
Дипломна робота

Засоби ідентифікації людини за голосом
на основі штучних нейронних мереж

Керівник: професор кафедри ММСА, д.т.н.
Мухін Вадим Євгенович

Виконав: студент групи КА-53
Троян Макар Романович

Об'єкт. Предмет. Мета.

Об'єкт дослідження:

Дискретизована акустична хвиля мовлення людини.

Предмет дослідження:

Методи ідентифікації людини за голосом

Мета дослідження:

Виявити найкращий спосіб ідентифікації людини за голосом та розробити на базі нього програмний продукт для демонстрації можливостей обраного методу.

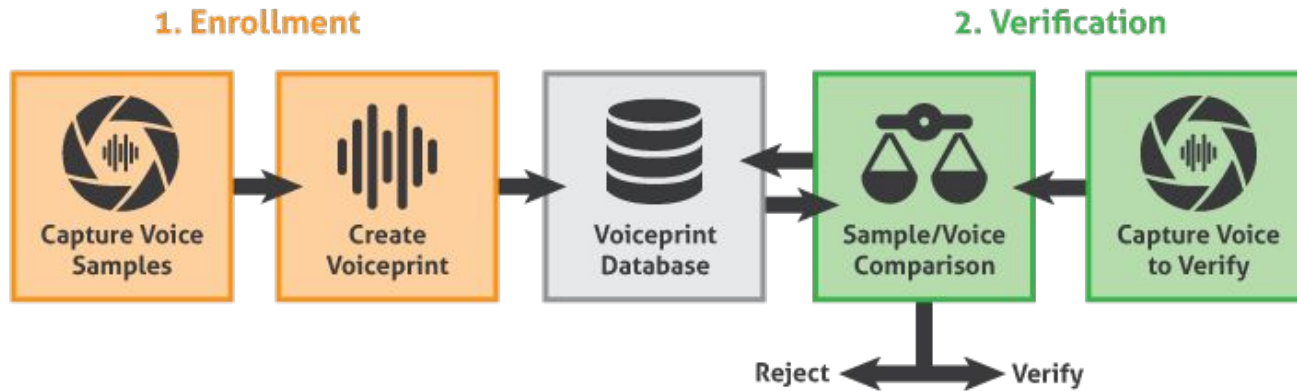
Постановка задачі

- Проаналізувати існуючі методи ідентифікації людини за голосом
- Дослідити різницю між класичними методами обробки сигналу та використанням глибоких штучних нейронних мереж
- Реалізувати програмний продукт для демонстрації можливостей обраного методу.

