

УДК 004.93

DOI [https://doi.org/10.24144/2616-7700.2022.41\(2\).118-127](https://doi.org/10.24144/2616-7700.2022.41(2).118-127)**Л. П. Добуляк¹, Д. О. Фербей², С. Б. Костенко³**

¹ Львівський національний університет ім. Івана Франка,
доцент кафедри математичного моделювання соціально-економічних процесів,
кандидат економічних наук, доцент
lesia.dobuliak@lnu.edu.ua
ORCID: <https://orcid.org/0000-0001-8665-8783>

² Львівський національний університет ім. Івана Франка,
магістр,
dariaferbey@gmail.com
ORCID: <https://orcid.org/0000-0003-2329-9693>

³ Львівський національний університет ім. Івана Франка,
доцент кафедри математичного моделювання соціально-економічних процесів,
кандидат фізико-математичних наук, доцент
svitlana.kostenko@lnu.edu.ua
ORCID: <https://orcid.org/0000-0003-0823-9801>

ВИКОРИСТАННЯ ГЛИБИННОГО НАВЧАННЯ У ЗАДАЧАХ КЛАСИФІКАЦІЇ ЗВУКІВ НАВКОЛИШНЬОГО СЕРЕДОВИЩА

У даній статті розглянуто різні аспекти, пов'язані з розпізнаванням звуків навколишнього середовища, що є прикладною задачею в багатьох сферах діяльності людини. На відміну від музики та мови, звук навколишнього середовища насичений шумом і не має ритму та мелодії музики або семантичної послідовності мови. Це ускладнює пошук спільних рис серед звукових сигналів навколишнього середовища. У даному дослідженні розв'язання задачі розпізнавання звуків базуються на використанні методів класифікації зображень. Для цього виконується перетворення кожного аудіо-запису вибірки у спеціальні зображення — спектрограму Мела, що є його компактним інформативним візуальним представленням. Щоб підвищити точність розпізнавання звуків, досліджуються різні методи збільшення навчального набору даних. В основі цих методів лежить створення нових екземплярів аудіозаписів шляхом деформації існуючих. За допомогою такого підходу ми можемо збільшити в рази кількість елементів набору даних, таким чином вирішити проблему його обмеженості. Для класифікації звуків навколишнього середовища з набору аудіоданих UrbanSound8K було використано глибинну згорткову нейронну мережу. Щоб оцінити якість (точність та втрати) представленої моделі було застосовано 10-кратну перехресну перевірку.

Ключові слова: класифікація, звуки навколишнього середовища, збільшення даних, спектрограма, згорткова нейронна мережа.

1. Вступ. Сучасні методи розпізнавання образів охоплюють методи класифікації та ідентифікації предметів, явищ, процесів, сигналів, ситуацій та подібних об'єктів, які характеризуються наборами деяких властивостей і ознак. Одним з важливих напрямків у цій області є розпізнавання звуків навколишнього середовища, незалежно від їх природи.

Однак через складність звуків навколишнього середовища, їх взаємного накладання та низки інших причини, дослідження щодо розпізнавання таких звуків значно відстали від вивчення мовлення та музики. Розпізнавання звуку безпосередньо впливає на виявлення різних явищ та процесів навколо нас і може широко використовуватися в багатьох галузях, таких як керування пристроями

без участі людини (автопілотування транспортних засобів, управління розумними будинками тощо), розпізнавання сигналів тривоги, ідентифікація збоїв у роботі механізмів та ін. А тому побудова надійних моделей, які здатні ідентифікувати в шумі певні звукові сигнали, представляє великий дослідницький інтерес.

2. Постановка завдання. Основна ідея роботи полягає в тому, щоб побудувати модель глибокого навчання, яка може класифікувати деякі визначені звуки навколишнього середовища.

Для досягнення зазначеної мети необхідно вивчити властивості звуку, особливості його подання у вигляді спектрограми — спеціального виду зображення, виконати навчання згорткової нейронної мережі на вибірці уривків звуків навколишнього середовища та оцінити якість проведеної класифікації. Щоб навчання мережі відбувалося на достатній за обсягом та репрезентативній тренувальній вибірці, дослідити методи збільшення набору використовуваних даних.

