



COMPARATIVE SPEECH EMOTION RECOGNITION IN MULTIPLE MODELS AND MULTIPLE DATASETS

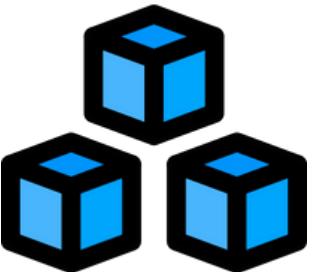
SUPASUN KHUMPRAPHAN

CE901-7-AU MSc Project and Dissertation

TABLE OF CONTENT

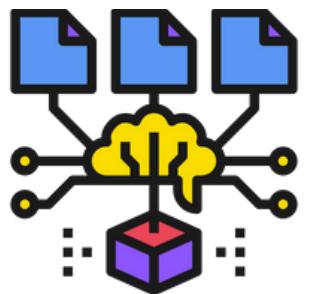
03	<u>Introduction</u>	08	<u>Machine Learning Step</u>	13	<u>Limitation</u>
04	<u>Advantage and Disadvages in every dataset</u>	09	<u>Deep learning step</u>	14	<u>Conclusion</u>
05	<u>Exploratory data analysis</u>	10	<u>Comparative default and hyper parameter</u>	15	<u>Link to what I do in this class each week.</u>
06	<u>Feature Extraction</u>	11	<u>Comparative result all model</u>		
07	<u>Sepeate Feature and label</u>	12	<u>Dicussion</u>		

