

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет комп'ютерної інженерії та управління
(повна назва)

Кафедра електронних обчислювальних машин
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА
Пояснювальна записка

Рівень вищої освіти другий (магістерський)

Метод виявлення БПЛА за допомогою
акустичних сигналів

(тема)

Виконав:

здобувач 2 року навчання,

групи СПм-23-4

Дмитро АЛЕЙНИК

(власне ім'я, прізвище)

Спеціальність

123 «Комп'ютерна інженерія»

(код і повна назва спеціальності)

Тип програми освітньо-наукова

(освітньо-професійна або освітньо-наукова)

Освітня програма

Системне програмування

(повна назва освітньої програми)

Керівник: доц. Віталій МАРТОВИЦЬКИЙ

(посада, власне ім'я, прізвище)

Допускається до захисту

Завідувач кафедри ЕОМ

(підпис)

Андрій КОВАЛЕНКО

(власне ім'я, прізвище)

2025 р.

Харківський національний університет радіоелектроніки

Факультет _____ комп'ютерної інженерії та управління _____

Кафедра _____ електронних обчислювальних машин _____

Рівень вищої освіти _____ другий (магістерський) _____

Спеціальність _____ 123 «Комп'ютерна інженерія» _____
(код і повна назва)

Тип програми _____ освітньо-наукова _____
(освітньо-професійна або освітньо-наукова)

Освітня програма _____ Системне програмування _____
(повна назва)

ЗАТВЕРДЖУЮ:

Зав. кафедри _____
(підпис)

“ _____ ” _____ 20__ р.

ЗАВДАННЯ

НА КВАЛІФІКАЦІЙНУ РОБОТУ

здобувачеві _____ Алейнику Дмитру Сергійовичу _____
(прізвище, ім'я, по батькові)

1. Тема роботи _____ Метод виявлення БПЛА за допомогою акустичних сигналів _____

затверджена наказом по університету від “ 21 ” квітня 2025 р. № 296 Ст

2. Термін подання здобувачем роботи до екзаменаційної комісії _____ 16 червня 2025 р.

3. Вхідні дані до роботи _____ Набір датасеті _____

4. Перелік питань, що потрібно опрацювати у роботі _____

Огляд сучасних методів виявлення БПЛА _____

Аналіз специфіки акустичних сигналів БПЛА _____

Формування датасету акустичних записів _____

Підготовка аудіоданих до навчання _____

Вибір і реалізація архітектур нейронних мереж _____

Навчання та валідація моделей _____

Оцінка ефективності розроблених моделей _____

5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій 17 слайдів

6. Консультанти розділів роботи (заповнюється за наявності консультантів згідно з наказом, зазначеним у п.1)

| Найменування розділу | Консультант (посада, прізвище, ім'я, по батькові) | Позначка консультанта про виконання розділу | |
|----------------------|--|---|------|
| | | підпис | дата |
| | | | |
| | | | |

КАЛЕНДАРНИЙ ПЛАН

| № | Назва етапів роботи | Строк / терміни виконання етапів роботи | Примітка |
|---|--|---|----------|
| 1 | Огляд методів виявлення безпілотних апаратів | 22.04.25-29.04.25 | |
| 2 | Вибір та обґрунтування методики дослідження | 30.04.25-05.05.25 | |
| 3 | Вибір інструментальних засобів | 06.05.25-09.05.25 | |
| 4 | Розробка моделей | 10.05.25-20.05.25 | |
| 5 | Проведення експериментів | 21.05.25-02.06.25 | |
| 6 | Оформлення матеріалів кваліфікаційної роботи | 03.06.25-05.06.25 | |
| 7 | Подання кваліфікаційної роботи керівникові та її попередній захист | 06.06.25-09.06.25 | |
| 8 | Подання кваліфікаційної роботи на рецензування | 10.06.25-12.06.25 | |
| | | | |

Дата видачі завдання “ 21 ” квітня 2025 р.

Здобувач

_____ (підпис)

Керівник роботи

_____ (підпис)

доц. Віталій МАРТОВИЦЬКИЙ

_____ (посада, власне ім'я, прізвище)

РЕФЕРАТ

Пояснювальна записка кваліфікаційної роботи: 54 с., 15 рис., 3 табл., 1 дод., 9 джерел.

БЕЗПЛОТНИЙ ЛІТАЛЬНИЙ АПАРАТ (БПЛА), АКУСТИЧНЕ РОЗПІЗНАВАННЯ, ГЛИБИННЕ НАВЧАННЯ, LSTM, BiLSTM, CNN, РЕАЛЬНИЙ ЧАС, АУДІОАНАЛІЗ, СИСТЕМА ВИЯВЛЕННЯ, НЕЙРОННІ МЕРЕЖІ.

Метою кваліфікаційної роботи є розробка системи виявлення безпілотних літальних апаратів (БПЛА) на основі аналізу їхніх акустичних сигналів із використанням методів глибинного навчання.

У ході виконання кваліфікаційної роботи було проведено дослідження акустичних характеристик звуків дронів з навантаженням та без нього, зібрано та оброблено відповідний аудіодатасет, а також реалізовано фільтраційний модуль для підготовки даних тривалістю одна секунда, що забезпечує роботу системи в режимі реального часу. Основну увагу приділено порівнянню ефективності різних типів рекурентних нейронних мереж (SimpleRNN, LSTM, BiLSTM, GRU), а також згорткової нейронної мережі (CNN). Моделі навчалися з поділом даних на тренувальну та валідаційну вибірки, а також були протестовані на незалежних тестових даних. Результати експериментів показали, що рекурентні архітектури LSTM та BiLSTM забезпечують найвищу точність класифікації у завданнях розпізнавання акустичних сигналів БПЛА. Розроблена система демонструє високу ефективність у виявленні дронів у режимі реального часу та може бути використана для охорони об'єктів, моніторингу повітряного простору та в інших прикладних задачах безпеки.

ABSTRACT

Master's thesis: 54 pages, 15 figures, 3 tables, 1 appendices, 9 sources.

UNMANNED AERIAL VEHICLE (UAV), DRONE DETECTION, ACOUSTIC RECOGNITION, DEEP LEARNING, LSTM, BILSTM, CNN, REAL-TIME SYSTEM, NEURAL NETWORKS, AUDIO SIGNAL ANALYSI.

The major goal of this thesis is to develop a drone detection system based on the analysis of their acoustic signals using deep learning methods. During the course of the work, the acoustic characteristics of drones both with and without payloads were studied, a relevant audio dataset was collected and processed, and a filtering module was implemented to prepare one-second-long audio segments to enable real-time operation of the system. The main focus was placed on comparing the performance of various types of recurrent neural networks (SimpleRNN, LSTM, BiLSTM, GRU) as well as a convolutional neural network (CNN). The models were trained with a split between training and validation datasets and tested on an independent test set. The experimental results showed that recurrent architectures such as LSTM and BiLSTM achieved the highest classification accuracy in the task of acoustic drone recognition. The developed system demonstrates high effectiveness in detecting drones in real-time and can be used for perimeter security, airspace monitoring, and other applied safety-related tasks.

ЗМІСТ

| | |
|---|----|
| СКОРОЧЕННЯ ТА УМОВНІ ПОЗНАКИ | 7 |
| ВСТУП | 8 |
| 1 ПІДХОДИ ДО ВИЯВЛЕННЯ БЕЗПЛОТНИХ ЛІТАЛЬНИХ АПАРАТІВ ЗА ДОПОМОГОЮ АКУСТИЧНИХ ДАНИХ | 9 |
| 1.1 Системи виявлення БПЛА | 9 |
| 1.2 Виявлення БПЛА на основі акустичних даних..... | 13 |
| 2 ОПИС АКУСТИЧНИХ ДАНИХ БПЛА..... | 18 |
| 3 МЕТОДИ ЦИФРОВОЇ ОБРОБКИ СИГНАЛУ | 22 |
| 3.1 Принцип представлення звукових даних | 22 |
| 3.2 Обробка акустичних сигналів у часовій області..... | 25 |
| 3.3 Короткочасне перетворення Фур'є (Short-Time Fourier Transform, STFT) | 27 |
| 4 ГЛИБИННІ МЕТОДИ НАВЧАННЯ ДЛЯ РОЗПІЗНАВАННЯ АКУСТИЧНИХ ДАНИХ БПЛА | 31 |
| 4.1 Згорткові нейронні мережі (CNN) у задачах розпізнавання звуку | 31 |
| 4.2 Рекурентні нейронні мережі (RNN) у задачах розпізнавання звуку | 34 |
| 4.3 Система розпізнавання та класифікації акустичних даних БПЛА в реальному часі | 37 |
| ВИСНОВКИ..... | 43 |
| ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ | 44 |
| ДОДАТОК А Графічний матеріал кваліфікаційної роботи..... | 45 |

СКОРОЧЕННЯ ТА УМОВНІ ПОЗНАКИ

БПЛА – безпілотний літальний апарат

CNN – згорткова нейронна мережа (англ., Convolutional Neural Network)

GRU – рекурентна нейронна мережа з контрольованими рекурентними одиницями (англ., Gated Recurrent Unit)

LSTM – довгострокова короткочасна пам'ять (англ., Long Short-Term Memory)

RNN – рекурентна нейронна мережа (англ., Recurrent Neural Network)

ВСТУП

Сучасні безпілотні літальні апарати (БПЛА) активно використовуються у військовій, цивільній та промисловій сферах. З одного боку, вони слугують інструментом спостереження, доставки, картографування та контролю, а з іншого – можуть бути джерелом загроз у разі несанкціонованого проникнення в охоронювані зони, використання для шпигунства або навіть диверсійних атак. У зв'язку з цим завдання оперативного виявлення і класифікації БПЛА набуває особливої актуальності для систем безпеки, протиповітряної оборони, охорони об'єктів критичної інфраструктури та громадського порядку.

Серед різних способів детекції БПЛА (радіолокаційні, оптичні, інфрачервоні тощо) акустичні методи вирізняються своєю пасивністю, низькою вартістю та здатністю працювати в умовах поганої видимості або за відсутності прямого візуального контакту. Акустичні сигнали, які генеруються двигунами та пропелерами дронів, мають характерні спектральні ознаки, що можуть бути використані для автоматизованого розпізнавання за допомогою методів цифрової обробки сигналів та штучного інтелекту.

Метою цієї кваліфікаційної роботи є розробка методу виявлення БПЛА на основі аналізу їх акустичних сигналів з використанням алгоритмів машинного навчання. У процесі дослідження передбачається зібрати акустичні дані різних типів БПЛА, провести їх обробку, виділити інформативні ознаки та побудувати модель класифікації, здатну в режимі реального часу розпізнавати наявність дронів у навколишньому середовищі.

Актуальність дослідження полягає у потребі розробки доступних, адаптивних та надійних засобів пасивного спостереження за повітряним простором, які можуть бути впроваджені як у військових, так і в цивільних системах моніторингу.