3. Огляд літератури. Поняття звуку, його особливості як фізичного явища, характеристики та можливість оцифрування подано в роботі Радзишевського А. Ю. [1]. Машинному навчанню та різним підходам до розпізнавання образів присвячені видання Кутковецького В. Я. [2] та Рашкевича Ю. М. [3]. Що таке глибоке навчання, глибокі нейронні мережі як засоби класифікації об'єктів та особливості застосування згорткових нейронних мереж для розпізнавання зображень детально викладено в працях Субботіна С. О. [4] та Новотарського М. А. [5].

Можливість переходу класифікації аудіо до класифікації звуків, особливості початкової підготовки аудіоданих, доцільність використання мел-спектрограми, а також структуру згорткової нейронної мережі у випадку класифікації звуків описано в [6]. Побудові згорткової нейронної мережі, яка здатна розпізнавати шум на будівельному майданчику присвячена робота [7]. Тут використовувалися аудіозаписи роботи різних екскаваторів. Мета такого дослідження — це виявлення небезпеки на робочому місці та моніторинг діяльності учасників будівництва. Характеристику колекціям аудіозаписів, придатним для використання у машинному навчанні, дано у [8].

З документацією по пакету Keras мови програмування Python, який містить інструментарій для обробки зображень та тексту, можна ознайомитися на офіційному сайті [9], а особливості його застосування щодо перетворення аудіоданих у спектрограму та виконання іншої підготовчої роботи з вхідними даними подані в [10].

4. Матеріали та методи. Наше сьогоденне навколишнє середовище переповнене різноманіттям звуків. Все, крім голосу та музики, у цьому різноманітті, можна віднести до особливого виду аудіо, серед якого є і звуки природного світу, наприклад, дзижчання комара, так і звуки роботи технічних пристроїв та механізмів, на кшталт гудіння холодильника.

Оскільки завданням цього дослідження є класифікація звуків, то необхідно розібратися з цим фізичним явищем та з'ясувати способи його опису та оцифрування з метою застосування глибокого навчання.

Ми знаємо, що звук — це коливальний рух частинок середовища, що поширюється у вигляді хвиль у різних середовищах та сприймається слухом. Часто звуковий сигнал повторюється через рівні проміжки часу, тому хвиля, що його

описує, має повторювальну (регулярну, періодичну) форму. Висота хвилі вказує на інтенсивність звуку та називається амплітудою, час, необхідний для того, щоб сигнал завершив одну повну хвилю, є періодом (рис. 1). Оберненою до періоду величиною є частота — це кількість періодів за одиницю часу.

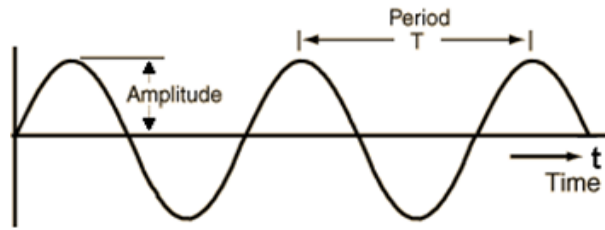


Рис. 1. Простий повторюваний сигнал з його амплітудою та часом.

Більшість звуків, які ми зустрічаємо, можуть не відповідати таким простим і регулярним періодичним моделям. Сигнали різних частот додаються разом, утворюючи сигнали з більш складними повторюваними шаблонами. Усі звуки, які ми чуємо, включаючи наш власний людський голос, складаються з таких сигналів. Прикладом такого складного сигналу є звук музичного інструменту, що зображено на рисунку 2.



Рис. 2. Форма хвилі складного повторюваного сигналу звуку музичного інструменту.

Щоб оцифрувати звукову хвилю, потрібно перетворити сигнал у серію чисел, шляхом вимірювання амплітуди звуку через фіксовані проміжки часу. Цей процес представлено на рисунку 3.

