

Класифікація емоцій в аудіозаписах української мови методами обробки природної мови

Автор: Жулкевський Владислав

студент МП КН 2

Науковий керівник: Ігнатенко Олексій
Петрович

Доктор фіз.-мат. наук



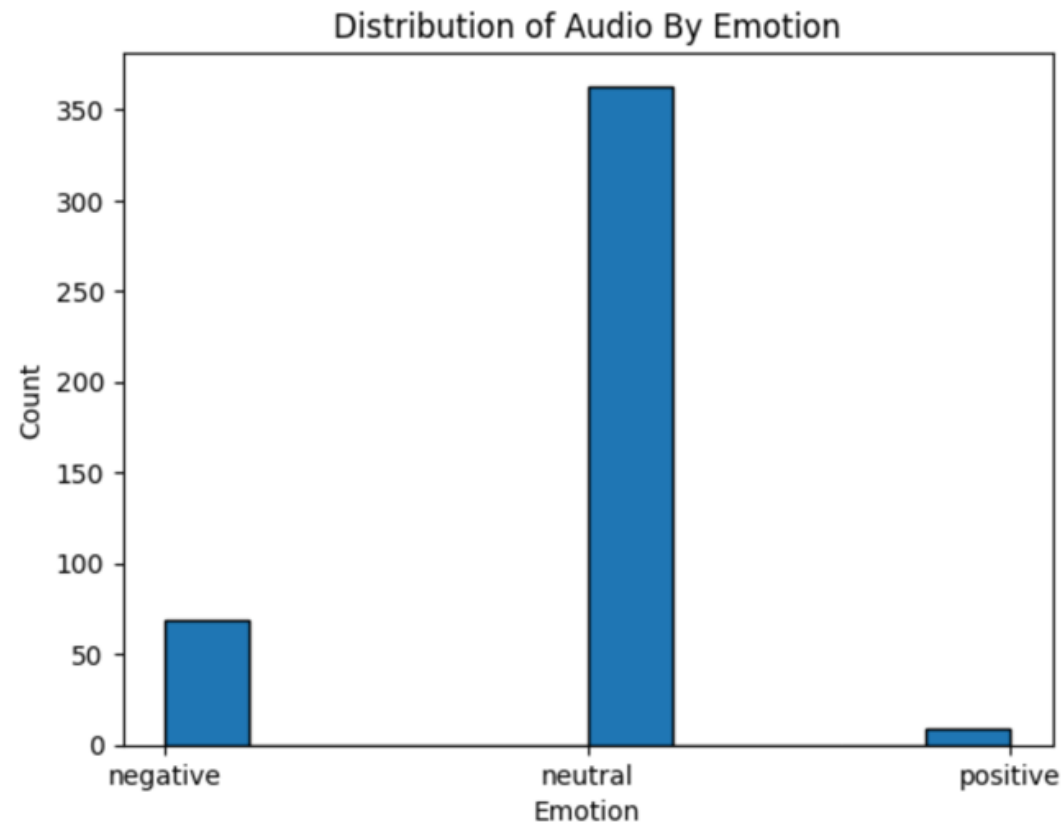
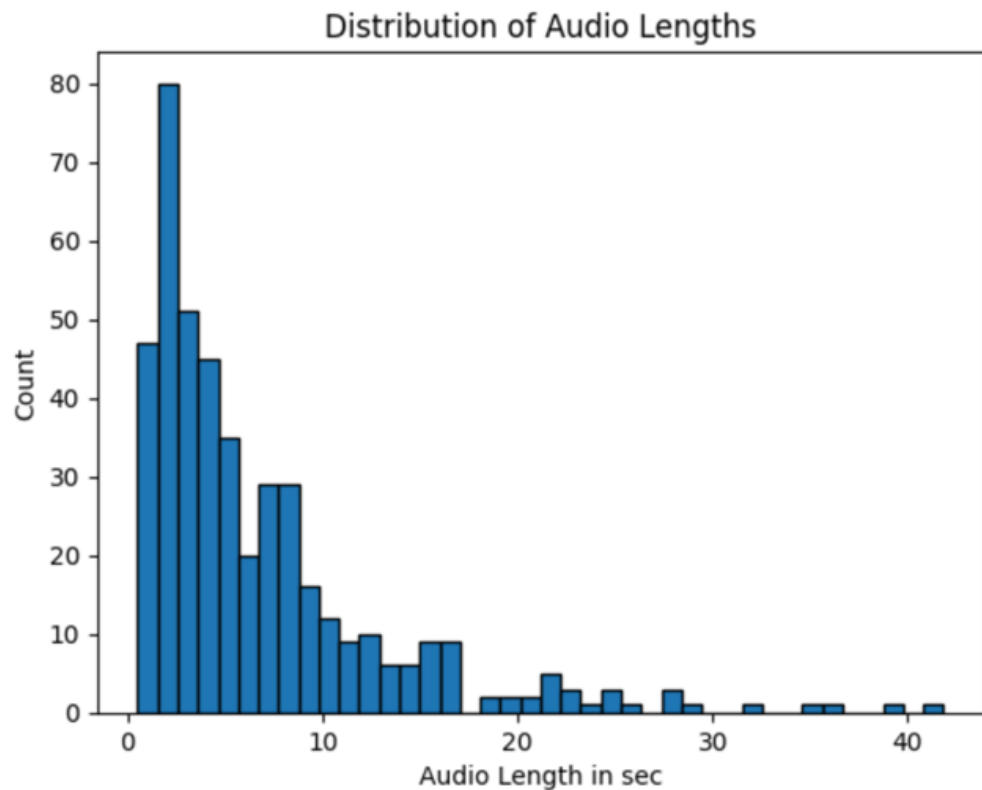
Тема