INTRODUCTION



COMPARATIVE 17 MODELS

Machine learning 11 models and deep learning 6 models



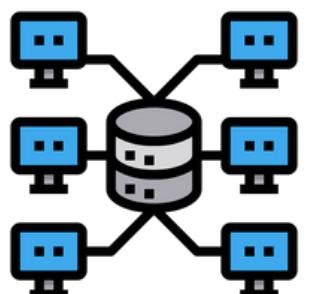
COMPARATIVE PREDICT

predict only emotion and predict emotion and gender



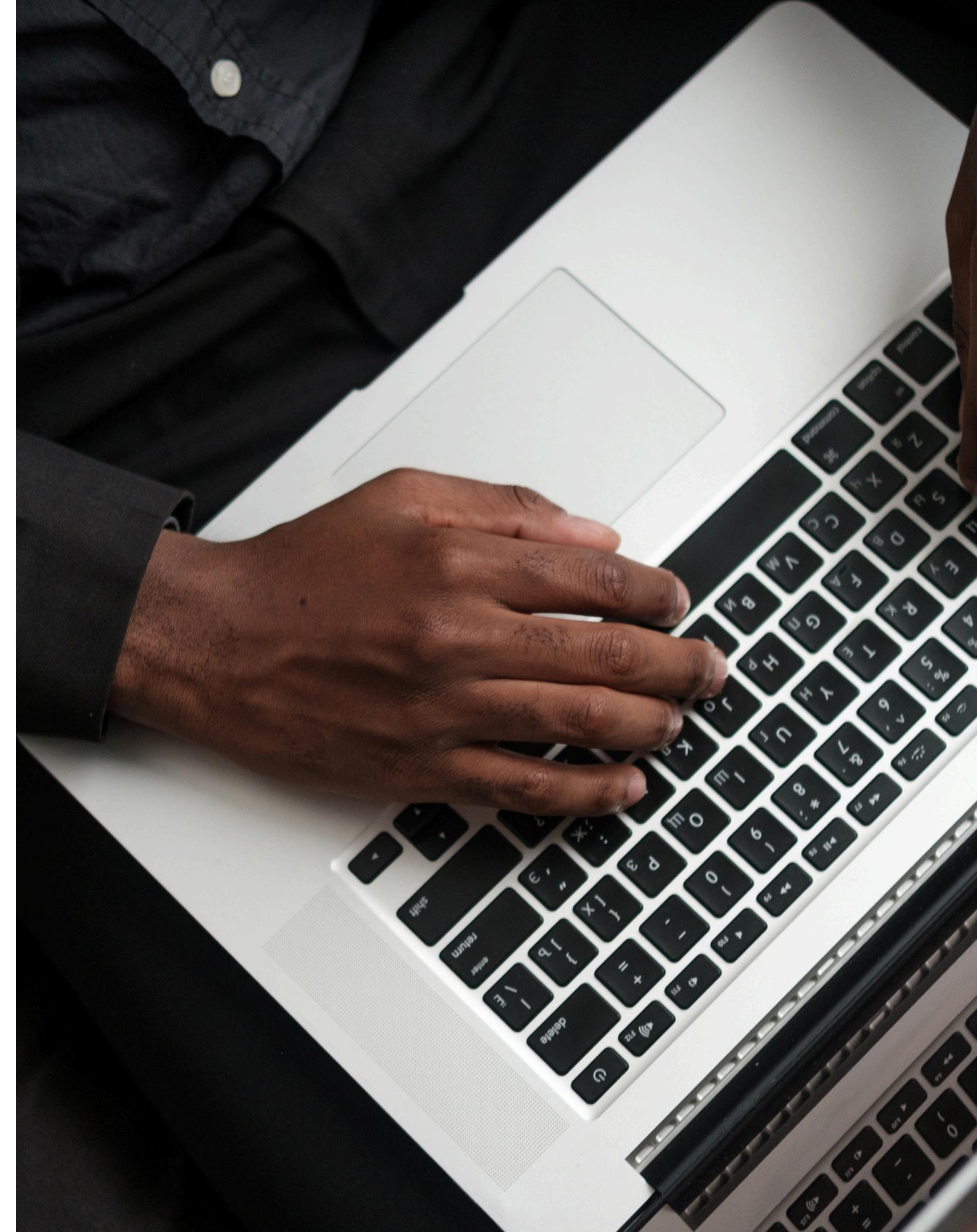
ENSEMBLE LEARNING

using Ensemble Learning to increase accuracy machine learning models



4 DATASET

using Ravdess dataset, Crema-D dataset, Tess dataset, Savee dataset



Advantage and Disadvantages in every dataset

RAVDESS DATASET:

Advantage: male and female voices have the same count in this dataset.

Disadvantage: There are only 1440 voices spoken.

CREMA-D DATASET:

Advantage: There are 7442 voices and there are voices from people of various accents, with 48 males and 43 females.

Disadvantage: There is some record that does not specify the gender and intensity of the user's voice.

SAVEE DATASET:

Advantage: this dataset is characterized by very high-quality sound.

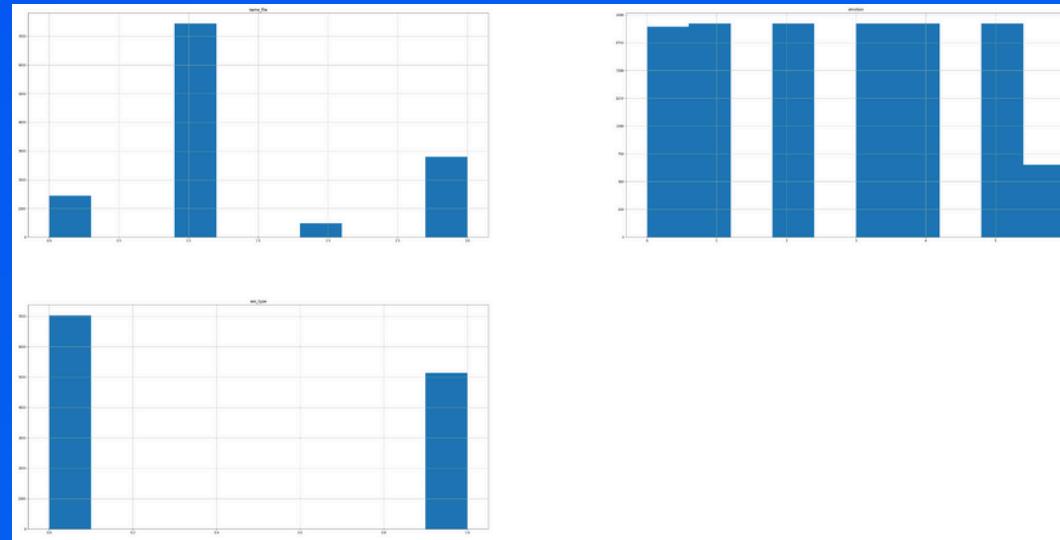
Disadvantage: It has only male voices and There are only 480 voices spoken and 4 speakers.

