

Міністерство освіти і науки України
Національний технічний університет України
“ Київський політехнічний інститут імені Ігоря Сікорського ”
“ Інститут прикладного системного аналізу ”

Дипломна робота бакалавра

на тему:

***«Система розпізнавання музичного стилю з використанням
глибинного навчання»***

Виконав:

Студент групи КА-31

Кулік О.П.

Науковий керівник:

к.т.н., доцент

Тимощук О.Л.

Київ-2017

Актуальність роботи

- ✓ Автоматична структуризація музики, що доступна в цифровій формі
- ✓ Рекомендаційні системи в музичних сервісах

Мета роботи:

Розробка системи розпізнавання музичного стилю композиції з використанням глибокого навчання

Об'єкт дослідження:

Розпізнавання музичного стилю

Предмет дослідження:

Методи класифікації звукових сигналів

Постановка задачі

- ✓ Провести огляд джерел існуючих методів отримання ознак звукового сигналу, методів класифікації та програмних засобів
- ✓ Обрати та обґрунтувати доцільність алгоритму розв'язання задачі
- ✓ Створити програмний продукт, що буде реалізовувати заданий алгоритм
- ✓ Провести тестування побудованої системи
- ✓ Зробити висновки та запропонувати рекомендації щодо подальших напрямів досліджень

Існуючі методи виділення ознак звукОВОГО сигналу

1) Мел-частотні кепстральні коефіцієнти

+ в результаті отримується невеликий набір значень, що представляють весь звуковий сигнал

- втрата деяких компонентів звукового сигналу, метод не дає загальну картину про композицію

2) Спектрограма

+ зберігає майже всі характеристики звукового сигналу

- таке представлення звукового сигналу є досить громіздким

Існуючі алгоритми класифікації

1) Метод опорних векторів

- + метод знаходить роздільну полосу максимальної ширини, що дозволяє надалі здійснювати кращу класифікацію
- + метод зводиться до рішення задачі квадратичного програмування, у випуклій області, яка завжди має єдине рішення
- повільне навчання
- метод чутливий до шумів та нормалізації даних

2) Дерева рішень

- + результати навчання легко інтерпретувати через деревоподібну структуру
- + висока швидкість навчання
- часто формуються занадто складні дерева, які мають ефект «перенавчання»
- дерева рішень не можуть гарантувати збіжності до оптимального рішення.

3) Згорткові нейронні мережі

- + стійкість до зашумлених даних
- + висока точність розпізнавання на специфічних даних
- потребують значних комп'ютерних ресурсів для ефективного навчання
- багато параметрів, які потрібно налаштовувати вручну