Кваліфікаційна робота полягає в розробці та аналізу моделей розпізнавання емоцій людей засобами обробки природньої мови. Основною метою роботи є дослідження алгоритмів машинного навчання, які можна буде використовувати для розпізнавання мовних сигналів та аналізу цих мовних сигналів на предмет виявлення емоцій співрозмовника.

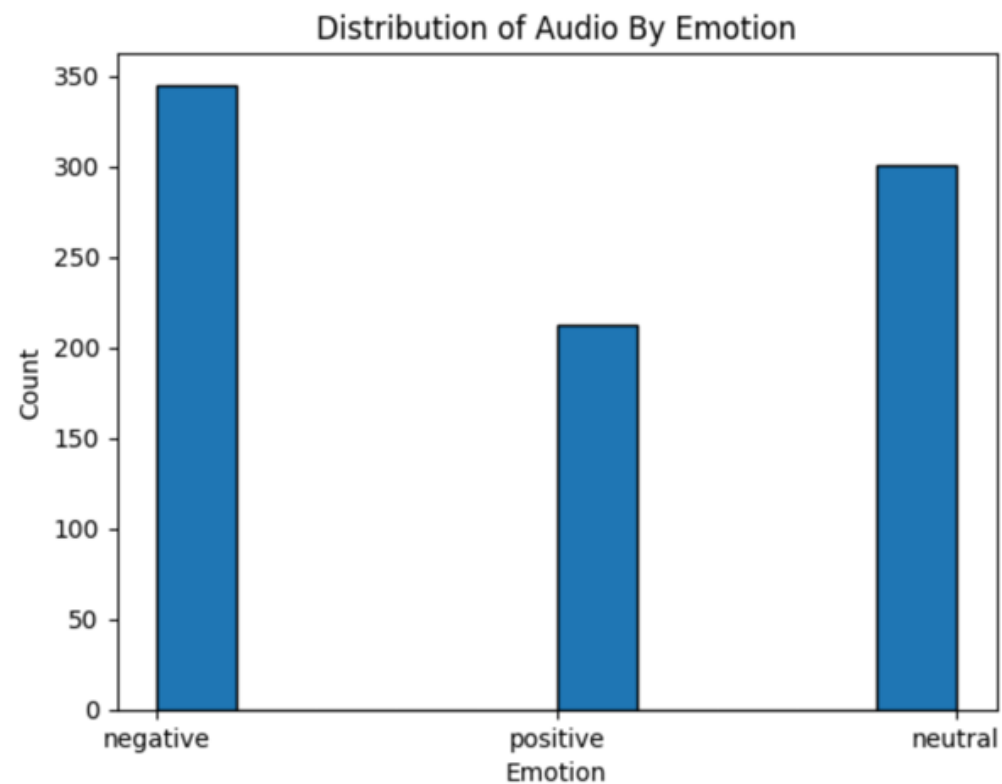
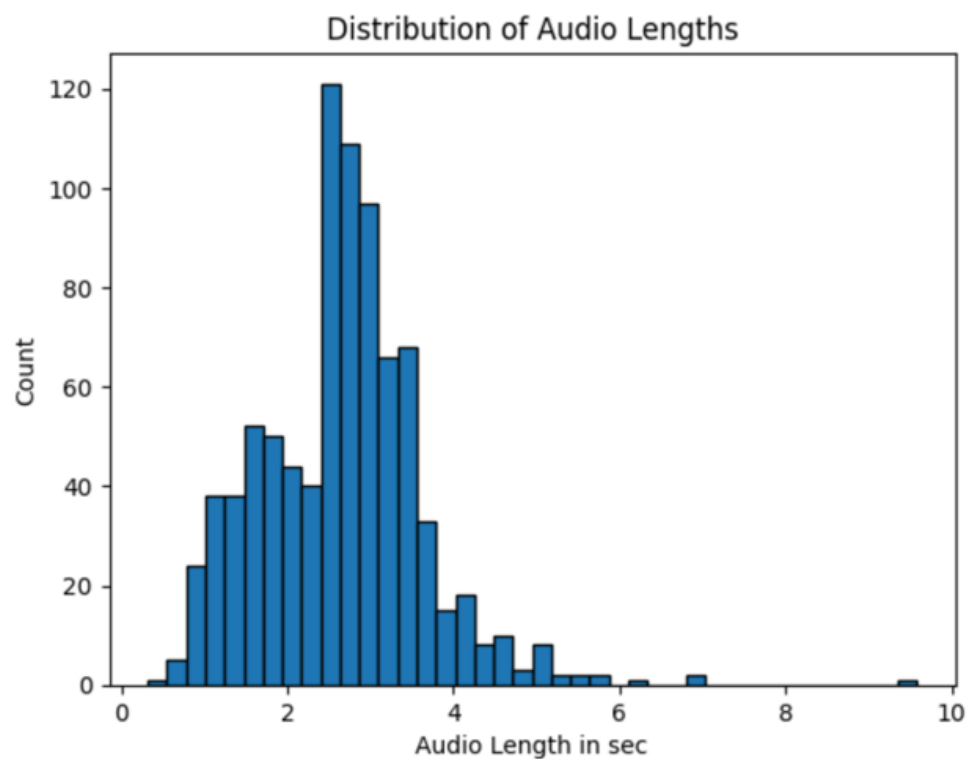
Опис проблеми

Наразі на ринку України немає досконалих рішень у сфері розпізнавання емоцій з аудіо природної мови. Є багато напрацювань у сфері Text-To-Speech та Speech-To-Text. Проте наразі не має достатніх досліджень для аналізу емоцій на основі ознак звукових сигналів української мови.