TESS DATASET:

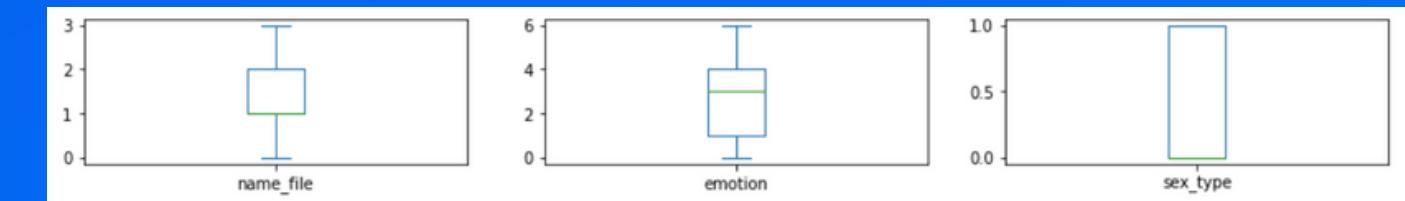
Advantage: recorded female voices as high-quality sound and There are 2800 voices from 14 speakers.

Disadvantage: recorded primarily male voices

EXPLORATORY DATA ANALYSIS

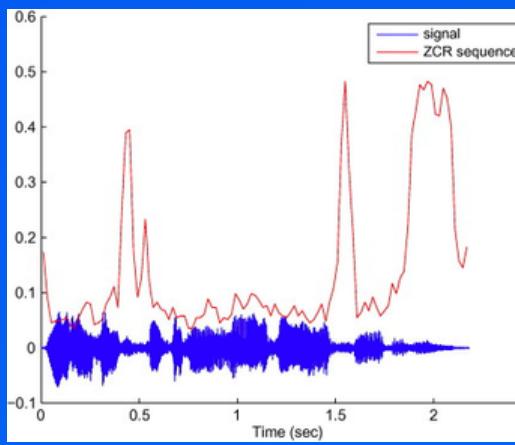


HISTOGRAM
PLOT

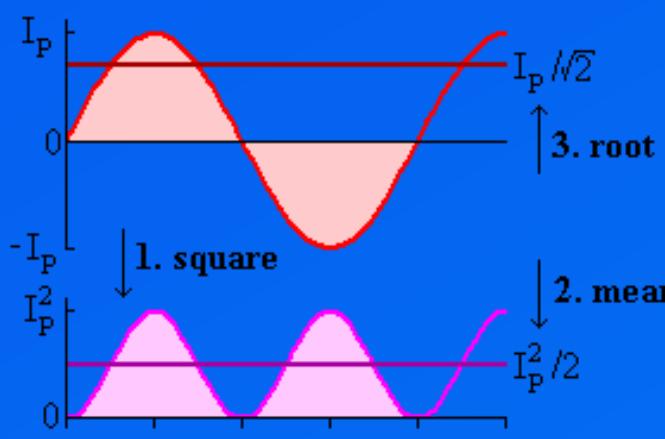


BOX PLOT

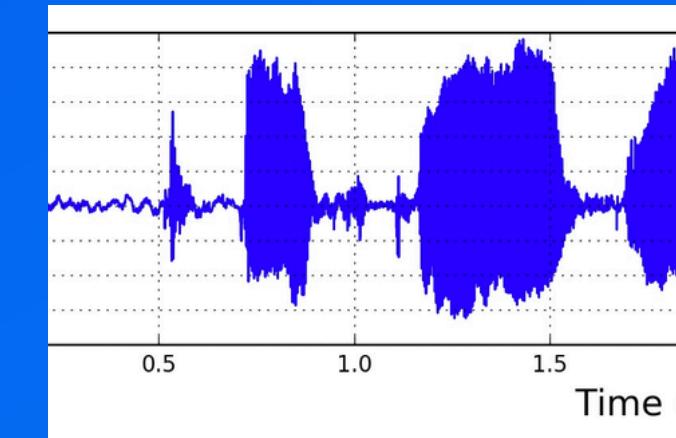
FEATURE EXTRACTION



ZERO CROSSING
RATE



ROOT MEAN
SQUARE



MEL FREQUENCY
CEPSTRAL
COEFFICIENT

SEPEATE FEATURE AND LABEL

ONLY EMOTION LABEL

1. we delete gender label
2. we will use every column without an emotion label to train model
3. we use emotion label to test model

