## Задача.

### Постановка задачі:

Створення системи штучного інтелекту розпізнавання мовлення, а саме - розпізнавання аудіо цифр від 0 до 9.

### Мотивація**:**

Розпізнавання мовлення приносить можливість машинам слухати і розуміти, що люди говорять або що користувач командує. Оскільки цифри зустрічаються у наших розмовах дуже й дуже часто, їх розпізнавання є першим кроком до створення повної AI системи.

### Формалізація задачі:

Для реалізації завдання використано бібліотеку Keras та фреймворк Tensorflow.

**Вхідні дані:** аудіофайл формату WAV

**Вихідні дані:** конкретне число типу Int - результат розпізнавання

**Навчальні дані:** аудіофайли формату WAV

### **Датасет**:

Free Spoken Digit Dataset (<https://github.com/Jakobovski/free-spoken-digit-dataset>), ліцензія [Creative Commons Attribution-ShareAlike 4.0 International](https://creativecommons.org/licenses/by-sa/4.0/)

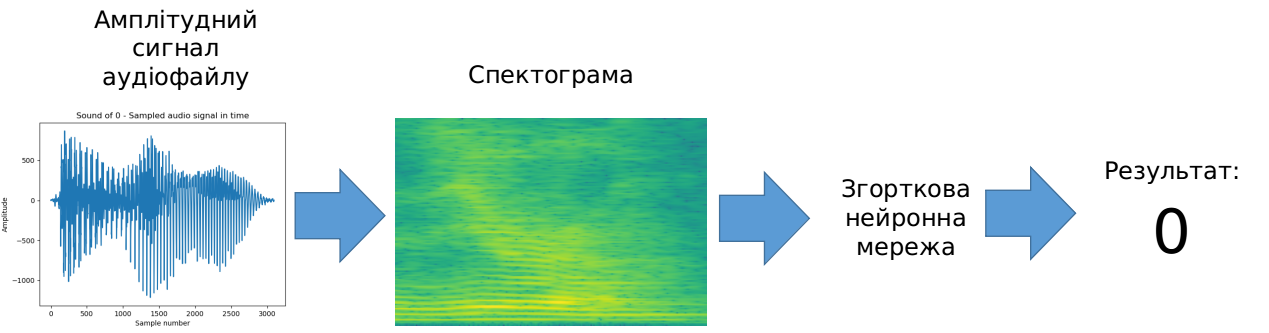
#### Характеристики

* 4 виконавці
* 2,000 записів (50 записів кожної цифри на спікера)
* Англійська вимова
* монофайли 8kHz, формат wav

**Попередня обробка даних перед використання в навчальній машині:**

* Виділяємо аудіосигнали з WAV файлів
* Перетворюємо аудіосигнали в спектограми розміру 25х17
* Зберігаємо спекторами і визначаєчаємо їх як наші дані для навчальної машини.

Процес перетворення



## Нейромежева модель.

### Тип нейронної мережі

Одним із способів розпізнавання аудіоданих є тренування розпізнавання спектограм сигналів. Саме тому, нейронною мережею було вибрано згорткову нейронну мережу(CNN)

