**ДЕРЖАВНИЙ УНІВЕРСИТЕТ**
**ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ**
НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ ТЕЛЕКОМУНІКАЦІЙ
Кафедра Мобільних та відеоінформаційних технологій

# КВАЛІФІКАЦІЙНА РОБОТА

## на тему: «Методи позиціонування по 5G-базових станціях високої точності»

на здобуття освітнього ступеня магістра
зі спеціальності 172 Телекомунікації та радіотехніка
освітньо-професійної програми Апаратура радіозв'язку, радіомовлення і телебачення

*Кваліфікаційна робота містить результати власних досліджень. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело*

(підпис) Евеліна ЮРЧУК (Ім'я, ПРІЗВИЩЕ)

Виконала: здобувачка вищої освіти
групи АРДМ-62
Евеліна ЮРЧУК

Керівник: Лариса ДАКОВА

Рецензент: А. Гриценевич

Київ – 2023

**ДЕРЖАВНИЙ УНІВЕРСИТЕТ**
**ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ**
**Навчально-науковий інститут телекомунікацій**

Кафедра Мобільних та відеоінформаційних технологій
Ступінь вищої освіти Магістр
Спеціальність 172 Телекомунікації та радіотехніка
Освітньо-професійна програма Апаратура радіозв'язку, радіомовлення і телебачення

**ЗАТВЕРДЖУЮ**
Завідувачка кафедрою МВТ
_____ Наталія РУДЕНКО
«  » _____ 2023 року

# ЗАВДАННЯ НА КВАЛІФІКАЦІЙНУ РОБОТУ

Юрчук Евеліні Ігорівні

1. Тема кваліфікаційної роботи: «Методи позиціонування по 5G-базових станціях високої точності»,
керівник кваліфікаційної роботи Лариса ДАКОВА, к.т.н., доцент,
затверджені наказом Державного університету інформаційно-комунікаційних технологій від 19.10.2023 р. №145.

2. Строк подання кваліфікаційної роботи 25.12.2023 р.

3. Вихідні дані до кваліфікаційної роботи:
   1. Безпроводові мережі.
   2. 5G-мережа.
   3. Методи машинного навчання.
   4. Науково-технічна література.

4. Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно розробити):
   1. Постановка проблеми та сценарії, які досліджуються;
   2. Теоретичні передумови;
   3. Методи та алгоритми;
   4. Оцінка ефективності.

5. Перелік ілюстративного матеріалу: презентація.

6. Дата видачі завдання _____

## КАЛЕНДАРНИЙ ПЛАН

| № з/п | Назва етапів кваліфікаційної роботи | Строк виконання етапів роботи | Примітка |
|---|---|---|---|
| 1. | Підбір науково-технічної літератури | 19.10.23 | Викон. |
| 2. | Постановка проблеми та сценарії, які досліджуються | 31.10.23 | Викон. |
| 3. | Теоретичні передумови | 15.11.23 | Викон. |
| 4. | Методи та алгоритми | 29.11.23 | Викон. |
| 5. | Оцінка ефективності | 07.12.23 | Викон. |
| 6. | Висновки, вступ, реферат | 11.12.23 | Викон. |
| 7. | Розробка презентації | 18.12.23 | Викон. |

**Здобувачка вищої освіти** _____ Евеліна ЮРЧУК
*(підпис)* *(Ім'я, ПРІЗВИЩЕ)*

**Керівник кваліфікаційної роботи** _____ Лариса ДАКОВА
*(підпис)* *(Ім'я, ПРІЗВИЩЕ)*

# РЕФЕРАТ

Текстова частина кваліфікаційної роботи на здобуття освітнього ступеня магістра: 83 стор., 26 рис., 4 табл., 23 дж.

*Мета роботи* – дослідження ефективності застосування методів машинного навчання в 5G-мережах для позиціонування по базовим станціям.

*Об'єкт дослідження* - процес машинного навчання в 5G-мережах для позиціонування по базовим станціям.

*Предмет дослідження* – телекомунікаційна 5G-мережа.

*Короткий зміст роботи*: У магістерській роботі проведено широке дослідження методів машинного навчання в 5G-мережах для позиціонування по базовим станціям. В роботі показано, позиціонування визнано важливим елементом застосуванням для довгострокової еволюції (LTE) та стільникових мереж п'ятого покоління (5G). Це пов'язано з його потенціалом для масових випадків комерційного використання, наприклад, автоматизації промисловості, віддаленої експлуатації та викликів надзвичайних ситуацій. Визначено, що алгоритми машинного навчання здатні позиціонувати обладнання користувача з похибкою горизонтального позиціонування менше 10 метрів у 80 відсотках тестових випадків. Для підвищення точності позиції за допомогою алгоритмів машинного навчання було застосовано фільтр Калмана. Результати показують, що це може значно підвищити точність позиції в сценаріях NLOS.

*КЛЮЧОВІ СЛОВА:* БЕЗПРОВОДОВА МЕРЕЖА, 5G-МЕРЕЖА, МАШИННЕ НАВЧАННЯ, ТЕЛЕКОМУНІКАЦІЙНА ІНФРАСТРУКТУРА, RANDOM FOREST, НЕЙРОННІ МЕРЕЖІ, МЕТОД, БАЗОВА СТАНЦІЯ

## ABSTRACT

Text part of the master's qualification work: 83 pages, 26 pictures, 4 table, 23 sources.

*The purpose of the work* - study of the effectiveness of the application of machine learning methods in 5G networks for positioning by base stations.

*Object of research* – the process of machine learning in 5G networks for positioning by base stations.

*Subject of research* – 5G telecommunication network.

*Summary of the work:* In the master's work, an extensive study of machine learning methods in 5G networks for positioning by base stations was carried out. The paper shows that positioning is recognized as an important element of applications for long-term evolution (LTE) and fifth generation (5G) cellular networks. This is due to its potential for mass commercial use cases such as industrial automation, remote operation and emergency calls. It was determined that machine learning algorithms are able to position the user's equipment with a horizontal positioning error of less than 10 meters in 80 percent of test cases. A Kalman filter was applied to improve position accuracy using machine learning algorithms. The results show that it can significantly improve position accuracy in NLOS scenarios.



*KEYWORDS:* WIRELESS NETWORK, 5G NETWORK, MACHINE LEARNING, TELECOMMUNICATION INFRASTRUCTURE, RANDOM FOREST, NEURAL NETWORKS, METHOD, BASE STATION

**ДЕРЖАВНИЙ УНІВЕРСИТЕТ**
**ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ**
**Навчально-науковий інститут телекомунікацій**

# ПОДАННЯ
# ГОЛОВІ ЕКЗАМЕНАЦІЙНОЇ КОМІСІЇ
# ЩОДО ЗАХИСТУ КВАЛІФІКАЦІЙНОЇ РОБОТИ
**на здобуття освітнього ступеня магістра**

Направляється здобувачка Юрчук Е.І. до захисту кваліфікаційної роботи
*(прізвище та ініціали)*

за спеціальністю 172 Телекомунікації та радіотехніка
освітньо-професійної програми Апаратура радіозв'язку, радіомовлення і телебачення

на тему: Методи позиціонування по 5G-базових станціях високої точності

Кваліфікаційна робота і рецензія додаються.

Директор ННІТ _____ Владислав КРАВЧЕНКО
*(підпис)* *(Ім'я, ПРІЗВИЩЕ)*

**Висновок керівника кваліфікаційної роботи**

Здобувачка Юрчук Евеліна Ігорівна показала гарну теоретичну та інженерну підготовку, уміння володіти технологіями нового покоління, вміло працювати з програмно-апаратними засобами мобільного зв'язку, довідковою і науково-технічною літературою. Працюючи над завданнями, які доручались керівником, проявила ініціативність, сумлінність та хист до інженерної роботи.

Все це дозволяє оцінити виконану кваліфікаційну роботу Юрчук Евеліни Ігорівни на оцінку «відмінно» та присвоїти їй кваліфікацію «Магістр з телекомунікацій та радіотехніки».

Керівник кваліфікаційної роботи _____ Лариса ДАКОВА
*(підпис)* *(Ім'я, ПРІЗВИЩЕ)*

“ ” _____ 2023 року

**Висновок кафедри про кваліфікаційну роботу**

Кваліфікаційна робота розглянута. Здобувачка Юрчук Е.І.
*(прізвище та ініціали)*

допускається до захисту даної роботи в Екзаменаційній комісії.

Завідувачка кафедри
Мобільних та відеоінформаційних технологій _____ Наталія РУДЕНКО
*(підпис)* *(Ім'я, ПРІЗВИЩЕ)*

“ ” _____ 2023 року

**ВІДГУК РЕЦЕНЗЕНТА**

на кваліфікаційну магістерську роботу

здобувачки вищої освіти Юрчук Евеліни Ігорівни

на тему: «Методи позиціонування по 5G-базових станціях високої точності»

**Актуальність:**

Позиціонування визнано важливим елементом застосуванням для довгострокової еволюції (LTE) та стільникових мереж п'ятого покоління (5G). Це пов'язано з його потенціалом для масових випадків комерційного використання, наприклад, автоматизації промисловості, віддаленої експлуатації та викликів надзвичайних ситуацій. Робота базується на даних про позиціювання та багаторазові радіовимірювання конкретного 5G пристрою.

Для моделювання зв'язку між вимірюваннями та положеннями розглянуто дві нелінійні моделі – це нейромережа та random forest моделі. Ці нелінійні моделі мають назву алгоритмами машинного навчання.

Тому тема обраних досліджень є актуальною та своєчасною.

**Позитивні сторони:**

Проведені дослідження свідчать про високий науково-технічний рівень використання інформаційних технологій в даному дослідженні.

Робота викладена науковою мовою, логічно й послідовно відбиває мету та поставлені в роботі завдання. Пояснювальна записка відповідає стандартам до її оформлення.

**Недоліки:**

1. В магістерській роботі нічого не говориться про можливі небезпеки діапазону 5G випромінювання, в тому разі при вирішуванні питань позиціювання.
2. Доцільно було б приділити більше уваги працям вітчизняних дослідників проблеми, а судячи з переліку посилань основну увагу приділено англомовним джерелам.

**Висновок:** кваліфікаційна робота заслуговує оцінку «відмінно», а здобувачка Юрчук Евеліна Ігорівна заслуговує присвоєння кваліфікації: «Магістр з телекомунікацій та радіотехніки».

**Рецензент:** Д.Т.Н., ДОЦ. А.Гриневич

*(науковий ступінь, вчене звання)* *(підпис)* *(Ім'я, ПРІЗВИЩЕ)*

**ЗМІСТ**

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ ..... 9
ВСТУП ..... 11
1 ПОСТАНОВКА ПРОБЛЕМИ ТА СЦЕНАРІЇ, ЯКІ ДОСЛІДЖУЮТЬСЯ ..... 13
1.1 Передумови ..... 13
1.2 Постановка проблеми ..... 17
1.3 Огляд науково-технічної літератури по темі роботи ..... 19
2 ТЕОРЕТИЧНІ ПЕРЕДУМОВИ ..... 22
2.1 Алгоритми машинного навчання ..... 22
2.2 Виявлення умов NLOS ..... 32
2.3 Фільтр Кальмана ..... 37
3 ЗАСТОСУВАННЯ МЕТОДІВ МАШИННОГО НАВЧАННЯ ДЛЯ ПОЗИЦІОНУВАННЯ ПО 5G-БАЗОВИХ СТАНЦІЯХ ..... 39
3.1 Опис даних ..... 39
3.2. Нейронні мережі ..... 49
3.3 Метод машинного навчання “random forest” ..... 52
3.4 Спосіб та реалізація методів виявлення умов NLOS ..... 54
3.5 Настройка фільтра Калмана ..... 56
4 ОЦІНКА ЕФЕКТИВНОСТІ ЗАПРОПОНОВАНИХ МЕТОДІВ ..... 59
4.1 Результати, отримані з використанням нейронних мереж ..... 59
4.2 Метод машинного навчання “random forest” ..... 66
4.3 Позиціонування в умовах LOS ..... 69
4.4 Виявлення в умовах NLOS ..... 73
4.5 Фільтр Кальмана ..... 73
4.6 Обговорення результатів ..... 77
ВИСНОВКИ ..... 80
ПЕРЕЛІК ПОСИЛАНЬ ..... 82
ДЕМОНСТРАЦІЙНІ МАТЕРІАЛИ ..... 85

# ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

| | | |
|---|---|---|
| 3GPP | 3rd Generation Partnership Project | - Проект партнерства третього покоління |
| 5G | 5th Generation cellular networks | - Стільникові мережі 5-го покоління |
| BLE | Bluetooth Low Energy | - Bluetooth з низьким рівнем енергії |
| BRSRP | Beam Reference Signal Received Power | - Отримана потужність опорного сигналу променя |
| CART | Classification And Regression Trees | - Класифікація та регресія Trees |
| CDF | Cumulative Distribution Function | - Функція накопичувального розподілу |
| CNN | Convolutional Neural Network | - Конволюційна нейронна мережа |
| CRLB | Cramer-Rao Lower Bound | - Нижня межа Крамера-Рао |
| DoD | Direction of Departure | - Напрямок руху |
| ECDF | Empirical Cumulative Distribution Function | - Емпірична функція кумулятивного розподілу |
| GPS | Global Positioning System | - Глобальна система позиціонування |
| GLRT | Generalized Likelihood Ratio Test | - Узагальнений тест на коефіцієнт ймовірності |
| LOS | Line of Sight | - Пряма видимість |
| MLE | Maximum Likelihood Estimator | - Оцінювач максимальної ймовірності |
| NLOS | Non-Line of Sight | - Не пряма видимість |
| NP | Neyman-Pearson | - Неймана-Пірсона |
| PDF | Probability Density Function | - Функція щільності ймовірності |
| RSS | Received Signal Strength | - Сила отриманого сигналу |
| TA | Timing Advance | - Випередження |
| TP | Transmission Point | - Точка передачі |

| | | |
|---|---|---|
| PD | probability of detection | Імовірність виявлення |
| PFA | probability of false alarm | ймовірність помилкової тривоги |
| UE | User Equipment | - Обладнання користувача |
| WGN | White Gaussian Noise | - Білий Гаусів шум |

# ВСТУП

Позиціонування визнано важливою особливістю стільникових мереж п'ятого покоління (5G) через велику кількість випадків комерційного використання, які отримали б користь від доступу до інформації. Позиціонування на базі радіо завжди було складним завданням у міських районах, де будівлі блокують та відбивають радіосигнал, викликаючи багатолунаюче поширення та умови сигналу нелінійного зору (NLOS). Один із підходів до управління NLOS полягає у використанні методів керування даними, таких як, наприклад, алгоритми машинного навчання на базі даних. Останній, як відомо, базується на променях, де навчальний набір даних із великої кількості вимірювань використовується для підготовки моделі, яка перетворює певне вимірювання в оціночні положення.

Робота базується на даних про позиціювання та багаторазові радіовимірювання (тестування) конкретного 5G пристрою. Точка передачі (ТП) у випробувальній установці має антену, яка надсилає промені як у горизонтальному, так і у вертикальному напрямах. Вимірювання – це результат порівняння опорного сигналу променя і прийнятої потужності (BRSRP) від променів і напрямку відходу (DoD) від набору променів з найбільшою силою прийнятого сигналу (RSS). Для моделювання зв'язку між вимірюваннями та положеннями розглянуто дві нелінійні моделі – це нейромережа та random forest моделі. Ці нелінійні моделі будуть називатися алгоритмами машинного навчання.

Алгоритми машинного навчання здатні розташувати обладнання користувача (UE) у регіонах NLOS з похибкою горизонтального позиціонування менше 10 метрів у 80 відсотках тестових випадків. Результати також показують, що важливо поєднувати інформацію променями з різних вертикальних шарів антен, щоб мати змогу виконувати позиціонування з високою точністю в умовах NLOS. Крім того, тести показують, що дані повинні бути розділені на оглядові

(LOS) та NLOS дані перед побудовою алгоритмів машинного навчання, щоб досягти достатньо хорошого позиціонування як в умовах LOS, так і в NLOS. Тому було розроблено узагальнений тест на коефіцієнт ймовірності (GLRT) для класифікації даних, що походять як в умовах LOS так і за умов NLOS. Імовірність виявлення (PD) алгоритмів становить близько 90%, коли ймовірність помилкової тривоги (PFA) становить лише 5%.

Для підвищення точності позиції за допомогою алгоритмів машинного навчання був взятий фільтр Калмана, вихід із алгоритмів машинного навчання як вхідний. Результати показують, що це може значно підвищити точність позиції в сценаріях NLOS.

**Метою магістерської роботи** є дослідження ефективності застосування методів машинного навчання в 5G-мережах для позиціонування по базовим станціям.

**Об'єкт дослідження** - процес передачі даних 5G-мережах.

**Предмет дослідження** – метод машинного навчання в 5G-мережах для позиціонування по базовим станціям.

**Методи дослідження** - системний метод, методі цифрової обробки інформації, метод моделювання, розрахунки на ЕОМ.

# 1 ПОСТАНОВКА ПРОБЛЕМИ ТА СЦЕНАРІЇ, ЯКІ ДОСЛІДЖУЮТЬСЯ

## 1.1 Передумови

Позиціонування визнано важливим елементом застосуванням для довгострокової еволюції (LTE) та стільникових мереж п'ятого покоління (5G). Це пов'язано з його потенціалом для масових випадків комерційного використання, наприклад, автоматизації промисловості, віддаленої експлуатації та викликів надзвичайних ситуацій. Це також є важливим тому, що в США ще з 2001 р. обов'язковою для локальних операторів мережі стала вимога щодо умов надання локальних послуг [1]. Радіозв'язок завжди був складним завданням у міських умовах, коли висотні будинки блокують та відбивають сигнал між користувальницьким обладнанням (UE) та точкою передачі (TP). Ці середовища прийнято називати міськими каньйонами. Схематичне зображення розташування двох UE, в міському каньйоні показано на рис. 1.1.

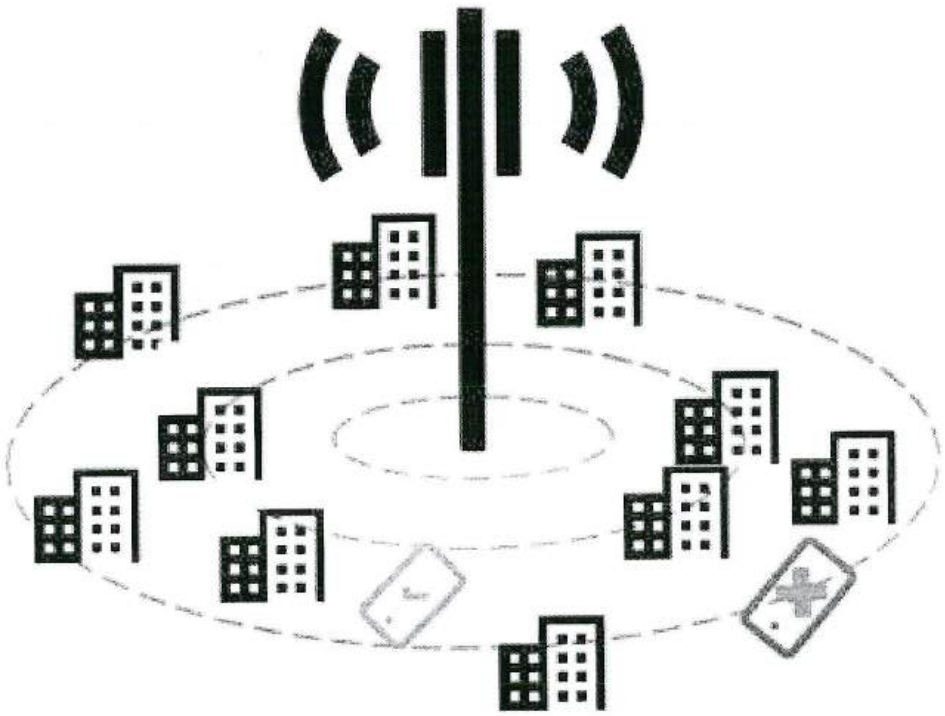


Рис. 1.1. Схематична картина позиціонування у міському районі

На рис. 1.1, як кажуть, перший UE (зелена галочка) знаходиться в полі зору (LOS), тобто коли між UE і TP є прямий і чіткий шлях. У той час, коли простір між UE і TP перекрито декількома висотними будівлями, як це стосується другого приймаючого пристрою (червоний хрест) на рис. 1.1, тобто UE знаходиться в нелінійній зоні (NLOS). У LOSi існує багато методів позиціонування, які можна використовувати для позиціонування UE з високою точністю. Можна навіть отримати високу точність позиціонування в ситуації навіть руху UE зі швидкістю до 100 км / год [2]. При цьому використовуються проміжок часу (TA), тобто затримка поширення сигналу між UE і TP, а також напрям відходу (DoD) між пучком з найсильнішим опорним сигналом променя, що приймає потужність (BRSRP) на TP і UE для оцінки позиції. Прикладом алгоритму для оцінки DoD є класифікація множинних сигналів [3]. Геометричні методи позиціонування LOS тут будуть називатися традиційними методами позиціонування.

На рис. 1.2-1.4 показані приклади традиційного методу позиціонування, коли розрахунковий DoD та відстань використовуються для позиціонування UE у міському районі. Це робиться для трьох різних шляхів, які UE пройшли в пробному стані. У цьому прикладі відстань між UE і TP вважається відомою. На рис. 1.3 показано розрахунковий DoD та справжню відстань, а на рис. 1.4 вони об'єднані для позиціонування UE.

Коли UE на рис. 1.4 потрапляє за будівлю (проходить перехрестя між Blafjallsgatan та Gronlandsgatan) та входить у регіон з умовами NLOS, оцінене положення UE (відображається як червоні точки) не є надійним. Це також можна побачити на рис. 1.2, де оцінка кута сильно коливається через 8 секунд.

Це коливання DoD обумовлює той факт, що традиційний метод позиціонування спрацьовує погано. Тому виникає необхідність розробки нових технологій позиціонування, які можуть працювати саме в умовах NLOS.

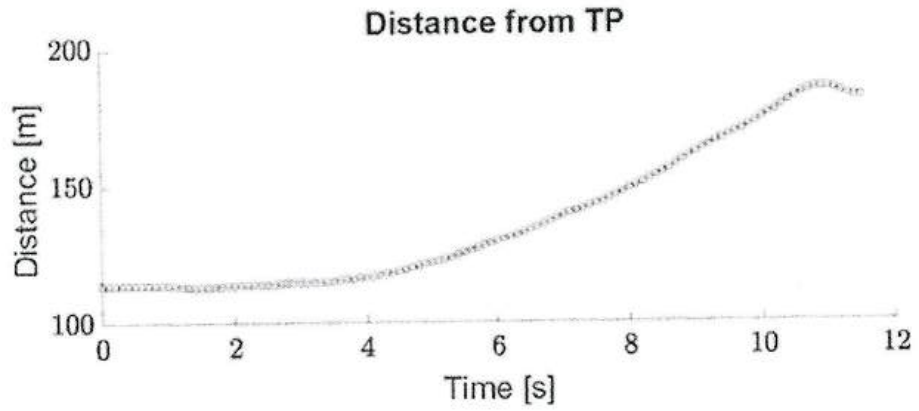



Рис. 1.2. Позиціонування за допомогою методу вимірювання напрямку відправлення та відстані. Цільова відстань від TP, обчислена за допомогою координат GPS

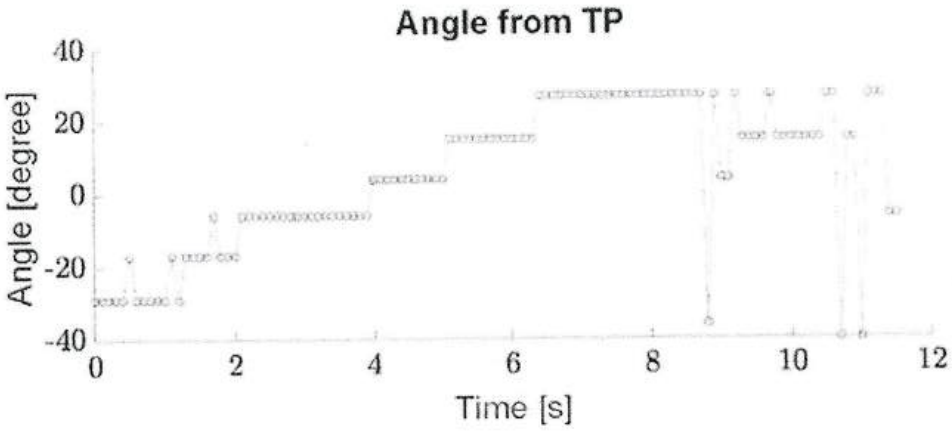



Рис. 1.3. Позиціонування за допомогою методу вимірювання напрямку відправлення та відстані. Цільовий кут від TP, обчислений за допомогою DoD за умов NLOS

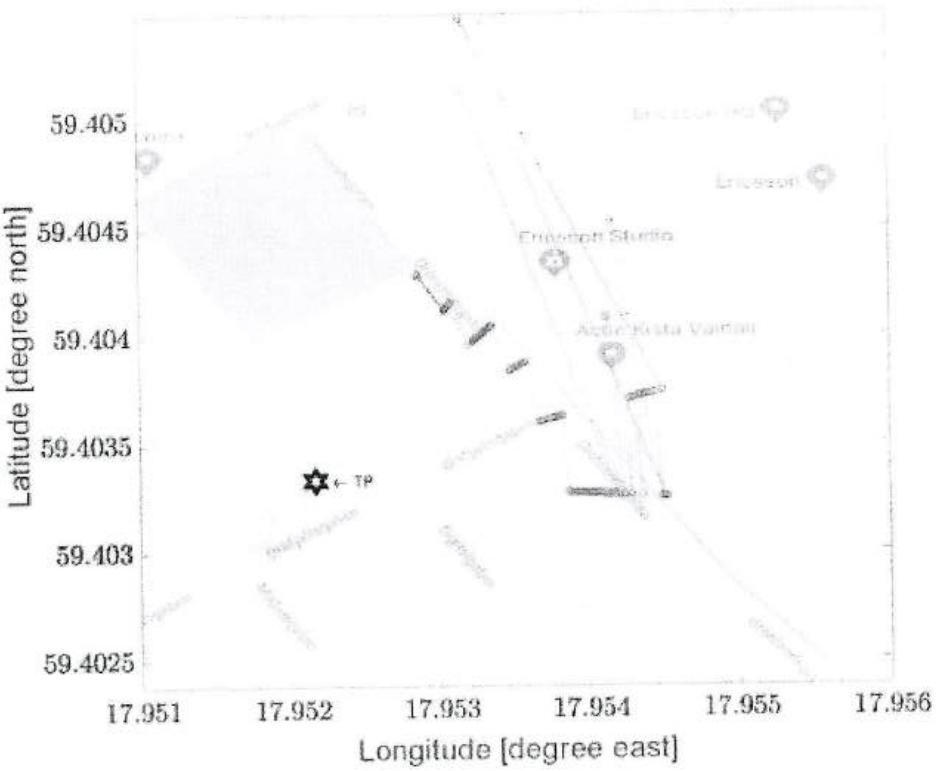



Рис. 1.4. Позиціонування за допомогою методу вимірювання напрямку відправлення та відстані. Орієнтовне положення UE у міському каньйоні, використовуючи розрахунковий кут та відому відстань між UE та TP

Один із підходів для визначення позиціонування в NLOS полягає у використанні методів керування даними, які в данній атестаційній роботі називаються алгоритмами машинного навчання [4]. Виявлення ситуації NLOS також є важливою особливістю, завдяки якій можна визначити, коли традиційні методи позиціонування можна, а коли і не можна використовувати.

