IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2347-6710

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Special Issue 2, April 2023

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

|| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | SJIF: 8.423|| Volume 12, Special Issue 2, April 2023 ||

8th National Conference on Advanced Computing Technologies (NCACT'23)

Organized by

Department of CSE, Velammal Engineering College, Chennnai, India

Investigation of Speech, Text, and Video Patterns for Depression Detection Using Machine Learning

Charumathi K[1], Nandhini S[1], Shivani A[1], Dr. P. Pritto Paul[2],

U.G. Student, Department of Computer Engineering, Velammal Engineering College, Chennai, Tamil Nadu, India

Associate Professor, Department of Computer Engineering, Velammal Engineering College, Chennai, Tamil Nadu, India

ABSTRACT:The massive and growing burden assessed on ultramodern society by depression has motivated examinations into early discovery through automated, scalable, and non-invasive styles, including those grounded on speech, video and text. Still, speech-grounded styles that capture articulatory information effectively across different recording biases and in natural surroundings are demanded. Here we presents a new multi-level attention-grounded network for multi-modal depression vaticination that fuses features from audio, video, and text modalities while learning the intra and inter-modality applicability. The multi-level attention reinforces overall literacy by opting for the most influential features within each modality for the decision timber. We perform a total trial to produce different retrogression models for audio, video, and text modalities. Evaluations of both landmark duration features and landmark n-gram features on the DAIC-WOZ and SH2 datasets show that they're largely effective, either alone or fused, relative to being approached.

KEYWORDS:Depression detection, Machine learning,Convolutional Neural Networks Algorithm, HAAR Cascade Algorithm, Support Vector Machine Algorithm**.**

I.INDRODUCTION

Depression, a major internal complaint reported to affect 10- 15 of the world's population, places severe health, security, productivity, and profitability burdens on ultramodern society. Beforehand discovery and treatment of depression can help relieve this profitable burden while adding to the productivity and quality of life of depressed individuals. still, treatment of depression is precious and frequently delayed due to the failure of trained cerebral clinicians and frequently the late opinion of internal complaint symptoms. likewise, the cost of early discovery by either spot or large-scale webbing is prohibitive due to the forenamed reasons. thus, indispensable technology-grounded webbing styles have been sought in the form of affordable, automatic systems, to grease large-scale early discovery and connect with timely intervention. The lack of effective depression webbing seeker technologies has attracted exploration attention for further than a decade. To date, there have been several studies on the automatic discovery of depression ranging from voice, facial video, EEG signals, head disguise, eye aspect, etc. Among these modalities, video, text, and speech, which have demonstrated promising effectiveness and effectiveness as an index of depression, remain especially non-invasive and fluently accessible. In this design, we present a new frame that invokes an attention medium at several layers to identify and prize important features from different modalities to prognosticate the position of depression. The network uses several low-position and mid-level features from audio, text, and video modalities. still, utmost studies to date on speech-grounded depression[2] discovery have primarily concentrated on laboratory-collected data, recorded from a single channel in a clean terrain. The adding relinquishment of smartphones coupled with the emergence of voice give unknown openings for new automated medical webbing styles through testing the mortal voice the capability to accumulate a sufficiently large volume of data to statistically model variations in speech patterns for depressed and non-depressed individualities across populations and audio recording bias type; and the capability to administer collectively acclimatized questionnaires, dissect voice samples and give clinical webbing feedback across large populations. Still, conventional features developed from the clean lab-grounded datasets may not generalize as well in real-world operations due to the dramatic differences in speech recording similar to noise conditions, handset tackle, and design protocols. This failure motivates the design of a new order of effective features for detecting depression in both surroundings. Although the essential relationship between verbal content and internal illness

|| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | SJIF: 8.423|| Volume 12, Special Issue 2, April 2023 ||

8th National Conference on Advanced Computing Technologies (NCACT'23)

Organized by

Department of CSE, Velammal Engineering College, Chennnai, India

position is more prominent, the visual features also play a vital part to reinstate the deep association of depression to facial feelings[1]. It has been observed that cases suffering from depression frequently have distorted facial expressions. g. eyebrow shuddering, dull smiles, lowering faces, aggressive aesthetics, confined lip movements, reduced eye blinks, etc. With the amount of proliferating video data from cameras in wearables and surveillance sectors, assaying facial feelings and sentiments[4] is the growing trend amongst the vision community.

