FREEDOM : Effective Early Depression Detection through Online Media Posts

Akshara Nair*, Ayush Agarwal†, Medha‡, Mohit Gupta§, Nidhi Verma¶, Pratik Chauhan‖
Dept. of Computer Science & Engineering
Indraprastha Institute of Information Technology Delhi, Delhi, India
akshara22008@iiitd.ac.in*, ayush22095@iiitd.ac.in†, medha22110@iiitd.ac.in‡, mohit22112@iiitd.ac.in§,
nidhi22044@iiitd.ac.in¶, pratik22118@iiitd.ac.in‖

***Abstract*—With an explosion in the availability of instant web-based services and enhancement in the pace of propagation of digital media content being published on an individual's social accounts, 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 aim of the proposed research-based project is the development and deployment of Artificial Intelligence (AI) based mathematical models to analyze uni-modal and multi-modal aspects of the input data to predict depression. The initial evaluation of baseline models involves scraping the data as well as the images and text from web platforms where user-based content is generated. The proposed research project suggests combining the predictions of distinct input modalities to provide efficient results on the media posts of an individual.**

***Index Terms*—Multimodal Classification, Natural Language Processing, Image Classification, Deep Learning, Depression Detection**

I. Problem Statement

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. Emotions such as anxiety, anger, fear, street, rage, and resentment are quite persistent, either directly or indirectly, and generally exist as an underlying emotion rather than being clear or dominant and hence are not easy to capture and predict in the content being published online. In addition, the problem is introduced through proposed solutions so far which utilize uni-modal textual or behavioral data-based physiological measures, which require in-depth domain knowledge and expertise. On the other hand, such procedures are prone to error due to human dishonesty while reporting emotions as sharing such emotions usually signifies the existence of negative stigma.

There is a need to capture such emotions efficiently without involving the individual oneself in the process and use the specified procedure for better ranking the content available on the web so as to ensure that people who travel through such web pages do not encounter an excessive amount of such content which triggers or enhances such negative emotions. The emotions, if identified, must be controlled and must not be frequent on the web, which might allow the user or any service provider to control the quality of the content being consumed.

II. Motivation

A. Enhancing the Procedural Pipeline

Even though there exists sufficient research in and around the domain, it is not possible to conclude the fact that the existent data set captures the dynamic relationships amongst the visual and textual media on the internet. The visual description of depression, as a set of features, is not necessarily similar across distinct images, posted by individuals who are all depressed. Hence, diversity across varied image samples must necessarily be included along with the goal of preserving the relevance between a visual and its textual description. There is a need to hence propose a complete solution to identify depression over online media and if the web page or post can be identified as depressive or non-depressive, it is extremely easy to incorporate the devised techniques while ranking the web pages or recommending relevant posts to the user.

B. Preliminary Diagnosis and Prediction

The procedure for appropriate clinical diagnosis of depression does not guarantee definite patient-based outcomes and results and involves a risk of inaccurate reporting, and treatment. It might or might not be possible for an individual to be self-aware of the fact and find the target help. In the due course of time, if any potential sufferer encounters a continuous stream of any such content on the web which triggers the emotion further, it may damage the mental health of the individual.

C. Objective Assessment

The procedure of accessing an individual's browsing activities or monitoring the content being scrolled would ultimately allow the implemented Deep Learning Techniques at the back end to process the real-time data and serve immediate predictions. The procedure will analyze the same and after a brief interval of time, it would be possible to analyze the behavior and pattern of searched web pages and scrolled posts.

D. *Personalized Results*

The proposed methods in the projects will be specific to user posts. In general, the publically available data set focuses only on particular kinds of images as facial features or postures whereas it is not necessary that an individual will always be able to infer a feeling of depression from such posts only. There is a need to train the Artificial Intelligence(AI) based model on a data set, which is directly relevant or related to the content being shared online.

III. Literature Review

Masud et al. [1] classified physical activities and geographic movement patterns using built-in phone sensors, such as the acceleration and Global Positioning System (GPS) sensors, respectively. A portion of the characteristics was chosen. With an accuracy of 87.2%, the SVM classifier was used to differentiate between the three different severity levels of depression (absence, moderate, and extreme)

Fukazawa et al. [2] gathered information from mobile phone sensors, including utilization of applications, brightness, acceleration, rotation, and orientation. They were combined by the creator to create higher-level feature vectors. The subjects' stress levels could be predicted using the fusions of these feature vectors.

