INTRODUCTION TO DEEP LEARNING

EMOTION DETECTION FROM URDU SPEECH

Presented By :

- Ikhlas Ahmed
- Shayaan Qazi
- Sameer Kamani
- Hunain Abbas

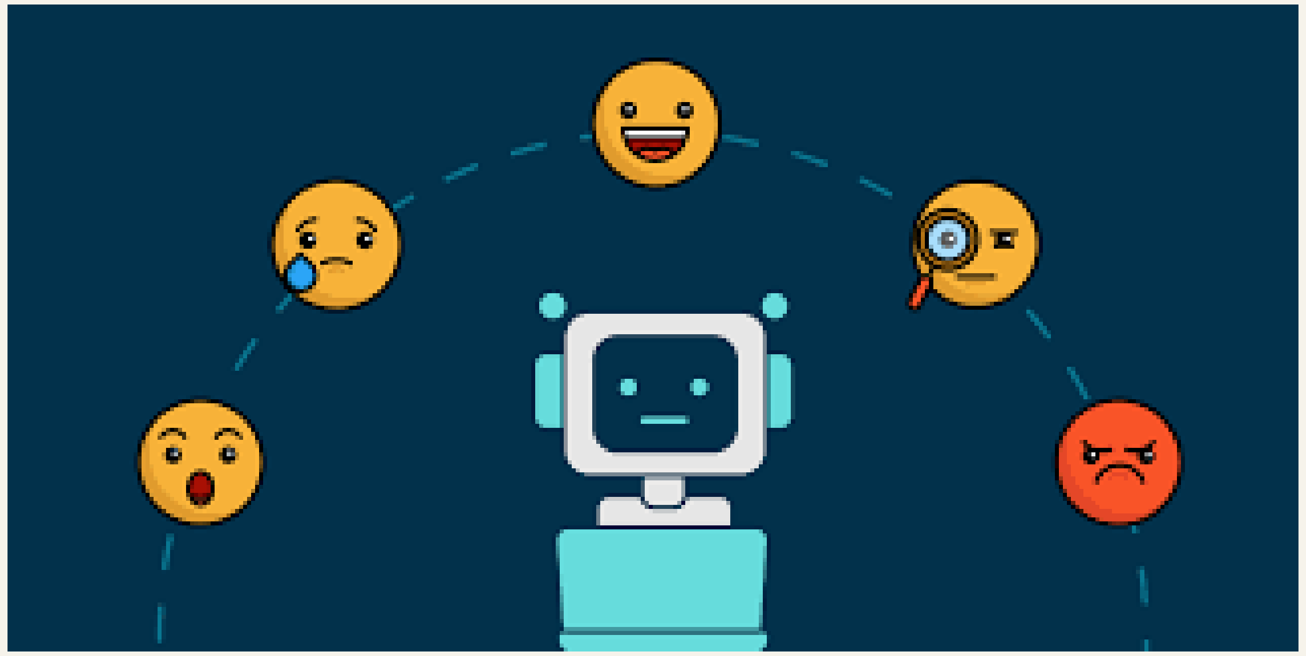
OVERVIEW

- Motivation
- Problem Statement
- Data
- Demographic
- Methodology
- Models
- Results
- Overview of results
- Comparing our work with previous work
- Future work
- References
- Conclusion

MOTIVATION

Empathetic AI-chatbots

Urdu recognizing AI Lagging Behind for 250M+ Speakers



Ethical Dilemma of AI Chatbots

Pakistan Population (LIVE)

PROBLEM STATEMENT

- Emotions are conveyed through speech using tone, pitch, and rhythm.
- While emotion detection has progressed in other languages, research on Urdu remains limited.
- This project aims to train a deep learning model that can detect emotions in Urdu speech.