Мета цієї магістерської роботи – дослідити варіанти використання алгоритмів машинного навчання для виконання позиціонування у міських каньйонах з умовами NLOS. Дослідження було зосередженим на позиціонуванні за допомогою ряду "даних знімків". Це означає, що вимірювання проводилися в одному часовому екземплярі в кожному місці UE. Вибір вхідних даних до алгоритмів машинного навчання став центральним, а також використовувалися різноманітні конструкції антен та їх діаграми випрмінювання при замірах цих

даних. Ця робота також враховувала виявлення різноманітних умов NLOS, а також досліджувала питання використання фільтрації виводу алгоритмів машинного навчання для підвищення точності позиціонування.

### 1.2 Постановка проблеми

Атестаційна робота має на меті відповісти на наступні питання:

1. Чи можливе вирішення питань позиціонування у міських каньйонах з умовами NLOS за допомогою використання алгоритмів машинного навчання? Якщо це так, яка буде очікувана точність позиціонування?

2. Чи можна розрізнити умови LOS та NLOS, використовуючи функції, пов'язані з DoD та BRSRP?

3. Чи можливо підвищити точність позиціонування, фільтруючи вихід алгоритмів машинного навчання?

Для дослідження цих питань використовувалися дані 5G-тестових випробувнь, тобто дані, сформовані з раннього прототипу стільникової мережі 5G з прототипами як ТР, так і UE. Дані використовувалися як для навчання та оцінки алгоритмів машинного навчання для позиціонування, так і для вивчення характеристик BRSRP та DoD в умовах NLOS. Дані походять від тесту, проведеного Ericsson AB в Кіста, Стокгольм. Сигнал від антени має несучу частоту 15 ГГц. Антена має 48 променів, розташованих у п'яти різних вертикальних шарах. На підставі даних випробуваної панелі встановлено, що UE завжди має зв'язок з ТП. На випробувальній панелі є лише одне UE, підключене до ТР, тому розміщення декількох UE в мережі та перешкоди між користувачами не враховувалися. Існує також обмеження кількості доступних даних для вивчення та оцінювання різних алгоритмів машинного навчання. Ця теза не буде враховувати моделювання навколишнього середовища та того, що викликає NLOS, та відмінності в характеристиках сигналу, наприклад, чи це дерево або будівля, що блокує сигнал між UE та ТР. Ефект різних погодних умов та

навколишнього середовища також не буде розглянуто.

Отже, умови LOS та NLOS можна розрізнити за допомогою функцій, які пов'язані з DoD і BRSRP у контексті безпроводового зв'язку і позиціонування. DoD вказує на напрям, в якому сигнал відправляється від пристрою до базової станції (або іншого приймача). Умова LOS характеризується тим, що DoD практично не має обмежень і може бути точно спрямованим на базову станцію без значних відхилень або розсіювання. У NLOS ситуація може бути іншою, DoD може мати відхилення або не бути спрямованим на базову станцію через взаємодію сигналу з перешкодами.

BRSRP вказує на силу сигналу, яку базова станція отримує від певного пристрою або приймача. Умова LOS характеризується тим, що BRSRP має стабільну і високу силу сигналу, оскільки сигнал вільно доходить до базової станції без значних втрат. Умова NLOS може вказувати на значні втрати сигналу та зменшену силу сигналу, оскільки сигнал повинен пройти через перешкоди.

Тобто аналіз DoD та BRSRP може допомогти визначити, чи є умови LOS або NLOS. Якщо DoD має малий відхил від напряму до базової станції, і BRSRP висока та стабільна, то це може свідчити про умови LOS. У випадку великого відхилення DoD та низької BRSRP, може виникати підозра на умови NLOS.

Однак слід зауважити, що для точного розрізнення між умовами LOS і NLOS може бути потрібний більший обсяг інформації та складні алгоритми обробки сигналу.

Позиціонування в умовах NLOS в містах може бути складним завданням через багато перешкод, таких як будівлі та інші структури, які можуть блокувати сигнали від базових станцій. Але за допомогою алгоритмів машинного навчання і сучасних технологій, таких як нейронні мережі, можливо досягти високої точності в позиціонуванні навіть в умовах NLOS. Ось деякі аспекти, які слід враховувати:

- Збір даних: перший крок - це збір даних про сигнали від базових станцій

та їхніх відбитків від перешкод. Для цього можуть використовуватися сучасні сенсори, такі як акселерометри, гіроскопи, сенсори відстані тощо.

• Навчання моделей: машинне навчання використовується для розробки моделей, які можуть передбачити місце розташування на основі зібраних даних. Це може бути задачею регресії, класифікації або сегментації, залежно від конкретної задачі.

• Алгоритми для виправлення впливу перешкод: оскільки сигнали можуть відбиватися від стін і перешкод, потрібні алгоритми для корекції цього впливу. Це може включати в себе використання геометричних обчислень та статистичних методів для оцінки траєкторії сигналу.

• Використання додаткових датчиків: крім сигналів мобільного зв'язку, можна використовувати інші датчики, такі як GPS, Wi-Fi, Bluetooth, акселерометри і живлення для поліпшення точності позиціонування.

• Апаратне забезпечення: для високоточного позиціонування може знадобитися спеціалізоване апаратне забезпечення, таке як антени з вузькою напрямною характеристикою.

• Передбачення об'єктів: іншою важливою задачею є передбачення руху об'єктів, які позиціонуються. Моделі машинного навчання можуть враховувати інформацію про рух, щоб поліпшити точність.

Використання алгоритмів машинного навчання для позиціонування в умовах NLOS є актуальним дослідницьким напрямком, і ці технології продовжують розвиватися.

### 1.3 Огляд науково-технічної літератури по темі роботи

Позиціонування в стільникових мережах є добре налагодженою темою дослідження. Багато спроб було зроблено за допомогою традиційних методів, які оцінюють рівень дози опромінювання та технічної допомоги. У роботі [2] позиціонування автомобіля, який рухається з високою швидкістю, здійснюється за

допомогою оцінки DoD від TP. Також [5] досліджує оцінку кута приходу та положення в стільникових мережах в умовах LOS, використовуючи прийнятий сигнал сигналу (RSS) від спрямованих антен. Тут їм вдається оцінити кут прибуття з точністю в три градуси і місце розташування з точністю до метра. Ці методи добре працюють в умовах LOS, де можна оцінити TA.

За умови NLOS ці методи більше не застосовуються, тому експерименти з використанням методів керування даними, були досліджені. У [4] метод машинного навчання для позиціонування в NLOS використовується гауссові процеси. Це робиться в приміщенні, використовуючи маяки з низькою енергією (BLE) RSS від Bluetooth.

Статистичний модельний підхід для позиціонування UE в міських каньйонах з умовами НЛЗ застосований у [1], де точність позиціонування становить менше 300 метрів. Їх випробування проводяться для міського середовища, тоді як у цій магістерської роботі передбачається розмір тесту на околиці. У середовищі, що має масштаби міста, метою є позиціонування UE у правій комірці, тоді як для позиціонування в масштабі сусідства мета - розташування UE в точній точці. Використання нейронних мереж для позиціонування у міських каньйонах описано в [6]. Вони використовують звивисті нейронні мережі (CNN) у поєднанні з позиціонуванням на основі відбитків пальців. Для порівняння, ця теза буде досліджувати використання байєсівських регульованих штучних нейронних мереж.

Теоретична основа для виявлення NLOS з використанням узагальненого тесту на коефіцієнт імовірності (GLRT) описана в [7], де вони припускають, що дисперсія в умовах NLOS більша, ніж в умовах LOS, використовуючи TA як сигнал для виявлення. Ця теза надихає теорію, представлену в [7], використовуючи при цьому RSS та DoD як сигнали для виявлення NLOS.

Фільтри також часто використовуються для позиціонування та відстеження UE в стільникових мережах. Наприклад, у [8] та [9] описано позиціонування з

використанням фільтрації частинок. У [10] автори обговорюють можливості та обмеження мобільного позиціонування у стільникових мережах. Одним з обмежень є труднощі пропозиції в умовах NLOS. Як фільтрування на основі моделі, так і сенсорний підхід для синтеза досліджено в роботі [10] з хорошими результатами. У цій тезі фільтрація результатів отриманних з допомогою алгоритмів машинного навчання буде використана для вирівнювання оцінки позиції.

# 2 ТЕОРЕТИЧНІ ПЕРЕДУМОВИ

## 2.1 Алгоритми машинного навчання

У цьому розділі викладене необхідне теоретичне підґрунтя для розуміння тем, висвітлених раніше.

У цьому розділі описані два різні алгоритми машинного навчання, а саме: Байєсовська регуляризована штучна нейронна мережа та random forest.

***Нейронні мережі***

Нейронні мережі (або штучні нейронні мережі) - це клас алгоритмів машинного навчання, які моделюють функціонування людського мозку для розв'язання завдань у широкому спектрі доменів, включаючи комп'ютерне бачення, обробку природних мов, розпізнавання мови, класифікацію, прогнозування та інші.

Компоненти нейронних мереж: ***нейрони*** (штучні нейрони) - це базові обчислювальні одиниці нейронної мережі. Кожен нейрон приймає вхідні сигнали, обчислює їхню вагову суму, застосовує функцію активації і генерує вихідний сигнал. ***Шари:*** нейрони організовані в шари. Вхідний шар приймає вхідні дані, вихідний шар генерує вихідний результат, а проміжні шари називаються прихованими і виконують обчислення між входом і виходом. ***Ваги:*** ваги визначають силу зв'язків між нейронами. Вони встановлюються під час процесу навчання і впливають на важливість входів для обчислення виходу нейрона. ***Функція активації:*** функція активації визначає вихід сигналу нейрона на основі його входів і ваг. Поширеною функцією активації є сигмоїдальна функція або функція ReLU (Rectified Linear Unit). ***Функція втрат:*** ця функція визначає різницю між прогнозованим виходом мережі і дійсними вихідними даними. Мета - зменшити цю втрату під час навчання. ***Зворотний прохід*** (Backpropagation): це алгоритм навчання, який використовується для оновлення ваг і покращення

точності прогнозів. Він включає в себе розповсюдження помилок назад через мережу і коригування ваг на основі цих помилок. ***Глибокі нейронні мережі:*** це нейронні мережі з багатьма прихованими шарами, які дозволяють моделювати складні взаємозв'язки у великих наборах даних. Глибокі нейронні мережі зазвичай використовуються для завдань глибокого навчання.

Нейронні мережі спочатку позначають спроби знайти математичне зображення обробки інформації в біологічних системах [11]. Ідея полягає в тому, щоб дозволити нейронній мережі вивчати шаблони під час фази навчання, для того щоб мережа змогла класифікувати нові вхідні дані після фази навчання. У розпізнаванні шаблонів модель нейронної мережі, що рухається вперед (також відома як модель багатошарового сприйняття) є потужним інструментом. Модель заснована на лінійній комбінації фіксованої кількості нелінійних параметричних функцій, званих базовими функціями. У свою чергу ці основні функції залежать від вхідних параметрів та регульованих ваг. Для двошарової нейронної мережі регульовані ваги позначаються як:

$$\mathbf{w} - w_{10}^{(1)}, \ldots, w_{MD}^{(1)}; w_{10}^{(2)}; \ldots, w_{KM}^{(2)}. \tag{2.1}$$

Регульовані ваги у вигляді $w_{ij}^{(n)}, j \neq 0$ називаються вагами, а $w_{i0}^{(n)}$ – як ухили. Ваги регулюються шляхом оптимізації заздалегідь визначеної вартості функції J(w). У загальному вигляді двошарові нейронні мережі задаються методом:

$$y_k(\mathbf{x}, \mathbf{w}) = \nu\Big(\sum_{j=1}^{M} w_{jk}^{(2)} h\Big(\sum_{i=1}^{D} w_{ji}^{(1)} x_i + w_{j0}^{(1)} x_0\Big) + w_{k0}^{(2)} z_0\Big), \tag{2.2}$$

де x = {xi}, i = 1, ..., D - вхідна змінна, а $\mathbf{y} = \{y_k\}_{k=1}^{K}$ - вихідна змінна, керована вектором w регульованих ваг, наведеним у (2.1). Параметри $x_o$ і $z_o$ називаються прихованими змінними [12], приховані змінні часто встановлюються

в одну. Вершина (1) означає, що регульовані параметри знаходяться в першому шарі, і тоді (2) означає, що вони знаходяться у другому шарі. Перетворення від першого до другого шару та від другого шару до вихідного здійснюється за допомогою диференційованих, нелінійних функцій активації h (•) та v (•) відповідно. Поширеною функцією активації є сигмоїдна функція [12, 13]. Тобто,

$$\nu(a) = \frac{1}{1 + e^{-a}}. \tag{2.3}$$

Змінні у другому шарі *zj, j = 1, ..., M* називаються прихованими одиницями або, у сукупності, як прихований шар. На рис. 2.1 наведена схема мережі для двошарової нейронної мережі з M нейронами на прихований шар, D вхідними змінними та K вихідними змінними [12].

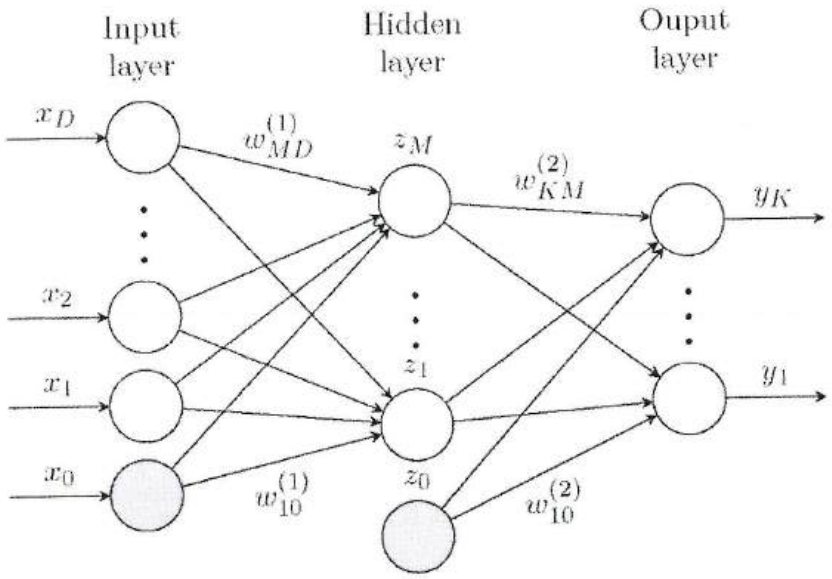



Рис. 2.1. Діаграма двошарової нейронної мережі. Вхідні, вихідні та приховані шари представлені вершинами, а ваги – ребрами

Під час фази навчання ваги та зміщення нейронної мережі оцінюються (вивчаються) шляхом мінімізації розглянутої вартісної функції J. Мінімізація функції витрат J є ітераційним процесом, що обумовлює необхідність надання

деяких критеріїв зупинки. Це часто робиться за допомогою валідаційного набору зразків, де процес навчання припиняється, коли помилка прогнозування валідаційного набору перестає зменшуватися. Перевищення – це явище, коли подальші ітерації призводять до збільшення помилки квадрата на наборі перевірки, але зменшення похибки на навчальному наборі. Одна з спроб подолати надмірну обробку полягає у використанні байєсівських регульованих штучних нейронних мереж [12, 13].

***Байєсові регуляризовані штучні нейронні мережі***

Баєсові регуляризовані штучні нейронні мережі (Bayesian Regularized Neural Networks) - це підхід до навчання нейронних мереж, який поєднує в собі техніки баєсової інференції та регуляризації для покращення якості моделей та забезпечення їхньої надійності. Основні ідеї та особливості цього підходу включають таке:

*Баєсова інференція*: В основі цього підходу лежить використання ймовірнісних розподілів для параметрів нейронної мережі замість точних значень параметрів. Це означає, що кожен параметр має свій власний розподіл ймовірностей, що визначає ймовірність кожного можливого значення параметра.

*Регуляризація*: Баєсові регуляризовані нейронні мережі використовують різні види регуляризації, такі як L1 або L2 регуляризація (також відома як weight decay), щоб контролювати складність моделі та запобігти перенавчанню. Регуляризація допомагає зменшити ризик перенавчання та покращити загальну узагальнюючу здатність моделі.

*Баєсова апроксимація:* Для навчання баєсових регуляризованих нейронних мереж використовують баєсову апроксимацію, таку як Variational Bayes або Монте-Карло Марковська цепля (MCMC) методи. Ці методи дозволяють наближено оцінювати розподіли параметрів, що робить їх більш обчислювально ефективними.

*Визначення невизначеності:* Один з головних вигод баєсових регуляризованих нейронних мереж - це здатність визначати невизначеність моделі. Оскільки кожен параметр має ймовірнісний розподіл, ми можемо отримувати додаткову інформацію про невизначеність прогнозів та передбачення моделі.

*Зменшення перенавчання*: Використання баєсового підходу та регуляризації допомагає знизити ризик перенавчання моделі на невеликих обсягах даних, тобто зробити модель більш стійкою до шуму та викидів.

У регульованих штучних нейронних мережах Байєса функція витрат J визначається як:

$$J(\mathbf{w}) = \beta \sum_{n=1}^{N} ||y(\mathbf{x}_n, \mathbf{w}) - \mathbf{t}_n||_2^2 - \alpha \sum_{n=1}^{N} w_n^2, \tag{2.4}$$

з гіперпараметрами α та β. В (2.4), $||\cdot||_2^2$ позначають евклідову норму у квадраті, а $\{x_n\}$ - це набір вхідних векторів, а $\{t_n\}$ позначає цільові значення, тобто справжні значення, а N - кількість точок даних тренувань у навчальному наборі L. Припускаючи, що умовний розподіл p (t|x) для однієї цілі є гауссовим із середньозалежним середнім значенням, заданим виведенням нейронної мережі у (x, w) з точністю 3, і що попередній розподіл по вагах, w

$$p(\mathbf{w}|\alpha) = \mathcal{N}(\mathbf{w}|\mathbf{0}, \alpha^{-1}\mathbf{I}). \tag{2.5}$$

Крім того, при припущенні, що пріоритет у (2.5) є гауссовим із нульовим середнім значенням та точністю [12, 13]. Нехай D = {ti,. . . $t_N$}, то

$$p(\mathbf{w}|\mathcal{D}, \alpha, \beta) \propto p(\mathbf{w}|\alpha) \prod_{n=1}^{N} \mathcal{N}(t_n|y(\mathbf{x}_n, \mathbf{w}), \beta^{-1}) \tag{2.6}$$

є отриманий задній розподіл не гауссовим, що обумовлено нелінійністю у ($X_n$, W). У (2.6) $\mathcal{N}(\mu, \sigma^2)$ позначає нормальний розподіл із середнім p та дисперсією $\delta^2$, а точність - 1 / $\delta^2$. Можна знайти наближення Гаусса (2.6), використовуючи наближення Лапласа з локальними максимумами, знайденими числовою оптимізацією. Наприклад, з використовуванням алгоритму Левенберга-Маркарда для вирішення нелінійних задач з найменшими квадратами та зворотного розповсюдження для ефективного обчислення похідних. Оптимізація логарифма заднього зводиться до задачі з мінімальними квадратами.

Якщо припустити, що гіперпараметри a та jd відомі, можна знайти максимум a posteriori (MAP) оцінювача, позначеного WMAP. З обчислень, проведених у [12], показано, що p (w) має лінійно-гауссову модель з гауссовим розподілом, а p (t | w) – розподіленою Гауссом. Використовуючи результат граничних та умовних розподілів Гаусса в [12], ймовірність досягнення цілі з огляду на вхід, набір цілей та гіперпараметри:

$$p(t|\mathbf{x}, D, \alpha, \beta) = \mathcal{N}(t|y(\mathbf{x}, \mathbf{w}_{MAP}), \sigma^2(\mathbf{x})), \tag{2.7}$$

де

$$\mathbf{g} = \nabla_{\mathbf{w}} y(\mathbf{x}, \mathbf{w})|_{\mathbf{w}\ \ \mathbf{w}_{MAP}}$$
$$\sigma^2(\mathbf{x}) \quad \beta^{-1} \mid \mathbf{g}^T(\alpha\mathbf{I} \mid \beta\mathbf{H})^{-1}\mathbf{g} \tag{2.8}$$

Значення H позначає матрицю Гессі, що компрометує другі похідні помилки суми квадрата, що містить w, а I - матриця тотожності відповідного розміру [12].

Баєсові регуляризовані нейронні мережі є потужним інструментом для різних завдань машинного навчання та великої кількості даних. Вони дозволяють отримувати надійні моделі зі збалансованою точністю та невизначеністю прогнозі

***Оптимізація гіперпараметрів***

Поки припускалося, що гіперпараметри a і ft відомі і фіксовані. Це не завжди так, і їх можна обчислити шляхом максимізації:

$$\ln p(\mathcal{D}|\alpha,\beta) \simeq -J(\mathbf{w}_{MAP}) + \frac{1}{2}\ln|(\alpha\mathbf{I}+\beta\mathbf{H})| - \frac{W}{2}\ln\alpha + \frac{N}{2}\ln\beta - \frac{N}{2}\ln 2\pi. \quad (2.9)$$

стосовно a і ft, де W – кількість параметрів у w. Максимізація (2.9) отримується, використовуючи аналогічні припущення, що і для $W_{MAP}$, детальніше див. [12].

Це призводить до значень гіперпараметрів, наведених як

$$\begin{aligned} \gamma &- \sum_{i-1}^{W}\frac{\lambda_i}{\alpha-\lambda_i} \\ \alpha &= \frac{\gamma}{\mathbf{w}_{MAP}^T\mathbf{w}_{MAP}} \\ \frac{1}{\beta} &= \frac{1}{N-\gamma}\sum_{n-1}^{N}\{y(\mathbf{x}_n,\mathbf{w}_{MAP})-t_n\}^2, \end{aligned} \quad (2.10)$$

при $\lambda_i$ є власним значенням βH. Щоб знайти $W_{MAP}$, α та β повинні бути відомі і навпаки. Це призводить до того, що оптимізація проводиться шляхом рекурсивного оновлення заднього розподілу та повторної оцінки гіперпараметрів [12, 13].

***Деревоподібні моделі***

Деревоподібні моделі зазвичай використовуються для класифікації та регресії. Ідея полягає у виконанні двійкових рішень функцій у наборі вводу, щоб розділити їх відповідно до різних класифікацій. Ці моделі часто називають

деревами класифікації та регресії (CART). Приклад однієї CART показаний на рис. 2.2. Приємною властивістю таких моделей на основі дерев, як CART, є те, що їх легко проілюструвати та інтерпретувати [12].

Направлене дерево на рис. 2.2 приймає два або три двійкових рішення для розділення вхідного простору {BRSRP, DoD} для оцінки виходу {posx, posy}. Він почнеться від кореневого вузла (червоний) і просувається через дерево, поки не знайде лист (білі вузли). Глибина дерева визначається як кількість вузлів, пройдених за найдовшим шляхом у дереві; для дерева на рис. 2.2 глибина чотири.

