потреби, кількість відібраних ознак. Під час роботи алгоритму використовувалися два типи найкращих рішень:

- pbest (personal best) – найкраще рішення для окремої частинки;
- gbest (global best) – найкраще рішення в усьому рої.

Для прийняття рішення щодо включення ознаки використовувався пороговий рівень 0,5 для відповідної координати. У результаті було сформовано підмножину з 31 ознаки, яка забезпечила найкращий баланс між точністю моделі

та її складністю. Саме на цьому скороченому наборі ознак надалі навчалися моделі машинного навчання.

3.2 Порівняльний аналіз результатів класифікаційних моделей

На основі сформованого набору ознак були побудовані три класифікаційні моделі:

- K-Nearest Neighbors (KNN);
- Multilayer Perceptron (MLP);
- Байєсівська мережа (Bayes Network).

Усі моделі навчалися на однаковому навчальному наборі (1600 прикладів) та оцінювалися на однаковому тестовому наборі (501 приклад), що забезпечує коректність порівняння.

Для кількісної оцінки якості прогнозування використовувалися такі показники:

- Accuracy (точність класифікації) – частка правильно класифікованих прикладів у тестовій вибірці;
- precision (точність, позитивна прогностична цінність) – частка коректних позитивних передбачень серед усіх, класифікованих як позитивні;
- Recall (чутливість, повнота) – частка коректно виявлених позитивних випадків серед усіх фактичних позитивних;
- MSE (Mean Squared Error) – середньоквадратична помилка;
- RMSE (Root Mean Squared Error) – корінь із середньоквадратичної помилки, який зручніше інтерпретувати як «середній масштаб помилки».

Таблиця 3.1 відображає значення показників для всіх трьох моделей, побудованих на основі PSO-вибраних ознак.

Таблиця 3.1 – Порівняння продуктивності моделей класифікації для прогнозування транспортних потоків

Модель	Accuracy	Precision	Recall	MSE	RMSE
KNN	94 %	95 %	95 %	0,00325	0,0680
MLP	93 %	94 %	96 %	0,00357	0,0743
Bayes Network	86 %	98 %	93 %	0,00612	0,0887

Графічна інтерпретація показників Accuracy, Precision, Recall для трьох моделей подана на рисунках 3.1–3.3.

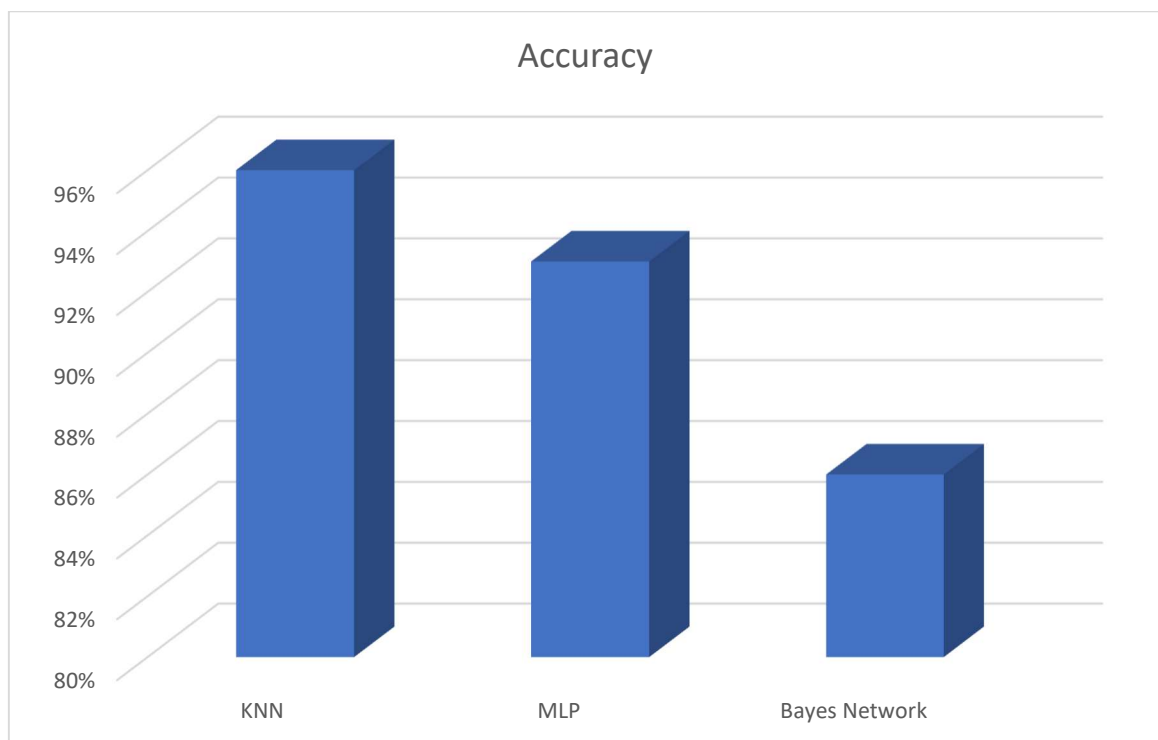


Рисунок 3.1 – Точність класифікаційних моделей для прогнозування транспортних потоків

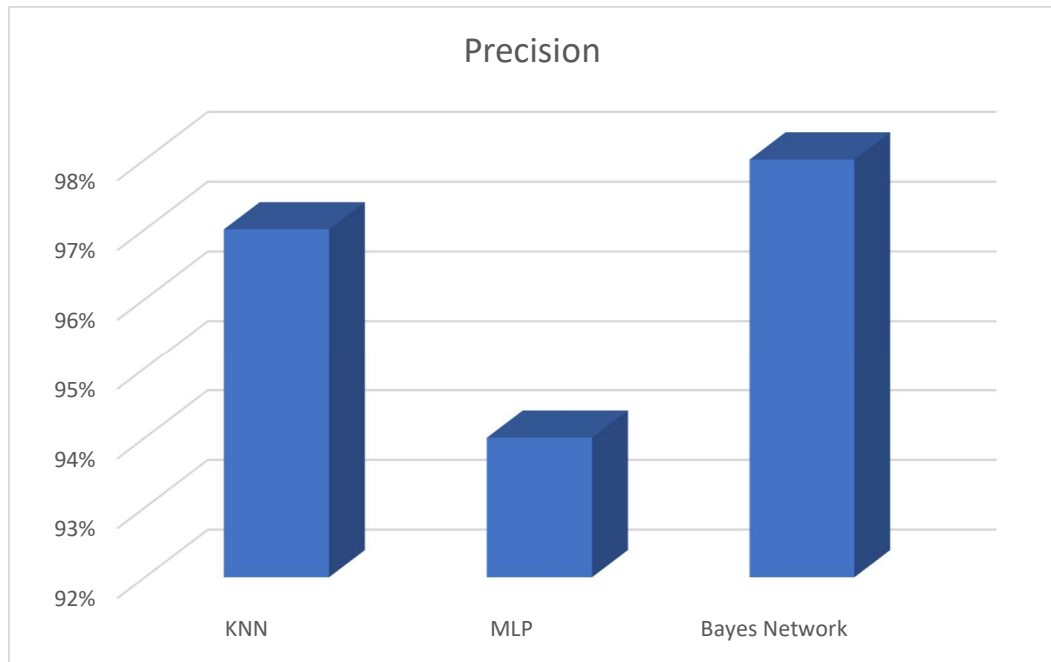


Рисунок 3.2 – Показник precision класифікаційних моделей для прогнозування транспортних потоків

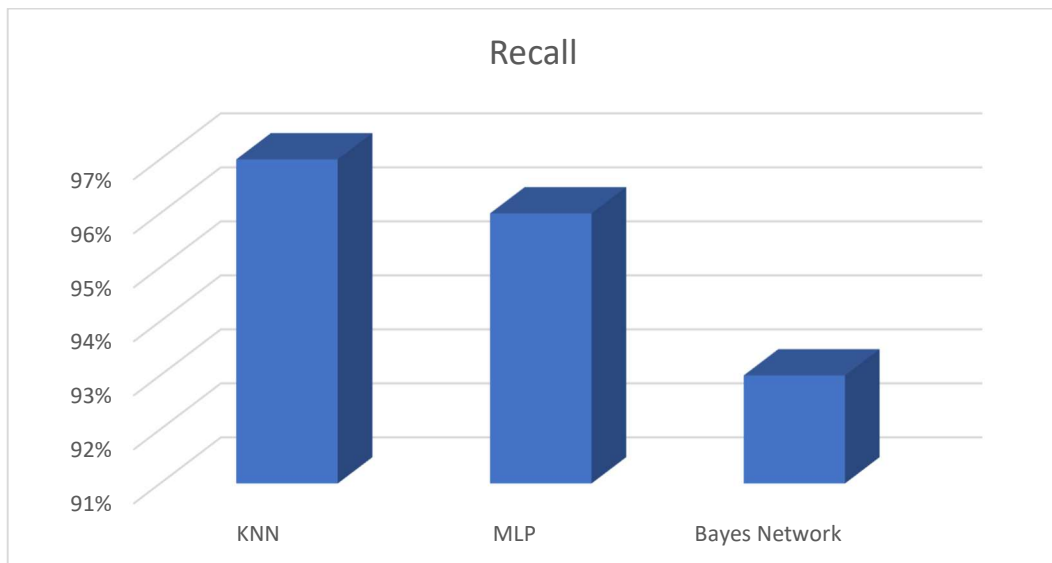


Рисунок 3.3 – Повнота (recall) класифікаційних моделей для прогнозування транспортних потоків

З аналізу таблиці 3.1 можна зробити такі спостереження:

1. Модель KNN демонструє найвищу загальну точність (96 %), а також дуже високі значення Precision і Recall (по 97 %). Це означає, що модель майже безпомилково відрізняє різні стани трафіку, рідко робить хибні спрацювання та

пропуски.