Вихідні дані

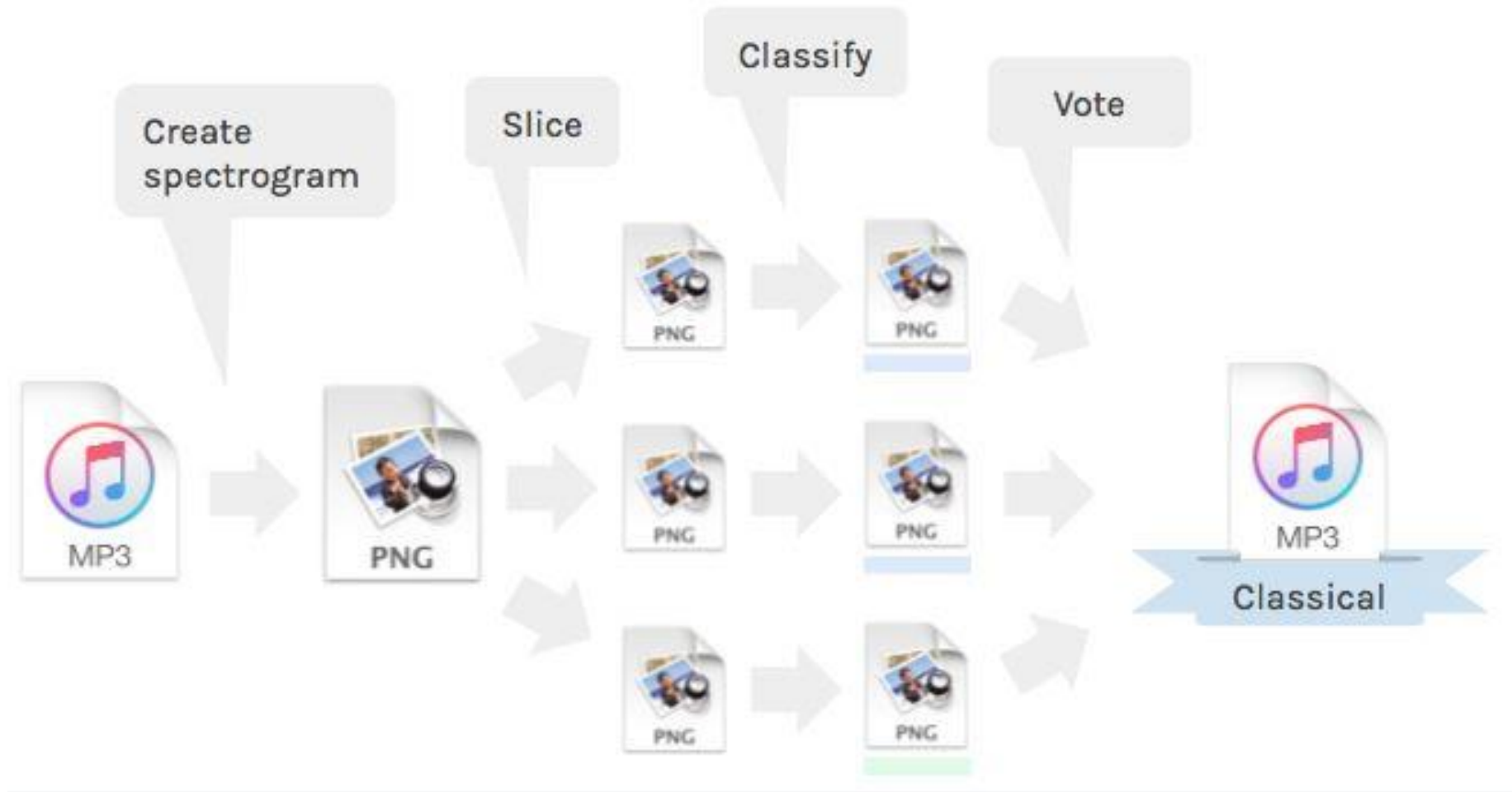
Набір даних: «GTZAN Genre Collection»

- 1000 композицій
- 10 стилів
- 100 композицій/стиль
- Тривалість кожної композиції: 30 с
- 80% – тренувальна вибірка
- 20% – тестова вибірка

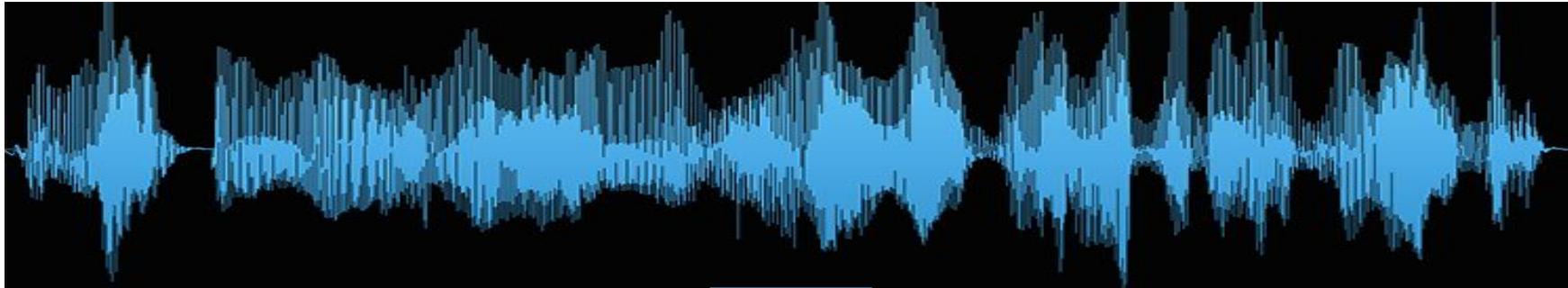
Представлені стилі:

- 1) blues
- 2) classical
- 3) country
- 4) disco
- 5) hip-hop
- 6) jazz
- 7) metal
- 8) pop
- 9) reggae
- 10) rock

