“Real Time Mental Health Analysis- Well Mind”

Synopsis-I

Submitted in the partial fulfilment of the requirement

for the award of a degree of

**BACHELORS OF SCIENCE**

**IN**

**COMPUTATIONAL STATISTICS & DATA ANALYTICS**

**(2022-2025)**

Submitted to

**DEPARTMENT OF COMPUTATIONAL STATISTICS AND DATA ANALYTICS**

**GURU NANAK DEV UNIVERSITY**

**AMRITSAR**



|  |  |  |
| --- | --- | --- |
| **SUBMITTED TO:** |  | **SUBMITTED BY:** |
| Dr. Sandeep Sharma (HOD) |  | Dipender Singh |
| Dr. Prabhsimran Singh (Coordinator) |  | 17332204549 |

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| **S. No** | **Title** | **Page Number** |
| 1. | INTRODUCTION | 3-4 |
| 2. | **WHY THIS PROJECT?** | 5 |
| 3. | **OBJECTIVES** | 6-7 |
| 4. | **FRAMEWORK DEVELOPMENT** | 8-11 |
| 5. | **PROJECT PROGRESS** | 11 |
| 6. | EXPLORATORY DATA ANALYSIS | 12-14 |

INTRODUCTION

Overview of the Project

Mental health is an essential aspect of human well-being, affecting emotional, psychological, and social stability. With the increasing prevalence of mental health disorders such as anxiety, depression, and stress, early detection and intervention are crucial for improving an individual's quality of life. Traditional methods for diagnosing mental health conditions rely on clinical assessments and self-reported questionnaires, which can be time-consuming, subjective, and limited in accessibility.

This project, Real-Time Mental Health Analysis, leverages Machine Learning (ML) and Deep Learning (DL) to analyze three key modalities: speech, facial expressions, and text to detect emotional states in real-time. By integrating multiple data sources, the system aims to provide a more comprehensive, automated, and objective approach to mental health assessment.

Motivation

The motivation behind this project arises from the increasing need for technology-driven mental health solutions. Mental health disorders often go unnoticed until they reach critical stages, mainly due to the lack of real-time monitoring tools and the stigma associated with seeking help. Some of the key reasons for developing this project include:

* Early Detection & Prevention: Early diagnosis can help in timely intervention and reduce the severity of mental health conditions.
* Objective Analysis: AI-driven analysis can eliminate human biases associated with traditional mental health assessments.
* Accessibility & Scalability: A technology-based system can provide assistance to a large number of individuals without the need for in-person clinical visits.
* Multi-Modal Approach: Combining speech, face, and text analysis allows for a more holistic understanding of an individual's emotional state.

Project Objectives

The primary goal of this project is to build a real-time, AI-powered mental health detection system that integrates multiple sources of emotional data. The system will:

* Analyze Speech Patterns: Detect emotional states based on tone, pitch, and modulation.
* Recognize Facial Expressions: Identify emotions using deep learning models applied to real-time facial images.
* Perform Text Sentiment Analysis: Extract emotional insights from text, including speech-to-text transcripts and user inputs.
* Provide Actionable Insights: Generate a mental health status report and suggest interventions based on detected emotional states.

Expected Outcome

At the end of this project, we aim to develop:

* A real-time AI model capable of detecting emotions from speech, face, and text with high accuracy.
* A user-friendly system that allows individuals to interact via voice or text and receive instant emotional analysis.
* A deployed application using Flask, Fast API, or another framework for real-time processing.
* An impactful tool that can assist healthcare professionals and individuals in monitoring mental health proactively.

This project represents a significant step toward the AI-driven future of mental health diagnostics, enabling early intervention and promoting well-being through cutting-edge machine learning and deep learning techniques.

**WHY THIS PROJECT?**

### Mental health remains a **stigmatized and often overlooked issue in India**, where societal norms discourage open discussions about emotional well-being. Despite the increasing awareness of mental health worldwide, many individuals in India still **hesitate to seek help** due to fear of judgment, cultural taboos, and a lack of understanding. This results in people **suffering in silence**, leading to worsening conditions like **anxiety, depression, and emotional distress** that often go undiagnosed and untreated.

One of the major challenges is that mental health is **not perceived as a priority** in many Indian households. Discussions around emotions and psychological struggles are often dismissed as **"just a phase"** or a sign of **weakness**, rather than being acknowledged as real and serious concerns. The shortage of mental health professionals, combined with the **high cost of therapy** and the reluctance to approach therapists, makes access to professional help difficult, especially in **rural areas** where mental health facilities are almost non-existent.