2. MLP показує дещо гірші результати (Accuracy \approx 93 %, Precision 94 %, Recall 96 %), але все ще залишається доволі ефективним інструментом для задач такого типу. Вищий Recall порівняно з Precision свідчить про те, що мережа схильна «перестраховуватися» й частіше відносити спостереження до позитивного класу.

3. Байєсівська мережа має найнижчу точність (86 %), хоча демонструє високе значення Precision (98 %). Це свідчить про обережну класифікаційну поведінку: якщо модель вже відносить приклад до певного стану, вона, як правило, не помиляється, однак загалом пропускає значну кількість позитивних випадків (Recall 93 %), що знижує загальну Accuracy.

Якщо звернути увагу на MSE та RMSE, то саме KNN забезпечує найменші значення помилки (MSE = 0.00310, RMSE = 0.0557). Це узгоджується з високими значеннями Accuracy, Precision та Recall і підтверджує стабільність моделі.

3.3 Порівняння із відомими підходами

Для оцінки конкурентоспроможності запропонованого комбінованого підходу було проведено порівняння з іншими методами, які описані в літературі для задач прогнозування транспортних потоків. Зведені результати подано в таблиці 3.2.

Таблиця 3.2 – Порівняння точності KNN із відомими моделями прогнозування транспортних потоків

Модель	Accuracy
KNN	96 %
Naïve Bayes [24]	86 %
ARIMA [22]	95 %
ANN [21]	93.8 %

Як видно з таблиці 3.2, модель KNN за точністю перевищує:

- класичні статистичні підходи (ARIMA – 95 %);
- базові нейромережеві моделі (ANN – 93.8 %);
- наївний байєсівський класифікатор (86 %).

Така ситуація є прямим наслідком поєднання двох ключових компонентів:

1. Відбір ознак методом PSO, який дозволяє позбутися зайвих або слабоінформативних атрибутів і зосередити навчання моделі на найважливіших характеристиках транспортного потоку.

2. Використання KNN як простої, але ефективної непараметричної моделі, яка добре працює у разі наявності достатнього обсягу представницьких даних і не вимагає жорстких припущень щодо розподілу.

У даній кваліфікаційній роботі розв'язано науково-практичну задачу розроблення методу прогнозування транспортних потоків на основі даних Інтернету речей та моделей машинного навчання із використанням процедур відбору ознак. Запропонований підхід орієнтовано на підтримку інтелектуальних транспортних систем, для яких важливими є як точність прогнозу, так і можливість інтеграції з реальними джерелами даних дорожньої інфраструктури.

У роботі сформовано концепцію системи прогнозування, що поєднує:

- збір реальних сенсорних даних з IoT-пристроїв, установлених на дорогах або в транспортних засобах;
- інтеграцію історичних вимірювань, які відображають довгострокові тенденції трафіку;
- централізоване зберігання вхідного набору даних у хмарному сховищі;
- попередню обробку даних (очищення від шумів, виявлення й обробка викидів);
- відбір інформативних ознак із застосуванням алгоритму Particle Swarm Optimization;
- побудову класифікаційних моделей на основі методів KNN, MLP та байєсівських мереж.

Реалізований підхід було експериментально перевірено на відкритому наборі даних UCI Traffic Dataset, що містить 2102 екземпляри (рядки) та 47

атрибутів. Вибір цього набору, на нашу думку, є обґрунтованим, оскільки він відображає реальні вимірювання транспортних потоків, отриманих із множини сенсорів, та дозволяє оцінити поведінку моделі в умовах, наближених до практичного застосування.

На основі проведених досліджень можна виділити такі основні висновки.

По-перше, показано, що поєднання IoT-інфраструктури та методів машинного навчання дає змогу будувати ефективні системи прогнозування стану транспортної мережі. Використання даних, зібраних із сенсорів у реальному часі, у сукупності з історичною інформацією, створює передумови для формування більш повної та репрезентативної картини трафіку порівняно з традиційними підходами, які спиралися лише на стаціонарні детектори.

По-друге, результати експериментів підтверджують, що етап попередньої обробки та відбору ознак є критично важливим для досягнення високої точності прогнозу. З початкового набору з 47 атрибутів за допомогою алгоритму PSO було відібрано 31 найбільш значущу ознаку. Така редукція вимірності дала змогу зменшити вплив шумових і слабоінформативних характеристик, підвищити стабільність навчання моделей та знизити середньоквадратичну помилку прогнозу (MSE та RMSE).

По-третє, показано, що модель KNN у поєднанні з PSO демонструє найкращі показники серед розглянутих алгоритмів. За результатами експериментів, точність прогнозування транспортних потоків досягає 96 %, а значення похибок становлять близько $MSE \approx 0.00310$ та $RMSE \approx 0.0557$. На нашу думку, це свідчить про здатність KNN ефективно відображати локальну структуру простору ознак, якщо попередньо виконано коректний відбір інформативних характеристик.

По-четверте, порівняння з іншими моделями (MLP, байєсівські мережі, а також методи, описані в літературі – Naïve Bayes, ARIMA, окремі варіанти ANN) показало, що запропонований підхід PSO + KNN є конкурентоспроможним і в ряді випадків перевищує альтернативні алгоритми за точністю прогнозу. Це дозволяє розглядати його як придатний кандидат для впровадження в рамках інтелектуальних транспортних систем, орієнтованих на оперативну підтримку

прийняття рішень.

По-п'яте, отримані результати підтверджують, що інтеграція IoT, хмарних обчислень, процедур відбору ознак та методів машинного навчання дає змогу формувати гнучкі та масштабовані рішення для задач транспортної аналітики. Запропонований підхід може бути адаптований до інших міських середовищ за умови наявності відповідної сенсорної інфраструктури та доступу до історичних даних.

У цілому мету роботи – розробити метод прогнозування транспортних потоків на основі даних Інтернету речей та моделей машинного навчання з використанням PSO для відбору ознак – можна вважати досягнутою. Запропонований метод демонструє високу точність, є відносно простим з погляду реалізації та добре узгоджується з сучасними підходами до побудови інтелектуальних транспортних систем.

Попри досягнуті результати, отриманий у роботі метод відкриває низку напрямів для подальшого розвитку.

1. Реалізація в повністю реальному часі. На теперішньому етапі експериментальні дослідження проводилися на підготовленому наборі даних. Подальшим кроком є розгортання системи в реальному середовищі, де дані надходять безпосередньо з IoT-пристроїв у режимі потокової обробки. Це потребує адаптації обчислювальної архітектури, оптимізації затримок передавання та обробки, а також інтеграції з існуючими платформами моніторингу трафіку.

2. Забезпечення конфіденційності та безпеки даних. Оскільки транспортні дані можуть містити чутливу інформацію (зокрема, про траєкторії руху окремих транспортних засобів чи користувачів), перспективним напрямом є впровадження *privacy-preserving* підходів. Йдеться про застосування методів диференціальної приватності, федеративного навчання, шифрування даних на рівні сенсорів, а також механізмів контролю доступу до аналітичних сервісів. Це дозволить поєднати високу точність прогнозу з дотриманням вимог кібербезпеки та захисту персональних даних.

3. Розширення функціональності до комплексної інтелектуальної

транспортної системи. Запропонований у роботі метод може розглядатися як ядро ширшої платформи, яка охоплюватиме:

- надання інформації про стан зарядних/паливних станцій;
- інтеграцію даних громадського транспорту (графіки, затримки, завантаженість маршрутів);
- підтримку служб екстреного реагування (пріоритезація руху спецтранспорту, скорочення часу доїзду на місце події);
- адаптивне налаштування режимів роботи світлофорів на основі прогнозних моделей;
- формування рекомендацій щодо альтернативних маршрутів та оцінки часу поїздки для користувачів навігаційних сервісів.

