

Multimodal Deep Learning based Framework for Detecting Depression and Suicidal Behaviour by Affective Analysis of Social Media Posts

Anshu Malhotra^{1,*}, Rajni Jindal¹

¹CSE Dept., Delhi Technological University, Delhi, India

Abstract

INTRODUCTION AND OBJECTIVES: Currently, no social media platform has deployed a real time system which can analyse users' state of mind based on day to day posts on continual basis and detect the onset of depression, suicidal or self harming behaviour etc. Platforms majorly rely on manual reporting of suicidal and self harming behavior. In this paper, we propose a real time, deep learning based system for affective analysis of a user's online social media posts of multimodal nature, with the objective of detecting onset of depression and suicidal or self harming behaviour; as depression often drives people to commit suicide or harm themselves physically.

METHODS: Joint representations are obtained by fusing the individual vector representations of multiple modalities from user's social media feed: text, image and videos. These vector representations are in turn obtained through state of the art approaches for each modality e.g. VGG-16 for feature extraction from images, word2vec for text and Faster R-CNN on video frames. These joint representations are used to obtain weighted average score which can be used for making the final classification using the Softmax prediction layer.

SIGNIFICANCE AND IMPACT: To the best of our knowledge, this is the first research where the use of deep learning techniques has been proposed for real time detection of onset of depression and suicidal behaviour by analysing multimodal user generated content.

Received on 16 December 2019; accepted on 10 January 2020; published on 31 January 2020

Keywords: Multimodal Deep Learning, Social Media Analysis, User Generated Content, Mental Health, Depression, Suicidal and Self harming Behaviour, Psychology

Copyright © 2020 Anshu Malhotra *et al.*, licensed to EAI. This is an open access article distributed under the terms of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>), which permits unlimited use, distribution and reproduction in any medium so long as the original work is properly cited.

doi:10.4108/eai.13-7-2018.164259

1. Introduction

Mental health of a person is his psychological, emotional and social well being which affects how a person thinks, feels and behaves. It is important throughout a person's lifetime as it governs the general well being of a person and his overall quality of life; it determines how a person handles stress, his decision making capabilities, how he interacts with people, relationships, his productivity at work and all other aspects of life¹. Mental health disorders can

occur to a person at any stage of his life, irrespective of gender, ethnicity and economic background². Some of the common mental disorders are: Depression, behavioural and mood disorders, suicidal and self harming behaviour, bipolar and borderline personality disorders, dissociative disorders, anxiety and panic attacks, hypomania, paranoia, phobias, schizophrenia, eating & sleep problems and others³.

The number of people suffering from mental disorders now days is increasing at an alarming rate due to multiple factors like fast paced lifestyle, lack of

*Corresponding author. Email: anshumalhotra5@gmail.com

¹<https://www.mentalhealth.gov/basics/what-is-mental-health>

²<https://www.medicalnewstoday.com/articles/154543.php>

³<https://www.mind.org.uk/>

emotional support system because of nuclear families and fewer friends or siblings, excessive work pressure, poor physical health, biological factors, family history, life experiences, failures & struggles, and many others. It has also led to an increase in number of suicide deaths recorded worldwide. As of 2016, about 15.5% of global population suffers from some mental or substance use disorder⁴. For India, this number is about 14.96% of its population. According to World Health Organization, about 27% of adult population in EU has suffered from some mental disorder during their lifetime⁵. About 1 in 6 adults in USA (18.3% of their population) experiences serious mental illness during an year⁶. Suicide is the top 10th reason for deaths in USA and 2nd most common cause of death amongst US population aged between 15-24⁷. India is among the countries with highest suicide rates for the youth aged between 15-29⁸. Teenagers are most active users of social media and there is a steep rise in cases of teenage depression worldwide; suicide deaths are increasing at an alarming rate in India and it is reported that one student commits suicide every hour where many a times he talks about ending life on various social media platforms⁹. Recently there have been few horrifying incidents of people even live streaming their suicide on social media and all of these people were found to be suffering from depression in past before they ended their life^{10 11 12}. We strongly feel, many lives could have been saved by the people in the person's social network if there was a real time system to detect such events and alert the people in network.

The subject of mental health is least talked about in comparison to physical health issues of human body. Mental health disorders go undetected for long, even by the person himself. They are ignored or neglected due to lack of awareness and understanding about them or due to social stigma. This further aggravates the mental agony, pain and frustration of the

patient and may incline him towards suicidal or self harming behaviour. Therefore the timely detection and treatment of mental disorders is of utmost importance or else it may lead to developing suicidal tendencies or self harming behaviour in a person and may also make him violent/aggressive towards other people. There have been many reported incidents like of mass shootings etc. where the suspect was suffering from poor mental health. Timely detection and awareness may help the person recover and restore back to normal life by extending required emotional support & care by friends & family and by providing medical/professional help as needed.