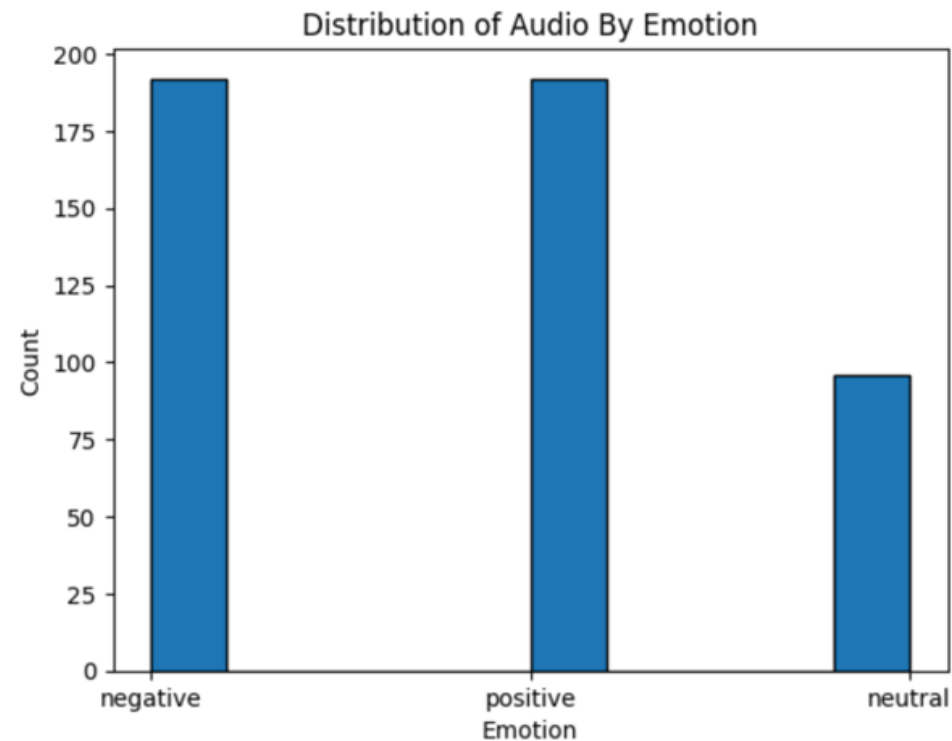
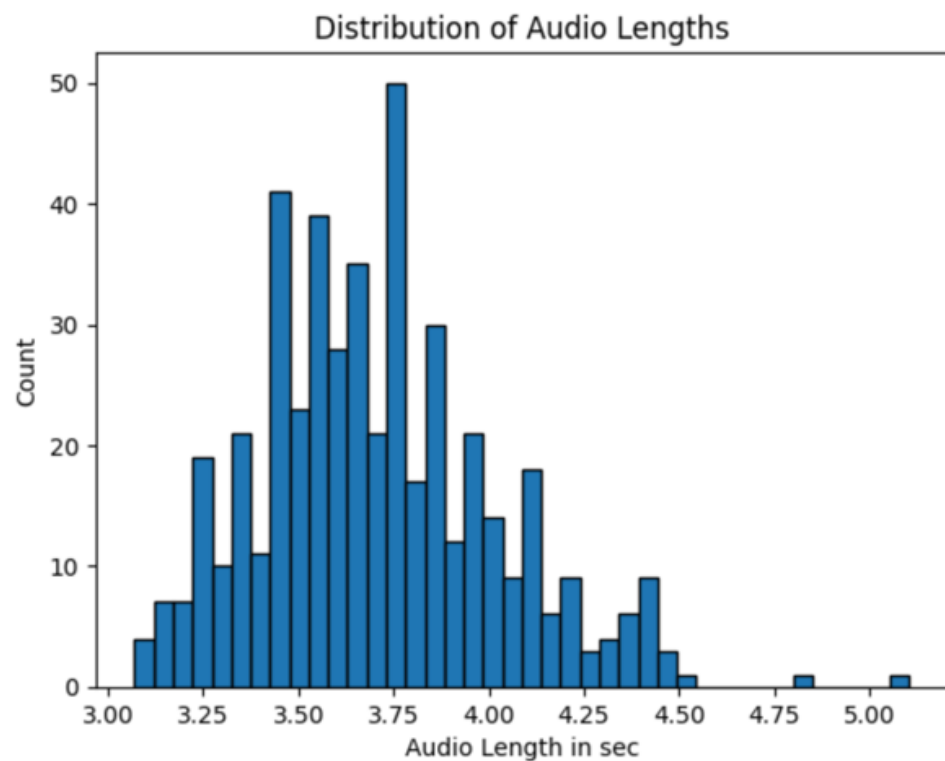
Збір даних – Датасет кол-центру