GENDER AND EMOTION LABELS

1. we will use every column without gender and emotion labels to train model
2. we use gender and emotion labels to test model

Standard Scalar

MACHINE LEARNING STEP

DEFAULT PARAMETER	TUNING PARAMETER	HYPERPARAMETER	ENSEMBLE LEARNING
<ul style="list-style-type: none">• train model with default parameter• K-Fold Cross Validation• predict accuracy score• visualize for classification report and confusion matrix	<ul style="list-style-type: none">• using the default parameter and multiple parameter to find the best parameter with RandomizedSearchCV• predict the best score, best parameter, best estimator	<ul style="list-style-type: none">• train model with tuning parameters• K-Fold Cross Validation• predict accuracy score• visualize for classification report and confusion matrix	<ul style="list-style-type: none">• use the best model 5 parameter for the vote and train model• K-Fold Cross Validation• predict accuracy score• visualize for classification report and confusion matrix

DEEP LEARNING STEP



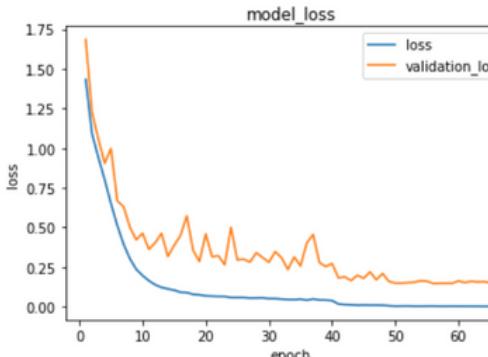
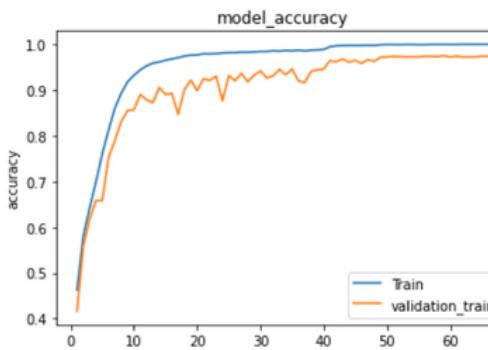
TRAIN MODEL

- Input layer with 2376 columns(feature)
- hidden layer (Each model has a different structure)
- output layer
- validation
- call back
- epoch
- batch size



CROSS ENTROPY

- Check whether this model is suitable for real use or not.



VIRTUALIZATION

- accuracy
- confusion matrix
- classification report

COMPARATIVE DEFAULT AND HYPER PARAMETER

ONLY EMOTION LABEL

algorithm	default parameter	hyperparameter	best accuracy
Decision Tree	49.26	47.18	49.26
Logistic Regression	43.36	42.58	43.36
Random Forest	77.10	80.82	80.82
K-Nearest Neighbors	61.21	76.49	76.49
support vector machine	61.02	61.02	61.02
Gaussian Naive Bayes	17.28	17.28	17.28
XGBoost	49.21	49.21	49.21
SGD	36.15	37.27	37.27
Multi-layer Perceptron	83.84	81.59	83.84
average model	53.17	54.83	55.40

GENDER AND EMOTION LABEL

algorithm	default parameter	hyperparameter	best accuracy
Decision Tree	44.54	38.17	44.54
Logistic Regression	40.11	40.71	40.71
Random Forest	78.10	85.41	85.41
K-Nearest Neighbors	53.54	74.53	74.53
support vector machine	53.48	53.47	53.48
Gaussian Naive Bayes	9.73	9.73	9.73
XGBoost	44.08	44.08	44.08
SGD	33.27	33.65	33.65
Multi-layer Perceptron	78.70	81.96	81.96
average model	48.39	51.30	52.01