Though it may not be always possible to accurately determine if a person is suffering from any mental disorder, however person's behaviour, feelings and mood may call out early signs of warning that may help e.g. uncharacteristic emotional outbursts of anger/frustration/aggression, feelings of helplessness and hopelessness, difficult feelings & behaviours, mood swings, negative or destructive thoughts & feelings among many others¹³. In real world as well, psychiatric or psychologists look for these signals during their interactions with patients and they try to gauge patient's moods, thoughts, feelings and behaviour by asking questions about them in different ways. The patient's responses help them analyse and diagnose his condition. In today's virtual world of online social media, people talk and express themselves more on "online" platforms than in person. Hence, we strongly believe affective analysis of a user's online social media posts can help spot these early signs of mental health disorders by trying to understand his state of mind on continual basis over an elongated period of time.

The latest research in the domain of machine learning and artificial intelligence is achieving unparalleled results by using deep neural networks like CNNs, RNNs, and LSTMs etc. for various supervised and unsupervised tasks. Such deep learning architectures excel in discovering patterns in unstructured data like images, videos and texts. These days, the majority of the content and text on the internet is highly unstructured as it is generated by users of the internet themselves who are not trained or professional content writers. The most common form of content created by the users on a daily basis are their online posts or comments on various social media like Twitter, Facebook etc. These posts generally comprise of short texts, images, videos or a mix of them. These posts by the user are a modern form of communication in today's virtual world and are like a window into the user's mind. If analysed properly, these can help decipher and understand the

⁴<https://ourworldindata.org/mental-health>

⁵<http://www.euro.who.int/en/health-topics/noncommunicable-diseases/mental-health/data-and-resources>

⁶<https://www.nimh.nih.gov/health/statistics/mental-illness.shtml>

⁷<https://www.nami.org/learn-more/mental-health-by-the-numbers>

⁸<http://businessworld.in/article/De-Stigmatizing-Suicide-And-Mental-Health-In-India/10-09-2017-125777/>

⁹<https://www.hindustantimes.com/health-and-fitness/every-hour-one-student-commits-suicide-in-india/story-7UFFhSs6h1HNgrN060FZ20.html>

¹⁰<https://indianexpress.com/article/india/gurgaon-man-commits-suicide-live-on-facebook-police-5287059/>

¹¹<https://www.ndtv.com/agra-news/agra-man-commits-suicide-live-streams-on-facebook-1881963>

¹²https://www.huffingtonpost.in/2017/04/04/a-24-year-old-live-streamed-his-suicide-on-facebook-before-jumpi_a_22024719/

¹³<https://www.mentalhealth.gov/basics/what-is-mental-health>

user's thoughts, emotions and state of mind. Online social media can be leveraged as a platform to detect onset of depression and such suicidal and self harming behaviour using machine learning and natural language based systems to decipher and understand user's thoughts & feelings he expresses through his social media posts. People's state of mind can be detected as they often talk of sadness, pain, grief, disappointment, sorrow and other similar negative/pessimistic feelings by share emoticons, pictures, videos and texts.

In this paper, we propose a real time, multimodal deep learning based framework for affective analysis of user's online social media posts with the objective to detect the onset of depression and suicidal or self harming behaviour.

2. Background Study and Related Work

2.1. Current State of the Art

Seeing such incidents on rise, even popular social media sites like Facebook are rolling out systems to detect depression and suicidal tendencies. In 2016, Facebook had rolled out a feature where in users can flag a post by another user that they happen to see as being of "suicidal or self-harming" nature by giving feedback from the drop-down option against the post¹⁴; following which numerous options and resources are listed like: alerting Facebook to review the post & reach out the person, resources like contact numbers of suicide prevention organizations, sending words of support and care to the user in distress. Facebook suggests that if someone seems in an immediate danger, local emergency services be called right away without any delay. However, such systems may not be successful in developing geographies like ours where technology adoption is still very low among the users and emergency services and agencies for them to be receive alerts and respond in real time. Also, such approaches that rely only on other users reporting and flagging suicidal and self-harming behaviour may not be adequate & sufficient to cater to billions of users across the clock; hence automated techniques should be researched to overcome the shortfalls of manual review which is not scalable in real time. Facebook is developing Artificial intelligence based tools to detect such user posts by analysing them and the responses of the people on those posts^{15 16}.