Дуже популярно генерувати безліч класифікаторів та агрегувати результат над ними. Цей метод називається ансамблевим навчанням. Два відомих приклади – це прискорення та мішок. Підвищуючи ці класифікатори, асоціюється з вагою, яка буде оновлюватися за допомогою ітераційного процесу. Щоб зрозуміти мішковину або прискорення завантаження, спочатку визначаються зразком завантажувальної програми.

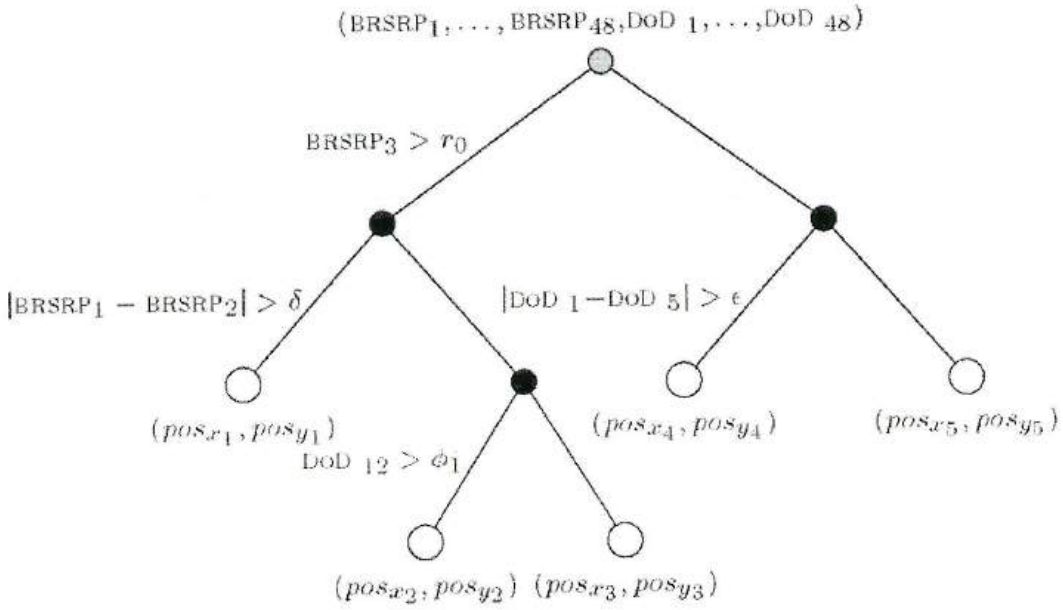



Рис. 2.2. Діаграма мережі над двійковим деревом рішень глибиною чотири.

Зразок завантажувальної програми - це створити набір XB, намалювавши N випадкових вибірок із набору даних X = {xi ... XN}. Це може призвести до того,

що деякі точки в X копіюються в XB, а деякі відсутні. Рисунок 2.3 є прикладом взяття трьох зразків завантажувальної програми з обмеженого навчального набору.

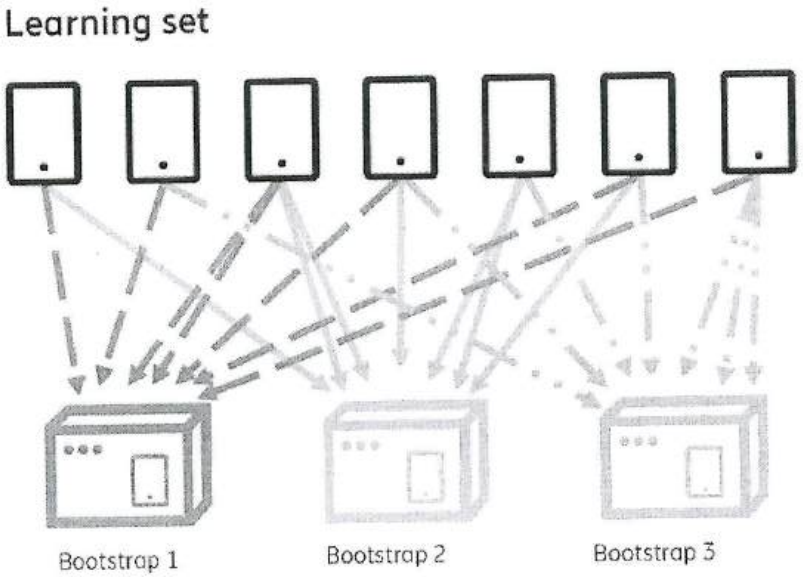



Рис. 2.3. Приклад взяття трьох зразків завантаження з навчального набору, що складається з семи точок даних

Ці точки даних представлені мобільними телефонами та зразками завантажувальної машини з коробками.

Слід визначити мішковину як взяття повторних зразків завантажувальної програми з навчального набору L, побудувати класифікатор у кожному з цих наборів і взяти підсумковий класифікатор як середнє значення для всіх цих менших класифікаторів. Використовуючи іграшковий приклад на рис. 2.3, на основі кожного ящика навчається одне класифікаційне дерево, і дерево мішків буде середнім типом цих класифікаторів [12, 15, 16].

Random Forest додає додаткового шару випадковості для переробки мішків. Окрім випадковості у відборі завантажувальних зразків, random forest також змінює структуру дерева між ітераціями. Для кожного kth дерева генерується вектор 0 k, який має такий самий розподіл, як і минулі випадкові вектори 0i, ...,

0k-i, але незалежно від них, цей вектор випромінювання визначає, як дерево розщепить, отже, вирішить будова дерева [17]. Тепер визначаєтьчя random forest як класифікатор, який складається з колекції модельованого на дереві класифікатора h (x, 0k), k = 1, ..., Ktree, де 0k – незалежні однаково розподілені випадкові вектори та кожне з дерев Ktree призначає одиничне LOS за найпопулярніший клас на вході x [16, 17]. Збіжність випадкових лісів доведено, використовуючи сильний закон великих чисел у [17]. Зважаючи на це, перевитрата не є проблемою для випадкових лісів, коли до лісу додається більше дерев.

Дві додаткових відомості також надаються random forest: мінливі значення та міра близькості. Змінна важливість вимірює, наскільки збільшується похибка прогнозування, коли дані, що знаходяться поза пакетом (OOB) для цієї змінної, змінюються, а інші залишаються незмінними. Класифікатор OoB визначається як класифікатори, навчальний набір Lk C £ не містить {t, x} G L. У матриці близькості елемент (i, j) повідомляє, як часто і-й і j-й елементи закінчуються на одном і тому ж листі. Це визначає міру близькості і може дати хорошу думку про структуру даних [16, 17].

Крім того, random forest є зручним для користувачів із лише двома регульованими параметрами (кількість дерев у лісі та кількість змінних у випадковому підмножині на кожному вузлі (глибина дерева)), це також дуже ефективний інструмент оцінки. Random forest міцний до шуму, швидший, ніж мішок та прискорення, дає корисну додаткову інформацію, таку як змінна важливість, і його легко паралелізувати [12, 16, 17].

## 2.2 Виявлення умов NLOS

Умови NLOS (Non-Line-of-Sight), що означають, що сигнал не може прямо дійти до базової станції або іншого пристрою через перешкоди, можуть бути виявлені та враховані за допомогою різних технологій та методів. Це важливо для забезпечення надійного функціонування безпроводних систем зв'язку та позиціонування, особливо в місцях з великою кількістю перешкод, таких як міста.

Методи виявлення умов NLOS:

- Аналіз сигналу: Використання аналізу сигналу для визначення, чи проходить сигнал через NLOS. Це може включати в себе аналіз амплітуди, фази та інших характеристик сигналу.

- Вимірювання затримки сигналу: Вимірювання затримки між відправленням сигналу та його прийманням може вказувати на присутність перешкод. Велика затримка може свідчити про NLOS.

- Використання антен з вузькими напрямними характеристиками: Антени з вузькими напрямними характеристиками можуть визначати напрям, з якого прийшов сигнал. Це може допомогти виявити перешкоди, що блокують прямий шлях сигналу.

- Застосування датчиків та інших сенсорів: Використання додаткових сенсорів, таких як інфрачервоні сенсори або лідари, може надати інформацію про наявність перешкод на шляху сигналу.

- Моделювання шляху сигналу: Використання програмного забезпечення для моделювання розповсюдження сигналу може допомогти виявити можливі шляхи сигналу та перешкоди на цих шляхах.

- Машинне навчання: Машинне навчання може використовуватися для аналізу даних сигналу та сенсорів з метою виявлення умов NLOS.

Для підвищення ефективності позиціонування у міських каньйонах центральне значення має виявлення умов NLOS. Надійне виявлення відкриває

можливість використовувати різні алгоритми позиціонування для LOS та NLOS. Менш обчислювально дорогі алгоритми LOS можна використовувати в умовах LOS, а більш обчислювальний більш важкий підхід, керований даними в умовах NLOS. Оскільки LOS та NLOS мають такі різні особливості, навчання алгоритмів позиціонування NLOS на даних LOS знизить продуктивність алгоритму. Є сенс ввести два різних підходи до статистичного виявлення: детектор Неймана-Пірсона (NP) та узагальнений тест коефіцієнта ймовірності (GLRT).

***Детектор Неймана-Пірсона***

Імовірність виявлення (PD) або потужність тесту та ймовірність помилкової тривоги (PFA) або рівень значущості або розміру тесту – це два важливих поняття при виявленні сигналу. З огляду на нульову гіпотезу Хо, альтернативну гіпотезу Привіт, а також набір спостережень x = {x [0], ..., x [N - 1]}, рівні $P_D$ та $P_{FA}$ задаються

$$P_D - \int_{R_1} p(\mathbf{x}; \mathcal{H}_1) d\mathbf{x}$$
$$P_{FA} = \int_{R_1} p(\mathbf{x}; \mathcal{H}_0) d\mathbf{x}, \tag{2.11}$$

де $R_1$ - значення, які відображаються у вирішенні $H_1$, або рівнозначно відхиляють $H_o$. Умовний PDF для вектора x у гіпотезі $H_1$ та умовний PDF для вектора x під гіпотезою Но позначаються відповідно p (x; $H_1$) та p (x; $H_o$). Теорема Неймана-Пірсона стверджує, що для максимізації $P_D$ для заданої $P_{FA}$, тоді приймають рішення про гіпотезу $H_1$, якщо

$$L(\mathbf{x}) = \frac{p(\mathbf{x}; \mathcal{H}_1)}{p(\mathbf{x}; \mathcal{H}_0)} > \gamma \tag{2.12}$$

де поріг γ знайдено з

$$P_{FA} = \int_{\{\mathbf{x}:L(\mathbf{x})>\gamma\}} p(\mathbf{x};\mathcal{H}_0)d\mathbf{x}. \tag{2.13}$$

Це називається детектором NP [18]. Таким чином, відбувається компроміс між бажаним $P_D$ (високий) та $P_{FA}$ (низький). На рис. 2.4 це показано за допомогою PDF-сигналу сигналу x, який є Гауссом як під $H_1$, так і $H_o$.

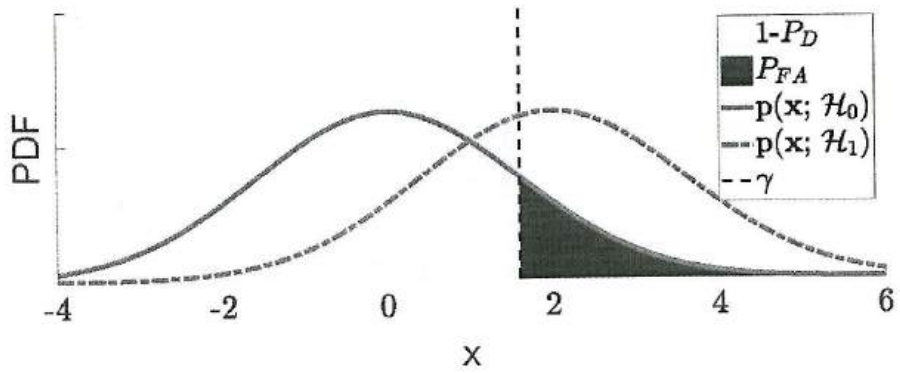



Рис. 2.4. Вірогідність виявлення та помилкової тривоги

Як видно з рисунка 2.4, збільшення $P_D$ призведе до збільшення $P_{FA}$ і навпаки. Це так, якби поріг γ перемістився праворуч, зелена зона була б меншою, але так би доповнила жовта зона під червоною кривою.

***Налаштування виявлення***

У цій роботі проблема виявлення NLOS була змодульвана як виявлення постійного струму (ПС) у білих гауссових шумах (WGN). Рівень постійного струму тут відповідає спаду BRSRP або більшої різниці DoD між елементами антени з найсильнішим BRSRP. Це означає, що x [n] - різниця ефіру BRSRP або DoD між найсильнішими пучками, див. Розділ 3.4. Нульова гіпотеза $H_o$, полягає в тому, що UE знаходиться в LOS, а альтернативна гіпотеза, Привіт, полягає в тому, що UE знаходиться в NLOS.

$$\mathcal{H}_0 : x[n] = w[n] \ , \ n = 1, \ldots, N-1$$
$$\mathcal{H}_1 : x[n] \quad A + w[n] \ , \ n \quad 1, \ldots, N-1 \tag{2.14}$$

В (2.14), де амплітуда при введенні умов NLOS A> 0 відома і w [n] WGN з відомою дисперсією $\delta^2$. За цих обставин x ~ N (0, $\delta^2$!) під $H_o$ і x ~ N (A, $\delta^2$I) під $H_1$, де N – нормальний розподіл, а I матриця тотожності [18].

Це призводить до налаштування, еквівалентного виявленню зміни середнього багатовимірного гауссового PDF. За допомогою детектора NP отримується правило виявлення: що повинні вирішити, якщо $H_1$

$$\frac{\frac{1}{(2\pi\sigma)^{\frac{N}{2}}} e^{\ \frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]\ \ A)^2}}{\frac{1}{(2\pi\sigma)^{\frac{N}{2}}} e^{-\frac{1}{2\sigma^2}\sum_{n-0}^{N}\ 1\ x^2[n]}} > \gamma, \tag{2.15}$$

які можна спростити до

$$\frac{1}{N}\sum_{n=0}^{N-1} x[n] > \frac{\sigma^2}{NA}\ln\gamma - \frac{A}{2} = \gamma'. \tag{2.16}$$

З цього також можна встановити залежність між $P_{FA}$, $P_D$ та порогом γ, використовуючи додаткову функцію кумулятивного розподілу Q (x), визначену як

$$Q(x) - \int_x^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}\, dt. \tag{2.17}$$

Для заданої $P_{FA}$ для обчислення $P_D$ та γ можна використовувати наступні

співвідношення

$$\gamma' = \sqrt{\frac{\sigma^2}{N}} Q^{-1}(P_{FA})$$
$$P_D = Q\left(Q^{-1}(P_{FA}) - \frac{NA^2}{\sigma^2}\right). \tag{2.18}$$

***Узагальнений тест на коефіцієнт ймовірності***

Розглянемо ту ж проблему, що була сформульована в (2.14), але параметр A невідомий, і в цьому випадку детектор NP не може бути використаний. Замість використання A заміним невідомий параметр його максимальною оцінкою ймовірності (MLE); такий підхід називається узагальненим тестом коефіцієнта ймовірності (GLRT). Нульова гіпотеза, Но, полягає в тому, що UE знаходиться в LOS, а альтернативна гіпотеза Ні полягає в тому, що UE знаходиться в NLOS.

$$L_G(\mathbf{x}) - \frac{p(\mathbf{x}; \hat{\theta}_1, \mathcal{H}_1)}{p(\mathbf{x}; \hat{\theta}_0, \mathcal{H}_0)} > \gamma. \tag{2.19}$$

**Налаштування виявлення**

Розглянемо рівень постійного струму в WGN з невідомою амплітудою $\theta_i$ = A і $\theta_0 = 0$, тоді гіпотеза стає

$$\mathcal{H}_0 : A = 0$$
$$\mathcal{H}_1 : A \neq 0. \tag{2.20}$$

У роботі [14] показано, що MLE з A дорівнює

$$\hat{A} - x - \frac{1}{N} \sum_{n=0}^{N-1} x[n], \tag{2.21}$$

і виявлення стає завданням: вирішити $H_1$, якщо

$$\bar{x}^2 > \frac{2\sigma^2 \ln \gamma}{N} = \frac{\gamma'}{N^2}. \tag{2.22}$$

Позначивши $P_r$ (•) як вірогідність події та використовуючи функцію Q (x), (2.17), і це

$$P_{FA} = Pr(|N\bar{x}| > \sqrt{\gamma'}; \mathcal{H}_0), \tag{2.23}$$

$\gamma$ і $P_D$ можна обчислити для даної $P_{FA}$ як

$$\sqrt{\gamma'} = \sqrt{\sigma^2 N} Q^{-1}\left(\frac{P_{FA}}{2}\right)$$
$$P_D = Q\left(Q^{-1}(P_{FA}/2) - \sqrt{N\hat{A}^2/\sigma^2}\right) - Q\left(Q^{-1}(P_{FA}/2) + \sqrt{N\hat{A}^2/\sigma^2}\right). \tag{2.24}$$

## 2.3 Фільтр Кальмана

Фільтр Калмана (Калман фільтр) та алгоритми машинного навчання (наприклад, нейронні мережі) можуть використовуватися разом для покращення точності позиціонування в безпроводних системах. Це особливо корисно в умовах, коли точне визначення місцезнаходження є критично важливим, наприклад, в автономних автомобілях, дронах або системах геолокації.

Для підвищення точності позиціонування було досліджено фільтрацію виходу з алгоритмів машинного навчання за допомогою фільтра Калмана. Фільтр Кальмана використовується для вирішення проблеми прогнозування, фільтрації та згладжування. Це робиться таким чином, що час обчислення є низьким, так що можна реалізувати фільтр Кальмана, який працює в режимі реального часу [13].

Фільтр Калмана та алгоритми машинного навчання як це може працювати:

1. Збір даних: Спочатку збираються дані про місцезнаходження, такі як сигнали від супутників GPS, Wi-Fi, Bluetooth, сигнали від датчиків (наприклад, акселерометрів або жироскопів), інші вимірювання або датчики, які можуть надавати інформацію про рух.

2. Апаратне відфільтрування: Початкові дані можуть бути досить

неточними або шумними. Фільтр Калмана може бути використаний для апаратного відфільтрування даних і визначення проміжних оцінок місцезнаходження.

3. Машинне навчання: Алгоритми машинного навчання, такі як нейронні мережі, можуть бути навчені на основі історичних даних про місцезнаходження та руху для покращення точності передбачення.

4. Інтеграція: В інтегрованому підході алгоритми машинного навчання використовуються для покращення оцінок місцезнаходження, отриманих від фільтра Калмана.

5. Корекція та адаптація: Фільтр Калмана може використовуватися для корекції передбачень, отриманих від алгоритмів машинного навчання, з урахуванням нових вимірювань. Це дозволяє системі бути адаптивною до змінних умов.

6. Вивід точного місцезнаходження: Зібрані та оброблені дані використовуються для визначення точного місцезнаходження об'єкта.

Такий комбінований підхід дозволяє покращити точність позиціонування, навіть у складних умовах, де один метод може бути недостатнім. Машинне навчання допомагає адаптувати систему до різних умов, а фільтр Калмана використовується для оптимального комбінування інформації та врахування шумів у даних.

У цій тезі використано аспект фільтрації фільтра Калмана. Якщо припустити, що шум технологічного процесу та вимірювальний шум є гауссовими, фільтр Калмана є найкращим можливим оцінювачем як лінійних, так і нелінійних оцінок [19].

Виявлення умов NLOS є важливим етапом в розробці алгоритмів та систем, які працюють у безпровідних мережах та системах позиціонування, оскільки це допомагає враховувати особливості розповсюдження сигналів та забезпечувати надійну комунікацію та точне позиціонування.

# 3 ЗАСТОСУВАННЯ МЕТОДІВ МАШИННОГО НАВЧАННЯ ДЛЯ ПОЗИЦІОНУВАННЯ ПО 5G-БАЗОВИХ СТАНЦІЯХ

## 3.1 Опис даних

У цьому розділі методи та різні алгоритми пояснюються так, як вони були застосовані у данному дослідженні. Також представлена повна постановка проблеми.

Застосування методів машинного навчання для позиціонування за допомогою 5G-базових станцій є однією зі сучасних технологічних тенденцій, яка дозволяє покращити точність визначення місцезнаходження об'єктів у мережах п'ятого покоління.

Ключові аспекти і переваги цього підходу:

*Зібрання та аналіз даних:* 5G-мережі забезпечують велику кількість даних, що передаються між базовими станціями та пристроями. Ці дані можуть включати параметри сигналу, такі як сила сигналу, затримка, напрямок, час приймання тощо. Машинне навчання може бути використане для аналізу цих даних та визначення, як вони пов'язані з місцезнаходженням.

*Використання нейронних мереж*: Нейронні мережі, зокрема глибокі нейронні мережі, можуть бути навчені на основі історичних даних про місцезнаходження та параметри сигналів для розробки моделей, які визначають залежності між цими факторами. Ці моделі можуть бути використані для передбачення місцезнаходження на основі нових даних.

*Ансамблі моделей*: Можна використовувати ансамблі моделей машинного навчання, такі як випадковий ліс або градієнтний бустінг, для покращення точності позиціонування. Ці ансамблі об'єднують декілька моделей, щоб вирішити завдання позиціонування більш точно.

*Адаптація до умов*: Машинне навчання може навчити систему адаптуватися

до різних умов, таких як зміни в розташуванні антен або завади в сигналі. Система може навчитися вирішувати проблеми позиціонування в умовах обмеженої видимості базових станцій або інших складних сценаріях.

*Реалізація в реальному часі:* Після навчання моделей машинного навчання вони можуть бути використані в реальному часі для визначення місцезнаходження об'єктів. Це особливо важливо для додатків, які вимагають негайного відслідковування місцезнаходження, наприклад, в автомобільній промисловості чи системах безпеки.

Це підходить для різних варіацій задач позиціонування, таких як:

- Висока точність: машинне навчання може допомогти досягти високої точності позиціонування в порівнянні з іншими методами, що робить його ідеальним для додатків, де важлива точність на рівні метрів або сантиметрів.

- Використання багатошляхового розповсюдження: у місцях зі складними умовами, такими як багатошляховий розповсюдження сигналу, методи машинного навчання можуть бути ефективними у вирішенні завдань позиціонування, таких як внутрішнє позиціонування в приміщеннях.

- Підвищення безпеки: використання машинного навчання для визначення місцезнаходження може покращити безпеку систем та додатків, забезпечуючи точні дані про місцезнаходження для служб безпеки та надзвичайних ситуацій.

Загалом, застосування методів машинного навчання для позиціонування в мережах 5G відкриває широкі можливості для покращення точності та надійності визначення місцезнаходження. Це може бути корисним у різних сферах, від телекомунікацій до автономних автомобілів та міського планування. Однак, разом з цим, важливо враховувати питання приватності та безпеки при зборі та використанні таких даних.

Дані перевірених тестів, які використовуються в цій тезі, сценарій, під час якого збиралися дані, та спосіб збирання даних – описані нижче. Також охарактеризовано характер обробки даних та процесу вибору функцій для

використання в алгоритмах машинного навчання та для їх виявлення. Оскільки одним із великих обмежень у цій кваліфікаційній роботі є відносно невеликий розмір наборів даних, будуть нижче також описані методи створення нового набору даних.

***Сценарій***

Сценарій, який використовувався: – це UE, що рухається по міській місцевості в місті Кіста, Стокгольм. UE – це автомобіль, оснащений антеною, яка зв'язує його з ТП. Він рухається зі швидкістю ходьби (близько 7 км / год). Ілюстрація сценарію та карти області позиціонування раніше була наведена на рисунках 1.1 та на рис. 1.2б відповідно.

На рис. 3.1 зображення UE та TP, використаних для проведенних випробувань. Носійна частота ТП становить 15 ГГц, а антена ТП, що використовується на пробній панелі, складається з двох антенних сіток 8 х 8. Частина антенних сіток використовується для цифрового формування випромінювання, що складається з 48 променів, горизонтальна ширина променя 6°, вертикальна ширина променя 5°. Балки розміщуються в п'ять різних вертикальних шарів з дев'ятьма десятьма балками в кожному шарі. Промінь-сітка показана на рис. 3.2.

У наборі даних, що реєструються, UE пройшов три різні шляхи. Всього існує близько 1200 точок даних, включаючи умови LOS та NLOS. Частина, що містить лише NLOS, становить близько 400 точок даних. Частина даних, де доступно кілька зразків для однієї позиції, дуже обмежена.



Рис. 3.1. Зображення UE і TP, використаного в тестовій системі

Дані, що реєструвалися, складаються з BRSRP для всіх 48 променів, вибірки з інтервалом 10 мс, та даних про положення, що реєструвалися приймачем глобальної системи позиціонування (GPS) зі швидкістю вибірки 10 Гц. Отже, відстань між послідовними вимірюваннями становить близько 0,19 м. Для отримання однакової довжини вхідного та цільового значення дані BRSRP усереднюються зпоказників кожних десяти вимірювань. Це може розглядатися як низькочастотна фільтрація входу. Варто зауважити, що похибка положення GPS для UE під відкритим небом може становити до 5 метрів і навіть більше в міських каньонах [20].

Із зареєстрованих GPS-позицій, заданих по довготі та широті, обчислювалась відстань від UE до TP. Потім система координат змінювалась від системи довготи / широти-координати до декартової системи координат з положенням TP за походженням, це та система координат, що використовувалася у цьому разі.