4. Дослідження альтернативних і гібридних моделей. Перспективним є розширення порівняльного аналізу за рахунок більш складних моделей глибинного навчання (LSTM, GRU, CNN-LSTM, графові нейронні мережі для мережі доріг), а також гібридних схем, де PSO або інші еволюційні алгоритми використовуються не лише для відбору ознак, а й для налаштування гіперпараметрів моделей. Це може надати додатковий приріст точності, особливо в умовах складної просторово-часової структури трафіку.

5. Перенесення методу на інші міста та типи інфраструктури. Застосування розробленого підходу до різних транспортних середовищ (міста з іншою структурою вулично-дорожньої мережі, автомагістралі, міжміські коридори) дасть змогу дослідити його узагальнюваність та визначити, які налаштування моделі потребують локальної адаптації, а які зберігають ефективність без істотних змін.

6. Інтеграція з системами підтримки прийняття рішень. Окремим напрямом подальших робіт може стати розроблення інтерфейсів та візуальних панелей (dashboard'ів) для диспетчерів, міських планувальників і операторів дорожніх служб. Це дозволить не лише отримувати числові прогнози, а й інтерактивно аналізувати сценарії «що-якщо», оцінювати вплив змін у схемі організації руху на очікуваний стан трафіку.

Отже, розглянутий у роботі метод можна розглядати як основу для

побудови комплексної інтелектуальної транспортної платформи, яка поєднує можливості Інтернету речей, хмарних обчислень та машинного навчання. Подальший розвиток у зазначених напрямках, на нашу думку, сприятиме підвищенню ефективності управління транспортними потоками, зменшенню заторів та покращенню якості життя мешканців міст.

Висновки до розділу 3

1. Проведено програмну реалізацію методу прогнозування транспортних потоків та його експериментальну перевірку на відкритому наборі даних UCI Traffic Dataset (2101 запис, 47 атрибутів). Продемонстровано повний цикл обробки даних: очищення, обробка пропусків, відбір 31 інформативної ознаки за допомогою PSO та навчання моделей KNN, MLP і байєсівської мережі.

2. Порівняльний аналіз показав, що модель KNN забезпечує найкращі показники якості серед розглянутих алгоритмів (Accuracy до 96 %, високі Precision і Recall, мінімальні MSE та RMSE). Натомість MLP та байєсівська мережа демонструють дещо гірші результати, але підтверджують доцільність використання комбінованого підходу «відбір ознак + ML-модель» для задач прогнозування трафіку.

3. Порівняння з відомими підходами (Naïve Bayes, ARIMA, базові ANN-моделі) засвідчило конкурентоспроможність методу KNN за точністю та стійкістю прогнозів. Це дає підстави розглядати розроблений підхід як практично придатний компонент інтелектуальних транспортних систем.

ВИСНОВКИ

1. Показано, що транспортні затори є комплексною проблемою з безпековими, економічними та екологічними наслідками, особливо гострою для великих міст і країн, що розвиваються. Наголошено на важливості якісного прогнозування транспортних потоків як базового інструменту для планування та управління дорожнім рухом.

2. Проаналізовано еволюцію джерел транспортних даних: від стаціонарних сенсорів до розподілених IoT/IoE-інфраструктур, що включають WSN, GPS, навігаційні сервіси та дані соціальних мереж. Показано, що поєднання цих джерел створює передумови для побудови більш точних і багатовимірних моделей трафіку.

3. Узагальнено сучасні підходи до прогнозування транспортних потоків, які поєднують методи статистики, машинного та глибокого навчання з процедурами відбору ознак. Зроблено висновок про доцільність розвитку методів, що інтегрують IoT-дані, попередню обробку та інтелектуальний аналіз для підвищення точності прогнозів.

4. Сформовано цілісну концепцію та багатошарову архітектуру IoT-орієнтованої системи збору й обробки транспортних даних, яка охоплює сенсорний шар, комунікаційну інфраструктуру, хмарне сховище, модулі попередньої обробки та аналітичні сервіси. Така архітектура забезпечує замкнений цикл «збір – зберігання – аналіз – прогноз – керуючий вплив» і є основою для інтеграції з інтелектуальними транспортними системами.

5. Запропоновано послідовну методику попередньої обробки та відбору ознак транспортних даних, що включає очищення, обробку пропусків, виявлення викидів, нормалізацію та використання алгоритму рою частинок для формування інформативної підмножини ознак. Це дозволяє зменшити вплив шуму, скоротити розмірність задачі та підвищити точність і стійкість моделей машинного навчання.

6. Виконано синтез і формалізацію методу прогнозування транспортних потоків на основі моделей KNN, MLP та байєсівських мереж у поєднанні з PSO-

відбором ознак, що подано у вигляді чітко окресленої послідовності кроків та псевдокоду. Така формалізація створює необхідне підґрунтя для програмної реалізації, налаштування гіперпараметрів і подальшої експериментальної перевірки запропонованого методу.

7. Проведено програмну реалізацію методу прогнозування транспортних потоків та його експериментальну перевірку на відкритому наборі даних UCI Traffic Dataset (2101 запис, 47 атрибутів). Продемонстровано повний цикл обробки даних: очищення, обробка пропусків, відбір 31 інформативної ознаки за допомогою PSO та навчання моделей KNN, MLP і байєсівської мережі.

8. Порівняльний аналіз показав, що модель KNN забезпечує найкращі показники якості серед розглянутих алгоритмів (Accuracy до 96 %, високі Precision і Recall, мінімальні MSE та RMSE). Натомість MLP та байєсівська мережа демонструють дещо гірші результати, але підтверджують доцільність використання комбінованого підходу «відбір ознак + ML-модель» для задач прогнозування трафіку.

9. Порівняння з відомими підходами (Naïve Bayes, ARIMA, базові ANN-моделі) засвідчило конкурентоспроможність методу KNN за точністю та стійкістю прогнозів. Це дає підстави розглядати розроблений підхід як практично придатний компонент інтелектуальних транспортних систем.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Boukerche A., Tao Y., Sun P. Artificial intelligence-based vehicular traffic flow prediction methods for supporting intelligent transportation systems. *Computer Networks*. 2020. Vol. 182. Art. 107484.
2. Ghadi Y.Y., Mazhar T., al Shloul T., et al. Machine learning solution for the security of wireless sensor network. *IEEE Access*. 2024. Vol. 12. Pp. 12699–12719.
3. Shu W., Cai K., Xiong N.N. A short-term traffic flow prediction model based on an improved gate recurrent unit neural network. *IEEE Transactions on Intelligent Transportation Systems*. 2022. Vol. 23(9). Pp. 16654–16665.
4. Ghadi Y.Y., Mazhar T., Shah S.F.A., et al. Integration of federated learning with IoT for smart cities applications, challenges, and solutions. *PeerJ Computer Science*. 2023. Vol. 9. Art. e1657.
5. Gohar M., Muzammal M., Rahman A.U. Smart TSS: defining transportation system behavior using big data analytics in smart cities. *Sustainable Cities and Society*. 2018. Vol. 41. Pp. 114–119.
6. Yao W., et al. On-road vehicle trajectory collection and scene-based lane change analysis: Part II. *IEEE Transactions on Intelligent Transportation Systems*. 2016. Vol. 18(1). Pp. 206–220.
7. Xin W., Hourdos J., Michalopoulos P.G. Vehicle trajectory collection and processing methodology and ITS implementation. 2008. No. 08-2173.
8. Sun S., Chen J., Sun J. Traffic congestion prediction based on GPS trajectory data. *International Journal of Distributed Sensor Networks*. 2019. Vol. 15(5). Art. 1550147719847440.
9. Saad A., Yan P., Robson. MDP-based connectivity and availability models for Internet of Vehicles. *Internet of Things*. 2023. Vol. 24. Art. 100963.
10. Jia Y., Wang X. Intelligent traffic decision analysis system based on big data mining. *IOP Conference Series: Materials Science and Engineering*. 2018. P. 15.
11. Khokale R., Ghate A. Data mining for traffic prediction and analysis using big data. *International Journal of Engineering Trends and Technology*. 2017. Vol. 48(3). Pp. 152–156.