¹⁴<https://techcrunch.com/2016/06/14/facebook-suicide-preventionz>

¹⁵<https://www.bbc.com/news/technology-39126027>

¹⁶<https://www.cnbc.com/2018/02/21/how-facebook-uses-ai-for-suicide-prevention.html>

2.2. Related Literature Survey

Many machine learning techniques have been previously used to understand user emotions and personality. Also, the use of social media content generated by the users has been explored to understand his behavior, emotions and personality. The social media has also been analyzed at scale for understanding health issues. Qualitative surveys and their statistical analysis have suggested difference in user behavior based on their mental health on social media platforms [2][3][4][5][6]. Since our work lies at the intersection of these three domains: social media content, (mental) health and use of machine learning techniques; hence we discuss the relevant literature from these three perspectives.

Desmet et al. have used lexical and semantic features to train binary SVM classifier which they deploy to understand 15 categories of emotions in suicide notes[7]. Though they have used only textual features and achieve moderate accuracy only in detecting various emotions, however their study motivates us to explore the techniques which may be designed for continuous detection of depressive moods and emotions which may help in preventing suicides.

User posts on social media have been used in the past to analyze various public health related issues that are being discussed on the platform. For example, study by Paul et al. created a statistical Ailment Topic Aspect Model to discover the health topics being discussed on social media in an unsupervised manner with minimal human intervention [8].

Social media are being increasingly used to understand user emotions. Boychuck et al. propose the use of SVMs (for text) and off the shelf commercial products (for facial expressions), to recognize emotions of aggression, violence etc. during football matches using textual and visual features from Instagram posts (not in real time) [9].

The work of Choudhary et al. is an extensive study on using social media platforms as a tool for measuring and predicting depression. Their major contribution is leveraging user's textual post for mental health surveillance using SVM classifiers and indicating depression index from user's social media posts. Crowdsourcing is used to collect the Twitter handles of users who give their consent for crawling their past posts, which are then used to extract user centric content and engagement features. Though not real time and unimodal, their SVM with radial basis function predicts the onset of depression with 70% accuracy[10][11].

Research by Reece et al. suggested that even the choice of image filters people use on social media can be indicative of depression e.g. in case of Instagram[1]; the authors have studied the various attributes of the

pictures posted on the media along with the other network properties.

Next we discuss few of the closest studies related to our proposed system and algorithm. Majumder et al. try to detect Big 5 personality traits from essays' sentences using fixed document level stylistic features and variable length sentence features using word2vec embedding; and 5 separate CNNs with single hidden layer are trained which give an accuracy between 50-60% [12]. The dataset used by them are essays which are structured and have minimal language errors as compared to user posts on social media. Also, user generated content (ugc) is generally multimodal in nature comprising of at least 2 modalities. The research done by Gucluturk et al. also tried to detect these Big 5 traits, however they used multimodal audio & visual signals from 10K YouTube videos of 15 seconds and trained deep residual networks with 17 layers each for audio & visual signals, which are then fused in a fully connected layer [13]. Though this model is only for short videos and may not work for multimodal user generated content on social media, however the use of deep neural nets gives much higher accuracy of the order of 90%. Lin et al. have explored the use of various classifiers and 4 layer deep neural network to understand the psychological state of the twitter users. They use two types of features - low level content based (for which they use cross auto encoders with CNN for feature extraction) and high level user behavior / engagement based features which are used to train DNN to predict users' stress levels [14].

Additionally, the use of advanced machine learning and computer vision techniques has also been explored for detecting psychological disorders by analysis of complex and multivariate data from brain images, EEGs, MRI, fMRI scans etc. Such studies have been successful in distinguishing patients with schizophrenia, bipolar disorder etc. from healthy patients. However, such clinical studies are not directly related with our proposed framework and hence excluded from the literature survey in the interest of space constraint.

2.3. Research Contributions of Proposed Framework

The techniques discussed in the previous section of literature survey have numerous limitations. None of the techniques discussed above are designed for real time detection of depression or suicidal behavior in user posts; most of the techniques are for lateral analysis of user posts to understand user behavior. Many of them are for very domain specific analysis. It is difficult to analyse unstructured multi modal content to structured or unimodal user content e.g. just text. Most of them do not analyse multi modal unstructured user generated content and hence miss out on a lot of information that can be useful in analysing user

behavior. Also, very few of them have explored the use of recent deep learning algorithms for analysis. Not much research has been done in applying deep learning techniques for analysis of multimodal User Generated Content for understanding depression and suicidal or self harming behavior, which a crucial & unaddressed health concern of the millennial world. Our proposed system framework discussed in detail in the next section, addresses these research gaps and challenges. The salient contributions of our research and proposed framework are:

- Fusing multimodal user generated content i.e. text, images and videos from user's social media posts and obtaining joint representations using deep learning based vectorization techniques
 - The proposed system uses deep learning techniques for classifying user generated multimodal content to detect depression and suicidal or self harming behavior
 - The proposed framework when implement and integrated by social media platforms can work in real time to the detect the onset of depression which may lead to suicides or self harming actions. The continuous evaluation of user posts over a time window can help understand user's behavioural pattern over past few months and help discover change in his mood / behaviour that may serve as an indicator of depression onset.
 - Deep learning techniques are known to outperform traditional machine learning techniques in case of error prone unstructured content as in our case i.e. user generated social media content which is not proof read published and has a lot of spelling errors, usage of emoticons, slangs, memes etcetera.
 - We have proposed the use of Deep learning techniques for feature generation from text, images and videos from user posts.
- To the best of our knowledge, this is the first research where the use of deep learning techniques has been proposed for detecting onset of depression and suicidal / self harming behaviour by analysing multimodal user generated content from social media. .

3. Proposed Framework

3.1. Social Media Platforms and Datasets

In our research paper the main contribution is that we have presented a novel idea using deep learning algorithms over multimodal user generated content to detect onset of depression and suicidal or self harming behaviour on various social media platforms in real time basis. Our proposed novel framework can be adopted by various social media networks for deployment and integration over their platforms. The three most popular social media platforms Facebook, Instagram and Twitter allow users to compose their post, status or tweet using all or any of these

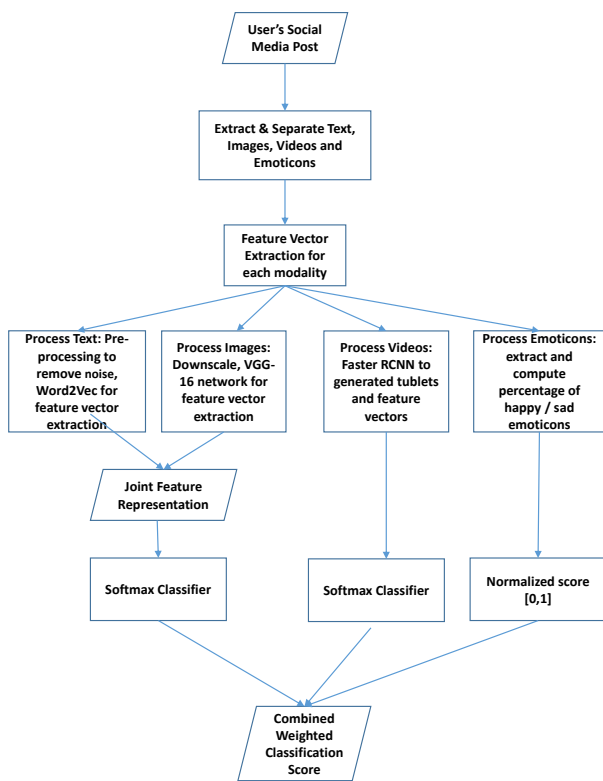


Figure 1. Flowchart for multimodal analysis of social media post.

modalities. Hence, the proposed framework may be adopted by these social networks for analysing the user's multimodal posts. Since in our framework, each modality is analysed independently and does not require inputs parameters from other modality; hence having all the types of content in each user post is not necessary. Vectorized feature sets are computed independently and separate deep learning algorithms are used for classification of each type of content. Hence it is not necessary to have all the modalities in each and every user posts, the available modalities can be analysed. Additionally, the framework will work even on other social media websites also where all modalities may not be supported. The platforms can train and deploy our proposed deep learning framework for different user personas representing different kinds of user clusters.

3.2. Flowchart and Algorithms

Our novel proposed framework for detection of depressive and suicidal behaviour, comprises of various independently working processing components required for pre processing and classification of multimodal user generated content on social media. These processing steps for achieving the research objective are briefly enlisted in flowchart in Figure 1 and the framework is illustrated in detail in Figure 2. Each of these processing components of the framework uses multiple, state of the art and domain specific algorithms and techniques to accomplish its task. These are discussed separately in the following subsections in depth.

3.3. Processing Images

Out of all modalities, Images and Text are predominant on social media; users post text and images more frequently as compared to videos. Rich textual and image content can be extracted from a user's social media profile. Hence, both of these act as a good metric to evaluate general well being of users mental health. Relevant features vector representations need to be extracted that can be used for classification of user posts. As a first step, the images in user posts are down-scaled for the purpose of consistency of further analysis.