II. RELATED WORK

The success obtained in face detection and recognition over the last decade of research is more, but the analysis of facial traits still remains a trending topic. Soft biometric traits, i.e. singular facial traits like the mouth, the nose, the hair, and so on, are considered an important field of investigation. In [9], an unsupervised clustering approachconsists of a neural network model for face attribute recognition based on transfer learning whose goal is grouping faces according to common facial features. It makes use of the features present in each cluster to provide a concise and comprehensive description of the faces belonging to each cluster and deep learning as a means for task prediction even in partially visible faces. In [10], real-life scenarios speech emotions are also important, for this purpose a sensor based on SVM verbal/ verbal sound was developed. A Prosodic Expression (PPh) automatic tagger was further employed to get the verbal/ verbal parts. For each member, the emotion and sound features were independently derived on basis of convolutional neural networks(CNNs)[5] and also concatenated to form a CNN-based general trait vector. Eventually, a sequence of CNN-based trait vectors for an entire dialog turn was fed to an attentive long short-term memory (LSTM) model to display an emotional sequence as a recognition result. Experimental results on the recognition of seven emotional traits in the NNIME (The NTHU- NTUA Chinese interactive multimodal emotion corpus) showed that the proposed system achieved a discovery delicacy of 52 percent outperforming the traditional styles. Analysis of landmark n-gram features and landmark duration features on the DAIC-WOZ and SH2 datasets show that they are highly efficient, either alone or when combined. Current speech processing methods basically cut the speech into short 10-20 millisecond frames before extracting. Lately, landmark-based traits display the discriminative information for speech-based depression detection, mainly using relatively simple counts of consecutive landmark Diagrams.

III. METHODOLOGY

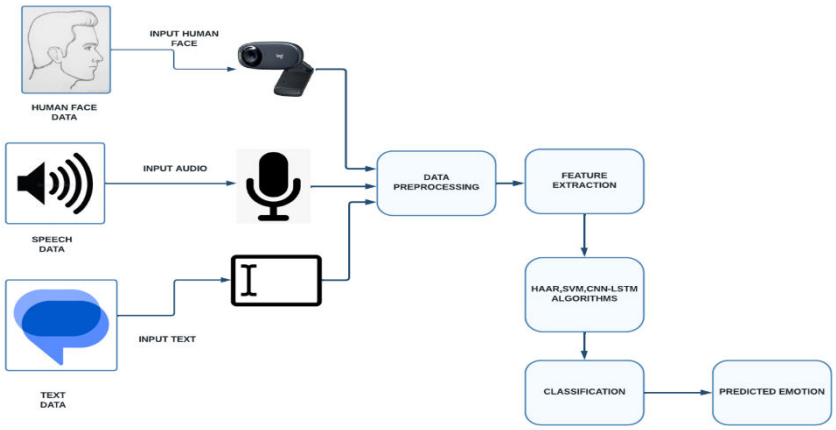
The huge need for depression discovery and the challenges[4] involved motivated the affective computing exploration community to use behavioral cues to learn to prognosticate depression, Post-traumatic stress complaints, and related internal diseases. Behavioralsimilar to facial recognition, and prosodic features from speech have proven to be excellent features for depression. In this design, we present a new frame that invokes attention medium at several layers to identify and prize important features from different modalities to prognosticate the position of depression. The network uses many low-position and mid-level features from text, audio, and video modalities. We show that attention to different situations gives us the rate of the significance of each point and modality, leading to better results. We perform several trials on each point from different modalities and combine several modalities. More precisely, we propose duration-grounded features, which are statistics calculated from two types of bigrams. To manage the challenges of changingdepression-related features[4], especially for demoralized recording conditions and different smartphones, herein we proposed two new sets of features grounded on speech milestones, which delivered promising results on two dramatically different datasets.

|| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | SJIF: 8.423|| Volume 12, Special Issue 2, April 2023 ||

8th National Conference on Advanced Computing Technologies (NCACT'23)

Organized by

Department of CSE, Velammal Engineering College, Chennnai, India

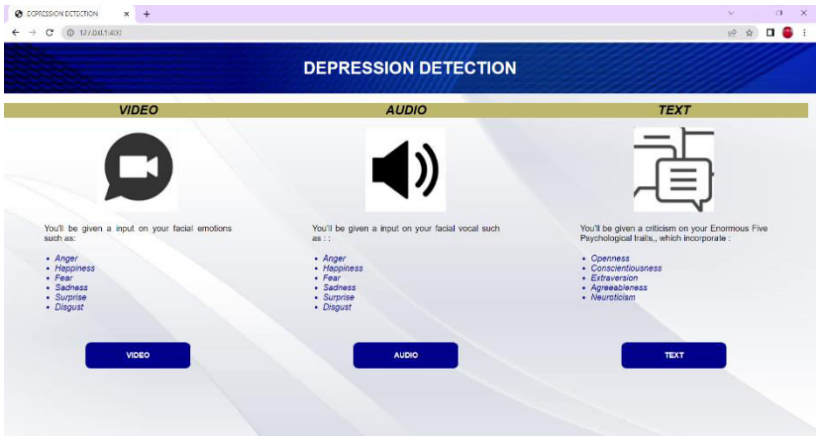


PROS:

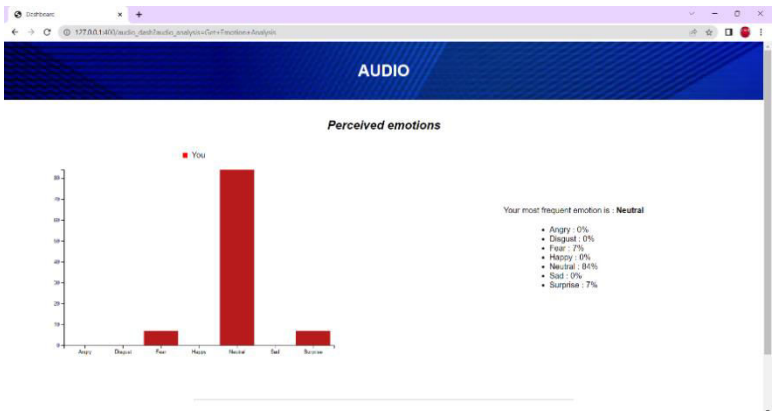
Overall, the proposed novel landmark-based features are very promising for depression detection not only for speech collected in a clean environment and a single channel but also for speech collected in the noisy environments from mobiles.

More sophisticated back-end classifiers can be examined to exploit the proposed features for improved performances.

IV. EXPERIMENTAL RESULTS



Fig(a)



Fig(b)

Fig(b) shows the result of depression detection based on audio

|| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | SJIF: 8.423|| Volume 12, Special Issue 2, April 2023 ||

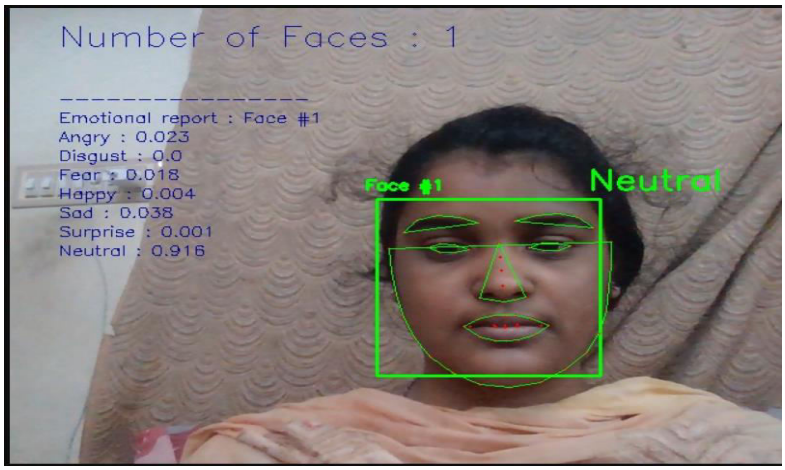
8th National Conference on Advanced Computing Technologies (NCACT'23)