1 ПІДХОДИ ДО ВИЯВЛЕННЯ БЕЗПЛОТНИХ ЛІТАЛЬНИХ АПАРАТІВ ЗА ДОПОМОГОЮ АКУСТИЧНИХ ДАНИХ

1.1 Системи виявлення БПЛА

На сьогоднішній день безпілотні літальні апарати (БПЛА), також відомі як дрони, стають дедалі популярнішими серед споживачів, оскільки вони забезпечують зручне рішення для багатьох повсякденних потреб. Їх усе частіше використовують у таких сферах, як сільське господарство, фотографія, кіновиробництво, правоохоронна діяльність, логістика та транспорт. Дрони є надзвичайно корисними, оскільки здатні досягати найвіддаленіших місць без присутності пілота на борту. Завдяки цій можливості дрони стали глобальним явищем і все ширше застосовуються у різних видах діяльності.

Сучасні технології розвиваються швидше, ніж будь-коли раніше. Наш стиль життя постійно змінюється завдяки інноваціям, які роблять його зручнішим, швидшим і ефективнішим. У цьому контексті БПЛА докорінно змінили авіаційну галузь, зробивши її безпечнішою, доступнішою та продуктивнішою. З розвитком технологій дрони стають компактнішими та зручнішими у використанні – дехто з них може навіть поміститися на долоні. Через свої невеликі розміри їх активно використовують компанії для різних завдань. Деякі моделі застосовуються у пошуково-рятувальних операціях. Також дрони набирають популярності в комерційному секторі завдяки своїй простоті використання, безпечності експлуатації та доступній ціні.

У військовій сфері дрони цінуються за здатність довго перебувати в повітрі. Водночас вони дедалі частіше використовуються приватними особами та професіоналами як сучасне рішення для домашнього та бізнес-використання. Хоча спочатку дрони створювалися для військових цілей,

сьогодні вони активно застосовуються урядами, бізнесом і приватними користувачами по всьому світу.

Попри численні переваги, БПЛА становлять потенційну загрозу для безпеки повітряного простору. Вони можуть створювати небезпеку для людей, об'єктів, стратегічно важливих зон і будівель. Такі загрози можуть виникати як внаслідок недосвідченості оператора, так і внаслідок навмисних атак, що можуть завдати суттєвої шкоди.

Було зафіксовано чимало випадків порушення охоронюваних територій. Зокрема, у 2015 році дрон впав на газон Білого дому через дії нетверезого держслужбовця; у Канаді 2017 року сталася аварія легкого літака з дроном; у Лондоні в 2018 році тисячі пасажирів були змушені скасувати рейси через підозрілий дрон над аеропортом Гатвік. У 2019 році кілька аеропортів у США, Великій Британії, Ірландії та ОАЕ зазнали серйозних збоїв через зафіксовану присутність дронів.

Дрони також часто використовують у системах доставки, для терористичної діяльності, контрабанди, перекидання заборонених речей у зони з обмеженим доступом (наприклад, тюрми, митниці), а також для фото- та відеозйомки з незвичних ракурсів. Через це дедалі більшої актуальності набуває проблема розпізнавання БПЛА з додатковим навантаженням.

Отже, актуальним є не лише завдання виявлення самого дрона, а й визначення його стану – зокрема наявності додаткового вантажу. Для запобігання зазначеним ризикам необхідно розробити превентивні стратегії. З огляду на стрімкий розвиток ринку БПЛА, важливо, щоб правила й нормативи змінювалися так само швидко, як і технології. Попри те, що регулювання може знизити кількість випадкових порушень, воно не здатне повністю запобігти навмисним атакам.

Для ефективної протидії загрозам необхідно створювати технології, які забезпечують:

- виявлення, класифікацію та відстеження дронів;
- їх придушення;

- збір доказів про факт порушення.

Кожен з цих напрямів є об'ємним та потребує окремого наукового підходу. Водночас саме виявлення БПЛА є першим і найважливішим завданням, необхідним для реалізації інших заходів.

Окремо варто відзначити актуальність виявлення БПЛА у прикордонних регіонах у зв'язку з їхнім використанням у воєнно-політичних цілях.

На сьогодні існують чотири основні методи виявлення дронів:

- за допомогою радарів;
- акустичних сенсорів;
- візуальних сигналів;
- радіочастотного аналізу (RF).

За останнє десятиліття розпізнавання об'єктів стало активно застосовуватися в системах штучного інтелекту, зокрема на основі методів машинного та глибокого навчання. Усі вищезазначені методи також почали вивчатися з використанням цих підходів.

Дрони можуть відрізнятися за розміром, формою, призначенням і станом. Тому одночасно з виявленням часто постає завдання класифікації – визначення типу, навантаження або призначення пристрою.

Радари вважаються ефективними для виявлення об'єктів на великих відстанях, але малорозмірні БПЛА з низькою швидкістю та непередбачуваною траєкторією часто залишаються непоміченими, оскільки радар здебільшого орієнтований на високошвидкісні цілі з баллістичними траєкторіями. Крім того, класифікаційні можливості радарів є обмеженими, оскільки характеристики БПЛА часто схожі з характеристиками птахів, що ускладнює розрізнення.

На тлі цього зростає зацікавленість до акустичних методів виявлення. Використовуючи масиви мікрофонів або сенсорів, такі системи здатні розпізнавати унікальні звукові характеристики роторів БПЛА навіть за

поганої видимості. Акустичні підходи на основі машинного та глибокого навчання стали новим етапом у дослідженні виявлення дронів.

Основною перешкодою залишається недостатня кількість відкритих акустичних даних для різних моделей БПЛА на різній висоті, швидкості та у різних шумових умовах. Запропоновані рішення наразі дозволяють виявляти дрони в радіусі до 150 метрів, що робить їх особливо актуальними для охорони стратегічних об'єктів або прикордонних територій.

Крім того, акустичні системи є ефективними як у денний, так і нічний час, адже не залежать від освітлення. Їхньою перевагою також є здатність розпізнавати стан дрона (наприклад, наявність додаткового навантаження), що суттєво розширює функціональність системи.

Радіочастотний (RF) підхід є ще одним способом ідентифікації та класифікації дронів. БПЛА можуть бути виявлені та класифіковані на основі їхніх радіочастотних характеристик. Крім того, в цій галузі набуває популярності підхід із використанням програмно-визначуваних радіо (SDR). RF-датчик, який розпізнає радіочастоти, виконує роль посередника між безпілотником та його контролером. При прослуховуванні сигналів контролера дрона RF-датчики, на відміну від акустичних, долають проблему обмеженого радіусу дії завдяки використанню антен з високим коефіцієнтом підсилення в поєднанні з високочутливими приймальними системами [1].

Проблема навколишнього шуму також вирішується за допомогою методів зниження шумів, зокрема використанням смугової фільтрації (bandpassing) [2].

У випадку виявлення дронів, які не передають RF-сигнали, можна використовувати недорогі сенсори на основі камер із алгоритмами комп'ютерного зору та акустичні сенсори. Дрони можна візуально розпізнати за допомогою фото сцени з камери – ці підходи є зрозумілими для людини, забезпечують прийнятну локалізацію, середній радіус дії та розумну вартість. Однак такий метод малоефективний у нічний час або за умов поганої видимості через хмари, туман, іній або забруднення.

Застосування тепловізійних камер може частково вирішити ці проблеми. Проте для військових цілей використовуються високоякісні тепловізори. Доступні на ринку комерційні тепловізори можуть втрачати ефективність за високої вологості або інших несприятливих погодних умов [3].

Кожен з описаних вище методів має власні переваги при застосуванні в певних умовах і з певною метою. Відповідно до цього обирається і сфера застосування методу.

У цій роботі розглядається вирішення проблеми інцидентів або проникнення дронів зі спеціальним (небезпечним) вантажем, що становить серйозну загрозу для життя людей у всьому світі. Основна увага зосереджена не на охороні території чи дальності виявлення, а саме на типі дронів, які проникають на захищену територію.

Як було зазначено раніше, акустичний метод розпізнавання є ефективним рішенням для виявлення та класифікації малих БПЛА.

1.2 Виявлення БПЛА на основі акустичних даних

Сьогодні проблема звукової класифікації БПЛА викликає особливий інтерес у науковій спільноті через її здатність виявляти стани дронів за наявності додаткового навантаження, у різних положеннях або для різних моделей [4]. Отже, ця робота спрямована на вивчення проблеми звукового розпізнавання БПЛА. До того ж значущість цього методу зростає завдяки його здатності оцінювати відстані до зон інтересу. Виявлення безпілотних літальних апаратів (БПЛА) у захищених зонах за допомогою акустичних сигналів розширює можливості точного виявлення потенційно небезпечних дронів для своєчасної активації системи безпеки.

Завдяки доступним цінам і розширеному функціоналу, дрони сьогодні стали набагато доступнішими, що сприяло їх поширенню у терористичних і злочинних атаках. Крім того, із розвитком технологій проектування та

виготовлення дронів стали ще доступнішими, а сфера їх можливого застосування стрімко зростає. Дрони особливо активно використовуються для таких цілей, як розпилення пестицидів, доставка їжі, пошуково-рятувальні операції, зганяння птахів у напрямку аеропортів, шпіонаж, допомога у надзвичайних ситуаціях, сільське господарство тощо, і цей список постійно поповнюється [5].

Поряд із численними варіантами застосування дронів в повітряному просторі зростають і загрози безпеці, зокрема: загроза самому повітряному простору, вторгнення в приватне життя, використання дронів як зброї, промислове шпигунство, зіткнення з транспортом, злам дронів. Один із таких інцидентів стався у жовтні 2016 року, коли терористи вбили двох військових, використовуючи дрон для транспортування вибухівки. Тривале використання такої зброї може призвести до масових жертв у мегаполісах, де легко вразити велику кількість людей.

Інший випадок стався 29 вересня 2022 року, коли дрон служби доставки Wing (підрозділ компанії Alphabet) зіткнувся з лініями електропередач у місті Браунс-Плейнс (Австралія), що призвело до відключення електроенергії для майже 2 000 клієнтів. Невідоме навантаження дрона зачепилося за повітряну лінію електропередач. У цьому випадку навіть зняти дрон із кабелю було неможливо. Хоча сам дрон не знеструмив лінію одразу, він передавав напругу до моменту приземлення, спалахнув полум'ям і вибухнув на землі. У результаті 2 000 місцевих жителів залишились без електрики на 45 хвилин, а ще 300 – протягом трьох годин, поки енергетики перевіряли лінії на пошкодження [6], (рисунок 1.1 а).

Кілька недавніх аварій дронів Amazon також були спричинені проблемами з двигунами та пропелерами [7], (рисунок 1.1 б).