Pushing Away Negative Emotions

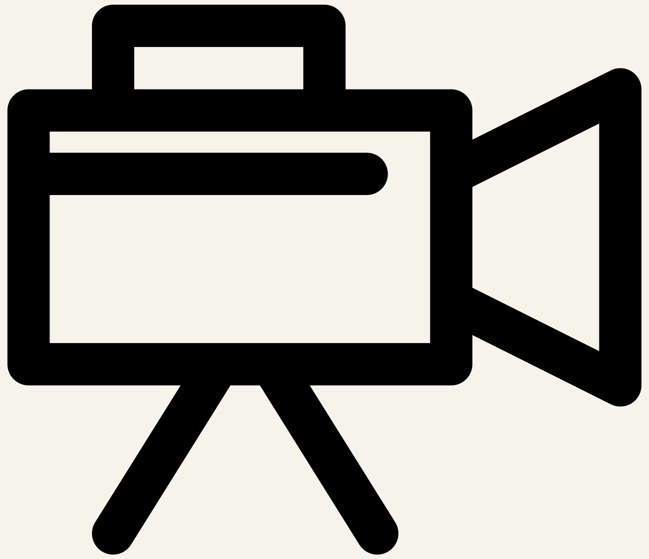
DATA

Total actors

24

Total Recordings

14,000+



Source

SEMOUR+

~10,000 training

~1,000 Validation

~3,000 testing

DEMOGRAPHIC

Male Actors

17 Actors

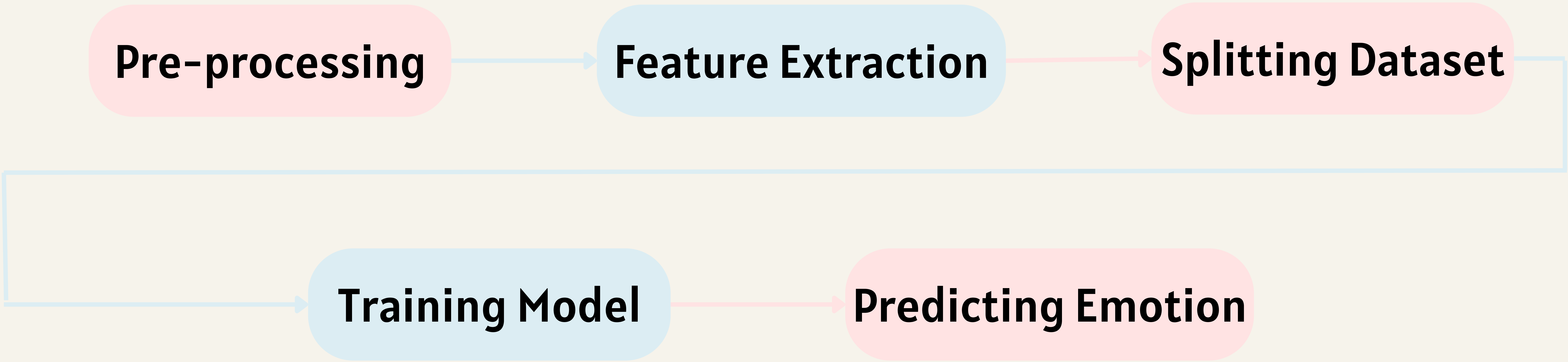