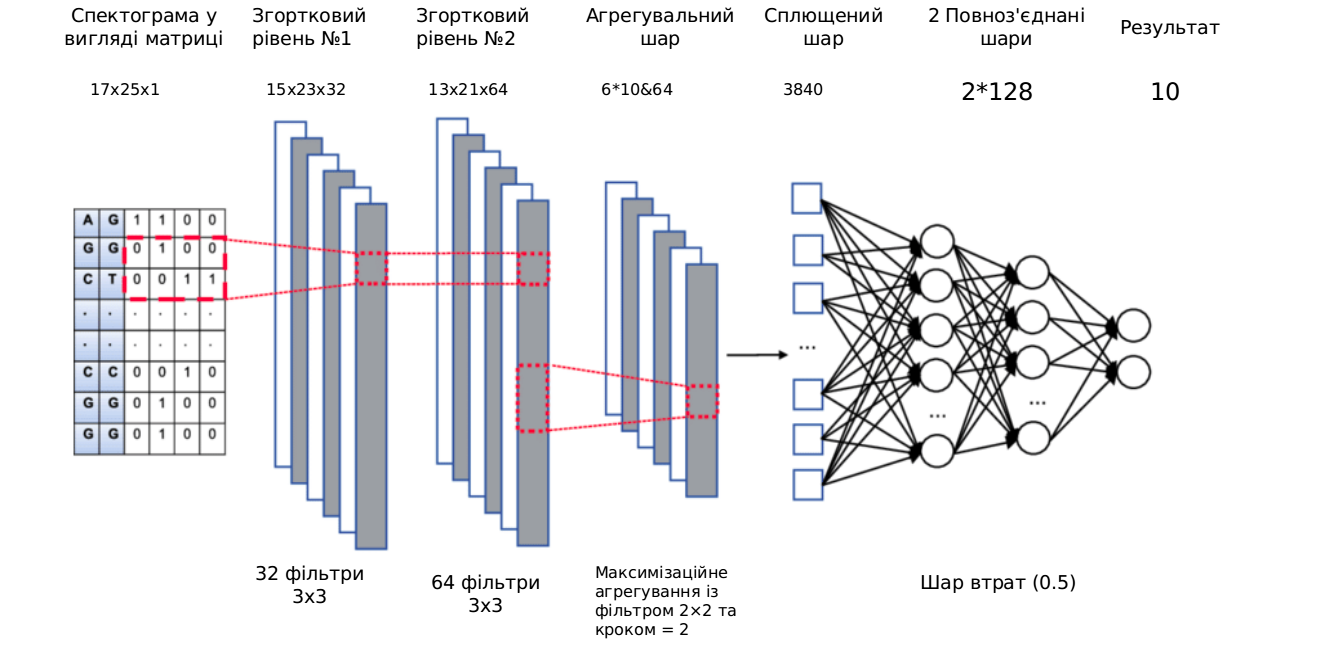
**Зго́рткова нейро́нна мере́жа** (*convolutional neural network*) — це клас глибинних штучних нейронних мереж прямого поширення, який успішно застосовувався до аналізу візуальних зображень.

### Структура мережі

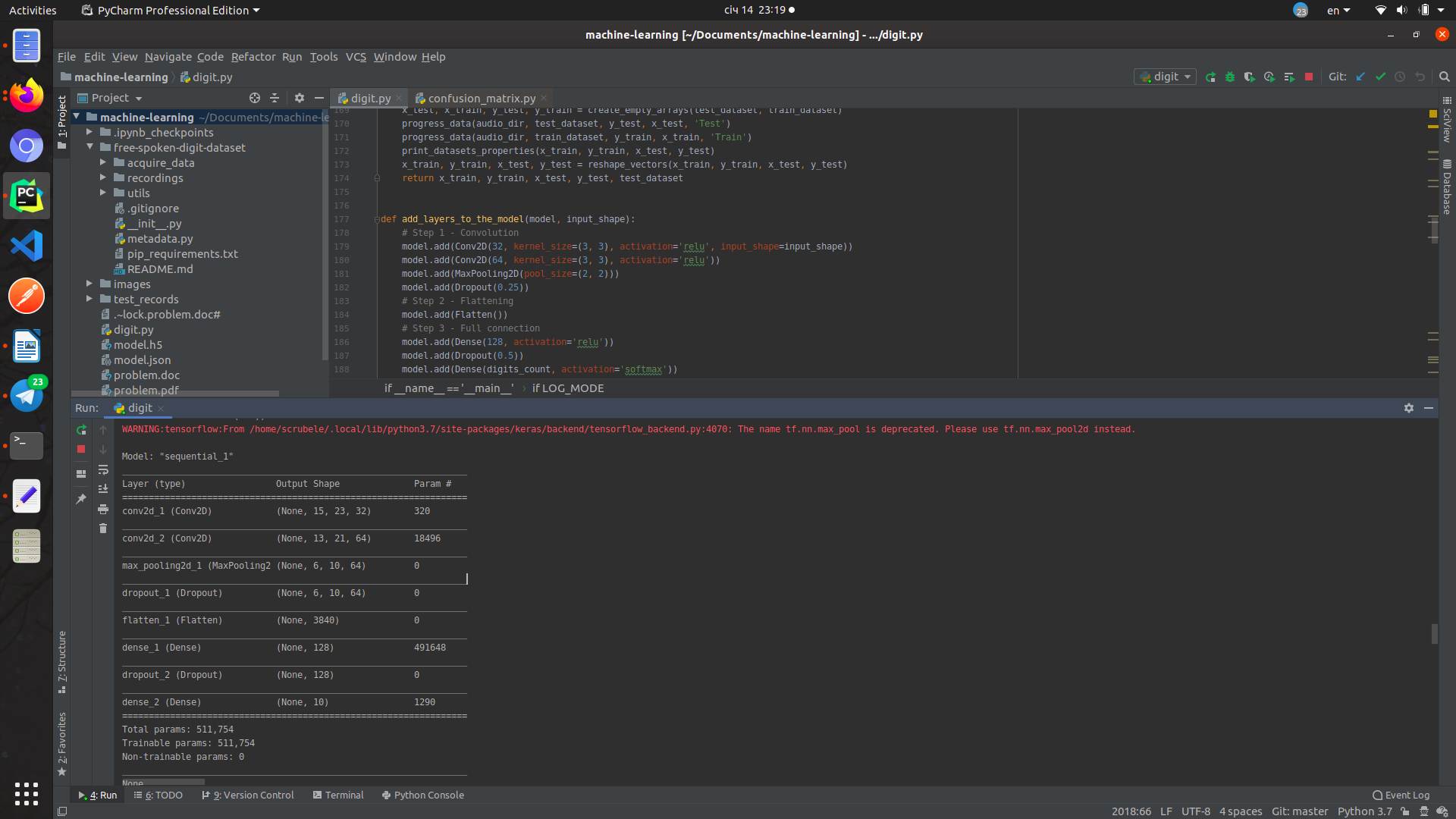
#### Вибрана модель:

Sequential (лінійній стек шарів)

#### **Характеристики шарів:**



Дані про шари, виведені програмою:



### Навчання мережі

Keras моделі тренуються на Numpy масивах, що складаються з вхідних даних та міток. Я поділила дані за співвідношенням:

* 10% - тестові дані
* 90% - дані тренування.

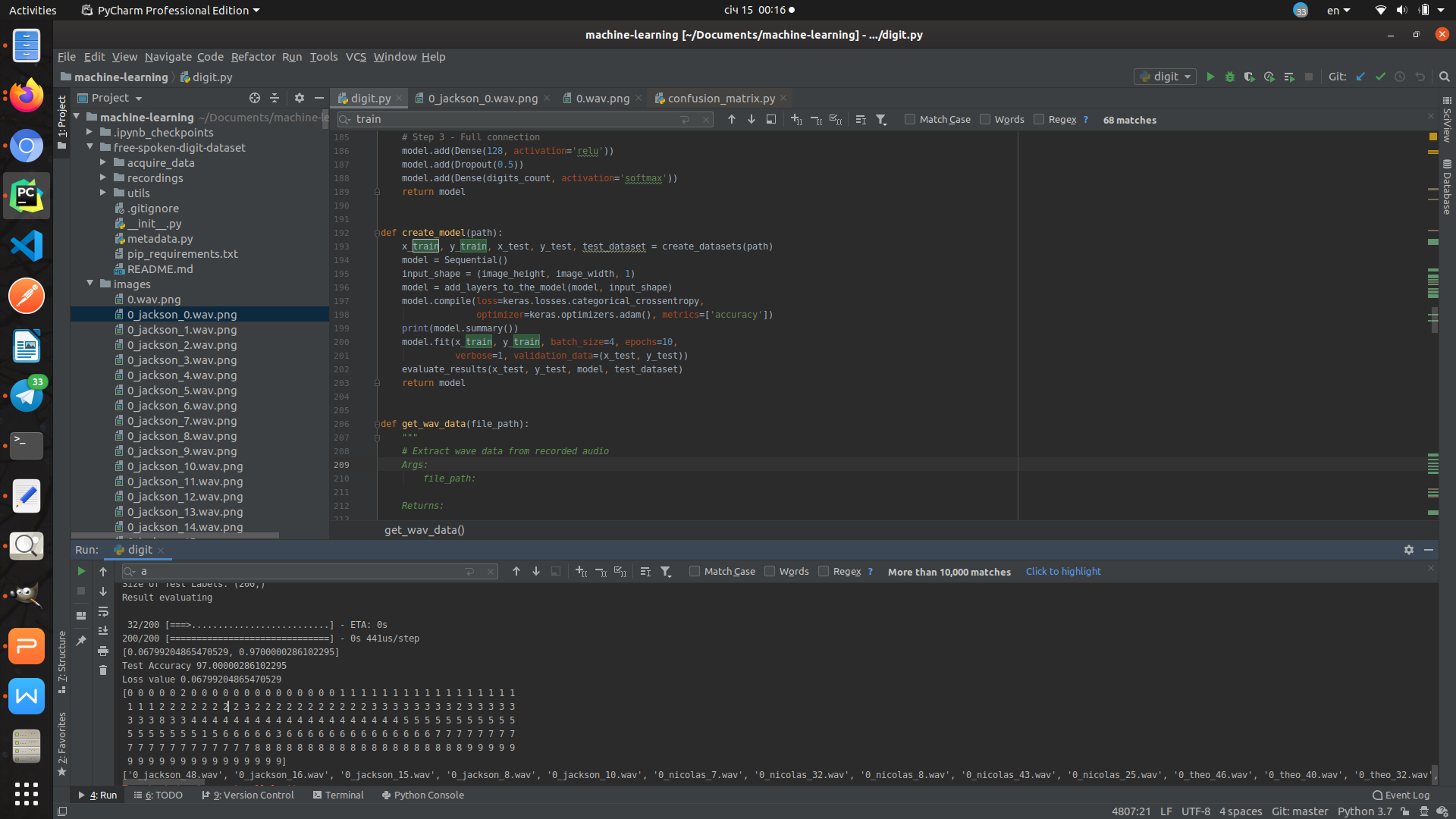
Я тренувала модель на фіксованій кількості епох. Приклад з коду:

*history = model.fit(x\_train, y\_train, batch\_size=4, epochs=1, verbose=1, validation\_data=(x\_test, y\_test))*

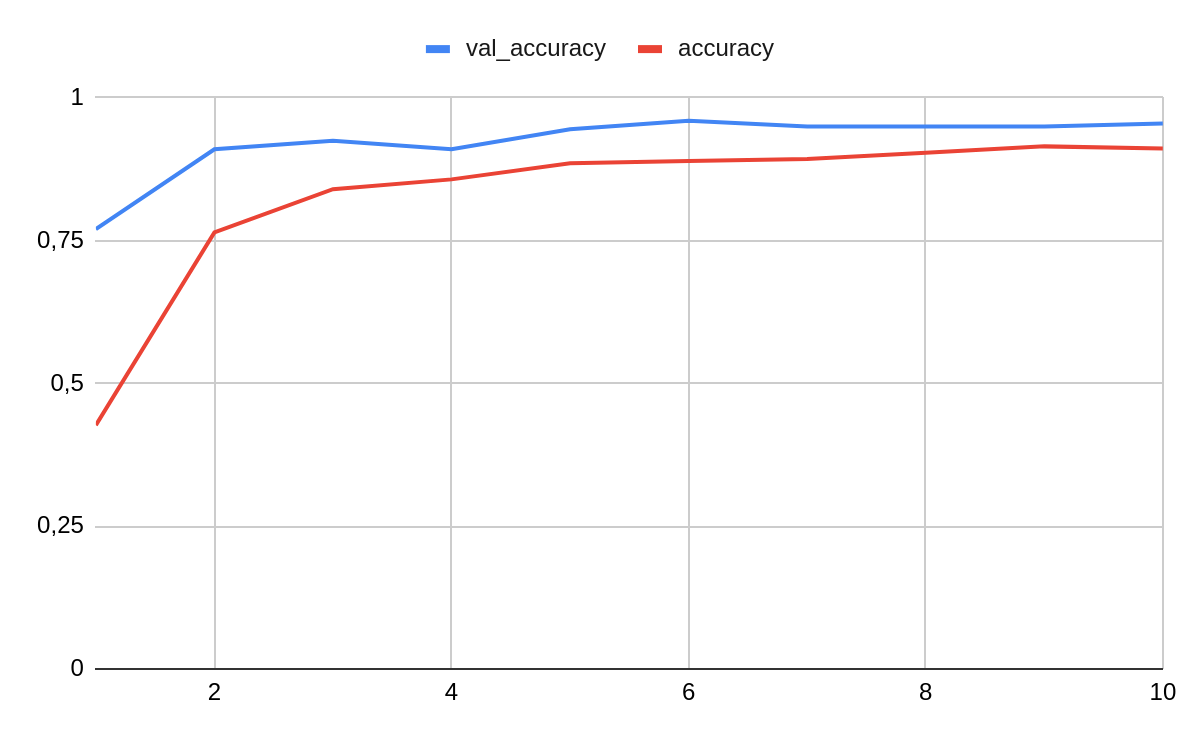
**epochs**: Integer. Кількість епох для навчання моделі. Епоха - це ітерація для всіх наданих даних x та y. Модель не навчається для ряду ітерацій, заданими епохами, а лише до досягнення епохи індексних епох.

