FREEDOM

EFFECTIVE EARLY DEPRESSION DETECTION THROUGH ONLINE MEDIA POSTS

(GROUP NO - 8)

AKSHARA NAIR - MT22008
AYUSH AGARWAL - MT22095
MEDHA - MT22110
MOHIT GUPTA - MT22112
NIDHI VERMA - MT22044
PRATIK CHAUHAN - MT22118

OVERVIEW

- PROBLEM STATEMENT
- MOTIVATION & CHALLENGES
- RELATED WORK & LIMITATIONS
- PROPOSED SOLUTION
  - SOLUTION
  - NOVELTY (DATASET AND PROCEDURE)
  - IMPACT
- CONCLUSION AND FUTURE SCOPE

THE AIM OF THE PROPOSED RESEARCH-BASED PROJECT IS THE DEVELOPMENT OF ARTIFICIAL INTELLIGENCE (AI) BASED MATHEMATICAL MODELS TO ANALYZE UNI-MODAL AND MULTI-MODAL ASPECTS OF THE INPUT DATA TO PREDICT DEPRESSION.

PROBLEM STATEMENT

There exists an emergent need to monitor if the content being posted or searched by an individual is indicative of the existence of early symptoms of depression.

Early and effective analysis and prediction through pre-learned representations of depression by visual or textual data form the basis of the problem of the proposed research-based project.

The presented work deals with executing and analyzing distinct techniques to identify the emotion of depression in uni-modal and multi-modal input as images and text and presenting the best architecture to identify the image or text which strongly emotes the identified or learned representations of the conveyed sentiment.

MOTIVATION & CHALLENGES

Motivation

1. It is uncertain whether existing data sets accurately capture the dynamic relationships between visual and textual media on the internet in the context of depression.
2. Visual representations and corresponding textual representations of underlying emotion of depression may differ across images posted by potentially depressed individual, so diverse image samples are necessary for accurate research.

Challenges

Social media often being noisy and unstructured contains high amount of irrelevant and redundant information. This make it difficult to extract features and patterns to extract features.

[1]: Masud, Mohammed T., et al. "Unobtrusive monitoring of behavior and movement patterns to detect clinical depression severity level via smartphone." Journal of biomedical informatics 103 (2020): 103371.
[2]: Fukazawa, Yusuke, et al. "Predicting anxiety state using smartphone-based passive sensing." Journal of biomedical informatics 93 (2019): 103151.
[3]: Wang, Qingxiang, Huanxin Yang, and Yanhong Yu. "Facial expression video analysis for depression detection in Chinese patients." Journal of Visual Communication and Image Representation 57 (2018): 228-233.
[4]: Williamson, James R., et al. "Vocal and facial biomarkers of depression based on motor incoordination and timing." Proceedings of the 4th international workshop on audio/visual emotion challenge. 2014.

RELATED WORK

01

MASUD ET AL.

Phone sensors like acceleration and GPS were used to classify physical activities and geographic movements. The SVM classifier distinguished between three depression severity levels with 87.2% accuracy.

02

FUKAZAWA ET AL.

Mobile phone sensors provided data on app usage, brightness, acceleration, rotation, and orientation, which were combined to create higher-level feature vectors. The fusion of these vectors accurately predicted the subjects' stress levels.

03

WANG Q ET AL.

Facial cue changes were measured using a person-specific active appearance model to detect 68-point landmarks, and statistical features were extracted from these landmarks to feed the SVM classifier. The classifier achieved a 78% test accuracy.

04

WILLIAMSON ET AL.

The study used facial movement and acoustic verbal signal feature sets, which were reduced in dimensionality using principal component analysis. The combination of primary feature vectors was categorized using the Gaussian mixture model.

[5]: Lin, Chenhao, et al. "Sensemood: depression detection on social media." Proceedings of the 2020 international conference on multimedia retrieval. 2020.
[6]: Chiong, Raymond, et al. "A textual-based featuring approach for depression detection using machine learning classifiers and social media texts." Computers in Biology and Medicine 135 (2021): 104499.
[7]: Zogan, Hamad, et al. "DepressionNet: A novel summarization boosted deep framework for depression detection on social media." arXiv preprint arXiv:2105.10878 (2021).
[8]: Gui, Tao, et al. "Cooperative multimodal approach to depression detection in twitter." Proceedings of the AAAI conference on artificial intelligence. Vol. 33. No. 01. 2019.

