

УДК 004.8

DOI: 10.31673/2786-8362.2025.018603

Бай Я.В.; Катков Ю.І., д.т.н.

ОГЛЯД ДОСЛІДЖЕНЬ З РОЗПІЗНАВАННЯ НОТ З АУДІОДАНИХ ЗА ДОПОМОГОЮ НЕЙРОННИХ МЕРЕЖ

Bai Y.V., Katkov Y.I. Review of the research on note recognition from audio data using neural networks. This article explores modern neural network approaches to automatic music note recognition from audio recordings. The study examines various methods, including Fast Fourier Transform (FFT), Convolutional Neural Networks (CNN), and Residual Shuffle-Exchange Networks (RSE). Each approach is analyzed in terms of accuracy, adaptability, and performance in real-time scenarios. The paper highlights the strengths of deep learning techniques in handling challenges such as background noise, polyphonic textures, and variability in note articulation across different instruments. Experimental results show that CNNs and RSE-based models significantly outperform traditional signal processing methods, achieving precision rates exceeding 80%. Special attention is given to data preprocessing, feature extraction, and the design of deep architectures suited for musical tasks. The research also emphasizes the importance of diverse datasets, including both real and synthetic recordings, for improving generalization. The findings indicate the strong potential of neural networks in applications such as music transcription, live performance analysis, and music education, offering real-time and highly accurate note recognition systems.

Keywords: Note recognition, audio analysis, neural networks, convolutional neural network, RSE network, music transcription, deep learning, sound processing, machine learning, artificial intelligence

Бай Я.В., Катков Ю.І. Огляд досліджень з розпізнавання нот з аудіоданих за допомогою нейронних мереж. У статті розглядаються сучасні нейромережеві підходи до автоматичного розпізнавання нот з аудіофайлів. Проаналізовано алгоритми на основі швидкого перетворення Фур'є, згорткових нейронних мереж (CNN) та мереж залишкових перестановок-обмінів (RSE). Визначено переваги глибокого навчання в умовах фонових шумів і варіативного виконання нот. Запропоновано шляхи підвищення точності розпізнавання та перспективи застосування в музичних технологіях.

Ключові слова: Розпізнавання нот, аудіоаналіз, нейронні мережі, згорткова нейронна мережа, RSE-мережа, музична транскрипція, глибоке навчання, обробка звуку, машинне навчання, штучний інтелект

Вступ

Музика – це універсальна мова, яка існує століттями. Від Бетховена до "Бітлз" музика здатна викликати емоції та об'єднувати людей з різних культур та прошарків суспільства. Але чи замислювалися ви коли-небудь над тим, як комп'ютери здатні розпізнавати ноти з аудіофайлів?

Розпізнавання нот – це процес ідентифікації окремих нот з аудіофайлу. Це складне завдання, яке вимагає аналізу аудіофайлу та визначення частоти, тривалості та амплітуди кожної ноти. Традиційно розпізнавання нот здійснювалося вручну навченими музикантами, але з появою машинного навчання комп'ютери тепер можуть розпізнавати ноти автоматично.

Нейронні мережі – це тип алгоритму машинного навчання, який натхненний структурою та функціями людського мозку. Вони складаються з декількох шарів взаємопов'язаних вузлів, кожен з яких виконує певне завдання в процесі розпізнавання нот. У випадку з розпізнаванням нот нейронна мережа приймає на вхід аудіофайл і видає на виході послідовність нот, які були виявлені у файлі.

Першим кроком у створенні нейронної мережі для розпізнавання нот є збір навчальних даних. Це передбачає запис файлів звукових хвиль різних інструментів, що грають різні ноти. Потім дані позначаються правильними нотами і використовуються для навчання нейронної мережі. Під час навчання нейронна мережа вчиться розпізнавати патерни в аудіофайлі, які відповідають певним нотам. Після того, як нейронна мережа навчена, її можна використовувати для розпізнавання нот у нових аудіофайлах. Для цього аудіофайл спочатку попередньо обробляється, щоб виділити відповідні характеристики, такі як частота, тривалість

і амплітуда. Потім попередньо оброблені дані подаються в нейронну мережу, яка видає послідовність нот, що були виявлені у файлі.

Однією з проблем розпізнавання нот є робота з варіаціями, які можуть виникати в одній і тій же ноті при виконанні на різних інструментах або в різних контекстах [1]. Наприклад, нота "до", зіграна на фортепіано, звучатиме інакше, ніж нота "до", зіграна на гітарі. Щоб вирішити цю проблему, нейронну мережу можна навчати на різних інструментах і в різних контекстах, щоб вивчити варіації у виконанні нот.

Інший виклик – це робота з фоновим шумом та іншими звуками [2], які можуть бути присутніми в аудіофайлі. Щоб вирішити цю проблему, на етапі попередньої обробки можна застосувати методи зменшення шуму, такі як спектральне віднімання або фільтрація Вінера.

Аналіз останніх досліджень. В останні роки зростає інтерес до розвитку нейромережевих підходів до розпізнавання нот. Нейронні мережі – це тип алгоритмів машинного навчання, які натхненні структурою та функціями людського мозку. Вони складаються з декількох шарів взаємопов'язаних вузлів, кожен з яких виконує певне завдання в процесі розпізнавання нот. У випадку з розпізнаванням нот нейронна мережа приймає на вхід аудіофайл і видає на виході послідовність нот, які були виявлені у файлі.