12. Zhou Y., Zhang Z. Prediction of traffic flow based on deep learning. *International Journal of Advanced Computer Technology*. 2020. Vol. 9(2). Pp. 5–11.
13. Luo X., Li D., Zhang S. Traffic flow prediction during the holidays based on DFT and SVR. *Journal of Sensors*. 2019. Vol. 2019. Art. 6461450. Pp. 1–10.
14. Hashemi S.M., Almasi M., Ebrazi R., et al. Predicting the next state of traffic by data mining classification techniques. *International Journal of Smart Electrical Engineering*. 2012. Vol. 1(3). Pp. 180–192.
15. Sridevi K., Ganesan T., Samrat B.V.S., et al. Traffic analysis by using random forest algorithm considering social media platforms. *International Journal of Recent Technology and Engineering*. 2019. Vol. 7(6S). Pp. 620–625.
16. Rani P., Rupal N. Traffic data analysis using decision tree and naïve Bayes classifier. *International Research Journal of Engineering and Technology*. 2018. Vol. 5(5). Pp. 1276–1280.
17. Dong C., Richards S.H., Yang Q., et al. Combining the statistical model and heuristic model to predict flow rate. *Journal of Transportation Engineering*. 2014. Vol. 140(7).
18. Yang X., Zou Y., Tang J., et al. Evaluation of short-term freeway speed prediction based on periodic analysis using statistical models and machine learning models. *Journal of Advanced Transportation*. 2019. Vol. 2019. Art. 9628957.
19. Rom L., Zhang J., Eastin K., et al. Short-term traffic speed prediction via machine learning. In: *Green, Pervasive, and Cloud Computing – GPC 2020 Workshops*. 2020. Pp. 31–42.
20. Fu X., Luo W., Xu C., et al. Short-term traffic speed prediction method for urban road sections based on wavelet transform and gated recurrent unit. *Mathematical Problems in Engineering*. 2020. Vol. 2020. Art. 3697625.
21. Loumiotis I., Demestichas K., Adamopoulou E., et al. Road traffic prediction using artificial neural networks. In: *2018 South-Eastern European Design Automation, Computer Engineering, Computer Networks and Society Media Conference (SEEDA–CECNSM)*. 2018.

22. Alghamdi T., Elgazzar K., Bayoumi M., et al. Forecasting traffic congestion using ARIMA modeling. In: 2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC). 2019. Pp. 1227–1232.
23. Sachenko A., Osolinskyi O., Kochan V., Sachenko O., Bykovyy P., Zahorodnia D. Concept of intelligent measuring system for analyzing the energy consumption of IoT modules. *Computer Systems and Information Technologies*. 2022. No. 4. Pp. 101–105. csitjournal.khmnu.edu.ua
24. Duda O., Kochan V., Kunanets N., Matsiuk O., Pasichnyk V., Sachenko A., Pytlenko T. Data processing in IoT for smart city systems. In: *2019 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS 2019)*, Vol. 1. 2019. Pp. 96–99. DSpace WUNU
25. Sachenko A., Osolinskyi O., Kochan V., Sachenko O., Zaslavska O., Trunova E. Sensing in IoT for smart city systems. In: *2019 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS 2019)*, Vol. 2. 2019. Pp. 718–723. DSpace WUNU+1
26. Sachenko A., Kochan V., Zaslavska O., Trunova E., Fesenko H., Duda O. ADC for energy measurement systems of microcontroller IoT modules. In: *2019 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS 2019)*, Vol. 2. 2019. Pp. 546–551.
27. Fesenko H., Kharchenko V., Sachenko A., Hiromoto R., Kochan V. An Internet of drone-based multi-version post-severe accident monitoring system: structures and reliability. In: Kharchenko V., Kor A., Rucinski A. (eds.) *Dependable IoT for Human and Industry: Modeling, Architecting, Implementation*. River Publishers. 2018. Pp. 197–217.
28. Ashraf S., Saleem S., Ahmed T., et al. Iris and foot based sustainable biometric identification approach. 2020.
29. Winata F., Jovanka I., Laurent A., et al. Traffic prediction: a comparison between the LSTM and multi-layer perceptron algorithm. In: 2022 2nd International

Conference on Intelligent Cybernetics Technology & Applications (ICICyTA). 2022. Pp. 12–16.

30. Zhang C., Sun S., Yu G. A Bayesian network approach to time series forecasting of short-term traffic flows. In: Proceedings of the 7th International IEEE Conference on Intelligent Transportation Systems. 2004. Pp. 216–221.

31. Traffic flow forecasting data set. UCI Machine Learning Repository. URL: <https://archive.ics.uci.edu/dataset/608/traffic+flow+forecasting>.

32. Дзядик Б., Мороз Ю., Шайнюк В. Підхід до аналізу мережевого трафіку та прогнозування транспортних потоків на основі Інтернет речей, блокчейну й глибокого навчання. Proceedings of the 2nd International Scientific and Practical Conference. November 26-28, 2025. С. 316–320.

33. Шайнюк В.О. Прогнозування транспортних потоків за допомогою Інтернету речей та машинного навчання. Збірник тез доповідей II Всеукраїнської науково-практичної конференції студентів, аспірантів та молодих вчених «Інтелектуальні комп'ютерні системи та мережі» (ІКСМ осінь 2025), м. Тернопіль, ЗУНУ, 20 травня 2025 р. Тернопіль, 2025. С. 27–30.

34. Комар М.П., Саченко А.О., Васильків Н.М., Загородня Д.І. Методичні рекомендації до виконання кваліфікаційної роботи з освітньо-професійної програми «Комп'ютерні науки» спеціальності 122 «Комп'ютерні науки» за другим (магістерським) рівнем вищої освіти. Тернопіль: ЗУНУ, 2024. 32 с.

35. Островерхов В.М., Біловус Л.І., Возьний К.З., Луцишин О.О., Монастирський Г.Л., Надвиничний С.А., Питель С.В., Шандрок С.К. Загальні методичні рекомендації з підготовки, оформлення, захисту та оцінювання кваліфікаційних робіт здобувачів вищої освіти першого (бакалаврського) і другого (магістерського) рівнів. Тернопіль: ЗУНУ, 2024. 83 с.

Додаток А
Копії публікацій

isu-conference.com



COLLECTION OF SCIENTIFIC PAPERS



ISSUE
№47

2ND INTERNATIONAL SCIENTIFIC
AND PRACTICAL CONFERENCE

**PROGRESSIVE
APPROACHES
IN SCIENCE
AND ENGINEERING**

NOVEMBER 26-28, 2025
COPENHAGEN, DENMARK





2nd International Scientific and Practical Conference
**«Progressive Approaches in Science and
Engineering»**

Collection of Scientific Papers

November 26-28, 2025
Copenhagen, Denmark

UDC 001(08)

Progressive Approaches in Science and Engineering: Collection of Scientific Papers with Proceedings of the 2nd International Scientific and Practical Conference. International Scientific Unity. November 26-28, 2025. Copenhagen, Denmark. 697 p.

ISBN 979-8-89704-979-0 (series)
DOI 10.70286/ISU-26.11.2025

The conference is included in the Academic Research Index ReserchBib International catalog of scientific conferences.

The collection of scientific papers presents the materials of the participants of the 2nd International Scientific and Practical Conference "Progressive Approaches in Science and Engineering" (November 26-28, 2025. Copenhagen, Denmark).

The materials of the collection are presented in the author's edition and printed in the original language. The authors of the published materials bear full responsibility for the authenticity of the given facts, proper names, geographical names, quotations, economic and statistical data, industry terminology, and other information.

The materials of the conference are publicly available under the terms of the CC BY-NC 4.0 International license.

ISBN 979-8-89704-979-0



© Participants of the conference, 2025
© Collection of Scientific Papers "International Scientific Unity", 2025
Official site: <https://isu-conference.com/>