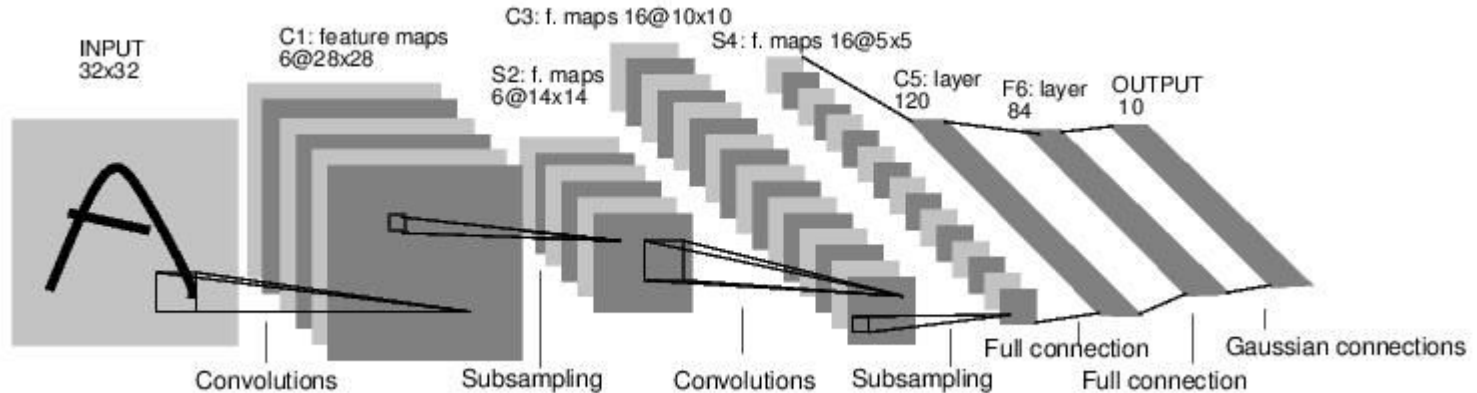
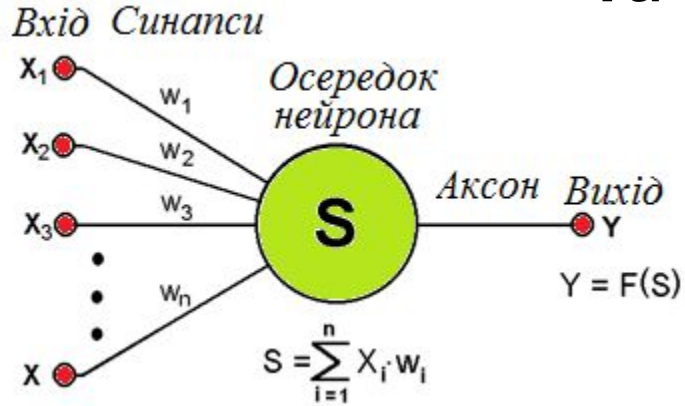
Актуальність проблеми



Загальна концепція системи

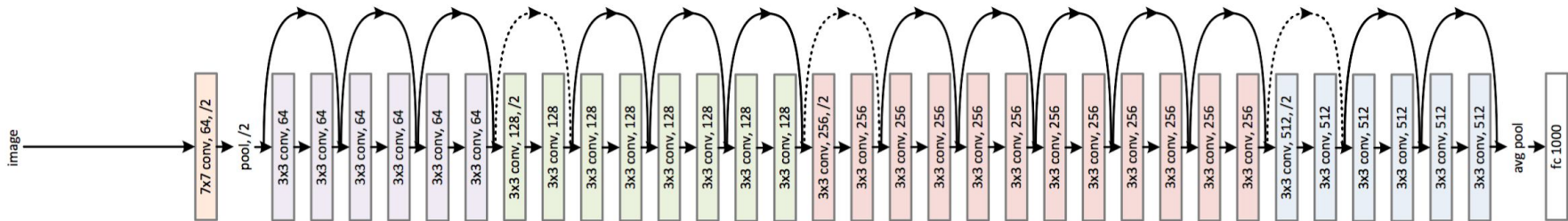


Приклад моделі штучного нейрона та згорткової нейромережі

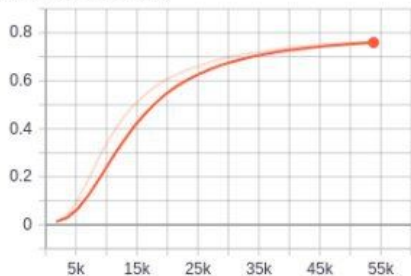


Вилучення ключових особливостей

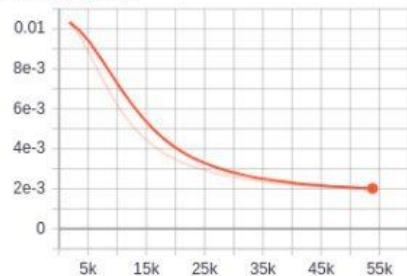
34-layer residual



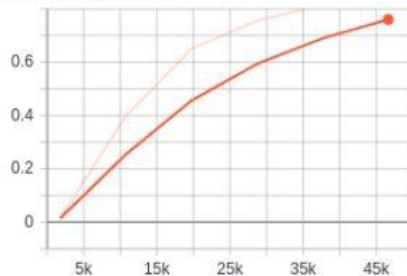
test_accuracy
tag: classifier/test_accuracy