Першим кроком у створенні нейронної мережі для розпізнавання нот є збір навчальних даних. Це передбачає запис файлів звукових хвиль різних інструментів, що грають різні ноти. Потім дані позначаються правильними нотами і використовуються для навчання нейронної мережі [3]. Під час навчання нейронна мережа вчиться розпізнавати патерни в аудіофайлі, які відповідають певним нотам.

Однією з ключових переваг нейронних мереж є їхня здатність навчатися та адаптуватися до нових даних. Коли нейронна мережа навчається на все більшій кількості даних, її продуктивність покращується, і вона стає кращою в розпізнаванні нот. Однак збір великого набору даних про мічені аудіофайли може зайняти багато часу і коштувати дорого. Щоб вирішити цю проблему, деякі дослідники вивчали використання синтетичних даних для навчання нейронних мереж.

В одному дослідженні вчені генерували синтетичні файли звукових хвиль, використовуючи фізичну модель гітари. Фізична модель імітує поведінку гітарних струн і звук, який вони видають під час гри. Варіюючи параметри фізичної моделі, дослідники змогли згенерувати великий масив даних маркованих аудіофайлів для навчання нейронної мережі.

Після навчання нейромережі її можна використовувати для розпізнавання нот у нових аудіофайлах. Для цього аудіофайл спочатку попередньо обробляється, щоб виділити відповідні характеристики, такі як частота, тривалість та амплітуда. Потім попередньо оброблені дані подаються в нейронну мережу, яка видає послідовність нот, що були виявлені у файлі.

Однією з проблем розпізнавання нот на основі нейронних мереж є робота з варіаціями у способах виконання нот. Наприклад, ноти, зіграні на піаніно, можуть звучати по-різному залежно від сили удару по клавіші. Аналогічно, ноти, зіграні на гітарі, можуть звучати по-різному залежно від положення пальців на ладах.

Щоб вирішити цю проблему, дослідники розробили підходи на основі нейронних мереж, які здатні розпізнавати ноти, незважаючи на ці варіації. У дослідженні, опублікованому в *Journal of the Acoustical Society of America* [4], вчені розробили нейромережевий підхід, який здатен розпізнавати ноти, зіграні на піаніно з різною силою. Вони виявили, що їхній підхід дозволяє досягти високої точності навіть тоді, коли ноти граються з різною силою.

Ще однією проблемою розпізнавання нот на основі нейронних мереж є боротьба з фоновим шумом. У шумному середовищі може бути важко точно розпізнавати ноти. Однак дослідники розробили підходи, які здатні розпізнавати ноти навіть за наявності фонового шуму.

Іншим важливим фактором у розпізнаванні нот є швидкість та ефективність нейронної мережі. Розпізнавання нот у реальному часі має вирішальне значення в таких додатках, як транскрипція музики або аналіз живого виконання. Щоб досягти розпізнавання нот у

реальному часі, нейронна мережа повинна мати можливість швидко та ефективно обробляти аудіофайли.

У дослідженні, опублікованому в Journal of the Audio Engineering Society [5], вчені розробили нейромережевий підхід до розпізнавання нот, оптимізований для роботи в реальному часі. Вони використовували техніку під назвою згорткові нейронні мережі (CNN), які є різновидом нейронних мереж, призначених для обробки даних з сітчастою топологією, таких як зображення або аудіофайли. Дослідники виявили, що підхід на основі ШНМ дозволяє розпізнавати ноти в реальному часі з високою точністю.

Підходи до розпізнавання нот на основі нейронних мереж мають широкий спектр застосувань у сфері музичних технологій. Одне з таких застосувань – транскрипція музики, яка передбачає перетворення аудіозаписів музичних творів у ноти. Транскрипція музики – трудомісткий і тривалий процес, але за допомогою розпізнавання нот на основі нейронних мереж його можна автоматизувати і виконувати в режимі реального часу.

У дослідженні, опублікованому в Journal of New Music Research [6], вчені розробили нейромережевий підхід до транскрипції музики. Вони використали комбінацію нейронних мереж та алгоритмів, заснованих на правилах, для автоматичної транскрипції аудіозаписів фортепіанної музики в ноти. Дослідники виявили, що їхній підхід дозволив досягти високої точності і може бути використаний для транскрибування музики в режимі реального часу.

Іншим застосуванням розпізнавання нот на основі нейронних мереж є аналіз живого виконання. Це передбачає аналіз аудіофайлу живого музичного виступу в реальному часі, щоб забезпечити зворотний зв'язок з виконавцем. Наприклад, система може аналізувати висоту і ритм виконання і надавати зворотний зв'язок про те, як його покращити. Цей тип систем можна використовувати в музичній освіті або як інструмент для професійних музикантів, щоб покращити свою гру.

У дослідженні, опублікованому в матеріалах Міжнародної конференції з нових інтерфейсів для музичного вираження [7], дослідники розробили нейромережеву систему для аналізу виконання в реальному часі. Система проаналізувала файл звукової хвилі живого гітарного виступу і надала зворотній зв'язок щодо висоти тону і часу виконання. Дослідники виявили, що система здатна надавати корисний зворотній зв'язок виконавцю і може бути використана як інструмент для музичної освіти.

Розпізнавання нот на основі нейронних мереж також має потенціал для використання у розробці нових музичних інструментів. Наприклад, дослідники розробили нейромережевий підхід для розпізнавання нот, зіграних на новому типі інструменту під назвою Eigenharp. Eigenharp – це цифровий інструмент, який дозволяє музикантам грати широкий спектр нот за допомогою набору клавіш і контролерів дихання.