Wang Q et al. [3] looked at how the facial cues of depressed and normal individuals changed in the same circumstance. (while displaying positive, neutral, and negative pictures). To measure the facial cue changes on the face, they used a person-specific active appearance model to detect 68-point landmarks. Statistical features are extracted from distances between feature points of the eyes, eyebrows, and corners of the mouth to feed the SVM classifier. The classifier achieved 78% test accuracy.

In order to identify psychomotor retardation, Williamson et al. [4] used feature sets drawn from facial movements and acoustic verbal signals. For dimensionality reduction, they used principal component analysis, and to categorize the combination of primary feature vectors, they used the Gaussian mixture model.

Chenhao Lin et al. [5] proposes a machine learning approach to detect depression among social media users based on their online behavior. The authors present a system called SenseMood that analyzes large datasets of user behavior on social media platforms, such as language use and engagement patterns, to identify users who exhibit symptoms of depression. The paper describes the various steps involved in developing the SenseMood system, including data collection and preprocessing, feature selection, and classification. The authors collected data from two social media platforms, Weibo and Twitter, and selected a set of features related to linguistic, semantic, and social aspects of user behavior. The authors used three different classification models to evaluate the performance of the SenseMood system: logistic regression, random forest, and support vector machines (SVM). They compared the results of these models to a baseline model that simply predicted the majority class. The results showed that the SenseMood system outperformed the baseline model and achieved high accuracy in detecting depression among social media users. The best performing model was SVM, which achieved an accuracy of 87.7% on the Weibo dataset and 82.8% on the Twitter dataset. The authors also conducted a feature analysis to identify the most informative features for detecting depression. They found that features related to emotional expressions, social network properties, and linguistic styles were the most important features for depression detection.

Raymond Chiong et al. [6] proposes a machine learning approach to detect depression based on text data from social media platforms. The authors developed a set of features related to linguistic and semantic aspects of user behavior and used machine learning classifiers to analyze large datasets of social media texts. The paper describes the various steps involved in developing the depression detection system, including data collection and preprocessing, feature selection, and classification. The authors collected data from two social media platforms, Twitter and Reddit, and selected a set of features related to sentiment, topic, and linguistic styles. The authors evaluated the performance of their depression detection system using several machine learning classifiers, including logistic regression, support vector machines (SVM), and random forests. The results showed that the SVM classifier outperformed the other classifiers and achieved an accuracy of 88.3% on the Twitter dataset and 84.7% on the Reddit dataset.

Hamad Zogan et al. [7] proposes a novel deep learning framework called DepressionNet for detecting depression on social media platforms. The framework uses a combination of convolutional neural networks (CNNs) and recurrent neural networks (RNNs) to analyze social media data and identify users who may be at risk of depression. The paper describes the various steps involved in developing the DepressionNet framework, including data preprocessing, feature extraction, and classification. The authors collected data from Twitter and Tumblr and used a summarization technique to reduce the dimensionality of the data and improve the performance of the model. The authors evaluated the performance of their DepressionNet framework using different performance metrics as accuracy, precision, recall and f1-Score with several machine learning classifiers, including SVM, logistic regression, random forest, XLNet, BERT, RoBERTa, BiGRU with attention and CNN with attention. The results showed that the DepressionNet framework outperformed the other classifiers and achieved an accuracy of 91.34% on the Twitter dataset and 87.63% on the Tumblr dataset. The authors also conducted a feature analysis to identify the most informative features for depression detection. They found that features related to emotion, social support, and linguistic styles were the most important features for depression detection.