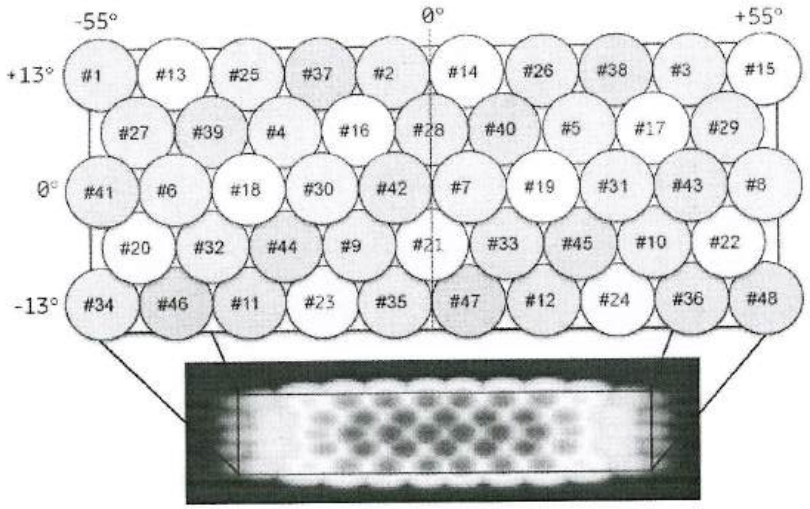



Рис. 3.2. Промінь-сітка ТП з несучою частотою 15 ГГц. У нижній частині малюнка, чим більше червоного кольору, тим більша потужність сигналу

***Вибір найкращого променя***

Кожен момент миттєво вибиралося n пучків з найвищим BRSRP. Кількість пучків n, які були досліджені в цій роботі, вибирався між п'ятьма і п'ятнадцятьма (рис. 3.3).

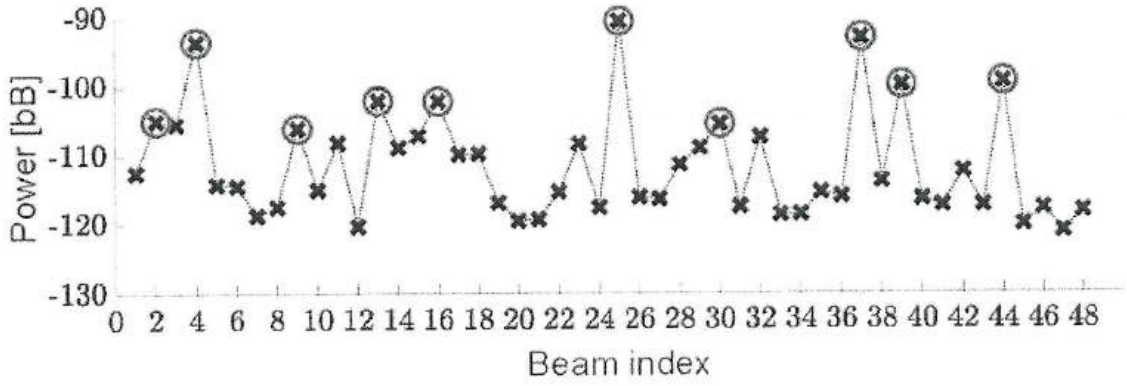



Рис. 3.3. Вибір променів з найбільшою потужністю сигналу. Тут обрано 10 балок з найвищим BRSRP

Це пов'язано з тим, що інформація з променів з низьким BRSRP є (як правило) недостовірною, і вона додає складності алгоритмам, забезпечуючи при цьому граничне підвищення продуктивності. Це також є наслідком

інтерпольованих обмежень та обмеження часу для випробувань. Приклад процесу вибору променя показаний на рис. 3.3.

***Обертання опорного кадру***

Під час тесту на позиціонування за допомогою нейронних мереж було виявлено велику різницю в продуктивності між напрямками x і y. Щоб мінімізувати помилку позиціонування:

$$\sqrt{(\widehat{pos}_x - pos_x^{true})^2 - (\widehat{pos}_y - pos_y^{true})^2}, \tag{3.1}$$

обертання опорного кадру було досліджено, де pos – оцінене положення. Результати показані на рис. 3.4а.

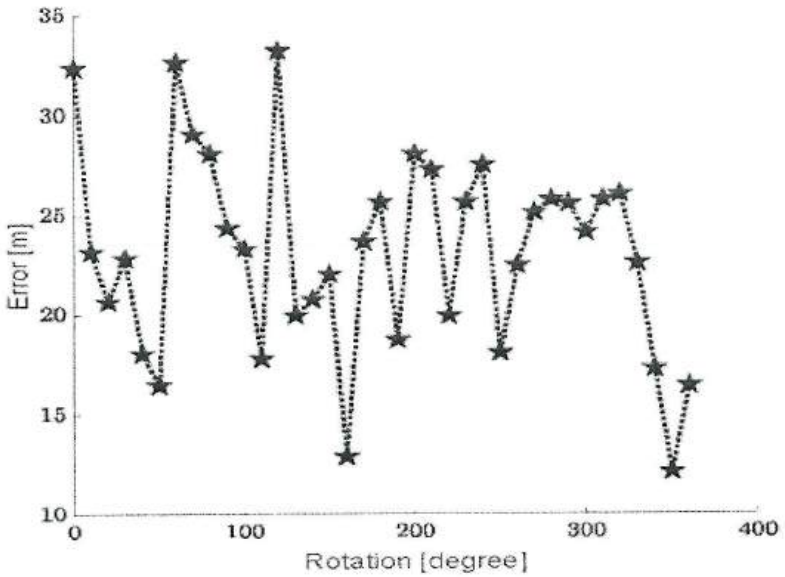



Рис. 3.4а. Продуктивність для різного кута повороту. На осі y – абсолютна похибка положення, а на осі x – кут повороту

З рис. 3.4а слідує, що обертання системи відліку обрано як 160 °. Обертання на 160 ° збігається з кутом різниці між опорним кадром для тракту UEs як функції

відстані від ТП і шляху, заданого в GPS координати.

На рис. 3.4б показано схематичне зображення обертання. Червоним на рис. 3.4b позначено покриття антени, а стара система координат - з чорною пунктирною віссю, а після обертання – синьою віссю.

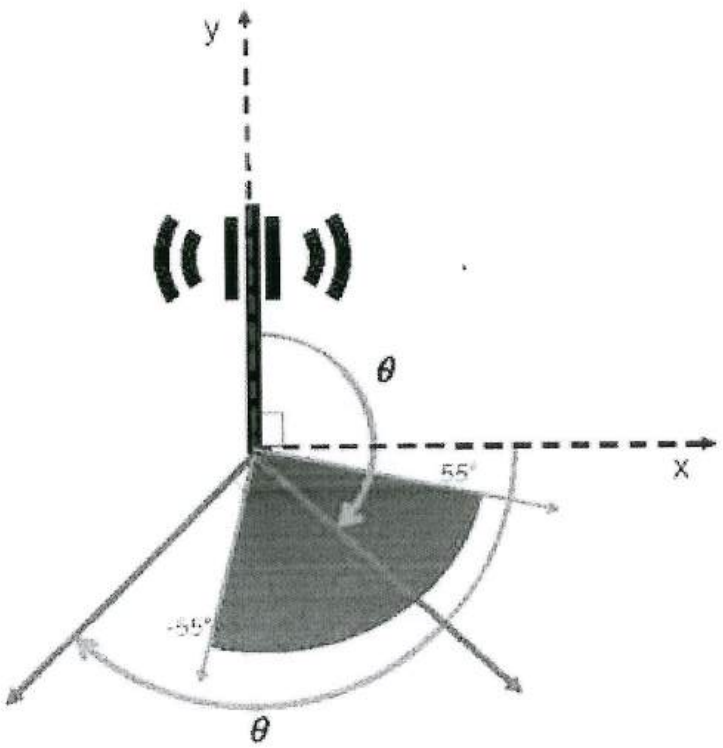



Рис. 3.4б. Обертання опорного кадру

Цей рисунок описує обертання опорного кадру. Червона дуга – це діапазон, в якому ТР має антени із покриттям, від -55 ° до 55 °.

***Вибір функції***

З BRSRP можна обчислити багато різних можливостей. Оцінка важливості функції проводилася за допомогою алгоритмів випадкових лісів, описаних у розділі 2.1. Тільки функції знімка використовуються як функції введення в алгоритми. У таблиці 3.1 наведені ознаки, використані в цій роботі.

***Оцінка напрямку вильоту***

Для розрахункового напрямку відправлення (DoD), тобто кута променя на антені, від якого відходить сигнал, слід обчислювалася модель променя антени. Цей розрахунок проводився за допомогою аналогічних алгоритмів, як у [2], з BRSRP як вхід до алгоритма. Розрахунковий коефіцієнт DoD для трьох різних променів показаний на рис. 3.5a, а нижня межа Креймера-Рао (CRLB), що оцінювач вибирає правильний кут із заданими двома альтернативами, обчислювався за результатами [5, 9] та показано на рис. 3.5b.

Таблиця 3.1.

Класи функцій, що були обрані як вхідні дані до алгоритмів машинного навчання

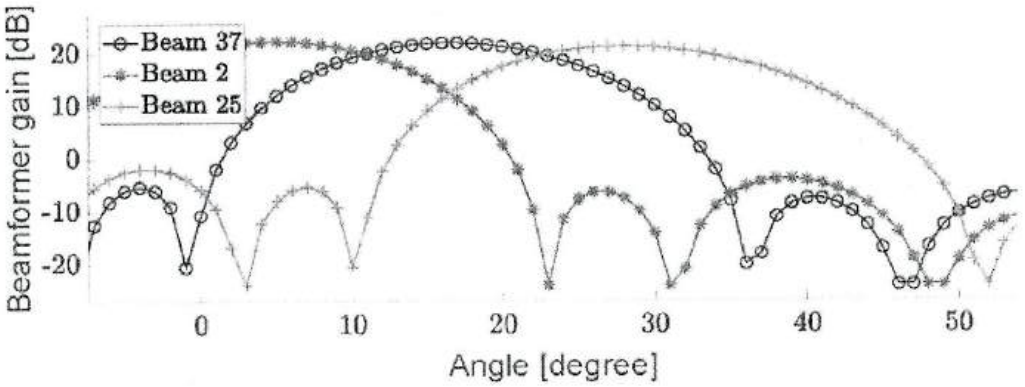
| Функція | Опис |
|---|---|
| BRSRP | Промінь опорного сигналу отримав потужність, яка визначена в [21]. |
| DOD | Напрямок відправлення. |
| DBRSRP | Difference in BRSRP - Різниця в BRSRP між найсильнішими та послідовними найсильнішими пучками. |
| DDOD | Різниця DoD між найсильнішими та послідовними найсильнішими пучками. |

Там, де на антені промені, чиї DoD є оцінкою на рис. 3.5a можна бачити на рис. 3.2. Промені вибрані таким чином, щоб вони знаходилися поруч один з одним в антені, щоб продемонструвати роздільну здатність DoD. Більш детально про алгоритм, який використовувався для оцінки DoD, можна знайти в [3, 4]. Для розрахунків CRLB на рис. 3.5b використовуються методи з [5, 9]. Там доведено, що CRLB для двох променів в якості кандидатів можна обчислити як

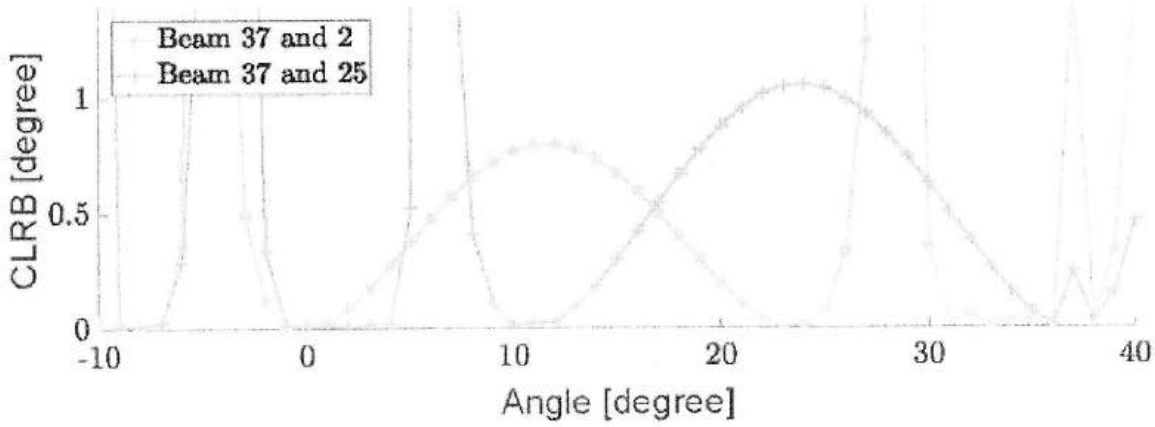
$$\mathrm{Var}(\phi) - \frac{2\sigma^2_{\mathrm{BRSRP}}}{\left(\frac{dH^{ij}(\phi)}{d\phi}\right)^2}, \tag{3.2}$$

де $\phi$ - оцінка DoD, а $\sigma^2_{BRSRP}$ – це дисперсія вимірювального шуму, а $H^{ij}(\phi)$ – різниця коефіцієнта підсилення пучка формувача між променем і та j.

Тут для варіантів вибрано дисперсію вимірювального шуму як 1 дБ. На рис. 3.5b видно, що коли оцінка потрапляє на бічну частку, CRLB стає дуже великою.




(а) Коефіцієнт посилення променя як функція кута для трьох суміжних зразків променя. Оцінка DoD є Argmax посилення формувача променю




(б) CRLB для DoD між значенням 37 променя і двома його сусідніми променями.

Рис. 3.5. Оцінка DoD та CRLB між оцінкою для одного променя та сусідніх променів

***Генерування більшого набору даних***

Як описано в розділі, кількість даних дуже мала. Тому досліджено способи генерації нових наборів даних. Досліджено та випробувано два різні способи генерування нових наборів даних. Це інтерполяція, що ґрунтується на різниці у швидкості вибірки, та використовуючи той факт, що антена має різні вертикальні шари.

*Інтерполяція*

Через різницю у швидкості вибірки між даними BRSRP та даними GPS, можна отримати більше даних, інтерполюючи дані GPS-приймача. Період вибірки BRSRP та вимірювань GPS становить 10 мс та 100 мс відповідно. Цей коефіцієнт десять між показниками вибірки дає можливість інтерполювати та створювати в десять разів більше даних. У цьому випадку десять послідовних BRSRP отримують однакове цільове значення.

*Поділ вертикальних балок шарів*

Горизонтальне положення UE представляє інтерес для цього завдання, і тому вертикальне положення не розглядатиметься. Оскільки антена має п'ять різних вертикальних шарів променів (див. рис. 3.2) одна ідея полягає в тому, щоб генерувати більше даних, вважаючи, що кожен вертикальний шар променями є незалежними наборами даних. Цей метод також збільшить кількість даних з коефіцієнтом п'ять.

***Показники ефективності***

Методика ефективності, що була використовувана тут, базується на тій, яка представлена в приміщенні дослідження щодо розміщення в приміщенні в проекті партнерства третього покоління (3GPP) [22]. Показник ефективності, згаданий у цьому звіті 3GPP, - це значення, при яких функція кумулятивного розподілу (CDF) помилки позиціонування досягає 40%, 50%, 70%, 80% і 90%. У цьому звіті фокус приділятиметься 80%.

Оскільки інформація про функцію щільності ймовірності (PDF) помилки позиціонування відсутня, CDF неможливо обчислити. Тому замість використання CDF в якості метрики ефективності використовується її неупереджена оцінка емпіричної функції кумулятивного розподілу (ECDF). Визначення ECDF

$$\hat{F}_n(t) = \frac{1}{n}\sum_{i=1}^{n} 1_{X_i<t}, \tag{3.3}$$

де $X_1,\ldots,X_n$ є незалежними, однаково розподіленими випадковими змінними зі спільним PDF $f_{X(y)}$, а $1_{Xi<t}$ - функція індикації події $X_i < t$, наприклад функція, яка є однією, коли подія $X_i < t$ відбувається а нуль інакше. Тоді може бути показано, використовуючи закон великих чисел, що $\hat{F}_n(t) \xrightarrow{a.c.} F(t)$, [23]. Зважаючи на примітну простоту, протягом решти звіту не буде проводитись різниця між CDF та ECDF.

### 3.2. Нейронні мережі

В MATLAB можна реалізовувати нейронні мережі завдяки наявності спеціального інструментарію для машинного навчання і глибокого навчання. Один із найбільш популярних інструментів для цього - це Neural Network Toolbox, який надає багато можливостей для створення та навчання нейронних мереж.

Кроки, для реалізації нейронних мереж в MATLAB:

*Створення нейронної мережі:* Використовуючи функції інструментарію Neural Network Toolbox, можна створити модель нейронної мережі. Це включає в себе визначення кількості шарів, кількість нейронів у кожному шарі, типи шарів (наприклад, пов'язані шари або звільнені шари), функції активації тощо.

*Навчання нейронної мережі:* MATLAB надає інструменти для навчання нейронних мереж на основі доступних даних. Ви можете використовувати різні

алгоритми оптимізації та функції втрат для досягнення найкращих результатів.

*Перевірка та тестування*: Після навчання мережі важливо перевірити її продуктивність і точність за допомогою тестового набору даних. MATLAB також надає інструменти для оцінки відповідності мережі вашим потребам.

*Використання нейронної мережі:* Після успішного навчання і тестування нейронної мережі, ви можете використовувати її для прогнозування, класифікації, розпізнавання образів, обробки сигналів, аналізу даних тощо.

*Оптимізація та налаштування:* По мірі розвитку проекту ви можете оптимізувати параметри мережі і вдосконалювати її продуктивність.

Нейронні мережі реалізовані в MATLAB за допомогою панелі інструментів нейронної мережі. Завдання стає проблемою розрахунку регресії, де метою є пристосування функції від вхідних функцій до цільових значень. Нейронна мережа, що використовується, має два прихованих шари та функцію активації h (•), що в першому шарі є tanh (), а функція активації у другому шарі - це чиста лінія.

***Розмір прихованих шарів***

Для визначення розміру прихованих шарів для використання його в нейронній мережі були реалізовані мережі з різною кількістю нейронів. Потім було обчислено ефективність усіх мереж за даними оцінювання та було обрано кількість нейронів чи розмір прихованого шару, які давали найкращі показники (як визначено у підрозділі 3.1) за їх складністю. Розміри прихованих шарів вибираються відповідно до рис. 3.6, а значення узагальнені в таблиці 3.2.

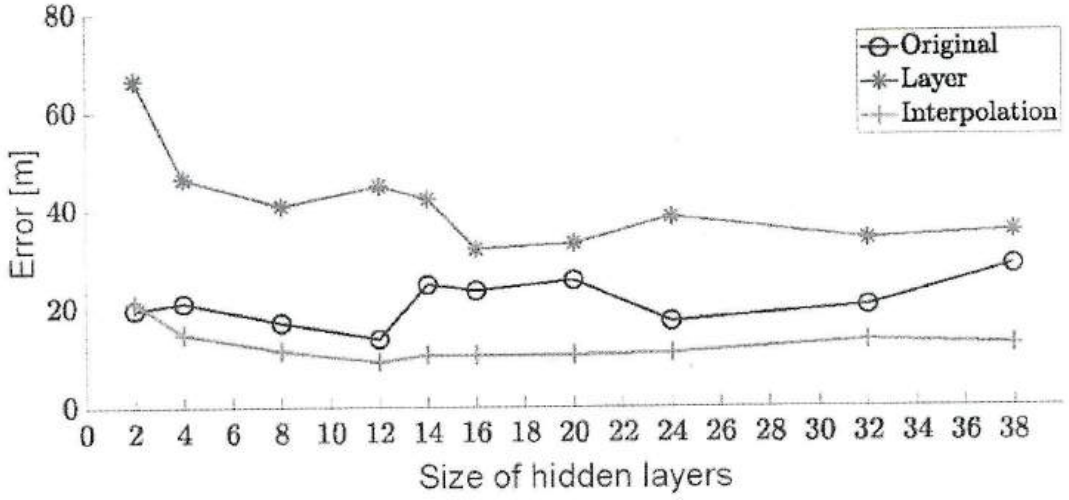



Рис. 3.6. Вибір кількості нейронів у прихованих шарах для різних наборів даних.

Таблиця 3.2.

Кількість нейронів у прихованому шарі для різних наборів даних.

| Набір даних, що містять: | Кількість нейронів у прихованих шарах |
|---|---|
| Оригінальні дані | 12 |
| Інтерполяційні дані | 12 |
| Відокремлений вертикальний шар променя | 14 |

***Об'єднання декількох мереж***

Для запобігання перетренованості та підвищення продуктивності використовується техніка поєднання результатів з декількох мереж. Кілька нейронних мереж з однаковою кількістю нейронів і шарів отримують однаковий набір даних про порядок навчання для навчання.

Після навчання всі мережі тестувалися з однаковими тестовими даними, а вихідний показник усереднювався по всіх мережах. Потім середні результати перетворювалися щодо справжньої позиції UE. Визначалася середня помилка як

$$\bar{e} = \sqrt{\left(\frac{1}{M}\sum_{m-1}^{M}\widehat{pos}_x^m - pos_x^{true}\right)^2 + \left(\frac{1}{M}\sum_{m-1}^{M}\widehat{pos}_y^m - pos_y^{true}\right)^2} \quad (3.4)$$

де M – кількість нейронних мереж, $(\widehat{pos}_x^m, \widehat{pos}_y^m),\ m = 1, \cdots, M$ – виходи з окремих мереж, $(pos_x^{true}, pos_y^{true})$ – це справжня позиція, і e – середня помилка. Існує багато інших прикладів успішного поєднання декількох мереж для підвищення продуктивності нейронних мереж [23].

***Попередня обробка функцій***

Для підвищення продуктивності нейронних мереж робиться деяка обробка входів і цілей. Методи, що використовуються в цій тезі, включають: нормалізацію даних, що знаходяться між [-1,1] з нульовою середньою величиною, та дисперсію одиниці разом із усуненням постійних ознак. Опис різних методів попередньої обробки та їх вплив на алгоритми машинного навчання можна знайти в [2]. Після підрахунку регресії вихід перетворюється на вихідні одиниці.

Загалом, MATLAB - це потужний інструмент для розробки та експериментів з нейронними мережами у вас на комп'ютері. Він надає багато можливостей для дослідження глибокого навчання і розробки різноманітних застосувань на основі нейронних мереж.

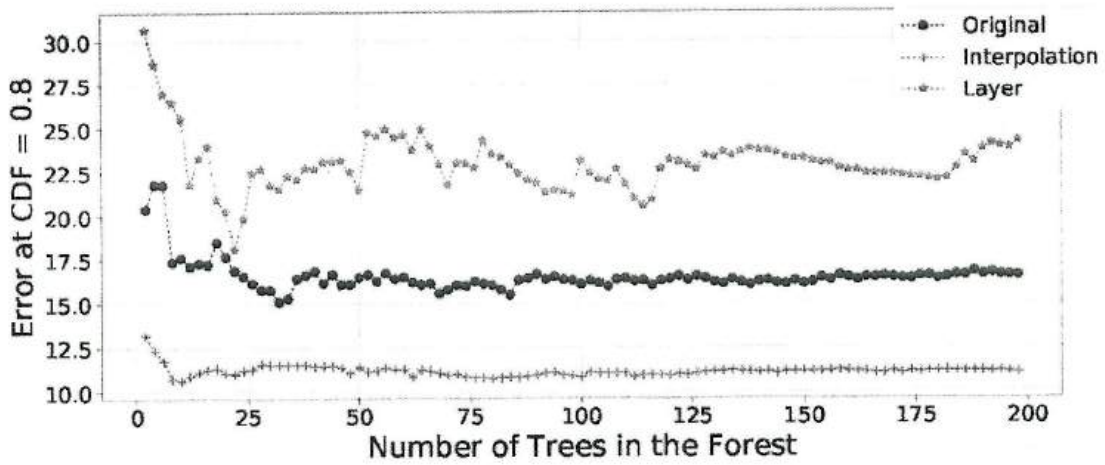
### 3.3 Метод машинного навчання "random forest"

Random forest реалізований у PYTHON за допомогою scikit learn. Проблема тут (як і у випадку з нейронними мережами) – це проблема регресії, і тому використовувалася бібліотека Random Forest Regression у scikit learn.