Given this scenario, there is an urgent need for **technology-driven solutions** that can bridge the gap between mental health awareness and early diagnosis. By leveraging **Artificial Intelligence (AI) and Deep Learning**, this project aims to develop a **real-time mental health detection system** that can analyze emotions through **speech, facial expressions, and text**. Such a system can act as a **first point of contact** for individuals struggling with mental health issues, offering **early detection, emotional analysis, and recommendations** without requiring them to step out of their comfort zone.

This project is not just about technological advancement; it is about **creating a supportive and accessible environment** where individuals can assess their mental well-being **privately and without stigma**. By integrating **AI-driven emotional analysis**, we hope to contribute to breaking the **cultural barriers** surrounding mental health discussions in India and make psychological well-being a **more openly addressed and accepted topic** in society.

**OBJECTIVES**

### **Primary Objectives**

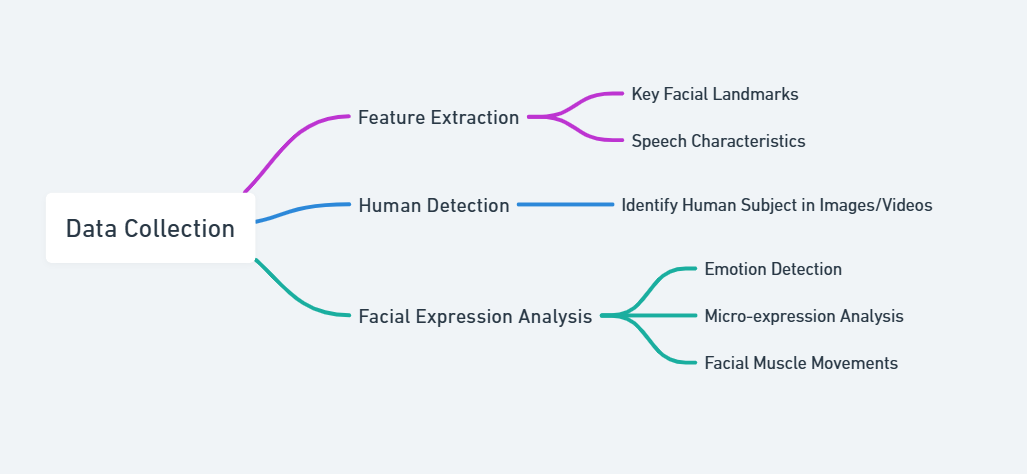
1. **Develop a System for Multi-Modal Emotion Recognition**
   * This project aims to create an AI-powered system capable of analyzing speech, facial expressions, and text sentiment to detect emotions.
   * By leveraging deep learning techniques, the system will process real-time inputs and classify emotional states with high accuracy.
2. **Build a Real-Time AI-Powered Mental Health Monitoring Tool**
   * The system will operate in real-time, enabling continuous monitoring of emotional states without delays.
   * It will use an optimized model for fast inference, making it suitable for live applications such as virtual therapy sessions.
3. **Improve Mental Health Detection Accuracy Using Multi-Modal Data Fusion**
   * Individual emotion recognition methods may have limitations, but combining speech, facial, and text analysis will enhance prediction accuracy.
   * The system will integrate these modalities using fusion techniques to provide a more comprehensive emotional assessment.

### **Secondary Objectives**

1. **Create an Interactive Interface for Real-Time Emotion Analysis**
   * A user-friendly interface will be designed for seamless interaction, allowing users to input speech, text, or facial expressions easily.
   * Real-time visualization will be incorporated, displaying emotion analysis results in an understandable format.
2. **Use Advanced Deep Learning Models for Improved Sentiment Classification**
   * State-of-the-art deep learning models such as Convolutional Neural Networks (CNN) for face emotion detection, Recurrent Neural Networks (RNN) for speech analysis, and Transformer models for text sentiment analysis will be employed.
   * Transfer learning and pre-trained models will be utilized to improve efficiency and reduce training time.
3. **Ensure Scalability and Robustness for Real-World Applications**
   * The system will be designed to function efficiently across different devices and platforms, including web and mobile applications.
   * Cloud-based deployment options will be explored to support large-scale usage and real-time processing demands.

By fulfilling these objectives, the project aims to provide a reliable and scalable AI-powered mental health detection system that can contribute to early emotional distress detection and intervention.

**FRAMEWORK DEVELOPMENT**

The mental health detection system collects real-time data from speech, facial expressions, and text using microphones, cameras, and textual input. It extracts key features such as facial landmarks, MFCCs, pitch, tone, and spectral features for accurate emotion recognition.

