



FINAL YEAR PROJECT REPORT

EMOTION DETECTION USING AUDIO AND VIDEO FILES

**In fulfillment of the requirement
For degree of
BS (COMPUTER SCIENCES)**

By

REHMAT ALI SHAH BUKHARI	67708 (BSCS)
UMER BIN AMIR SULTANI	67759 (BSCS)
SHER ALI AHMAR	68089 (BSCS)

SUPERVISED

BY

MISS AMNA IFTIKHAR

BAHRIA UNIVERSITY (KARACHI CAMPUS)

FALL-2023

DECLARATION

We hereby declare that this project report is based on our original work except for citations and quotations which have been duly acknowledged. We also declare that it has not been previously and concurrently submitted for any other degree or award at Bahria University or other institutions.

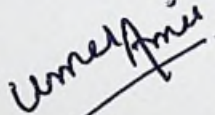
Name : Rehmat Ali Shah Bukhari

Reg No. : 02-134201-013

Signature :  _____

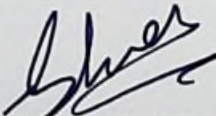
Name : Umer Bin Amir Sultani

Reg No. : 02-134201-073

Signature :  _____

Name : SherAli Ahmar

Reg No. : 02-134201-099

Signature :  _____

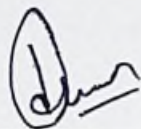
Date : _____

APPROVAL FOR SUBMISSION

We certify that this project report entitled “**EMOTION DETECTION USING AUDIO AND VIDEO FILES**” was prepared by **Rehmat Ali Shah Bukhari, Umer Bin Amir Sultani, SherAli Ahmar** has met the required standard for submission in partial fulfilment of the requirements for the award of Bachelor of Computer Science at Bahria University.

Approved by,

Signature : _____



Supervisor : Miss Amna Iftikhar

Date

:

24/01/24

According to the Intellectual Property Policy of Bahria University BUORIC-P15 amended on April 2023, Bahria University's copyright of this report is maintained. Any material used or derived from this report must be acknowledged always.

© 2023 Bahria University. All right reserved.

Specially dedicated to
My beloved grandmother, mother and father
Rehmat Ali Shah Bukhari
My beloved grandmother, mother and father
Umer Bin Amir Sultani
My beloved grandmother, mother and father
Sher Ali Ahmar

ACKNOWLEDGEMENT

We want to thank all those who have assisted in successfully completing this project. My research supervisor, Miss Amna Iftikhar, whom I would like to extend my thanks for her priceless advice, direction and being so patient with me during the period of the study.

Additionally, our thanks go to our caring parents and friends for their supportive assistance and words of encouragement.

ABSTRACT

Human emotions play a critical role in interpersonal communication, influencing our decisions, actions, and reactions. The ability to accurately detect emotions can significantly enhance human-computer interaction, providing more empathetic and responsive systems. However, traditional methods of emotion detection, relying on facial expressions or textual data, often miss the nuanced cues embedded in the tonal variations of human speech. Consequently, there is a pressing need to develop efficient and accurate models capable of detecting emotions from speech, which can be particularly challenging given the subtlety and complexity of acoustic patterns associated with different emotions. This project seeks to bridge the gap between human emotions and machine understanding by devising a solution that can accurately detect emotions embedded in human speech. Recognizing the importance of catering to the emotional needs and states of users, our research has been geared towards ensuring that machines can not only comprehend but also respond aptly to the emotional cues present in speech. By successfully addressing this challenge, we aim to revolutionize human-machine interactions, making them more intuitive, empathetic, and responsive. Such advancements have profound implications, especially in sectors like healthcare, where understanding a patient's emotional state can influence treatment decisions, or in entertainment, where user experience can be enhanced manifold by tailoring content based on detected emotions. Furthermore, in the realm of personalized assistance, machines that can understand and respond to user emotions can provide a more tailored and enriching experience, paving the way for a future where our digital interactions are as nuanced and fulfilling as our human ones.