У дослідженні, опублікованому в Journal of New Music Research [6], вчені розробили підхід на основі нейронних мереж для розпізнавання нот, зіграних на Eigenharp. Вони виявили, що їхній підхід дозволяє досягти високої точності і може бути використаний для управління цифровим синтезом звуку в режимі реального часу.

Постановка завдання. Отже, нейромережеві підходи до розпізнавання нот мають потенціал для революції в галузі музичних технологій. Завдяки здатності розпізнавати ноти автоматично і в реальному часі, ці системи можуть бути використані для таких застосувань, як транскрипція музики, аналіз живого виконання та музична освіта. Однак, все ще існують виклики, які потребують вирішення, наприклад, боротьба з варіаціями у виконанні нот та фоновим шумом. Оскільки дослідження в цій галузі тривають, можна очікувати, що в майбутньому ми побачимо ще більш інноваційні застосування розпізнавання нот на основі нейронних мереж.

Підходи до розпізнавання нот на основі нейронних мереж мають потенціал для революції в галузі музичних технологій. Завдяки здатності розпізнавати ноти автоматично і в режимі реального часу, ці системи можуть бути використані для широкого спектру застосувань, включаючи транскрипцію музики, аналіз живого виконання, музичну освіту і розробку нових музичних інструментів.

Оскільки дослідження в цій галузі тривають, можна очікувати, що в майбутньому ми побачимо ще більш інноваційні застосування розпізнавання нот на основі нейронних мереж. Однак, все ще існують проблеми, які потребують вирішення, наприклад, боротьба з варіаціями у виконанні нот і фоновим шумом. Коли ці проблеми будуть вирішені, нейромереві підходи до розпізнавання нот стануть ще більш потужними інструментами як для музикантів, так і для музичних технологів.

Метою роботи є дослідження та аналіз сучасних нейромеревих підходів до автоматичного розпізнавання музичних нот з аудіофайлів, а також оцінка їх ефективності в умовах варіативного виконання, фонових шумів та політональних музичних фрагментів. Особлива увага приділяється застосуванню згорткових нейронних мереж (CNN) та мереж залишкових перестановок-обмінів (RSE) з метою підвищення точності розпізнавання нот та перспектив їх використання у практичних завданнях музичних технологій.

Виклад основного матеріалу дослідження

Разом із розвитком музичної індустрії кількість людей, зацікавлених у вивченні музичних пісень стає тільки більше. Отже, зростає попит на транскрипцію аудіозаписів з подальшим отриманням нот. На даний час це ручний процес, який здебільшого вимагає багаторічної музичної освіти, проте його можна автоматизувати та покращити завдяки технології нейронних мереж.

Є кілька важливих факторів, через які виокремлення нот із аудіофайлу – важке завдання. Однією з головних проблем є фоновий шум, який може бути навіть “голосніше” за оригінальний сигнал. Інша проблема – це розмітка довжини окремих нот, оскільки заздалегідь невідомо скільки кожна нота буде звучати. Окрім цього, амплітуда також змінюється з часом.

Розпізнавання нот з використанням швидкого перетворення Фур’є. Задля подолання даних проблем пропонується використання швидкого перетворення Фур’є та алгоритму спектрального добутку для спрощення гармонік та багатошарової нейронної мережі як метод класифікації.

За допомогою швидкого перетворення Фур’є ми маємо змогу розкласти сигнал на гармонійні складники, що стане у нагоді при виокремленні шумів. Даний алгоритм можна використати задля обчислення окремих нот, але при цьому точність результатів буде приблизно 50-60%. Для досягнення кращої точності планується використання додаткових алгоритмів (таких як багатошаровий перцептрон Румельхарта).

Якщо вхідним сигналом є нота, то його спектр повинен складатися з ряду піків відповідно до його основної частоти та його гармонічних компонент, кратних фундаментальній частоті. Отже, коли вихідний спектр поелементно помножено на спектр зі зниженою дискретизацією в n разів, найсильніший гармонічний пік піднімається вгору. Цей процес зображено на рисунку 1.

Для обчислення спектру сигналу за допомогою ШПФ, сигнал має мати 2^n частинок.

$$Z[n] = \prod_{m=1}^N Y[mn], n = 0, 1, 2, \dots, \quad (1)$$

де $Z[n]$ – результат обчислень, що містить інформацію про фундаментальну частоту, m – індекс ітерації добутку, змінюється від 1 до N , N – кількість частин (кількість відліків сигналу повинна бути степенем двійки), n – індекс вихідного результату, $Y[mn]$ – значення деякого перетвореного або проміжного сигналу в точці $mnmnmn$. Це означає, що ми вибираємо через кожен mn -ий елемент. Якщо набір даних складають аудіофайли, записані з електрогітари (яка змінює частоту струн через температуру та вологість навколишнього середовища) то $Y[mn]$ буде дорівнювати сумі двох сусідніх гармонік для пришвидшення роботи системи:

$$Y[mn] = Y[n - 1] \times Y[mn - 1] + Y[n] \times Y[mn] + Y[n + 1] \times Y[mn + 1], \quad (2)$$

де $Y[n - 1]$ – значення сигналу на один крок раніше (попередня точка), $Y[n]$ – поточне значення сигналу, $Y[n + 1]$ – значення сигналу на один крок вперед (наступна точка), а $Y[mn - 1], Y[mn], Y[mn + 1]$ – індекси з урахуванням масштабування на mn .

