National College of Ireland

Multimodal Stress Analysis Using Traditional and Deep Learning Models

MSc Research Project
MSc in Data Analytics

Chethan Sureshbabu
Student ID: x21235091

School of Computing
National College of Ireland

Supervisor: Mrs. Harshani Nagahamulla

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: Chethan Sureshbabu

Student ID: x21235091

Programme: MSc in Data Analytics **Year:** 2024

Module: MSc Research Project

Supervisor: Mrs.Harshani Nagahamulla

Submission Due Date: 31/01/2024

Project Title: Multimodal Stress Analysis Using Traditional and Deep Learning Models

Word Count:6588

Page Count: 21

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Chethan Sureshbabu

Signature: ……

Date: 31st January 2024

……

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Multimodal Stress Analysis Using Traditional and Deep Learning Models

Chethan Sureshbabu
X21235091

Abstract

The current field of stress analysis demands sophisticated diagnostic systems for accurate assessments, particularly in cardiovascular health. Traditional methods, like Echocardiography (ECG), are often considered time-consuming and require specialized knowledge. This project is a pioneering effort to transform stress analysis by integrating audio, text, and image datasets. The goal is to create a comprehensive framework that can help healthcare professionals, regardless of their specialization, in making swift and accurate stress-related diagnoses. Our approach is designed to simplify stress analysis by employing a multi-data approach. Robust data mining methodologies are developed for extensive medical datasets. A hybrid model, integrating feature selection and classification algorithms, is proposed to identify crucial stress-related features and categorize stress levels. Model performance is evaluated based on accuracy, precision, and F1-score. The integrated model, utilizing audio, image, and text data, effectively identified key stress-related features from different datasets. The Late Fusion model achieved a good classification accuracy of 80%, 80% weighted average F1-score and precision values of 86% and 73% for non-stress and stress classes. The combination of audio, image, and text data showcases the comprehensive nature of stress analysis. Comparative analysis reveals that the model's accuracy (80%) surpasses conventional diagnostic methods (ECG) and aligns with contemporary stress analysis frameworks. This research provides an efficient model for stress analysis, integrating data from different sources to enable healthcare professionals to make stress-related diagnoses more effectively. Future work will focus on fine-tuning the model for optimal performance across varied stress scenarios.

Keywords- Stress Analysis, Multi-Data Approach, Integrated Model, Late Fusion model.

1 Introduction

The task of Stress analysis is to classify data into Stressed or Not Stressed. As of now most of the work on Stress analysis is done on textual data. With the rise of social media, people started sharing information in the form of video, audio, and text. So multimodal data have been required with changes in conventional sources of communication Abburi et al. (2016). Among these challenges, the emergence of stress-related issues in online communities has become a main point of concern. Using traditional Machine Learning methods for stress analysis often falls short of capturing the complex and dynamic nature of stress in the digital realm instead we can use deep learning neural networks that can best suit these kinds of problems. The analysis of text, image, audio, and video can be used to work on the emotions and analyse the emotions of the persons Aggarwal et al. (2020). The problem with using only text for analysing the stress lies in that it might miss important signals. Text doesn't show things like tone of voice or facial expressions, which are important for understanding stress. But when we are integrating multiple modalities such as audio, image, and video can provide

a more comprehensive approach to stress analysis by capturing a broader spectrum of emotional cues.

1.1 Background:

According to the Researchers around 72% of people utilize social media applications, and the most frequently used ones are Facebook, Instagram, Twitter, Snapchat to interact and share their thoughts with others Selvadass et al. (2022). Diagnostic systems, such as Echocardiography, play a crucial role in identifying stages of cardiovascular diseases but it requires highly skilled professionals. Similarly, the stress analysis also demands a state-of-the-art approach. By delving into the state of the art, we identify gaps in current methodologies and propose a comprehensive approach for stress analysis.

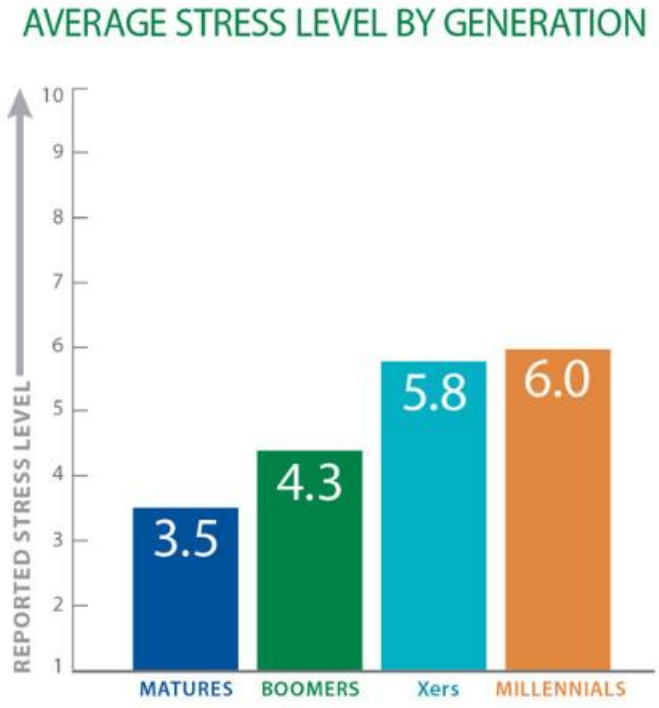

Figure 1: stress level among various generations .

Figure 1 shows the increase in stress level among various generations as reported by the American Psychological Association.

1.2 Motivation:

The motivation behind choosing this topic for research is to understand and address the impact of stress on online platforms. Most of the old ways of doing things focused on using only text, audio, videos, or images. But here's the thing life is a mix of all three. Our research found that the current methods are struggling with combining them. Our approach is a new way to analyse stress by making use of text, audio, and images. By combining text, audio, and images, so that the user can get a more Comprehensive Understanding with each mode providing unique insights. We can easily identify whether a person is in happy mode, sad mode, angry mode, or stressed mode by seeing facial expressions, but if we talk about the

text, we can't properly capture the emotions Mehta et al. (2019) Furthermore using all three modes improves the accuracy of stress detection.

1.3 Research Question and Objectives:

Can a fusion of text, audio, and visual data, analysed through machine learning methods, provide a sophisticated classification of stress levels in social media content with improved accuracy?

To solve this question, we break this down into smaller research goals:

1. To understand existing approaches, methodologies, and technologies employed in stress analysis across diverse datasets.

2. To create a comprehensive framework that unifies the preprocessing steps and machine learning models for each dataset—sentiment 140 tweets, facial expressions in AffectNet, and emotional tones in RAVDESS.

3. To implement the Fusion: Execute the designed framework, integrating insights from text, audio, and visual data to classify stress levels.

4. To assess the performance of the fused model, accuracy, sensitivity, and specificity.

1.4 Contribution:

The main contribution of this research is the development of a stress analysis system that combines the predictions of text, audio, and image data and provides a comprehensive understanding of stress levels.

1.5 Structure of the Paper:

In the following sections, we will discuss the state of the art, present the methodology employed, and critically evaluate the performance of our proposed stress analysis system.

2 Related Work

2.1 Stress detection using Text data:

Mounika et al. (2019) aim to detect stress levels in students through the social media platform. And as input, they have extracted data from Twitter. They have considered emojis in sentiment analysis, constructing an emoji-sentiment dictionary, and ranking based on occurrences. considered Recurrent Neural Network (RNN) for sentiment analysis. The model aims to detect stress in students using RNN, Results from that is suggesting that two out of three students may experience stress. They are looking to predict stress using deep neural networks for improved accuracy for future work. Overall, this work contributes to the field of stress detection And provides useful insights into the mental well-being of students.