Датасет студії звукозапису

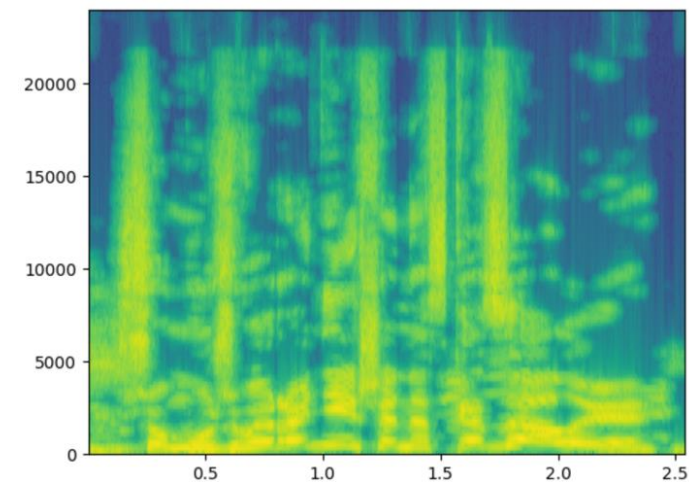
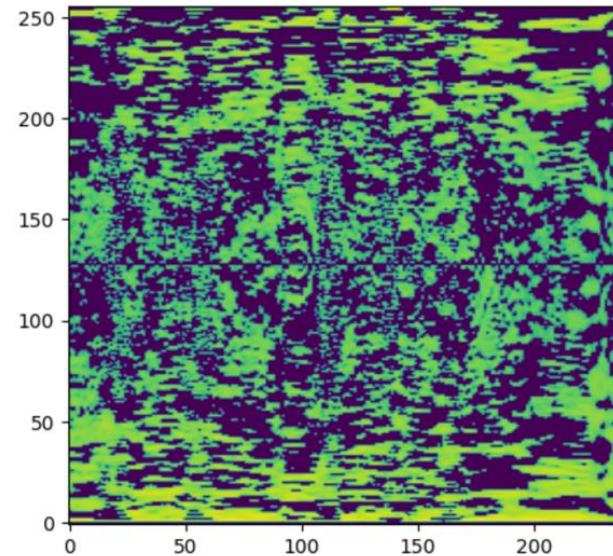
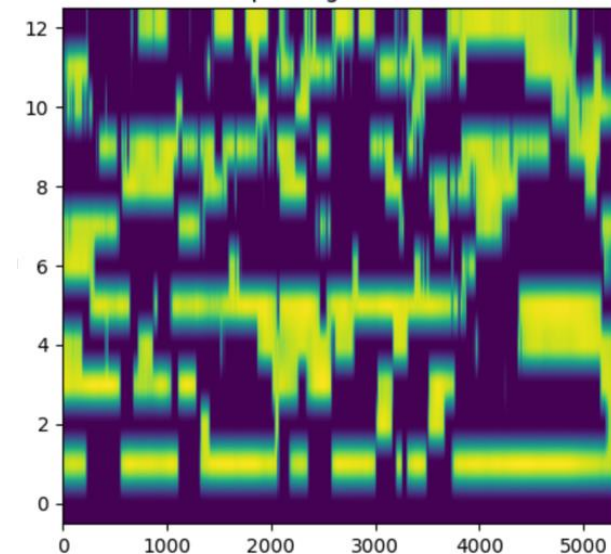
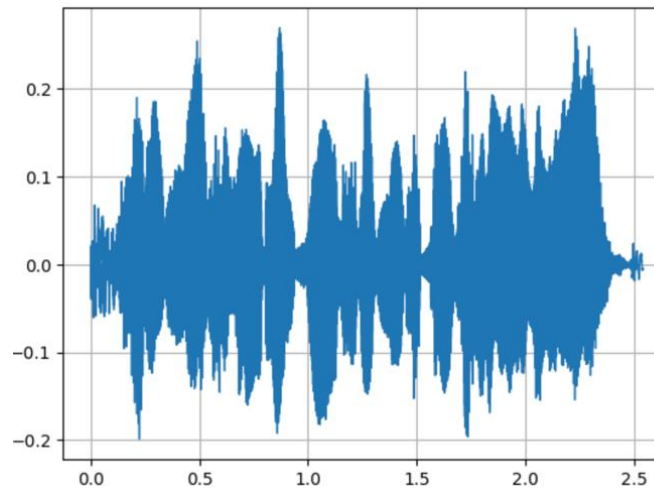