Mamrosh V.S. IMPROVED METHODOLOGY FOR DEFECT IDENTIFICATION IN MULTIPLAYER GAMES CASE STUDY OFF THE GRID.....	301
Липа А., Савка А. МЕТОДИ МАШИННОГО НАВЧАННЯ ТА ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ ДЛЯ ПРОГНОЗУВАННЯ РИЗИКІВ ТА УПРАВЛІННЯ ПОРТФЕЛЕМ ПРОЄКТІВ.....	305
Галин В., Аравець Р., Сичов Р. ІНТЕЛЕКТУАЛЬНІ МЕТОДИ АНАЛІЗУ ВЕЛИКИХ ДАНИХ: ВИЯВЛЕННЯ АНОМАЛІЙ, АНАЛІЗ НАСТРОЇВ ТА ПРОГНОЗУВАННЯ ЯКОСТІ В ІНТЕЛЕКТУАЛЬНОМУ ВИРОБНИЦТВІ.....	310
Sharovalova S., Chyzh Ye. ARCHITECTURAL APPROACHES TO IMPLEMENTING A ROLE- BASED ACCESS CONTROL (RBAC) MODEL FOR MODERN WEB PLATFORMS.....	314
Дзядик Б., Мороз Ю., Шайнюк В. ПІДХІД ДО АНАЛІЗУ МЕРЕЖЕВОГО ТРАФІКУ ТА ПРОГНОЗУВАННЯ ТРАНСПОРТНИХ ПОТОКІВ НА ОСНОВІ ІНТЕРНЕТ РЕЧЕЙ, БЛОКЧЕЙНУ Й ГЛИБОКОГО НАВЧАННЯ.....	316
Шелег Я.П. ОБМЕЖЕННЯ SAST ІНСТРУМЕНТІВ ПРИ ДЕТЕКЦІЇ КОНТЕКСТНО-ЗАЛЕЖНИХ ВРАЗЛИВОСТЕЙ ТА LLM- АЛЬТЕРНАТИВА.....	321
Maiko D.R., Pohorilets V.M., Maiko T.S. COMPARATIVE ANALYSIS OF CONTAINERIZATION AND VIRTUALIZATION TECHNOLOGIES IN CLOUD INFORMATION SYSTEMS DEPLOYMENT.....	323
Юрченко В.О. ПЕРЕВІРКА КОРЕКТНОСТІ ВІДПОВІДЕЙ АГЕНТІВ ШТУЧНОГО ІНТЕЛЕКТУ.....	327
Sharovalova S., Huryn I. PERSONALIZED RECOMMENDATIONS BASED ON THE PROCESSING OF TEXT DATA AND USER BEHAVIOR PATTERNS..	329
Кім В. ІЄРАРХІЯ КЕШІВ І ПОНЯТТЯ FALSE SHARING.....	333

ПІДХІД ДО АНАЛІЗУ МЕРЕЖЕВОГО ТРАФІКУ ТА ПРОГНОЗУВАННЯ ТРАНСПОРТНИХ ПОТОКІВ НА ОСНОВІ ІНТЕРНЕТ РЕЧЕЙ, БЛОКЧЕЙНУ Й ГЛИБОКОГО НАВЧАННЯ

Дзядик Богдан

здобувач вищої освіти

Мороз Юрій

здобувач вищої освіти

Шайнюк Вадим

здобувач вищої освіти

Кафедра інформаційно-обчислювальних систем і управління
Західноукраїнський національний університет, Україна

Сучасні інформаційно-комунікаційні системи, побудовані на базі Інтернету речей (Internet of Things, IoT), характеризуються високою динамічністю, географічною розподіленістю та величезними обсягами даних, що безперервно генеруються різними сенсорами, пристроями та транспортними засобами. На зміну простим телеметричним сценаріям приходять складні когнітивні IoT-системи (Cognitive IoT, CIoT), у яких дані не лише збираються та передаються, але й інтерпретуються, аналізуються та використовуються для автономного прийняття рішень за допомогою методів штучного інтелекту [1–3].

Такі системи дедалі частіше стають частиною критичної інфраструктури: інтелектуальних транспортних систем, енергомереж, систем безпеки, промислових комплексів. Будь-яке спотворення даних, збої в мережевому трафіку чи некоректна робота моделей машинного навчання можуть призвести до суттєвих економічних втрат, зниження рівня безпеки або навіть до катастрофічних наслідків. При цьому класичні підходи до забезпечення безпеки мереж зосереджуються переважно на рівні каналів зв'язку та криптографічних протоколів, не охоплюючи в повній мірі довіру до самих даних і до результатів аналітики [1, 2].

Додатковим викликом є масове шифрування мережевого трафіку. З одного боку, це необхідна умова забезпечення конфіденційності користувачів; з іншого – традиційні методи глибокого аналізу пакетів стають малоефективними або непридатними, оскільки вміст пакетів недоступний. Це ускладнює виявлення шкідливої активності, класифікацію застосунків та забезпечення якості обслуговування. У відповідь на це активно розвиваються методи класифікації зашифрованого трафіку на основі статистичних ознак і машинного навчання, причому найперспективнішими довели себе підходи на основі глибоких нейронних мереж [5, 6].

У той же час інтелектуальні транспортні системи використовують IoT-інфраструктуру для моніторингу стану дорожньої мережі: вимірювання швидкості, інтенсивності потоку, щільності транспорту, виявлення заторів та

ДТП. Дані надходять із дорожніх сенсорів, GPS-пристроїв, бортових систем, метеостанцій, відеокамер тощо. Їх інтеграція та аналіз за допомогою моделей машинного та глибокого навчання (зокрема MLP, LSTM, байєсівських мереж) дозволяють будувати точні прогнози транспортних потоків, оптимізувати налаштування світлофорів, маршрути громадського транспорту та системи інформування водіїв [7–9].

В оглядах безпеки IoT-систем виділяються численні атаки: підміна пристроїв, несанкціонований доступ, відмова в обслуговуванні (DoS/DDoS), компрометація шлюзів, ін'єкція некоректних даних тощо [1, 2]. Запропоновано різні протоколи автентифікації, схеми розподілу ключів і легковагові криптографічні алгоритми, проте більшість рішень фокусуються на сеансовому рівні (захист каналу зв'язку), тоді як проблема довіри до даних та моделей III часто залишається поза увагою.

Концепція Cognitive IoT (CIoT) розширює класичний IoT за рахунок поєднання сенсорних даних, контекстної інформації, знань і механізмів навчання, що дозволяє системі адаптуватися до змін середовища [3]. Водночас саме активне використання моделей III робить CIoT вразливим до специфічних атак на дані та моделі: data poisoning, коли зловмисник підмішує у тренувальну вибірку шкідливі приклади; adversarial attacks, коли вхідні дані модифікуються таким чином, щоб змусити модель помилитися; model stealing, коли зловмисник намагається відтворити параметри моделі [4].

Серед технологій, що здатні підвищити прозорість і довіру, особливе місце займає блокчейн – розподілений реєстр транзакцій, який гарантує незмінність записів та дозволяє відстежувати походження даних. Ряд робіт пропонує інтегрувати блокчейн з IoT для реєстрації подій, конфігурацій пристроїв, логів доступу та навіть параметрів ML-моделей [1, 2]. Проте залишається відкритим питання ефективної інтеграції блокчейна з обмеженими за ресурсами IoT-вузлами та з високорівневими ML-пайплайнами.

Класифікація мережевого трафіку традиційно виконувалася на основі аналізу вмісту пакетів, сигнатур або портів. Шифрування (наприклад, TLS) робить payload недоступним, а використання динамічних портів та тунелювання обходить прості евристики. У відповідь розвиваються методи, що аналізують метадані: довжину пакетів, час між пакетами, напрямок передачі, статистичні розподіли [5].

Класичні ML-підходи будують вектор ознак на основі цих статистичних характеристик і застосовують алгоритми Random Forest, SVM, KNN тощо [5, 6]. Вони досягають прийнятної точності, але якість значною мірою залежить від ручної інженерії ознак. У сучасних роботах демонструється, що глибокі нейронні мережі – зокрема CNN, LSTM та їх комбінації – здатні автоматично виділяти релевантні просторово-часові патерни в послідовностях пакетів і досягати значно вищої точності без необхідності глибокого ручного налаштування ознак [5, 6].

CNN добре працюють з локальними шаблонами (наприклад, характерними комбінаціями довжин чи інтервалів), тоді як LSTM моделюють довгострокові

залежності у послідовностях. Гібридні архітектури CNN+LSTM дозволяють поєднати обидві переваги: CNN-частина виконує попереднє вилучення ознак, а LSTM-частина – їх часову агрегацію. Результати таких моделей, за даними [5, 6], демонструють точність, яка перевершує класичні ML-методи, зокрема для задач розпізнавання застосунків і VPN-каналів на зашифрованому трафіку.

У сфері транспортних систем довгий час домінували класичні моделі часових рядів (ARIMA, SARIMA) та регресійні підходи. Однак вони погано враховують нелінійність, сезонність і залежність від зовнішніх факторів (погода, події, дорожні роботи). Із поширенням IoT виникла можливість збирати багатоджерельні дані: сигнали зі стаціонарних сенсорів, GPS-треки, дані з навігаційних сервісів, метеодані тощо. На основі таких даних почали розроблятися моделі машинного навчання та глибокого навчання [7].

Порівняння результатів MLP, LSTM та інших DL-архітектур показує, що глибокі моделі здатні краще вловлювати складні часові взаємозв'язки та взаємодії між параметрами трафіку, особливо при короткостроковому прогнозуванні (кілька хвилин або десятків хвилин уперед) [7]. Байєсівські мережі застосовуються для моделювання ймовірнісних залежностей та оцінки невизначеності прогнозів [8]. Для експериментів широко використовуються відкриті набори даних, такі як Traffic Flow Forecasting Dataset з репозиторію UCI, що дозволяє відтворювати результати й порівнювати моделі в однакових умовах [9].