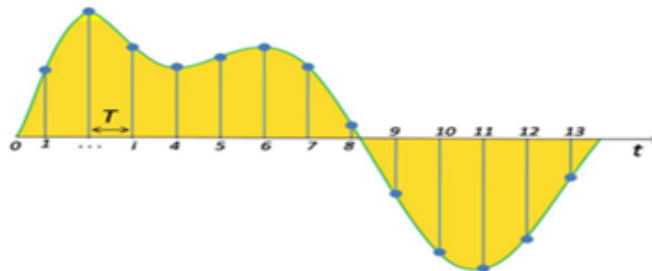


Рис. 3. Вимірювання амплітуди звуку через регулярні проміжки часу.

Результати кожного такого вимірювання називається вибіркою, а частота дискретизації — це кількість вибірок за секунду. Наприклад, загальна частота

дискретизації становить близько 44 100 вибірок в секунду. Це означає, що 10-секундний музичний кліп матиме 441 000 значень.

Для візуального зображення спектру частот сигналу в часі використовуються спектрограма [11]. Як правило, спектрограма — це двовимірна діаграма, де по горизонтальній осі представлено час, по вертикальній осі — частота коливань, а амплітуда на певній частоті в конкретний момент часу представлена інтенсивністю, тобто кольором відповідної точки. Спектрограма використовує різні кольори для позначення величини амплітуди або сили кожної частоти: чим яскравіше колір, тим вище енергія сигналу.

Спектрограми формуються із звукових сигналів за допомогою перетворень Фур'є [12]. Перетворення Фур'є розкладає сигнал на його складові частоти і визначає амплітуду кожної частоти, присутньої в сигналі. Таким чином, спектрограма розбиває тривалість звукового сигналу на менші часові сегменти, а потім застосовує перетворення Фур'є до кожного сегмента, щоб визначити частоти і на завершення об'єднує перетворення Фур'є для всіх цих сегментів в єдиний графік.

Кожен вертикальний «зріз» спектрограми, по суті, є спектром сигналу в цей момент часу і показує, як потужність сигналу розподіляється по кожній частоті, знайденій у сигналі в цей момент.

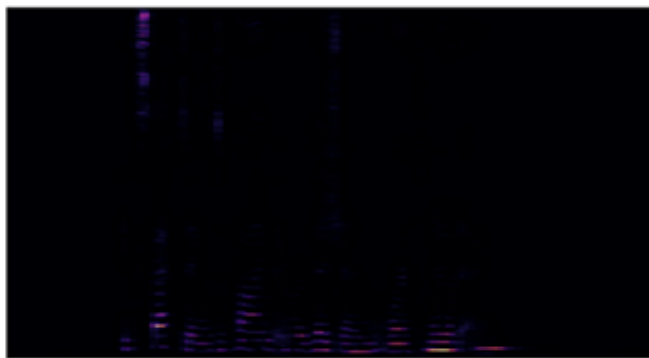


Рис. 4. Звичайна спектрограма.

Як ми бачимо, спектрограма не дає нам багато інформації (рис. 4). Це відбувається через те, що люди сприймають звук у вузькому діапазоні частот і амплітуд. Для збільшення наглядності спектрограми була розроблена шкала Мела, у відповідності до якої кількісна оцінка звуку базується на статистичній обробці даних про суб'єктивне сприйняття висоти звукових тонів великої кількості слухачів [13]. Тобто, в її основі є психо-фізичне сприйняття звуку людиною, і вона є логарифмічною залежністю от частоти.

На основі шкали Мела будується спектрограма Мела (рис. 5), яка є аналогом звичайної спектрограми, проте використовує на вертикальній осі замість частот спеціальні величини мели, перехід до яких здійснюється за допомогою застосування до спектрограми мел-фільтрів [14].

Очевидно, що мел-спектрограма є більш інформативною ніж звичайна спектрограма. Таким чином можна зробити висновок, що спектрограма Мела є компактним інформативним представленням аудіосигналу, схожим на відбиток пальця людини.

