

# Інформаційна технологія розпізнавання нот з використанням нейронних мереж

**Магістерська кваліфікаційна робота**  
**спеціальність 122 «Комп'ютерні науки та інформаційні технології»**

**Виконав:** студент гр. 2КН-12м Кучеровський Ю.В.

**Керівник:** к. т. н., доцент Арсенюк І.Р.

# Мета та завдання дослідження

**Метою дослідження** є підвищення точності розпізнавання нот музичних композицій на основі цифрового аудіо-запису за рахунок використання інформаційної технології на основі згорткової нейронної мережі.

Для досягнення поставленої мети необхідно розв'язати такі наступні завдання:

- здійснити аналіз сучасного рівня розвитку інформаційної технології розпізнавання нот у цифровому звуці;
- розглянути та проаналізувати існуючі методи розв'язання задачі розпізнавання нот, які задовольняють мету даної магістерської кваліфікаційної роботи;
- розробити математичну модель для інформаційної технології розпізнавання нот музичних композицій;
- навести стадії інформаційної технології та на їх основі розробити структуру та алгоритм роботи програмного засобу;
- виконати програмну реалізацію запропонованої інформаційної технології розпізнавання нот;
- провести тестування програмного продукту та виконати аналіз отриманих результатів.

# Об'єкт, предмет та методи дослідження

**Об'єкт дослідження** – процес розпізнавання нот музичних інструментів з використанням інформаційних технологій.

**Предмет дослідження** – програмні засоби визначення нот музичних композицій за допомогою згорткової нейронної мережі.

## **Методи дослідження**

У роботі використані наступні методи наукових досліджень: методи оброблення цифрової інформації (вейвлет-перетворення, згортка), теорія штучних нейронних мереж для реалізації інформаційної технології розпізнавання нот, методи математичної статистики для обрахунків результатів отриманих за допомогою програмного засобу, програмування на мовах високого рівня, методи паралельних обчислень.

## **Постановка задачі дослідження**

У загальному вигляді задача розпізнавання нот музичних інструментів визначається наступним чином: для всієї музичної композиції необхідно ідентифікувати в будь-який момент часу ноту, що звучить. Відповідно, вхідні дані представляють собою цифрову музичну композицію в форматі .wav чи .mp3, а в результаті виходить послідовність нот, програних у різних часових інтервалах.

# Наукова новизна одержаних результатів

**полягає в наступному:**

- розроблено спеціалізовану функцію втрат глибокої згорткової нейронної мережі;
- розроблено спеціалізовану архітектуру (з урахуванням музичної теорії) глибокої згорткової нейронної мережі.

# Практичне значення одержаних результатів

**полягає у наступному:**

отриманий досвід та результати можуть потенційно застосуватися для вирішення суміжних задач multi-label класифікації одновимірних періодичних сигналів (наприклад, в радіотехніці, сейсмології, задачах спектрального аналізу тощо).

# АКТУАЛЬНІСТЬ ЗАДАЧІ

Один з найважливіших типів мультимедійного контенту – це великий обсяг цифрової музики, що розповсюджується через Інтернет. Це мотивує дослідників розробляти нові методи вилучення музичної інформації (MIR), які можуть бути корисними для музичних пошукових систем Інтернету, музичних рекомендаційних систем, систем розділення та розпізнавання музичних інструментів у композиції, систем автоматичного транскрибування, систем автоматичної категоризації та генерації музики. Зважаючи на це, досить актуальною є задача підвищення точності розпізнавання нот музичних композицій на основі цифрового аудіо-запису за рахунок використання інформаційної технології на основі штучних нейронних мереж, зокрема згорткової нейронної мережі.

# Математична модель функції втрат

На практиці при застосуванні комбінації сигмоїди (sigmoid) і перехресної ентропії присутні обмеження:

1. Продуктивність обчислень.
2. Переповнення змінних при великих позитивних вхідних значень.
3. Зміщення помилки в бік негативних міток (в наборі даних кількість позитивних значень для міток на порядок менше ніж негативних значень).

$x$  – вхідні дані,  $z$  – мітки,  $q$  – вага для позитивних значень.