Age 20-40

Female Actors

7 Actors

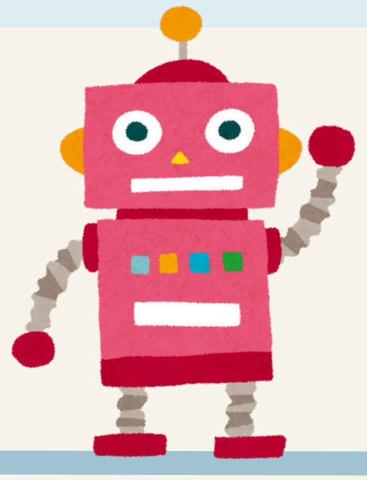
Age 20-40

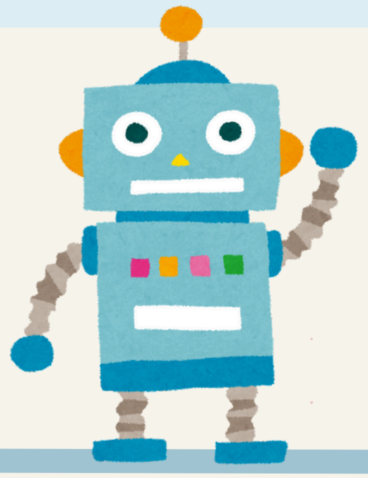
METHODOLOGY



OUR MODELS

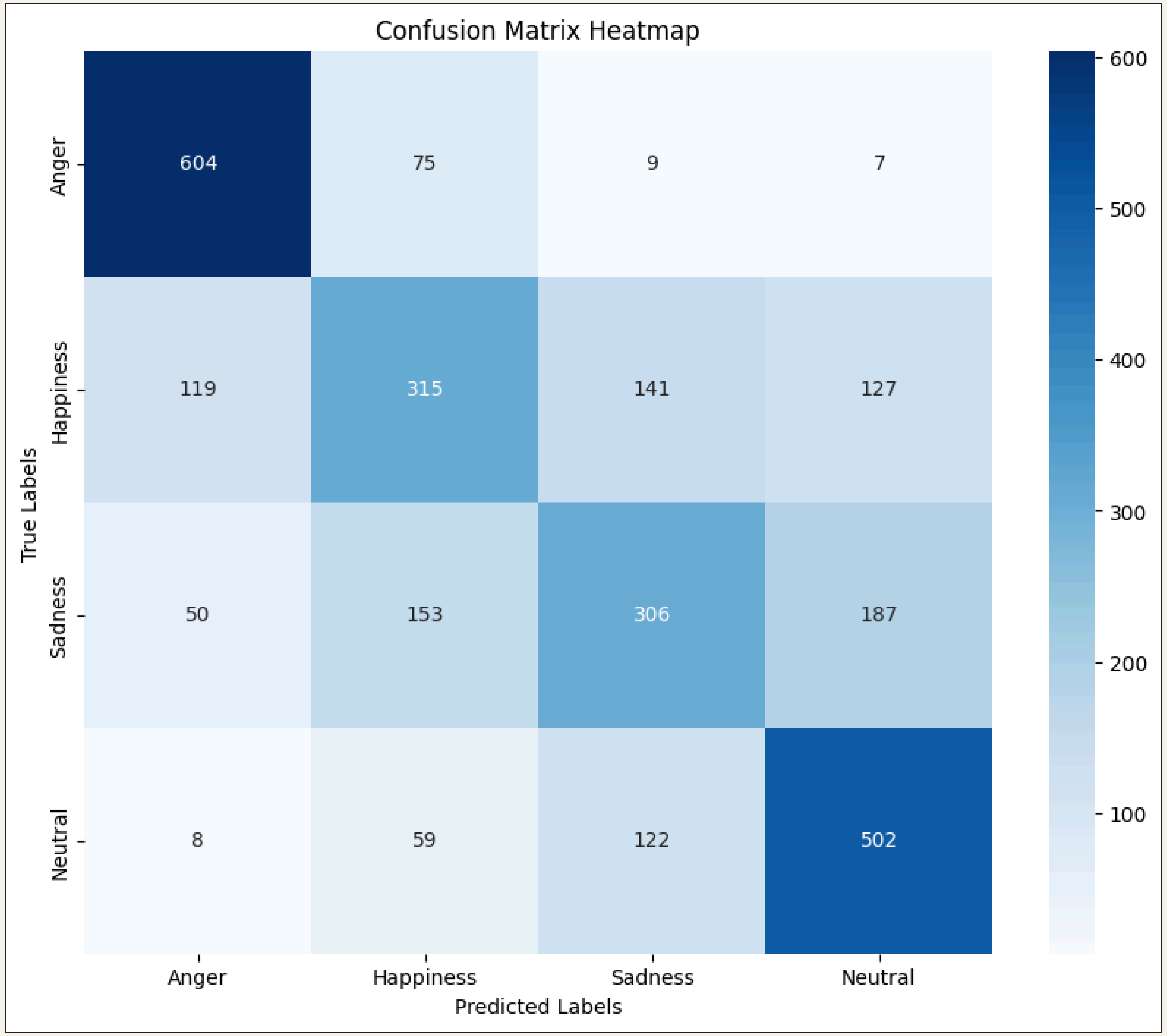
- SVM
- CNN
- RESNET50
- HUBERT
- WAV2VEC 2.0
- KNN