Human detection ensures that only relevant subjects are analyzed using deep learning models like YOLO or MTCNN. Facial expression analysis detects emotions, micro-expressions, and facial muscle movements to assess psychological states. Speech emotion recognition employs CNN-LSTM, while text sentiment analysis uses NLP models like BERT or RoBERTa to interpret emotional context.

Fusion techniques such as weighted averaging and decision-level fusion combine predictions from speech, facial, and text data to enhance accuracy. Machine learning and deep learning models are trained and evaluated using performance metrics like accuracy, precision, recall, and F1-score.

The system is deployed in a real-time environment with an intuitive interface providing live mental health assessments. Graphical dashboards display emotional trends, enabling early detection and intervention. This integrated AI-driven framework offers a comprehensive and scalable solution for real-time mental health analysis.

**Data Collection**

* The facial recognition part of this project uses a custom dataset with images showing different emotions like Anger, Contempt, Disgust, Fear, Happiness, Neutral, Sadness, and Surprise. These images were taken in a controlled setting and labeled manually to make sure they are correctly categorized. Since the data is created and sorted by hand, it is more accurate and free from errors that can be found in pre-made datasets.

To analyze these images, deep learning models like Convolutional Neural Networks (CNNs) are used. The system looks at facial features, muscle movements, and small changes in expressions to understand emotions better. To make the model work well in different conditions, techniques like image rotation, flipping, and brightness changes are applied. This helps the model learn how to recognize emotions even when the lighting or face position is different.

Once the images are processed, the model predicts the person’s emotion by matching it with the trained categories. This allows the system to detect emotions in real-time, helping in mental health analysis. When combined with speech and text analysis, facial recognition makes the system more accurate and useful in understanding a person’s mental state.

* The system does not use a predefined dataset; instead, it processes real-time textual input directly from users. This dynamic approach allows for flexible and adaptive data handling, ensuring that the model can analyze a wide range of text types. Various NLP libraries and models facilitate this real-time text processing.

The transformers library provides pre-trained models for *text summarization* and *sentiment analysis*, enabling efficient processing of textual data. TextBlob extracts polarity and subjectivity scores, determining the emotional tone and distinguishing between factual and opinion-based content. spaCy is employed for *Named Entity Recognition (NER)*, allowing the system to identify key entities such as names, locations, and organizations within the text.

For sentiment classification, the nltk package, specifically the *SentimentIntensityAnalyzer (VADER)*, is utilized to provide detailed sentiment scores, making the system capable of analyzing both short and long texts. Additionally, textstat computes readability scores using Flesch-Kincaid metrics, ensuring that the system can assess text complexity and suitability for different audiences.

* For the voice emotion recognition model, we are utilizing existing datasets that contain labeled speech recordings representing various emotional states, such as anger, happiness, sadness, fear, surprise, contempt, disgust, and neutrality. These datasets provide a strong foundation for training deep learning models in speech-based emotion classification.

We are currently exploring and selecting publicly available speech emotion datasets, such as:

RAVDESS (Ryerson Audio-Visual Database of Emotional Speech and Song)

* CREMA-D (Crowd-Sourced Emotional Multimodal Actors Dataset)
* IEMOCAP (Interactive Emotional Dyadic Motion Capture Dataset)
* TESS (Toronto Emotional Speech Set)

These datasets contain professionally recorded and labeled audio samples, ensuring high-quality training data for our model.

Once the dataset is finalized, we will process the speech data by extracting essential acoustic features such as Mel-Frequency Cepstral Coefficients (MFCCs), chroma features, spectral contrast, pitch, and zero-crossing rate. These extracted features will be used to train deep learning models like CNN-LSTMs, Wav2Vec, and Transformer-based architectures for accurate classification of speech emotions.

This approach enables us to build a robust and scalable voice emotion recognition model that will be integrated into our multi-modal mental health detection system, working alongside facial and text-based emotion analysis to provide comprehensive insights.

### **PROJECT PROGRESS**

In the past two months, I have made good progress in building the foundation for this project. To develop a better understanding of machine learning and deep learning, I have practiced coding in frameworks like TensorFlow and PyTorch. I have also worked on speech processing, facial emotion recognition, and sentiment analysis, which are key parts of the project.

I initially started working with Kaggle datasets for facial emotion recognition, but they did not give the expected results due to low-quality images and inaccurate labels. After facing challenges with these datasets, I decided to create my own dataset by capturing and manually labeling images representing different emotions like anger, happiness, sadness, fear, surprise, contempt, disgust, and neutrality. This has improved the accuracy of the facial emotion recognition model.