Jadhav et al. (2019) have done stress detection using textual data (tweets, comments, chats). They have used LSTM, BLSTM, and SVM Algorithms for stress detection. Facial recognition involves landmark detection, ROI localization, feature extraction using GLDM, and classification using Naive Bayes. Speech recognition uses MFCC and Teager Energy for feature extraction and classification with GMM. Attention-based LSTM model has been used for detecting psychological stress. In conclusion, BLSTM has less error rate than LSTM and SVM shows better accuracy only in the case of labelled data. LSTM and BLSTM can work on both labelled and unlabelled datasets. Bidirectional LSTM shows the best accuracy for stress detection. Their future work is implementing stress detection systems based on text, speech, and facial expressions.

Giuntini et al. (2021) aims to detect stress using raw data extracted from social media, particularly Twitter. That raw data has undergone a meticulous cleaning and preparation process to ensure quality. They have used Sequential Pattern Mining to track temporal behaviour patterns in depressive users on social media. The data underwent preprocessing steps such as feature extraction, context analysis, and temporal analysis. TROAD Framework: The core of the system, TROAD utilizes a computational approach that identifies optimal intervals, extracts emotional and contextual features, and models these features into time windows for recognizing sequential patterns in depressive user behaviour. The TROAD framework showcases strong sequence patterns with a minimum of 70% support, 81% confidence, and 69% sequential confidence, considering periods of silence between users' posts. Without accounting for silent periods, the rules still hold ground with 70%, 86%, and 38% in terms of support, confidence, and sequential confidence. Then the findings are visually represented through sentiment percentages (positive, negative, neutral) using tools such as Matplotlib. In conclusion, The TROAD framework emerges as a promising tool for clinical specialists, offering a unique perspective into the temporal evolution of emotional behaviour in users dealing with depression on social media. The authors suggest extending the work to predict stress using deep neural networks for improved accuracy.

Shaw et al. (2022) have proposed a Model for automatic stress detection and used deep learning-based models for automatic feature extraction. For this study, they have used Multichannel CNN, GRU, and BERT models for detecting stress. Multichannel CNN emerges as the top-performing architecture, achieving an impressive accuracy of 97.5%. Precision, recall, and f-score values are reported as 96.8%, 97.5%, and 97.2%, respectively. In conclusion, they have successfully implemented a model for automatic stress detection from Twitter text. Multichannel CNN proves to be the best model achieving high accuracy and reliability in identifying tweets expressing psychological stress. According to the authors, Future work is to explore deep learning models for stress detection for improved accuracy in stress detection of the messages of Twitter data.

Rastogi et al. (2022) the authors have recognized stress detection as a fundamental task in assessing mental health. They have used four high-quality datasets for stress detection from textual inputs on Twitter and Reddit. Study involving rule-based and machine learning approaches to assess stress detection performance. They have examined the existing systems for stress detection and have Identified the limitations and areas for improvement in current methodologies. The authors have discussed that they have adopted Transformer-based models for superior performance. future research is on stress detection from social media texts and making advancements in the field with a focus on reliable stress detection models.

Selvadass et al. (2022) Utilization of supervised machine learning algorithms to identify stress in social media posts. Focus on posts without explicit keywords like "stress" or "tension" to capture subtle expressions of stress. Identifying stress-inducing factors such as inadequacy, resentment, and seclusion arising from virtual socialization. Implementation of

two textual-based feature extraction methods BERT and TF-IDF. Applying machine learning classifiers to analyse sentiment and categorize posts into 'stress' and 'non-stress.' They have achieved the highest accuracy of 75.80% with the BERT fine-tuned model. Utilized metrics to identify the optimal model for stress analysis in social media. discussed the importance of stress in individuals before advising or treating them. Suggestions for future research directions, exploring advanced machine learning approaches, and expanding the study to a broader spectrum of social media platforms.

2.2 Stress detection using visual or image data:

Cacciatori et al. (2023) will be using visual data like a user's video stream from the webcam for stress detection. In this paper, they have discussed the development and experimental process of a platform for facial expression recognition and stress analysis. They have implemented neural network algorithms for facial stress detection. The goal is to enhance the accuracy of facial expression recognition and increase response speed. Special attention was given to overcoming the hardware limitations associated with running ANN. The implemented algorithms look to improve performance to ensure efficient real-time analysis without compromising on accuracy.

According to the research conducted by Upadhyay et al. (2020) Facial Feature Extraction for Emotion Recognition. The authors have discussed about the importance of facial features in the field of emotion recognition they have done a detailed exploration of the facial features. The research provides an extensive overview of different facial feature extraction methods. They have tried to identify the strengths, limitations, and application scenarios of each method, aiding researchers in making informed choices. The authors have discussed the real-world implications of the facial feature extraction methods in detail. They have addressed some of the key challenges while dealing with facial feature extraction methods.

Kulatilake et al. (2022) Examination of user behaviour on selected social media platforms (Facebook and Twitter). Implementation of machine learning-based sentiment analysis to identify negativity and depressiveness in "Sinhala" content. They have introduced a chatbot communicating with users in the "Singlish" language helping the Enhancement of user interaction and support through the chatbot. The authors have recognized that depression is a widespread issue, especially among students and teenagers. Identifying social media over-usage as a major contributor to procrastination and depression. they have done the mobile application with four key components:

Facial emotion tracking, Eye aspect ratio analysis, Identification of user fatigue, Procrastination tracking. Plan for AutoML approaches to enhance analysis efficiency. They have discussed Shifting from traditional machine learning to AutoML approaches for improved efficiency.

2.3 Stress detection using text and audio data:

Abburi et al. (2016) for this research they have used Diverse product reviews shared on social media in multiple modalities audio, text, and video. Preparing and cleaning the data for analysis, and then ensuring a standardized format across modalities as it is a mix of both audio and text data. Extracting Mel Frequency Cepstral Coefficients (MFCC) features specifically from stressed significant regions. Detection based on the strength of excitation in the audio input. Employing Gaussian Mixture Models (GMM) classifier to develop a sentiment model using these features. They have observed that MFCC features extracted from stressed significance regions outperform features from the entire audio input.

Computing textual features using Doc2vec vectors from the transcript of the audio input. Developing a sentiment model using a Support Vector Machine (SVM) classifier for textual features. The authors have found that the performance has increased by combining both audio and text features in sentiment detection. In conclusion, they have successfully implemented a multimodal sentiment detection system. further enhancements are incorporating additional modalities (like video) and refining feature extraction techniques for comprehensive sentiment analysis.

2.4 Stress detection using text, audio, and image/video data:

Aggarwal et al. (2023) for this research the authors used social media posts like text, audio, and videos. They have Cleaned and standardized data for analysis across multiple modalities. used machine learning and artificial intelligence for emotion classification in written text. Employing methods to analyze emotions conveyed through audio content. Extending emotion classification to images and videos using advanced algorithms. Evaluating the intensity of emotions expressed in the content. In conclusion, they have successfully implemented emotion classification across a wide range of social media content. For future work they have decided to Investigate potential applications in online shopping, healthcare, and social media.