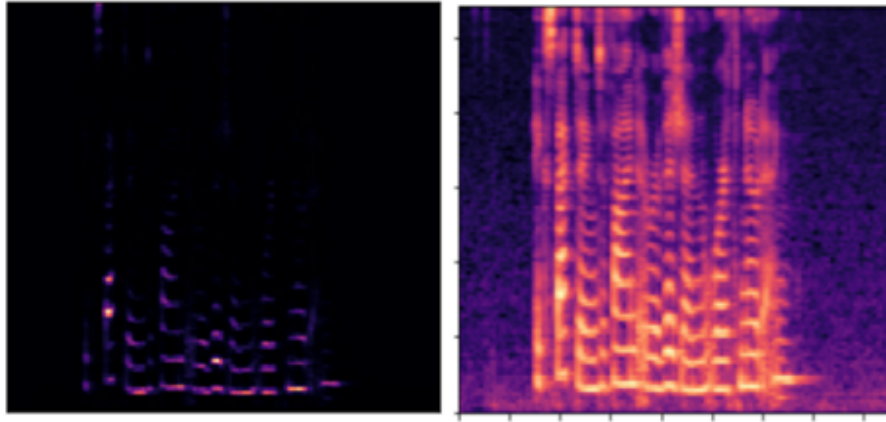


Рис. 5. Звичайна спектрограма та побудована на її основі спектрограма Мела.

Глибинне навчання, як галузь машинного навчання, передбачає моделювання високорівневих абстракції із декількома обробними шарами, де здійснюються різні перетворення. Існують різні архітектури глибинного навчання, серед яких вагоме місце займають згорткові нейронні мережі.

Згорткові нейронні мережі — це клас глибинних штучних нейронних мереж прямого поширення, які можуть розпізнавати й класифікувати зображення, а тому широко використовуються для аналізу візуальних даних [15]. Термін «згортка» в позначає математичну функцію згортки, яка є особливим видом лінійної операції, в якій дві функції перемножуються, щоб отримати третю функцію, яка виражає, як форма однієї функції змінюється іншою. Простіше кажучи, два зображення, які можна представити у вигляді матриць, перемножуються, щоб отримати вихід, який використовується для вилучення ознак із зображення.

Якість побудови будь-якої моделі машинного навчання залежить від репрезентативності навчальної вибірки. Основними проблемами й перешкодами для дослідження міських звуків є відсутність у відкритому доступі колекцій з достатньою кількістю маркованих аудіоданих для тренування нейронної мережі. Для урізноманітнення наявних екземплярів звуків та наповнення таких колекцій використовуються різні методики штучного створення даних на основі існуючих [16]. Ці методики можна застосувати як до необробленого аудіозапису перед створенням спектрограми, так і до його спектрограми.

В даному дослідженні використовуються такі дві методики:

- додавання фоновий гауссовий (білий) шуму [17] до спектрограми,
- зміщення часу: зсуває спектрограму у часі вправо, при цьому та частина спектрограми, яка попадає поза фіксований розмір часової рамки, відсікається.

При генеруванні нового зразка звуку можливі застосування різних комбінацій описаних методів до одної і тієї ж вхідної спектрограми. Результат застосування різних комбінацій методів для створення нових екземплярів, що застосовані до спектрограми звуку гавкання собаки, показаний на рисунку 6.

5. Основний результат. Проведене дослідження базувалось на вибірці уривків звуків UrbanSound8K [18]. Це набір даних містить 8732 аудіофай-

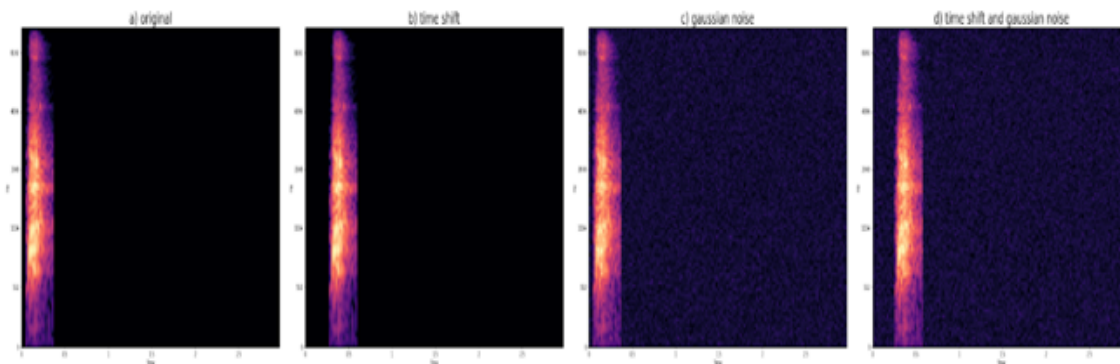


Рис. 6. Результати застосування до спектрограми звуку гавкання собаки методів створення нових екземплярів.