Рисунок 1.1 – Приклад аварій дронів

Як видно з вищенаведеного, останніми роками зросла кількість випадків використання безпілотних літальних апаратів (БПЛА) у військових конфліктах. У роботі [8] наводиться розширений список інцидентів із дронами у військових та інших ситуаціях. Загалом, невійськові БПЛА часто ставали причиною інцидентів, які створювали загрозу для авіації, людей або майна на землі. Існує занепокоєння щодо безпеки польотів, оскільки зіткнення з дроном може призвести до пошкодження двигуна літака. Також мали місце підтвержені випадки порушення авіаційних правил любителями, які керували дронами з порушенням правил повітряної безпеки.

Це підтверджує важливість ідентифікації та класифікації БПЛА, які завжди будуть ключовими завданнями. Акустичний сенсорний метод може бути ефективним рішенням для виявлення та класифікації дронів. З огляду на поширення багатоцільових технологій, що дозволяють користувачам створювати власні дрони, і майже неможливість їх централізованого контролю, інші методи стають малоефективними.

У військовій сфері дрони можуть виявлятися за допомогою високоточних радарних систем, але вони є надзвичайно дорогими та непридатними для застосування у міських умовах. Крім того, існують комерційні рішення, які поєднують радар, RF, камери та тепловізори. Однак наведені вище інциденти з дронами потребують визначення моделей дронів, відстаней до об'єктів та характеру їх навантаження.

Акустичний сенсорний метод є оптимальним технічним рішенням цих завдань. Якщо вивчати дрони за їхніми звуковими підписами, то можна визначити модель, стан та положення БПЛА. Це можливо завдяки тому, що різні моделі дронів мають різні двигуни, які створюють різний гул, відповідно – різний спектр частот. Таким чином, можна зібрати достатньо звукових даних для обробки методами глибинного навчання в рамках штучного інтелекту. Крім того, якщо дрон несе додатковий вантаж, навіть у межах однієї моделі, характеристики звуку змінюються через підвищене навантаження на двигун.

Підсумовуючи всі перелічені фактори, можна стверджувати, що дослідження звукового розпізнавання дронів є ефективним напрямом. Застосування методів машинного та глибинного навчання, які належать до найсучасніших напрямів штучного інтелекту, є найнадійнішим рішенням для обробки акустичних даних. Розпізнавання таких об'єктів на основі великої кількості кількісних патернів досягається навчанням нейронних мереж за зібраними прикладами.

Звуки присутні практично в будь-якій діяльності – як людській, так і техногенній – як у приміщенні, так і на вулиці. Технічно задача звукового розпізнавання є складною, оскільки звуковий сигнал динамічний і складний, але інтерес до цього напрямку зростає завдяки широкому спектру застосувань.

Внутрішні звуки (indoor) характерні для офісного, житлового чи освітнього середовища, тоді як зовнішні звуки (outdoor) – для міських або природних умов. До прикладів людських звуків належать: автоматичне розпізнавання мовлення (ASR), ідентифікація мовця (SID), аналіз музики (MIR). А звуки, які утворюються рухомими об'єктами, є результатом роботи двигунів або їхньої взаємодії з середовищем – наприклад, звуки автомобілів, мотоциклів, потягів, літаків і дронів.

Розпізнавання таких звуків у цілях безпеки є одним із ключових завдань. У звуковому розпізнаванні застосовуються як класичні методи

машинного навчання, так і сучасні методи глибинного навчання. Основою успішного розпізнавання є правильна обробка звукового сигналу, зокрема – поділ сигналу на фрейми, що дозволяє виділити цільовий звук і виключити сторонні.

Звукові ознаки формуються на основі акустичних характеристик сигналу, таких як гучність, висота тону, тембр. Широко використовуються також спектрограми, мел-спектрограми, коефіцієнти мел-частотної кепстральної характеристики (MFCC) та їх похідні.

Центральною темою цієї кваліфікаційної роботи є виявлення звукових відмінностей об'єктів, зокрема – дронів. Визначення об'єкта за його звуком стало ефективним способом класифікації, а наукові дослідження підтверджують, що правильна обробка акустичних параметрів сигналу дозволяє досягати високої точності розпізнавання.

Методи глибинного навчання для розпізнавання звуків стали предметом численних досліджень. Вони відрізняються від класичних підходів тим, що в них використовуються ознаки з часової, частотної та перцептивної областей, які подаються на вхід алгоритмам машинного навчання.

2 ОПИС АКУСТИЧНИХ ДАНИХ БПЛА

У процесі дослідження підозрілої діяльності БПЛА було запропоновано створити більш надійну систему виявлення цих літальних апаратів на основі розпізнавання їхніх акустичних даних, яка б могла ідентифікувати їхні звуки в різних ситуаціях, включаючи стани з додатковим навантаженням. Запропонована система акустичного виявлення БПЛА має на меті розробити інтелектуальний аудіосенсор, здатний у реальному часі виявляти появу БПЛА у певних зонах або на територіях за їхнім звуком – як під час польоту, так і під час польоту з додатковим вантажем.

Розробка цієї системи включає два основні етапи: підготовку акустичних даних БПЛА для різних моделей і умов; та побудову архітектури системи розпізнавання в реальному часі. У цьому розділі розглядається етап збору акустичних даних від БПЛА, оскільки для початку дослідження необхідно сформувати початкову базу даних.

Початковий етап стосувався підготовки акустичних даних БПЛА шляхом запису звуків їхніх польотів, а також польотів із вантажем різних моделей і з різною вагою вантажу, на різних відстанях – 0,5 м та 100 м від мікрофона. Процес запису набору даних БПЛА було реалізовано шляхом проведення польотів БПЛА моделей DJI Phantom 1 та DJI Phantom 2 з вантажем і без нього на відстанях від 0,5 до 100 метрів від мікрофона.

Дані збиралися протягом кількох різних сезонів. Під час запуску деяких БПЛА поблизу було чути шум вантажного потяга, мотоциклів, автомобілів, вантажівок Gator, а також фонові шуми з людськими голосами. Під час тестування також фіксувалися звуки вітру, шелестіння листя та інші навколишні шуми, дані про які також були зібрані з метою відокремлення звуків БПЛА від хибно негативних спрацьовувань.

Під час запису використовувався DJI Phantom 2 для запуску навантаженого БПЛА, який переносив вантаж у вигляді 0,5 кг пластиліну.

Модель БПЛА Syma X20, яка часто використовується для розваг, була протестована як у навантаженому стані (з металевим акумулятором масою 0,425 кг), так і без навантаження. При оцінюванні впливу навантаження на ці аматорські дрони враховувалась імовірність шкоди у разі помилки керування.

Крім того, для запису акустичних даних без навантаження були протестовані інші доступні моделі БПЛА, зокрема Tarantula x6 та Syma x5c, на відстанях від 1 до 40 метрів. Інші моделі БПЛА, включно з DJI Phantom 1, 2, 3, 4, DJI Phantom 4 Pro, Mavic Pro та Qazdrone, також були запуснені з параметрами, наведеними в таблиці 2.1, і їхні звуки були додані до датасету.

Таблиця 2.1 – Опис моделей БПЛА, які включені до датасету

| Модель БПЛА | Вантажопідйомність (кг) | Дальність (м) |
|--------------------------|-------------------------|---------------|
| DJI Phantom I | – | 2–100 |
| DJI Phantom II | – | 2–100 |
| DJI Phantom II | 0.5 | невідомо |
| DJI Phantom III | – | невідомо |
| DJI Phantom III | 0.454 | невідомо |
| DJI Phantom IV | – | 2–100 |
| DJI Phantom IV | 0.4 | 2–100 |
| DJI Phantom IV Pro | – | невідомо |
| DJI Phantom IV Pro | Січ.36 | невідомо |
| DJI Phantom квадрокоптер | – | невідомо |
| Mavic Pro | 0.156–0.256 | невідомо |
| Syma x5 | – | 1–40 |
| Syma x20 | 0.425 | 1–40 |

Усі ці БПЛА були записані з використанням мікрофона з глибиною роздільної здатності 16 біт при частоті 44 100 Гц, рухаючись вгору і вниз, вперед і назад на різних швидкостях залежно від їхніх технологічних

характеристик, починаючи з досить близької відстані до мікрофона та прилеглої автостоянки. Решта інформації була зібрана з безкоштовних і відкритих ресурсів. Процес збору аудіофайлів із публічних джерел, таких як "www.zaplast.com" та "www.sound-ideas.com", вимагав набагато більше зусиль. Це пов'язано з тим, що наша система прогнозування була заснована на акустичному сенсорі, здатному слухати на частоті 44100 Гц з глибиною 16 біт, а звуки навантажених БПЛА були виявлені лише на аматорських відео та оброблені за допомогою спеціального конвертера на частоті 44100 Гц з глибиною 16 біт. Решта даних із відкритих джерел також була перетворена з різних форматів на частоту дискретизації 44 000 Гц із глибиною 16 біт у режимі мікрофона "моно" з розширенням ".wav", оскільки наша модель була створена для отримання аудіоданих саме у форматі wav.

DJI Phantom 2 та його навантажені стани були єдиною моделлю БПЛА, яка розглядалася в попередніх дослідженнях, що мала таке обмеження. Це дослідження має на меті дослідити вплив акустичних даних різних моделей БПЛА на проблему складних звуків БПЛА та їхніх навантажених станів. Метою дослідження є з'ясувати, як акустичні дані різних моделей БПЛА впливають на розпізнавання навантаження БПЛА для різних моделей і ваг.

Таблиця 2.2 – Тривалість набору звукових даних БПЛА

| Класи | Загальна тривалість, с | Навчальна вибірка, с | Вибірка для прогнозу, с |
|-----------------------|------------------------|----------------------|-------------------------|
| БПЛА з навантаженням | 1513 | 1413 | 100 |
| БПЛА без навантаження | 3334 | 3234 | 100 |
| Фоновий шум | 2765 | 2765 | 100 |
| Усього | 7612 | 7312 | 300 |

У цілому, вся інформація про записи БПЛА була зібрана та поділена на три категорії: «Без навантаження», «З навантаженням» та «Фоновий шум». У трьох папках містяться всі ці 31 записаний і зібраний звук. Тривалість записаних звуків дронів варіювалася від кількох секунд до понад п'яти хвилин. У таблиці 2.2 подано загальний огляд тривалості зібраних звуків для кожного класу (в секундах).

Звуки БПЛА з відкритих джерел включали кілька звукових файлів у режимі «стерео». Під час експерименту деякі звуки, які видавали дрони Qazdrone, DJI Phantom 2, DJI Phantom 4 та DJI Phantom 4 Pro, були записані за допомогою мікрофонів пристроїв Apple, таких як iPhone 13B та iPad AIR 2020. Використовуючи спеціально створений фільтр, усі звукові файли були перетворені на частоту 44 100 Гц у режимі «моно».

3 МЕТОДИ ЦИФРОВОЇ ОБРОБКИ СИГНАЛУ

3.1 Принцип представлення звукових даних

У кваліфікаційній роботі дослідження охоплює обробку та розпізнавання звуків об'єктів, зокрема звуків БПЛА, різних моторизованих об'єктів і навколишнього середовища. У цьому розділі розглядається теоретичне основоположне представлення загального аудіосигналу, а також способи його подання для подальшого вивчення акустичних сигналів, які можуть створюватися різними об'єктами або фоновими джерелами.

