International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 05 Issue: 05 | May - 2021 ISSN: 2582-3930

Depression Detection Using Convolutional Neural Network From speech

[1] Gautam Uppal, [2] Rishabh Tyagi

Computer Science Engineering Department

SRM Institute of Science and Technology, NCR Campus

ABSTRACT

Early identification and treatment of misery is fundamental in advancing abatement, forestalling backslide, and diminishing the enthusiastic weight of the illness. Current determinations are fundamentally abstract, conflicting across experts, and costly for people who might be in earnest need of help. The affective computing community and the A.I. field have shown a developing interest in arranging programmed frameworks. The speech features have helpful data for the analysis of sadness. In any case, physically arranging and area information square measure still important for the decision of the component, that makes the strategy work exceptional and abstract. As of late, deep-learned component upheld neural organizations have shown better execution than hand-crafted choices in multiple zones. This paper proposes a novel way to deal with computerized depression recognition in audio utilizing Convolutional Neural Network (CNN) and multipart intelligent preparing. Trials led on Audio Video Emotion Challenge 2013 and 2014 depression data sets show that our methodology is hearty and successful for the conclusion of sorrow when contrasted with cutting edge sound based techniques. In evaluation, data were applied to residual CNNs in the form of spectrograms—images auto-generated from audio samples. In explore directed, information were applied to remaining CNNs as spectrograms—pictures auto-produced from sound examples.

INTRODUCTION

As per the World Health Organization (WHO), more than 256 million people are measurable to experience the ill effects of depression [1], which assortment is universally developing, especially at a muddled age. Logically alluded to as Major depression disorder(MDD), depression could be an issue portrayed by an infrequent state of mind, low vanity, loss of interest, low energy, partner in nursing torment while not a straightforward reason for an extended measure of your time. It's adverse effects on an individual's family, work, dozing, and admission propensities. In outrageous cases, as reportable by [2], 1/2 all finished suicides are related with burdensome and elective mind-set issues. Because of that, few analysts have focused on creating frameworks to analyze and prevent this problem to help specialists and clinicians to help patients as by and by as achievable.

Lately, some AI techniques are arranged using sound signals for depression investigation. In the interim, there is an abundance of examination, that recommends that voice patterns have an inside and out relationship with feeling and stress. Handsewn choices are attempted to get better for assessing depression seriousness. Be that as it may, there region unit a few constraints of handcrafted alternatives for depression scale forecast. To style handsewn alternatives needs stores of exertion (i.e., space information, work and time, and so on) for example, MFCCs territory unit wide utilised in programmed discourse and loud-speaker acknowledgment errands. Be that as it may, in the event that we tend to

planned handsewn alternatives like MFCCs, we ought to consistently have task-explicit information of depression and to gather such information is long. At last, it's difficult to choose partner degree satisfactory tool stash to separate the alternatives. Various reachable tool stash region unit wide wont to remove low-level alternatives, as open-source Speech and Music Interpretation by Large-space Extraction, Cooperative Voice Analysis Repository for Speech Technologies, Speech Signal Processing Toolkit, KALDI, Yet Another Audio Feature Extractor, and Open Emotion and Recognition Toolkit.
Through AI point of view, depression investigation is a relapse or grouping issue (e.g., in Audio Video Emotion Challenge 2013 [5] and Audio Video Emotion Challenge 2014 [6]). We will probably anticipate the depression score called BDI–II of a person from acquired sound. As in the DCC sub-challenge, the English-speakers information base DAIC-WOZ is utilized for the framework assessment.

DATA SET

All audio recordings and associated depression metrics were provided by the DAIC-WOZ Database, which was com-piled by USC Institute of Creative Technologies and released as part of the 2016 Audio/Visual Emotional Challenge and Workshop (AVEC 2016). The dataset consists of 189 sessions, averaging 16 minutes, between a participant and virtual interviewer called Ellie (Figure 1), controlled by a human interviewer in another room via a \Wizard of Oz" approach. Prior to the interview, each participant completed a psychiatric questionnaire (PHQ-8), from which a binary\truth" classification (depressed, not depressed) was derived.

LITERATURE REVIEW

Different depression acknowledgment approaches have been proposed in the DSC of the audio video workshop.