RELATED WORKS

05

LIN, CHENHAO, ET AL

The paper discusses a system that uses natural language processing and machine learning to detect depression from social media posts. It analyzes various linguistic and behavioral patterns in the posts to identify individuals who may be at risk of depression.

06

CHIONG ET AL.

This paper discusses about a method for detecting depression using machine learning classifiers and social media texts. The approach is based on textual features extracted from social media posts, and the results show that the method has a high accuracy in detecting depression.

07

ZOGAN ET AL.

The paper presents a novel deep learning framework called DepressionNet, which uses summarization techniques to detect depression on social media. The framework shows promising results in detecting depression and can be useful for early intervention and support.

08

GUI, TAO ET AL.

The paper proposes a cooperative multimodal approach to detect depression in Twitter by combining text and image features. The approach achieves high accuracy in depression detection and can be useful for early identification and intervention of depression.

ABOUT DATASET

Dataset Image-1 and Image-2

- Dataset created by **scraping images** and text from the web using **BeautifulSoup** in Python.
- Images obtained from sources such as **ShutterStock.**
- **For Image-1**
  - **Total count of** Image samples are **2473**
  - The count of samples **in Train, Validation and Test Set** is as **1583 Images, 395 Images and 495 images** for two classes.
  - **Only images data is scrapped without considering the relation with caption(text), we named this dataset as Image-1.**

- **For Image-2**
  - **Total count of Image samples are 5475.**
  - The samples in **Train, Validation and Test Set are 4380 Images, 547 Images and 548 Images** for two classes.
  - **Direct association between text(caption) and images in the dataset** but here we are **utilizing only images for our unimodal training** and we named the data as Image-2

ABOUT DATASET

Dataset Text - 1 and Text - 2

- Text obtained from Reddit subreddits using Beautiful Soup from python.
- Used tags like depression, sadness etc. for depressive and joy, happy etc. for non-depressive posts.

Dataset Text - 1:

No direct association between text and images in the dataset which is scrapped. Total count of text 'depression' was relatively low, hence we add additional 'suicide samples' from 'Suicide Watch' to counteract the imbalanced dataset.

For Dataset Text - 2:

This refers to the text-only samples taken from the multimodal dataset containing image-caption pairs corresponding to the same post scrapped from Reddit.

ABOUT DATASET

Dataset DF

- Scraped **visuals and corresponding text from the same source**, where **text is the description of the image.**
- Advantage of the dataset is the ability to incorporate and **learn the relation between image and text.**
- **In total we have extracted 5475 images, and their corresponding description.**
- **Appropriate training is intended to produce effective results**

Dataset for Inference

- **20** samples of text as well as images from real posts are collected where the image is associated with the text is used for inference which is fed to unimodal and multimodal setup to get the results.

PROPOSED SOLUTION

Two main approaches: **Unimodal and Multimodal Deep Learning based Architectures.**

Unimodal

Models implemented for Image classification: **CNN, ResNet50, VGG19, InceptionV3, & Xception.**

Models Implemented for text classification:**Bi-LSTM, SVM using non-contextual and contextual embeddings of fasttext and fined-tuned BERT model.**

Multimodal

Three approaches used**:**

- **ResNet-50+Distill-BERT**
- **VGG-16 +BERT**
- **Inception v3+BERT**

SOLUTION

- Novelties in the **data set and methodologies** such as exploration of both **unimodal and multimodal** .
- **Utilization of related images and captions** for classification of new posts as depressed or not depressed.
- Potential enhancement of model's ability to **generalize well to test data** on social media when trained on **scraped and fused data from the web**.

Our proposed model also gives **personalized experience** to user.

NOVELTY

As we move from unimodal to multimodal approach, we observed that:

- **Improved Performance:**
- **Enhanced Robustness**
- **Enhanced Interpretability**

.