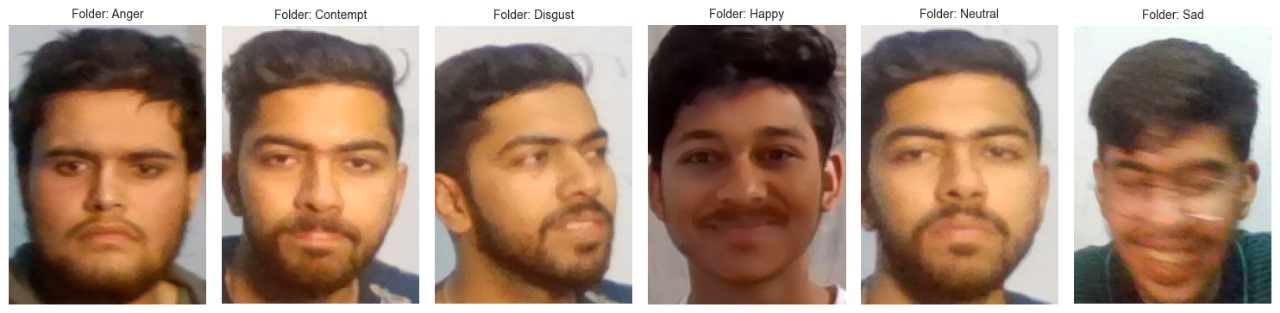
I have studied several **research papers** on how AI can be used for mental health detection. These papers helped me understand different techniques for **emotion recognition from speech, face, and text**, as well as the challenges and improvements needed in existing models.

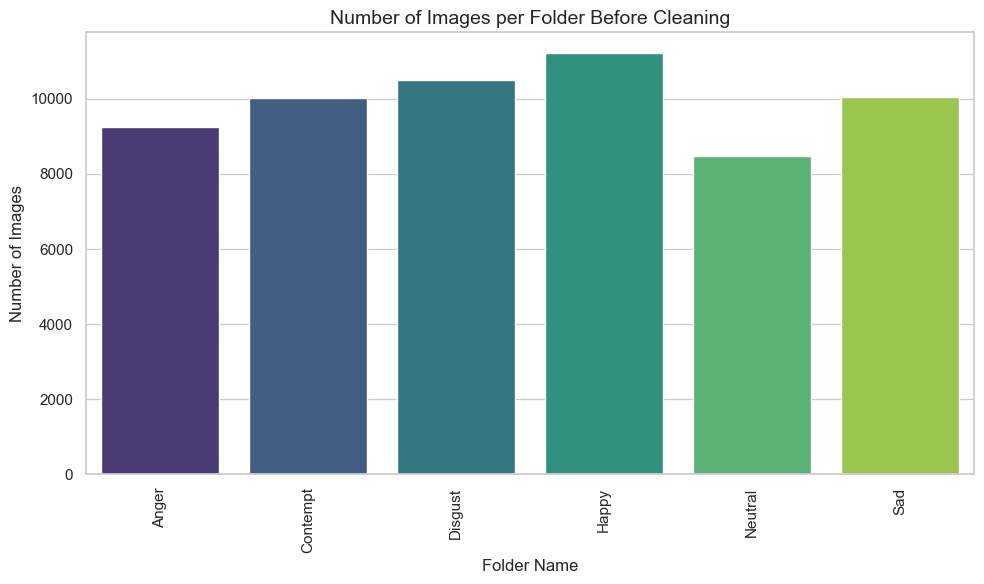
A major focus has been on **finding the right datasets**. I have explored and selected **publicly available datasets** such as **RAVDESS, CREMA-D, IEMOCAP, FER-2013,** toensure high-quality training data. I have also researched **preprocessing techniques** to prepare this data for model training.

Apart from technical work, I have also spent time **learning about mental health**, including **how emotions are expressed through speech, facial expressions, and text**. This has helped in designing a system that can **detect mental health conditions more accurately**.