Звичайна перехресна ентропія в комбінації з sigmoid :

$$z * -\log(\text{sigmoid}(x)) + (1 - z) * -\log(1 - \text{sigmoid}(x))$$

Для вирішення проблеми (3) – мультиплікативно збільшуємо помилку для додатних значень

$$q * z * -\log(\text{sigmoid}(x)) + (1 - z) * -\log(1 - \text{sigmoid}(x))$$

Для вирішення проблеми (2) – використовуємо sigmoid для  $x < 0$

$$\text{sigmoid}(x) = \frac{e^x}{e^x + 1}$$

Запропоновано наступні оптимізації:

1. Алгебраїчне спрощення операцій.
2. Застосування комбінованої функції сигмоїди (sigmoid) в якій експонента завжди обчислюється для від'ємних значень.
3. Використання додаткового аргументу (ваги) для збільшення значень позитивної помилки.

Для вирішення проблеми (2) – використовуємо sigmoid для  $x > 0$

$$\text{sigmoid}(x) = \frac{1}{1 + e^{(-x)}}$$

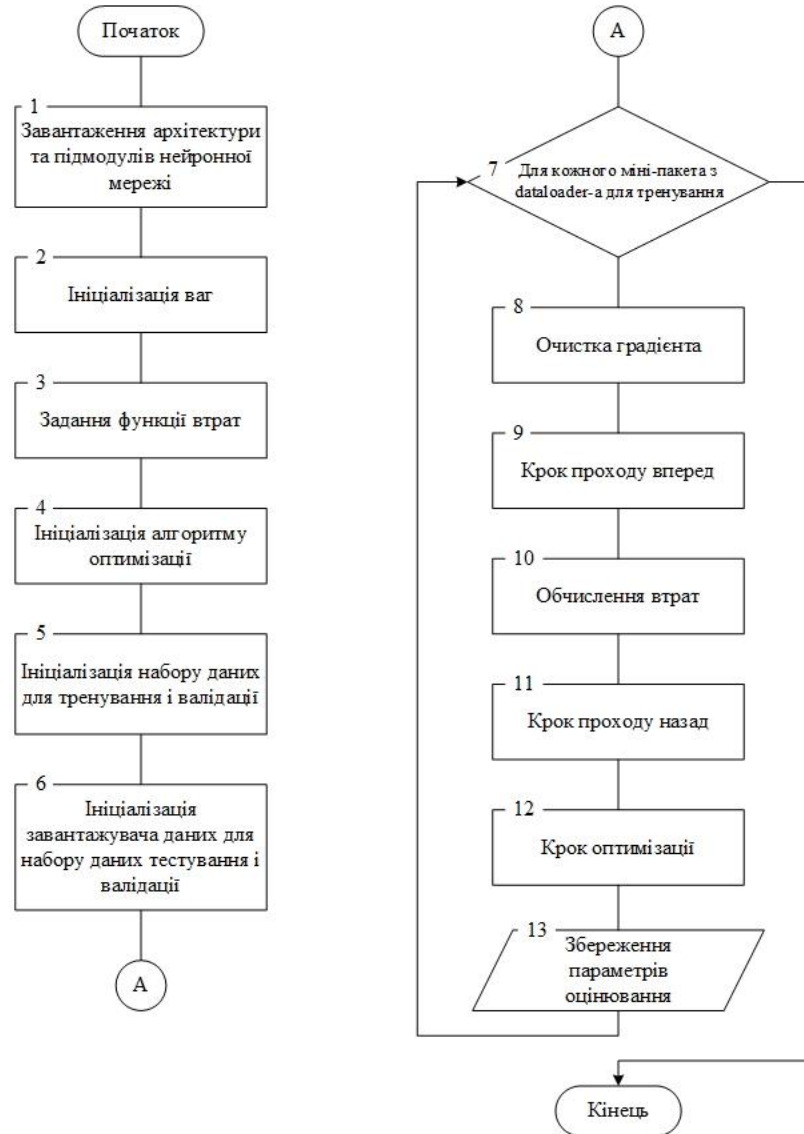
Комбінуємо для загального випадку:

$$\text{для } x < 0 \quad \log(e^x + 1) * (z * (q - 1) + 1) - q * z * x$$

$$\text{для } x > 0 \quad (\log(e^{(-x)} + 1) + x) * (z * (q - 1) + 1) - q * z * x$$

$$\text{для } x \in \mathbb{R} \quad (z * (q - 1) + 1) * (\log(\exp(x - \max(x, 0)) + 1)) - q * z * x$$