лів, які відносяться до одного з 10 різних класів (air_conditioner, car_horn, children_playing, dog_bark, drilling, engine_idling, gun_shot, jackhammer, siren, street_music).

Збільшення даних було виконано в режимі онлайн шляхом застосування методів додавання гауссового шуму та зміщення часу до вхідних спектрограм перед тренуванням моделі. Обидва методи застосовують випадково згенеровані значення параметрів деформації зображень (спектограм).

Представлений набір даних було збільшено та перетворено у відповідні спектрограми Мела. Приклади спектрограм Мела звуку кондиціонеру, сигналу автомобіля та звуки гри дітей із зазначеного набору подано на рисунку 7.

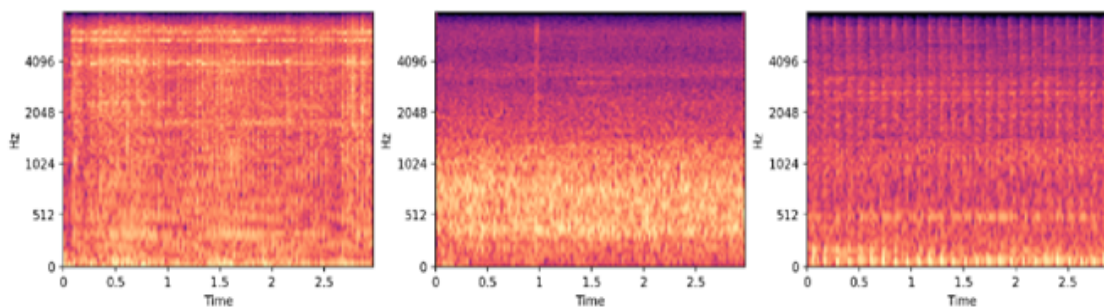


Рис. 7. Приклади спектрограм Мела.

Для обробки цих зображень було використано модель згорткової нейронної мережі, яка містить: Conv2D (згортковий шар), MaxPooling2D (максимальне об'єднання), Activation (функція активації) та Dense Layer (щільний шар) [19].

Згортковий шар безпосередньо має справу з вхідними даними. Вхід кожного шару Conv2D і MaxPooling2D є 3D тензори (висота, ширина, RGB-канал кольору). Застосування згорткового шару призводить до зменшення розмірів зображення.

Структура використаної згорткової нейронної мережі, особливості параметрів виводу результатів роботи кожного шару мережі та їх кількість зображені

```

Model: "sequential"

```

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 124, 124, 24)	624
max_pooling2d (MaxPooling2D)	(None, 41, 41, 24)	0
activation (Activation)	(None, 41, 41, 24)	0
conv2d_1 (Conv2D)	(None, 38, 38, 36)	13860
max_pooling2d_1 (MaxPooling2D)	(None, 19, 19, 36)	0
activation_1 (Activation)	(None, 19, 19, 36)	0
conv2d_2 (Conv2D)	(None, 17, 17, 48)	15600
activation_2 (Activation)	(None, 17, 17, 48)	0
global_average_pooling2d (GlobalAveragePooling2D)	(None, 48)	0
dense (Dense)	(None, 60)	2940
dropout (Dropout)	(None, 60)	0
dense_1 (Dense)	(None, 10)	610

```

Total params: 33,634
Trainable params: 33,634
Non-trainable params: 0

```

Рис. 8. Архітектура згорткової нейронної мережі.

на рисунку 8.