IMPACT

UNI-MODAL CLASSIFICATION

Image Classification

- **Pre-processing techniques used** with specified parameter values: Rescale=1./255 & shear range=0.2, zoom range=0.2 and horizontal flip=True to avoid overfit.
- Pre-trained models used: **ResNet50, VGG19, InceptionV3, and Xception.**
- Learning rate set as 0.01.
- Early Stopping has been found in **VGG19, Inception and Xception** model, after **27, 18, and 14** epochs respectively
- SGD (Stochastic Gradient Descent) used as optimizer.
- Loss function used is Binary Cross-Entropy.

RESULT FOR TEST SET (IMAGE)

UNI-MODAL

TABLE III: Performance Metrics for Testing Set of DI-1 for Image Classification

Models & Metrics	Accuracy	Precision	Recall	F-1 Score
CNN	0.40	0.40	1.00	0.57
ResNet50 [9]	0.71	0.61	0.78	0.67
VGG-19 [10]	0.85*	0.75	0.93*	0.83*
InceptionV3 [11]	0.84**	0.79**	0.82**	0.79**
Xception [12]	0.84**	0.85*	0.73	0.78

TABLE IV: Performance Metrics for Testing Set of DI-2 for Image Classification

Models & Metrics	Accuracy	Precision	Recall	F-1 Score
ResNet50 [9]	0.70	0.68	0.87*	0.77
VGG-19 [10]	0.73	0.87*	0.63	0.72
InceptionV3 [11]	0.82*	0.83**	0.86**	0.85*
Xception [12]	0.80**	0.81	0.84	0.82**

UNI-MODAL CLASSIFICATION

Text Classification:

Three distinct techniques were employed for processing and classification using various textual embeddings and model combinations -

- **FastText embeddings** used to produce **subword embeddings** for c**apturing meaning of complex morphological and out-of-vocabulary (OOV) words**. **SVM classifier used for training**
- **Pre-trained BERT model embeddings** used for **contextualized information**.SVM classifier used on BERT embeddings.
- **Bert-based-cased' tokenizer** used to **tokenize input samples for BERT model**.**Fine-tuning of pre-trained BERT model** on specific task of depression classification to leverage its ability to **capture contextual information.**

RESULTS FOR TEST SET(TEXT)

UNI-MODAL

TABLE V: Performance Metrics for Test Set of Text-1 for Text Classification

Models & Metrics	Accuracy	Precision	Recall	F-1 Score
BI-LSTM (Baseline)	0.92	0.89	0.93	0.92
FastText + SVM	0.96	0.97	0.92	0.95
BERT-Embeddings + SVM	0.96	0.95	0.95	0.95
BERT-Fined Tuned Model	0.98	0.98	0.98	0.98

TABLE VI: Performance Metrics for Test Set of Text-2 for Text Classification

Models & Metrics	Accuracy	Precision	Recall	F-1 Score
FastText+SVM	0.83	0.80	0.88	0.84
BERT-Embeddings+SVM	0.87	0.99	0.82	0.0.90
BERT-Fined Tuned Model	0.95	0.95	0.95	0.95

MULTI-MODAL CLASSIFICATION

Both Image and Text into consideration for predicting depression. Train, test, and validation split used in all the three setups are same which is 80:10:10.
Hyper-Parameters used: lr=1e-5, num_epochs = 10, dropout=0.5

Setup-1

ResNet-50 for image and Distill-BERT for text.