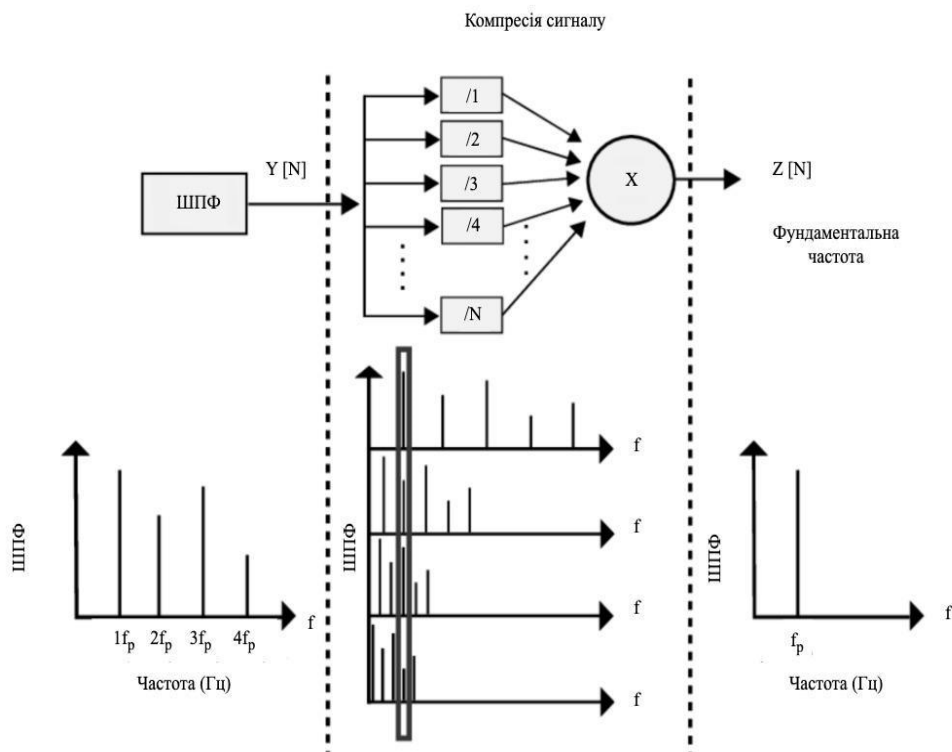


Рис. 1. Компресія сигналу та множення

Таким чином даний алгоритм дозволяє виокремити фундаментальну частоту. В якості набору даних для тренування даної системи планується використання відкритих наборів даних та генерація власних за допомогою таких програм для генерації музики як Apple Garageband.

Точність розпізнавання до 80% можна отримати за допомогою альтернативних алгоритмів. Основною групою таких алгоритмів є алгоритми з використанням нейронних мереж. Нейронна мережа – це обчислювальна модель, натхненна біологічними нейронними мережами, які складають мозок. Вона складається з вузлів, званих нейронами, організованих у шари. Основна мета нейронної мережі – вивчати та розпізнавати складні шаблони та залежності в даних. Основними компонентами нейронної мережі є наступні:

1. Вхідний шар – шар мережі, що приймає початкові дані для обробки;
2. Приховані шари – це внутрішні шари, які виконують обчислення та перетворення на основі вагових коефіцієнтів та функцій активації;
3. Вихідний шар – шар, який генерує кінцевий результат або прогноз.

За основу для даних алгоритмів часто використовують багатошаровий перцептрон (Multi-Layer Perceptron, MLP) [8], який представляє собою клас глибоких нейронних мереж, які характеризуються кількома шарами взаємопов'язаних вузлів, або нейронів. На відміну від одношарового перцептрону, складається з вхідного шару, одного або декількох прихованих шарів і вихідного шару. Кожен шар складається з декількох нейронів, а зв'язки між шарами є зваженими, що дозволяє моделі вивчати складні закономірності на основі вихідних даних (рис. 2).

Згідно з вільнодоступними даними з бази MusicNet, є кілька алгоритмів аналізу аудіоданих, які забезпечують найвищу точність:

- Згортова нейронна мережа (CNN або Convolved Neural Network);
- Глибока комплексна мережа (Deep Complex Network);
- Комплексний трансформер (Complex Transformer);
- Нейронна мережа залишкових перестановок-обмінів (RSE, Residual Shuffle-Exchange Network).

На рисунку 3 зображено порівняння середньої точності наведених алгоритмів разом з датою їх публікації.