Аудіосигнал – це представлення будь-якого звуку, яке часто складається або з послідовності бінарних значень для цифрових сигналів, або зі змінного рівня електричної напруги для аналогових сигналів .

По суті, звук виникає тоді, коли вібрації об'єкта передаються через середовище і досягають барабанної перетинки. У фізиці звук – це хвиля тиску. У результаті вібрації об'єкта молекули повітря в його безпосередній близькості також починають вібрувати, створюючи серію звукових хвиль, які резонують у середовищі.

Вібрація моторизованих об'єктів може виникати внаслідок руху компонентів двигуна. БПЛА – це один із прикладів таких об'єктів. Звук існує незалежно від здатності людини його сприймати, у той час як фізіологічне визначення звуку також враховує здатність суб'єкта чути його (рисунок 2.1).

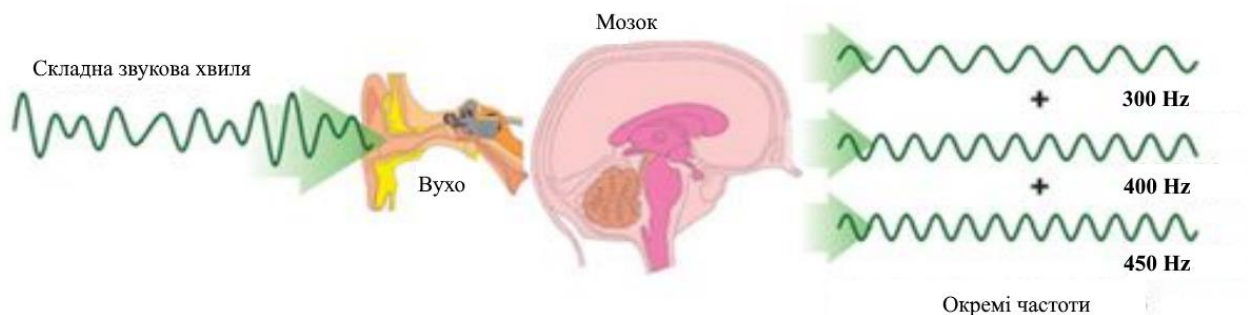


Рисунок 2.1 – Процес утворення звуку та система сприйняття людини

Отже, звукова хвиля виникає в результаті вібрації об'єкта, яка створює звук. Навколишнє середовище (повітря, вода або тверда речовина) зазнає дії хвилі тиску, що спричиняє коливальний рух частинок. Звук передається далі середовищем унаслідок руху сусідніх частинок, які починають вібрувати під впливом хвилі. Коливальні частинки повітря викликають вібрацію дрібних компонентів людського вуха, що призводить до сприйняття звукових хвиль [57].

Траєкторія цих частинок подібна до синусоїдальних хвиль. У зв'язку з існуванням цього фізичного явища, дослідження звуку в формі хвиль є загальноприйнятим. Звукові хвилі часто спрощують, описуючи як синусоїдальні хвилі, що мають спільні властивості, такі як частота, довжина хвилі, амплітуда, звуковий тиск або інтенсивність тощо.

Загальне формулювання синусоїдальної хвилі, наведене на рисунку 2.2, можна подати у вигляді рівняння:

$$y(t) = A \sin(2\pi ft + \varphi) = A \sin(\omega t + \varphi), \quad (2.1)$$

де A – амплітуда, максимальне відхилення функції від нуля;

f – частота, кількість коливань (циклів), що відбуваються за одну секунду часу;

$\omega = 2\pi f$ – кутова частота, швидкість зміни аргументу функції, вимірюється в радіанах за секунду.

φ – це фаза, яка вказує (у радіанах), у якій точці свого циклу перебуває коливання при $t = 0$. Коли φ не дорівнює нулю, форма хвилі, очевидно, зміщується в часі на величину φ/ω секунд. Від'ємне значення означає затримку, а додатне – випередження.

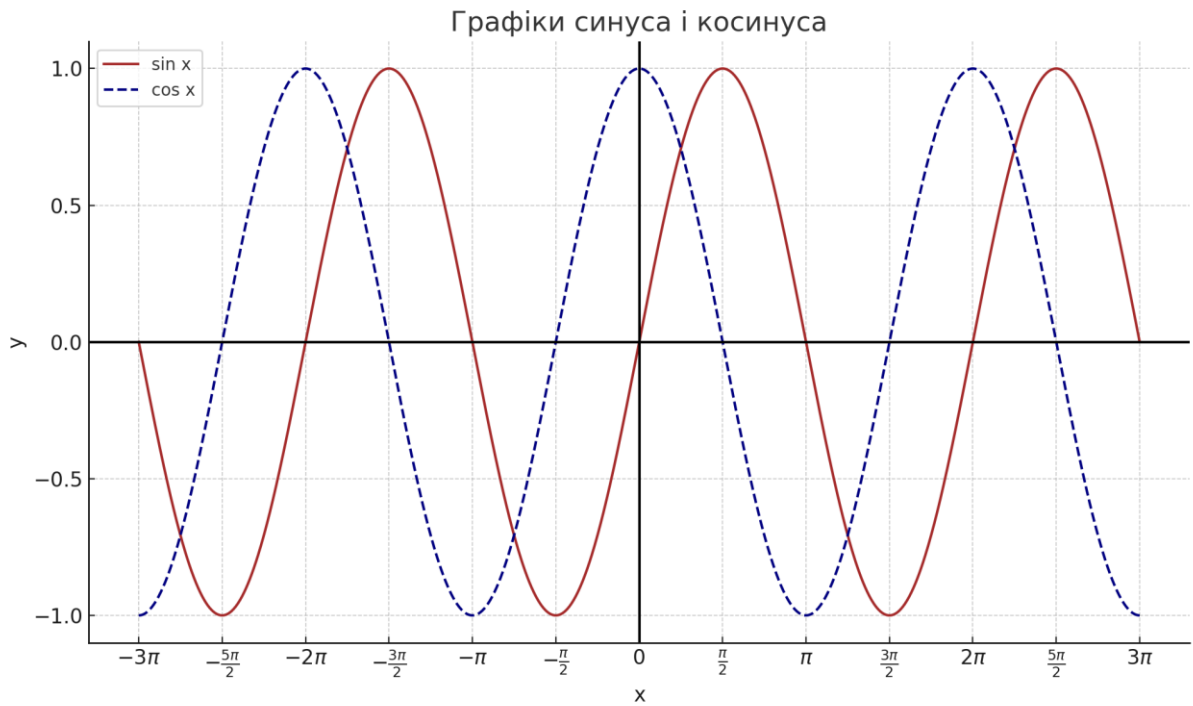


Рисунок 2.2 – Представлення синусоїдальної хвилі

Оскільки синусоїда зберігає свою форму при поєднанні з іншою синусоїдою тієї ж частоти та довільної фази й амплітуди, вона має важливе значення у фізиці. Ця властивість є унікальною для цієї періодичної хвилі. Вона робить синусоїду акустично особливою та суттєвою в аналізі Фур'є.

Загалом, амплітуда і частота є основними характеристиками звукових сигналів. Амплітуда описує величину сигналу, наприклад гучність аудіосигналу. Частота – це кількість повторень хвилі звукового тиску за секунду, і вимірюється в герцах.

Оскільки в природі багато повітряних частинок, які коливаються через суміжні частинки, математична сума кількох або багатьох синусів чи косинусів може відтворювати звуки певних об'єктів. Додавання кількох синусоїд утворює унікальну форму хвилі, яка змінює тембр звуку. Для людського вуха звук, утворений більше ніж однією синусоїдою, має чутні гармоніки.

Підсумовуючи згадані явища, можна сказати, що звуки є складними сигналами.

Складні сигнали можуть аналізуватися за допомогою моделей дискретного або неперервного часу. Це два різні підходи до моделювання змінних, що залежать від часу.

Неперервний сигнал – це сигнал, що змінюється з часом і має неперервну область визначення. Таким чином, область визначення функції є незліченною множиною. Сама функція не обов'язково є неперервною. Неперервний у часі сигнал (також відомий як аналоговий сигнал) — це сигнал з постійною амплітудою та тривалістю, який має значення в кожен момент часу.

У свою чергу, дискретний у часі сигнал має зліченну область визначення, як-от натуральні числа. Так само, як звуки можуть бути представлені дискретними сигналами, саме дискретні системи будуть створені для обробки цих аудіосигналів.

3.2 Обробка акустичних сигналів у часовій області

Обробка сигналів може здійснюватися в будь-якій області, оскільки звукові сигнали можуть бути представлені як в аналоговому, так і в цифровому форматах. Аналогові процесори працюють безпосередньо з електричними сигналами, тоді як цифрові – оперують бінарними представленнями сигналу. Тому варто коротко розглянути поняття аналогових і цифрових сигналів. Аналоговий звук – це електричний сигнал, що відображає зміну тиску звукової хвилі. У цифровому ж вигляді звукова хвиля представлена у вигляді послідовності чисел (дискретна функція). У найпростішому випадку цифрове представлення виконується за допомогою комп'ютерів і мікропроцесорів.

Хоча перетворення аналогового сигналу в цифровий може бути з втратами, цифрова обробка звуку має значно більшу ефективність та гнучкість, тому сучасні аудіосистеми переважно використовують саме цей підхід. Одним із прикладних напрямів обробки аудіосигналів є зберігання

звуку, тобто запис і відтворення, що дозволяє зафіксувати звуки голосу, музики чи навколишнього середовища на носіях для подальшого відтворення.

Іншим поширеним застосуванням є стиснення аудіоданих. Мета – зменшити обсяг збереження або пропускну здатність для передачі аудіо. Існує два типи стиснення: без втрат (lossless) і з втратами (lossy). Наприклад, аудіофайли формату WAV – це кодування без втрат, яке дозволяє зберегти повну інформацію про сигнал. Формати MP3, AAC, Ogg Vorbis чи FLAC – приклади аудіокодерів із стисненням з втратами. У рамках даного дослідження для запису та обробки аудіосигналів було обрано саме формат WAV, який дозволяє працювати з "сирими" (raw) звуковими даними.

Обробка аудіосигналу починається з таких сирих даних. Оскільки звук за своєю природою є неперервним у часі, першим етапом є перетворення його в дискретний сигнал шляхом семплювання. Семплювання – це процес перетворення неперервного сигналу на дискретний. Кожна вибірка (sample) – це значення сигналу в певний момент часу. Таким чином, можна представити сигнал у вигляді вектора, де кожен елемент – це значення амплітуди в конкретний момент.

З математичної точки зору аудіосигнал у часовій області – це одномірний вектор, що містить значення амплітуди в часі. Такий вектор є прикладом часових рядів – послідовності значень, впорядкованих у часі. Функцію сигналу можна представити як $x(t)$, де t – час. Таким чином, сигнал можна візуалізувати на графіку як залежність амплітуди від часу.