The feature vector extraction in case of images from user's posts is performed using a VGG-16 network, a popular deep learning approach. The VGG-16 is a 16 layer architecture. The selected model is pre-trained on ImageNet dataset. It has 13 layers of convolutions, followed by three fully connected layers. The interesting attribute of VGG networks is: convolution from filters of spatial size 3×3 . Using filters with spatial size of 3×3 , reduces number of trainable parameters drastically. In our proposed architecture, the last three layers (i.e. Fully Connected layers) are removed, as they are trained for the task of classification of ImageNet dataset, which is not what we need to do. Rather we want to extract the feature vector representation from images that will be used for classification of user posts. Hence, a global average pooling layer is added after 13th convolutional layer, which outputs a feature vector of dimension 512×1 , that can be used for classification of user posts.

The extracted feature representation is passed through a fully connected (FC) layer, which is later combined with textual representation to obtain a joint representation for textual and image based feature.

3.4. Processing Text

Other than pictures, text is another critical modality which is usually always available for analysis of any social media post. Usually the available text from social media in the form of tweets, status updates, and posts, is

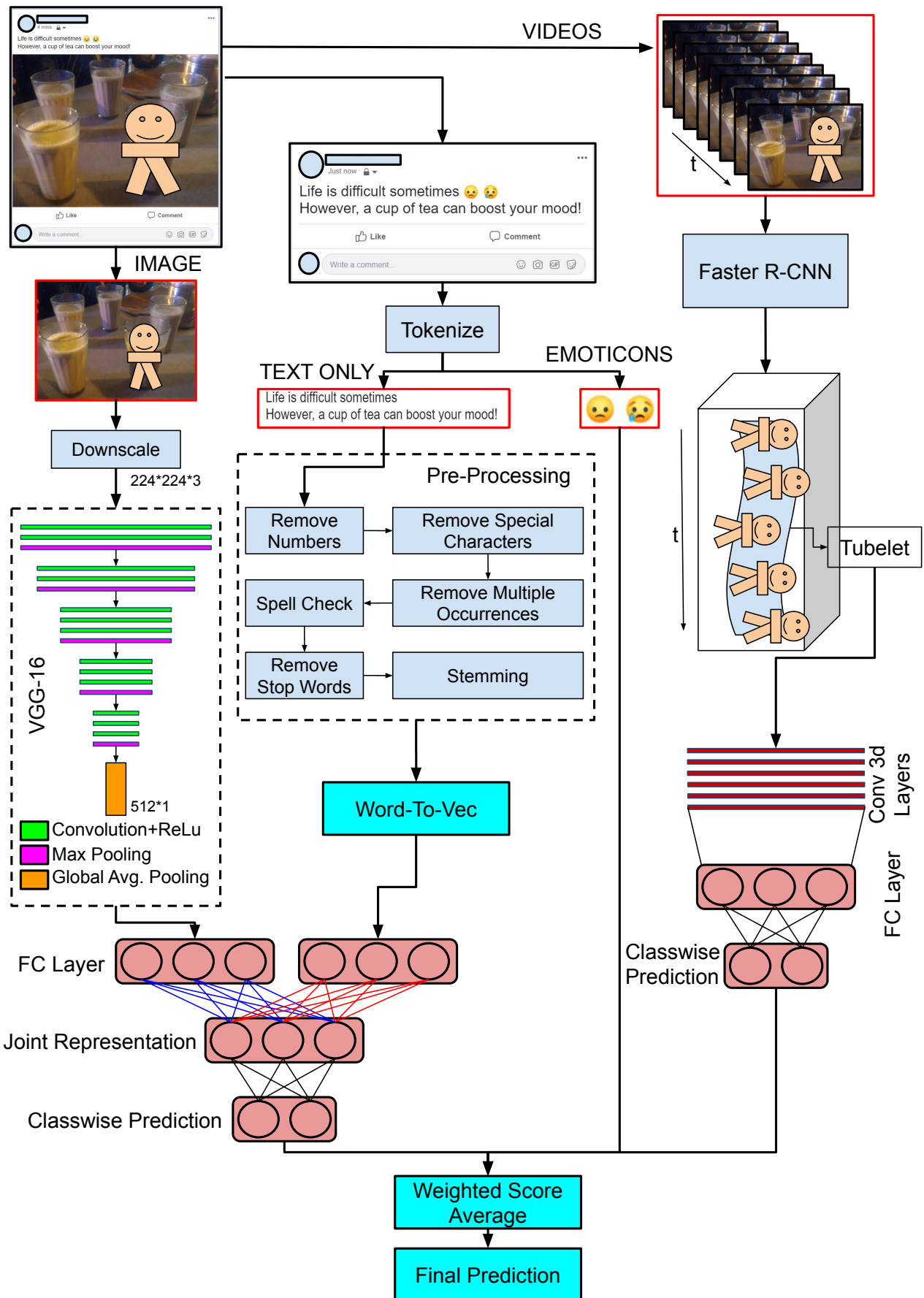


Figure 2. Proposed model for multimodal analysis of social media post.