Setup-2

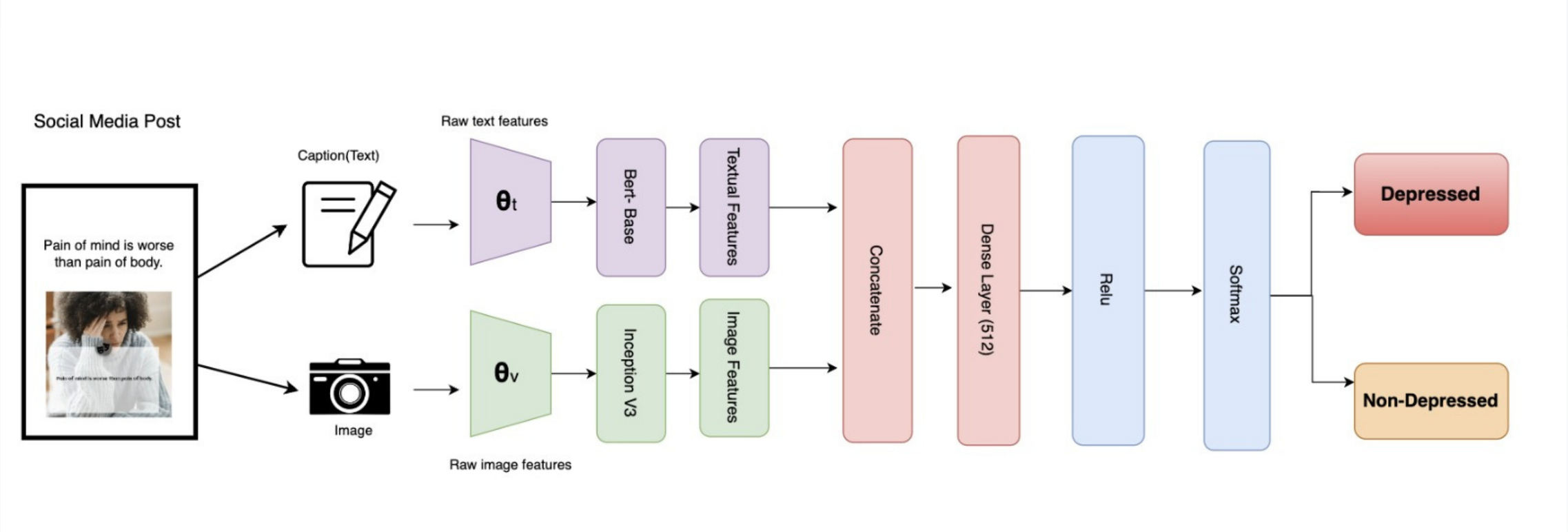
VGG-16 for image and BERT for text.

Setup-3

InceptionV3 for image and BERT for text.

MULTI-MODAL ARCHITECTURE

Both image and text into consideration for predicting depression



RESULTS FOR TEST SET (IMAGE+TEXT)

MULTIMODAL

TABLE VII: Performance Metrics for Multi-modal Setup on Test Set

Models & Metrics	Accuracy	Precision	Recall	F-1 Score
Setup-1	96.53	97.0	96.0	96.0
Setup-2	96.71	97.0	97.0	97.0
Setup-3	96.89	97.0	97.0	97.0

RESULTS FOR TEST SET (IMAGE+TEXT)

MULTIMODAL

TABLE VII: Performance Metrics for Multi-modal Setup on Test Set

Models & Metrics	Accuracy	Precision	Recall	F-1 Score
Setup-1	96.53	97.0	96.0	96.0
Setup-2	96.71	97.0	97.0	97.0
Setup-3	96.89	97.0	97.0	97.0

RESULT FOR INFERENCE

TABLE VIII: Performance Metrics for Inference Data

Models & Metrics	Accuracy	Precision	Recall	F-1 Score
Image Classification	0.40	0.40	0.40	0.39
Text Classification	0.45	0.66	0.45	0.52
Multimodal Classification	0.75	0.75	0.77	0.74