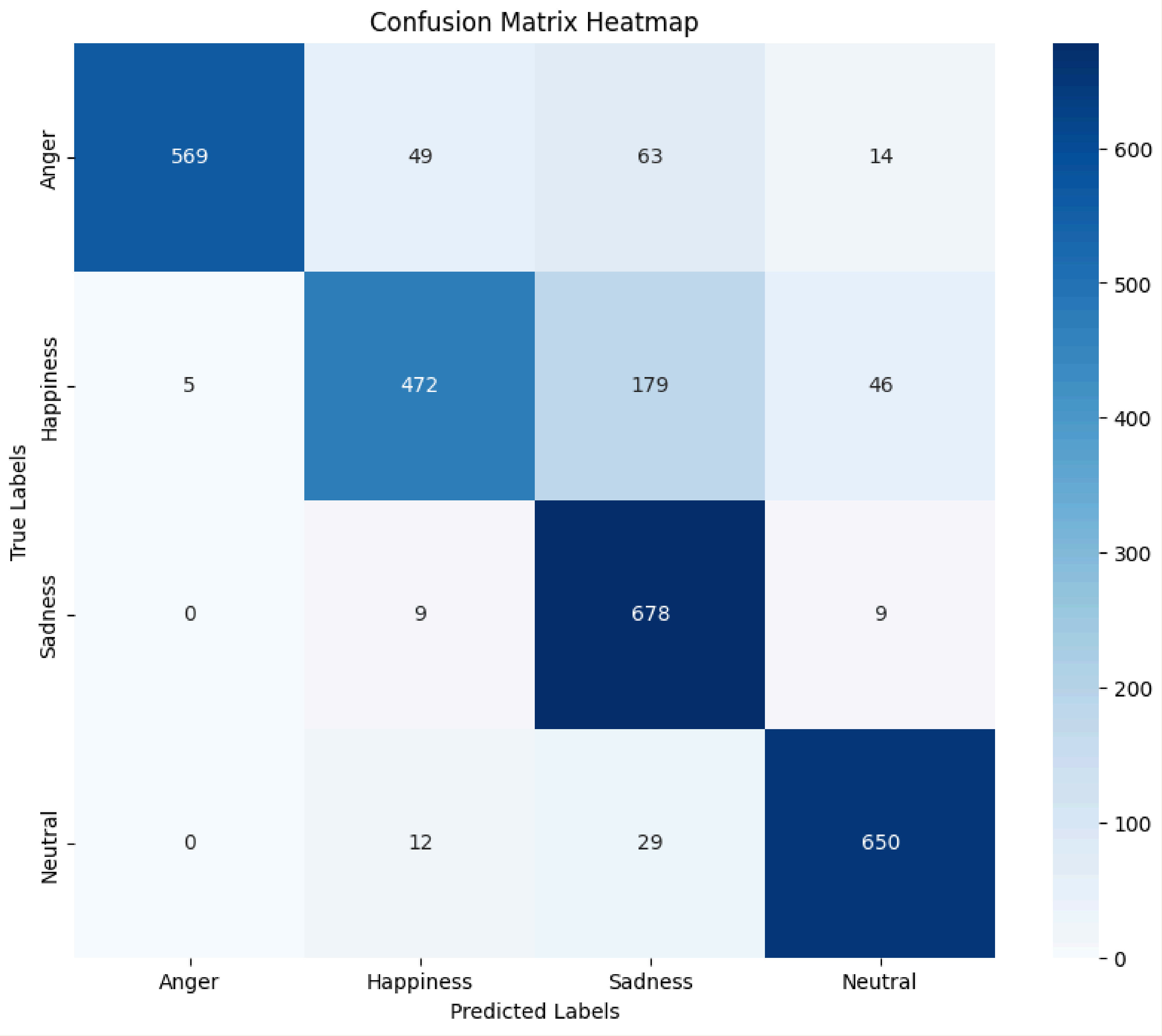
SVM RESULTS - 62%

Aspect	Details
Kernel	Linear
C (Regularization)	0.1
Feature Type	MFCC



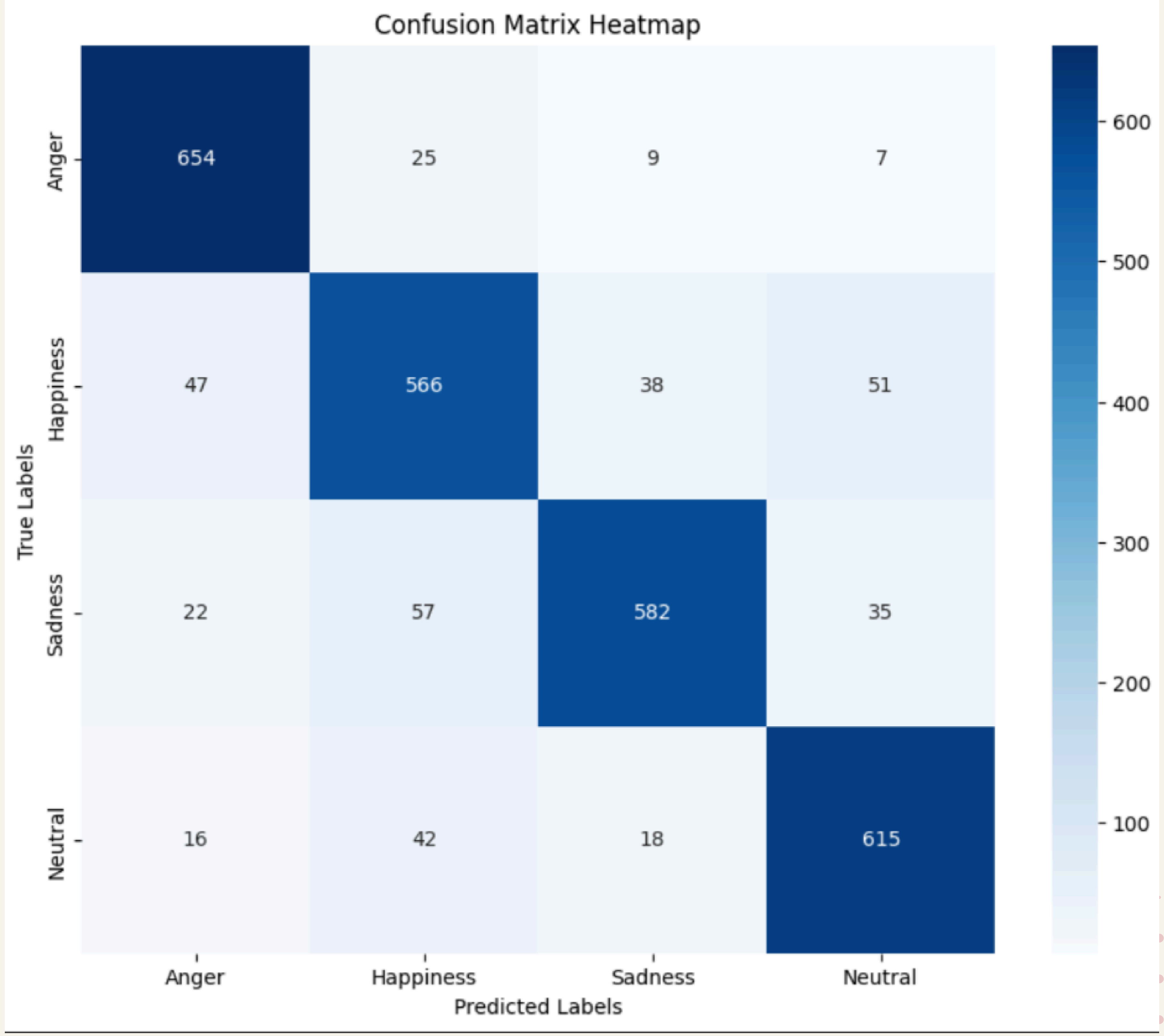
CNN RESULTS - 85.09%

Aspect	Details
Number of Layers	9 (3 Conv + 3 Pool + 2 FC + 1 Output)
Epochs	20
Batch Size	32
Optimizer	Adam
Learning Rate	0.001
Loss function	Sparse Categorical Crossentropy