2.5 Stress detection using various approaches:

Tajuddin et al are working on social media microblogs for this study of stress detection. The old conventional methods like psychologist interviews seem laborious and unworkable. To detect stress more accurately, this research offers an Effective Stress Detection system using hybrid ontology. The need for stress management for both individuals and organizations are discussed in the paper. And they have made use of probabilistic models like GSHL and tree alignment algorithms. For future work, they will be focusing on increasing the accuracy of stress detection. Additionally, Tajuddin et al. (2020) aim to extend stress detection to multilingual languages.

Joshi et al. (2033) suggests the connection between mental disorders and social media activities, emphasizing that individuals with mental health issues tend to use more online platforms. they have mentioned that the users express their emotional states through public posts, tweets, and related YouTube content. A model is developed to alert users to negative patterns in their exhibited state of mind. They have employed the LSTM algorithm, a deep learning method, to identify users' emotions among six defined categories. For this, the authors have gathered user data from four sources that are Twitter, YouTube, Gmail, and a personal journal. The Bidirectional LSTM algorithm processes this data for identifying and storing emotional states for further analysis and to understand the stress.

Mehta et al. (2019) in this research they discussed Emotion Intensity Recognition providing insights into the selection process and the variables considered for emotion intensity recognition. A study is conducted to explore various machine learning algorithms for emotion intensity recognition and capturing and categorizing different levels of emotional intensity. And employed evaluation metrics to assess the performance of each machine-learning algorithm including accuracy, precision, recall, and other relevant metrics. They have discussed the significance of accurate emotion intensity recognition. Future work is to do some more research in the field of emotion intensity recognition.

Rajendar et al. (2022) Address the psychological stress affecting individuals and its impact on work, academic, and health aspects. They have Explored recent advancements in Human Stress Level Detection Systems (HSLDS). Combining data from different data sources,

including human physiology, physical characteristics, social media posts, and micro-blogs. Utilized deep learning and machine learning models for stress detection. The performance is evaluated based on precision, recall, accuracy, and other relevant metrics. They have addressed issues in stress detection of existing methods, and the need for further exploration in this field. Contribution to the understanding of stress detection systems and their impact on various aspects of human life.

3 Research Methodology

The research methodology for stress analysis aligns with the KDD (Knowledge Discovery in Databases). The methodology involves several key stages:

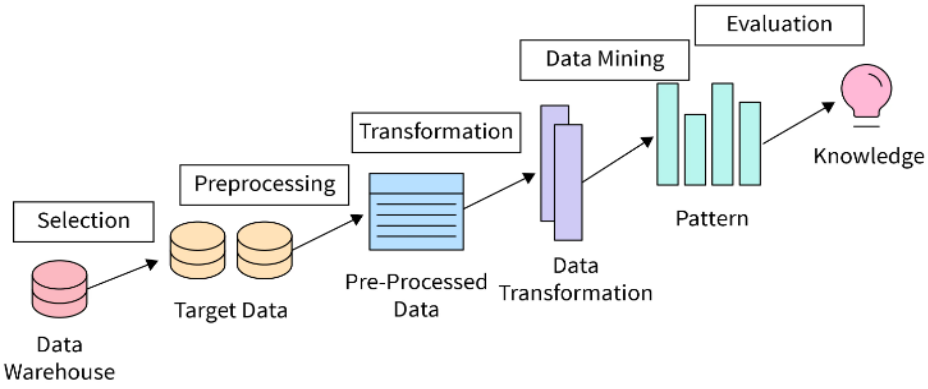

Figure 2: Process involved in KDD Methodology

In conducting this research on stress analysis, a multi-modal approach has been used to analyse stress in social media content using text, audio, and visual data. The methodology was informed by a comprehensive review of related work in stress analysis and multimodal fusion. For data collection, datasets of text, audio recordings, and images have been used. The research procedure involved building individual models for Audio, image, and Text data utilizing machine learning algorithms.

3.1 Dataset Collection and Exploration:

3.1.1 Dataset Selection:

Three distinct datasets were chosen to represent different modalities: Sentiment140 for text, Facial Expressions Training Data for images, and RAVDESS Emotional Speech Audio for audio. All three datasets are publicly available on Kaggle.

3.1.2 Data Understanding:

The nature of each dataset was comprehensively explored, including the sentiment140 dataset with 1.6 million tweets, Facial Expressions Training Data based on Affect Net-HQ, and RAVDESS Emotional Speech Audio containing audio expressions of various emotions.

Sentiment140 Dataset (Text)					
Sentiment	**IDs**	**Date**	**Flag**	**User**	**Text**
0 or 4 (high-stressed or low-stressed)	Unique identification numbers	Timestamp when the tweet was posted	Query tag	Twitter username	Actual content of the tweet

Facial Expressions Training Data (Image)			
Index	**File path (pth)**	**Emotion Label (e.g., anger, contempt, disgust)**	**RelFCs**
Index column	File path of the image	Emotion labels associated with the image	Relevance factor associated with each image

RAVDESS Emotional Speech Audio (Audio)	
File Path (path)	**Emotion**
File path of the audio	Integer labels representing different emotions

Table 1: Information of Text, Audio, and Image data

3.2 Preprocessing and Feature Extraction:

3.2.1 Textual Data:

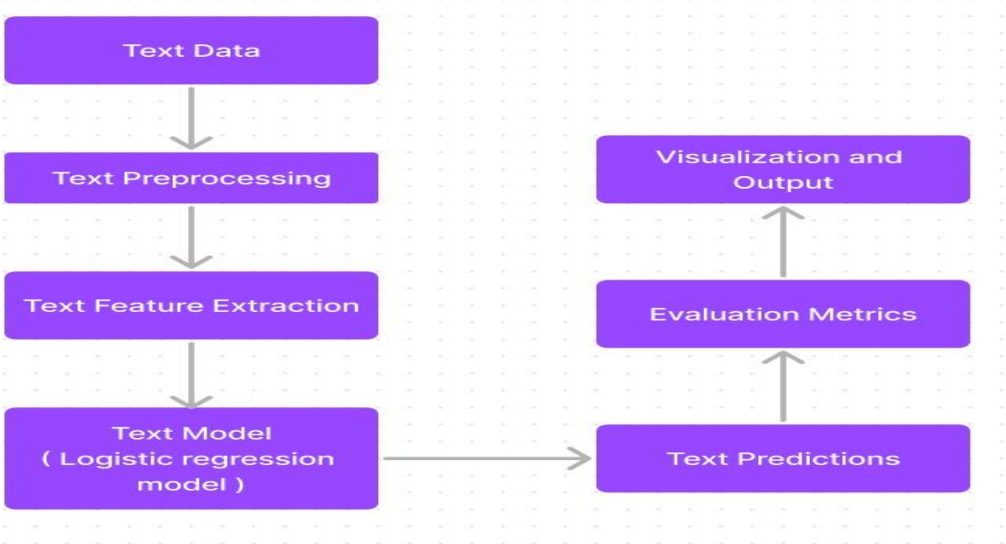

Figure 2: Text model workflow diagram

First, the columns were renamed for better clarity: ["sentiment", "ids", "date", "flag", "user", "text"]. Then the sentiment classes (0 and 4) were mapped to human-readable labels ("high stressed" and "low stressed"). Various text preprocessing steps were applied to clean the tweet text, such as converting text to lowercase, replacing URLs with 'URL', replacing emojis with corresponding meanings, Replacing Twitter usernames with 'USER', removing non-alphanumeric characters, changing consecutive letters with two occurrences, and removing stop words. The TF-IDF vectorizer was utilized to convert the preprocessed text into numerical features. The vectorizer was configured with a word n-gram range of (1, 2) to capture unigrams and bigrams. The maximum number of features were limited to 500,000. Figure 2 describes all the steps used for text model building.