Щоб оцінити точність моделі, використали 10-кратну перехресну перевірку. Для оцінки використали такі показники як точність та втрата. Для функції втрат вибирається категоріальна втрата перехресної ентропії, а точність — це відсоток правильно класифікованих екземплярів. Для кожного класу втрата визначається як мінімальне значення втрат серед всіх епох у процесі перевірки. Аналогічно, точність для кожного класу розраховується шляхом отримання найкращого значення точності класифікації у кожну епоху.

Розглянемо приклад роботи моделі — класифікацію звуків з першої групи набору даних (звуків кондиціонера). Порівняємо точність та втрати на тестових та на тренувальних даних. На рисунку 9 зображено залежність точності та втрати від кількості епох тренування моделі. Синя крива — це результати для тренувальної вибірки, а червона для тестувальної.

Точність, розрахована програмою, становить 81%, а величина втрат стано-

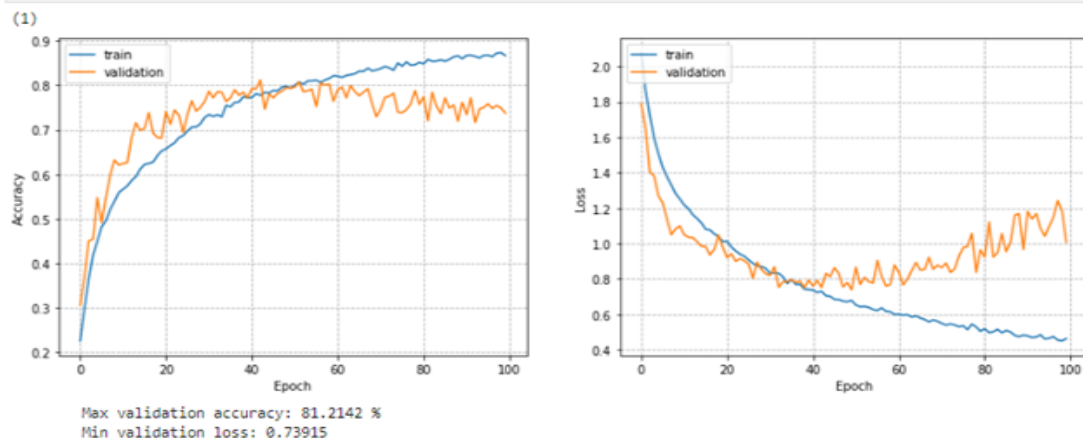


Рис. 9. Оцінка якості роботи згорткової нейронної мережі щодо класифікації звуків кондиціонера.

вить 0,74, що засвідчують графіки на рисунку 9.

6. Висновки та перспективи подальших досліджень. Під час дослідження було встановлено, що згорткові нейронні мережі, розроблені спеціально для розпізнавання зображень, можна успішно навчити класифікувати звуки навколишнього середовища використовуючи спектральне представлення відповідних аудіозаписів. Розроблено модель глибокого навчання, яка може класифікувати звуки навколишнього середовища кількох різних класів. Експерименти з набором даних UrbanSound8K, що містить звуки десяти класів, показують, що середня точність й середня втрата при класифікації об'єктів кожного з класів, досягають значень 75% і 92,4% відповідно.

Розроблена модель має як переваги, так й недоліки. Перевагою є низька складність структури нейронної мережі в порівнянні з іншими методами класифікації звуків. Недоліком залишається відносно невелика точність. Тому, у подальшому можна продовжити вивчати різні підходи для класифікації звуків з метою покращення точності.

Дана робота виокремила перспективи подальших досліджень. Зокрема потребує додаткової уваги пошук ефективних методів збільшення даних, які доцільно застосовувати для різних видів звуків, з метою їх використання для побудови моделей класифікації. Крім того, існує потреба дослідити вплив на якість класифікації звуків застосування нейронних мереж різної топології, взявши до уваги особливості їх реалізації та впровадження.