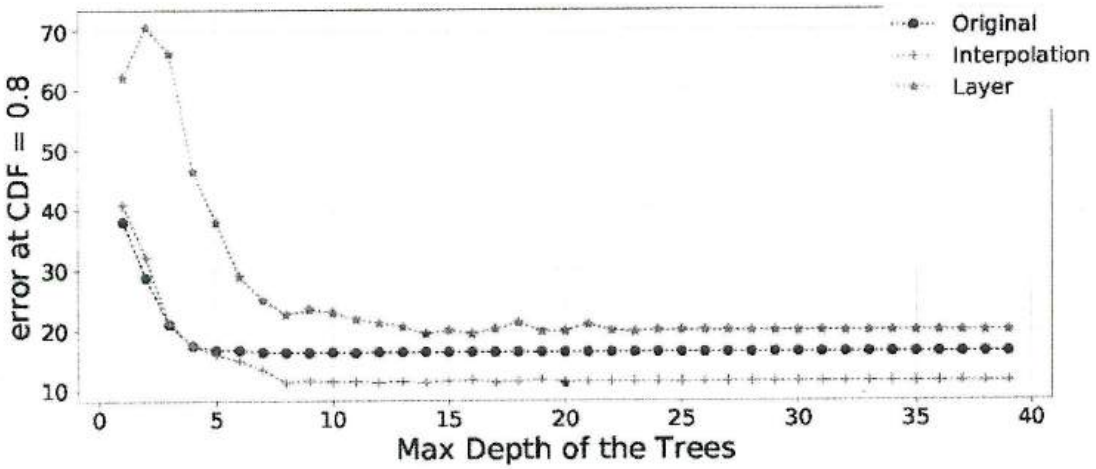
***Кількість «дерев у лісі»***

Існує поріг у кількості дерев, що використовуються у випадковому лісі,

коли збільшення кількості дерев не призводить до значного покращення продуктивності. Це за результатами, знайденими в [8]. Щоб знайти таку ситуацію в кількох деревах, було зроблено симуляцію. Результат можна побачити на рис. 3.7a. З цього видно, що після 125 дерев продуктивність збігається, отже, кількість дерев вибирається до 125 для всіх наборів даних.




а) Результати моделювання того, скільки дерев оптимально в лісі. Видно, що продуктивність збігається після того, як ліс складається з 125 дерев для всіх наборів даних.




(b) Результати моделювання глибини дерев. Видно, що продуктивність конверсує близько 10 дерев для всіх наборів даних.

Рис. 3.7. Вибір параметрів для Random forest.

***Глибина дерев***

Досліджено різну глибину дерев і результати показані на рис. 3.7b. Варто зазначити, що зміна глибини не впливає суттєво на похибку помилки позиціонування. Оскільки продуктивність сходить після глибини десяти дерев для всіх наборів даних, глибина вибирається як десять.

***Попередня обробка функцій***

Подібно до нейронної мережі, попередня обробка функцій була зроблена для підвищення продуктивності. Особливості введення орієнтовані навколо середнього значення та нормалізуються на одиницю дисперсії.

## 3.4 Спосіб та реалізація методів виявлення умов NLOS

У цьому підрозділі описані спосіб та реалізація описаних у розділі 2.2 методів виявлення умов NLOS. Далі пазано які сигнали використовувалися для визначення, яким чином параметри A і A2 отримані і апроксимовані, а також, як обчислювався поріг Y і PD для даного PFA. Алгоритми виявлення були реалізовані в MATLAB.

***Вибір сигналу***

Два різних сигнали, доступних для виявлення NLOS, є BRSRP і DoD. Перший сигнал, що використовується для виявлення, xi [n], використовує різницю BRSRP між променями, а другий сигнал, X2 [n], використовує різницю в DoD між променями. У першому випадку за допомогою BRSRP сигнал визначається як

$$x_1[n] = \frac{1}{M-1} \sum_{\hat{}}^{M} \mathrm{BRSRP}_1[n] - \mathrm{BRSRP}_m[n]_{\cdot} \cdot \tag{3.5}$$

де BRSRPi [n] - BRSRP найсильнішого променя в момент часу n, а BRSRPm [n] m = 2,. . . , M - BRSRP з M найпотужніших пучків. З цього зміщення сигналу вибирається такий, що сигнал в умовах LOS становить близько нуля. Це так, що

коли UE ввійде в область з умовами NLOS, буде крок в сигналі. Сигнал також нормалізується, щоб мати амплітуду одиниці при вході в NLOS. У другому випадку за допомогою DoD сигнал визначається як

$$x_2[n] = \frac{1}{M-1}\sum_{m=2}^{M} |\mathrm{DoD}_1[n] - \mathrm{DoD}_m[n]|. \tag{3.6}$$

де $DoD_1[n]$ - DoD найсильнішого променя в момент часу n, а DoD m [n] m = 2, ..., M - DoD M послідовних найсильніших пучків. Сигнал також нормалізується, щоб мати амплітуду одиниці при вході в NLOS. Для обох сигналів кількість променів M вибирається як 9. Цей вибір ґрунтується на дослідженні даних, а при додаванні нових променів не надається нова корисна інформація.

***Вибір параметра***

Використовуючи детектор NP, амплітуду A вибирали, вивчаючи xi [n] та X2 [n] та те, як вони поводяться там, де очікуються умови NLOS. Дисперсія a2 вибиралася як максимальна дисперсія сигналів, xi [n] і X2 [n], після класифікації, якщо вони були в умовах NLOS як для детектора NP, так і для GLRT. У таблиці 3.3 перераховані всі параметри, що використовуються для виявлення NLOS.

Таблиця 3.3

Вибір параметрів для виявлення NLOS.

| | BRSRP | | DoD | |
|---|---|---|---|---|
| | NP | GLRT | NP | GLRT |
| A | 0.5 | - | 0.5 | - |
| $a^2$ | 0.05 | 0.05 | 0.15 | 0.15 |
| $P_{FA}$ | $3 \cdot 10^{-5}$ | $3 \cdot 10^{-5}$ | $3 \cdot 10^{-5}$ | $3 \cdot 10^{-5}$ |
| N | 7 | 7 | 10 | 10 |

У таблиці 3.3 N посилається на кількість променів на класифікацію, яка вибирається таким чином, щоб роздільна здатність для виявлення була достатньо

точною, але все ще не було зафіксовано тенденцію сигналу. Ймовірність помилкової тривоги PFA вважається конструкцією. З PFA поріг ймовірність виявлення PD обчислювалася відповідно до (2.18) та (2.24). На рис. 3.8 показана робоча характеристика приймача для двох детекторів NP.

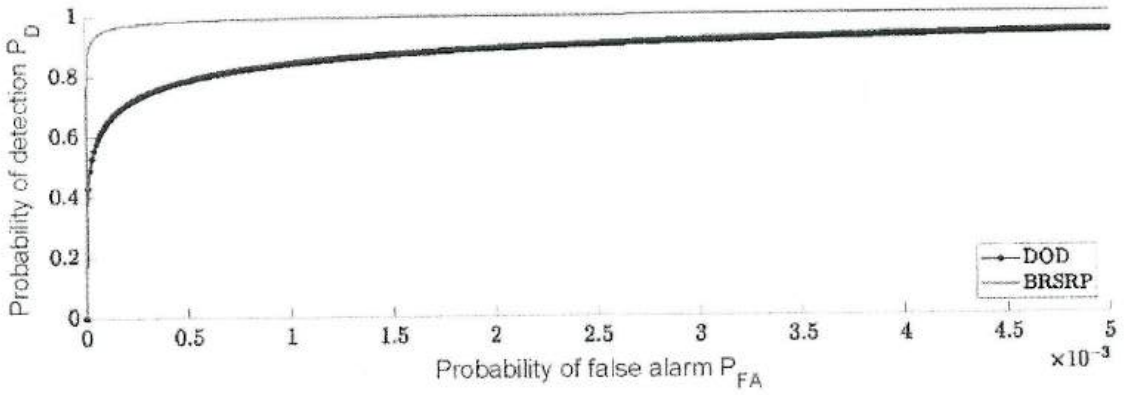



Рис. 3.8. Операційна характеристика приймача для моделі, що використовується в детекторі NP.

### 3.5 Настройка фільтра Калмана

У цьому підрозділі описана настройка фільтра Калмана та вказані його початкові параметри. Було також визначено вибір моделі простору стану та причина їх вибору. Фільтр Кальмана реалізований за алгоритмами, знайденими в [19]. Настройка фільтра Калмана перед вибором моделі простору стану - це важливий крок у процесі впровадження фільтрації Калмана для статистичного оцінювання системи. Для успішної роботи фільтра Калмана важливо правильно обрати модель простору стану, яка відображає динаміку системи.

Для налаштування фільтра Калмана перед вибором моделі потрібно виконати такі дії:

1. Перед настройкою фільтра Калмана, потрібно зібрати дані про систему.

Це може бути вимірювання реальних параметрів системи, таких як положення, швидкість, або будь-які інші змінні, які ви хочете оцінити за допомогою фільтра Калмана.

2. Проаналізувати зібрані дані для визначення характеристик системи. Розглянути структуру даних, статистичні характеристики, можливі джерела шуму та інші фактори.

3. Визначити модель простору стану яка визначає, як змінні системи залежать одна від одної в часі. Вона може бути лінійною чи нелінійною. Важливо правильно вибрати цю модель, оскільки вона визначає, яким чином система рухається та як вона взаємодіє з шумом.

4. Оцінити параметри моделі. Після визначення параметрів моделі простору стану, наприклад, матриці переходу і матриці вимірювань, можливо визначати, як система переходить від одного стану до іншого та як вимірювання впливають на стан.

5. Ініціалізувати фільтр Калмана, встановивши початкові значення математичного сподівання та коваріації для оцінок стану. Ці початкові значення можуть бути визначені на підставі початкових умов та точності даних.

6. Провести оцінку та корекцію за необхідністю. Потрібно запустити фільтр Калмана на обробку даних і оцінити стан системи з часом. За необхідністю можливо скоригувати фільтр на основі нових вимірювань.

7. Перевірка і підтвердження того, на скільки добре фільтр Калмана відповідає очікуванням та наскільки точні оцінки стану. Відповідно до результатів, можливо внести зміни в модель простору стану або параметри фільтра.

***Модель стану та простору***

Про те, що UE рухається з постійною швидкістю, раніше згадується у підрозділі 3.1. Ця інформація використовується при створенні моделі простору стану. Використовуючи відношення відстані для об'єкта, що рухається, з

постійною швидкістю матриця системи та матриця шуму моделювалася відповідно до [20-23]. Час вибірки, Ts, положення GPS становить 100 мс.

***Налаштування***

Як зазначено в підрозділі 2.3, процес шуму і вимірювання шуму передбачаються гаусовим, тобто, помилка оцінки машинного навчання алгоритми мають гауссово розподіл. Припускалося, що матриця коваріації шуму технологічного процесу становить 0,01 • I2x2м2, а матриця коваріації шуму вимірювань - 25 • I2x2м2. Вибір попереднього шуму випливає з того, що є довіра до алгоритмів машинного навчання, хоча відомо, що GPS може мати помилку ра до 5 м, отже, шум вимірювань становить 25 м2. Параметри ініціалізації вибираються як перше положення з нульовою швидкістю, [posx, i, posy, i, 0, 0], а початкова матриця коваріації обрана як 100 • I4x4, на основі досліджень різних значень початкової матриці коваріації.

# 4 ОЦІНКА ЕФЕКТИВНОСТІ ЗАПРОПОНОВАНИХ МЕТОДІВ

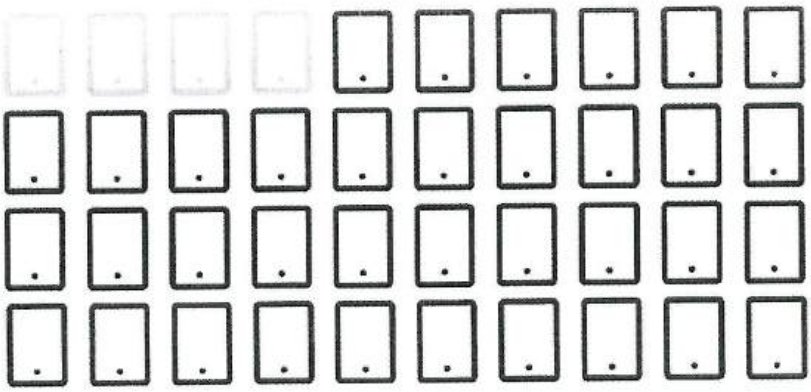
## 4.1 Результати, отримані з використанням нейронних мереж

У цьому розділі представлені результати оцінки алгоритмів машинного навчання. Вектор характеристик, що використовується в цих експериментах, складається з BRSRP для n пучків з найвищим BRSRP, визначених як найкращі пучки, DoD з n найкращих пучків, різниці BRSRP між найкращим променем і послідовними і різниці DoD між найкращим променем і послідовні. Для набору даних, де вертикальні шари променя в антені відокремлюються, кількість кращих променів встановлюється на п'ять. Для вихідних даних та інтерпольованих даних кількість найкращих променів встановлюється на десять. Ці цифри вибрані таким чином, що продуктивність не покращиться значно, коли додаються більше кращих променів.

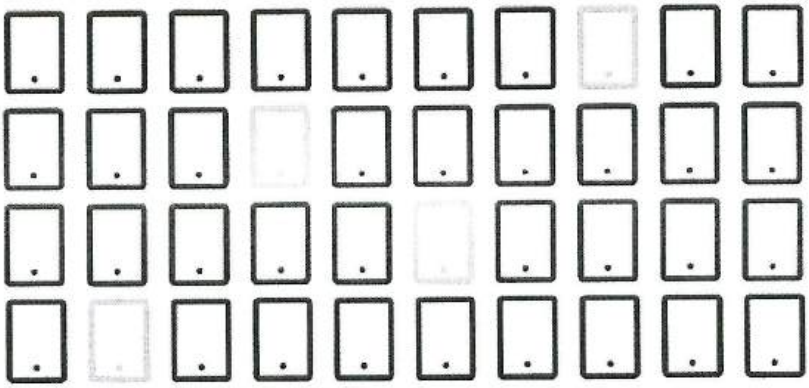
Для оцінювання використовувалося близько 10% даних, а решта 90% - для навчання. Продуктивність алгоритмів машинного навчання визначається як значення, коли CDF помилки позиціонування дорівнює 0,8. У додатку В наведена таблиця, що підсумовує результати всіх алгоритмів машинного навчання.

Існує два різних способи розділення набору даних на навчальний набір і тестовий набір, це послідовно або випадково, див. рис. 4.1. Для порівняння досліджувалося позиціонування в умовах LOS за допомогою алгоритмів машинного навчання, розроблених аналогічно алгоритму, який використовується для позиціонування NLOS.

Представлені результати виявлення умов NLOS, як з використанням детектора NP, так і з використанням GLRT. Цей розділ глави також включає результати дослідження фільтрації результатів з алгоритмів машинного навчання для підвищення точності позиціонування.



(a) Послідовно



(b) Випадково

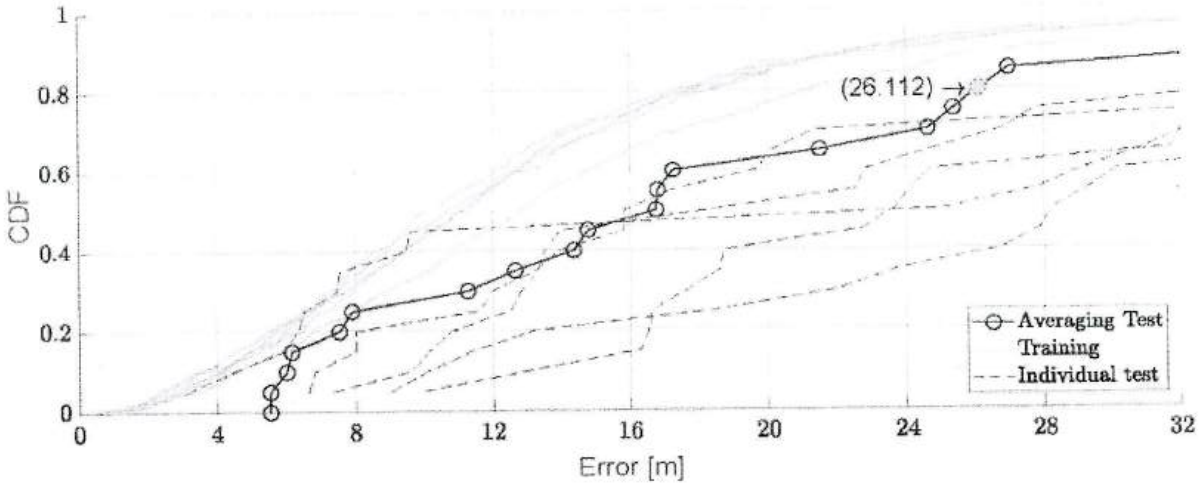
Рис. 4.1. Приклад двох способів поділу набору даних для оцінювання та вивчення даних послідовно (а) та випадковим чином (б). Помаранчеві телефони використовуються як набір для тестування

У цьому підрозділі описані результати, отримані з використанням нейронних мереж. На рис. 4.2-4.4 суцільною лінією є продуктивність, що поєднує вихід із понад 100 нейронних мереж, зелена – це ефективність даних про навчання, а чорна – окремі показники нейронних мереж. Для презентації були відібрани п'ять індивідуальних тестів та навчальних виступів. Кількість нейронів у прихованих шарах вибиралася відповідно до підрозділу 3.2.1, опис проектування нейронної мережі можна знайти там також.

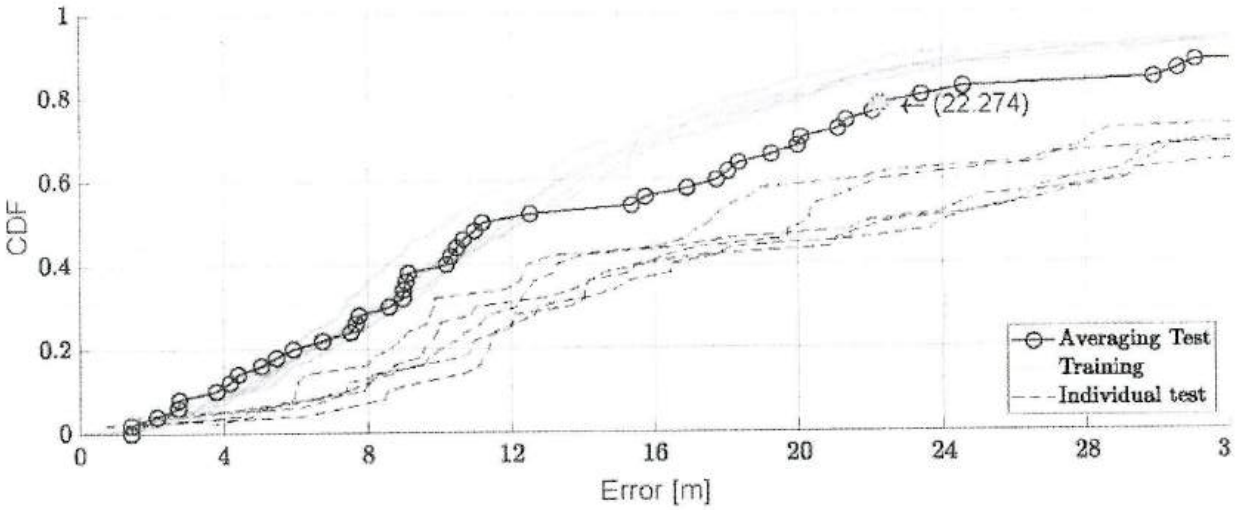
***Оригінальні дані***

Результати для нейронних мереж з навчальним набором, що складається з оригінального набору даних, показані на рис. 4.2. Результативність, що

оцінюється з послідовно та випадковим розділенням набору для навчання та тестування, показана на рисунках 4.2a та рис. 4.2b відповідно.




(a) Ефективність, оцінювана на тестових даних, послідовно відокремлених від набору даних.



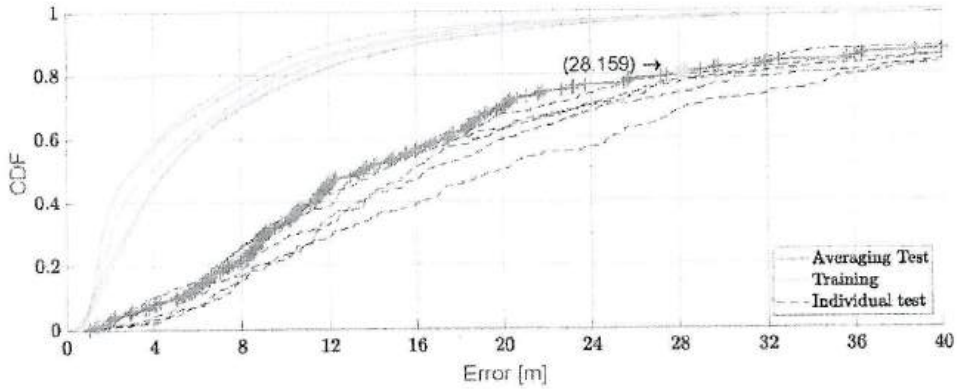

(b) Ефективність оцінюється на даних тесту, відокремлених випадковим чином від набору даних

Рис. 4.2. CDF помилки позиціонування для нейронних мереж, підготовлених та оцінених з оригінальним набором даних. Можна побачити, що між виконанням двома способами, послідовно та випадковим чином, є певна різниця показників для розділення набору даних у навчальному наборі та наборі оцінювання

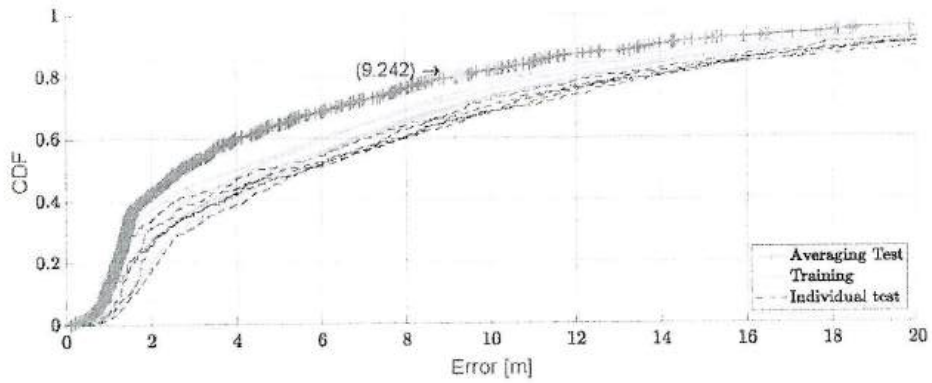
Як видно, ефективність усереднення результатів у кількох нейронних мережах підвищує ефективність позиціонування на оціночних даних. Можна також побачити, що продуктивність схожа на тестовий набір, відокремлений послідовно та випадковим чином від початкового набору даних.

***Інтерполяція даних***

Результати нейронних мереж з набором даних, що складається з даних про інтерполяцію, показані на рис. 4.3. Результативність, оцінена послідовним та випадковим розділенням набору для навчання та тестування, показана на рисунках 4.3a та рис. 4.3b відповідно.




(a) Ефективність, оцінювана на тестових даних, послідовно відокремлених від набору даних.

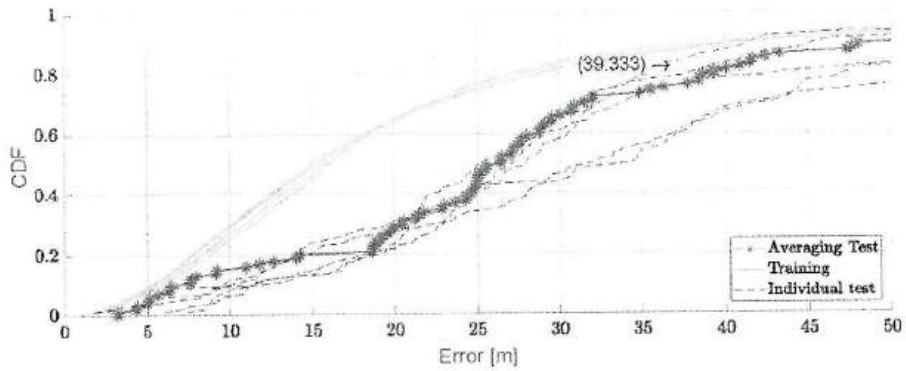



(b) Ефективність оцінюється на даних тесту, відокремлених випадковим чином від набору даних.

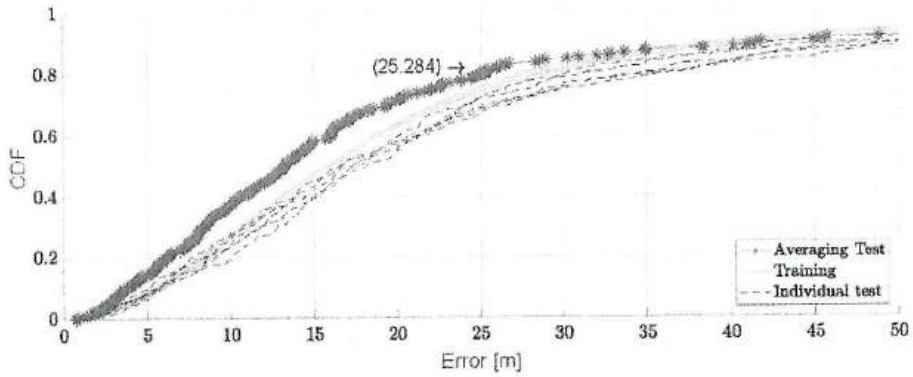
Рис. 4.3. CDF помилки позиціонування нейронних мереж, підготовлених та оцінених на наборі даних, створеному з використанням різниці у швидкості вибірки між BRSRP та GPS.

Завдяки кращій якості для тестового набору, розділеного випадковим чином, шкала на осі x вдвічі перевищує розмір на рис. 4.3a порівняно з рис. 4.3b

На рис. 4.3a спостерігається дивна поведінка CDF. Результати є дуже поганими порівняно з рисунком 4.3b, і CDF є майже прямою лінією після певної кількості оцінених позицій. Це може бути наслідком невеликого набору даних про навчання. Оскільки набір для навчання невеликий, набір для навчання може пропускати бали, аналогічні цільовим значенням, наведеним у наборі даних тесту. Це призводить до того, що всі оцінки можуть закінчуватися в одній точці, далеко від справжнього значення. На рис. 4.3b використання декількох мереж підвищує продуктивність.