Tao Gui et al. [8] proposes a cooperative multimodal approach for detecting depression in social media data, specifically Twitter. The authors used both text and image data to develop a deep learning model that can detect depression in users' tweets. The paper describes the various steps involved in developing the cooperative multimodal approach, including data collection and preprocessing, feature extraction, and classification. The authors collected data from Twitter and used

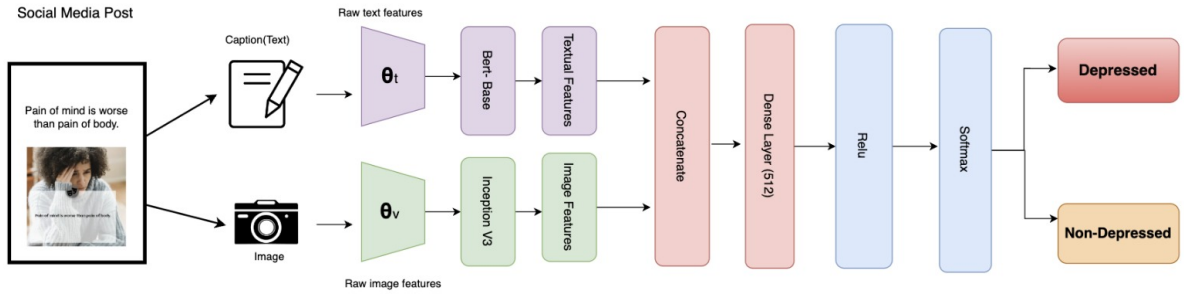


Fig. 1: Proposed Multi-Modal Architecture for predicting Depression in the given image and text for social media posts.

both text and image features for depression detection. The paper proposes a new method for detecting depression using a combination of textual and visual data from social media. The traditional diagnosis of depression requires a face-to-face conversation with a medical doctor, limiting the identification of potential patients. Therefore, social media data has become an important source for monitoring public health issues, including depression. The proposed method uses multi-agent reinforcement learning to cooperatively select indicator texts and images, which are more beneficial to the classifier. The experimental results demonstrate that the proposed method achieved better performance than existing methods by a large margin. The selected posts can indicate depression effectively. The results showed that the cooperative multimodal approach outperformed the other classifiers and achieved an accuracy of 88.9%. The authors also conducted a feature analysis to identify the most informative features for depression detection. They found that features related to emotion, social support, and image content were the most important features for depression detection.

IV. Novelty

The following novelties are being introduced to the data set, methodologies for comparative study, and result-based evaluation to allow the diverse but simplified application of the discussed research efficiently. To derive results for user-based media content, it is necessary to introduce and include such data, which is highly relevant to web-based content but is not specific to the user or feature and hence is easy to generalize. Hence the web pages are crawled to collect such data which is relevant to general images of depression. The introduced novelty is to explore all possible uni-modal and multi-modal architectures for using pre-trained models for generating immediate features for pronounced classification of the input image and text to the relevant class as depressed or not depressed. The final results are to find out whether the proposed model trained on the scraped and fused data is able to fit well on the actual content posted on social media. The discussed techniques also contribute to minimizing the probability of the model miss-classifying the actual content on web on the basis of irrelevant learning through the scraped data. In this way, the proposed novelty may enhance the ability of the model to generalize well to the test data on social media when trained on data scraped from the web itself and learn from related images and the corresponding captions for classifying a new post as depressed or not depressed.

V. Methodology

A. *Dataset*

The Dataset has been created by scraping the images and text from the web using Python library *BeautifulSoup*. Initially, the dataset consists of the scraped images from distinct sources **ShutterStock**, and text from **Reddit** for the purpose of uni-modal analysis. It is important to note here that there does not exist an associated relation between the texts and images. But for utilizing the research on the real-time content being pushed on the internet, there is a significant but not guaranteed relation of an inter-dependence between the sentiment extracted through the visuals posted online and the textual caption of the content being posted as online media.

1) ***Image1 and Image2****:* The initial pre-processing of the text includes the elimination of certain redundancies where we process extracted images and convert the data to a usable form. The following problems were identified where one is there is a strip of watermark on the Shutter-Stock website, and the second is the shape of the images, some images are rectangular in shape, and some are square, which might be the problem with the training algorithm. So to solve this issue we first crop the watermark strip, then resize the images to 224*224, and after that convert all images into gray-scale, just to reduce the computation cost. After that, we created a data frame that stores the image path, the text, and the label 0 for non-depression, and 1 for depression. The entire data set is split into the train, test, and validation files with the split ratio of 80:10:10 stratified with label columns. The Dataset **Image1** consists of 2473 Image samples and the samples in Train, Validation and Test Set as 1583 Images, 395 Images and 495 Images for 2 Classes. The Dataset **Image2** consists of 5475 Image samples and the samples in Train, Validation and Test Set as 4380 Images, 547 Images and 548 Images for 2 Classes.