Organized by

Department of CSE, Velammal Engineering College, Chennnai, India

Fig(c)

Fig(c) shows the result of depression detection based on text



Fig(d)

Fig (d) shows the result of depression detection based on video

V. CONCLUSION

Amulti-level attention grounded early emulsion network which fuses audio, video, and text modalities to prognosticate inflexibility of depression. For this task, we observed that the attention network gave the loftiest weights to the text modality and nearly equal weightage to audio and video modalities. The use of multi-level attention led us to gain significantly better results in all individual and emulsion models compared to both the birth and state art. Using attention over each point and each modality had a twofold advantage overall. Originally, this gives us a deep and better understanding of the significance of each point within a modality towards depression vaticination. Further, attention reduced the network's overall computational complexity and also , the training and test time.

REFERENCES

[1]. M.-I. Georgescu, R. T. Ionescu, and M. Popescu, ''Local learning with deep and handcraftedfeatures for facial expression recognition,'' IEEE Access, vol. 7, pp. 64827–64836, 2019.
[2]. K.-Y. Huang, C.-H. Wu, Q.-B. Hong, M.-H. Su, and Y.-H. Chen, ''Speech emotionrecognition using deep neural network considering verbal and nonverbal speech sounds,'' inProc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), May 2019, pp. 5866–5870.

|| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | SJIF: 8.423|| Volume 12, Special Issue 2, April 2023 ||

8th National Conference on Advanced Computing Technologies (NCACT'23)

Organized by

Department of CSE, Velammal Engineering College, Chennnai, India

[3]. D. Kollias, P. Tzirakis, M. A. Nicolaou, A. Papaioannou, G. Zhao, B. Schuller, I. Kotsia, andS. Zafeiriou, ''Deep affect prediction inthe-wild: Aff-wild database and challenge, deeparchitectures, and beyond,'' Int. J. Comput. Vis., vol. 127, pp. 1–23, Jun. 2019.

[4]. D. Kollias, A. Schulc, E. Hajiyev, and S. Zafeiriou, ''Analysing affective behavior in thefirst ABAW 2020 competition,'' 2020.

[5]. Z. Du, S. Wu, D. Huang, W. Li, and Y. Wang, ''Spatio-temporal encoderdecoder fullyconvolutional network for video-based dimensional emotion recognition,'' IEEE Trans. Affect.Comput., early access, Sep. 10, 2019.

[6]. P. Barros, N. Churamani, and A. Sciutti, ''The FaceChannel: A light-weight deep neuralnetwork for facial expression recognition,'' in Proc. 15th IEEE Int. Conf. Autom. Face GestureRecognit. (FG), Buenos Aires, AR, USA, Apr. 2020.

[7]. M. Koujan, L. Alharbawee, G. Giannakakis, N. Pugeault, and A. Roussos,''Real-time facialexpression recognition 'in the wild' by disentangling 3D expression from identity,'' in Proc. 15thIEEE Int. Conf. Autom. Face Gesture Recognit. (FG), Buenos Aires, AR, USA, May 2020.

[8]. G. Zhao, H. Yang, and M. Yu, ''Expression recognition method based on a lightweightconvolutional neural network,'' IEEE Access, vol. 8, pp. 38528–38537, 2020.

[9]. A. F. Abate, P. Barra, S. Barra, C. Molinari, M. Nappi, and F. Narducci, ''Clustering facialattributes: Narrowing the path from soft to hard biometrics,'' IEEE Access, vol. 8, pp.9037–9045, 2020.

[10]. H. Wang and S. Hou, ''Facial expression recognition based on the fusion of CNN and SIFTfeatures,'' in Proc. IEEE 10th Int. Conf. Electron. Inf. Emergency Commun. (ICEIEC),2020.

INNO SPACE
SJIF Scientific Journal Impact Factor

SJIF: 8.423

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INTERNATIONAL CENTRE

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details