Датасет Ravdess



Представлення ознак аудіо

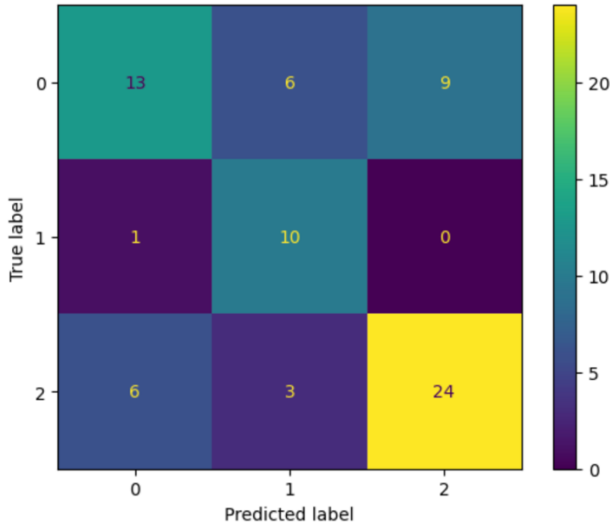
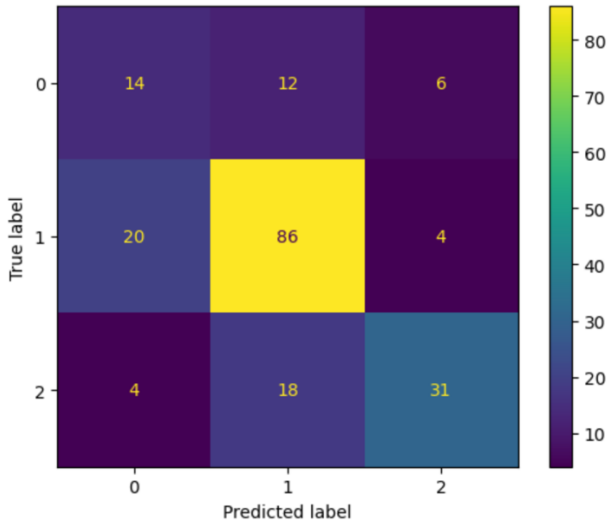
Представлення даних з аудіофайлу тривалістю 2.5 секунди та частотою дискретизації 48кГц у формі а) хвилі в частотному розрізі, б) мелспектрограми, в) MFCC, г) LFCC