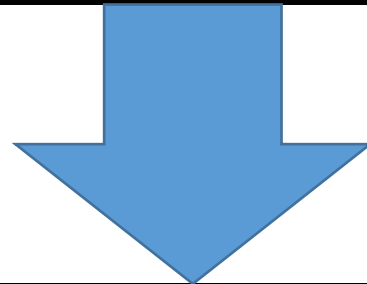
Функціональна схема роботи системи



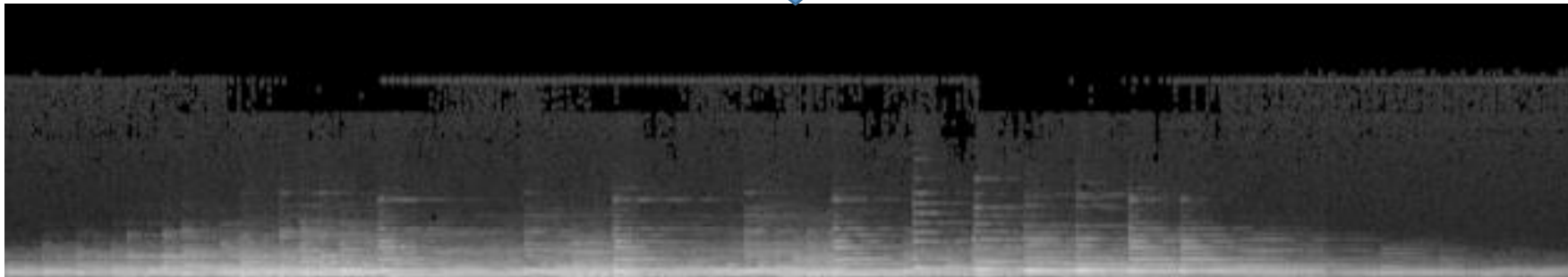
Етап 1: Побудова спектрограми



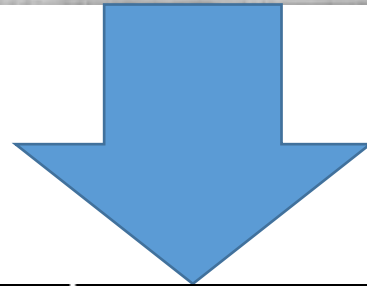
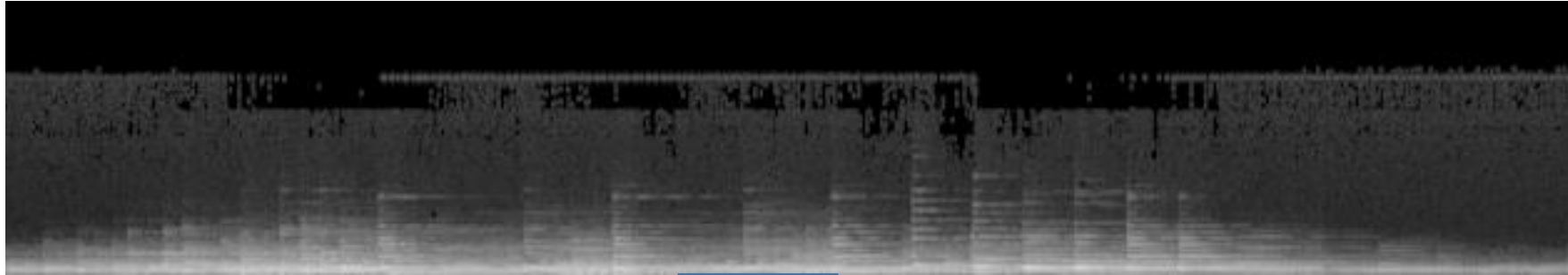
Віконне перетворення
Фур'є