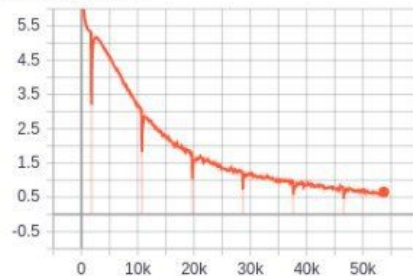
test_loss
tag: classifier/test_loss



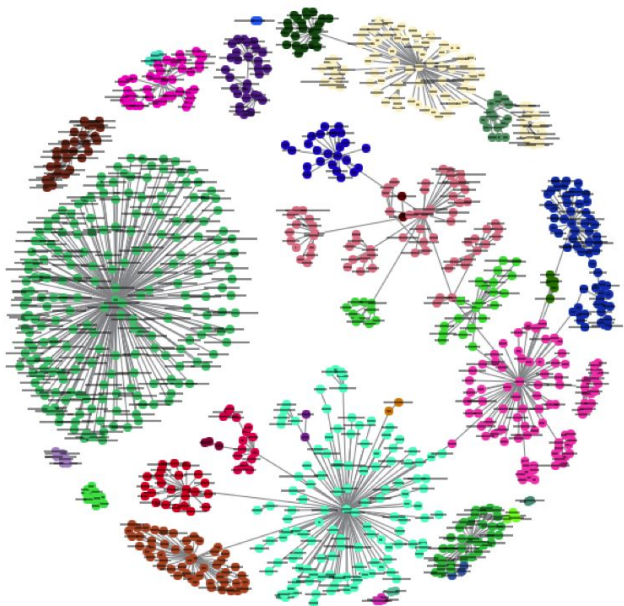
train_accuracy
tag: classifier/train_accuracy