Під час обробки сигналів у часовій області важливо дотримуватися теореми Найквіста-Шеннона, яка стверджує, що частота семплювання F_{smax} бути більшою за подвоєне максимальне значення частоти сигналу f_{max} . Це дозволяє уникнути ефекту аліасингу – накладання частот. У цифрових пристроях та програмних середовищах для машинного навчання ця умова повинна обов'язково виконуватись.

Однак часовий домен надає лише інформацію про зміну амплітуди з часом. Щоб отримати більше інформації про природу звуку, необхідно перейти до аналізу сигналу в частотній області.

3.3 Короткочасне перетворення Фур'є (Short-Time Fourier Transform, STFT)

Усі чутні звукові сигнали, що зустрічаються в природі, можуть бути представлені як сукупність чистих синусоїд різної частоти. Математичний метод, відомий як перетворення Фур'є, використовує декомпозицію сигналу на окремі частотні складові, щоб визначити його спектральний склад. Згенеровані синусоїди в перетворенні Фур'є для сигналу як функції часу мають комплексне значення, де уявна частина відповідає фазовому зсуву синусоїди, а абсолютне значення – амплітуді відповідної частотної компоненти.

Оскільки аудіосигнал є дискретним, до нього застосовується дискретне перетворення Фур'є (DFT). Це є реалізація класичного перетворення Фур'є для дискретних сигналів.

Загалом, математичну основу вищезгаданого перетворення DFT, що дозволяє перейти з часової області у частотну, описує наступне рівняння:

$$X_k = \sum_{n=0}^{N-1} x_n e^{\frac{-2\pi i k n}{N}}. \quad (2.2)$$

Дискретне перетворення сигналу може бути представлене за допомогою комплексних чисел і тригонометричних хвиль із комплексними значеннями. Найефективнішим методом обчислення дискретного перетворення Фур'є (DFT), який дозволяє перейти від часової області до частотної, є швидке перетворення Фур'є (FFT, Fast Fourier Transform).

FFT використовується для представлення сигналу в частотній області та його подальшого аналізу, що описується рівнянням[48, с. 39]:

$$f(x) = \int_{-\infty}^{\infty} f(x)e^{-2\pi ikx} dx. \quad (2.3)$$

На практиці, якщо FFT застосовується до аудіофайлу певної довжини, воно надає інформацію лише про частоти та амплітуди цього аудіосигналу. Обробка сигналу лише за частотою та амплітудою недостатня для створення потоку даних для задач машинного навчання. З цією метою обробка аудіосигналу із збереженням інформації одночасно з часової та частотної областей дозволяє отримати розширене уявлення про вміст аудіосигналу.

Короткочасне перетворення Фур'є (STFT) дозволяє виконувати таку обробку сигналу із збереженням часово-частотної інформації. Оскільки аудіосигнал постійно змінюється, передбачається, що він не змінюється істотно в межах коротких інтервалів часу, що спрощує процес обробки. Тому розбиття довжини вхідного сигналу на малі часові інтервали дозволяє витягувати з них інформацію, пов'язану з частотними характеристиками. Ці частини називають фреймами.

Зазвичай тривалість фреймів становить від 20 до 40 мс і вони формуються із вхідного аудіосигналу. Якщо фрейм занадто довгий, сигнал може надто сильно коливатися всередині нього; якщо занадто короткий – вибірок буде недостатньо для надійної спектральної оцінки. Це реалізується теоретично за допомогою математичного методу короткочасного перетворення Фур'є (рисунок 2.3).

STFT є перетворенням, пов'язаним із Фур'є, і використовується для визначення синусоїдальних частот і фазового складу коротких інтервалів сигналу, що змінюється з часом. Нехай функція є представленням будь-якого аудіосигналу (рисунок 2.3). Ця функція $x(t)$ ділиться на певні «короткі часові сегменти». Далі для кожного сегмента обчислюється FFT (рисунок 2.3).

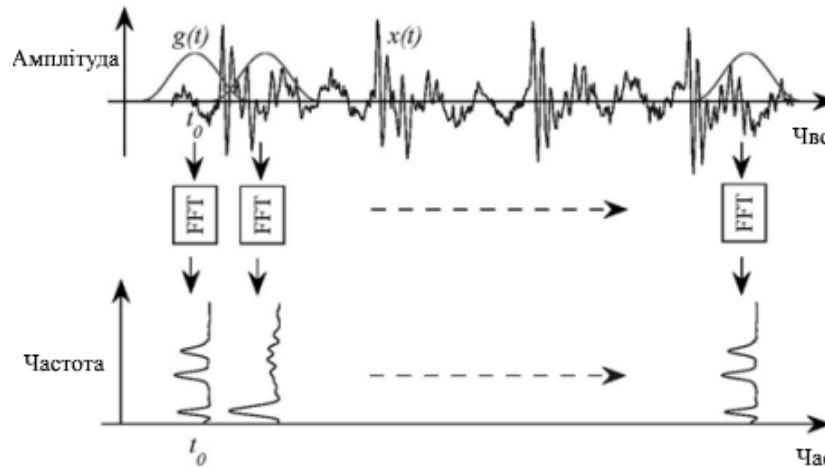


Рисунок 2.3 – Процес розрахунку ШПФ для аудіосигналів

Малі сегменти на рисунках 2.3 і 2.4 вибираються за спеціальним правилом. Це називається віконуванням (Windowing). «Вікно» в обробці сигналів – це функція (форма), яка є ненульовою протягом певного проміжку часу і дорівнює нулю до і після цього періоду. За винятком ненульової частини вікна, де воно накладається на інший сигнал, множення його на інший сигнал дає в результаті 0. Віконування найчастіше використовується у спектральному аналізі, тобто у процесі спостереження короткого фрагмента більшого сигналу та дослідження його

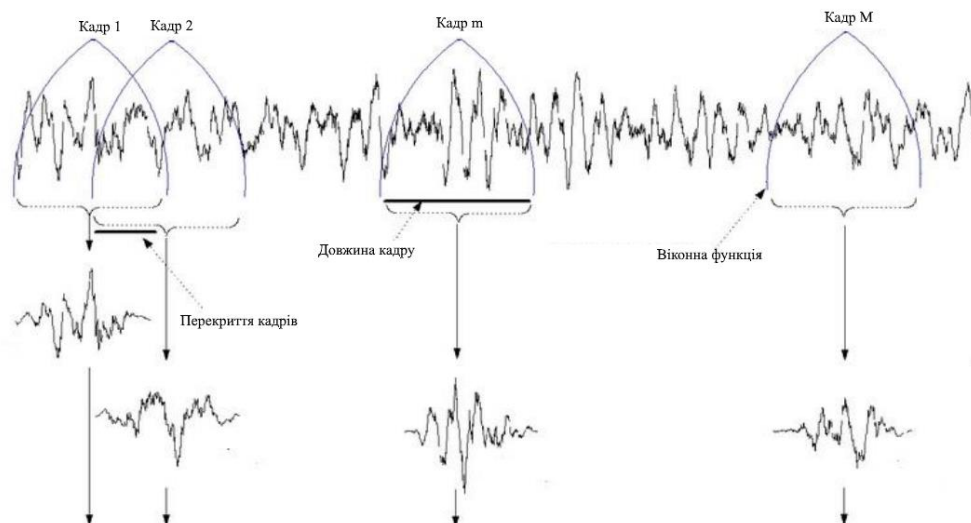


Рисунок 2.3 – Візуальне представлення розрахунку фреймів

2.4 Мел-спектрограми (Mel-Scale Spectrograms)

Мел-спектрограма є однією з найпоширеніших форм представлення аудіосигналів у задачах розпізнавання звуків, зокрема для аналізу акустичних даних БПЛА. Вона базується на перетворенні STFT, але зі шкалою частот, адаптованою до людського сприйняття звуку – мел-шкалою.

Мел-шкала (від англ. *mel* – мелодія) – це логарифмічна шкала, яка наближає сприйняття висоти звуку людиною. Частоти нижче 1000 Гц сприймаються майже лінійно, тоді як частоти вище – логарифмічно. Перетворення з частоти f (в Гц) у мел-одиноці виконується за формулою:

$$\text{mel}(f) = 2595 \cdot \log_{10} \left(1 + \frac{f}{700} \right), \quad (2.4)$$

іншими словами, мел-шкала – це психоакустичне лінійне представлення частоти. Кожна з мембран кохлеї вуха вібрує на певну компоненту частоти, виконуючи роль смугового фільтра у слуховій системі людини. У 1937 році Стівенс, Фолькманн і Ньюмен запропонували одиницю висоти тону "Mel" як спосіб відтворення цих характеристик у звуковій обробці [20, с. 11].

Mel – це перцептивна шкала висот тону, які слухачі сприймають як рівновіддалені одна від одної. В ході численних досліджень було з'ясовано, що слухова система людини сприймає частоти лінійно при значеннях нижче 1000 Гц, і логарифмічно – при значеннях вище 1000 Гц. Основна мета мел-шкали полягає у відображенні цієї особливості.

Мел-шкала зазвичай використовується для побудови мел-спектрограм, а в окремих випадках – для розрахунку спектрограм коефіцієнтів MFCC (Mel-Frequency Cepstral Coefficients). Спектрограми коефіцієнтів MFCC розраховуються за допомогою дискретного перетворення Фур'є (DFT) на основі значень мел-спектрограм.

4 ГЛИБИННІ МЕТОДИ НАВЧАННЯ ДЛЯ РОЗПІЗНАВАННЯ АКУСТИЧНИХ ДАНИХ БПЛА

4.1 Згорткові нейронні мережі (CNN) у задачах розпізнавання звуку

Згорткові нейронні мережі (CNN, Convolutional Neural Networks) є одним із типів глибоких нейронних мереж, які застосовуються в різних галузях, таких як розпізнавання об'єктів, комп'ютерний зір, розпізнавання звуку та обробка природної мови (NLP). Основною структурною особливістю CNN є наявність стандартної нейронної мережі, яка складається з шару семплювання та багатьох згорткових шарів.

CNN здебільшого розроблялися для розпізнавання зображень за ознаками у двовимірному просторі. Вхідними даними для таких мереж є зображення або спектрограми, які подаються у вигляді 2D-матриць ознак. Кожен шар мережі навчається розпізнавати різні аспекти зображення або спектрограми. Кожне навчальне зображення (або спектрограма) обробляється фільтрами з різною роздільною здатністю, а результат згортки кожного фільтра подається на вхід наступного шару.

Починаючи з базових ознак, таких як контраст, границі або освітленість, фільтри поступово навчаються розпізнавати складніші характеристики об'єктів. CNN демонструє високу здатність до навчання, вимагає мінімальної попередньої обробки сигналів і вже успішно застосовується для розпізнавання рукописного тексту, облич, об'єктів та звуків. Архітектура CNN складається з трьох основних частин: вхідного шару, набору прихованих шарів і вихідного шару (рисунок 4.1).