*2) **Text1 and Text2**:* For the Unimodal Text Classification task, top posts from a subreddit were initially used with tags similar to the ones used for extracting images. However, the number of posts extracted related to depression was relatively low. To address this, additional 5000 texts related to suicide obtained from a collection of posts from the "SuicideWatch" and "depression" subreddits on the Reddit platform were incorporated. Overall,in Text-1 dataset, 13311 posts related to depression and non-depression topics were collected. To ensure a fair evaluation of models, the dataset was divided into three parts using a stratified split of 80:10:10.The Dataset Text-2 consists of 5475 text samples which consists of 4380 texts in train, 547 texts in validation and 548 texts in test sets respectively.

*3) **Image+Text and Inference**:* The project utilizes the visuals scraped and the corresponding text from the same source and the text is the description of the corresponding image. An advantage of the new data set is the ability to incorporate and learn the relation between the image and text. In total we have extracted 5475 images, and their corresponding description. The tags we used to extract the related images were depression, sadness for scraping depression posts, and happiness, joy for scraping non-depression posts. This final dataset is referred as DF, whose image and textual samples are separately reffered as Image2 and Text2 respectively, to re-evaluate the unimodal pipelines.

B. Uni-Modal Classification Architectures

*1) **Image Classification**:* The techniques being used for pre-processing techniques are used with the specified values for the parameters as rescale=1./255, shear_range=0.2, zoom_range=0.2, horizontal_flip=True as transformations. The target image size is set as 128*128*3. The Pre-Trained Models used in our project are as ResNet50, VGG19, InceptionV3 and Xception. The learning rate is set as 0.01. However, to prevent the case of overfitting the model and to keep a check on the redundant update of loss and hence the inability of the model to learn further, the patience for decay is kept as 10 and the learning rate reduces by a factor of 0.50, if there is no significant learning for 10 consecutive epochs. Early Stopping has been found in VGG19, Inception and Xception model, after 27, 18, and 14 epochs respectively, whereas the maximum number of epochs is set to 50. SGD (Stochastic Gradient Descent) is being used as Optimizer and the loss function being used is Binary Cross-Entropy.

*2) **Text Classification**:* The paper proposes various pre-processing techniques using the NLTK library like whitespace, URL, user mention, number, and emoji's removal, lowercase conversion, tokenisation, stopword removal, punctuation removal, spelling checking, and lemmatisation. For dataset T-1, it applies all the above-mentioned techniques for further classification along the pipeline. However, dataset T-2, on which only the BERT model is applied, lowercasing, punctuation removal etc. is avoided as the BERT is sensitive to the contextual semantics added through cases, punctuations, etc.

Three distinct models were employed to process and classify the curated data using a variety of embeddings. The initial model used FastText embeddings to produce subword embeddings that could capture the meaning of complex morphological and out-of-vocabulary (OOV) words. The resultant embeddings were trained using a support vector classifier. However, as FastText embeddings produce non-contextual embeddings, a second model was developed that employed pre-trained BERT model embeddings that are contextualized and can provide more precise contextual information. This model also utilized a support vector machine classifier on the BERT embeddings. The pre-trained BERT model was fine-tuned on the dataset to enhance the classification performance. Firstly, a 'Bert-based-cased' tokenizer was used to tokenize the given sample into the correct input format for Bert, and then Fine-tuning the pre-trained BERT model on the specific task of depression classification leveraged its ability to capture contextual information and further enhance the model's performance.

C. Multi-modal Classification Architectures

In this configuration, we have taken both image and text into consideration for predicting depression. For this multi-modality configuration, we have performed experiments on three setups. **Setup-1** is a combination of **ResNet-50** for image and **Distill-BERT** for text, in **Setup-2**, we have **VGG-16** for images and **BERT** for text, and in **Setup-3**, we have **Inception_v3** for images and **BERT** for text. The architecture is built using PyTorch. First, the data is pre-processed, and create data loaders. The dataset is examined for the class imbalance problem in the training data. The dataset has 2496 depression labels and 1184 non-depression labels. The hyper-parameters are set as a weight decay of 0.01 to update the loss to handle this imbalance problem and improve the model's performance. For fusing the different modalities, we have used the ***Late fusion*** approach concatenating the feature vectors of both image and text, then applying MLP. The dimension of the base image is 224*224*3, and for text is String.