3.2.2 Image Data:

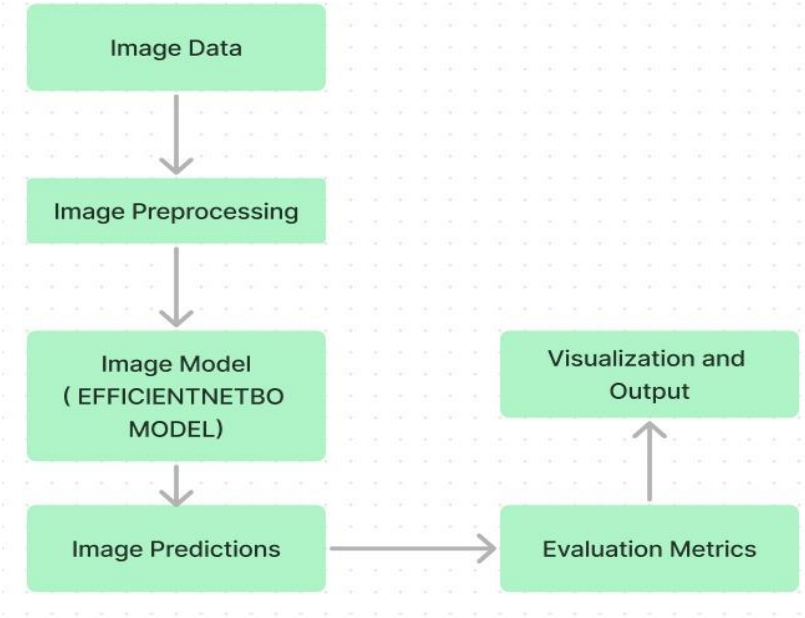

Figure 3: Image Model workflow diagram

Existing Emotion labels have been mapped to stress levels, then rows with missing values have been dropped, z-score was applied to remove outliers from the 'relfcs' column, Plotted The distribution of stress levels were plotted and the count of instances for each stress level were displayed, min-max scaling was applied to the 'relfcs' column, The dataset has split into training and testing sets, data augmentation was applied to the numerical features using the image data generator. Then images were loaded and preprocessed for both training and testing, Data augmentation was applied to the training images using the image data generator, and rotation, width shift, height shift, shear, zoom, and horizontal flip to augment training images were applied. Figure 3 describes the steps we used for image model building.

3.2.3 Audio Data:

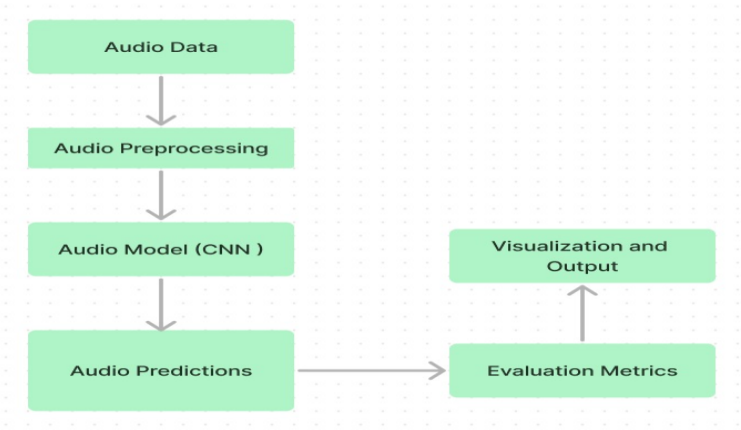

Figure 4: Audio Model workflow diagram

Label Mapping: emotion labels were mapped to stress levels ('low_stress' and 'high_stress').
Feature Extraction Function: A function was defined to extract audio features (MFCCs, chroma, mel, contrast, tonnetz) using the librosa library.
Feature Extraction:
Features were extracted from each audio file and a Data Frame was created, Installing the 'resampy' library, dropping rows with missing features, encoding stress level labels using Label Encoder, converting encoded labels to categorical using to categorical, and standardizing the features using a standard scaler. Figure 4 describes the workflow for building the audio model.

3.3 Individual Model Training:

3.3.1 Text Model:

For the text data, three models: Bernoulli Naive Bayes, Linear Support Vector Classification, and Logistic Regression have been implemented and evaluated. The best model selected based on the evaluation matrix is selected for the Late Fusion model. The best model was the Logistic Regression with an accuracy of 83%. Therefore, we have used the Logistic Regression for the Late Fusion model.

3.3.2 Image Model:

The EfficientNetB0 model was implemented for image classification with a binary output, Designing a convolution neural network (CNN) architecture. The model architecture included a pre-trained base model followed by dense layers. We trained the model on image data, achieving a test accuracy of 75.91%.

3.3.3 Audio Model:

The audio data undergoes stress classification using a neural network model (CNN), comprising two dense layers with dropout to prevent overfitting. The model is trained on audio data for 50 epochs, resulting in an accuracy of 83.68% on the test set.

3.4 Late Fusion Model Development:

Following the construction of individual models for the Text, Audio, and Image data, the features from the individual models were concatenated to create a unified feature set for late fusion. The concatenated features were standardized using a Standardized standard scaler. The reason behind scaling features is to ensure uniformity and facilitate model training. The

late fusion model employed a logistic regression model which was trained on the standardized feature set.

4 Design Specification

The design specification describes the architecture of the late fusion model, involving the integration of text, audio, and image features. The late fusion process was implemented through concatenation and standardization of features. The requirements included compatibility of features across modalities and the incorporation of a logistic regression model for final classification.