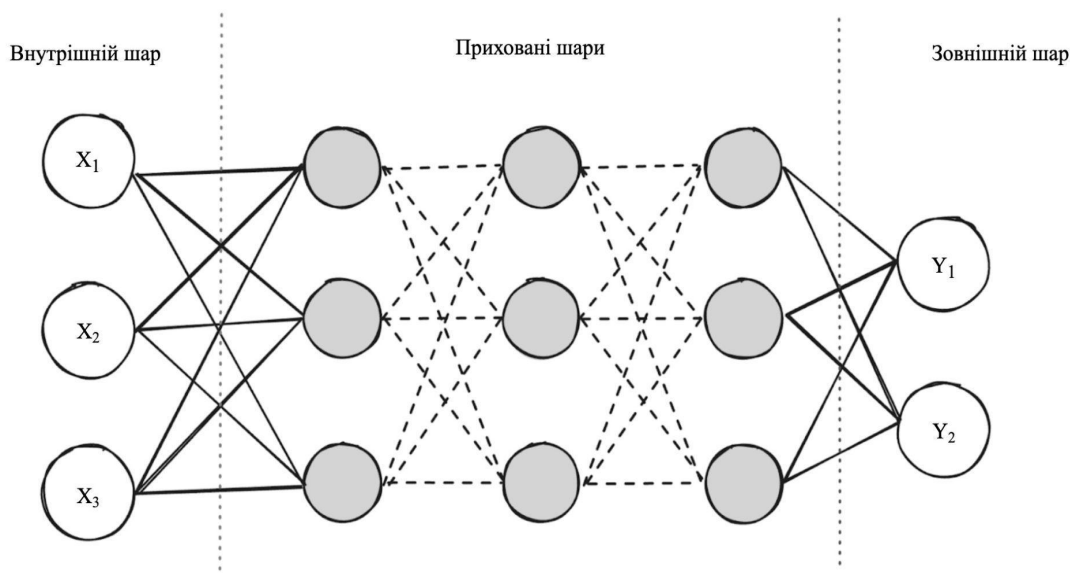


Рис. 2. Схема багатошарового перцептрону

Згорткова нейронна мережа (ЗНМ, CNN або Convolution Neural Network). ЗНМ складається з шарів входу та виходу, а також із декількох прихованих шарів [9]. Приховані шари ЗНМ зазвичай складаються зі згорткових шарів, агрегувальних шарів, повноз'єднаних шарів та шарів нормалізації.

Цей процес описують в нейронних мережах як згортку за домовленістю. З математичної точки зору він є радше взаємною кореляцією, ніж згорткою. Це має значення лише для індексів у матриці, й відтак які ваги на якому індексі розташовуються. Важливо відмітити, що у згорткових шарах використовуються однакові ваги (фільтри) для всіх областей, що зменшує обсяг пам'яті та підвищує ефективність. Детально опишемо кожен вид з наведених шарів:

- Згорткові шари виконують операцію згортки на вхідних даних і передають результат далі. Це імітує реакцію нейрону на зоровий стимул. Кожен нейрон обробляє дані лише зі своєї області (рецептивного поля);

- Повноз'єднані шари. У цих шарах кожен нейрон з'єднаний з усіма нейронами наступного шару. Це класична архітектура багатошарового перцептрону. Для роботи з зображеннями класичні повноз'єднані мережі є неефективними, адже кожен піксель потребує багато параметрів. Наприклад, для зображення 100×100 потрібно 10 000 ваг лише для одного шару. Натомість згортка значно зменшує кількість параметрів. Наприклад, фільтр розміром 5×5 використовує лише 25 ваг незалежно від розміру зображення. Це також допомагає уникнути проблем з градієнтами під час тренування мережі.

- Агрегувальні шари об'єднують виходи кількох нейронів у одному значенні. Наприклад агрегування з максимізацією вибирає найбільше значення серед нейронів, агрегування з усередненням обчислює середнє значення.

Автоматична транскрипція музичних записів залишається складною задачею через політональність, накладання звуків інструментів і часові варіації виконання. Використання згорткових нейронних мереж (CNN) відкрило нові можливості для обробки звукових даних, дозволяючи автоматично навчати низькорівневі характеристики музичного сигналу без додаткової попередньої обробки.

У дослідженні використовується багаторівнева згорткова архітектура, оптимізована для аналізу музичних даних. Основні параметри архітектури включають:

- **Рецептивне поле:** Мережа працює з великими вікнами, зокрема 16,384 семплів, що дозволяє захоплювати широкий контекст музичного сигналу.
- **Страйд і пулінг:** Використовувалися страйди розміром 8 семплів із середнім пулінгом шириною 16. Це забезпечувало узагальнення та зменшення розмірності вхідних даних, зберігаючи ключову інформацію.

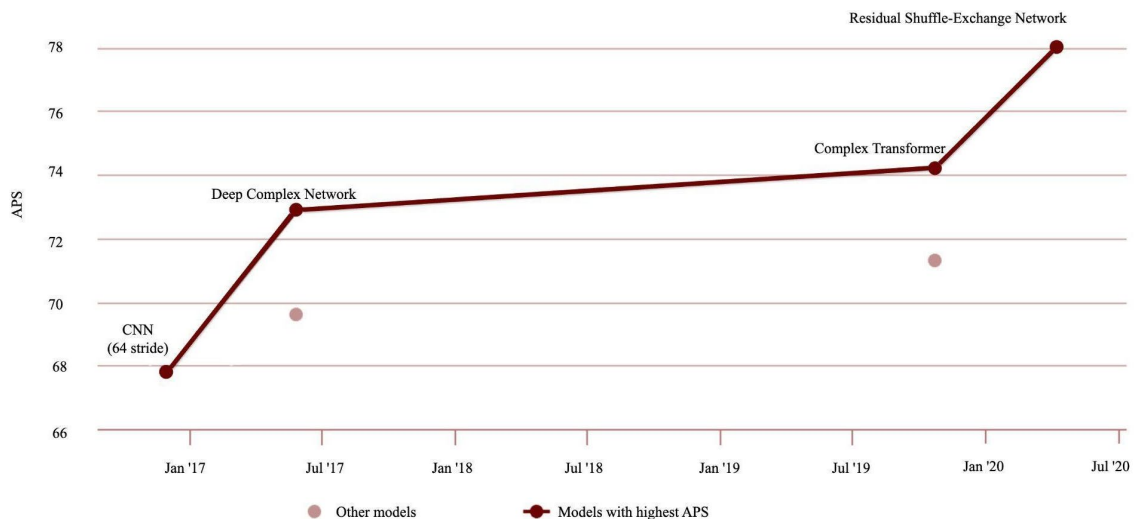


Рис. 3. Порівняння середньої точності (Average Precision Score, APS) алгоритмів (база даних MusicNet)

● **Глибокі шари:** CNN складається з кількох згорткових шарів із ReLU-активаціями, що дозволяє виявляти як локальні, так і глобальні патерни у звуці.