1. In the Audio Video Emotion Challenge 2013 workshop, Williamson [9] embraced the mix of eigenvalue spectra and coordination highlights to break down the connection between the audio practices and depression value. They planned a Gaussian flight of stairs relapse framework to foresee the Beck Depression Index–II scores for every sound information. Principal Component Analysis is additionally utilized for measurement decrease. At long last, the creators gave the base presentation on the test sets with 7.42 as RMSE and 5.75 as MAE.

2. Moore[10] investigated prosodics, the audio lot, and boundaries removed straightforwardly from the outpur signal to segregate the discouraged discourse. They removed around two hundred prosodics, audio parcel, and output signal and made an interpretation of them into two thousand insights for study.

3. Nicholas[11] given an extensive and thorough decision about the evaluation and conclusion of the depression and the self destruction. They evaluated the significant qualities of paralinguistic discourse influenced by depression and self destruction. They broke down the patterns which were utilized in grouping and relapse issues. At long last, they gave a top to bottom conversation about the current constraints and difficulties.

4. In [12], [13], the creators examined the connection between vocal prosody and change in depression seriousness over the long haul. They introduced three theories: (1) Naive audience members can recognize the discouraged members and wellbeing controls from vocal chronicles;

(2) the quantitative highlights of vocal prosody can catch changes from the determination of the depression; and (3) relational connections can likewise happened in the seriousness of depression assessment system. At last, they approved the speculations by tests. The outcomes show that the investigation of audio samples is an important device for depression examination.

5. Williamson[14] investigated the connection and reciprocal attributes by separating highlights from the discourse source, framework, and prosody. They melded the distinctive element space to get a superior exhibition. At last, they consolidated Gaussian flight of stairs relapse with ELM classifiers, and get a 8.12 as RMSE.

6. A ton of works (Malam et al., 2017; Trotzek et al., 2017; Sadeque et al., 2017; Almeida et al., 2017) use dictionary includes that show utilization of explicit words like enthusiastic words, estimation words, and explicit terms related for medication or diagnosis. This words generally are taken from pre-characterized word references.

TOOLS USED

1. **NumPy** :- NumPy is a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays.

2. **pyAudioAnalysis** :- pyAudio Analysis, an open-source Python library that provides a wide range of audio analysis procedures including: feature extraction, classification of audio signals, supervised and unsupervised segmentation and content visualization

3. **SciPy** :- SciPy is a free and open-source Python library used for scientific computing and technical computing. SciPy contains modules for optimization, linear algebra, integration, signal and image processing.

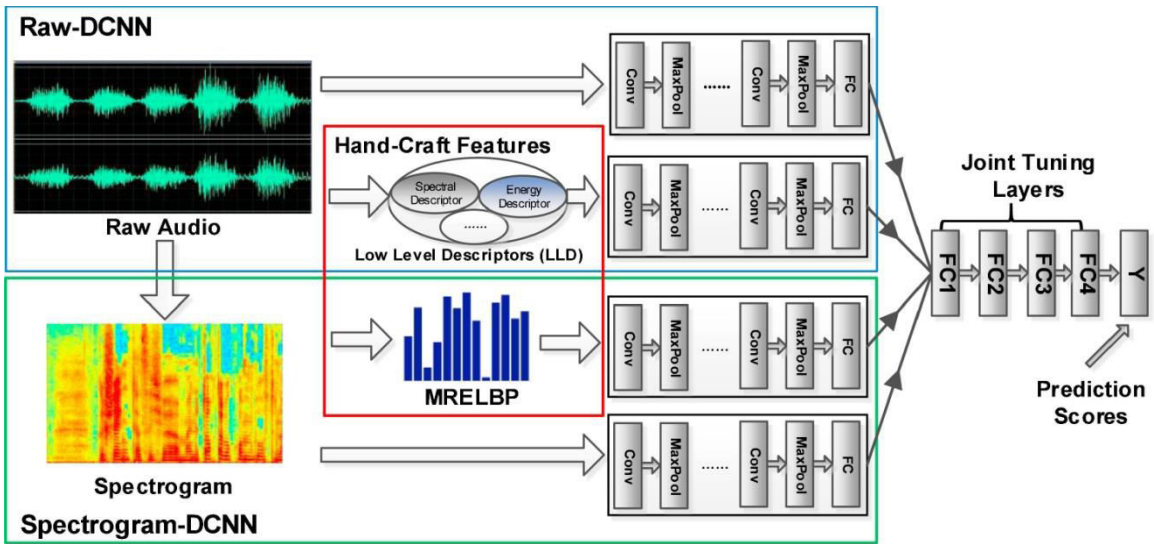
4. **Scikit-learn** :- Scikit-learn is a free software machine learning library for the Python programming language. It features various classification, regression and clustering algorithms including support vector machines.
5. **Matplotlib** :- Matplotlib is a plotting library for the Python programming language. It provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits like Tkinter, wxPython, Qt, or GTK.