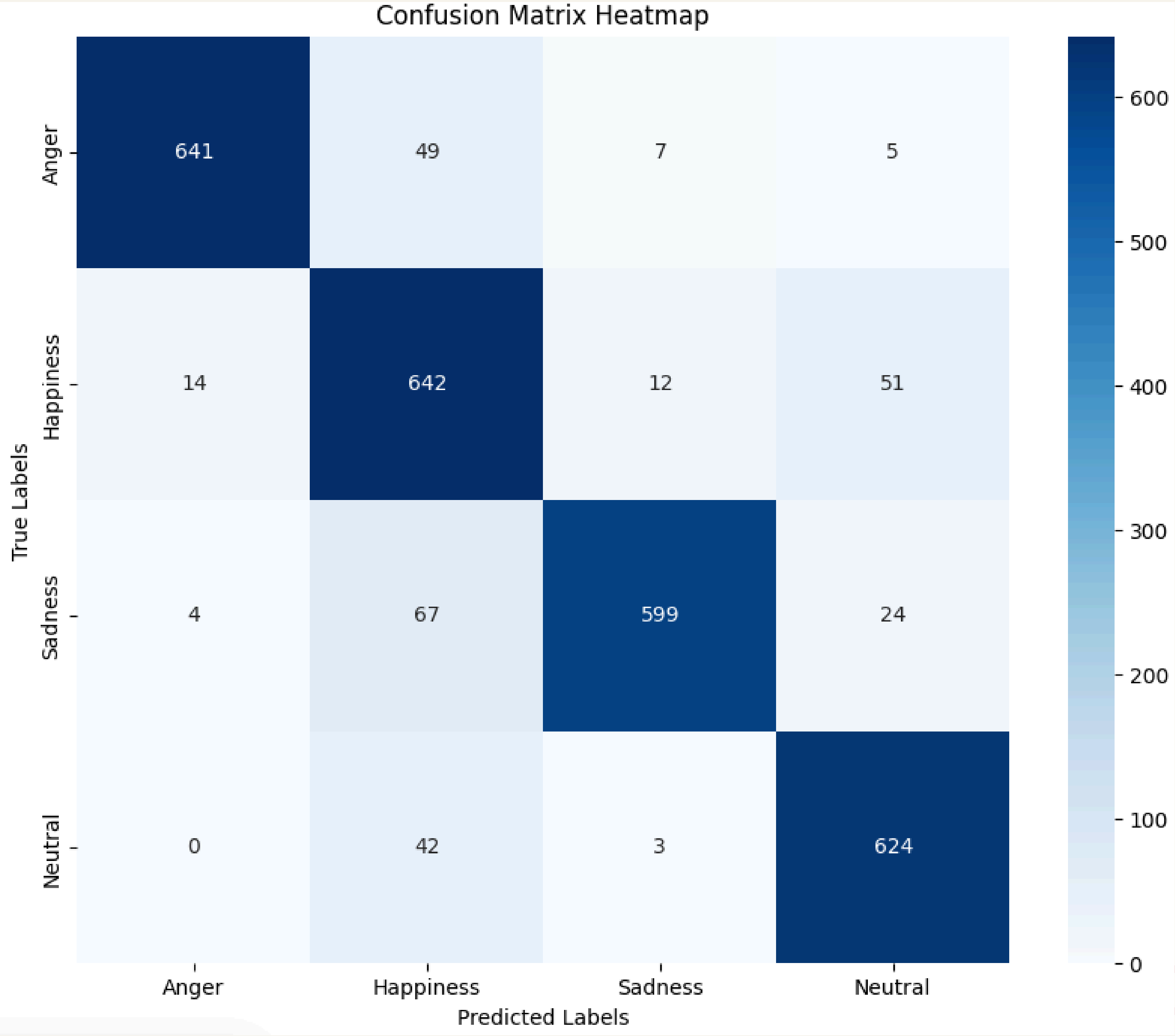
RESNET50 RESULTS - 86.82%

Aspect	Details
Epochs	30
Batch Size	32
Optimizer	Adam
Learning Rate	0.001
Loss function	Sparse categorical cross entropy



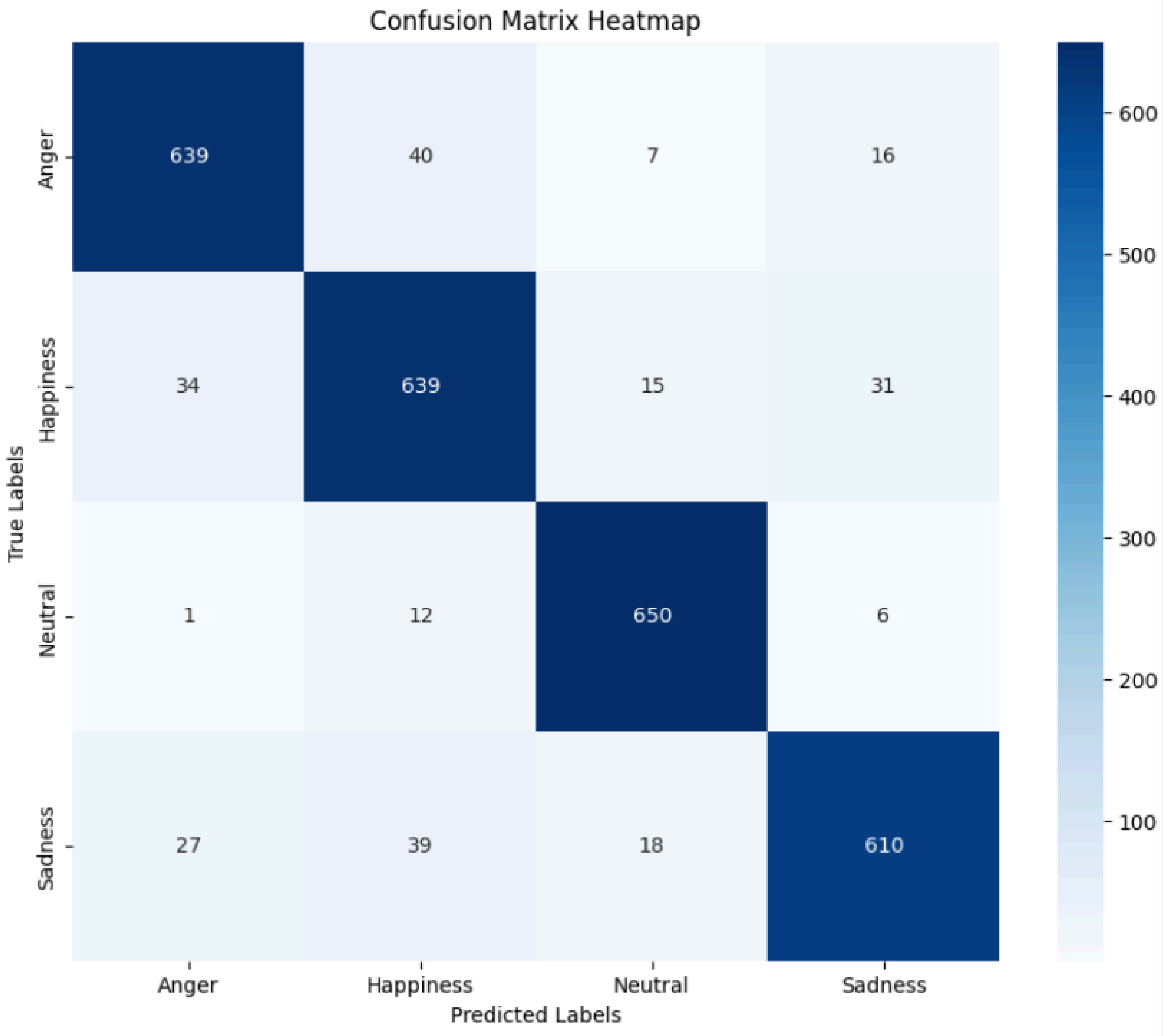
HUBERT RESULTS - 90%

Aspect	Details
Model Type	hubert-large-ls960-ft
Epochs	10
Batch Size	16
Optimizer	AdamW
Learning Rate	3e-5, Cosine Scheduler
Loss function	Sparse Categorical Cross-Entropy