Недоліком більшості рішень є відсутність тісного зв'язку з питаннями безпеки даних та довіри до джерел, а також слабка інтеграція з підсистемами мережевого моніторингу. Це відкриває простір для комплексних рішень, де прогнозування транспортних потоків спирається на захищені CIoT-дані та узгоджене з методами аналізу мережевого трафіку.

Архітектура CIoT із блокчейн-шаром довіри. Запропонована архітектура складається з кількох рівнів. На сенсорному рівні розташовані IoT-пристрої, які вимірюють фізичні параметри (температуру, швидкість, положення, стан мережі тощо). Вони передають дані на граничні вузли (edge), де здійснюється попередня агрегація, фільтрація та, за потреби, локальна аналітика.

На хмарному рівні розгортається блокчейн-мережа, вузли якої отримують агреговані дані від граничних пристроїв у вигляді транзакцій. Кожна транзакція включає хеш даних, часову мітку, ідентифікатор джерела та метадані про сценарій збору. Після досягнення консенсусу блок додається до ланцюга, забезпечуючи незмінність історії.

ML-пайплайни працюють з «офчейн»-сховищами, які зберігають фактичні масиви даних, але прив'язуються до блокчейн-записів через контрольні суми й ідентифікатори. Це дозволяє відтворювати експерименти, контролювати цілісність датасетів та відслідковувати зміни моделей (нові версії, перенавчання) через запис у блокчейн відповідних подій [1–4].

Таким чином, блокчейн виступає шаром довіри, а CIoT-інфраструктура – джерелом багатих даних для моделей прогнозування й класифікації.

Глибокі моделі для класифікації зашифрованого трафіку. Метод класифікації зашифрованого трафіку включає кілька етапів. На першому етапі формується датасет, який містить набори мережевих сесій, записаних у реальних умовах (або з публічних репозиторіїв), де кожна сесія має мітку класу (тип застосунку, сервісу чи протоколу) [5, 6].

Другий етап – виділення ознак. Із кожної сесії обчислюються послідовності довжин пакетів, інтервалів часу між ними, напрямків (вхідний/вихідний), кількість пакетів у потоці, статистичні агрегати (середнє, дисперсія, квантілі). Отримані послідовності перетворюються у вектори або матриці фіксованої довжини, придатні для подачі в глибоку модель.

Третій етап – моделювання. Застосовується гібридна CNN+LSTM-архітектура, де:

- CNN-шари витягують локальні шаблони зі структури потоку (наприклад, характерні «сигнатури» довжин пакетів і інтервалів);

- LSTM-шари моделюють часові залежності та довгострокові залежності в послідовності;

- вихідний Dense-шар з softmax-активацією виконує класифікацію.

Четвертий етап – оцінювання якості. Моделі порівнюються з класичними ML-алгоритмами (Random Forest, SVM, KNN) за точністю, повнотою, прецизійністю, F1-мірою та AUC. Результати експериментів у літературі демонструють суттєвий приріст точності глибоких моделей, особливо у задачах розрізнення близьких за поведінкою типів трафіку.

Такий підхід може бути інтегрований у CIoT-архітектуру як інтелектуальний модуль мережевої безпеки, що працює на основі статистики зашифрованих пакетів.

Прогнозування транспортних потоків. Для прогнозування транспортних потоків пропонується IoT-орієнтована архітектура, де:

- датчики та бортові пристрої формують потоки даних (швидкість, інтенсивність, координати, метеоумови);

- на граничних вузлах виконується локальне згладжування, виявлення аномалій, усунення пропусків;

- у хмарній частині дані інтегруються, масштабуються та подаються в ML-моделі [7–9].

Для побудови моделей застосовуються:

- MLP – як базова нелінійна модель для регресії;

- LSTM – для моделювання часових залежностей і сезонності в трафіку;

- байєсівські мережі – для врахування невизначеності та ймовірнісних залежностей [7, 8].

В якості навчальної та тестової вибірок можуть використовуватися публічні дані, наприклад, Traffic Flow Forecasting Dataset з UCI [9], де наявні часові ряди інтенсивності руху, виміряні в різні часові проміжки. Після навчання моделі порівнюються за RMSE, MAE, MAPE та іншими метриками. Практика показує, що LSTM зазвичай перевершує MLP та класичні моделі часових рядів, особливо в короткостроковому прогнозуванні.

Вбудувавши такі моделі у CІoT-архітектуру, можна створити інтелектуальну службу прогнозування, яка використовує довірені дані з блокчейн-реєстру та забезпечує проактивне керування дорожнім рухом.

Отже, в даному дослідженні:

1. Показано, що інтеграція блокчейн-технології з ML-пайплайнами у CІoT-системах дозволяє підвищити довіру до даних та результатів аналітики, забезпечуючи незмінність, простежуваність і відтворюваність моделей та експериментів;

2. Проаналізовано та узагальнено сучасні підходи до класифікації зашифрованого трафіку. Доведено доцільність застосування глибоких нейронних мереж, які здатні забезпечувати високу точність класифікації без доступу до вмісту пакетів.

3. Розглянуто IoT+ML-підхід до прогнозування транспортних потоків, де багатоджерельні дані з сенсорів і бортових пристроїв аналізуються за допомогою моделей MLP, LSTM та байєсівських мереж. Продемонстровано, що LSTM-моделі на основі відкритих датасетів здатні забезпечувати високу точність короткострокового прогнозування.

Список використаних джерел

1. Das, A. K., Zeadally, S., & He, D. (2018). Taxonomy and analysis of security protocols for Internet of Things. *Future Generation Computer Systems*, 89, 110–125.
2. Wazid, M., Das, A. K., Shetty, S., Gope, P., & Rodrigues, J. J. P. C. (2021). Security in 5G-enabled Internet of Things communication: Issues, challenges, and future research roadmap. *IEEE Access*, 9, 4466–4489.
3. Wu, Q., Ding, G., Xu, Y., Feng, S., Du, Z., Wang, J., & Long, K. (2014). Cognitive Internet of Things: A new paradigm beyond connection. *IEEE Internet of Things Journal*, 1(2), 129–143.
4. Ren, K., Zheng, T., Qin, Z., & Liu, X. (2020). Adversarial attacks and defenses in deep learning. *Engineering*, 6(3), 346–360.
5. Shi, H., Li, H., Zhang, D., Cheng, C., & Cao, X. (2018). An efficient feature generation approach based on deep learning and feature selection techniques for traffic classification. *Computer Networks*, 132, 81–98.
6. Zhang, Z., Kang, C., Fu, P., Cao, Z., Li, Z., & Xiong, G. (2017). Metric learning with statistical features for network traffic classification. In *2017 IEEE 36th International Performance Computing and Communications Conference (IPCCC)* (pp. 1–7). IEEE.
7. Winata, F., Jovanka, I., & Laurent, A. (2022). Traffic prediction: A comparison between the LSTM and multi-layer perceptron algorithm. In *2022 2nd International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA)* (pp. 12–16). IEEE.
8. Zhang, C., Sun, S., & Yu, G. (2004). A Bayesian network approach to time series forecasting of short-term traffic flows. In *Proceedings of the 7th International IEEE Conference on Intelligent Transportation Systems* (pp. 216–221). IEEE.
9. UCI Machine Learning Repository. (2020). Traffic flow forecasting data set [Data set]. University of California, Irvine.

ЗАХІДНОУКРАЇНСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ФАКУЛЬТЕТ КОМП'ЮТЕРНИХ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ
КАФЕДРА КОМП'ЮТЕРНОЇ ІНЖЕНЕРІЇ



Комп'ютерна
Інженерія



**III ВСЕУКРАЇНСЬКА НАУКОВО-ПРАКТИЧНА
КОНФЕРЕНЦІЯ СТУДЕНТІВ, АСПІРАНТІВ ТА
МОЛОДИХ ВЧЕНИХ
«ІНТЕЛЕКТУАЛЬНІ КОМП'ЮТЕРНІ СИСТЕМИ ТА
МЕРЕЖІ»**