train_loss
tag: classifier/train_loss



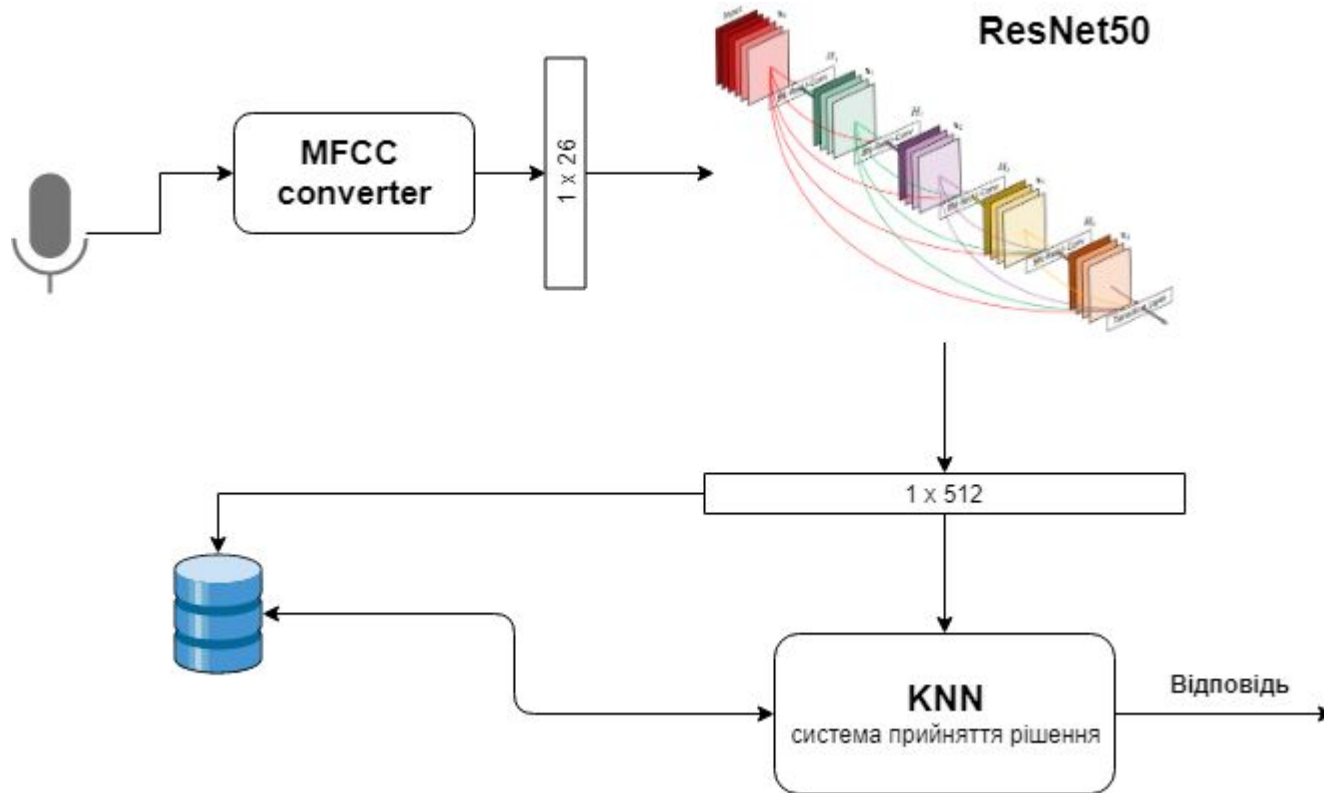
knn - метод прийняття рішення



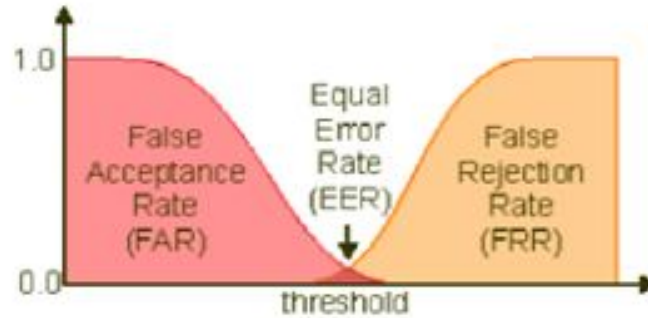
$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}}$$

Правило: отримується $n=3$ (embeddings) зі вхідного зразку (10 секунд). Для кожного шукається $k=10$ найближчих сусідів, зваженим за дистанцією голосуванням обирається “переможець” для кожного з n частин. Якщо хоча б два “переможця” співпадають, приймається рішення про ідентифікацію, в іншому випадку відхиляється.

Архітектура системи



Критерій якості



EER (Equal Error Rate) - рівень помилок біометричної системи, при якому

FRR (ступінь помилкової відмови)