До найпоширеніших шарів відносяться:

- згортковий шар (convolutional layer);
- шар активації (activation layer);
- шар підвибірки (pooling layer).

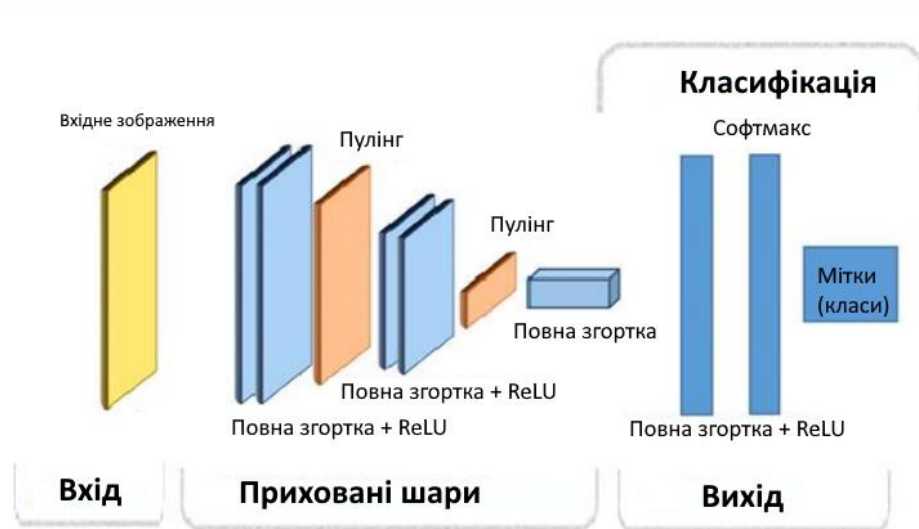


Рисунок 4.1 – Загальна архітектура мережі

Основним компонентом CNN є згортковий шар, який несе найбільше обчислювальне навантаження. Під час згортки до вхідних даних застосовуються фільтри, кожен з яких активується для певних ознак сигналу.

Наступним шаром є шар активації, який, завдяки використанню нелінійних функцій, таких як ReLU, sigmoid, softmax чи tanh, дозволяє мережі виявляти складні залежності. Найчастіше в CNN використовується функція активації ReLU (рисунок 4.1), яка прискорює навчання, замінюючи від'ємні значення на нуль, а додатні залишаючи без змін.

Приховані шари (hidden layers) у CNN відповідають згортковим шарам, кожен з яких формує карту ознак (feature map або activation map), яка передається на наступний рівень обробки.

На відміну від повнозв'язних нейромереж, CNN значно ефективніші при роботі з великими розмірностями вхідних даних (наприклад, зображення або спектрограми великої роздільності). Це досягається завдяки використанню локальних зв'язків (receptive fields) та меншої кількості параметрів, що також дозволяє уникнути проблеми зникнення або вибуху градієнтів при навчанні.

Шар Pooling йде наступним у списку. Завдяки нелінійній процедурі зменшення розмірності (downsampling), pooling знижує кількість параметрів,

які мережа має навчити, спрощуючи при цьому вихідні дані. У згорткових нейронних мережах (CNN) можуть також використовуватись як стандартні згорткові шари, так і локальні або глобальні pooling-шари. За допомогою методу об'єднання виходів груп нейронів одного шару в один нейрон наступного шару, pooling-шари зменшують розмірність даних.

Локальне pooling здійснюється зазвичай на невеликих фрагментах (наприклад, з розміром плитки 2x2). Глобальне pooling охоплює всю карту ознак (feature map) – кожен нейрон впливає на результат. Найпоширенішими типами pooling є max pooling та average pooling. У max pooling вибирається максимальне значення серед локального кластеру нейронів, тоді як average pooling обчислює середнє значення.

Після навчання ознакам у багатьох шарах CNN переходить до етапу класифікації. Передостанній шар – це повнозв'язний шар (fully connected layer), який формує вектор розмірності N , де N – максимальна кількість класів, які модель може передбачити. Цей вектор містить ймовірності належності об'єкта до кожного з класів. У повнозв'язаних шарах усі нейрони одного шару з'єднані з усіма нейронами іншого шару – аналогічно до звичайної багат шарової перцептронної мережі (MLP). Перед остаточною класифікацією згорткова карта ознак «розплющується» (flattening) і передається до повнозв'язаного шару.

Кінцевий вихід класифікації надає останній класифікаційний шар CNN. Протягом еволюції задачі розпізнавання об'єктів було створено багато моделей CNN. До них належать: LeNet, AlexNet, ResNet, GoogleNet / Inception, MobileNetV1, ZfNet та глибинні CNN (Depth-based CNNs).

У задачах розпізнавання звукових сигналів часто використовувались прості типи згорткових шарів, побудовані на кількох шарах.

Інфраструктура CNN є адаптивною для обробки зображень завдяки своїй структурі та функціональності. Однак у випадку змінних у часі сигналів, таких як звук, доцільно розглянути теоретичні основи рекурентних нейронних мереж (RNN), яким присвячено наступний підрозділ.

4.2 Рекурентні нейронні мережі (RNN) у задачах розпізнавання звуку

Початковим та найпростішим варіантом штучної нейронної мережі була нейронна мережа прямого поширення сигналу (feedforward neural network). У цій мережі дані передаються лише в одному напрямку – від вхідних вузлів через будь-які наявні приховані шари до вихідних вузлів. У мережі немає зворотних зв'язків або циклів. Нейронні мережі прямого поширення сигналу виявилися привабливими у задачах розпізнавання звуку, що підтверджується численними дослідженнями. Крім того, мережі цього типу стали популярними для розв'язання задач передбачення, таких як розпізнавання зображень, комп'ютерний зір, розпізнавання мовлення, виявлення звуків та інші, оскільки використовують кілька прихованих шарів для максимального навчання з вхідних даних.

Основною проблемою використання лише одного прихованого шару є перенавчання (overfitting). Збільшення кількості прихованих шарів може зменшити перенавчання та покращити узагальнення моделі. Із додаванням шарів нейронні мережі стають глибокими (Deep FNN). Проте, основним недоліком глибоких нейронних мереж прямого поширення є експоненційне зростання часу навчання, що робить їх використання малопрактичним у багатьох задачах.

У відповідь на ці функціональні обмеження були розроблені рекурентні нейронні мережі (RNN). Вони є підкласом мереж прямого поширення, але мають здатність працювати з послідовними та часовими даними, виявляючи довготривалі залежності та характеристики. На відміну від FFNN, у RNN вхід для кожного нейрона прихованого шару містить затримку у часі. Поточна ітерація в RNN залежить від історичних даних. Наприклад, щоб передбачити слово в реченні, потрібно враховувати попередні слова. Таким чином, RNN опрацьовують вхідні дані протягом часу, використовуючи ваги та довільну довжину входу. Обчислення цієї моделі враховують історичний контекст, а розмір не залежить від обсягу даних.

Недоліком RNN є низька швидкість обробки, що вважається її слабкою стороною. З метою вирішення цього недоліку було запропоновано кілька варіантів RNN. На сьогодні найпопулярнішими для задач прогнозування є чотири типи обчислювальних осередків: проста RNN, LSTM (Long Short-Term Memory), BiLSTM (бінаправлена LSTM) та GRU (Gated Recurrent Unit).

Прості рекурентні нейронні мережі (Simple RNN) – це підклас нейронних мереж, здатних розпізнавати послідовні дані. Ці мережі широко використовуються у середовищі програмування Python з бібліотеками Keras. У бібліотеках Keras стандартні рекурентні мережі відомі під назвою "SimpleRNN".

Проста RNN складається з трьох шарів: вхідного, прихованого та вихідного, як зображено на рисунку 4.2.

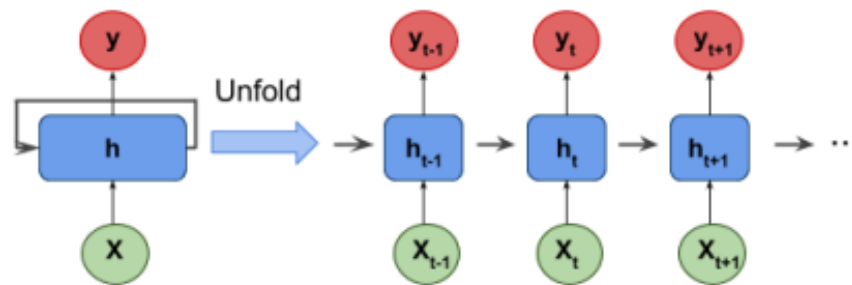


Рисунок 4.2 – Проста структура RNN та її розгорнута (розгорнута) форма

Згідно з основним принципом роботи Simple RNN, вузли мережі пов'язуються між собою так, щоб поточна інформація враховувала попередній стан: вихід шару нейронної мережі в момент часу t подається на вхід того самого шару на наступному кроці часу $t + 1$.

Вхідні дані являють собою послідовність векторів у часі t , наприклад:
 $\dots, x_{t-1}, x_t, x_{t+1} \dots$

Вхідні блоки у Simple RNN з повним з'єднанням передають дані до прихованих блоків прихованого шару. Приховані одиниці (units) в прихованому шарі мають вигляд $h_{t-1}, h_t, h_{t+1} \dots$ і пов'язані між собою в часі через рекурентні (циклічні) з'єднання.

На рисунку 4.2 концепція розгортання (unfolding) структури RNN представлена у вигляді послідовних кроків для випадку:

- множинних часових кроків вхідних даних ..., X_{t-1} , X_t , X_{t+1} ...;
- множинних часових кроків внутрішнього стану h_{t-1} , h_t , h_{t+1} ...;
- множинних часових кроків вихідних даних u_{t-1} , u_t , u_{t+1}

Продуктивність і стабільність мережі можна підвищити шляхом ініціалізації прихованих модулів невеликими ненульовими значеннями. Однак ці мережі мають і недоліки. Основними недоліками Simple RNN є проблеми зникання та вибуху градієнта. З метою подолання цих недоліків стандартної мережі RNN були розроблені мережі LSTM.

Мережі з довготривалою короткочасною пам'яттю (LSTM – Long Short-Term Memory) є вдосконаленням стандартних рекурентних нейронних мереж (RNN) і були запропоновані для вирішення проблем зникнення та вибуху градієнтів, що виникають під час навчання звичайних RNN.

Головна особливість LSTM полягає у впровадженні спеціальних модулів, які називаються LSTM-комірками (LSTM cells). Кожна така комірка здатна зберігати інформацію впродовж довгого періоду часу завдяки внутрішньому механізму контролю – вентилям (gates), які регулюють потік інформації:

- forget gate: вирішує, яку частину старої інформації слід зберегти або видалити з комірки пам'яті;
- input gate: визначає, яку нову інформацію потрібно зберегти в комірці пам'яті;
- output gate: контролює, яку інформацію з комірки пам'яті потрібно передати на вихід.