Список використаної літератури

1. Радзишевский А. Ю. Основы аналогового и цифрового звука. Москва : Вильямс, 2006. 281 с.
2. Кутковецкий В. Я. Розпізнавання образів : навчальний посібник. Миколаїв : Вид-во ЧНУ ім. Петра Могили, 2017. 420 с.
3. Рашкевич Ю. М., Ткаченко Р. О., Цмоць І. Г., Пелешко Д. Д. Нейро-подібні методи, алгоритми та структури обробки зображень у реальному часі: монографія. Львів: Вид-во Львів. політехніки, 2014. 256 с.
4. Субботін С. О. Нейронні мережі. Теорія та практика : навч. посіб. Житомир : Вид. О. О. Євенок, 2020. 184 с.

5. Новотарський М. А., Нестеренко Б. Б. Штучні нейронні мережі : обчислення. Праці Інституту математики НАН України. Т. 50. Київ : Ін-т математики НАН України, 2004. 408 с.
6. Palanisamy K., Singhania D., Yao A. Rethinking CNN Models for Audio Classification. URL: <https://arxiv.org/pdf/2007.11154.pdf> (дата звернення: 06.09.2022).
7. Maccagno A., Mastropietro A., Mazziotta U., Scarpiniti M., Lee Y.-Ch., Uncini A. A CNN Approach for Audio Classification in Construction Sites. URL: https://iris.uniroma1.it/bitstream/11573/1444275/2/Maccagno_postprint_CNN-Approach_2021.pdf (дата звернення: 06.09.2022).
8. Varazida N. 40 Open-Source Audio Datasets for ML. URL: <https://towardsdatascience.com/40-open-source-audio-datasets-for-ml-59dc39d48f06> (дата звернення: 10.09.2022).
9. Keras API reference. URL: <https://keras.io/api/> (дата звернення: 10.09.2022).
10. Lillelund C. Classify MNIST Audio using Spectrograms-Keras CNN. URL: <https://www.kaggle.com/code/christianlillelund/classify-mnist-audio-using-spectrograms-keras-cnn/data> (дата звернення: 06.09.2022).
11. Voddatpati V., Petef A., Rasmusson J., Lundberg L. Classifying environmental sounds using image recognition networks. *Procedia computer science*. 2020. 112. P. 2048–2056.
12. Перелигін Б. В., Ткач Т. Б., Гор'єв С. А. Спектрально-часовий аналіз даних моніторингу. Одеса : ТЕС, 2018. 15 с.
13. Cotton C. V., Ellis D. Spectral vs. spectro-temporal features for acoustic event detection. *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*. IEEE. 2011. P. 69–72.
14. Swedia E. R., Mutiara A. B., Subali M., and Ernastuti. "Deep Learning Long-Short Term Memory (LSTM) for Indonesian Speech Digit Recognition using LPC and MFCC Feature", in 2018 Third International Conference on Informatics and Computing (ICIC), Palembang, Indonesia. 2018. P. 1–5.
15. Bahmei B., Birmingham E., Arzanpour S. Cnn-rnn and data augmentation using deep convolutional generative adversarial network for environmental sound classification. *IEEE Signal Processing Letters*, 2022. 29. P. 682–686.
16. Gaussiannoise. URL: <https://hasty.ai/docs/mp-wiki/augmentations/gaussian-noise> (дата звернення: 10.09.2022).
17. Salamon J., Jacoby C., Bello J. P. "UrbanSound8k Dataset". Urban Sound Datasets. URL: <https://urbansounddataset.weebly.com/urbansound8k.html> (дата звернення: 06.09.2022).
18. Salamon J., Bello J. P. "Deep Convolutional Neural Networks and Data Augmentation for Environmental Sound Classification submitted. 2016. URL: <https://arxiv.org/pdf/1608.04363.pdf> (дата звернення: 06.09.2022).

Dobuliak L. P., Ferbei D. O., Kostenko S. B. Using deep learning for environmental sound's classification.