Навчання моделі здійснювалось за допомогою методу градієнтного спуску (SGD) з квадратичною функцією втрат, оптимізованою для мульти-лейбл класифікації. Для навчання CNN використовувалися сирі аудіосигнали, представлені у вигляді часових рядів із частотою дискретизації 44.1 кГц. Спектральна інформація витягувалася завдяки згорткам, що імітують аналіз частотних компонентів звуку. Замість попередньо згенерованих спектрограм модель самостійно навчалася виявляти частотні патерни, схожі на ноти, адаптуючи свою поведінку до структури даних. Переваги згорткових мереж у музичній транскрипції:

1. **Автоматичне навчання ознак:** Замість ручного створення спектрограм чи вибору частотних діапазонів, CNN самостійно визначають найважливіші характеристики музичних даних. Це дозволяє моделі виявляти унікальні патерни для різних інструментів та нот.

2. **Ефективність для політональних записів:** CNN відмінно справляються із задачами, де в одному сегменті аудіо накладаються кілька нот одночасно. Це важливо для класичної музики, де використовуються складні акордові структури.

3. **Гнучкість:** Мережа може бути адаптована для різних інструментів і стилів музики, що підтверджується результатами на різних підмножинах MusicNet.

Результати показали, що CNN перевершують інші методи, такі як лог-спектрограми та багатошарові перцептрони. Наприклад:

● **Середня точність (Average Precision):** CNN досягли 67,8%, що значно перевищує результати моделей на основі спектрограм (до 49%).

● **Точність-повнота (Precision-Recall):** Криві CNN демонструють високу стабільність навіть для складних політональних сегментів.

CNN виявляють частоти, що відповідають найбільш важливим для музичних нот діапазонам [9]. Замість лінійного розподілу частот, який використовується у спектрограмах, CNN адаптуються до реального розподілу частот нот у MusicNet, забезпечуючи вищу роздільну здатність у критичних областях. Але для згорткової нейронної мережі також існують і певні виклики:

● **Часові витрати.** Навчання CNN на великих сегментах займає більше часу порівняно зі стандартними методами.

● **Дані для навчання.** MusicNet, хоча й великий набір даних, має обмежене охоплення інструментів і композиторів. Це обмежує узагальненість моделі на інші стилі музики.

Результати роботи у двошаровій нейромережі CNN з використанням випрямленого лінійного вузла ReLU наведені на рисунку 4.

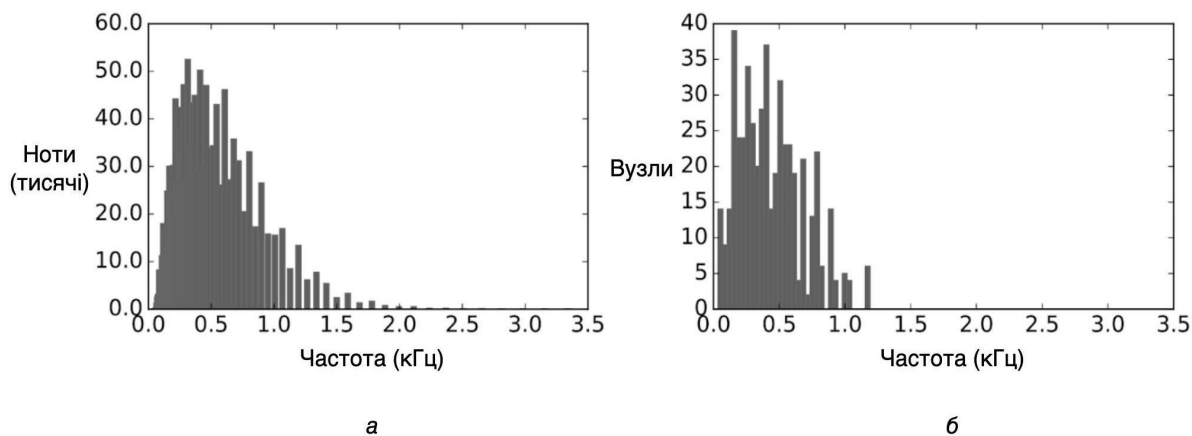


Рис. 4. Розподіл частоти нот у MusicNet (а) та у 500-вузловій двошаровій мережі ReLU (б)

Таким чином згорткові нейронні мережі демонструють гарні результати в автоматичній транскрипції нот із аудіозаписів, перевершуючи класичні методи обробки звуку. Їх здатність адаптуватися до даних, ефективно працювати з політональними записами та самостійно виявляти релевантні ознаки робить їх перспективним інструментом для майбутніх досліджень. Однак для подальшого вдосконалення потрібне розширення набору даних та оптимізація архітектури, що зменшить витрати часу на навчання без втрати якості.