The model is trained for 10 epochs with the specified hyper-parameters as Adam optimizer, Drop-Out value 0.5, and a learning rate of 1e-5. The proposed work includes a T-SNE (T-Distributed Neighbourhood Embedding) plot to visualize the high-dimension features. The model. We have changed the classification head of our model to a simple Linear layer just to get the fused vector of both modalities. The feature vector has the dimensions as $\mathbb{R}^{1280}$. The dataset is sub-sampled to randomly retrieve 50 samples of each class from the test set to plot the T-SNE plot. After concatenating the feature vectors of both classes, the dimension of the feature vector is as $\mathbb{R}^{100*1280}$ rows. The T-SNE plot is given in Fig. 15

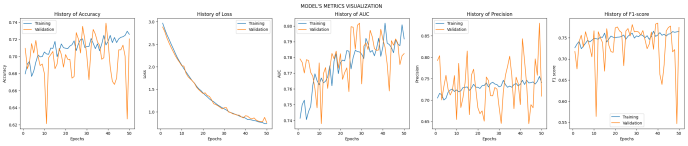


Fig. 2: Plot of Evaluation Metrics for Training and Validation Data of ResNet50 for Image2

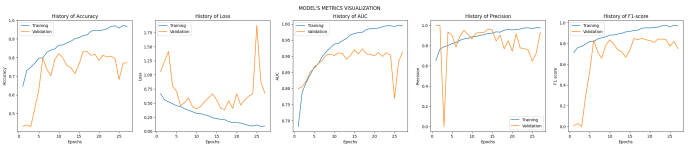


Fig. 3: Plot of Evaluation Metrics for Training and Validation Data of VGG19

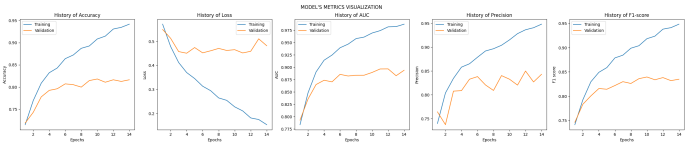


Fig. 4: Plot of Evaluation Metrics for Training and Validation Data of Xception

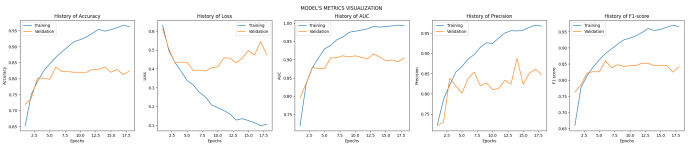


Fig. 5: Plot of Evaluation Metrics for Training and Validation Data of Inception

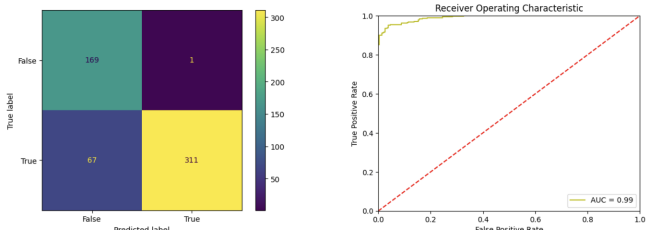


Fig. 6: Evaluation metric for Non-Contextual+SVM Metric

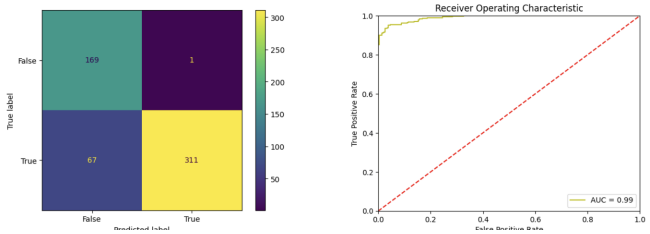


Fig. 7: Evaluation Metric Contextual + SVM evaluation metric

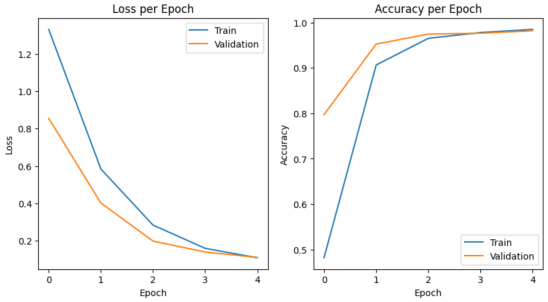


Fig. 8: Plot of Loss and Accuracy for BERT Text for DT-2

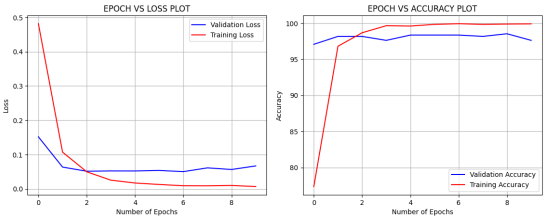


Fig. 9: Plot of Loss and Accuracy for Setup-1

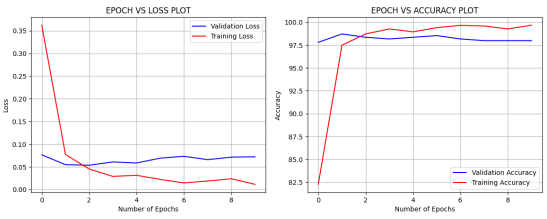


Fig. 10: Plot of Loss and Accuracy for Setup-2

D. *Explainability*