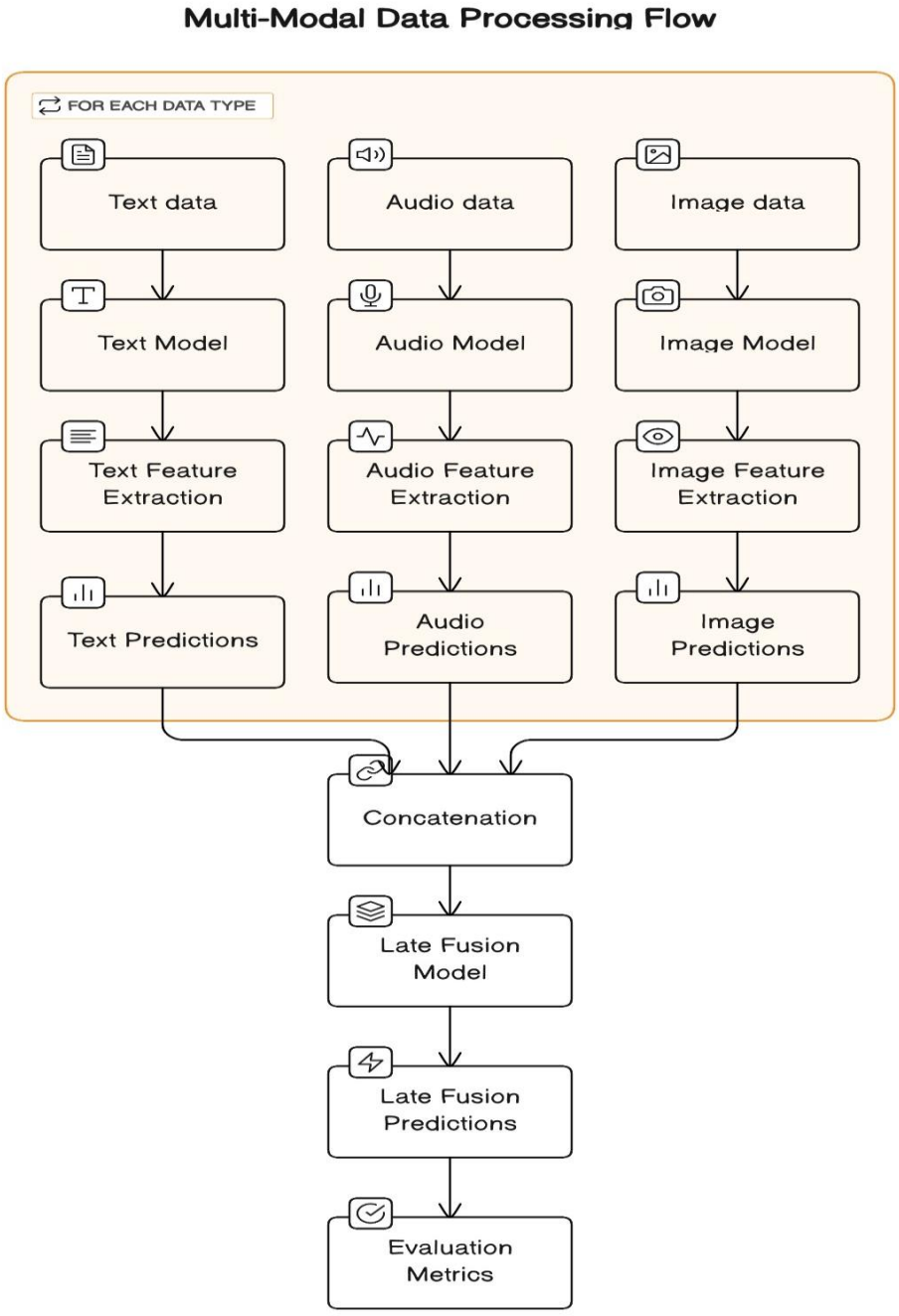

Figure 5: Late Fusion Model Architecture

Figure 5 above represents the architecture of the Late Fusion model.

The design specification of the Late Fusion Model is as follows:
The data used for the late fusion model were taken from the test data of Audio, image, and text datasets. After building the models individually the models were then saved to use for the late fusion model. The saved models were loaded, and the test data of audio, image, and text underwent some pre-processing.

Audio Processing:
Feature Extraction: A specialized classification algorithm, designed for audio data, Which utilizes a convolutional neural network (CNN) was used. This CNN model was been used to process and extract features from the RAVDESS Emotional Speech Audio dataset and capture speech emotions.

Image Processing:
Feature Extraction: Image preprocessing techniques such as normalization and resizing were used on facial expression training data. The Efficient Net-based model was employed for feature extraction, and then capturing complex visual details required for stress analysis.

Text Processing:
Feature Extraction: preprocessing for the Sentiment140 dataset included data handling, cleaning, and tokenization. The TF-IDF vectorization technique was employed for feature extraction, converting the textual information into a numerical format, and preserving the essential information.

Late Fusion Model:
Feature Concatenation: The features extracted from the audio, image, and text models were concatenated, creating an integrated feature set. This step ensured the integration of diverse modalities and prepares the data for the final classification stage.

Standardization: A standard scaler was applied to standardize features, aligning the scales across modalities. Standardization is a crucial step when combining information from different modalities.

Classification Algorithm: A logistic regression model was used as the final classification algorithm. It takes advantage of the standardized feature set to provide a robust and interpretable framework for stress analysis.

The Audio, Image Processing, and Text Processing components each have their feature extraction methods and provide individual predictions. The Late Fusion Model combines these predictions by concatenating features and standardizing them. The Final Predictions are the result of combining information from the Audio, Image, and Text models.

5 Implementation

In this research project, the late fusion model was implemented, by bringing together the trained individual models for audio, image, and text data. The primary outputs of this implementation phase include, the late fusion model generates a transformed feature set, a representation of information from different modalities. Features from the audio, image, and

text models were concatenated and standardized, forming the basis for the final stress analysis.

Individual classification models for audio, image, and text data were developed and fine-tuned during the implementation phase. Specific algorithms for each dataset were chosen, these played a crucial role in contributing predictions to the late fusion model.

The late fusion model was trained by integrating predictions from audio, image, and text models using the standardized feature set. This model played a crucial role in stress analysis in social media content.

Toolset and Languages:

Some tools such as scikit-learn, TensorFlow, and joblib were used. Scikit-learn was used for the development and fine-tuning of machine learning models, TensorFlow was used to create the convolutional neural network for image processing, and joblib was used for handling model persistence.

Python programming was used as the primary language for coding and implementing the late fusion model. Its vast libraries and frameworks provided a strong and reliable environment for machine learning projects.

Throughout the implementation, the main focus was on efficiency and compatibility. The selected tools and languages ensured smooth integration and execution of the late fusion model, with outputs adaptable across various platforms and environments.

Steps for Reproducibility and Hyperparameter Optimization:

The experiments were conducted through multiple iterations, systematically adjusting architecture, hyperparameters, and data preprocessing. Detailed documentation, version control, and random seed setting were employed to ensure reproducibility. For the Late Fusion Model, saved models with consistent initialization and dependencies were utilized, allowing for the replication of the exact training conditions and preprocessing steps during inference.

The optimal hyperparameters were selected through a manual approach due to practical considerations. Automated searches like grid search or random search were deemed time-consuming and resource-intensive given the dataset's size and complexity. The manual tuning process involved an iterative exploration, starting with default values and making incremental adjustments based on real-time model performance, balancing trade-offs between effectiveness and computational efficiency.

6 Results and discussion

Evaluation metrics, including accuracy, precision, recall, and F1-score were used to assess the performance of individual models and the late fusion model. The results were compared with previous works in stress analysis and existing methodologies in each modality. This methodology ensures a systematic and structured approach to extracting meaningful insights from diverse multimodal datasets.

6.1 Text Model Evaluation

	precision	recall	f1-score	support
high stressed	0.83	0.82	0.83	39986
low stressed	0.82	0.83	0.83	40014
accuracy			0.83	80000
macro avg	0.83	0.83	0.83	80000
weighted avg	0.83	0.83	0.83	80000