# Алгоритм тренування/валідації нейронної мережі



# Алгоритм тестування нейронної мережі





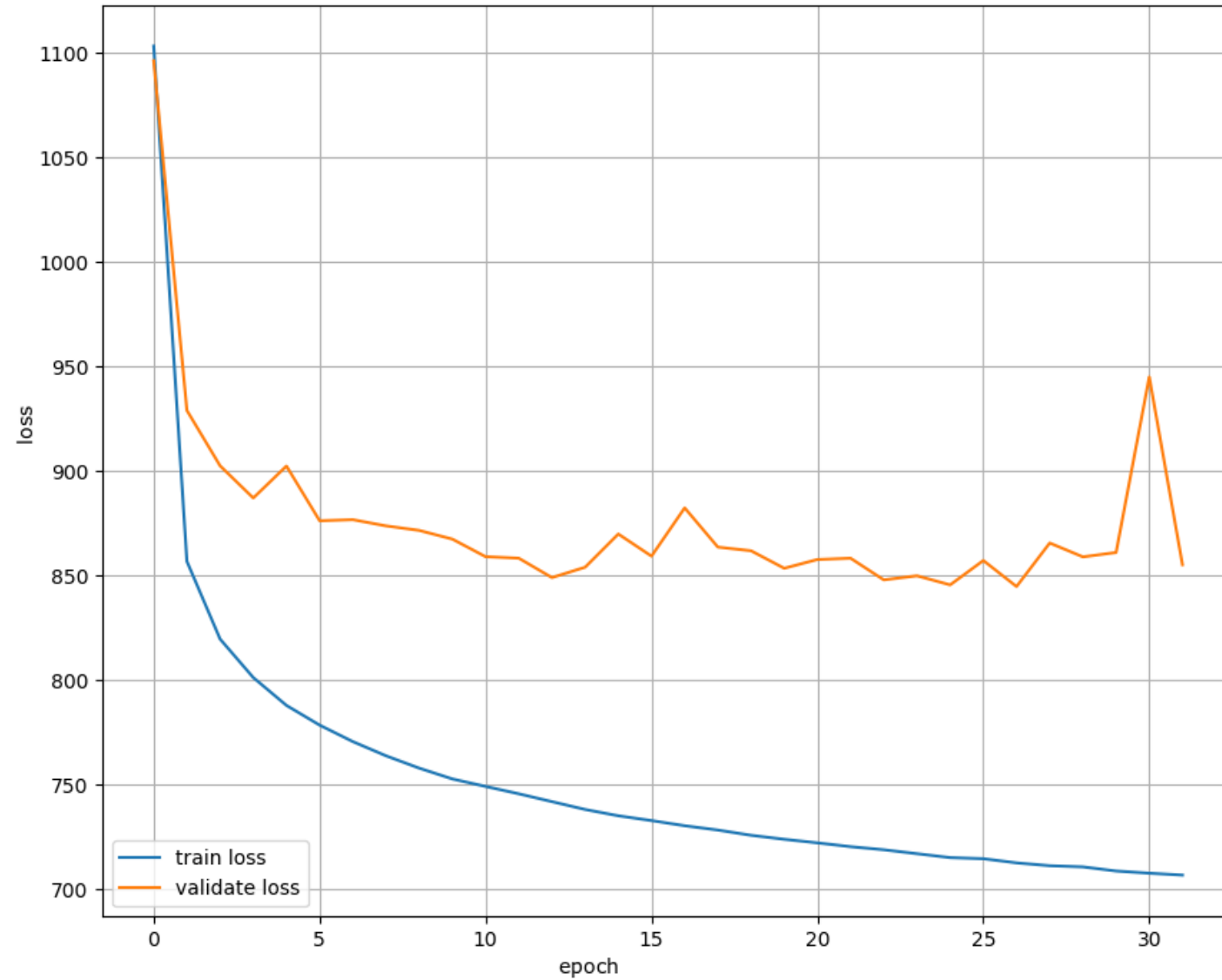
# Алгоритм оцінювання нейронної мережі



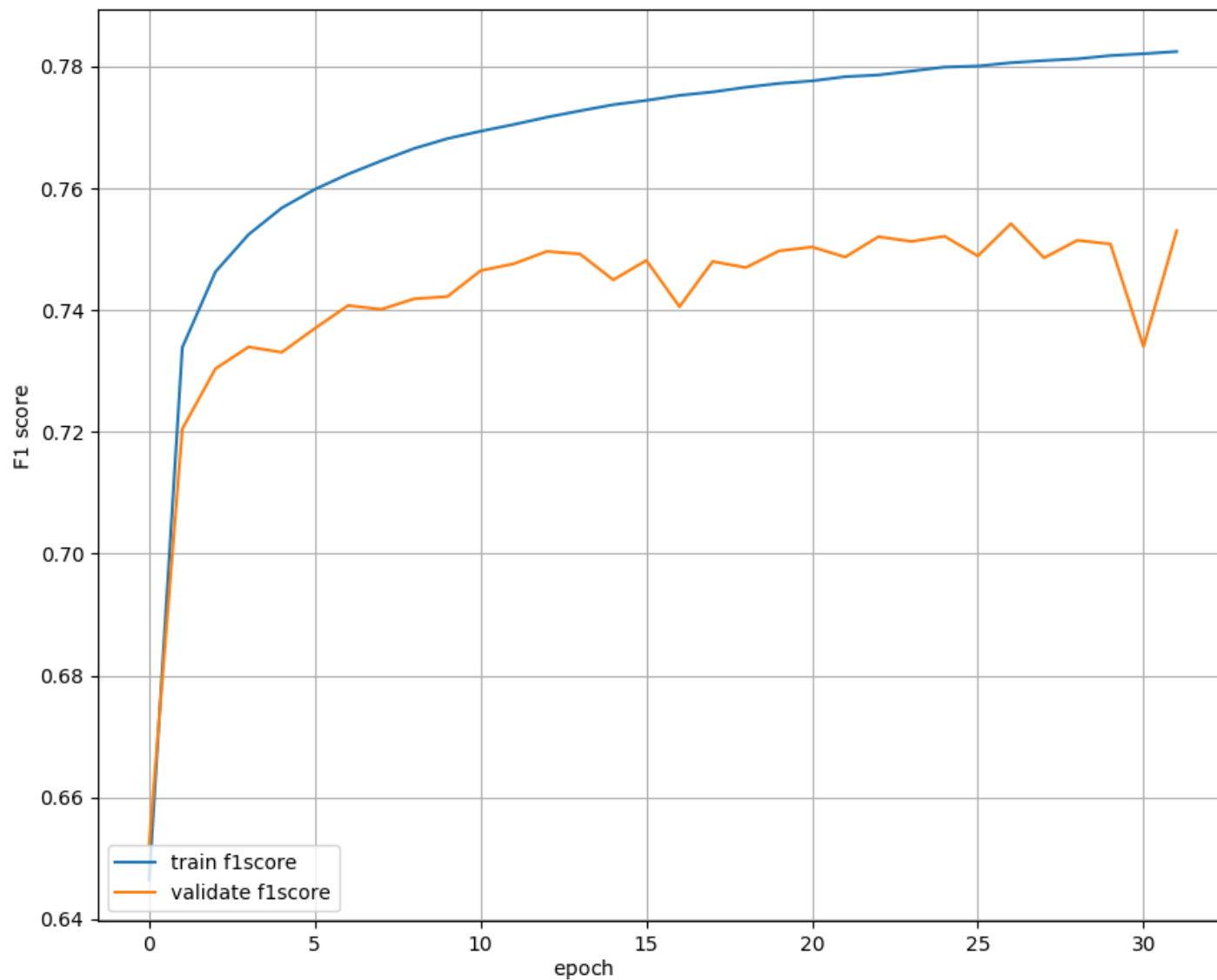
# Структура інформаційної технології розпізнавання нот з використанням нейронних



# Криві навчання нейронної мережі по значенню функції втрат



# Криві навчання нейронної мережі по метриці F1 score



# Організація набору даних

За основу взяті оригінальні записи 6151 композицій 873 авторів в форматі Ogg та відповідні їм караоке звукозаписи в форматі MIDI.