short and highly unstructured. Hence before extracting the feature vector representations from these text, the text needs to be pre-processed so as to minimize the noise. The techniques utilized for pre-processing and for extracting feature vector representations from texts are discussed below.

Pre-processing. The first and foremost task is to tokenize text by the space characters and other delimiters. Post tokenization, emoticons are separated. These emoticons are treated as a separate modality and are processed distinctly which is discussed in a later section. For the tokenized text, the sequential steps for pre-processing are:

- **Removing numbers:** Unnecessary numbers may lead to misclassifications. Hence, as the first step, all numbers are removed from the tokenized text as usually numerical values don't contribute in expression of mood, feelings or thoughts; numerical values are commonly used for sharing factual information e.g. phone number, address and others.

- **Removing special characters:** Since the aim to classify words, special characters in data can be lead to additional noise and impact the model training and classification. Thus, all ASCII characters other than a-z and A-Z characters are removed.

- **Removing multiple occurrences and Spell Correct:** One of the most prominent characteristics in short texts like social media posts, SMSs are repeated characters and typographical spelling errors; e.g "Heyyyyy" and "Heellooooo". Such repeated characters from words that don't occur in English Dictionary are removed.

- **Spell check:** All the previous cleaning steps were performed with the aim to bring the mis spelt word near to its correct spelling. For this, wherever possible dictionary based spell correction is used to replace the words with nearest matching dictionary words. Closeness is defined in terms of edit distance. The vocabulary of dictionary words must be built by combining multiple sources: language dictionaries e.g. Oxford, Wikipedia, Encyclopedia to include words from different domains of study, psychology based studies to include words characterizing emotional and mental well being of humans.

This master vocabulary or master dictionary built above using multiple sources, is used for spell correction of the pre processed words from user posts. For each of the word, we look for words in our master vocabulary to find a set of words with least edit distance, with which the original word can be replaced. This edit distance can be 0 if the word is located precisely in our dictionary. From the set of words with the same least edit distance, we choose the one with the highest probability in our vocabulary, i.e., the word that is most frequent in our vocabulary.

Input word: theC

Words with edit distance 0: none

Words with edit distance 1: the and they

Thus, we replace our input word theC with the one with higher probability out of the and they.

- **Removing stop words:** Since stop words don't add much meaning and information while compare two sentences or documents, hence stop words are removed using the standard stop word list.

- **Stemming:** Many words are derived from the same root word, with some changes in plurality, tense, etc. The basic meaning is similar to the parent word. For instance, the parent word of "children" is "child". Thus, in this step, we map all such words to their parent words by stemming.

Feature Extraction. Once the text is pre-processed by all the steps discussed above, the next requirement is to represent the text in a vectorized form which can serve as a feature representation for classifying user posts. For this, Word2Vec can be utilized to learn the word embeddings or features representations of dimension size of few hundreds. The word2vec document representation of words occurring in user posts may be obtained through Skip Gram model or CBOW model. The word vectors can be modified using SGD to capture the contextual information as well.

3.5. Fusing Image and Text Feature Representations

Once the vectorized feature representation for text is obtained, the representation is passed through a Fully Connected (FC) layer. Followed by Fully Connected layer, joint representation is obtained by passing image feature representation and text feature representation together to the next Fully Connected layer. Note that, there might be scenarios where both text and image might not be available simultaneously. To incorporate this scenario while learning joint representations, we drop some connections (marked in blue and red lines in Fig. 2) of text, or image, or both of them during training. This ensures that the next Softmax prediction layer is able to classify user posts, even if some information is missing.

The following Softmax layer as indicated in Figure 2, is used for binary classification of user posts from the joint representation of text and image and detect & flag social media posts which are indicative of depression or suicidal behaviour.