***ІКСМ
ОСІНЬ 2025***

25 ЛИСТОПАДА 2025



KI.WUNU.EDU.UA/CONFERENCE/

ТЕРНОПІЛЬ

2025



ЗМІСТ

<i>Березька К. М., Цимбалюк Л. В.</i> Цифрові засоби формування логічного мислення у процесі підготовки до ТЗНК	9
<i>Ковтуненко А.Р.</i> Мультимодальна висхідна сегментація об'єктів за текстовим запитом	11
<i>Андрухів Б.І., Воронній В.А.</i> Сучасні технології створення програмних засобів генерування звуків природніх мов	13
<i>Квітень Д.О.</i> Алгоритми класифікації режимів енергоспоживання для зниження пікових навантажень в розумному будинку.....	15
<i>Савка А.П.</i> Управління портфелем проєктів з використанням засобів штучного інтелекту	17
<i>Луца А.В.</i> Методи машинного навчання для прогнозування та управління ризиками в інфраструктурних проєктах.....	21
<i>Мороз Ю.П.</i> Нейромережева модель глибокого навчання для класифікації мережевих пакетів	24
<i>Шайнюк В.О.</i> Прогнозування транспортних потоків за допомогою Інтернету речей та машинного навчання.....	27
<i>Дзядик Б.-Д.Ю</i> Інтеграція блокчейн-технології та штучного інтелекту для аналізу великих даних у середовищі Інтернету речей.....	30
<i>Сичов Р.С.</i> Модель машинного навчання для аналізу та прогнозування якості в процесах інтелектуального виробництва	33
<i>Каравець Р.О.</i> Аналіз настроїв в соціальних мережах на основі технологій великих даних	37
<i>Галин В.А.</i> Методи динамічного та статичного виявлення аномалій у великих даних.....	39
<i>Горяча І.В.</i> Автоматизований підхід до огляду літератури з використанням великих мовних моделей	43
<i>Киричук Д.О.</i> Дослідження ефективності застосування Slicing Aided Hyper Inference для виявлення малих об'єктів на зображеннях високої роздільної здатності	45
<i>Гуда Ю.Ю.</i> Застосування методів машинного навчання для прогнозування запахів на основі молекулярної структури	48
<i>Загрійчук В. І.</i> Аналіз способів автоматизації ділової комунікації в організаціях.....	50
<i>Панасюк Н.Р.</i> Метод та засоби відлагодження програмного забезпечення для інтелектуальних давачів наземної мобільної робототехнічної платформи.....	52
<i>Чайківська І.Р.</i> Модель та засоби оцінки дизайну ІТ-продуктів.....	55

Шайнюк В.О.
 магістрант 2 курсу ФКІТ ЗУНУ
 Науковий керівник к.т.н., професор Кочан В.В., кафедра ІОСУ ЗУНУ

ПРОГНОЗУВАННЯ ТРАНСПОРТНИХ ПОТОКІВ ЗА ДОПОМОГОЮ ІНТЕРНЕТУ РЕЧЕЙ ТА МАШИННОГО НАВЧАННЯ

Вступ. Проблема транспортних заторів є одним із ключових викликів сучасних міст [1]. Затори мають виразну безпекову, економічну та екологічну складові. Вони підвищують імовірність дорожньо-транспортних пригод, призводять до пошкодження транспорту, травмування учасників руху й іноді до летальних наслідків. Водії та пасажери зазнають втрат часу, збільшується споживання пального та викиди шкідливих речовин, погіршується якість життя в урбанізованих зонах. Тому управління транспортними потоками та зниження рівня заторів слід розглядати не лише як інженерне, а як комплексне соціально-економічне завдання.

Одним із базових інструментів підтримки рішень у цій сфері є прогнозування транспортних потоків. Його метою є оцінка майбутніх змін у використанні різних видів транспорту, інтенсивності руху на окремих ділянках та завантаженості маршрутів. Вихідним етапом виступає збір даних про поточний стан руху (інтенсивність, швидкість, затримки, тривалість поїздки) і супровідні параметри середовища. На основі цих даних формуються набори ознак, які можуть включати демографічні характеристики, структуру зайнятості, тарифи, вартість пального, показники економічної активності та метеорологічні умови [2].

Традиційно транспортні моделі спиралися на дані з окремих сенсорів – індукційних петель, фотоелектричних та п'єзоелектричних датчиків, установлених уздовж доріг [2]. Такі системи давали змогу формувати часові ряди інтенсивності та будувати короткострокові прогнози. З поширенням методів машинного навчання стало можливим враховувати складні нелінійні залежності, взаємодії між параметрами трафіку та інформацію з альтернативних маршрутів, що підвищило точність прогнозів.

Сучасний розвиток транспортної галузі пов'язаний із широким впровадженням Інтернету речей (IoT) [4] та інтелектуальних транспортних систем (ITS) [5]. Масове розгортання датчиків у дорожній інфраструктурі та транспортних засобах забезпечує безпрецедентні обсяги даних, які оновлюються майже в реальному часі й охоплюють позицію, швидкість, тип транспортного засобу, стан покриття, погоду тощо. Поєднання IoT-інфраструктури з алгоритмами машинного навчання, на нашу думку, відкриває якісно нові можливості для побудови прогнозних моделей транспортних потоків.

Отже, постає науково-практичне завдання розробити метод прогнозування транспортних потоків на основі даних Інтернету речей та алгоритмів машинного навчання, який забезпечуватиме підвищену точність прогнозів завдяки інтеграції динамічних сенсорних даних, спеціалізованих процедур обробки й відбору ознак та адаптації моделі до змінних умов функціонування транспортної системи.

Постановка задачі. Попри значний прогрес у розвитку ITS, наявні рішення мають низку обмежень, які суттєво звужують їхню ефективність у реальних умовах. Ці проблеми пов'язані як із вибором технічної інфраструктури збору даних, так і з обробкою великих масивів інформації та особливостями алгоритмічних підходів. Існуючі ITS стикаються з комплексом проблем: високою вартістю розгортання WSN-інфраструктури, труднощами обробки великих обсягів різномірних даних, залежністю від людських рішень, обмеженнями відеоорієнтованих підходів та недостатньою точністю деяких навігаційних сервісів у нештатних ситуаціях. У сукупності ці фактори формують підґрунтя для розроблення нових підходів до прогнозування транспортних потоків, які б ефективно поєднували можливості Інтернету речей, методів машинного навчання та процедур відбору ознак.

Об'єкт дослідження – процеси формування та динаміки транспортних потоків в умовах функціонування інтелектуальних транспортних систем на базі Інтернету речей. Предмет дослідження – методи та моделі прогнозування транспортних потоків на основі даних, отриманих від IoT-пристроїв, зокрема архітектура системи збору й обробки даних, процедури

попередньої обробки та відбору ознак і алгоритми машинного навчання для прогнозування стану трафіку. Метою дослідження є розробка підходу до прогнозування транспортних потоків на основі даних Інтернету речей та машинного навчання, який забезпечує підвищення точності та стійкості прогнозів за рахунок інтеграції сенсорних даних, процедур попередньої обробки та відбору інформативних ознак.

Основний матеріал. Ефективність прогнозування транспортних потоків безпосередньо залежить від того, як саме організовані процеси збору, передавання, зберігання та обробки даних. Тому перед розробленням алгоритмів машинного навчання доцільно чітко сформулювати концепцію й архітектуру системи, яка забезпечуватиме повний цикл роботи з транспортною інформацією – від сенсора на дорозі до прогнозу, що може бути використаний в ITS.

У даному дослідженні архітектура системи збору й обробки даних розглядається як багатоплатформна IoT/LoE-орієнтована платформа, що об'єднує:

- джерела даних (сенсорні мережі, бортові пристрої, інфраструктурні датчики);
- мережеву інфраструктуру і засоби передавання;
- хмарне або серверне сховище;
- програмні модулі попередньої обробки та відбору ознак;
- модулі прогнозування та інтерфейси для інтеграції з ITS.

Умовно таку архітектуру (рисунок 1) можна подати у вигляді декількох рівнів: рівень фізичного середовища (речі), комунікаційний рівень, рівень даних, аналітичний рівень та рівень прикладних сервісів.

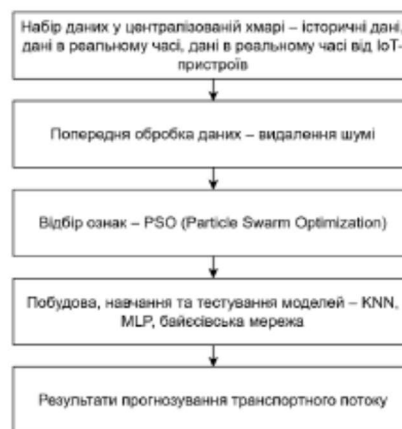


Рисунок 1 – Архітектура системи прогнозування транспортного потоку на основі IoT та методів машинного навчання

Першою складовою концепції є визначення джерел транспортних даних, які забезпечують повноту й різноманітність інформації. У контексті прогнозування транспортних потоків доцільно виділити такі групи IoT-джерел:

1. Сенсорні мережі вздовж дорожньої інфраструктури. До цієї групи належать: індукційні петлі в дорожньому полотні, радарні та ультразвукові датчики, п'єзоелектричні сенсори, локальні метеостанції (температура, опади, видимість). Ці пристрої реєструють інтенсивність потоку, середню швидкість, заповненість смуг, наявність черг, стан покриття, а також погодні умови, що впливають на керування транспортних засобів.