Environmental sound recognition is an applied problem in many spheres of human activity. Unlike music and speech, ambient sound is saturated with noise and lacks the rhythm and melody of music or the semantic consistency of speech. This makes it difficult to find common features among environmental sound signals. In order to improve the accuracy of sound recognition, this paper proposes different methods for increasing the training data set. The basis of these methods is the creation of new instances of audio recordings by deformation of existing ones. Using this approach, we can increase the number of data set elements several times and solve the problem of its limitation. In order to convert the data sample into an image, the Mel spectrogram is used, which is a compact informative representation of the audio signal, which is convenient to use in further work. A deep convolutional neural network was used to classify environmental sound from the UrbanSound8K audio dataset. A 10-fold cross-validation was applied to assess the quality (accuracy and loss) of the presented model.

Keywords: classification, environmental sounds, data augmentation, spectrogram, convolutional neural network.

References

1. Radzyshevsky, A. Yu. (2006). *Fundamentals of analog and digital audio*. Moskow: Williams [in Russian].
2. Kutkovetskyj, V. Ya. (2017). *Recognition of images*. Mykolaiv: [in Ukrainian].
3. Rashkevych, Yu. M., Tkachenko, R. O., Cmoc, I. G., & Peleshko, D. D. (2014). *Neural-like methods, algorithms and frameworks for real-time image processing*. Lviv [in Ukrainian].
4. Subbotin, S. O. (2020). *Neural networks. Theory and practice*. Jytomyr [in Ukrainian].
5. Novotarskyj, M. A., & Nesterenko, B. B. (2004). Artificial neural networks: calculations. *Proceedings of the Institute of Mathematics of the National Academy of Sciences of Ukraine*. (Vol. 50). Kyiv [in Ukrainian].
6. Palanisamy, K., Singhania, D., & Yao, A. Rethinking CNN Models for Audio Classification. Retrieved from: <https://arxiv.org/pdf/2007.11154.pdf>
7. Maccagno, A., Mastropietro, A., Mazziotta, U., Scarpiniti, M., Lee, Y.-Ch., & Uncini, A. A CNN Approach for Audio Classification in Construction Sites. Retrieved from: https://iris.uniroma1.it/bitstream/11573/1444275/2/Maccagno_postprint_CNN-Approach_2021.pdf
8. Barazida, N. 40 Open-Source Audio Datasets for ML. Retrieved from: <https://towardsdatascience.com/40-open-source-audio-datasets-for-ml-59dc39d48f06>
9. Keras API reference. Retrieved from: <https://keras.io/api/>
10. Lillelund, C. Classify MNIST Audio using Spectrograms/Keras CNN. Retrieved from: <https://www.kaggle.com/code/christianlillelund/classify-mnist-audio-using-spectrograms-keras-cnn/data>
11. Boddapati, V., Petef, A., Rasmusson, J., & Lundberg, L. (2020). Classifying environmental sounds using image recognition networks. *Procedia computer science*, 112, 2048–2056.
12. Perelygin, B. V., Tkach, T. B., & Gorjev, S. A. (2018). *Spectral-temporal analysis of monitoring data*. Odesa [in Ukrainian].
13. Cotton, C. V. & Ellis, D. (2011). Spectral vs. spectro-temporal features for acoustic event detection. *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*. *IEEE*, 69–72.
14. Swedia, E. R., Mutiara, A. B., Subali, M., & Ernastuti. (2018). Deep Learning Long-Short Term Memory (LSTM) for Indonesian Speech Digit Recognition using LPC and MFCC Feature. *Third International Conference on Informatics and Computing (ICIC)*. Palembang. Indonesia, 1–5.
15. Bahmei, B., Birmingham, E., & Arzanpour, S. (2022). Cnn-rnn and data augmentation using deep convolutional generative adversarial network for environmental sound classification. *IEEE Signal Processing Letters*, 29, 682–686.
16. Gaussiannoise. Retrieved from: <https://hasty.ai/docs/mp-wiki/augmentations/gaussian-noise>
17. Salamon, J., Jacoby, C., & Bello, J. P. "UrbanSound8k Dataset". Urban Sound Datasets. Retrieved from: <https://urbansounddataset.weebly.com/urbansound8k.html>
18. Salamon, J., & Bello, J. P. (2017). "Deep Convolutional Neural Networks and Data Augmentation for Environmental Sound Classification", submitted. Retrieved from: <https://arxiv.org/pdf/1608.04363.pdf>

Одержано 13.10.2022