COMPARATIVE RESULT

ALL MODEL

ONLY EMOTION LABEL

algorithm	angry	disgust	fear	happy	neutral	sad	surprise	accuracy
Decision Tree	55.17	46.66	45.25	45.94	50.37	52.74	47.91	49.26
Logistic Regression	54.16	43.84	34.30	38.93	37.87	49.43	49.43	43.36
Random Forest	88.39	80.36	69.98	77.54	84.42	84.00	81.43	80.82
K-Nearest Neighbors	78.85	67.99	71.56	73.42	77.82	86.31	85.41	76.49
support vector machine	73.24	60.47	43.93	58.42	59.86	68.40	67.04	61.02
Gaussian Naive Bayes	5.07	19.37	10.29	24.13	2.44	21.67	78.03	17.28
XGBoost	67.44	55.82	27.70	38.99	47.04	58.95	48.86	49.21
SGD	43.16	37.82	31.06	30.37	35.80	48.51	29.35	37.27
Multi-layer Perceptron	88.79	86.78	83.90	80.56	82.41	79.18	88.44	83.84
EL with hard vote	92.12	88.61	85.15	86.61	90.20	89.16	93.56	88.91
EL with soft vote	93.06	88.21	84.89	88.22	90.38	92.20	95.64	89.82
CNN	98.33	96.59	96.70	96.84	98.61	97.48	97.91	97.44
LSTM	95.73	93.45	93.73	94.33	96.35	96.43	95.64	94.99
gated recurrent unit	96.39	93.97	92.54	95.04	95.16	97.09	95.45	95.03
Deep neural network	84.18	78.46	71.70	78.50	83.73	80.37	85.98	79.30
artificial neuron network	74.38	67.27	67.67	69.17	72.36	71.38	79.54	69.85
CNN and LSTM	94.66	89.13	86.41	86.74	95.03	84.93	96.21	89.39
average model	75.47	70.28	64.51	68.45	70.57	74.01	77.40	70.78

GENDER AND EMOTION LABEL

algorithm	female	male	accuracy
Decision Tree	49.13	38.24	44.54
Logistic Regression	45.45	34.22	40.71
Random Forest	88.44	81.26	85.41
K-Nearest Neighbors	79.21	68.11	74.53
support vector machine	59.25	45.57	53.48
Gaussian Naive Bayes	10.61	8.52	9.73
XGBoost	52.37	32.71	44.08
SGD	44.32	19.04	33.65
Multi-layer Perceptron	84.69	78.22	81.96
EL with hard	90.24	80.77	86.24
EL with soft	90.89	85.77	88.64
CNN	98.32	96.17	97.35
LSTM	95.82	91.93	94.00
gated recurrent unit	94.45	89.69	92.34
Deep neural network	77.47	65.11	71.59
artificial neuron network	78.84	67.67	73.50
CNN and LSTM	95.96	91.35	93.87
average model	72.67	63.19	68.56

DISCUSSION



USE RANDOMIZEDSEARCHCV INSTEAD GRIDSEARCHCV

we can not use grid search because using time so much. it takes too long

MACHINE LEARNING MODEL

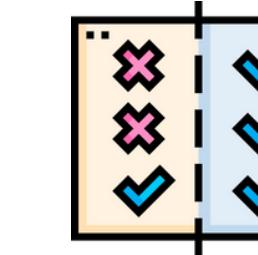
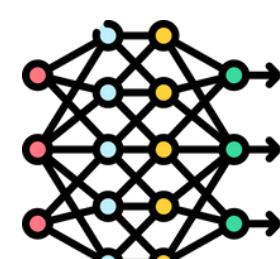
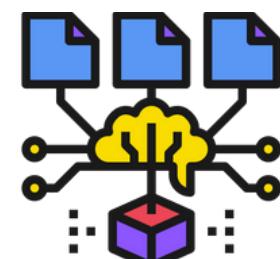
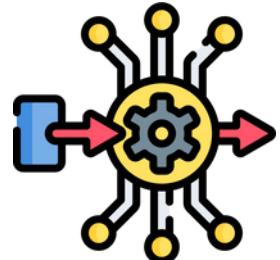
the best model in machine learning is Multi-layer Perceptron.