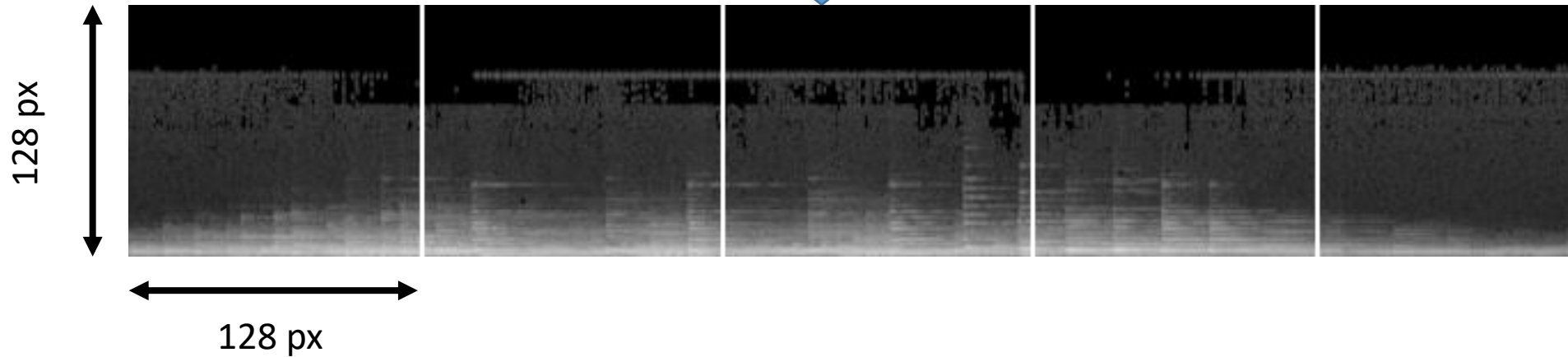
$$F(m, \omega) = \sum_{n=-\infty}^{\infty} f[n] \omega[n - m] e^{-j\omega n}$$



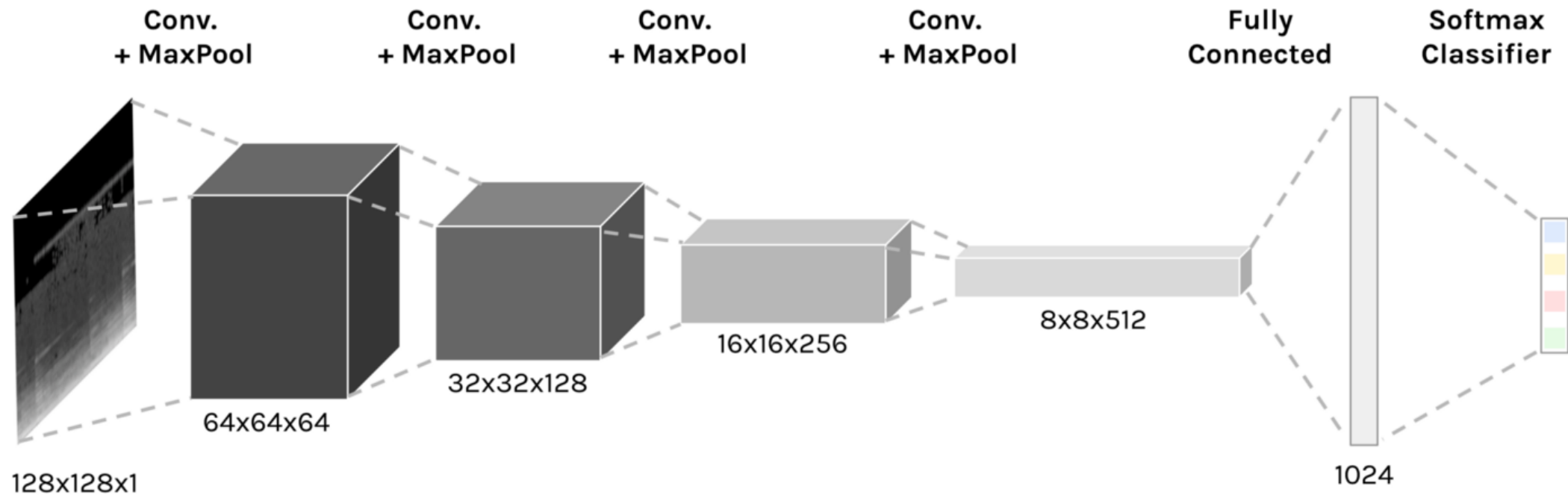
Етап 2: Розбивка спектрограми



128 px = 2,56 c



Етап 3: Класифікація спектрограм



Згорткова нейронна мережа

Основний принцип – згортка:

$$(f * g)[m, n] = \sum_{k, l} f[m - k, n - l] \cdot g[k, l],$$

де f – вихідна матриця зображення;
 g – ядро (матриця) згортки.