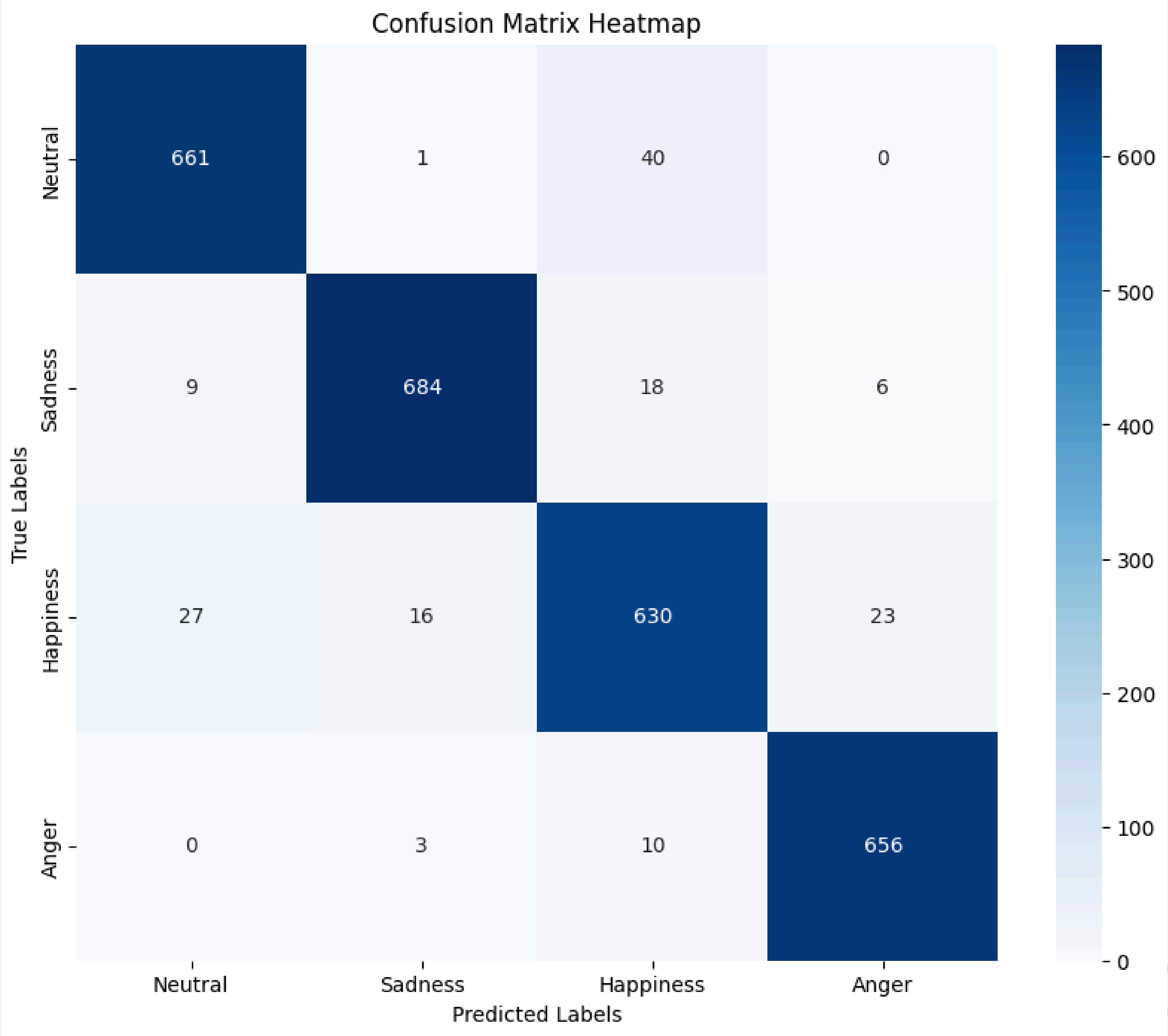
KNN RESULTS – 91.16%

Aspect	Details
Number of Neighbors	5
Learning Curve Metric	Accuracy



WAV2VEC 2.0 RESULTS - 94.50%

Aspect	Details
Model Type	wav2vec2-xls-r-300m
Epochs	20
Batch Size	32
Optimizer	AdamW
Learning Rate	0.00003, Cosine Scheduler
Loss function	Sparse Categorical Cross entropy



SUMMARY

Model	Result
SVM	62%
CNN	85.09%
Resnet50	86.82%
Hubert	90%
KNN	91.16%
Wav2vec2.0	94.5%

COMPARING WITH PAST WORK

- A similar paper [5] tested on the same 4 emotions, but on a smaller dataset.

