

Володимир Кучинський, Наталія Процах*

¹ НЛТУ України, Львів, Україна, 23kuchynskyi.v@nltu.lviv.ua

² НЛТУ України, Львів, Україна, ORCID: 0000-0002-3034-5902,
protsakh@nltu.edu.ua

Використання технологій машинного навчання для ідентифікації видів птахів за голосом на прикладі моделі BirdNET

Анотація. В даній роботі досліджено перспективи використання технологій машинного навчання та оброблення звукових сигналів з метою автоматичної ідентифікації видів птахів за голосами. Здійснено огляд найновіших досягнень в цій галузі. Проаналізовано існуючі технологічні рішення для полегшення визначення птахів в польових умовах. Створено набір даних аудіозаписів голосів птахів України на основі репозиторію Xeno-canto Розгорнуто модель BirdNET та досліджено її роботу на обмеженій регіональній вибірці у розрізі окремих видів та таксономічних груп.

Ключові слова – визначення птахів, вид, обробка сигналів, машинне навчання, нейронна мережа

В останні роки з'явилася низка наукових публікацій та практичних проєктів, пов'язаних із застосуванням різних інформаційних технологій машинного навчання та оброблення сигналів за допомогою конволюційних нейромереж (CNN) для автоматичної ідентифікації птахів за зображеннями чи звуковими сигналами. Деякі з запропонованих моделей на сьогодні є досить ефективними, але точність визначення окремих видів залежить від різних факторів, зокрема регіону, систематичної групи птахів, якості зображення чи звуку тощо. Тому залишається актуальним пошук шляхів вдосконалення існуючих моделей з урахуванням регіональних та таксономічних особливостей, а також створення на їх базі зручних інтерактивних інструментів для автоматичного визначення птахів в польових умовах.

Конволюційні нейронні мережі (CNN) відіграють ключову роль у обробці і розпізнаванні зображень. Оскільки спектрограма це теж зображення, яке представляє частотно-часову характеристику вхідного сигналу, то методи з сучасного розпізнавання зображень можна застосовувати і до класифікації по аудіодорі-

жці. Відповідно, можна використати вже існуючі передові архітектури моделей для класифікації зображень (такі як EfficientNet) у задачі класифікації аудіодоріжок. Саме такий підхід було реалізовано у моделі BirdNET.

Характеристика моделі. BirdNET на сьогодні вважається одним з найбільш ефективних інструментів автоматичної ідентифікації птахів за голосами. На базі цієї моделі створено мобільний додаток *BirdNET Sound ID* для Android та IOS, а також дослідницьку платформу *BirdNET Analyzer* для автоматичного оброблення звукозаписів, зібраних за допомогою автономних записуючих пристроїв в процесі пасивного акустичного моніторингу. BirdNET є спільним проектом Центру природоохоронної біоакустики К. Лізи Янг при Корнельській лабораторії орнітології та Технологічного університету Хемніца. Остання версія моделі 2023 року вже може ідентифікувати понад 6000 видів птахів з усього світу [2].

Програма BirdNET забезпечує автоматизоване визначення видів птахів на основі сонограм. Користувач завантажує аудіозапис голосу птаха, який перетворюється на спектрограму. Модель аналізує звукозапис, розбивши його на 3-секундні інтервали, і для кожного інтервалу повертає назву виду або видів та значення впевненості (у %) [1]. Для тренування моделі використано головним чином аудіозаписи голосів птахів з цифрових репозитаріїв Xeno-canto та бібліотеки Маколея (Macaulay Library) (понад 90% всіх даних), ряд наборів даних звукового ландшафту зі змагань BirdCLEF, а також частину записів з мобільного додатку [4].

В цій роботі, модель BirdNET взята за основу для розгортання і проведення дослідів на даних. Її було обрано оскільки це найкраще задокументована і публічно доступна модель, з хорошою точністю та практикою розгортання на мобільних пристроях. Також у репозитарії моделі описано бібліотеку *birdnetlib*, яка суттєво спрощує розгортання натренованих моделей [2].