Згортковий шар:

$$x^l = f(x^{l-1} * k^l + b^l),$$

де x^l – вихід шару l ;

$f()$ – функція активації;

b – коефіцієнт зсуву;

$*$ – оператор згортки входу x з ядром k .

$$x_j^l = f\left(\sum_i x_i^{l-1} * k_j^l + b_j^l\right),$$

де x_j^l – карта ознак j (вихід шару l);

f – функція активації;

b – коефіцієнт зсуву для карти ознак j ;

k – ядро згортки номер j ;

x_i^{l-1} – карта ознак попереднього шару.

Підвибірковий шар:

$$x^l = f(a^l \cdot \text{subsample}(x^{l-1}) + b^l)$$

де $\text{subsample}()$ – операція вибірки локальних максимальних значень.

Багатошаровий перцептрон:

$$x_j^l = f\left(\sum_i x_i^{l-1} \cdot w_{ij}^{l-1} + b_j^{l-1}\right),$$

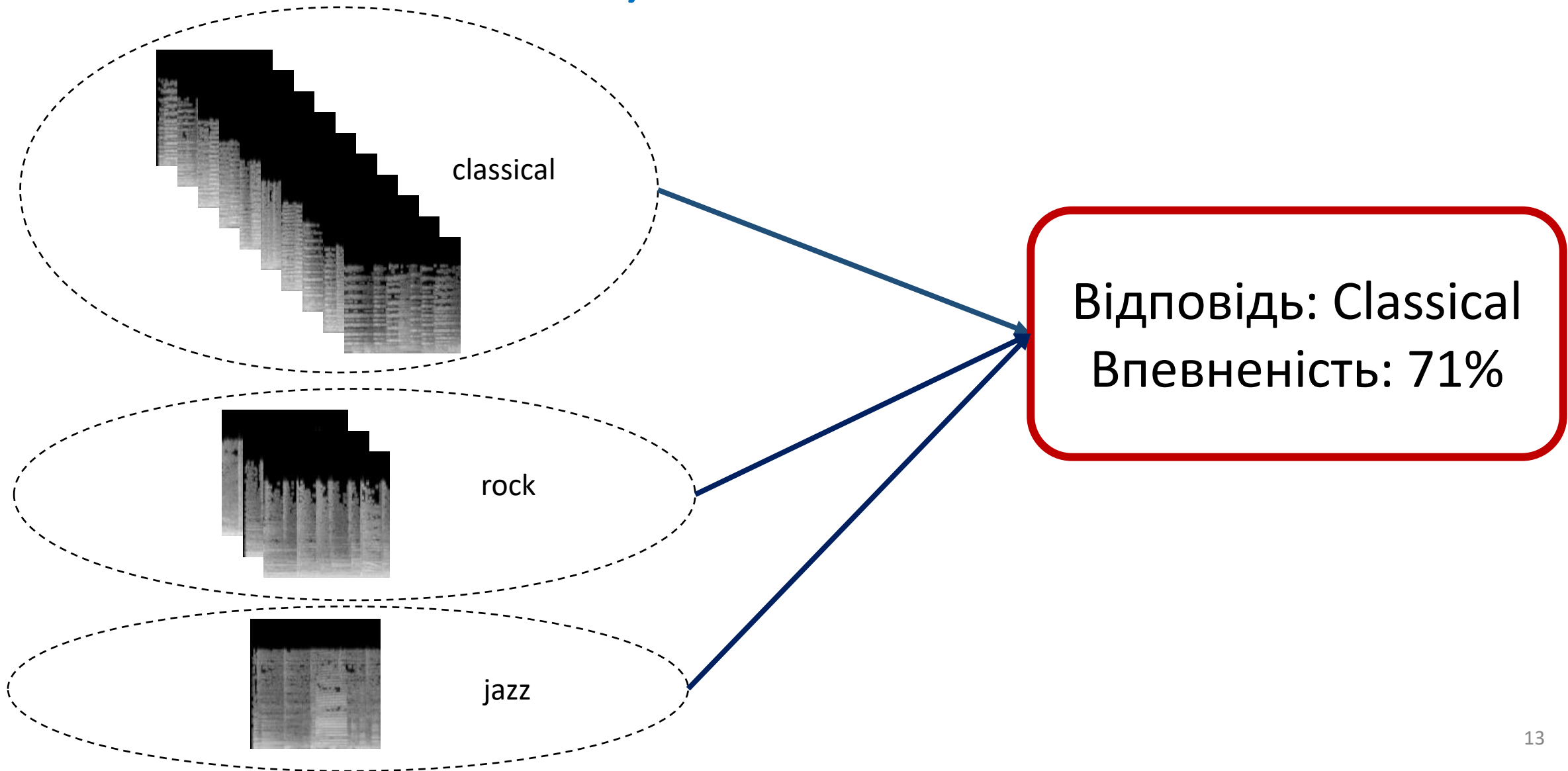
де x^l – вихід шару l ;