(a) Ефективність, оцінювана на тестових даних, послідовно відокремлених від набору даних.




(b) Ефективність оцінюється на даних тесту, відокремлених випадковим чином від набору даних.

Рис. 4.4. CDF помилки позиціонування для нейронних мереж, підготовлених та оцінених за допомогою набору даних, де вертикальні промені шару в ТП розділені

***Поділ вертикальних шарів пучка***

Можна помітити, що є більш ніж на десять метрів кращі показники, коли навчальний набір і тестовий набір розділені випадковим чином.

Результати нейронних мереж з набором даних, в яких розділені різні вертикальні шари, показані на рис. 4.4. Результативність, оцінена послідовно та випадковим поділом набору для навчання та тестування, показана на рис. 4.4a та рис. 4.4b відповідно.

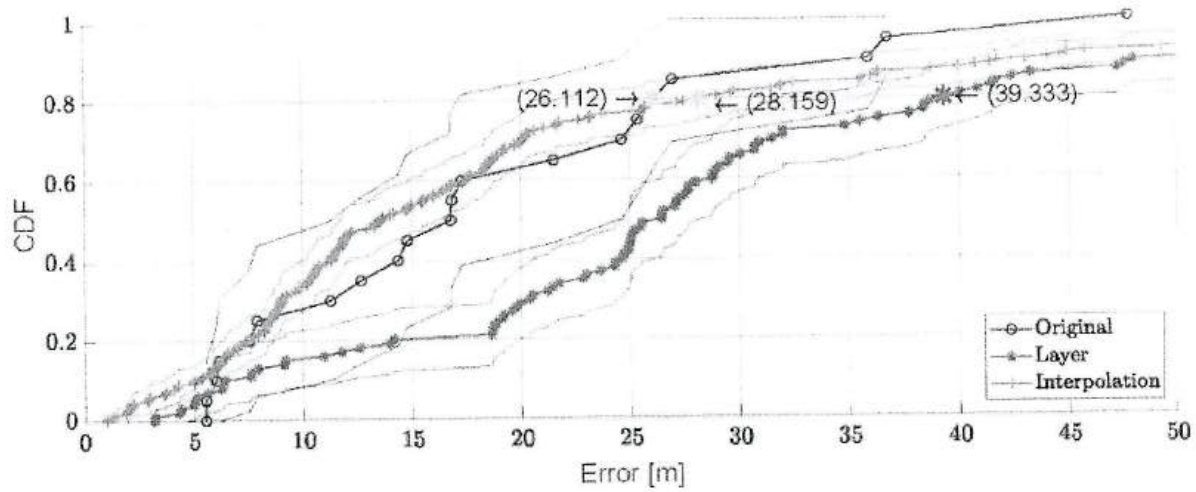
Як очікується, комбінування виходу з декількох нейронних мереж підвищує продуктивність. Слід звернути також увагу, що продуктивність, яка використовує промені лише з одного шару, значно гірша, ніж оцінка на початковому наборі даних (див. рис. 4.2 та рис. 4.4), де використовується інформація з декількох шарів. Це вказувало б на те, що використання інформації з декількох шарів має важливе значення для позиціонування з високою точністю.

***Порівняння навчальних наборів***

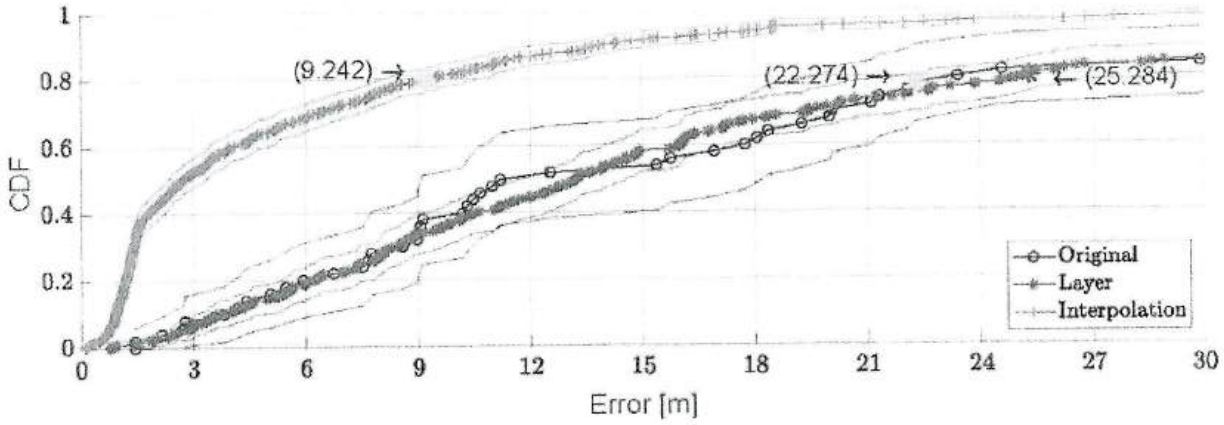
На рис. 4.5 показано ефективність для різних навчальних наборів; як для випадковим чином і послідовного поділу навчання і тестування даних. Діаграми такі ж, як синя лінія на рис. 4.2, пурпуровий на рис. 4.3 та червона на рис. 4.4, тобто точність позиціонування після поєднання виводу з декількох нейронних мереж.

Оскільки точність позиціонування значно краща для випадкового поділу набору для навчання та тестування, шкала на осі x для випадково відокремлених менша.

95% довірчий інтервал для CDF, обчислений за формулою Greenwood, також показаний для всіх наборів даних. Довіра представлена двома рядками, в тому ж кольорі, що і дані, які представлені набором даних. Інтервал довіри для вихідного набору даних значно більший, ніж для інших; це може бути пов'язано з обмеженою кількістю точок даних у вихідному наборі даних. Точка, де CDF досягає 0,8, виділена на рис. 4.5a та на рис. 4.5b.




(а) Ефективність, оцінювана на тестових даних, послідовно відокремлених від набору даних.




(б) Ефективність оцінюється на даних тесту, відокремлених випадковим чином від набору даних.

Рис. 4.5. Порівняння продуктивності для нейронних мереж, підготовлених та оцінених різними наборами даних.

## 4.2 Метод машинного навчання “random forest”

У цьому підрозділі представлений результат визначення позиціонування за допомогою алгоритму випадкових лісів. Показані визначені як продуктивність для різних вхідних функцій, так і ранжування їх важливості.

Використання методу "Random Forest" для позиціонування:

- Збір даних місцезнаходження: Першим кроком у використанні методу "Random Forest" для позиціонування є збір відповідних даних місцезнаходження. Це може бути інформація, отримана від супутникових систем GPS, Wi-Fi, Bluetooth, стільникових даних та інших джерел.

- Створення навчального набору даних: необхідно підготувати навчальний набір даних, який включає в себе ознаки (фактори), які можуть бути корисними для визначення місцезнаходження. Ознаки включають сигнали з різних джерел, сили сигналів, інформацію про сусідні точки доступу, часові параметри.

- Створення моделі "Random Forest": будується модель "Random Forest", яка є ансамблем дерев рішень. Кожне дерево в ансамблі навчається на підмножині навчального набору даних з випадковим вибором ознак. Це дозволяє моделі бути стійкою до перенавчання та забезпечує високу точність.

- Навчання та тестування моделі: ьодель навчається на навчальних даних, а потім перевіряється на тестових даних для оцінки її точності та продуктивності. Важливо враховувати, що дані для тестування повинні бути незалежними від даних для навчання.

- Визначення місцезнаходження: Після навчання моделі вона може бути використана для визначення місцезнаходження на основі вхідних даних. Модель "Random Forest" надає імовірнісний результат, що дозволяє враховувати невизначеність в позиціонуванні.

- Оцінка результатів: Оцінка результатів включає в себе аналіз точності позиціонування, виявлення можливих викидів та вдосконалення моделей.

***Важливе значення***

Метод машинного навчання "Random Forest" є потужним і ефективним інструментом для вирішення завдань позиціонування в мережах 5G та інших контекстах. Цей метод належить до категорії ансамблевих алгоритмів машинного навчання і використовується для класифікації, регресії та інших завдань, включаючи позиціонування, завдяки його високій точності та здатності працювати з багатовимірними даними.

Використовуючи алгоритм випадкового лісу, можна отримати рейтинг ознак. Ось наскільки важливими є функції для отримання прогнозованого значення. Особливості тестування: BRSRP, DoD, різниця між променями з найвищим BRSRP та послідовними (DBRSRP), а також різниця в DoD між найкращим променем і послідовним (DDOD) для всіх n обраних кращих променів. Результати наведені в Таблиці 4.1, де більший бал вказує на те, що функція важливіша, виділено особливість з найвищим значенням

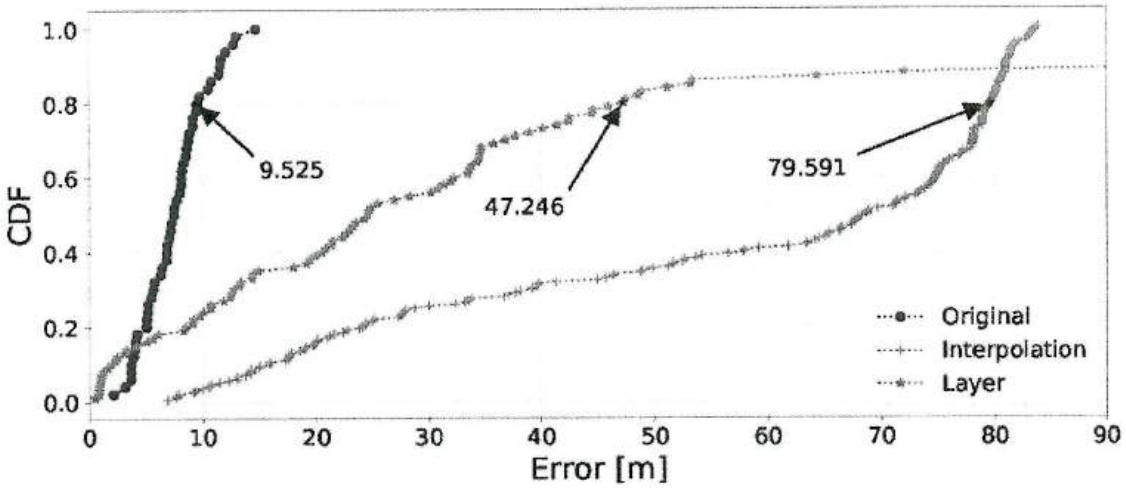
Таблиця 4.1.

Рейтинг важливості функції, сформований алгоритмом випадкових лісів для вихідного набору даних. Найкращий промінь – це промінь з найвищим BRSRP.

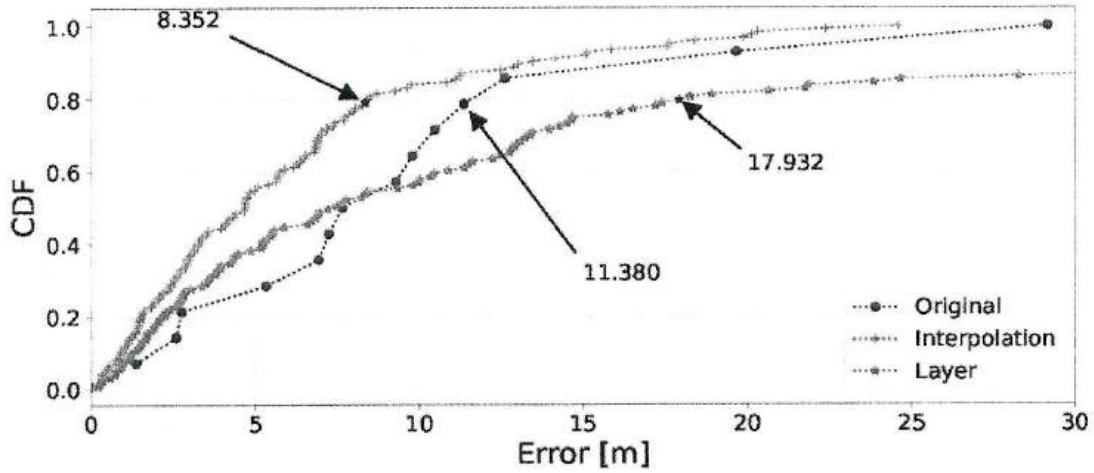
| Кращий промінь № | BRSRP | DoD | DDOD | DBRSRP |
|---|---|---|---|---|
| 1 | 0.0034 | 0.0102 | - | - |
| 2 | 0.0049 | 0.0161 | 0.0220 | 0.0032 |
| 3 | 0.0043 | 0.0052 | 0.0026 | 0.0218 |
| 4 | 0.0037 | 0.0251 | 0.0028 | 0.0348 |
| 5 | 0.0062 | 0.0652 | 0.0073 | 0.0041 |
| 6 | 0.0022 | 0.0071 | 0.0093 | 0.0031 |
| 7 | 0.0049 | 0.0086 | 0.0040 | 0.0118 |
| 8 | 0.0106 | 0.0016 | 0.0058 | 0.0186 |
| 9 | 0.0170 | 0.0019 | 0.0058 | 0.0128 |
| 10 | 0.0069 | 0.0040 | 0.0057 | **0.6196** |

Варто зазначити, що для всіх наборів даних остання особливість, яка виділяється, має дуже велике значення. Це відкриває для дослідження, якщо дані з більшої кількості променів покращать продуктивність.

***Порівняння навчальних наборів***




(а) Ефективність, оцінювана на тестових даних, послідовно відокремлених від набору даних.




(б) Ефективність оцінюється на даних тесту, відокремлених випадковим чином від набору даних.

Рис. 4.6. Порівняння ефективності для випадкових лісів, підготовлених та оцінених різними наборами даних.

Рисунок 4.6 показує, в результаті виконання алгоритму випадкового лісу оцінюється на різних наборах даних. Ці ліси навчаються та оцінюються на навчальних наборах та наборах для тестування, які розділяються послідовно та випадковим чином. Варто зауважити: низька ефективність інтерполяції щодо послідовного розділення даних про навчання та тестування порівняно з випадковим чином розділеними. Це може бути пов'язано з нестачею даних або відсутністю даних про навчання на певній ділянці вздовж шляху, по якому проходять курс UE. Це твердження підтверджується тим, що похибка позиціонування майже рівна між 10 м і 60 м, порівняно з 60 м і далі.

Варто зауважити – низька точність позиціонування, коли набір для навчання та тестування відокремлюються послідовно порівняно з навчальним набором та набором тестування, розділеними випадковим чином. Також помічені хороші показники для оригінального набору даних порівняно з нейронними мережами, рис. 4.2.

### 4.3 Позиціонування в умовах LOS

Позиціонування в умовах прямої видимості (Line of Sight, LOS) є одним із методів визначення місцезнаходження абонентського обладнання або пристрою в бездротовій мережі. Умови LOS створюють ідеальне середовище для позиціонування, оскільки вони передбачають прямий шлях між передавачем і приймачем без значних перешкод і відбиттів сигналу.

Основні аспекти позиціонування в умовах LOS:

1. Принцип роботи: Умови LOS означають, що сигнал може проходити безпосередньо від передавача до приймача без перешкод. В цьому випадку вимірюються час затримки (Time of Flight, TOF) або інші характеристики сигналу, щоб визначити відстань між передавачем і приймачем.

2. Геометрична точність: Умови LOS надають високу геометричну точність

позиціонування. Оскільки сигнал подорожує прямим шляхом, вимірювана відстань є точною і дозволяє визначити місцезнаходження з великою точністю.

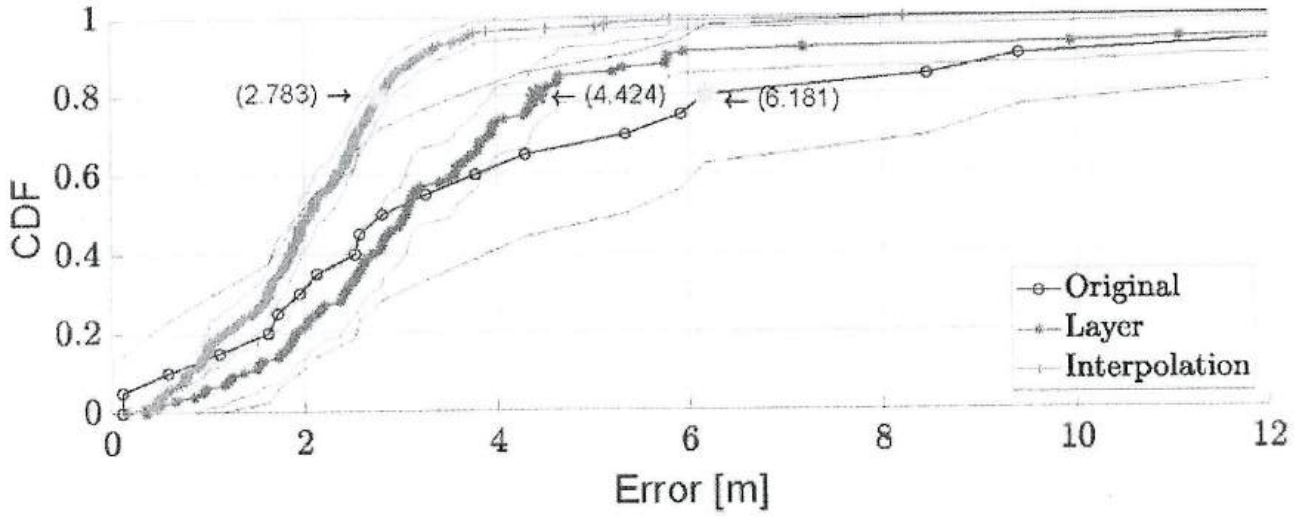
3. Використання тригонометричних методів: Для позиціонування в умовах LOS використовують тригонометричні методи, такі як трилатерація або метод суміжних кутів (Angulation). Ці методи базуються на вимірюванні кутів або відстаней між приймачем і декількома відомими передавачами.

4. Системи з однією антеною: В умовах LOS можливе використання систем з однією антеною для позиціонування. Простота системи з однією антеною спрощує розгортання та використання.

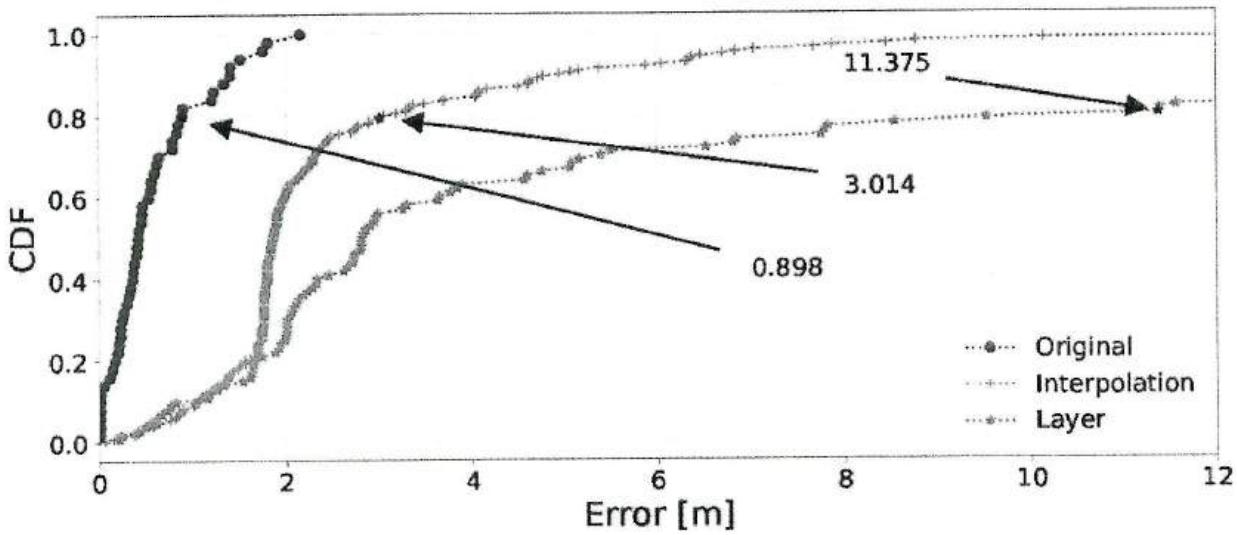
5. Застосування в геодезії та навігації: Умови LOS використовуються в геодезії, навігації, військовому дослідженні та інших галузях, де вимагається висока точність позиціонування.

Позиціонування в LOSi проводить LOS за допомогою аналогічних алгоритмів, як у NLOS. На рис. 4.7 навчання та оцінювання розділені послідовно, а на рис. 4.8 випадковим чином. На рис. 4.7a та рис. 4.8a показані результати позиціонування за допомогою нейронних мереж у LOSi. Нейронні мережі мають ті самі параметри, що і ті, що використовуються для позиціонування в NLOS. Ті ж параметри, що й були використовани для позиціонування в NLOS, використовувались як для нейронних мереж, так і для випадкових лісів. Результати навчання випадкового лісу за даними LOS можна спостерігати на рисунках 4.7b та на рис. 4.8b.

Дані про навчання та оцінювання відокремлюються послідовно від наборів даних в умовах LOS. Варто зауважити, що кращі показники для умов LOSa порівняно з умовами NLOS.




(а) Нейронна мережа



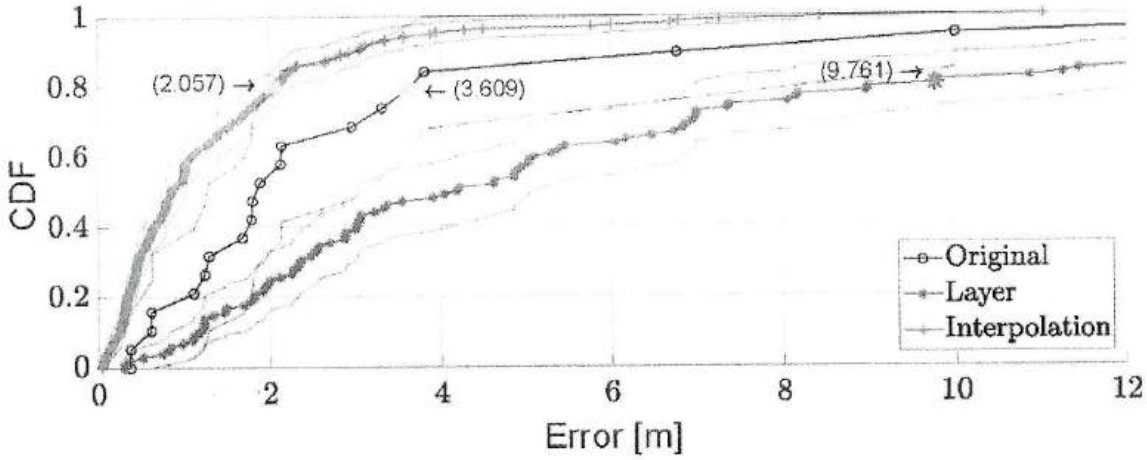

(б) Random forest

Рис. 4.7. Порівняння між різними алгоритмами позиціонування для даних LOS.

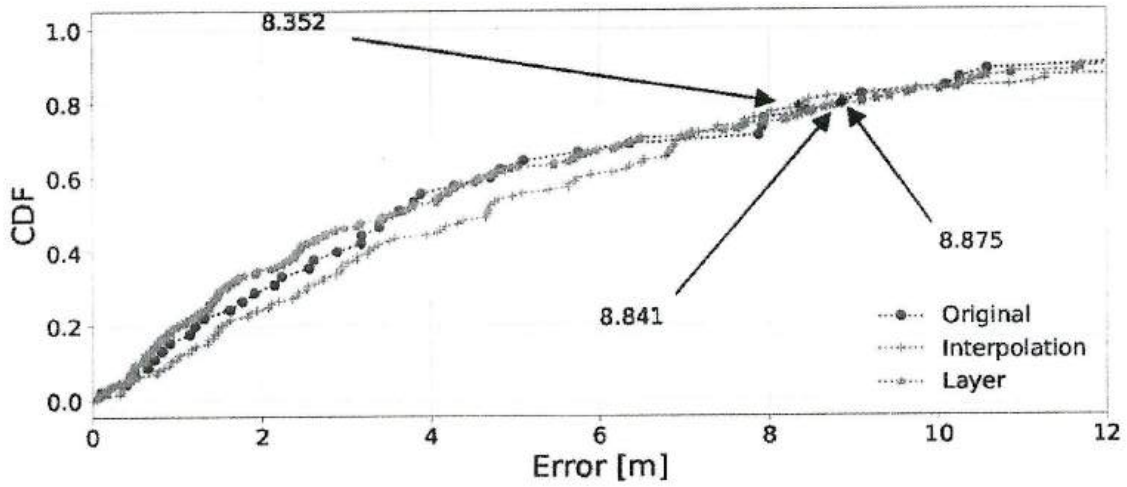
Дані про навчання та дані про оцінку випадково відокремлюються від наборів даних в умовах LOS. Варто помітити, що кращі показники для умов LOS порівняно з умовами NLOS.