Завдяки цим вентилям, LSTM можуть ефективно навчатися на довгих послідовностях, що є критично важливим для задач розпізнавання звуків, особливо у випадку обробки акустичних сигналів БПЛА (дронів), де часовий контекст відіграє важливу роль.

У порівнянні зі звичайними RNN, LSTM забезпечують кращу здатність до запам'ятовування важливої інформації на тривалих інтервалах часу, а також стабільніше та точніше навчаються на складних послідовних даних.

4.3 Система розпізнавання та класифікації акустичних даних БПЛА в реальному часі

Загалом, вивчення розпізнавання звуків дронів розпочалося з запису їхніх звуків. Це зумовлено тим, що звуки дронів необхідні як початкові дані для початку дослідження. Основна увага в роботі приділяється розробці системи розпізнавання, здатної працювати в режимі реального часу. Під "режимом реального часу" в цьому дослідженні мається на увазі система, адаптована для розпізнавання аудіофайлів тривалістю 1 секунда. Для цього вона повинна відповідати вимогам обробки аудіоданих із застосуванням глибинного навчання. Інакше кажучи, аудіодані для моделі глибинного навчання необхідно розділити на групи для «навчання» моделі та «валідації» для перевірки її надійності. Насамперед усі записані звуки дронів були розподілені на 3 основні класи відповідно до їх змісту. Це клас «Дрон з навантаженням» зі спеціальним вантажем, який імітує підозрілі дрони, клас «Дрон без навантаження» та клас «Фоновий шум». Зібрані звуки для цих трьох різних класів були збережені з їхньою початковою тривалістю у трьох окремих папках. Однак тривалість аудіозаписів має бути скоригована таким чином, щоб відгук системи розпізнавання можна було отримувати щосекунди. Отже, потрібно, щоб тривалість раніше записаних і зібраних аудіоданих становила рівно 1 секунду. Для виконання цього завдання був створений «спеціальний фільтраційний блок», який повторно адаптує записи звуків дронів різного стану та тривалості до заданих умов і розміщує їх у відповідних папках .



Рисунок 4.3 – Алгоритм адаптації звукозапису БПЛА для систем реального часу

Акустичні дані БПЛА були адаптовані перед вивченням етапу розпізнавання звуків БПЛА на основі аналізу частотних діапазонів. Спочатку акустичні дані вивчалися в часовій області. Наш клас фонового шуму складається зі звуків багатьох моторизованих об'єктів. Ці звуки були зібрані з метою запобігання хибному розпізнаванню, яке могло б виникнути через можливість плутанини під час розпізнавання.

На рисунку 4.4 представлено діаграму класів вибірки.

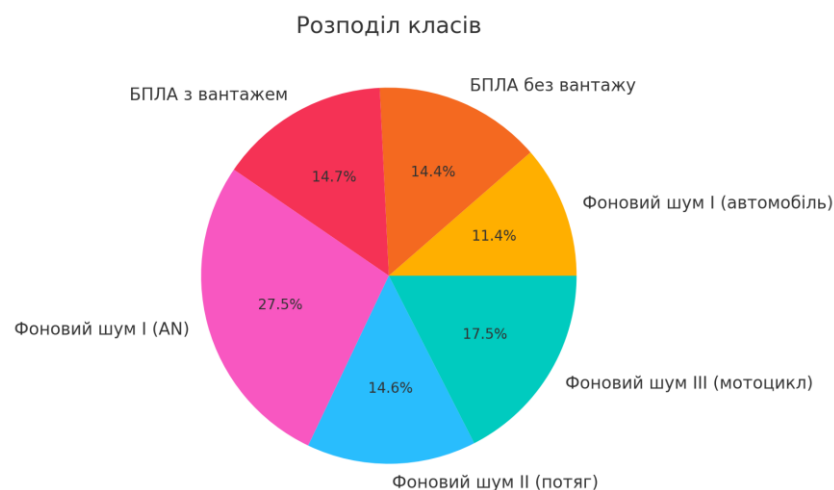


Рисунок 4.4 – Діаграма класів вибірки

В роботі було протестовано чотири рекурентні нейронні мережі: SimpleRNN, LSTM, BiLSTM та GRU. Також була навчена структура моделі CNN. Експериментальна робота виконувалась у програмі на Python на процесорі Intel(R) Core (TM) i5-8265U з тактовою частотою 1.60 ГГц.

Розподіл акустичних даних БПЛА здійснювався у співвідношенні 70 на 30: 70% від загальної кількості отриманих звукових файлів було використано для навчання, а решта 30% – для валідації, при цьому ці файли не були показані моделям під час навчання. Окрім цих 30% валідаційних файлів, окремо було збережено по 100 аудіофайлів тривалістю 1 секунда для кожного класу, всього 300 файлів для трьох класів, щоб перевірити, наскільки добре та стабільно моделі здатні розрізняти кожен клас.

У результаті навчання точність розпізнавання моделей, розроблених для систем реального часу, була ретельно перевірена на 30% валідаційних даних за допомогою графіків точності розпізнавання. Спочатку всі моделі навчалися протягом 30 епох із використанням 32 RNN-комірок. Після аналізу графіків було вибрано область “хорошого узгодження” при 25 епохах, і кожну модель було повторно навчено протягом 25 епох перед збереженням у форматі Pickle з розширенням ".h5". Ефективність класифікації 300 окремо збережених файлів перевірялася за допомогою F1-міри, повноти (recall) та точності (accuracy). Час навчання моделей подано на рисунку 4.5 у секундах.

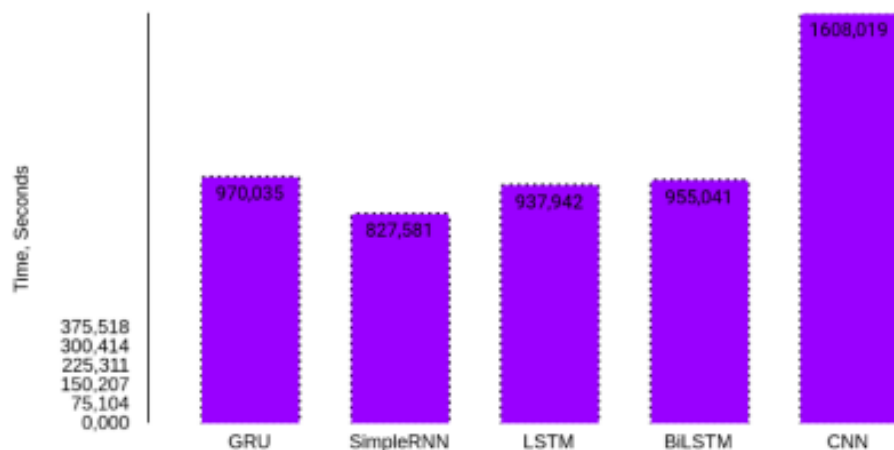


Рисунок 4.5 – Час навчання з запропонованими моделями

Для оцінки практичної придатності можливостей розпізнавання типів комірок RNN усі шари архітектури моделей RNN залишалися незмінними після навчання. Було проведено два експериментальні тести на цих навчених RNN-моделях. Під час навчання кожної моделі були сформовані відповідні архітектури. Оскільки пріоритет роботи дослідження архітектур, що складаються з RNN-мереж, на рисунках нижче показано архітектуру моделей RNN після навчання.

```

Model: "SimpleRNN"
Layer (type)                Output Shape                Param #                    Connected to
-----
stft_3_input (InputLayer)    [(None, 16000, 1)]         0
stft_3 (STFT)                (None, 100, 257, 1)       0                          stft_3_input[0][0]
magnitude_3 (Magnitude)      (None, 100, 257, 1)       0                          stft_3[0][0]
apply_filterbank_3 (ApplyFilter) (None, 100, 128, 1)       0                          magnitude_3[0][0]
magnitude_to_decibel_3 (Magnitu (None, 100, 128, 1)       0                          apply_filterbank_3[0][0]
batch_norm (LayerNormalization) (None, 100, 128, 1)       256                        magnitude_to_decibel_3[0][0]
reshape (TimeDistributed)    (None, 100, 128)          0                          batch_norm[0][0]
td_dense_tanh (TimeDistributed) (None, 100, 64)          8256                       reshape[0][0]
SimpleRNN (SimpleRNN)        (None, 100, 32)           3184                       td_dense_tanh[0][0]
skip_connection (Concatenate) (None, 100, 96)          0                          td_dense_tanh[0][0]
SimpleRNN[0][0]
dense_1_relu (Dense)          (None, 100, 64)           6288                       skip_connection[0][0]
max_pool_id (MaxPooling1D)    (None, 50, 64)            0                          dense_1_relu[0][0]
dense_2_relu (Dense)          (None, 50, 32)            2080                       max_pool_id[0][0]
flatten (Flatten)             (None, 1600)              0                          dense_2_relu[0][0]
dropout (Dropout)            (None, 1600)              0                          flatten[0][0]
dense_3_relu (Dense)          (None, 32)                51232                      dropout[0][0]
softmax (Dense)               (None, 3)                 99                         dense_3_relu[0][0]
-----
Total params: 71,235
Trainable params: 71,235
Non-trainable params: 0
  
```

Рисунок 4.6 – Архітектура, отримана при компіляції простої RNN-мережі

```

Model: "LSTM"
Layer (type)                Output Shape                Param #                    Connected to
-----
stft_2_input (InputLayer)    [(None, 16000, 1)]         0
stft_2 (STFT)                (None, 100, 257, 1)       0                          stft_2_input[0][0]
magnitude_2 (Magnitude)      (None, 100, 257, 1)       0                          stft_2[0][0]
apply_filterbank_2 (ApplyFilter) (None, 100, 128, 1)       0                          magnitude_2[0][0]
magnitude_to_decibel_2 (Magnitu (None, 100, 128, 1)       0                          apply_filterbank_2[0][0]
batch_norm (LayerNormalization) (None, 100, 128, 1)       256                        magnitude_to_decibel_2[0][0]
reshape (TimeDistributed)    (None, 100, 128)          0                          batch_norm[0][0]
td_dense_tanh (TimeDistributed) (None, 100, 64)          8256                       reshape[0][0]
LSTM (LSTM)                  (None, 100, 32)           12416                      td_dense_tanh[0][0]
skip_connection (Concatenate) (None, 100, 96)          0                          td_dense_tanh[0][0]
LSTM[0][0]
dense_1_relu (Dense)          (None, 100, 64)           6288                       skip_connection[0][0]
max_pool_id (MaxPooling1D)    (None, 50, 64)            0                          dense_1_relu[0][0]
dense_2_relu (Dense)          (None, 50, 32)            2080                       max_pool_id[0][0]
flatten (Flatten)             (None, 1600)              0                          dense_2_relu[0][0]
dropout (Dropout)            (None, 1600)              0                          flatten[0][0]
dense_3_relu (Dense)          (None, 32)                51232                      dropout[0][0]
softmax (Dense)               (None, 3)                 99                         dense_3_relu[0][0]
-----
Total params: 80,547
Trainable params: 80,547
Non-trainable params: 0
  