Their best

KNN -82.5%

Our best

WAV2VEC - 94.5%

Table 8 Comparison with related work.

Papers	Languages	Training technique	Features extraction techniques	Emotions	Classifier used	Accuracy
Tripathi & Beigi (2018)	English and German	Speaker dependent	RNN	Anger, happiness, neutral and sadness	RNN with three layers	71.04%
Kaminska, Sapinski & Anbarjafari (2017)	Polish	Speaker dependent independent	MFCC, BFCC, RASTA, energy, formants, LPC and HFCC	Sadness, happiness, anger, neutral, joy, fear and surprise	SVM and k-NN	81%
Rajisha, Sunija & Riyas (2016)	Malayalam	Speaker dependent	MFCC, STE and pitch	Neutral, anger, happiness and sad	ANN and SVM	78%
Ali et al. (2013)	Urdu	Speaker dependent	Duration, intensity, pitch and formants	Anger, sadness, happiness and comfort	Neive Bayes	76%
Abbas, Zehra & Arif (2013)	Urdu	Speaker dependent	Intensity, pitch and formants	Anger, sadness, happiness and comfort	SMO, MLP, J48 and Neive Bayes	75%
Latif et al. (2018)	Urdu	Speaker independent	LLDs low level descriptor	Happiness, sadness, anger and neutral	SVM, logistic regression and RF	83%
Sinith et al. (2015)	English Malayalam and	Speaker dependent	MFCC, pitch and energy	Anger, neutral sadness and happiness	SVM	70%
Our work	Urdu (with disgust emotion)	Speaker dependent	MFCC, LPC, energy, pitch, zero crossing, spectral flux spectral centroid, spectral roll off	Anger, disgust, happiness, sadness and neutral	k-Nearest Neighbours	73%
Our work	Urdu (without disgust emotion)	Speaker dependent	MFCC, LPC, energy, pitch, zero crossing, spectral flux spectral centroid, spectral roll off	Anger, happiness, sadness and neutral	k-Nearest Neighbors	82 .5%

(Asghar, Sohaib, Iftikhar, Shafi, & Fatima, 2022)

WHAT NEXT?

- Experiment with other feature extraction methods for better accuracy
- Use other Augmentation technique for better accuracy
- Classify Models on more emotions or on other languages

REFERENCES

- [1] S. Latif, "Cross Lingual Speech Emotion Recognition: Urdu vs. Western Languages," 2018. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8616972.
- [2] B. B. Al-onazi, "Transformer-Based Multilingual Speech Emotion Recognition Using Data Augmentation and Feature Fusion," Applied Sciences, vol. 12, no. 18, 2022. [Online]. Available: https://www.mdpi.com/2076-3417/12/18/9188.
- [3] R. Shaik, "Sentiment Analysis with Word-Based Urdu Speech Recognition," Journal of Ambient Intelligence and Humanized Computing, 2021. [Online]. Available: https://link.springer.com/article/10.1007/s12652-021-03460-x.
- [4] Sehar, U., Kanwal, S., Dashtipur, K., Mir, U., Abbasi, U., & Khan, F. (2021). Urdu sentiment analysis via multimodal data mining based on deep learning algorithms. IEEE Access, 9, 153072-153086.
- [5] Asghar, A., Sohaib, S., Iftikhar, S., Shafi, M., & Fatima, K. (2022). An Urdu speech corpus for emotion recognition. PeerJ Computer Science, 8, e954.
- [6] Ullah, A., Khan, K. U., Khan, A., Bakhsh, S. T., Rahman, A. U., Akbar, S., & Saqia, B. (2024). Threatening language detection from Urdu data with deep sequential model. PLOS ONE, 19(6), e0290915.
- [7] Mateen, M., & Bawany, N. Z. (2023). Deep Learning Approach for Detecting Audio Deepfakes in Urdu. NUML International Journal of Engineering and Computing, 2(1).
- N. Al-Sibai, "Lonely Teens Are Making Friends With AI," Futurism, 2023. [Online]. Available: https://futurism.com/the-byte/lonely-teens-friends-with-ai.
- D. Robson, "Can AI Chatbot Therapists Do Better Than the Real Thing?," The Guardian, 2024. [Online]. Available: https://www.theguardian.com/lifeandstyle/2024/mar/02/can-ai-chatbot-therapists-do-better-than-the-real-thing.

Habib University | 2024

THANK YOU