### Результати застосування нейронної мережі

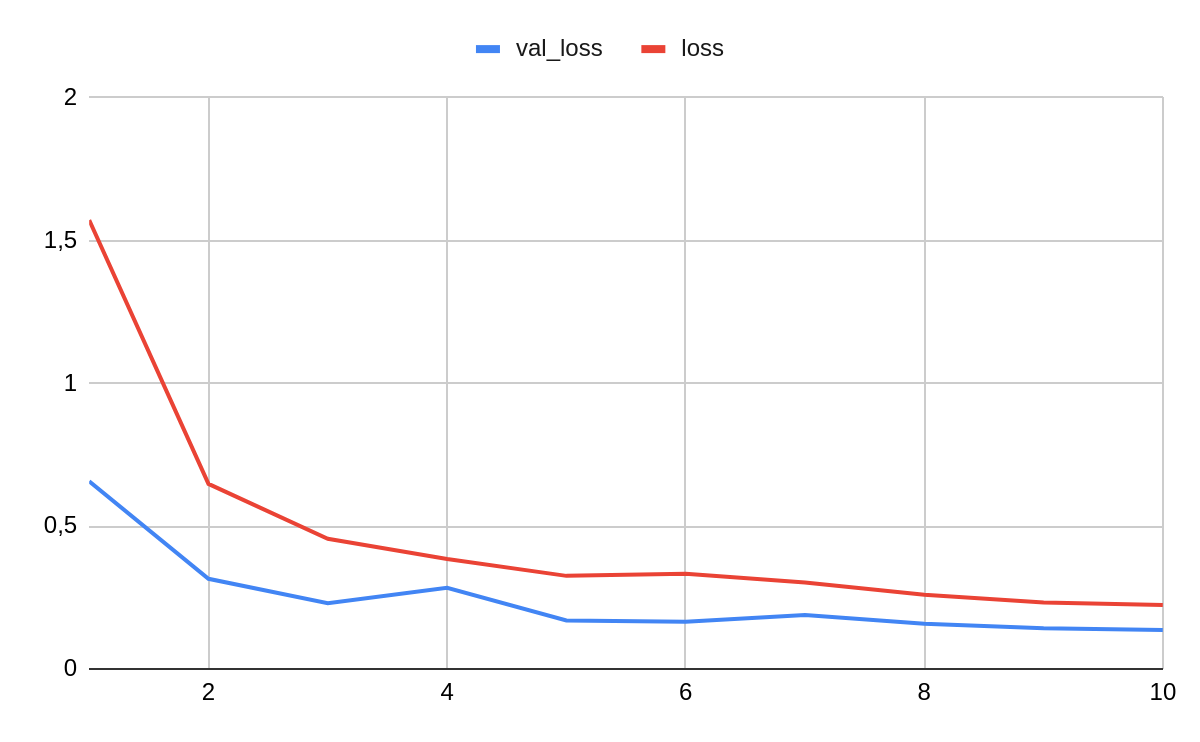
### **Кінцеві** результати:



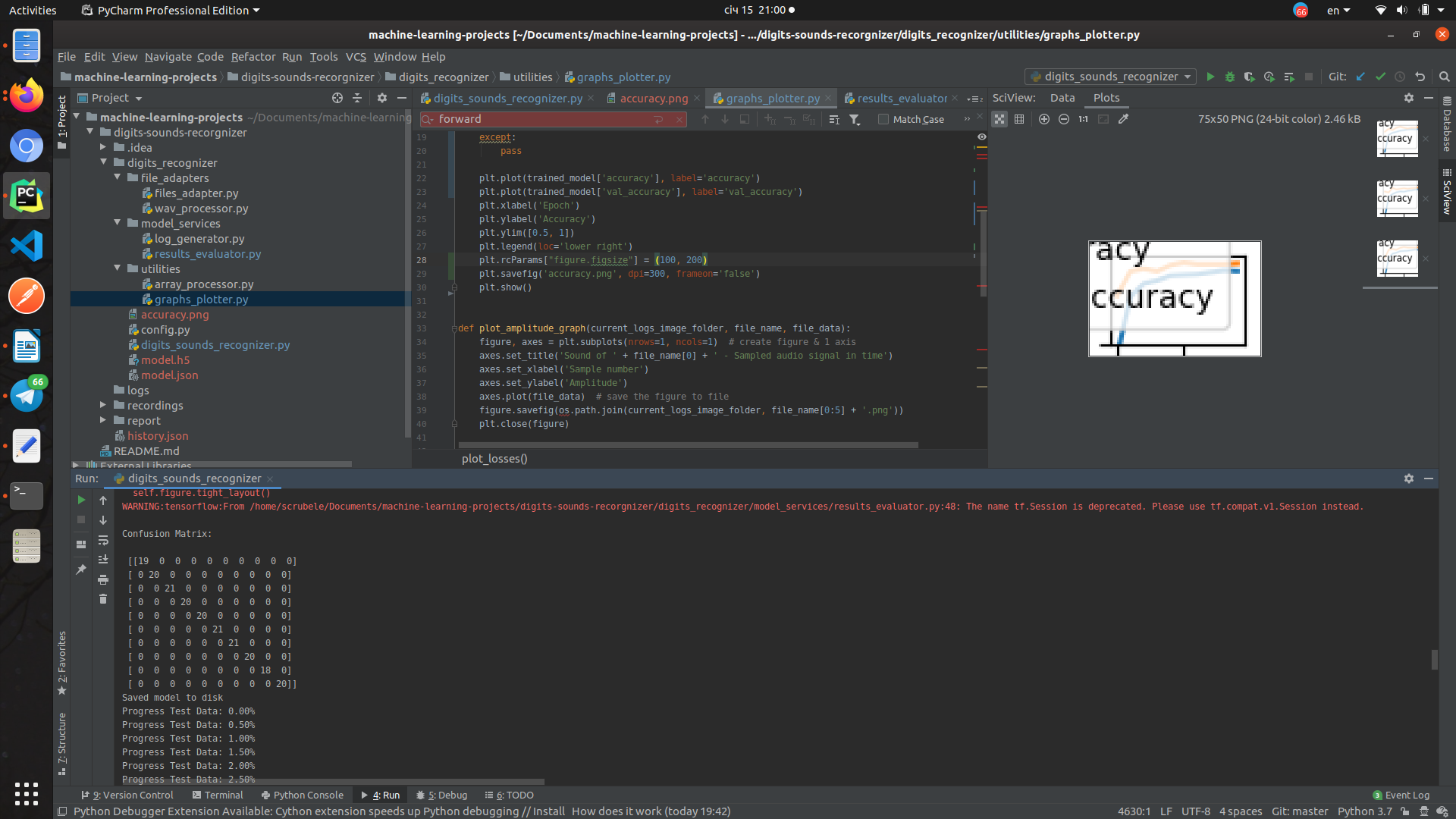
### Графік точності:



### Графік функції втрат:



### **Матриця** помилок:



### Висновки

Обробку даних проведено успішно. Отримала високу точність та мале значення функції втрат, що свідчить про хороше розпізнавання аудіоданих.