TABLE OF CONTENTS

DECLARATION	ii
APPROVAL FOR SUBMISSION	iii
ACKNOWLEDGEMENTS	vi
ABSTRACT	vii
TABLE OF CONTENTS	viii
LIST OF TABLES	xii
LIST OF FIGURES	xiii

CHAPTER

1	INTRODUCTION	1
	1.1 Background	1
	1.2 Problem Statements	2
	1.2.1 Variability in Emotional Expression	2
	1.2.2 Acoustic Complexity	2
	1.2.3 Data Availability and Quality	2
	1.2.4 Multimodal Challenges	2
	1.2.5 Real-time Processing	2
	1.3 Aims and Objectives	3
	1.4 Scope of Project	4
	1.4.1 Comprehensive Analysis	4
	1.4.2 Instant Feedback Mechanism	4
	1.4.3 Flexibility and Future-Readiness	4
	1.4.4 User-Centric Design	4
	1.4.5 Detailed Emotional Insight	4
	1.4.6 Advancement in AI Capabilities	5
	1.5 Future Scope	5
2	LITERATURE REVIEW	6
	2.1 To Design And Develop Advance Speech Emotion Recognition Using MLP Classifier With Evolutionary Librosa Library	6
	2.2 Speech Emotion Recognition using RNNs	7
	2.3 Emotional speech Recognition using CNN and Deep learning techniques	8

2.4	EEG-based emotion classification based on Bidirectional Long Short-Term Memory Network	8
2.5	Speech emotion recognition system using gender dependent CNN	10
2.6	Human Emotions Detection and Classification Model from Speech using CNN	12
2.7	Hybrid Time Distributed CNN-transformer for Speech Emotion Recognition	12
2.8	Hybrid LSTM-Transformer Model for Emotion Recognition from Speech Audio Files	13
2.9	CCNN Architecture for Speech Emotion Recognition in Noisy Conditions	14
2.10	A Compact Deep Learning Model for Robust Facial Expression Recognition	15
2.11	Real Time Facial Expression Recognition using Deep Learning	15
2.12	A 3D-convolutional neural network framework with ensemble learning techniques for multi-modal emotion recognition	16
3	REQUIREMENT SPECIFICATIONS	18
3.1	System Enhancement Overview	18
3.1.1	Existing System	18
3.1.2	Proposed System	18
3.2	Requirement Specifications	18
3.2.1	Emotion Analysis	18
3.2.2	Non-Functional Requirements	19
3.2.2.1	Performance	19
3.2.2.2	Security	19
3.2.2.3	Scalability	19
3.2.2.4	User Interface	19
3.2.2.5	Sustainability	19
4	DESIGN AND METHODOLOGY	20
4.1	Methodology	20
4.2	Algorithms	21
4.3	Dataset Used	24

4.3.1	RAVDESS Contributions	24
4.3.2	TESS Contributions	24
4.3.2.1	Modality	25
4.3.2.2	Vocal Channel	25
4.3.2.3	Emotion	25
4.3.2.4	Emotional Intensity	25
4.3.2.5	Statement	25
4.3.2.6	Repetition	25
4.3.2.7	Actor	25
4.4	Architecture of System	25
4.4.1	Dataset Input	25
4.4.2	Feature Extraction	25
4.4.3	Feature Storage	26
4.4.4	Model Training	26
4.4.5	Testing with Flask	26
4.5	Model Training and Evaluation	27
4.5.1	Performance Metrics	27
4.5.2	Balanced F1 Scores	27
4.5.3	Challenging Classes	27
4.5.4	Minor Performance Drop	27
4.5.5	Model Reliability	28
4.6	Modules Discussion	29
4.6.1	Dataset Input	29
4.6.2	Feature Extraction	29
4.6.3	Feature Storage	30
4.6.4	Model Training	30
4.6.5	Testing with Flask	30
5	SYSTEM IMPLEMENTATION	31
5.1	UI Initialization	31
5.2	Options to Record	31
5.3	Recording and Analysis	32
5.4	Flask Integration	34
6	SYSTEM TESTING AND EVALUATION	35
6.1	CNN-LSTM EVALUTION	35

6.2	XCeption EVALUTION	37
7	CONCLUSION	42
7.1	Future Work	42
7.1.1	Incorporation of Additional Modalities	42
7.1.2	Expanding the Dataset	42
7.1.3	Refinement of Models	42
7.1.4	Real-world Applications	42
7.1.5	Addressing Ethical Concerns	43
	REFERENCES	44