Figure 6: Classification Report

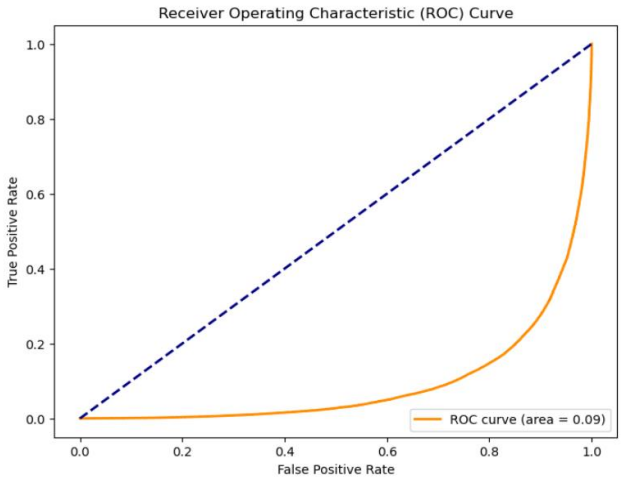

Figure 6: Text ROC curve

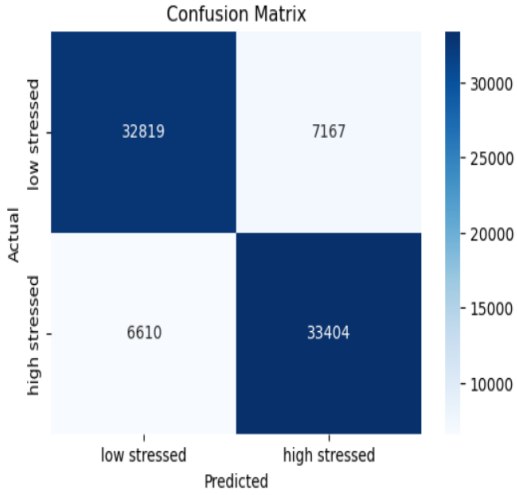

Figure 7: Text Confusion Matrix

The Logistic Regression model for the text data shows good performance achieving an overall accuracy of 83% and well-balanced precision, recall, and F1 scores for both low-stressed and high-stressed classes. A detailed true positives, true negatives, false positives, and false negatives is provided by the confusion matrix. The model's ability to discriminate between classes is illustrated by the ROC curve, and the AUC value indicates the effectiveness of the model in distinguishing between high-stressed and low-stressed instances. In general, the text model shows good performance in stress classification.

6.2 Image Model Evaluation

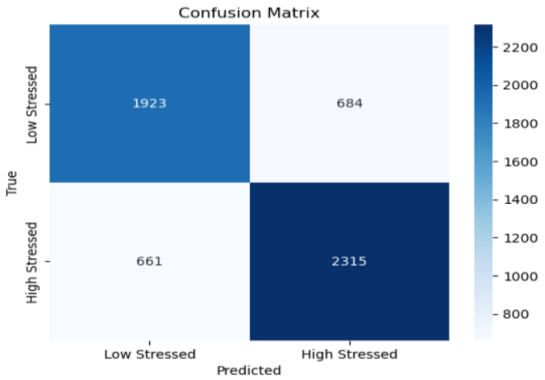

Figure 7: Image Confusion Matrix

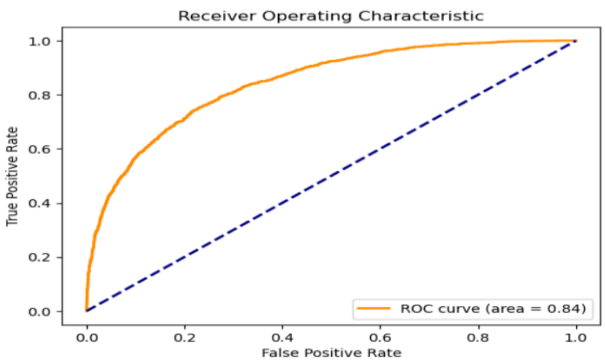

Figure 8: Image ROC curve

	precision	recall	f1-score	support
0	0.74	0.74	0.74	2607
1	0.77	0.78	0.77	2976
accuracy			0.76	5583
macro avg	0.76	0.76	0.76	5583
weighted avg	0.76	0.76	0.76	5583

Figure 9: Image Classification Report

A balanced performance is demonstrated by the image model, achieving an overall accuracy of 76%. Precision, recall, and F1 scores are consistent for both low-stressed and high-stressed classes. The model's ability to discriminate between classes is illustrated by the ROC curve, with an AUC of 0.82. A detailed breakdown of true positives, true negatives, false positives, and false negatives is provided by the confusion matrix, offering insights into the model's performance across stress levels. In general, the image model is doing well in stress classification based on the provided metrics.

6.3 Audio Model Evaluation

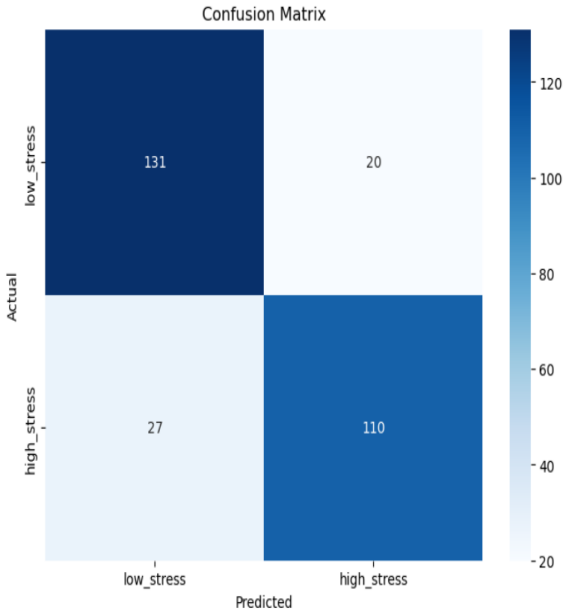

Figure 10: Audio Confusion Matrix

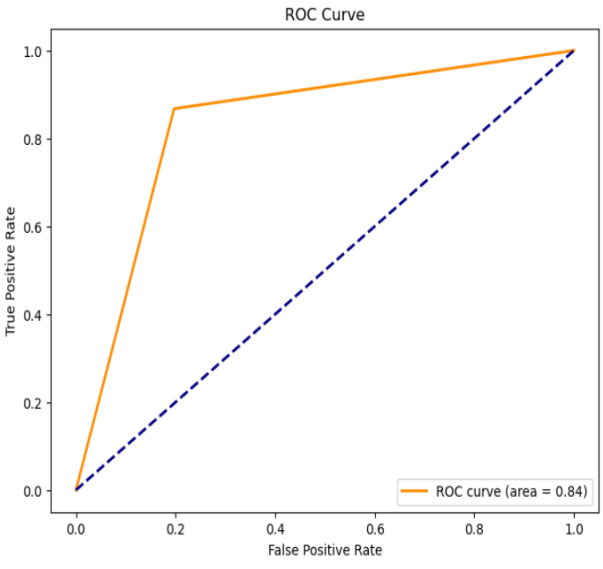

Figure 11: Audio ROC curve

```
Classification Report:
                precision    recall  f1-score   support

  high_stress       0.83      0.87      0.85       151
   low_stress       0.85      0.80      0.82       137

     accuracy                           0.84       288
    macro avg       0.84      0.84      0.84       288
 weighted avg       0.84      0.84      0.84       288

Accuracy: 0.8368055555555556
```

Figure 12: Audio Classification Report

The Audio model's overall accuracy of 84% demonstrates a high level of performance. Both low-stress and high-stress classes show a well-balanced combination of precision and recall. The trade-off between the true positive rate and the false positive rate is indicated by the ROC curve, and the model's discriminatory power is illustrated by the AUC. Overall, it can be observed that the audio model seems to be effective in distinguishing between low-stress and high-stress classes.

6.4 Late Fusion Model Evaluation

```
Binary Classification Model Performance:
              precision    recall  f1-score   support

           0       0.86      0.75      0.80       159
           1       0.73      0.85      0.79       129

    accuracy                           0.80       288
   macro avg       0.80      0.80      0.79       288
weighted avg       0.80      0.80      0.80       288
```

Figure 13: Late fusion classification report

A good balance between precision and recall for both classes is achieved by the late fusion model, with an overall accuracy of 80%.