75%

METRIC VISUALIZATION

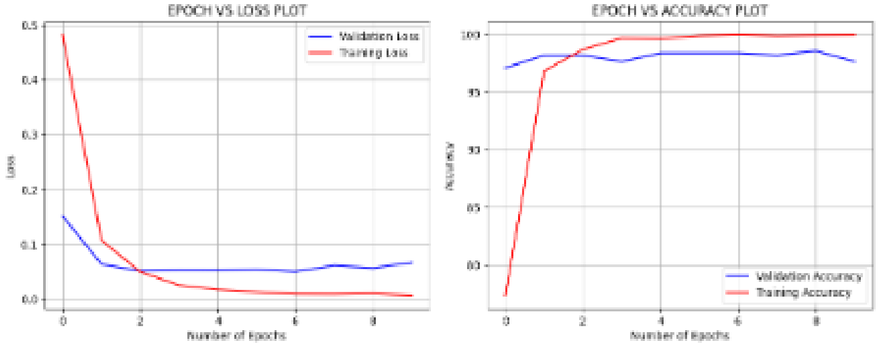


Fig. 9: Plot of Loss and Accuracy for Setup-1

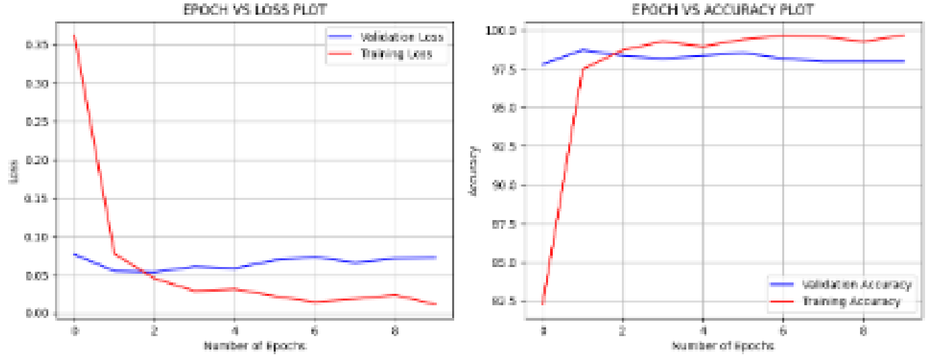


Fig. 10: Plot of Loss and Accuracy for Setup-2

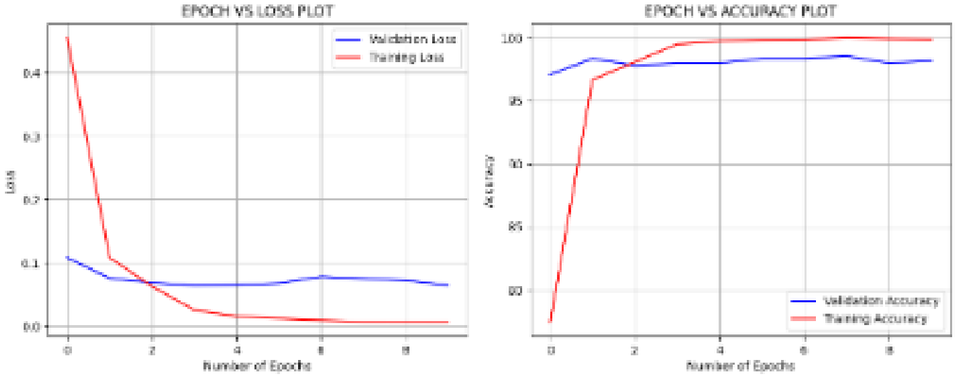
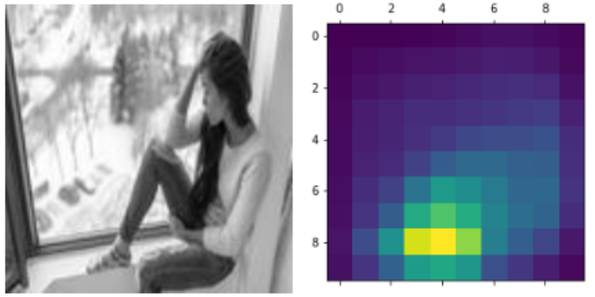


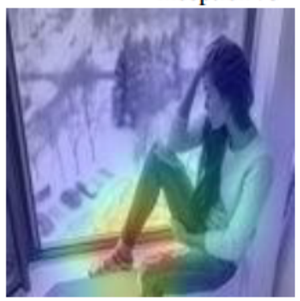
Fig. 11: Plot of Loss and Accuracy for Setup-3

EXPLAINABILITY



(a) Original Image (Label - Depressed) in the Dataset Image1

(b) Gradient Heat-Map for Original Image as Input to InceptionV3

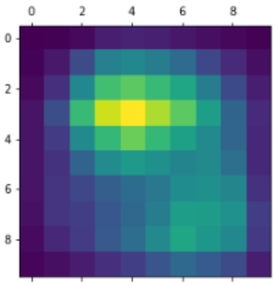


(c) Overlapped Image for visualizing learned features in an Input Image

Fig. 14: GRAD-CAM as a tool for Explainability

- **GRAD-CAM (Gradient-weighted Class Activation Mapping)** is a technique used for visualizing the activation regions during image classification tasks.
- Provides a heatmap overlay highlighting the areas that deems important for making its classification decision.
- Here in both the image of depressed as well as non-depressed, **GRAD-CAM** finds the important regions of images that helps in classifying the image as depressed or non-depressed .