$f()$ – функція активації;

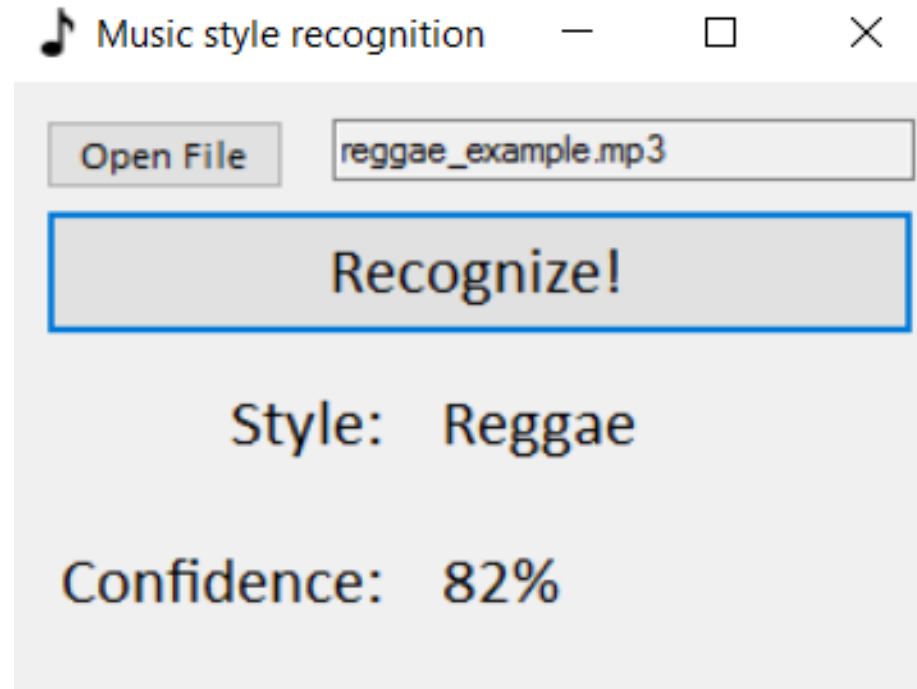
b – коефіцієнт зсуву;

w – матриця вагових коефіцієнтів.

Етап 4: Голосування



Інтерфейс програмного продукту



Аналіз якості отриманих результатів

Критерій	Значення
Точність згорткової мережі на частинах спектрограм	72%
Точність всієї системи розпізнавання музичного стилю	76%

Матриця помилок системи

actual pred. \	classical	country	disco	hip-hop	jazz	rock	blues	reggae	pop	metal
classical	17									
country		15				1				
disco			15	1		4	2	1	2	
hip-hop			2	17				3	1	
jazz	2				18				1	
rock	1	3			1	12	3		1	6
blues		1			1		13			
reggae		1	1	2		3	2	16		
pop			2						15	
metal										14
<i>Style acc:</i>	85%	75%	75%	85%	90%	60%	65%	80%	75%	70%

Висновки

- ✓ Проведено огляд існуючих методів виділення ознак звукового сигналу, методів класифікації та програмного забезпечення
- ✓ Розроблено програмний продукт, що дозволяє розпізнавати музичний стиль композиції
- ✓ Проведено аналіз якості створеної системи, точність розпізнавання всієї системи складає 76%

Шляхи подальшого розвитку

- ✓ Побудова ще однієї нейронної мережі для бінарної класифікації музичної композиції та інших звукових сигналів
- ✓ Вдосконалення системи голосування
- ✓ Розширення тренувальної вибірки за рахунок інших наборів даних
- ✓ Створення дерева нейронних мереж для зменшення кількості класів для кожної нейронної мережі

Дякую за увагу!