3.6. Processing Videos

Though images and text dominate the social media posts, Videos can also play a vital role in detecting depression and suicides. In recent times, many users have posted live stream video of their suicide on social media sites such as Facebook. Also, at times the users may share video or voice messages also on their

social media profiles talking about their concerns and troubles. In order to process such videos, the algorithm should be able to: (i) process videos in real time, (ii) perform person/object detection, and (iii) accomplish classification of video into 2 classes, self harming / suicidal or not. Keeping the above tasks in mind, the videos are processed with 3 sequential modules, as described below:

Person detection using Faster-RCNN. To first step is to locate, detect and identify the person of interest i.e. the user in the videos shared by him. As the foremost task, the video stream is converted to frames. However, detection at each frame level is not be a feasible option due to intensive computation. Hence, every 20th frame is selected for analysis and person detection & identification.

For each of the selected frames, Faster-RCNN (Region Convolutional Neural Network) is used for detection. Faster-RCNN has a convolutional neural network to generate feature maps for each frame. A ZF-net architecture is used to generate features maps, which are further fed to the Region Proposal Network (RPN). The RPN creates region proposals, from which the person/object is detected. Each proposal feature map is fed as input to the fully connected (FCs) layers, which predicts bounding-box for the object and class probabilities. The model is initially pre-trained using Pascal VOC dataset.

Using bi-linear interpolation from coordinates of every 20th frame, the coordinates of the middle 10th frame is found. This way, we process videos at six fps (frames per second), however, perform computation at merely three fps. Finally, we create an ensemble of the detected person over the video to create tubelet. This tubelet with the detected person/object would be used to track its activity throughout the video.

Feature generation from tubelets. The tubelet generated in the previous step is fed as input to 3D convolution layers (conv3d). Conv3d layers, in addition to the regular spatial locations, perform convolutions over an extra dimension. This dimension is time depth in our case. Performing convolutions of frames over the time frame encodes temporal context, which plays an essential role in understanding how the person moves across the video and detect his activities.

Classification using FCs. Once the conv3d layers encode the tubelet to obtain feature representation, the classification is performed using the FC (Fully Connected layer). The FC layer is followed by a two-node Softmax layer, which does the binary classification of video activity into two classes: depression/suicidal or not.

3.7. Processing Emoticons

Emoticons are an important way of expressing a person's mood, feelings and thoughts. From the tokenized text, we had pruned out emoticons. Here we discuss how these emoticons may be utilized as features for understanding user feelings and thoughts. All the other three modalities utilized Softmax prediction layer to output probability scores denoting the prediction classes i.e. Depressive / Suicidal behaviour vs. normal behaviour. These three previous scores are normalized in the range [0,1] for each class. Hence, the score from emoticons needs to be in the range of [0,1] as well. The class wise scores are computed as follows:

$$DepressionorSuicidal = \frac{\#sadEmoticons}{\#totalEmoticons} \quad (1)$$

$$NormalorHealthy = \frac{\#happyEmoticons}{\#totalEmoticons} \quad (2)$$

The score is 0.5 each if there are no emoticons.

3.8. Fusion Feature Vector and Classification

At the last step, the final probability score for classification is obtained by weighted average of the four scores obtained above for each modality: text, image, videos and emoticons as discussed in the above subsections. The user post belongs to the class with the highest score which used the predict and flag the post as indicative or depression /suicidal behaviour or not.

4. Discussion

Though social media platforms have systems for user sentiment analysis, opinion mining and also provide him with content recommendation & suggestions; however not many platforms have an objective to detect the onset of depression and suicidal or self harming behaviour. Recently some platforms have deployed manual reporting based widgets where other users can report posts indicating such behaviour e.g. if they see someone talking about committing suicide or posting pictures repeatedly standing on the edge of a roof or something similar. However, manual reporting based approaches are not reliable due to various reasons the number of people who happen to see that post and are observant or even care to report. Also, it is important to understand the user psychology over an elongated period of time. Hence automated machine learning based techniques need to be explored so that all user posts of multimodal nature can be analysed over a time window to better comprehend the user's mental state. The handful of approaches that are being used currently, rely majorly on analysis of textual content and for detecting live streaming videos where a user may be committing suicide; both of which are then marked to human moderators to take further actions