Порівнюючи рисунок 4.5 та рисунок 4.6 з результатами на рис. 4.8, можна легко помітити, що продуктивність для нейронних мереж в LOS набагато краща, якщо в HLOS спрацьовує ефективність, навіть така ж хороша або краща, ніж невизначеність для GPS, а для випадкових лісів – виступи схожі між LOS та

NLOS. Це може бути наслідком того, що нейронні мережі будують моделі для даних, що простіше зробити в умовах LOS, тоді як random forest класифікується шляхом узгодження даних із подібними ознаками.




(а) Нейронна мережа




(б) Random forest

Рис. 4.8. Порівняння різних алгоритмів позиціонування на даних LOS.

Загалом, умови LOS створюють ідеальні умови для точного позиціонування, і цей підхід широко використовується у різних додатках, де важлива точність і надійність визначення місцезнаходження.

### 4.4 Виявлення в умовах NLOS

На рис. 4.9 є цифра з GLRT з використанням різниці BRSRP і DoD. Як виявлення NP детектора, так і GLRT дали подібні результати. Розраховувалася ймовірність виявлення PD та ймовірність помилкової сигналізації PFA. Використовуючи сигнал, що складається з різниці BRSRP PD = 88%, і з використанням різниці в DoD PD = 76%. Для обох сигналів PFA = 5%.

Варто зазначити, що при виявленні за допомогою DoD відсутні великі частини наприкінці шляху 3, у північно-західному куті. Це може бути пов'язано з тим, що висота блокуючої будівлі невелика, тому сигнал може майже проходити над будівлею або через неї, що призводить до того, що значна частина простору не буде виявлена (визначена) як NLOS.

На рис. 4.10 показано детектор, застосований до шляху 1. Тут показані обідва сигнали xi [n] і X2 [n], а також детектор NP і GLRT для обох сигналів.

Видно, що два сигнали схожі на те, що детектор NP та GLRT дають однакові результати виявлення.

### 4.5 Фільтр Кальмана

На рис. 4.11 показаний результат використання фільтра Кальмана на виході алгоритмів машинного навчання. CDF відфільтрованих і нефільтрованих сигналів показані з довірчим інтервалом 95%, розрахованим за формулою Greenwood. Налаштування фільтра та модель простору стану описані в Розділі 3.5.

На рис. 4.11a виходи з нейронних мереж фільтрують за допомогою фільтра Калмана, а на рис. 4.11b фільтрують вихід з випадкового лісу. Набір тестових даних для алгоритмів машинного навчання – це послідовне розділення набору для навчання та тестування вихідного набору даних. З малюнка 4.11 зрозуміло, що фільтрація виходу з нейронних мереж підвищує ефективність позиціонування,

тоді як для випадкових лісів не спостерігається значного покращення. Причину цього можна побачити, вивчаючи ефект фільтрації та вихід з алгоритмів машинного навчання, показаний на рис. 4.12.

Фільтрується лише вихід із алгоритмів машинного навчання, оцінених на тестовому наборі з оригінального набору даних, послідовно відокремлених. Це пов'язано з тим, що для фільтрації потрібно мати кілька послідовних зразків часу для відстеження фільтра.

На рис. 4.12 $y_{true}$ – це справжня відстань від TP, y – вихід від алгоритмів машинного навчання, xf відфільтрований сигнал і xp прогнозований сигнал від фільтра Калмана. Зрозуміло, що оцінка положення за допомогою випадкових лісів упереджена, тоді як оцінка в нейронних мережах такою не є. Державно-просторова модель, яка використовується для фільтра Калмана, не може забезпечити зміщення, отже, не покращує точність визначення положення.

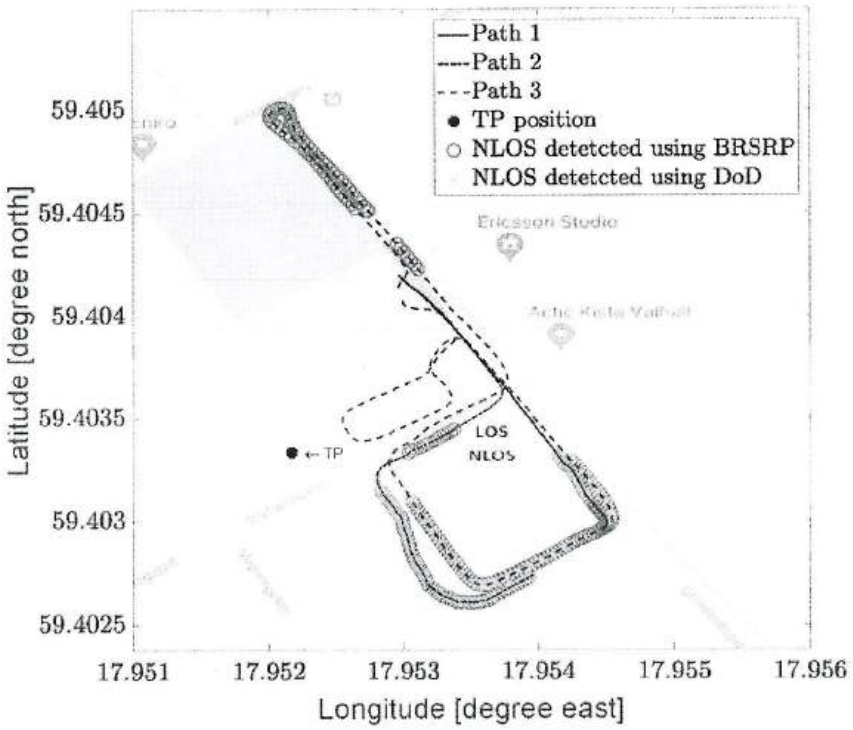



Рис. 4.9. Виявлення NLOS за допомогою обох сигналів, що складаються з різниці BRSRP і DoD. Жовті області є справжньою класифікацією умов LOS для даної ТП. Це визначається в результаты вивченням карти поєднаним зі знаннями про навколишнє середовище.

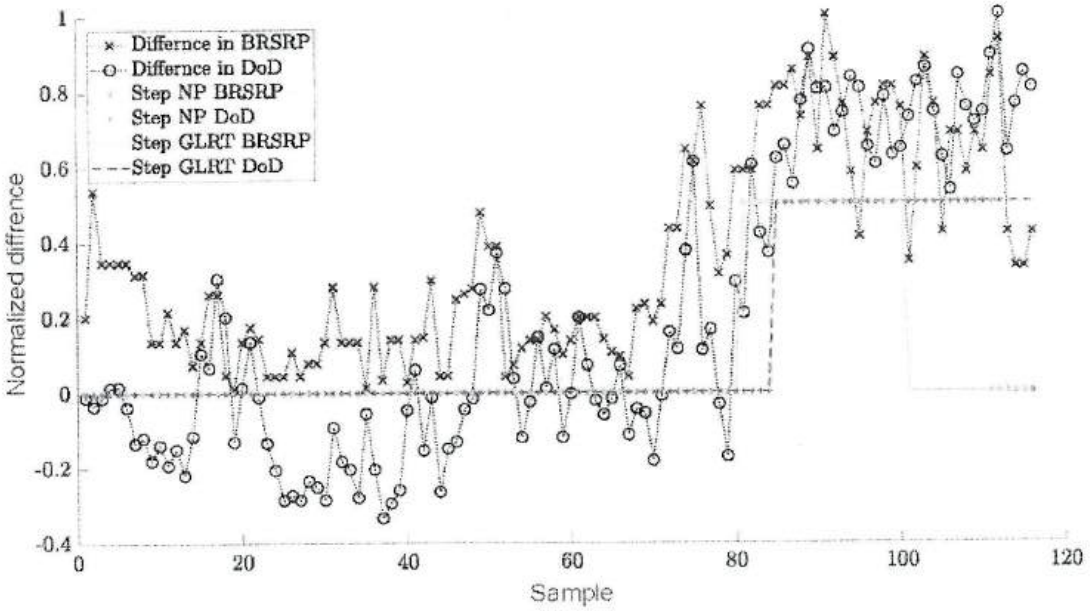
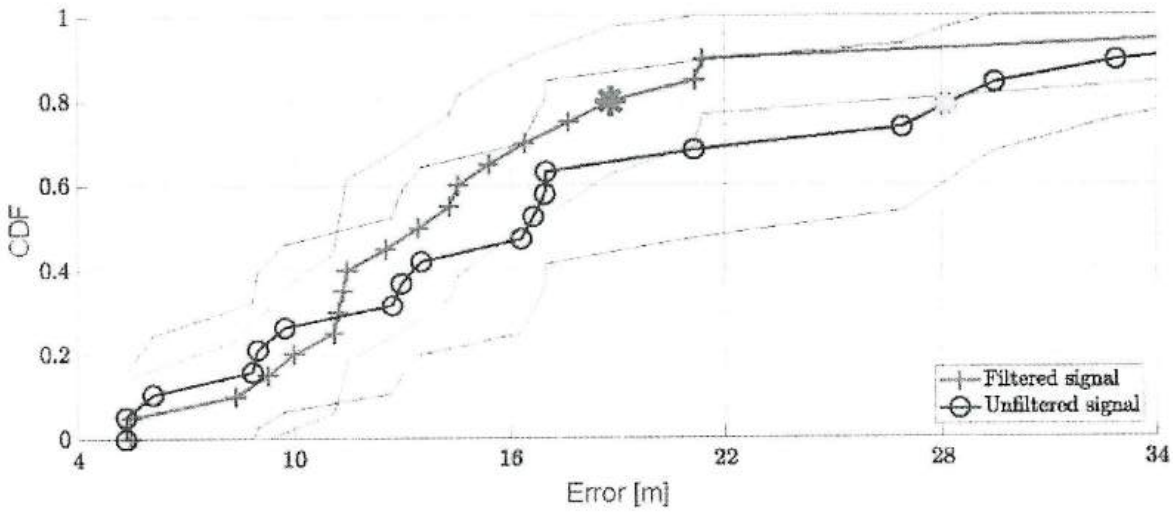
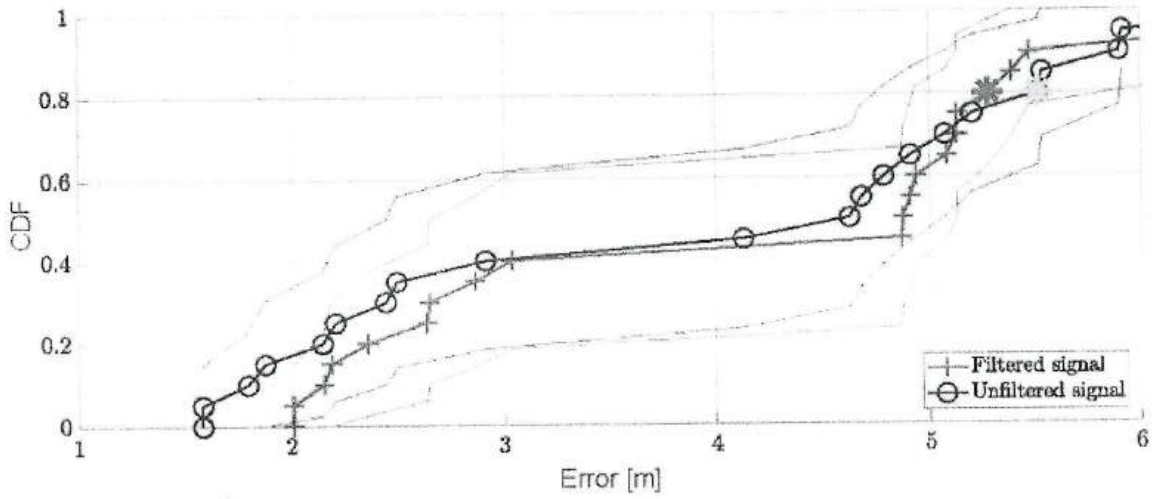



Рис. 4.10. Сигнали, що використовуються для виявлення NLOS по шляху 1

Ефект фільтрації алгоритмів машинного навчання за допомогою фільтра Калмана. Це вертикальна відстань від TP до UE на осі y. Від'ємне значення вказує, що UE знаходиться на південь від TP, а додатне значення – що UE знаходиться на північ від TP.




(а) Фільтр Кальмана, застосований до виходу з нейронних мереж




(b) Фільтр Кальмана, застосований до виходу з random forest

Рис. 4.11. Результати фільтрації Кальмана на виході алгоритмів машинного навчання

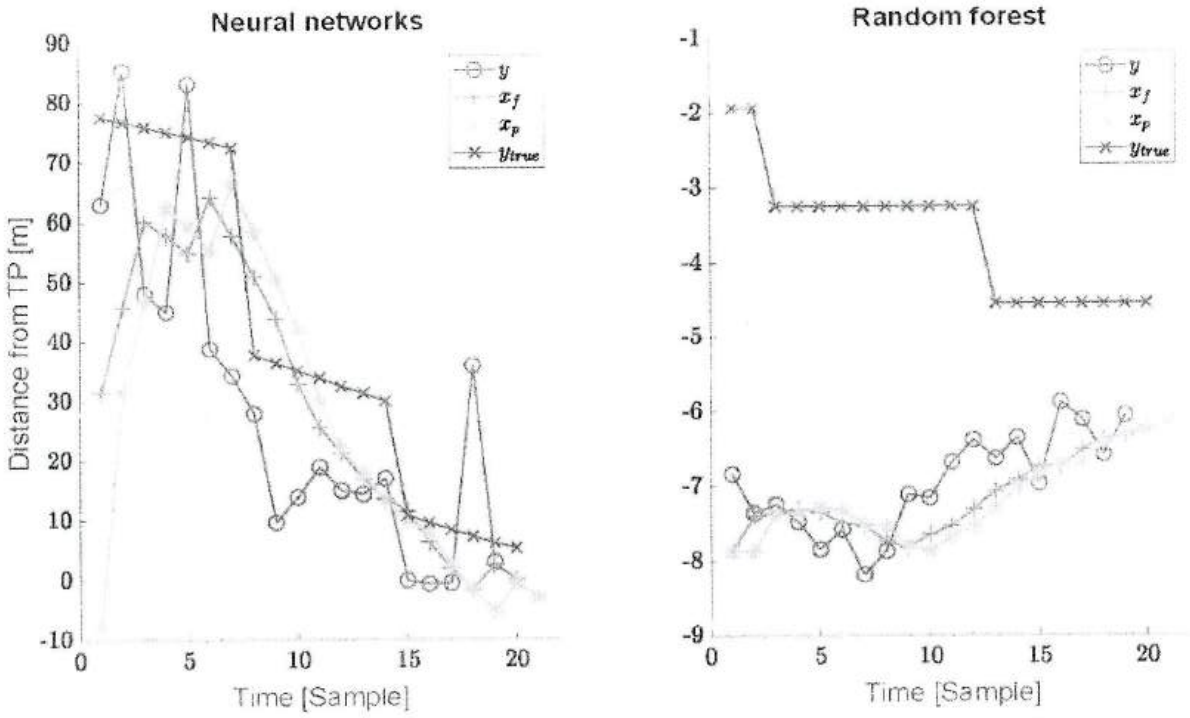



Рис. 4.12. Ефект фільтрації алгоритмів машинного навчання за допомогою фільтра Калмана

## 4.6 Обговорення результатів

Далі резюмуємо питання, що розглядалося, отримані результати у 4 розділі, завершумо оглядом мети та постановкою проблеми та пропонуємо напрямки подальшої роботи.

***Нейронні мережі***

З рисунка 4.5 видно, що використання інтерполяції даних для навчання дає найкращі результати; це відбувається з точністю позиціонування нижче 10 м. Цей результат є випадковим розділенням начального та навчального набору (див. рис. 4.3). Порівняння результатів, отриманих з оригінальним набором даних та інтерпольованим набором даних, показує, що випадкове розділення набору тестових даних дає кращі показники порівняно з оцінюванням на тестуванні та навчальним набором, розділеними послідовно. Зважаючи на це, тільки використання набору даних з інтерпольованими даними має кращу ефективність у точності позиціонування порівняно з тестами, оціненими на вихідному наборі даних. Також варто ще раз зазначити, що невизначеність GPS, яка в кращому випадку становить близько декількох метрів, а роздільна здатність сітки зразків становить 0,19 м.

Вивчення розриву в точності між оцінкою даних про навчання та даними тестування на рис. 4.2-4.4 (різниця між зеленими та суцільними лініями), висновків щодо оптимальної продуктивності та якщо нейронні мережі є надмірно обладнаними чи ні, можна зробити. Невеликий розрив вказує на те, що мережа наближається до оптимальної продуктивності та не є надмірно обладнаною. При усередненні оцінки позиції понад 100 у нейронних мережах розрив між тестовою та навчальною ефективністю невеликий. Це і тоді, коли набір тестування та навчання розділені випадковим чином і послідовно. Він дає уявлення, що для випадково розділених даних нейронна мережа узагальнює оцінку положення.

Ще одне зауваження полягає в тому, що якщо нейронні мережі матимуть

більше даних, що використовуються для тренувань, продуктивність може бути значно покращена. Це випливає з того, що оцінка на сукупності даних, що містить інтерполяцію даних, дає таку хорошу ефективність для тестових даних, розділених випадковим чином. Набір даних з інтерполяцією у поєднанні з випадковим поділом повторює сценарій, у якому є більший набір даних.

***Random forest***

З рисунка 4.6 видно, що вихідні дані дають найкращі показники у випадку використовування алгоритму випадкових лісів. Ефективність, оцінена на оригінальному наборі даних, значно краща для випадкового лісу порівняно з нейронною мережею. В цілому random forest дає набагато кращі показники, ніж нейронні мережі. На додаток до цього тренування випадкового лісу швидше, ніж тренування нейронної мережі. Тренування випадкового лісу займає пару годин, а нейронної мереж – займає дні.

Деякі цікаві результати можна побачити в підрозділі 4.2, можна побачити, що для всіх наборів даних одна особливість виділяється за своєю важливістю. Різниця в BRSRP між найсильнішим і деяким послідовним променем антени є найбільш важливою особливістю для всіх наборів даних, крім набору даних, що складається з інтерпольованих точок даних. Можна також побачити, що майже всі функції використовуються однаково, що вказує на те, що всі вони корисні для алгоритмів машинного навчання.

***Виявлення в умовах NLOS***

Алгоритми виявлення узгоджуються з тим, де очікується NLOS, що можна вирішити, вивчаючи карту та оточення. Це свідчить про хороший вибір сигналу, який слід використовувати для виявлення. На рис. 4.10 видно, що детектор NP і GLRT дають однакове виявлення, що вказує на те, що знання амплітуди є зайвим.

Ймовірність виявлення (PD) висока, тоді як ймовірність помилкової тривоги (PFA) низька, що вказує на те, що алгоритми виявлення працюють добре. Зважаючи на це, PD та PFA далеко не такі хороші, як теоретичні значення,

наведені у підрозділі 3.4. Це вказує на те, що модель не є ідеальною. Слід також зазначити, що справжня класифікація робиться дуже інтуїтивно, базується більше на вивченні карті та навколишнього середовища, саме так визначаються будівлі що блокують сигнал. Це призводить до того, що справжня класифікація є дуже невизначеною.

***Фільтр Кальмана***

З рисунка 4.11a видно, що використання фільтра Калмана в поєднанні з алгоритмом машинного навчання може підвищити точність позиціонування. Тим не менш, як показано на рис. 4.11b, це не завжди може бути так. У цьому випадку це пов'язано з ухилом у виході випадкового лісу, який не враховується державно-просторовою моделлю. Збільшення розміру набору даних, ймовірно, вирішить проблему з ухилом в оцінці. Це, оскільки упередження відбувається від відсутніх точок даних у навчальному наборі для алгоритмів машинного навчання. Якщо немає прямої відповідності між балами в наборі даних про навчання та справжнім цільовим значенням, може статися, що всі оцінки мають упередження, це швидше характерно для невеликого набору з навчанням.

# ВИСНОВКИ

Повертаючись до постановки проблеми в розділі 1, виявлено три питання, на які повинна була відповісти ця робота. Чи можливо використовувати алгоритми машинного навчання для позиціонування у міських каньйонах? Чи можна розрізнити NLOS та LOS? І чи може фільтрація підвищити продуктивність алгоритмів машинного навчання?

Результати в розділі 4 показують, що алгоритми машинного навчання є чудовим інструментом для позиціонування UE в стільниковій мережі, розташованій в міському каньйоні з умовами NLOS. Особливо алгоритм random forest отримав чудові результати, незважаючи на обмежений набір даних. Оскільки точність позиціонування random forest у міських каньйонах з умовами NLOS має аналогічну точність, як GPS-позиціонування під відкритим небом, яка становить кілька метрів, для покращення позиціонування, що проводиться random forest, можуть знадобитися деякі нові вимірювання.

Багато часу було витрачено на генерування більшої кількості даних для алгоритмів машинного навчання, деякі більш успішні, ніж інші. Інтерполяція, що використовує різницю частоти дискретизації сигналу для створення нових наборів даних, показує перспективні результати. Для підтвердження того, що новий набір даних копіює реальні дані, слід здійснити порівняння двох наборів даних однакового розміру, одного з реальними даними та одного з новим набором даних.

Одним із висновків, що можна зробити з генерування нових наборів даних, є те, що знання конструкції антени є корисним для позиціонування за допомогою алгоритмів машинного навчання. Цей висновок робиться виходячи з того, наскільки гірші показники роботи при розгляді поділу вертикального променя шару в антені.

Умови LOS та NLOS можна розрізнити за допомогою функцій, які пов'язані

з DoD і BRSRP у контексті безпроводового зв'язку і позиціонування. DoD вказує на напрям, в якому сигнал відправляється від пристрою до базової станції (або іншого приймача). BRSRP вказує на силу сигналу, яку базова станція отримує від певного пристрою або приймача. Умова LOS характеризується тим, що BRSRP має стабільну і високу силу сигналу, оскільки сигнал вільно доходить до базової станції без значних втрат. Умова NLOS може вказувати на значні втрати сигналу та зменшену силу сигналу, оскільки сигнал повинен пройти через перешкоди.

Якщо DoD має малий відхил від напряму до базової станції, і BRSRP висока та стабільна, то це може свідчити про умови LOS. У випадку великого відхилення DoD та низької BRSRP, може виникати підозра на умови NLOS.

Однак слід зауважити, що для точного розрізнення між умовами LOS і NLOS може бути потрібний більший обсяг інформації та складні алгоритми обробки сигналу. Можна також сказати, що важливо відокремити позиціонування в LOS та NLOS. Це показано з кращою ефективністю позиціонування нейронних мереж у LOS порівняно з NLOS.

Виявлення NLOS працює добре. Особливо з використанням інформації про BRSRP, як з детектором NP, так і з GLRT. PD становить майже 90%, тоді як PFA складає лише пару відсотків. Недолік цього результату полягає в тому, що справжні значення є отриманими вручну, спостерігаючи за міською місцевістю. Для розробки нових алгоритмів виявлення NLOS виходи двох досліджуваних детекторів можуть бути використані як справжнє виявлення.

Для фільтрації результатів за допомогою машинного способу навчання можна підвищити точність позиціонування. Це справедливо, якщо модель простору станів моделюється правильно. Тут для нейронних мереж припущення про те, що системний шум поширювався Гауссовським, не маючи розумних швів розумних після вивчення вихідних даних, отже, фільтр Калмана працює добре. У той час як для random forest, ухил в оцінці означав, що фільтр Калмана не покращив продуктивність позиціонування.

# ПЕРЕЛІК ПОСИЛАНЬ

1. Roos T., Myllymaki P., Tirri H. A statistical modeling approach to location estimation. IEEE Transactions on Mobile Computing, 2020. August, 99(1) P. 59–69.

2. Beam-based vehicular position estimation in 5G radio Access / Zhang X. та ін. Special issue in "Localization in Current and Emerging Networks", WCNC, 2018.

3. J. Font-Segura and X. Wang, "GLRT-based spectrum sensing for cognitive radio with prior information", IEEE Trans. Commun., vol. 58, no. 7, pp. 2137-2146, Jul. 2016.

4. Zhao Y. Position Estimation in Uncertain Radio Environments and Trajectory Learning.: Linköping studies in science and technology. Thesis No. 1172, 2017.

5. Ash J.N., Potter Lee C. Sensor network localization via received signal strength measurements with Directional Antennas. URL : https://www.researchgate.net/profile/Lee_Potter/publication/228708555_Sensor_network_localization_via_received_signal_strength_measurements_with_directional_antennas/links/0a85e537a538ddb6c5000000/Sensor-network-localization-via-received-signal-strength-measurements-with-directional-antennas.pdf /. (дата звернення: 13.10.2023).

6. Deep Convolutional Neural Networks for Massive MIMO Fingerprint-Based Positioning / Vieira J. та ін. URL: https://arxiv.org/pdf/1708.06235.pdf . (дата звернення: 18.10.2023).

7. Borris J., Hatrack P., Manclayam N. B. Decision theoretic framework for NLOS identification. URL: https://www.vtvt.ece.vt.edu/research/references /uwb/ranging_mobile_location/decision_theoretic_framework.pdf. (дата звернення: 02.11.2023).

8. Particle filters for positioning, navigation, and tracking / F. Gustafsson. URL; https://www.researchgate.net/publication/3318231_Particle_Filters_for_Positioning_Navigation_and_Tracking. (дата звернення: 17.11.2023).

9. Заїка В. Ф., Варфоломеєва О. Г., Домрачева К. О., Гринкевич Г. О. Телекомунікаційні системи та мережі наступного покоління: навчальний посібник. Київ: ДУТ, 2019. 315 с, URL; http://www.dut.edu.ua/uploads/l_1762_12226724.pdf. (дата звернення: 13.11.2023).