Результати CNN

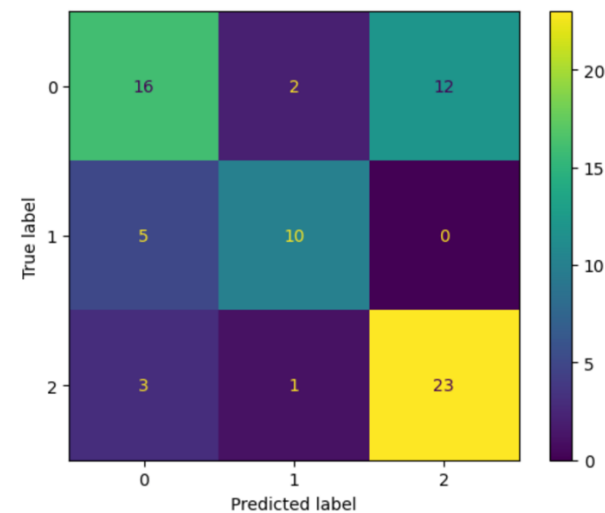
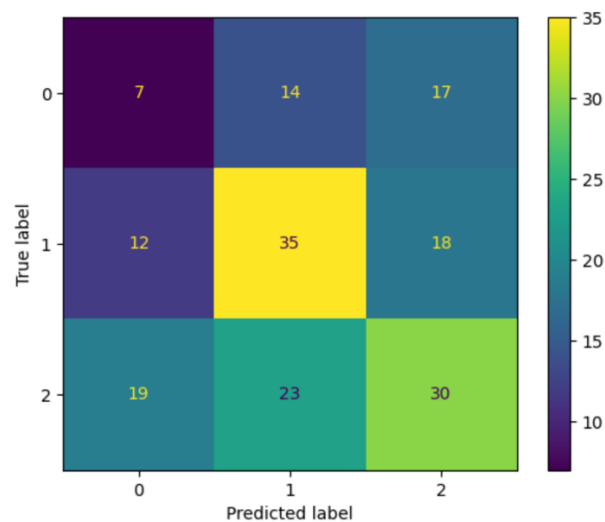
Dataset	Metric	LFCC	MFCC	MelSpectrogram
Call center + voice acting	Accuracy	0.5684523837906974	0.6026785714285714	0.5372023837906974
	F1Score	0.3869362899235317	0.4614541360310146	0.38173965045384
RAVDESS	Accuracy	0.5416666666666666	0.6770833333333334	0.59375
	F1Score	0.42312441269556683	0.5296610395113627	0.5126200318336487

Dataset	Metric	Base MFCC 64, Copying, LR = 0.001, Without normalization	LR - 0.0001	With normalization	MFCC 16	Padding
Call center + voice acting	Accuracy	0.6026785714285714	0.5758928571428571	0.6383928571428571	0.5595238123621259	0.5639880980764117
	F1Score	0.4614541360310146	0.4172184807913644	0.4280605741909572	0.4880585031849997	0.378029784985951
RAVDESS	Accuracy	0.6770833333333334	0.7083333333333334	0.7083333333333334	0.625	0.59375
	F1Score	0.5296610395113627	0.6770833333333334	0.6843018929163615	0.6191267172495524	0.46497655908266705



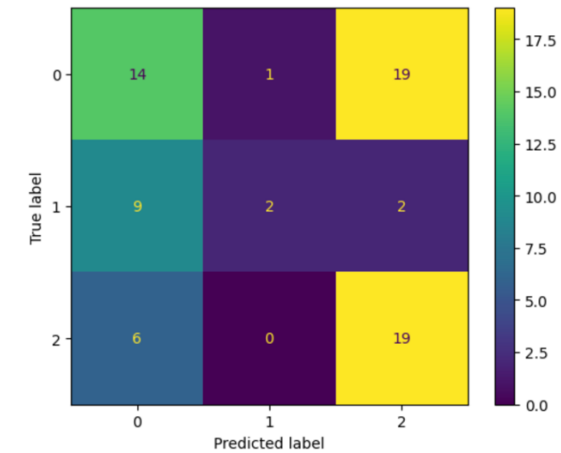
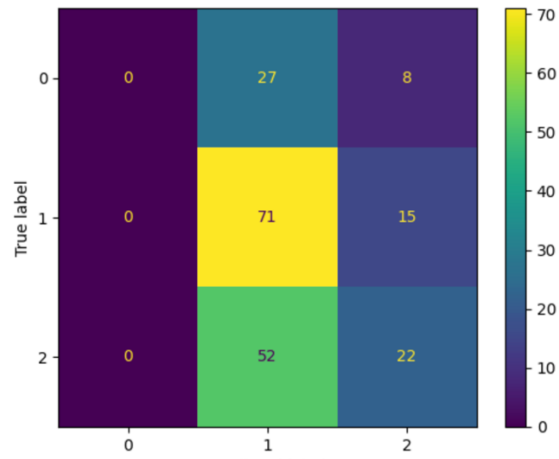
Результати RNN