Модель BirdNET останньої доступної версії станом на зараз (2.4) працює наступним чином:

1. спочатку відбувається обробка даних. Модель натреновано для класифікації на аудіодоріжках фіксованої довжини (3 секунди) з частотою дискретизації 48 КГц, тобто на вхід подається вектор довжиною 144 тис. відліків. Через це, аудіофайли більшої довжини розбиваються на 3-секундні інтервали, кожен з яких окремо подається на вхід моделі;
2. архітектура моделі виглядає наступним чином:
 - 2.1. нормалізація значень до діапазону від -1 до 1;
 - 2.2. шар MelSpecLayerSimple: обчислення двох спектрограм Мела: одна з діапазоном частот [500, 1500], інша [0, 3000]. Конкатенація результатів у двоканальний вихід, тензор розміром $96 \times 511 \times 2$. Довжина вікна для віконного перетворення Фур'є (STFT): 1024.;
 - 2.3. конволюційний шар, а також шар batch normalization, max pooling і average pooling. Після цих шарів тензор має бути того розміру, який прийнятний для першого блоку EfficientNetB0;
 - 2.4. efficientnetb0 блоки: це основна частина (backbone) CNN моделі для класифікації зображень. Використано 4 блоки, з певними змінами до архітектури порівняно з класичною реалізацією EfficientNetB0, таких як інші розмірності шарів, проте основні ідеї цієї архітектури (відповідні розміри та MBConv структуру);
 - 2.5. замінений конволюційний шар а також global average pooling, які визначають розмір ембедингу (в останній версії це 1024);
 - 2.6. останній шар: лінійний шар, розмір якого відповідає кількості класів, які розпізнає модель (6522 в останній версії). Значення на кожному з нейронів відповідають впевненості що на аудіодоріжці було чути птаха саме цього виду. При дотреновуванні моделі (transfer learning), ваги цього шару занулюються та вивчаються наново, а розмірність змінюється і відповідає кількості класів у тренувальному наборі даних;
3. під час тренування, точність моделі оцінюється за допомогою двох метрик: AUROC та AUPRC.

В GitHub репозитарії, автори не опублікували повний код тренування для відтворення тої моделі, яка доступна у репозитарії. Проте доступний сам файл з вагами і архітектурою моделі, а також можливість дотреновувати модель, змінюючи розмірність та ваги лише останнього шару [2].

Програмна реалізація. Створено інтерактивний Jupyter-ноутбук та Python-скрипти, в яких реалізована взаємодія з моделлю через birdnetlib та проведено всі дослідження. Завантажено треновану модель BirdNET_GLOBAL_6K_V2.4_Model_FP32.tflite останньої версії (2.4). Це модель, яка підтримує класифікацію 6522 видів птахів, де ваги представлено 32-бітними значеннями з плаваючою крапкою та займає 50 Мб. Формат tflite підтримується бібліотекою TensorFlow Lite. У репозитарії також є квантовані моделі з меншими типами даних (таких як INT8), використання яких може бути доцільним на вбудованих пристроях.

Для проведення досліджень сформовано набір даних аудіозаписів голосів птахів України (257092 записів 309 видів) на основі безкоштовного цифрового репозиторію голосів птахів Xeno-canto [6].

Ми провели експеримент з дослідження точності моделі на 10 найбільш поширених видах птахів нашого регіону, які мають яскраву, впізнавану пісню. Для кожного з цих видів була випадковим чином сформована вибірка із 100 записів, які були проаналізовані моделлю з метою ідентифікації виду. Точність визначення розраховувалась як відношення правильно ідентифікованих записів до їх загальної кількості.