like alerting local authorities for help. As per our knowledge such system exists only for Facebook where the individual posts and live streams are analysed independently i.e. on a per post basis with the goal of preventing suicides. However we believe continuous analysis of user multimodal posts over an elongated time period is necessary to detect the early signs of depression which may cause suicide at a late stage. The system must learn and update user's mental well being profile continuously and discover the changes in his online communication, mood and behaviour. Prioritization is also required so as the detect cases based on severity and those in immediate danger. Real time actions based on priority can then be triggered so that appropriate timely help can be extended. One example of trigger can be alerting the family members listed by the user in his network and the close friends with whom he does maximum check-ins so that they can reach out to him with immediate help in the real world. In other scenarios where immediate action may not be required, trigger actions can be: showing user self help content and pages, connecting to the user with a chat bot to converse and provide psychological assistance; such chat bots may be designed with the help of medical professionals and psychologists. To meet these objectives, we feel a framework to analyse the mental well being of a user should be deployed over social media platforms since they have become a quintessential part of day to day communication and expression. The framework must account the multiple ways the user may express his emotions, feelings and thoughts i.e. through images, videos, text, emojis, stickers etc. and all of these features must be utilized to monitor user's mental state over time. Our proposed framework has been designed to meet the expectations discussed above from such a system that detects onset of depression to prevent suicidal or self harming behavior. The salient contributions of the framework being: fusing multimodal user generated content i.e. text, images, emoticons and videos from user's social media posts and obtaining joint representations using deep learning based vectorization techniques, using deep learning based algorithms to classify posts with the objective to detect depression and suicidal behavior related posts. The proposed framework when implement and integrated by social media platforms can work in real time to the detect the onset of depression which may help in preventing suicides and other self harming actions. However, it is important to understand that the system's accuracy shall depend on user's willingness to allow analysis of his entire content over an elongated period of time. The system when deployed and integrated with various social media platforms must work in a non-pervasive manner to discover changes in user communication, behavior and

mood and must not violate the privacy concerns of a user.

References

- [1] REECE AG and DANFORTH CM. *Instagram photos reveal predictive markers of depression*. EPJ Data Science. 2017 Dec;6(1):1-2.
- [2] PARK M, CHA C and CHA M. *Depressive moods of users portrayed in Twitter*. InProceedings of the ACM SIGKDD Workshop on healthcare informatics (HI-KDD) 2012 Aug 12 (Vol. 2012, pp. 1-8).
- [3] PARK M, McDONALD DW and CHA M. *Perception differences between the depressed and non-depressed users in twitter*. InSeventh International AAAI Conference on Weblogs and Social Media 2013 Jun 28.
- [4] PAUL MJ and DREDZE M. *You are what you tweet: Analyzing twitter for public health*. InFifth International AAAI Conference on Weblogs and Social Media 2011 Jul 5.
- [5] MORENO MA, JELENCHICK LA, EGAN KG, COX E, YOUNG H, GANNON KE and BECKER T. *Feeling bad on Facebook: Depression disclosures by college students on a social networking site*. Depression and anxiety. 2011 Jun;28(6):447-55.
- [6] GOSLING SD, AUGUSTINE AA, VAZIRE S, HOLTZMAN N and GADDIS S. *Manifestations of personality in online social networks: Self-reported Facebook-related behaviors and observable profile information*. Cyberpsychology, Behavior, and Social Networking. 2011 Sep 1;14(9):483-8.
- [7] DESMET B and HOSTE V. *Emotion detection in suicide notes*. Expert Systems with Applications. 2013 Nov 15;40(16):6351-8.
- [8] PAUL MJ and DREDZE M. *Discovering health topics in social media using topic models*. PloS one. 2014;9(8).
- [9] BOYCHUK V, SUKHAREV K, VOLOSHIN DV and KARBOVSKII VA. *An Exploratory Sentiment and Facial Expressions Analysis of Data from Photo-sharing on Social Media: The Case of Football Violence*. InICCS 2016 Jan 1 (pp. 398-406).
- [10] DE CHOUDHURY M, GAMON M, COUNTS S and HORVITZ E. *Predicting depression via social media*. InSeventh international AAAI conference on weblogs and social media 2013 Jun 28.
- [11] DE CHOUDHURY M, COUNTS S and HORVITZ E. *Social media as a measurement tool of depression in populations*. InProceedings of the 5th Annual ACM Web Science Conference 2013 May 2 (pp. 47-56).
- [12] MAJUMDER N, PORIA S, GELBUKH A and CAMBRIA E. *Deep learning-based document modeling for personality detection from text*. IEEE Intelligent Systems. 2017 Mar 27;32(2):74-9.
- [13] GÜÇLÜTÜRK Y, GÜÇLÜ U, VAN GERVEN MA and VAN LIER R. *Deep impression: Audiovisual deep residual networks for multimodal apparent personality trait recognition*. InEuropean Conference on Computer Vision 2016 Oct 8 (pp. 349-358). Springer, Cham.
- [14] LIN H, JIA J, GUO Q, XUE Y, LI Q, HUANG J, CAI L, and FENG L. *User-level psychological stress detection from social media using deep neural network*. InProceedings of the 22nd ACM international conference on Multimedia 2014 Nov 3 (pp. 507-516).