Dataset	Metric	LFCC	MFCC	MelSpectrogram
Call center + voice acting	Accuracy	0.4439102571744185	0.5112179494821109	0.45993589896422166
	F1Score	0.35035193195709813	0.41896237432956696	0.39933934120031506
RAVDESS	Accuracy	0.5125	0.525	0.675
	F1Score	0.4365427404642105	0.4736339569091797	0.6241839826107025



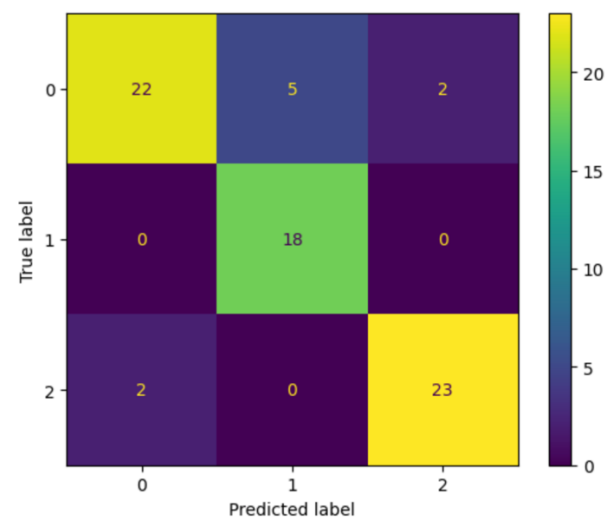
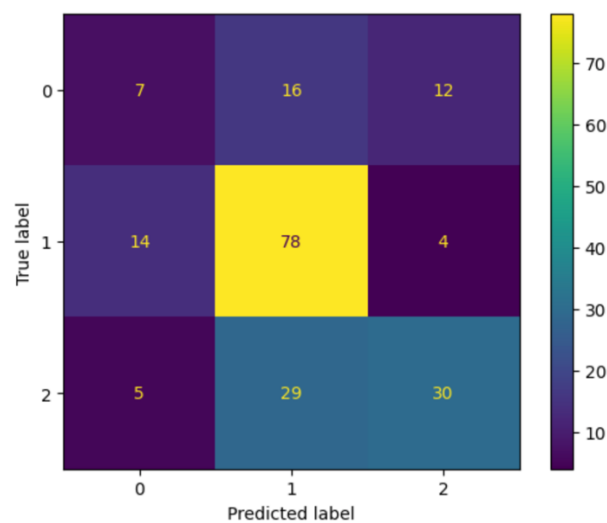
Результати Transfer Learning з моделлю ResNet

Dataset	Call center + voice acting dataset	RAVDESS
Accuracy	0.5669642857142857	0.5
F1Score	0.3368342880691801	0.37497853934764863



Transformers

Dataset	Call center + voice acting			RAVDESS
Train params	LR 0.005 + 10 epochs	LR 0.005 + 5 epochs	LR 0.0005 + 5 epochs	LR 0.005 + 5 epochs
Accuracy	0.5692307692307692	0.6153846153846154	0.5743589743589743	0.875
F1Score	0.5389010989010988	0.5386617398987871	0.5147650417215636	0.8760791532443627



Порівняльна таблиця всіх моделей розпізнавання емоцій

	CNN	RNN	ResNet	Transformer
Accuracy	0.6383928571428571	0.5112179494821109	0.5669642857142857	0.6153846153846154
F1Score	0.4280605741909572	0.41896237432956696	0.3368342880691801	0.5386617398987871

	CNN	RNN	ResNet	Transformer
Accuracy	0.7083333333333334	0.675	0.5	0.875
F1Score	0.6843018929163615	0.6241839826107025	0.37497853934764863	0.8760791532443627

Дякую за увагу!
Готовий відповідати на ваші запитання!