2. Бортові пристрої транспортних засобів. Сюди належать: GPS-трекери, телематичні модулі, модулі OBD/CAN, вбудовані навігаційні системи. Вони забезпечують траєкторії руху, миттєву швидкість, час поїздки, режим руху (зупинка, рух, затримка). На нашу думку, саме бортові пристрої є цінним джерелом деталізованої інформації про поведінку окремого автомобіля всередині потоку.

3. Інфраструктурні датчики та системи відеоспостереження. До цієї категорії належать: «розумні» світлофори, камери відеоспостереження на перехрестях і магістралях, сенсори на

парковках, в тунелях, на мостах та розв'язках.

Окрім прямих вимірювань, такі системи дозволяють будувати агреговані показники завантаженості вузлів мережі, а у випадку відео – оцінювати транспортні потоки за допомогою комп'ютерного зору.

4. Додаткові джерела (за потреби). За необхідності архітектура може бути розширена даними: метеослужб (регулярні погодні зведення), відкритих міських порталів (плани ремонтів, перекриття руху), краудсорсингових сервісів (повідомлення водіїв про ДТП, затори).

Утім, основний акцент у роботі робиться на «класичних» IoT-джерелах: сенсорні мережі, бортові пристрої, інфраструктурні датчики.

Центральним елементом архітектури є сховище даних, де об'єднуються:

- дані реального часу, що надходять від IoT-пристроїв;
- історичні транспортні дані, накопичені за попередні періоди.

Централізоване хмарне сховище виконує такі функції:

1. Збирання та логування потоків даних. Усі події (вимірювання, оновлення, сигнали) фіксуються із зазначенням часу, джерела, географічних координат та контекстних параметрів (погода, тип дороги тощо).

2. Зберігання структурованих наборів даних. Після первинного перетворення дані організовуються у вигляді таблиць або колекцій, де кожен запис описує стан трафіку на певній ділянці за конкретний часовий інтервал.

За своєю суттю сховище виконує роль єдиного джерела правди для всіх модулів системи: як для блоків попередньої обробки, так і для аналітичних сервісів.

Для розуміння логіки роботи платформи важливо описати основні інформаційні потоки:

1. Від сенсорів до хмарного сховища:

- сенсор або бортовий пристрій генерує вимірювання;
- вимірювання передається до локального шлюзу;
- шлюз агрегує дані та відправляє їх до хмари або центрального сервера;
- у сховищі вимірювання логуються та прив'язуються до відповідних часових та просторових індексів.

2. Від сховища до модуля попередньої обробки:

- за визначеним розкладом або тригером формується вибірка даних для обробки (наприклад, за останню годину чи добу);
- дані передаються в модуль *preprocessing*, де здійснюється очищення, обробка пропусків, нормалізація, виявлення викидів.

3. Від модуля попередньої обробки до модуля відбору ознак.

- очищений набір даних передається в модуль *feature selection*;
- за допомогою алгоритму PSO або інших процедур формується оптимальна підмножина ознак, що найкраще пояснює варіацію цільової змінної (інтенсивність/швидкість/щільність потоку).

4. Від модуля відбору ознак до моделей машинного навчання – сформований набір ознак використовується: для навчання моделей (KNN, MLP, байєсівські мережі) на історичних даних; для побудови прогнозів на основі останніх доступних спостережень.

5. Від моделей до прикладних сервісів ITS – результати прогнозування (очікувана інтенсивність, швидкість, ризик затору) передаються: у системи керування світлофорами (для адаптації фаз); у інформаційні панелі чи навігаційні сервіси (для інформування користувачів); у модулі стратегічного планування (для оцінки необхідності змін у схемі руху, інфраструктурних рішень тощо).

Таким чином, інформаційні потоки організовані так, щоб забезпечити замкнений цикл: «збір – зберігання – аналіз – прогноз – керуючий вплив – нові дані». Це дозволяє системі поступово вдосконалюватися за рахунок накопичення досвіду та донавчання моделей.

Висновки. Сформовано цілісну концепцію та багатопшарову архітектуру IoT-орієнтованої системи збору й обробки транспортних даних, яка охоплює сенсорний шар, комунікаційну інфраструктуру, хмарне сховище, модулі попередньої обробки та аналітичні сервіси. Така архітектура забезпечує замкнений цикл «збір – зберігання – аналіз – прогноз – керуючий вплив»

і є основою для інтеграції з інтелектуальними транспортними системами.

Список літератури

1. Boukerche A., Tao Y., Sun P. Artificial intelligence-based vehicular traffic flow prediction methods for supporting intelligent transportation systems. *Computer Networks*. 2020. Vol. 182. Art. 107484.
2. Ghadi Y.Y., Mazhar T., al Shloul T., et al. Machine learning solution for the security of wireless sensor network. *IEEE Access*. 2024. Vol. 12. Pp. 12699–12719.
3. Shu W., Cai K., Xiong N.N. A short-term traffic flow prediction model based on an improved gate recurrent unit neural network. *IEEE Transactions on Intelligent Transportation Systems*. 2022. Vol. 23(9). Pp. 16654–16665.
4. Ghadi Y.Y., Mazhar T., Shah S.F.A., et al. Integration of federated learning with IoT for smart cities applications, challenges, and solutions. *PeerJ Computer Science*. 2023. Vol. 9. Art. e1657.
5. Gohar M., Muzammal M., Rahman A.U. Smart TSS: defining transportation system behavior using big data analytics in smart cities. *Sustainable Cities and Society*. 2018. Vol. 41. Pp. 114–119.

Дзядик Б.-Д.Ю.

магістрант 2 курсу ФКІТ ЗУНУ

Науковий керівник к.т.н., професор Кочан В.В., кафедра ІОСУ ЗУНУ

ІНТЕГРАЦІЯ БЛОКЧЕЙН-ТЕХНОЛОГІЙ ТА ШТУЧНОГО ІНТЕЛЕКТУ ДЛЯ АНАЛІЗУ ВЕЛИКИХ ДАНИХ У СЕРЕДОВИЩІ ІНТЕРНЕТУ РЕЧЕЙ

Вступ. Стрімкий розвиток Інтернету речей (IoT) зумовлює безпрецедентне зростання обсягу, швидкості та різноманітності даних, що генеруються сенсорами, вбудованими пристроями та кіберфізичними системами. У таких умовах традиційні централізовані підходи до збирання та опрацювання даних стикаються з обмеженнями масштабованості, затримок, довіри й кіберстійкості [1]. Паралельно із цим, блокчейн-технологія пропонує незмінний розподілений реєстр транзакцій і механізми децентралізованого узгодження, а методи штучного інтелекту (ШІ) та машинного навчання (МН) забезпечують здатність виявляти закономірності, прогнозувати стани та підтримувати прийняття рішень у режимі наближеному до реального часу. Інтеграція блокчейна та ШІ у середовищі IoT постає як перспективний напрям, що поєднує гарантії цілісності та простежуваності даних із когнітивними можливостями аналітики великих даних [2, 3].

Постановка задачі. Когнітивний Інтернет речей (Cognitive IoT, CIoT) формується як еволюційна модель мережі, що розвиває ідеї IoT. Подібно до класичного IoT, у CIoT взаємодіють фізичні й віртуальні об'єкти з мінімальною участю людини, проте комунікація та керування ґрунтуються на контекстно-обізнаному циклі «сприйняття → дія» [4]. CIoT застосовує підхід *understanding-by-building*: система навчається як із соціальних мереж, так і з фізичного середовища, зберігаючи семантичні подання знань у відповідних базах; надалі вона адаптується до невизначеності та змін за допомогою ресурсоефективних методів прийняття рішень.

Незважаючи на активний розвиток кожної зі складових – IoT, блокчейна та ШІ – їхня інтеграція у цілісну архітектуру залишається складною задачею. Виникає низка проблем: як забезпечити довіру до даних на всьому шляху їхнього життєвого циклу; яким чином мінімізувати ризики отруєння даних і маніпуляцій під час навчання моделей; як синхронізувати децентралізоване збирання та зберігання з вимогами до продуктивності й затримок; як масштабувати консенсус без втрати пропускнуої здатності; як адаптувати МН-пайплайни до відмовостійких, енергоефективних та ресурсно-обмежених сценаріїв «країв» мережі. Відсутність узгодженої моделі інтеграції спричиняє фрагментацію рішень і ускладнює їх практичне впровадження в промислових, міських, аграрних та критичних інфраструктурах.

Актуальність даного дослідження підсилюється зростанням частоти інцидентів, пов'язаних із порушенням цілісності й конфіденційності даних, та потребою у відтворюваних, верифікованих аналітичних результатах. У контексті IoT навіть невелика частка спотворених або