*1) **GRAD-CAM Visualisation for Unimodal Image Classification**:* GRAD-CAM has been used as a part of the set of tools used to iteratively improve the model by visualizing whether or not, the model is able to learn appropriate representation. In the below examples for Fig. 12 and Fig. 13, the model is able to clearly identify and learn from certain patches of images for a particular label. However, the model may fail under certain circumstances as it might not be able to learn clear distinguishing features from target patches as shown in 14.

E. *T-SNE Plots for Multi-modal Fusion*

The T-SNE Plot for fused embeddings for generated feature vectors as extracted from the proposed multi-modal architecture is shown in Figure. 15

The architecture of the ResNET50 and DistillBERT is given in Fig. 2

VI. Results & Evaluations

TABLE I: Performance Metrics for Validation Set of Image-1 for Image Classification

Models & Metrics	Accuracy	Precision	Recall	F-1 Score
CNN	0.40	0.40	1.0	0.56
ResNet50 [9]	0.72	0.61	0.80	0.69
VGG-19 [10]	0.86*	0.75	0.98*	0.83*
InceptionV3 [11]	0.84	0.78*	0.82**	0.81**
Xception [12]	0.85**	0.85*	0.76	0.77

VII. Conclusion

The project aims to rank online media posts for depression, which can contain multimodal and hidden, subtle contexts.This might not be captured by unimodal models or models trained on just clinical, psychological datasets, which is also confirmed by our unimodal results.Hence, we created a novel

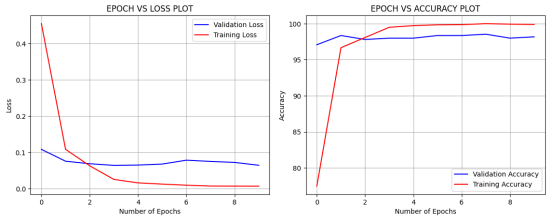


Fig. 11: Plot of Loss and Accuracy for Setup-3

TABLE II: Performance Metrics for Validation Set of Image-2 for Image Classification

Models & Metrics	Accuracy	Precision	Recall	F-1 Score
ResNet50 [9]	0.72	0.70	0.86*	0.74
VGG-19 [10]	0.77	0.93*	0.64	0.75
InceptionV3 [11]	0.82*	0.84**	0.84**	0.84*
Xception [12]	0.81**	0.84**	0.83	0.83**

TABLE III: Performance Metrics for Testing Set of Image-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 Image-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**