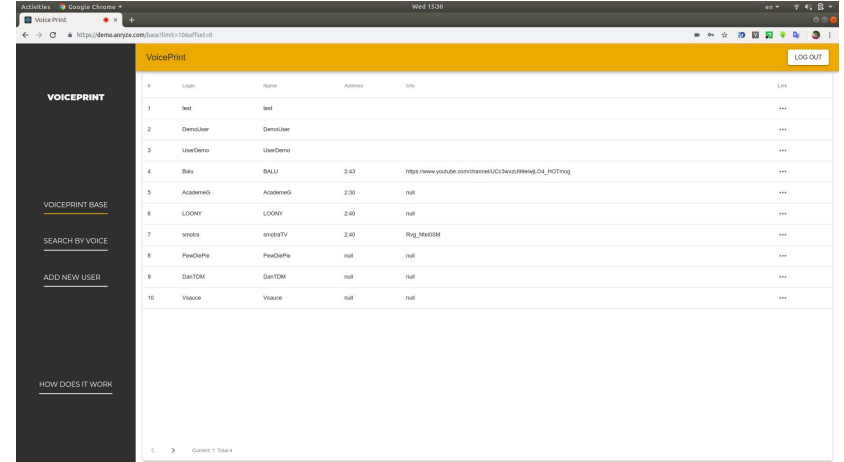
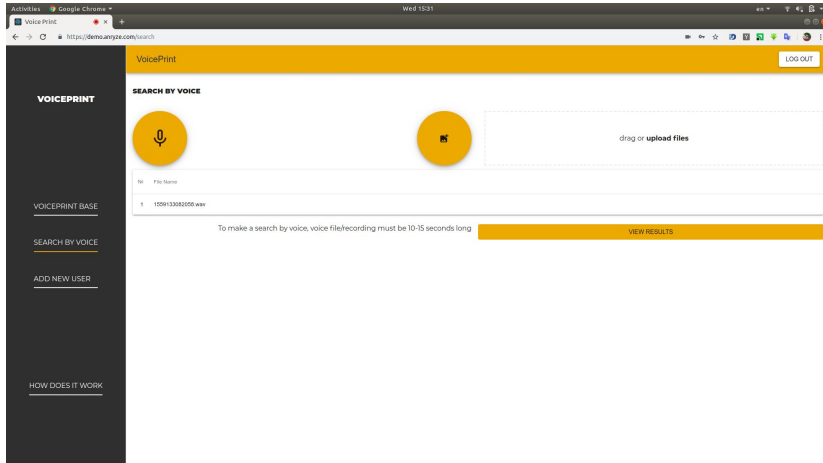
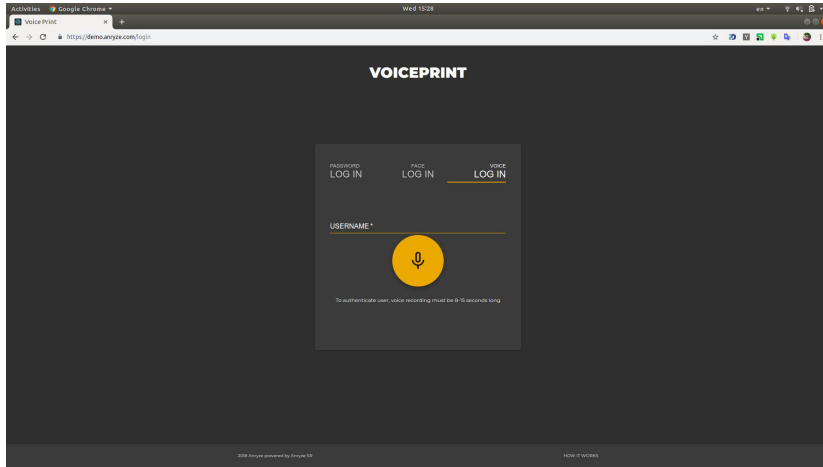
=

FAR (ступінь помилкового прийняття)

Результати роботи системи

Конфігурація embedder-y	EER
VGG-M (1)	5.94
ResNet-34 (1)	5.04
ResNet-34 (2)	5.11
ResNet-34 (3)	4.83
ResNet-50 (1)	4.19
ResNet-50 (2)	4.43
ResNet-50 (3)	3.95

Приклади роботи системи



У цій дипломній роботі мною було проаналізувати існуючі методи ідентифікації людини за голосом на основі MFCC та глибоких нейронних мереж, досліджено різницю між такими методами обробки сигналу як MFCC та LPCC для подальшого використання з глибокими штучними нейронними мережами. Також було реалізовано систему ідентифікації людини за голосом на основі згорткової нейронної мережі та knn . Для демонстрації можливостей обраного методу було реалізовано програмний продукт у вигляді веб додатку.

Подальші дослідження

- Дослідити більш глибоку нейронну мережу для отримання вектору ознак голосового прикладу мовника.
- Застосувати спеціальні методи для попередньої обробки аудіо запису: очищення від шумів та виявлення періодів на запису без мовлення людини.
- Провести аналіз роботи системи на різних конфігураціях обчислювальних машин для виявлення мінімально необхідних характеристик пристрою, на якому можна запускати програмний продукт.

Дякую за увагу