10. Gustafsson F., Gunnarsson F. Mobile positioning using wireless networks: possibilities and fundamental limitations based on available wireless network measurements. URL; https://www.researchgate.net/publication/3321638_Mobile_positioning_using_wireless_networks_Possibilities_and_fundamental_limitations_based_on_available_wireless_network_measurements. (дата звернення: 15.11.2023).

11. Полоневич О. В., Косенко В. Р., Сторчак К. П., Ткаленко О.М. Інформаційні мережі. Київ: ДУТ, 2019. 94 с. URL; http://www.dut.edu.ua/ua/lib/1/category/96/view/1175 (дата звернення: 17.11.2023).

12. Bishop C. M. Pattern Recognition and Machine Learning. Springer, 2016. URL; https://www.springer.com/gp/book/9780387310732. (дата звернення: 17.11.2023).

13. Livingstone D. J. Artificial Neural Networks Method and Applications. URL;https://books.google.com.ua/books?id=hfp6nVbCV8C&pg=PA306&lpg=PA306&dq=D.+J.+Livingstone.+Artificial+Neural+Networks+Method+and+Applications.&source=bl&ots=xZHqYvYV&sig=ACfU3U1S0ThrOGtHU8z2RC4ztRq13pPyyQ&hl=com&sa=X&ved=2ahUKEwiU98miopzpAhWKyKYKHRKgA2EQ6AEwA3oECAUQAQ#v=onepage&q=D.%20J.%20Livingstone.%20Artificial%20Neural%20Networks%20Method%20and%20Applications.&f=false . (дата звернення: 19.11.2023).

14. Воробієнко П. П., Нікітюк Л. А., Резніченко П. І. Телекомунікаційні та інформаційні мережі. Київ: САМІТ-Книга, 2010. 708 с.

15. Breiman L. Bagging predictors. URL; https://link.springer.com/content/pdf/10.1023/A:1018054314350.pdf . (дата звернення: 19.11.2023).

16. Liaw A., Wiener M. Classification and regression by random forest. URL; https://www.researchgate.net/publication/228451484_Classification_and_Regression_by_RandomForest. (дата звернення: 19.11.2023).

17. L. Breiman. Random forests. URL; https://link.springer.com/content/pdf/10.1023/A:1010933404324.pdf. (дата звернення: 20.11.2023).

18. Kay S. M. Fundamentals of Statistical Signal Processing Volume II: Detection Theory. Prentice Hall PTR, 1998. 672 c,

19. Gustafsson F., Ljung L., Millnert M. Signal Processing; Student litteratur AB, 2013. 406 p.

20. GPS Accuracy. Technical report, National Coordination Office for Space-Based Positioning, Navigation, and Timing, 2017. Information collected 5 April 2018.

21. 3GPP TS 38.215 V15.1.0 (2018-03) Technical Specification 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; NR; Physical layer measurements (Release 15). URL; https://panel.castle.cloud/view_spec/38215-f10/. (дата звернення: 20.11.2023).

22. LTE Technical Specifications and Technical Reports for Evolved Packet System (EPS) based 3GPP system (3GPP TS 21.201 version 13.0.0 Release 13). URL; https://www.etsi.org/deliver/etsi_ts/121200_121299/121201/13.00.00_60/ts_121201v130000p.pdf. (дата звернення: 22.11.2023).

23. Основні положення теорії штучних нейронних мереж URL: https://dl.nure.ua/pluginfile.php/634/mod_resource/content/2/content/content1.html (дата звернення: 22.11.2023).

# ДЕМОНСТРАЦІЙНІ МАТЕРІАЛИ

**ДЕРЖАВНИЙ УНІВЕРСИТЕТ**
**ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ**

Презентація по магістерській роботі на тему:

# Методи позиціонування по 5G-базових станціях високої точності

Виконала
Юрчук Евеліна Ігорівна

# МЕТА РОБОТИ

- ***Об'єкт дослідження*** – процес передачі даних 5G-мережах.

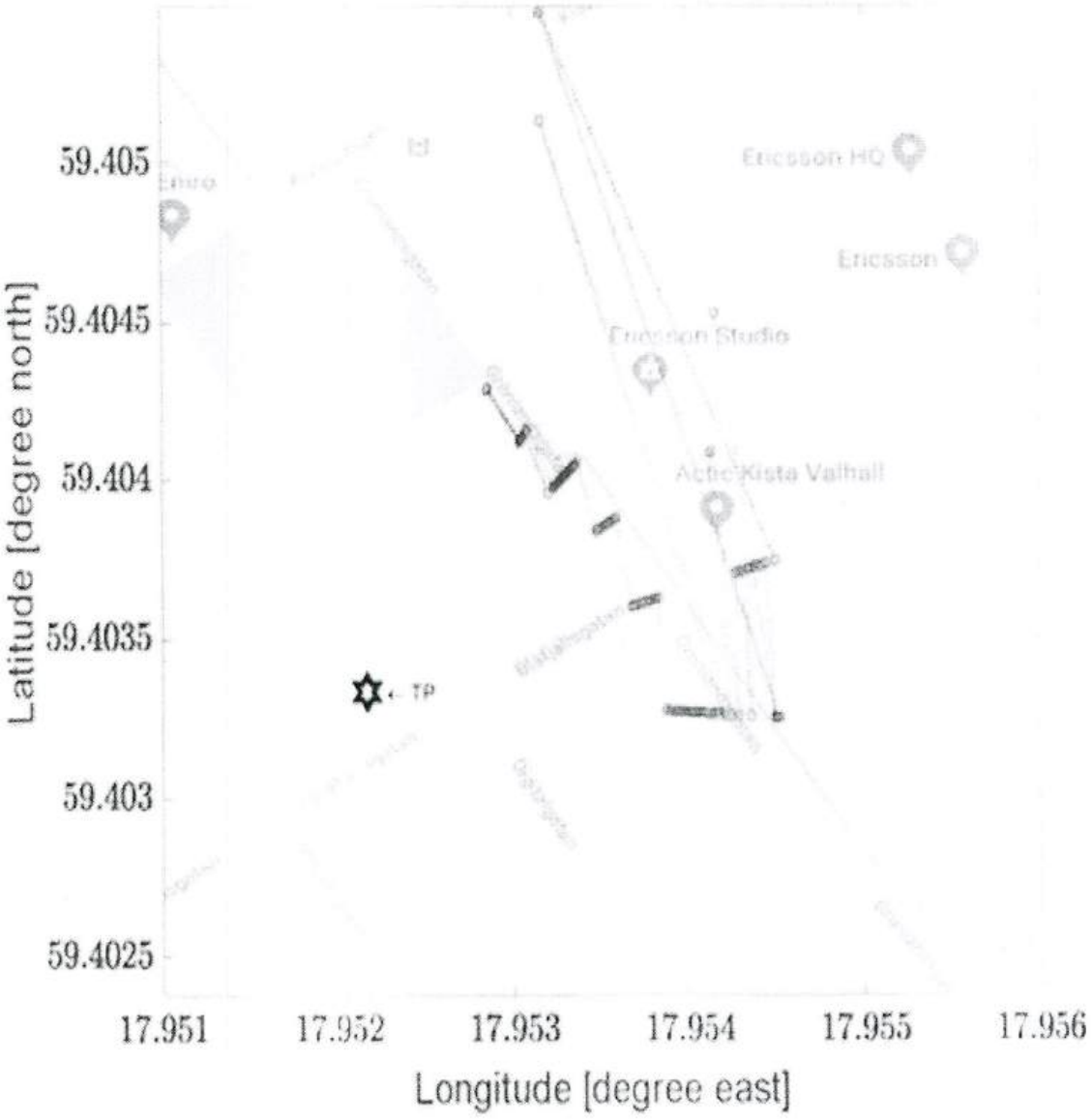
- ***Предмет дослідження*** - метод машинного навчання в 5G-мережах для позиціонування по базовим станціям.

- ***Мета роботи*** – дослідження ефективності застосування методів машинного навчання в 5G-мережах для позиціонування по базовим станціям.

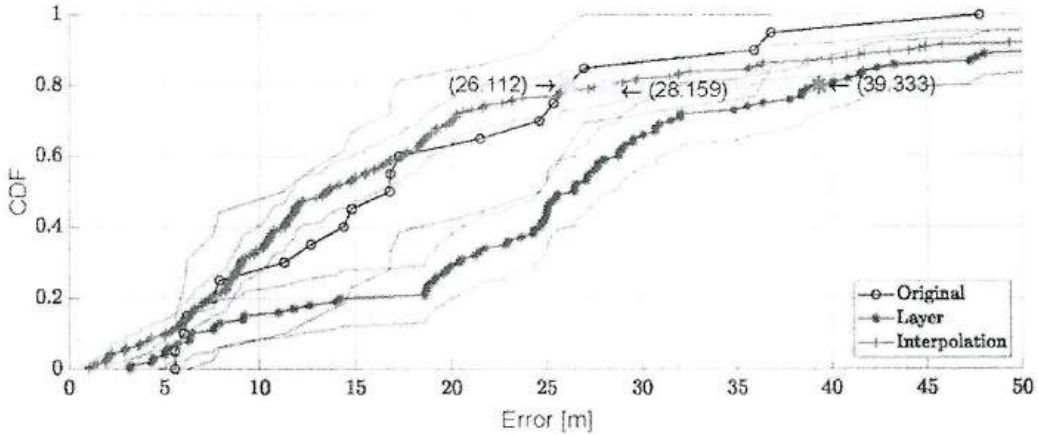
# Схематична картина позиціонування у міському районі



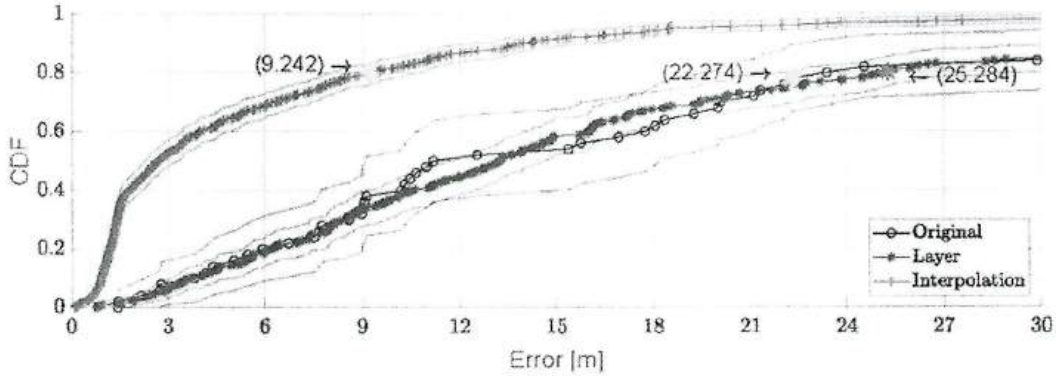
# Позиціонування



# Нейронні мережі

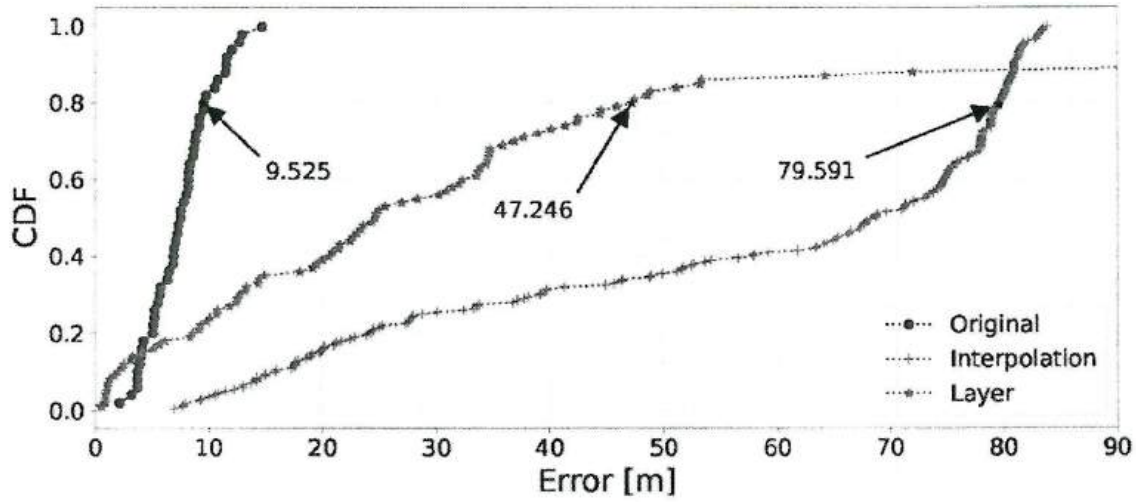



(а) Ефективність, оцінювана на тестових даних, послідовно відокремлених від набору даних.

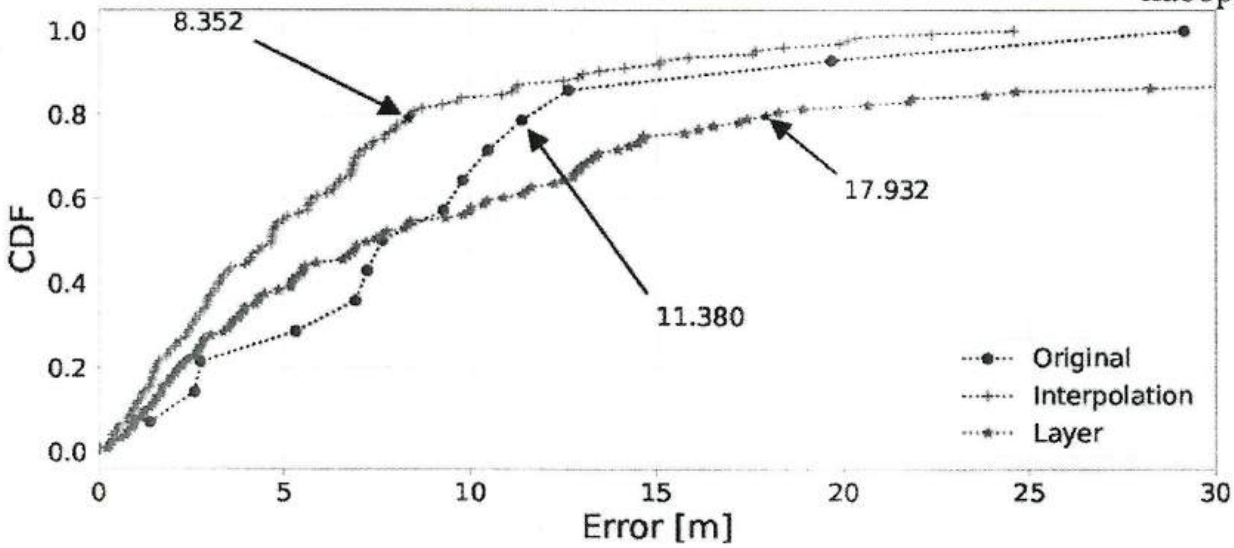



(б) Ефективність оцінюється на даних тесту, відокремлених випадковим чином від набору даних.

# Random forest

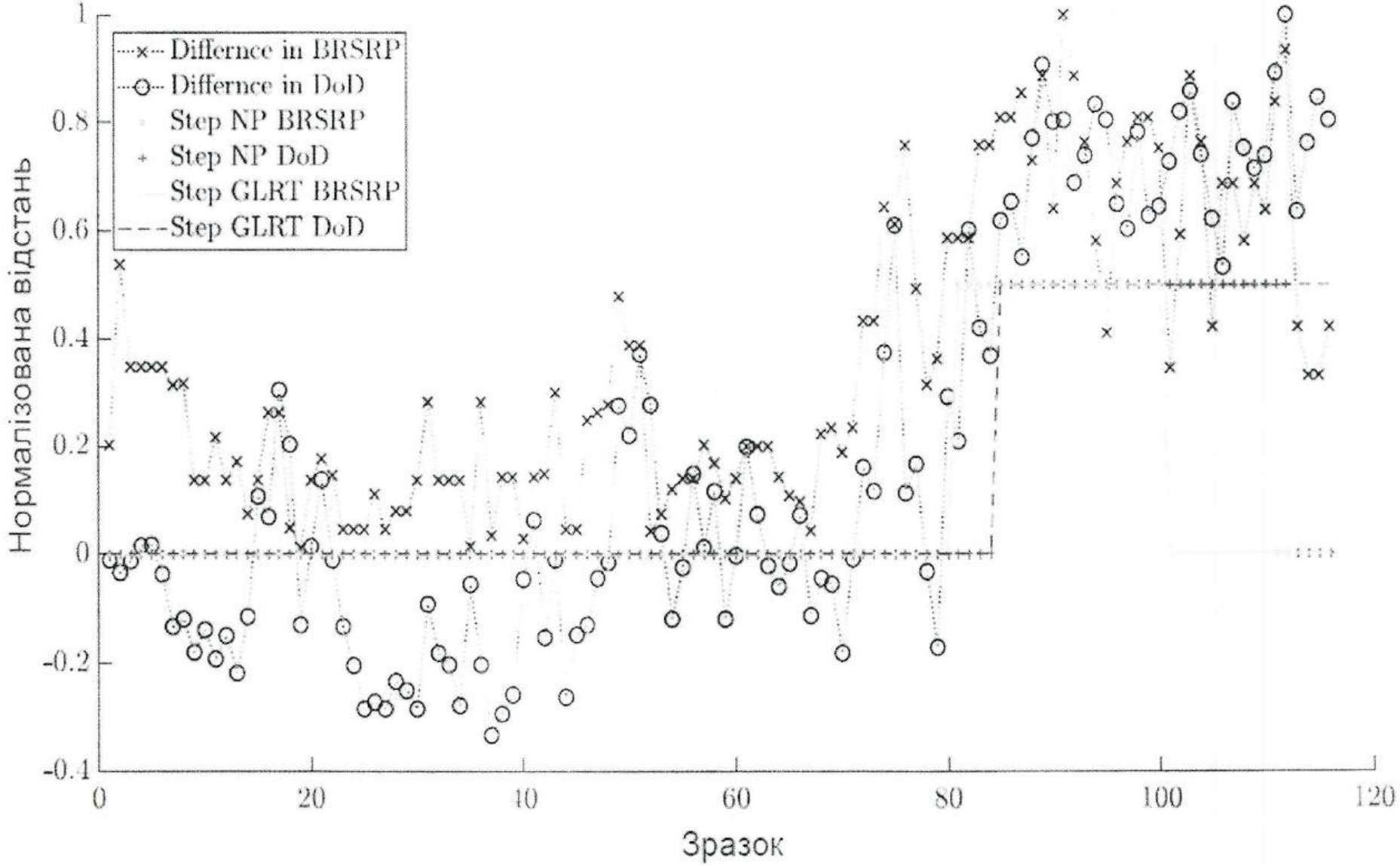



(а) Ефективність, оцінювана на тестових даних, послідовно відокремлених від набору даних.

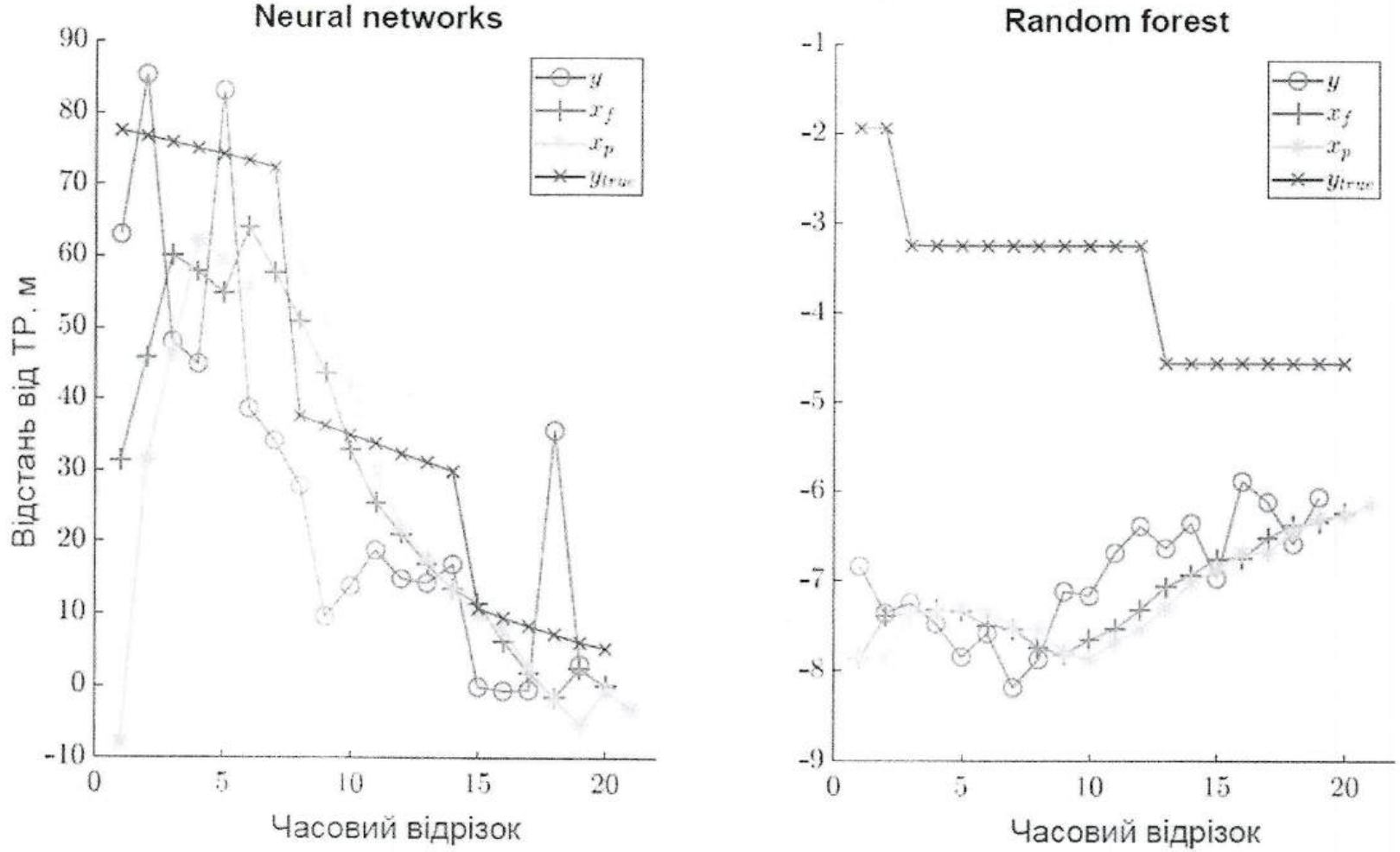



(б) Ефективність оцінюється на даних тесту, відокремлених випадковим чином від набору даних.

# Виявлення в умовах NLOS



# Ефект фільтрації алгоритмів машинного навчання за допомогою фільтра Калмана




**Y true** – це справжня відстань від ТР, **y** – вихід від алгоритмів машинного навчання, **xf** відфільтрований сигнал і **xp** прогнозований сигнал від фільтра Калмана. Зрозуміло, що оцінка положення за допомогою random forest упереджена, тоді як оцінка в нейронних мережах такою не є.

# Висновки

У магістерській роботі проведено широке дослідження методу машинного навчання в 5G-мережах для позиціонування по базовим станціям.

Результати роботи показують, що алгоритми машинного навчання є чудовим інструментом для позиціонування обладнання користувача в стільниковій мережі, розташованій у міській місцевості з умовами не прямої видимості.

Особливо алгоритм random forest отримав чудові результати, незважаючи на обмежений набір даних. Оскільки точність позиціонування random forest у міських каньйонах з умовами NLOS має аналогічну точність, як GPS-позиціонування під відкритим небом, яка становить кілька метрів, для покращення позиціонування, що проводиться random forest, можуть знадобитися деякі нові вимірювання.

# Апробація результатів магістерської роботи

- Основні результати магістерської роботи доповідалися на Всеукраїнській науково-практичній конференції «Telecommunication: problems and innovation», Київ: ДУІКТ, 20 грудня 2023 р.

- Основні наукові результати магістерської роботи опубліковано в науковому журналі “Наукові записки ДУІКТ, – 2023, – №2”.

**Протокол аналізу звіту подібності науковим керівником**

Заявляю, що я ознайомився (-лась) з Повним звітом подібності, який був згенерований Системою виявлення і запобігання плагіату щодо роботи:

**Автор:** Евеліна ЮРЧУК

**Назва:** Методи позиціонування по 5G-базових станціях високої точності

**Координатор:** Лариса ДАКОВА

**Підрозділ:** ННІТ

**Коефіцієнт подібності 1** :10.6

**Коефіцієнт подібності 2** :6.2

**Тривога:** 5

**Після аналізу Звіту подібності констатую наступне:**

☐ виявлені в роботі запозичення є сумлінними і не мають ознак плагіату. Тому робота визнається самостійною і допускається до захисту;

☐ виявлені в роботі запозичення не мають ознак плагіату, але їх надмірна кількість викликає сумніви щодо цінності роботи і самостійності її автора. Роботу направити на доопрацювання;

☐ виявлені в роботі запозичення є недобросовісними і мають ознаки плагіату або в ній містяться навмисні спотворення тексту, що вказують на спроби приховування недобросовісних запозичень. У зв'язку з чим, робота не допускається до захисту.

..... .....

*Дата* *Підпис Наукового керівника*