```

Рисунок 4.7 – Архітектура, отримана при компіляції

```

Model: "BiLSTM"
-----
Layer (type)                Output Shape          Param #   Connected to
-----
stft_2_input (InputLayer)   [(None, 16000, 1)]  0        stft_2_input[0][0]
stft_2 (STFT)               (None, 100, 257, 1) 0        stft_2_input[0][0]
magnitude_2 (Magnitude)     (None, 100, 257, 1) 0        stft_2[0][0]
apply_filterbank_2 (ApplyFilter) (None, 100, 128, 1) 0        magnitude_2[0][0]
magnitude_to_decibel_2 (Magnitu) (None, 100, 128, 1) 0        apply_filterbank_2[0][0]
batch_norm (LayerNormalization) (None, 100, 128, 1) 256       magnitude_to_decibel_2[0][0]
reshape (TimeDistributed)   (None, 100, 128)    0        batch_norm[0][0]
td_dense_tanh (TimeDistributed) (None, 100, 64)    8256      reshape[0][0]
bidirectional_lstm (Bidirection) (None, 100, 64)    24832     td_dense_tanh[0][0]
skip_connection (Concatenate) (None, 100, 128)   0        td_dense_tanh[0][0]
bidirectional_lstm[0][0]
dense_1_relu (Dense)        (None, 100, 64)    8256      skip_connection[0][0]
max_pool_1d (MaxPooling1D)  (None, 50, 64)    0        dense_1_relu[0][0]
dense_2_relu (Dense)        (None, 50, 32)    2080      max_pool_1d[0][0]
flatten (Flatten)          (None, 1600)      0        dense_2_relu[0][0]
dropout (Dropout)          (None, 1600)      0        flatten[0][0]
dense_3_relu (Dense)        (None, 32)        51232     dropout[0][0]
softmax (Dense)            (None, 3)         99        dense_3_relu[0][0]
-----
Total params: 95,811
Trainable params: 95,811
Non-trainable params: 0
-----
Python Console History

```

Рисунок 4.8– Архітектура, отримана при компіляції

```

Model: "GRU"
-----
Layer (type)                Output Shape          Param #   Connected to
-----
stft_2_input (InputLayer)   [(None, 16000, 1)]  0        stft_2_input[0][0]
stft_2 (STFT)               (None, 100, 257, 1) 0        stft_2_input[0][0]
magnitude_2 (Magnitude)     (None, 100, 257, 1) 0        stft_2[0][0]
apply_filterbank_2 (ApplyFilter) (None, 100, 128, 1) 0        magnitude_2[0][0]
magnitude_to_decibel_2 (Magnitu) (None, 100, 128, 1) 0        apply_filterbank_2[0][0]
batch_norm (LayerNormalization) (None, 100, 128, 1) 256       magnitude_to_decibel_2[0][0]
reshape (TimeDistributed)   (None, 100, 128)    0        batch_norm[0][0]
td_dense_tanh (TimeDistributed) (None, 100, 64)    8256      reshape[0][0]
GRU (GRU)                   (None, 100, 64)    24960     td_dense_tanh[0][0]
skip_connection (Concatenate) (None, 100, 128)   0        td_dense_tanh[0][0]
GRU[0][0]
dense_1_relu (Dense)        (None, 100, 64)    8256      skip_connection[0][0]
max_pool_1d (MaxPooling1D)  (None, 50, 64)    0        dense_1_relu[0][0]
dense_2_relu (Dense)        (None, 50, 32)    2080      max_pool_1d[0][0]
flatten (Flatten)          (None, 1600)      0        dense_2_relu[0][0]
dropout (Dropout)          (None, 1600)      0        flatten[0][0]
dense_3_relu (Dense)        (None, 32)        51232     dropout[0][0]
softmax (Dense)            (None, 3)         99        dense_3_relu[0][0]
-----
Total params: 95,139
Trainable params: 95,139
Non-trainable params: 0
-----
Python Console History

```

Рисунок 4.9– Архітектура, отримана при компіляції

На початковому етапі було досліджено графіки точності моделей. Це було необхідно для оцінки надійності навчених моделей на даному наборі даних. Другим етапом стало детальне прогнозування для 300 «односекундних аудіофайлів», які були попередньо збережені окремо. Для точної оцінки здатності розпізнавати окремі класи була побудована матриця плутанини.

У таблиці 4.1 нижче наведено середні значення результатів розпізнавання для RNN-мереж. Також на рисунку 4.10 подано графік точності розпізнавання, отриманої під час навчання на кожній епосі.

Таблиця 4.1 – Порівняння точності моделей SimpleRNN, LSTM, BiLSTM, GRU та CNN на 128-100 вимірних мельспектрограмах

| Навчені моделі | Точність, % |
|---------------------------|-------------|
| Проста RNN (SimpleRNN) | 98 |
| LSTM | 97 |
| Двостороння LSTM (BiLSTM) | 97 |
| GRU з 32 осередками | 98 |
| CNN-структура, як у [1] | 94 |
| GRU з 64 осередками | 98 |

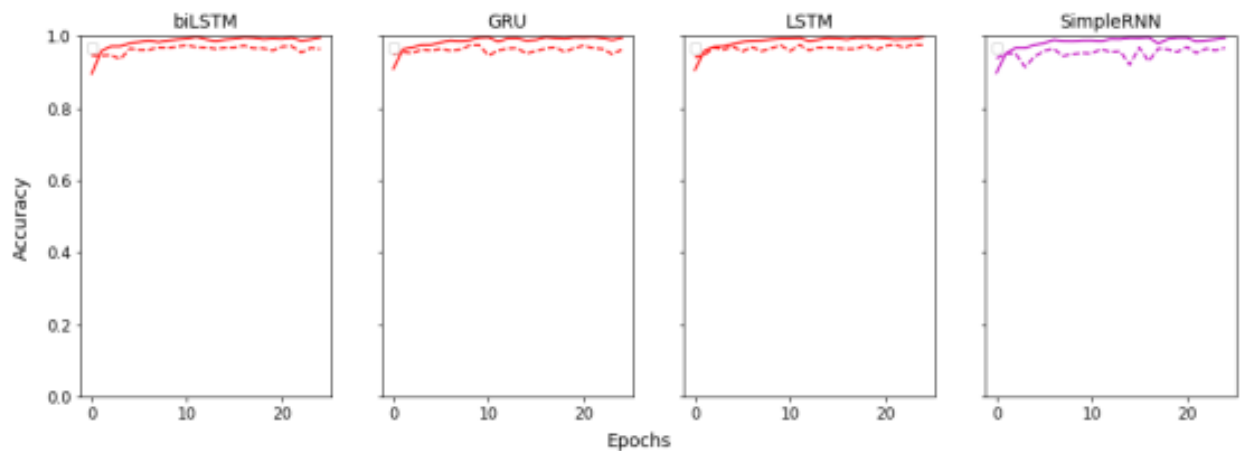


Рисунок 4.10 –Результати навчання

У таблиці 4.1 також наведено результати точності розпізнавання, отримані за допомогою моделі CNN.

ВИСНОВКИ

У цій кваліфікаційній роботі було досліджено проблему розпізнавання звуків безпілотних літальних апаратів (БПЛА) за допомогою методів глибинного навчання. Основна увага приділялася розробці та порівнянню ефективності різних типів рекурентних нейронних мереж (SimpleRNN, LSTM, BiLSTM, GRU), а також згорткової нейронної мережі (CNN).

На першому етапі була проведена обробка та підготовка аудіоданих. Було сформовано датасет з трьох класів: «БПЛА з навантаженням», «БПЛА без навантаження» та «Фоновий шум». Для забезпечення можливості розпізнавання в режимі реального часу всі звукові файли були обрізані до тривалості в одну секунду. Також було розроблено спеціальний фільтраційний блок для стандартизації тривалості записів.

Далі було проведено навчання моделей з поділом даних на тренувальну (70%) і валідаційну (30%) вибірки. Окремо було збережено по 100 тестових файлів на кожен клас для незалежного оцінювання. Навчання моделей здійснювалося з фіксованими архітектурами, після чого було обрано оптимальну кількість епох (25) для досягнення найкращого співвідношення між точністю та стабільністю.

Результати показали, що найкращі результати продемонстрували моделі LSTM і BiLSTM, які краще за інші справлялися із розпізнаванням коротких звукових фрагментів та виявили вищу точність класифікації. CNN-модель також показала конкурентоспроможні результати, однак у завданнях із часовою залежністю рекурентні моделі виявилися ефективнішими.

Проведені експерименти підтверджують, що використання глибинного навчання для акустичного розпізнавання БПЛА є перспективним напрямом. Розроблені моделі здатні забезпечити стабільне розпізнавання у режимі реального часу та можуть бути використані як основа для побудови системи виявлення дронів у прикладних задачах безпеки та моніторингу.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Ezuma M., Erden F., Anjinappa C.K. et al. Micro-UAV Detection and Classification from RF Fingerprints Using Machine Learning Techniques // Proceed. of the IEEE AERO. – Big Sky, MT, USA, 2019. – P. 1-13.
2. Seidaliyeva U., Akhmetov D., Ilipbayeva L. et al. Real-Time and Accurate Detection in a Video with a Static Background // Sensors. – 2020. – Vol. 20, Issue 14. – P. 3856-1-3856-18.
3. Taha B., Shoufan A. Machine Learning-Based Drone Detection and Classification: State-of-the-Art in Research // Proceed. IEEE Access. – 2019. – Vol. 7. – P. 138669-138682.
4. Li S., Kim H., Lee S.D. et al. Convolutional Neural Networks for Analyzing Unmanned Aerial Vehicles Sound // Proceed. 18th internat. conf. on Control, Automation, and Systems (ICCAS). – Daegwallyeong, 2018. – P. 862-866.
5. Vemula H.C. Multiple Drone Detection and Acoustic Scene Classification with Deep Learning. – Dayton: Wright State University, 2018. – 149 p.
6. Delivery drone crashes into power lines, causes outage // https://www.theregister.com/2022/09/30/delivery_drone_crashes_into.
7. When Amazon drones crashed, the company told the FAA to go fly a kite // <https://www.businessinsider.com/amazon-prime-air-faa-regulators>.
8. List of unmanned aerial vehicles-related incidents // https://en.wikipedia.org/wiki/List_of_unmanned_aerial_vehicles-related.
9. Мартовицький В.О., Шеховцов О.В., Алейник Д.С., Пахомова Є.В. та Іванченко Д. І. «ПІДХІД ДО ВИЯВЛЕННЯ ТА КЛАСИФІКАЦІЇ РАДІОКЕРОВАНИХ МОДЕЛЕЙ ЗА ЇХ РАДІОСИГНАЛОМ» Вісник Херсонського національного технічного університету» для розміщення у № 2 (2025).