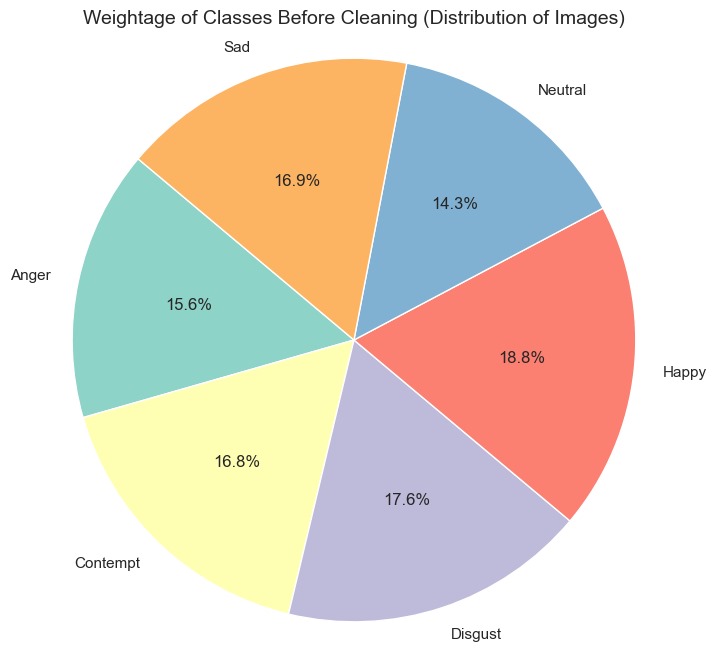
EXPLORATORY DATA ANALYSIS

The dataset contains **manually categorized facial images** representing emotions like **Anger, Contempt, Disgust, Fear, Happy, Neutral, Sad, and Surprise**. All images are **grayscale** and resizedto **48x48 pixels** for consistency in model training.



Sample Images for Each Emotion  
This grid of images represents sample faces from each emotion category in the dataset: Anger, Contempt, Disgust, Fear, Happy, Neutral, Sad, and Surprise. The purpose of this visualization is to ensure that the dataset contains clear and well-distinguishable facial expressions. By observing the images, we can check for image quality, lighting conditions, and variations in expressions. 

**Emotion Class Distribution Chart**  
This bar chart shows the **number of images in each emotion category**. It helps to analyze whether the dataset is **balanced** or if some emotions have significantly more or fewer images than others. If an emotion class has very few images, the model might struggle to learn it properly, leading to biased predictions.

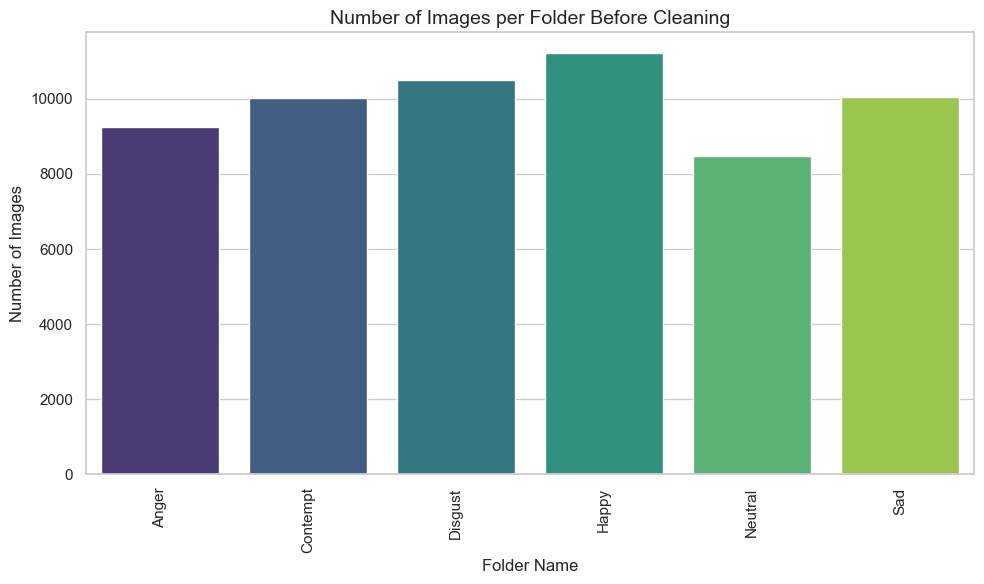


Observations

If some emotions have much fewer images than others, we might need to collect more data or apply data augmentation to balance the dataset.

A well-balanced dataset ensures better generalization and prevents the model from being biased toward a dominant class.

**Resized Sample Images for Model Training**  
This visualization shows the **processed images after resizing them to a uniform size (48x48 pixels)**. Since deep learning models require consistent input sizes, resizing is a crucial preprocessing step. By ensuring that all images are of the **same resolution**, the model can process them efficiently without distortion or information loss.



sss

\*\* For detailed code EDA.ipynb file is uploaded to the Git Hub repository \*\*