Отримані результати ми порівняли з показниками точності, які наводить в своїй роботі один з авторів розробки Stefan Kahl [3]. Порівняння продемонструвало, що у 4 видів наш результат виявився вищим, а в 6 – нижчим ніж наведений в першоджерелі (табл.1). На загал, можна стверджувати, що чим характернішою і більш структурованою є пісня, тим вищу точність демонструє модель при ідентифікації птаха. Найбільш проблемними є так звані види-пересмішники, які мають здатність імітувати голоси інших птахів, а також сто-

ронні звуки. До таких належить наприклад шпак, для якого модель демонструє найнижчу точність (0,42) порівняно з іншими видами.

Таблиця 1. Точність визначення найбільш поширених видів птахів за результатами тестування моделі в порівнянні з даними S. Kahl, 2020 [4]

Наукова назва	Українська назва	Розмір вибірки	Точність визначення	
			експеримент	літературні дані
<i>Fringilla coelebs</i>	Зяблик	100	0.73	0.58
<i>Motacilla alba</i>	Плиска біла	100	0.76	0.66
<i>Serinus serinus</i>	Щедрик	100	0.81	0.86
<i>Phoenicurus phoenicurus</i>	Горихвістка звичайна	100	0.74	0.67
<i>Ficedula albicollis</i>	Мухоловка білошия	100	0.92	0.82
<i>Turdus merula</i>	Дрізд чорний	100	0.58	0.66
<i>Erithacus rubecula</i>	Вільшанка	100	0.67	0.70
<i>Sturnus vulgaris</i>	Шпак	100	0.42	0.51
<i>Phylloscopus collybita</i>	Вівчарик-ковалик	100	0.77	0.85
<i>Luscinia luscinia</i>	Соловейко східний	100	0.77	0.83

Висновок. Отже, використання методів класифікації спектрограм звуків за допомогою конволюційних нейромереж (CNN) є достатньо ефективним для вирішення поставлених завдань. BirdNET на сьогодні – одна з кращих моделей автоматичного розпізнавання голосів птахів, оскільки вона має досить високу точність і широкі можливості для вдосконалення. Точність моделі є видоспецифічною, тобто залежить від особливостей вокалізації різних видів птахів. Подальші дослідження можуть бути спрямовані на дотренування моделі для ідентифікації окремих груп птахів з більшою точністю та меншим об’ємом моделі, а також визначення оптимальних порогових значень впевненості для полегшення автоматичного виявлення окремих видів.

Список використаних літературних джерел

- [1] Височин М.О., Корнієнко Т.М., Струс Ю.М. та ін. (2023) Перші результати акустичного моніторингу фонових видів птахів в Українських Карпатах: Біорізноманіття, екологія та експериментальна біологія., 25, 2, 50-62
- [2] BirdNET analyzer for scientific audio data processing (2024) - <https://github.com/kahst/BirdNET-Analyzer>
- [3] Kahl S. (2020) Identifying Birds by Sound: Large-scale Acoustic Event Recognition for Avian Activity Monitoring, Dissertation. Chemnitz University of Technology, Chemnitz, Germany.
- [4] Kahl S, Wood CM, Eibl M, Klinck H (2021) BirdNET: A deep learning solution for avian diversity monitoring. *Eco Inform* 61: 101236
- [5] Wood, C.M., Kahl, S. (2024) Guidelines for appropriate use of BirdNET scores and other detector outputs. *Journal of Ornithology*, 165, 777–782. <https://doi.org/10.1007/s10336-024-02144-5>
- [6] [Xeno-canto, 2024] Xeno-canto (2024). Sharing bird sounds from around the world. <https://www.xeno-canto.org/about/xeno-canto>.

Bird species identification by sound using machine learning methods: a case of BirdNET

Volodymyr Kuchynskyy, Nataliya Protsakh

This project explores the possibilities of applying machine learning and signal processing methods to bioacoustics with a goal to automate the identification of bird species by their sounds. A review of the newest achievements in this area was conducted. Existing technical solutions meant to simplify bird detection in the wild were analyzed. A dataset of audio recordings for birds of Ukraine based on the Xeno-canto repository. A BirdNET model was deployed and tested on a limited regional sample, split by species and taxonomy groups.