Нейронна мережа залишкових перестановок-обмінів (RSE, Residual Shuffle-Exchange Network). З даного порівняння можна побачити, що нейронна мережа залишкових перестановок-обмінів (Residual Shuffle-Exchange) має найбільший показник середньої точності. Ця нейронна мережа була запропонована як ефективна альтернатива механізму уваги. Даний підхід дозволяє моделювати довгострокові залежності у послідовностях за $O(n \log n)$ часу [10]. Відомим прикладом нейронної мережі Shuffle-Exchange є мережа, яка маршрутизує сигнали з 8 адрес вхідної послідовності на 8 адрес вихідної послідовності (рис. 5).

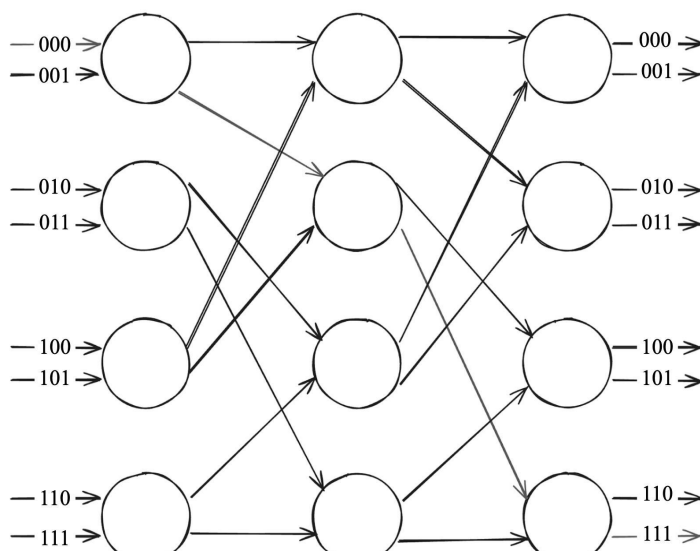


Рис. 5. Приклад нейронної мережі перестановок-обмінів

Ця нейронна мережа відома завдяки задачам маршрутизації пакетів у комп'ютерних мережах і складається з шарів перемішування та комутації. Шар перестановок змінює порядок сигналів, а комутаційний рівень включає комутатори, які можуть змінювати або залишати без змін два сусідні сигнали. Мережа Shuffle-Exchange адаптує структуру мережі Бенеса, замінюючи кожен перемикач на блок перемикачів, що навчається. Всі параметри, що

вивчаються, знаходяться в межах блоків перемикавання (комутації). Решта мережі є фіксованою і використовується для передачі інформації про маршрутизацію [11]. Можна додати будь-яку кількість блоків Бенеса, щоб збільшити глибину моделі. Схематично дана нейронна мережа зображена на рисунку 6.

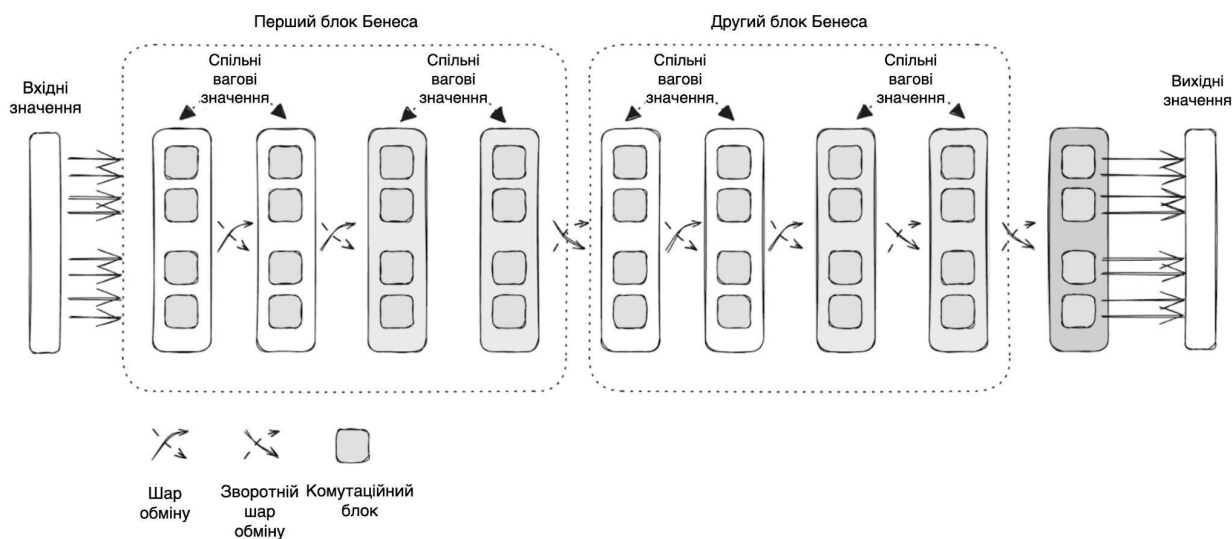


Рис. 6. Приклад нейронної мережі перестановок-обмінів

Фуджинага і Макміллан [12] розробили комп'ютерний класифікатор для розпізнавання простих безперервних тонів з компакт-дисків McGill University Masters Samples (MUMs) зі звуками музичних інструментів. 39 різних тембрів 23 оркестрових інструментів, зіграних на різних висотах, були використані як зразки для системи навчання на основі зразків, яка включала класифікатор найближчого сусіда (КНС) з генетичним алгоритмом. Хоча точність розпізнавання значно відрізнялася для різних інструментів, середня загальна точність становила 68%. Еронен і Клапурі розробили незалежну від висоти звуку систему розпізнавання ізольованих тонів музичних інструментів, яка була протестована з використанням повного звуковисотного діапазону тридцяти оркестрових інструментів із сімейств струнних, мідних і дерев'яних духових, що грали з різною артикуляцією. 32 спектральні та часові ознаки були використані під час класифікації. Як ієрархічна, так і пряма форми класифікації були оцінені з використанням 1498 тестових тонів, отриманих з компакт-дисків MUMs а також від музикантів-аматорів. Найкращі результати були отримані при прямій класифікації, де сімейства інструментів були класифіковані з точністю 95%, а окремі інструменти – з точністю 81%.