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

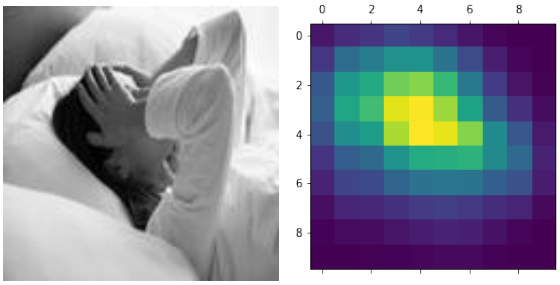
Models & Metrics	Accuracy	Precision	Recall	F-1 Score
BI-LSTM (Baseline)	0.92	0.89	0.93	0.92
FastText + SVM	0.96	0.97	0.92	0.95
BERT-Embeddings + SVM	0.96	0.95	0.95	0.95
BERT-Fined Tuned Model	0.98	0.98	0.98	0.98

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

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



(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. 12: GRAD-CAM as a tool for Explainability

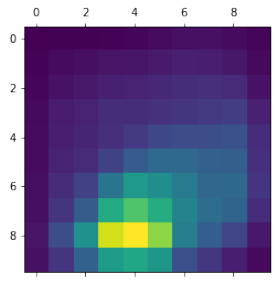
TABLE VIII: Performance Metrics for Inference Data

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

multimodal dataset scrapped from real-world online posts using information retrieval techniques.Also, we propose novel, deep fusion based model which being trained on multi-modal data of real-world media posts, shall be robust to identify different variations of depression personalised to different posts. The inference models,

Based on the experiments and results, it can be concluded that both text and image play a role in determining the presence of depression in social media posts. Overall findings indicate that using both text and image features in a multi-modal approach outperforms using only text or image features in isolation.The performance of these models can be improved by using more advanced language models such as BERT, which can capture more complex contextual information in the text. On the other hand, for image classification models also showed that images can be a useful source of information in detecting depression in social media posts. The models based on pre-trained convolutional neural networks (CNNs) were able to achieve high levels of accuracy in detecting depressive posts based on the image content. However, when text and image are combined, features in our multimodal approach, we observed a significant improvement in the overall classification performance. This indicates that both text and image features provide complementary information in detecting depression

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

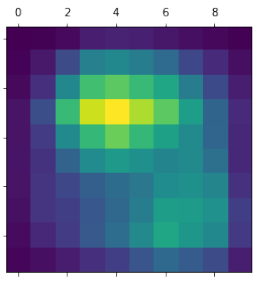


(b) Gradient Heat-Map for Original Image as Input to InceptionV3

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

Fig. 13: GRAD-CAM as a tool for Explainability

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

in social media posts. Therefore, in a real-world scenario, it would be beneficial to consider both text and image features when determining the presence of depression in social media posts. This multimodal approach can provide a more comprehensive and accurate understanding of the content and context of social media posts related to depression, which can in turn be used to inform mental health interventions and support services.

VIII. Future Scope

Dataset quality can be improved by adding more diversified data like scrapping data from Linkedin, Instagram as it consists of several distinct types of depression posts as an individual sharing the post of a layoff or an emergent need for a job will consist of an underlying emotion of depression, stress, and anxiety. The scope hence proposes to use diverse forms of social platforms to incorporate different causes of depression. Modalities such as audio, video, or physiological data can also be used to improve the performance of the model. Extending the model to include these modalities or exploring how to fuse multiple modalities together. In this project a simple concatenation method has been used to fuse the image and text features. However, there are several other fusion strategies such as attention-based fusion, cross-modal retrieval-based fusion, or late fusion that could be explored to further improve the performance of the model. The proposed research result pave the way for deploying the model on social media platforms such as Twitter or Facebook to detect early signs of depression in users' posts. This can be achieved by integrating the model into the social media platform's algorithms to automatically flag posts that indicate depression or recommend mental health resources to users who may be at risk.

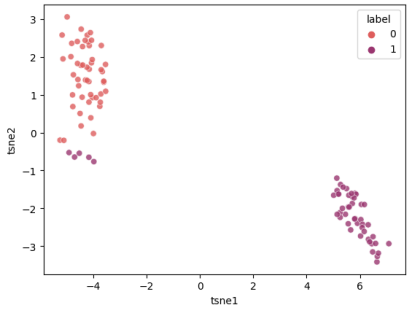


Fig. 15: T-SNE Plot for High Dimension Feature Representation by features from Setup-1

References

[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.

[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.