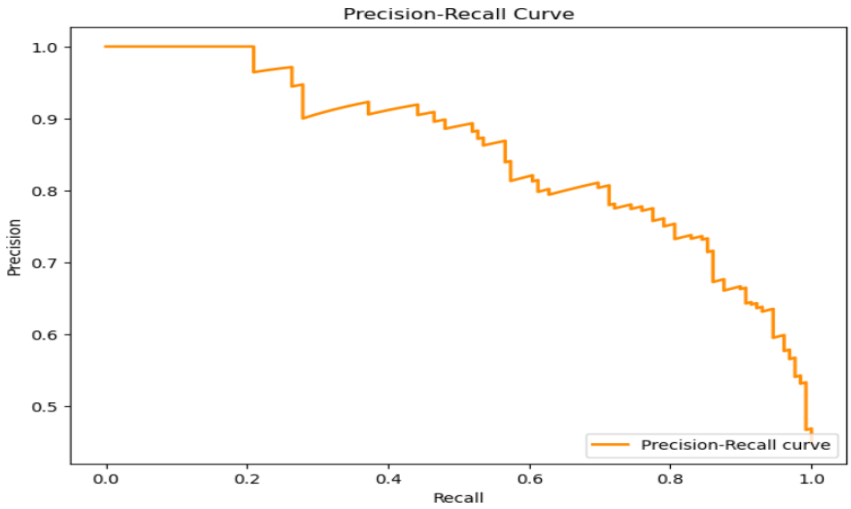

Figure 14: Late fusion Precision-Recall curve

The trade-off between precision and recall at different thresholds is illustrated by the precision-recall curve. The model's ability to balance precision and recall is measured by the area under the curve (AUC-PR: 0.851) .

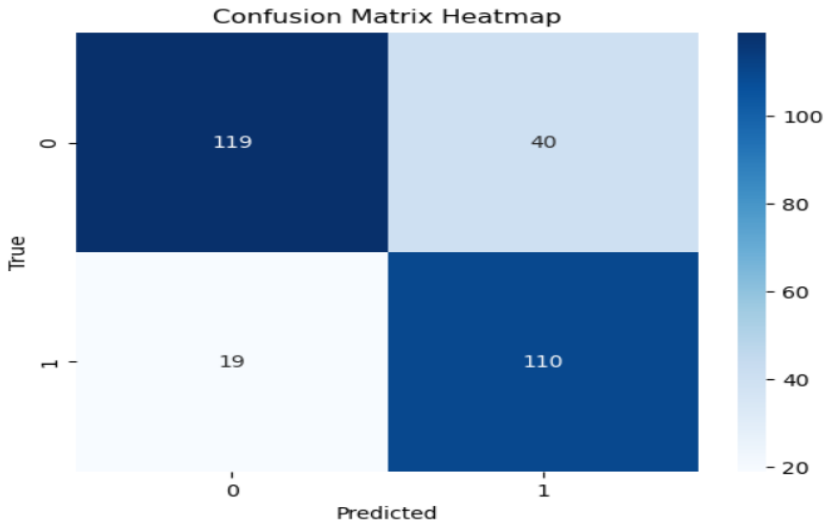

Figure 15: Late fusion Confusion Matrix

The confusion matrix provides a visual representation of the model's performance, showing the true positives, true negatives, false positives, and false negatives.

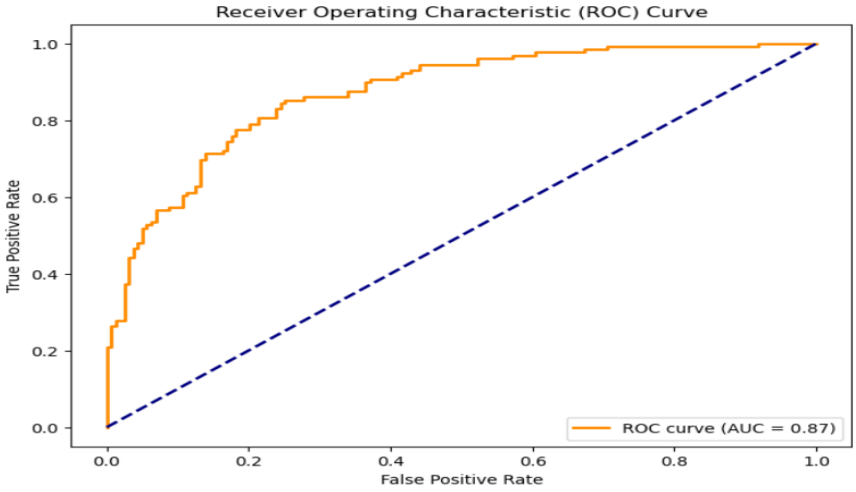

Figure 16: Late fusion ROC curve

The model's ability to discriminate between positive and negative instances is evaluated by the ROC curve. The model's discriminatory power is indicated by the area under the ROC curve (AUC-ROC: 0.871).

A Log -Loss of 0.4533 was obtained, where Logarithmic loss helps to measure the performance of the classification model and the prediction is a probability value. A lower value indicates better performance. A Brier score of 0.1462, helps in assessing the accuracy of probabilistic predictions, and a lower score indicates better calibration and accuracy.

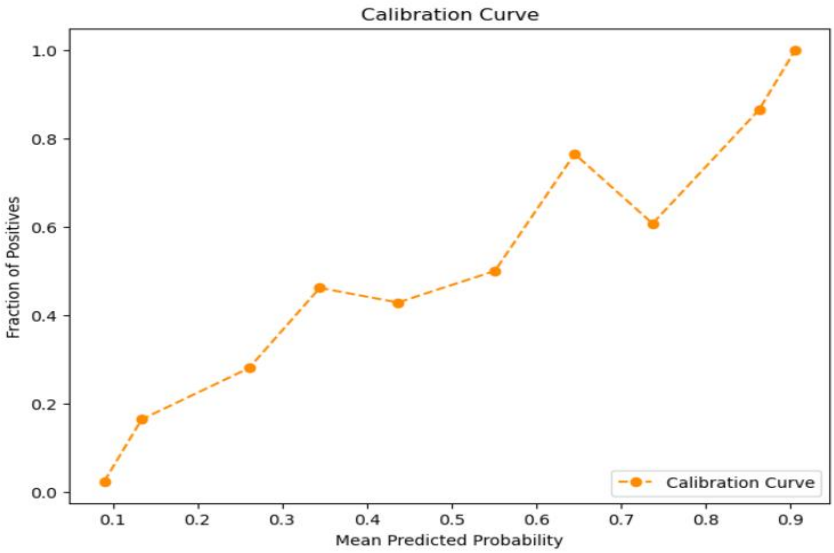

Figure 17: Calibration Curve

The model's predicted probabilities are assessed against the actual outcomes by the calibration curve. Overall based on the following results, the classifying the stress levels by model is observed to be performed well.

In the text model evaluation, the Logistic Regression model demonstrated commendable performance with an overall accuracy of 83%. The balanced precision, recall, and F1 scores for both low-stressed and high-stressed classes indicate the model's effectiveness in stress classification. The ROC curve and AUC value further affirm the model's discriminatory power.

Moving to the image model, a balanced performance was observed with an overall accuracy of 76%. Precision, recall, and F1 scores remained consistent for both stress classes, and the ROC curve illustrated the model's ability to discriminate between instances with an AUC of 0.82.

The audio model showcased high performance with an overall accuracy of 84%, demonstrating a well-balanced combination of precision and recall for both stress classes. The ROC curve and AUC provided additional evidence of the model's discriminatory power.