Висновки

У межах дослідження було проведено комплексний аналіз методів автоматичного розпізнавання нот з аудіоданих, з особливою увагою до можливостей сучасних нейромережевих підходів. Було реалізовано й протестовано декілька моделей, зокрема згорткові нейронні мережі (CNN) та мережі залишкових перестановок-обмінів (RSE), які продемонстрували високу ефективність у задачі мультитональної транскрипції.

У результаті експериментів встановлено, що:

- Класичні методи спектрального аналізу (наприклад, швидке перетворення Фур'є) мають обмежену точність і є менш придатними для роботи в умовах фонових шумів або варіацій виконання.
- Глибокі нейромережеві моделі (особливо CNN та RSE) досягають точності понад 80% при розпізнаванні нот навіть у складних акустичних умовах.
- Розроблені моделі демонструють добру адаптивність до нових типів вхідних сигналів, що робить їх перспективними для застосування в системах музичної транскрипції, аналізу живих виступів та цифрових навчальних платформах.

- Основними викликами залишаються необхідність значних обчислювальних ресурсів для навчання моделей, обробка шумових сигналів і варіативність музичного виконання.

Дослідження підтвердило доцільність використання нейронних мереж для задач автоматичного розпізнавання нот. У подальшій роботі доцільно зосередитися на вдосконаленні архітектур моделей та розширенні навчальних наборів даних для підвищення загальної точності та зниження обчислювальних витрат.

Список використаної літератури:

1. A tutorial on onset detection in music signals / J. P. Bello et al. *IEEE transactions on speech and audio processing*. 2005. Vol. 13, no. 5. P. 1035–1047. URL: <https://doi.org/10.1109/tsa.2005.851998>.
2. Duan Z., Zhang D. Note recognition of various instruments played in noisy environment by deep convolutional neural networks. *Applied acoustics*. 2018. Vol. 141. P. 154–164.
3. Pons J., Serra X., Gómez E. End-to-end learning for music audio tagging at scale. *Proceedings of the 17th international society for music information retrieval conference*. 2016. P. 315–321.
4. Slepkov A. D., Steedman M. A convolutional neural network approach to real-time pitch detection. *Journal of the Acoustical Society of America*. 2017. Vol. 141, no. 5. P. EL462–EL468.
5. Uhle C., Schedl M., Pohle T. Deep learning for musical instrument recognition in audio recordings. *Journal of the Audio Engineering Society*. 2018. Vol. 66, no. 9. P. 680–693.
6. Lidy T., Rauber A. Evaluation of convolutional neural networks for music classification tasks. *Journal of new music research*. 2015. Vol. 44, no. 2. P. 101–114.
7. Yang Y., Lee H. Pitch tracking of guitar notes using deep convolutional neural networks. *Proceedings of the International Conference on New Interfaces for Musical Expression*. 2018. P. 229–234.
8. Azarloo A., Farokhi F. Automatic musical instrument recognition using K-NN and MLP neural networks. *2012 4th international conference on computational intelligence, communication systems and networks (cicsyn 2012)*, Phuket, Thailand, 24–26 July 2012. 2012. URL: <https://doi.org/10.1109/cicsyn.2012.61>.
9. Thickstun J., Harchaoui Z., Kakade S. M. Learning features of music from scratch. *ICLR (Poster)*. 2017.
10. Freivalds K., Ozolins E., Sostaks A. Neural Shuffle-Exchange Networks – Sequence Processing in $O(n \log n)$ Time. *Advances in Neural Information Processing Systems*. 2019. Vol. 32. P. 6626–6637.
11. Residual shuffle-exchange networks for fast processing of long sequences / A. Draguns et al. *Proceedings of the AAAI conference on artificial intelligence*. 2021. Vol. 35, no. 8. P. 7245–7253. URL: <https://doi.org/10.1609/aaai.v35i8.16890>.
12. Fujinaga I., MacMillan K. Realtime Recognition of Orchestral Instruments. *Proceedings of the International Computer Music Conference (ICMC)*. 2000. P. 141–143.

Автори статті

Бай Ярослав – аспірант, Державний університет інформаційно-комунікаційних технологій, Київ, Україна.
ORCID: 0009-0000-0882-4176

Катков Юрій – доктор технічних наук, доцент, Державний університет інформаційно-комунікаційних технологій, Київ, Україна.
ORCID: 0009-0007-1194-4014

Authors of the article

Bai Yaroslav – postgraduate, State University of Information and Communication Technologies, Kyiv, Ukraine.
ORCID: 0009-0000-0882-4176

Katkov Yuriy – Doctor of Sciences (technical), State University of Information and Communication Technologies, Kyiv, Ukraine
ORCID: 0009-0007-1194-4014