Композиції поділені на фрагменти (семпли) триванням 2 сек. які зберігаються в директорії `features_samples` в форматі WAV (сирцеве аудіо) на основі даних обчислюється MD5-сума, яка служить ім'ям семплу (файлу).

Кожен фрейм являє собою бінарний вектор зі 108 елементів (9 основних октав (для перекриття діапазону фортепіано) \* 12 звуків в октаві (Scientific pitch notation) в якому звучання відповідного звуку позначається як 1 в відповідній позиції (індексу), якщо звук на момент фрейму не звучить то значення встановлюється рівним 0. Фрейми зберігаються в вигляді numpy – матриці розміром 108 \* 128 в директорію `targets_samples` і служать цільовими даними.

Композиції поділено на групи (використання семплів одної композиції в різних групах – не дозволяється):

- `train` – основна група для тренування ~25ГБ семплів;
- `validate` – група для валідації (перевірки моделі після кожної епохи) – 1ГБ;
- `test` – група підтвердження роботоспроможності моделі 2ГБ;

# Тестування

- В результаті перевірки роботоспроможності моделі на групі семплів розміром в 2 ГБ, які не використовувалися в процесі тренування/валідації, був отриманий задовільний результат – 0,729 по метриці F1 score.
- Варто зазначити, що для перевірки роботоспроможності було відібрано одні з найскладніших композицій гітаристів-віртуозів Інгві Мальмстіна, Стіва Вея, Джимі Хендрікса, та ін.
- На простіших композиціях з меншою кількістю інструментів та складних ефектів F1 score збігається з результатами валідації – 0,752.

# ЕКОНОМІЧНА ЧАСТИНА

Було виконано оцінювання комерційного потенціалу розробки інформаційної технології розпізнавання нот з використанням нейронних мереж.

- Загальні витрати на виконання науково-дослідної роботи - 59 220 грн.
- Приведена вартість всіх чистих прибутків підприємства від реалізації результатів наукової розробки - 509784,3 грн.
- Абсолютна ефективність вкладених інвестицій – 450564 грн.
- Відносна ефективність вкладених в наукову розробку інвестицій - 105 %
- Термін окупності вкладених у реалізацію наукового проекту інвестицій - 0,95 року, що свідчить про доцільність фінансування нової розробки.

# Апробація результатів роботи та публікації

1. XLV науково-технічній конференції факультету інформаційних технологій та комп'ютерної інженерії Вінницького національного технічного університету (м. Вінниця, Україна, 2016 р.)
2. X міжнародна науково-практична конференція «ІНТЕРНЕТ-ОСВІТА-НАУКА-2016» (ІОН-2016)
3. XLVI науково-технічній конференції факультету інформаційних технологій та комп'ютерної інженерії Вінницького національного технічного університету (м. Вінниця, Україна, 2017 р.)



# ВИСНОВОК

- Показано актуальність задачі розпізнавання нот та виконано аналіз предметної області.
- Наведено математичну функції втрат та розроблено алгоритм функціонування програмного модуля розпізнавання нот. Обґрунтовано вибір платформи та мови програмування. Розроблено інструкцію користувача програмного продукту.
- Протестовано розроблену програму та здійснено аналіз отриманих результатів. Результати, отримані під час тестування роботи програми, підтвердили правильність її функціонування.
- Всі задачі поставлені в завданні виконані в повному об'ємі.
- Мета дослідження — підвищення точності розпізнавання нот музичних композицій на основі цифрового аудіо-запису за рахунок використання інформаційної технології на основі згорткової нейронної мережі досягнута (0,752 для простих музичних композицій та 0,752 для складних мультиінструментальних композицій).

Дякую за увагу!