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 20^{th} 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 20^{th} frame, the coordinates of the middle 10^{th} 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:

$$DepressionorSuicidal = \frac{\#sadEmoticons}{\#totalEmoticons} \tag{1}$$

$$NormalorHealthy = \frac{\#happyEmoticons}{\#totalEmoticons} \tag{2}$$

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).