ENSEMBLE LEARNING MODEL

Ensemble Learning with soft vote can increase performance for machine learning model.

DEEP LEARNING MODEL

The best model in this research is the Convolutional neural network



COMPARATIVE PREDICTION ONLY EMOTION LABEL AND GENDER LABELS

predict only emotion label has accuracy more than emotion and gender labels
GENDER



female labels have good accuracy than male labels

SOUND

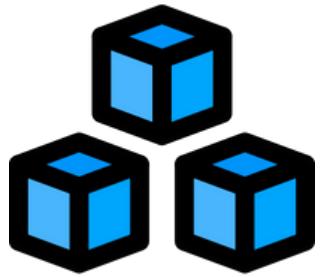
surprise sound is the most accurate, while the fear sound is the least accurate



SOME RESEARCH ACCURACY GOOD THAN THIS RESEARCH

2 research have 98 percent accuracy

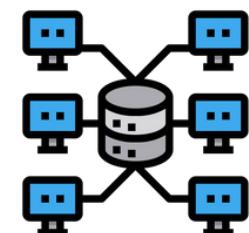
LIMITATION



HAS A MORE POWERFUL GPU

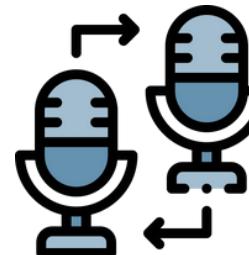
- improve accuracy in machine learning because it can use GridSearchCV
- improve accuracy in Deep Learning because it can try out deep learning models much faster for training models many times.

CONCLUSION



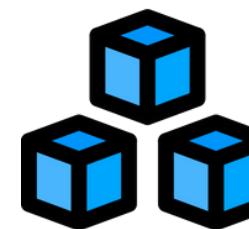
4 DATASET

Ravdess dataset, Crema dataset, Savee dataset, and TESS dataset



FEATURE EXTRACTION

Zero Crossing Rate, Root Mean Square, and the Mel Frequency Cepstral Coefficient



TRAIN MODEL WITH 17 MODELS

The best model in this research is the Convolutional neural network



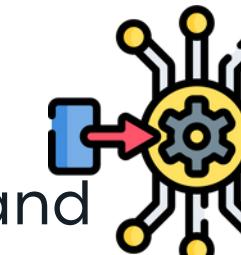
EMOTION

happiness, disgust, fear, happy, neutral, sad, and surprise. The least accurate predictive emotion label is fear emotion but the best emotion to predict is a surprise.



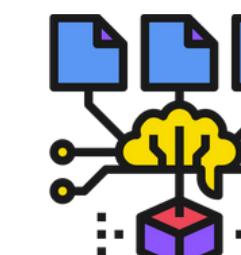
GENDER VOICE

Female voices are more accurate than male voices.



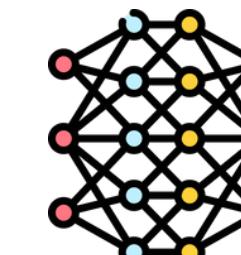
MACHINE LEARNING MODEL

The best model in machine learning is Multi-layer Perceptron
only emotion label: 83.84 percent
gender and emotion label: 81.96 percent



ENSEMBLE LEARNING MODEL

The best model in Ensemble Learning is Ensemble Learning with soft vote
only emotion label: 89.82 percent
gender and emotion label: 88.64 percent



DEEP LEARNING MODEL

The best model in Deep Learning is the Convolutional neural network
only emotion label: 97.44 percent
gender and emotion label: 97.35 percent

LINK TO WHAT I DO IN THIS CLASS EACH WEEK.

- <https://gitlab.com/micsupasun/CE901/-/tree/main>
- https://github.com/micsupasun/university_of_essex/tree/main/MSc_project_and_dissertation

**THANK YOU FOR WATCHING
MY PRESENTATION!**



I HOPE YOU LIKED IT :D