In the late fusion model evaluation, a good balance between precision and recall was achieved, resulting in an overall accuracy of 80%. The precision-recall curve showcased the model's ability to maintain a trade-off between these metrics at different thresholds, with an AUC-PR of 0.851. The ROC curve, with an AUC-ROC of 0.871, further emphasized the model's discriminatory power.

Notably, the Log-Loss of 0.4533 and Brier score of 0.1462 contribute valuable insights into the probabilistic predictions, indicating a lower value for better model performance. The calibration curve visually represented the model's predicted probabilities against actual outcomes, affirming the reliability of the stress level classifications.

Combining the strengths of the text, image, and audio models in the late fusion approach proved effective, resulting in a more accurate and robust classification of stress levels. The synergistic integration of information from multiple modalities enhanced the overall model performance, showcasing the potential of the late fusion model in achieving superior stress classification accuracy.

6.5 Discussion

The integration of audio, image, and text data for stress analysis is incorporated in this project. individual models for each dataset have been developed by us and successfully integrated them using a late fusion model. This approach extends the scope beyond singular analysis, providing a better understanding of stress. Integrating text, audio, and image data for the late fusion model will have a few challenges in terms of feature compatibility and standardization. Ensuring that predictions from individual models were synchronized and combining them effectively required careful consideration of each modality's unique characteristics. Despite all these challenges, the late fusion model successfully provided a consistent analysis of stress. The strengths of our approach lie in the fusion methodologies and the use of comprehensive performance metrics. However, the model tends to be overcautious, impacting specificity. This cautious nature is one of the limitations of the late fusion model that will be addressed in future iterations. Additionally, the current models' performance might be influenced by the specific characteristics of the dataset used. The developed models exhibit good generalizability. The multimodal fusion approach can be adapted to various scenarios requiring a detailed understanding derived from diverse data types. The confidence in our stress analysis results is the high accuracy of the model with 80%, and an 80% weighted average F1-score and precision values of 86% and 73% for non-stress and stress categories, respectively. This corresponds well with existing benchmarks in stress analysis and is supported by medical literature. The validity is further strengthened by the comprehensive evaluation metrics.

7 Conclusion and Future Work

The primary research question addressed in this research on stress analysis was whether a thorough classification of stress levels on social media content could be offered through the integration of text, audio, and image data can machine learning methods. The objectives were to develop individual models for text, audio, and image stress analysis, then use those three models and then integrate them into a late fusion model and evaluate the performance using suitable evaluation metrics. The work in this project includes training models on their respective datasets, ensuring feature compatibility, and implementing the late fusion methodology. This research work has been successful in answering the research question. The late fusion model shows a classification accuracy of 80% and an 80% weighted average F1-score along with precision values of 86% and 73% for non-stress and stress categories, Highlighting the effectiveness of combining information from multiple modalities for sophisticated stress classification in social media content. The successful integration of text, audio, and image data for stress analysis is included in the key findings of this project, decent performance metrics achieved comparable to existing benchmarks. Improved accuracy in terms of classifying the stress can be achieved because of this integration. However, the model tends to be overcautious, and there is a need for further improvement.

Future work will be focused on improving the accuracy of the late fusion model and then addressing overcautious predictions and exploring ways to enhance specificity without

compromising sensitivity. Further commercialization lies in deploying the model as a stress analysis tool. Also incorporating automated searches such as grid search or random search could be explored to further enhance model accuracy by systematically exploring a broader range of hyperparameter combinations.

References

Mounika, S. N., Kumar Kanumuri, P., K. N. Rao, and Manne, S., 2019. "Detection of Stress Levels in Students using Social Media Feed." In: 2019 International Conference on Intelligent Computing and Control Systems (ICCS). Madurai, India: IEEE, pp.1178–1183.

Giuntini, F. T., et al., 2021. "Tracing the Emotional Roadmap of Depressive Users on Social Media Through Sequential Pattern Mining." IEEE Access, 9, pp.97621–97635.

Tajuddin, M., Kabeer, M. and Misbahuddin, M., 2020. "Analysis of Social Media for Psychological Stress Detection using Ontologies." In: 2020 Fourth International Conference on Inventive Systems and Control (ICISC). Coimbatore, India: IEEE, pp.181–185.

Shaw, B., Saha, S., Mishra, S. K. and Ghosh, A., 2022. "Investigations in Psychological Stress Detection from Social Media Text using Deep Architectures." In: 2022 26th International Conference on Pattern Recognition (ICPR). Montreal, QC, Canada: IEEE, pp.1614–1620.

Joshi, M., Mahajan, C., Korgaonkar, T., Raul, N. and Naik, M., 2023. "Mental Health Analysis using Deep Learning of Social Media Data gathered using Chrome Extension." In: 2023 4th International Conference for Emerging Technology (INCET). Belgaum, India: IEEE, pp.1–8.

Cacciatori, F., Nikolaev, S., Grigorev, D. and Archangelskaya, A., 2023. "On Developing Facial Stress Analysis and Expression Recognition Platform." In: 2023 International Conference on Artificial Intelligence Science and Applications in Industry and Society (CAISAIS). Galala, Egypt: IEEE, pp.1–6.

Mehta, D., Siddiqui, M. F. H. and Javaid, A. Y., 2019. "Recognition of Emotion Intensities Using Machine Learning Algorithms: A Comparative Study." Sensors, 19(8), p.1897.

Upadhyay, V. and Kotak, D., 2020. "A Review on Different Facial Feature Extraction Methods for Face Emotions Recognition System." In: 2020 Fourth International Conference on Inventive Systems and Control (ICISC). IEEE, pp.15–19.

Abburi, H., Shrivastava, M. and Gangashetty, S. V., 2016. "Improved multimodal sentiment detection using stressed regions of audio." In: 2016 IEEE Region 10 Conference (TENCON). Singapore: IEEE, pp.2834–2837.

Aggarwal, V., Kaur, H., Sharma, D. and Singhal, A., 2023. "Emotion Classification of Social Media Posts using Artificial Intelligence and Machine Learning." In: 2023 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES). Greater Noida, India: IEEE, pp.999–1004.

Kulatilake, T. T., et al., 2022. "PRODEP: Smart Social Media Procrastination and Depression Tracker." In: 2022 17th International Workshop on Semantic and Social Media Adaptation & Personalization (SMAP). Corfu, Greece: IEEE, pp.1–6.

Rastogi, A., Liu, Q. and Cambria, E., 2022. "Stress Detection from Social Media Articles: New Dataset Benchmark and Analytical Study." In: 2022 International Joint Conference on Neural Networks (IJCNN). Padua, Italy: IEEE, pp.1–8.

Rajendar, S., et al., 2022. "An Extensive Survey on Recent Advancements in Human Stress Level Detection Systems." In: 2022 6th International Conference on Electronics, Communication and Aerospace Technology. Coimbatore, India: IEEE, pp.1550–1554.

Selvadass, S., Bruntha, P. M. and Priyadharsini, K., 2022. "Stress Analysis in Social Media using ML Algorithms." In: 2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT). Tirunelveli, India: IEEE, pp.1502–1506.

S. Jadhav, A. Machale, P. Mharnur, P. Munot, and S. Math (2019), "Text-Based Stress Detection Techniques Analysis Using Social Media," 2019 5th International Conference On Computing, Communication, Control And Automation (ICCUBEA), Pune, India, pp. 1-5. doi: 10.1109/ICCUBEA47591.2019.9129201.