(a) Original Image (Label - Not Depressed) in the Dataset Image1

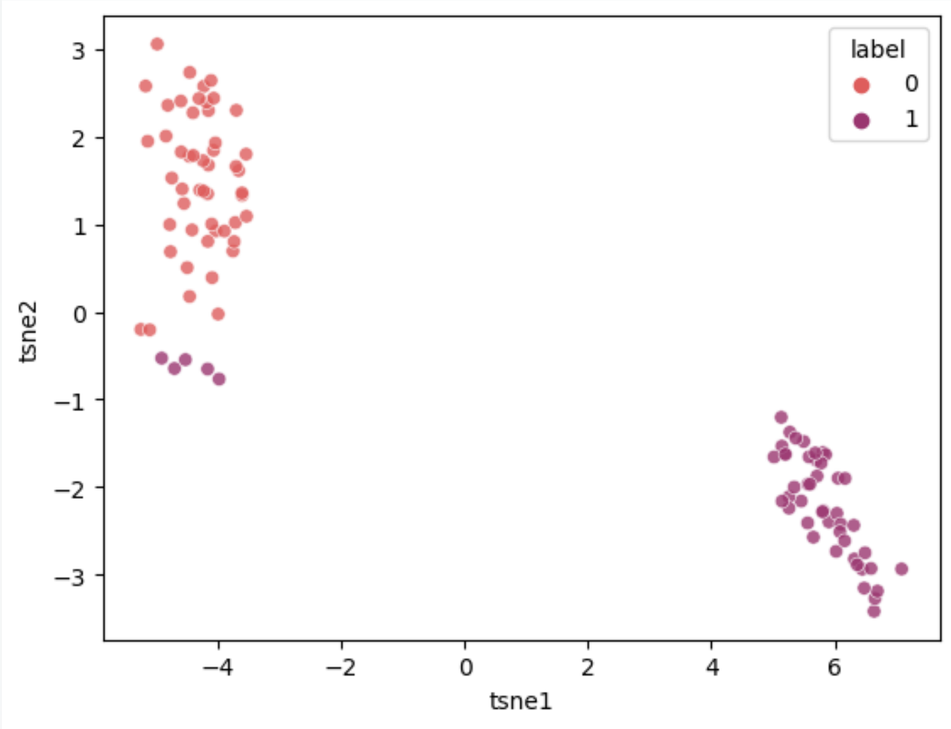


(b) Gradient Heat-Map for Original Image as Input to InceptionV3

(c) Overlapped Image for visualizing learned features in an Input Image

Fig. 15: GRAD-CAM as a tool for Explainability

EXPLAINABILITY



The **Scatter Plot** is able to visulaise the high dimensional features by creating a plot of the features being reduced to a 2-D The plots allow us to separate data that cannot be separated by any straight line, The probabilities of a data point to pick up the most probable neighbour, leads to the formation of two distinct clusters here, which ensures that the features generated for two data classes, are easily distinguishable when further processed.

CONCLUSION

The project aims to detect depression on social media and web pages which contain hidden and multimodal semantics to indicate depression.

The proposed novel, multimodal dataset prepared from scrapping online media posts , significantly helped prepare robust model to capture depression on the internet.

The exploratory observations across three modalities led to the proposed multimodal deep fusion based model architecture using BERT-based and Inception-based embeddings which generalised well to classify posts for depression.

FUTURE SCOPE

- Dataset quality can be improved by adding more diversified data from social media platforms like LinkedIn and Instagram as incorporating different causes of depression can improve the performance of the model.

- Modalities such as audio, video, or physiological data can also be used to improve the model's performance.

- Other fusion strategies such as attention-based fusion, cross-modal retrieval-based fusion, or late fusion can be explored to further improve the performance of the model.

- Deploying the model on social media platforms can help detect early signs of depression in users' posts and recommend mental health resources.

THANK YOU