6. **CNN** :- A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other.

ARCHITECTURE

Feauture style or feauture extraction assumes a vital part in depression investigation assignments. During this work we will in general blend hand-crafted alternatives in with DL choices for assessing the seriousness of depression. For hand-lingered choices, we will in general concentrate the LLD from the

crude brief snippets and MRELBP choices from the spectrograph of sound. Next we will in general utilize DCNN to straightforwardly take in the deep-learned alternatives from the crude sound and pic pictures. At long last, we will in general depict the arranged joint adjusting philosophy to blend the 4 streams for a definitive depression forecast. The proposed structure for programmed depression acknowledgment is shown below.



The above figure is the framework for automatic Depression recognition, where Raw-DCNN (Top) takes raw audio signals and low level descriptors (LLD) as input, while the Spectrogram-DCNN (Bottom) uses texture features as input. The red box in above figure is Hand-Crafted features. Other two arrows are Deep-Learned features. The predicted depression score is computed by aggregating or averaging the individual predictions per frame from four DCNNs.

WORK DONE

Audio Splitting

The data-set contains 192 audio sessions between a animated virtual interviewer called Ellie, and the participant. The features of audio segments of the participants are useful for classification, the segments are split by silence removal and then separated by speaker diarization. these audio segments are of varying lengths such that there is spread in data.

Data Imbalance

In the data-set, the number of non-depressed subjects is about four times larger than that of depressed ones, which can introduce a classification ”non-depressed” bias. Additional bias can occur due to the considerable range of interview durations from 7-33 minutes because a larger volume of signal from an

individual may emphasize some characteristics that are person specific.To rectify this imbalance, audio segments are randomly sampled in equal numbers.

Spectrogram Conversion

The sampled audio segments are then converted to spectrogram images of size 512 X 512 pixels. These images are put into different folders corresponding different classes and then the folders are split into training and validation data with ratio 8:2.

Image Preprocessing

These images are converted to TensorFlow tensor using flow from directory of image processing datagenrator method. The imae tensor is normalised and fed into the Convolutional Neural Network.

Convolutional Neural Network Architecture

The Convolutional Neural Network consists of six layers:

2D Convolutional layer 1:

```
Input= 512X512X3 image tensor
Activation = ReLU
No of filters = 32
Filter size = 5X
Strides = 0
Padding = 0
```

MaxPooling layer :

```
Pooling size = 4X
Strides = 4
```

2D Convolutional layer 2 :

```
Activation = ReLU
No of filters = 32
Filter size = 3X
L2 regulariser (0.01)
```

MaxPooling layer :

```
Pooling size = 1X
Strides = 1,
```

Flatten layer

Dense layer 1 :

```
Nodes = 128
Activation = linear
L2 regulariser (0.01)
```

Dropout layer :

```
Dropout = 60%
```

Dense layer 2 :

```
Nodes = 256
Activation = ReLU
L2 regulariser (0.001)
```

Dropout layer :

```
Dropout = 80%
```

Dense layer 3 : (output)

```
Node = 1
Activation = Sigmoid
```

CONCLUSION

Depression might be a genuine mental unsettling influence. Computer helped innovations are examined to help therapists inside the appraisal of sadness volume. To support the precision of programmed depression recognition from discourse gestures, we will in general make a venture on an approach which is totally founded on DL and traditional technique, that we tend to used to beat the challenges brought about by planning the hand-created highlights for depression recognition. Inside the projected technique, we will in general utilize the crude and spectrograph DCNN to show the trademark data of discouragement. Also, we tend to furthermore projected to receive joint normalization dimensions, to blend the crude and spectrograph Deep CNN, which may enhance the presentation of sorrow acknowledgment. Exploratory outcomes on 2 depression data sets i.e, Audio-Visual Emotion Challenge 2013 and 2014, showed that the methods got prevalent execution contrasted and different sound based methodologies for depression acknowledgment.

FUTURE ENHANCEMENTS

In future work, we will assess the presentation of our technique for an alternate arrangement of sound example sizes (10, 20, 30, 40 sec, and 1.8 min) to guarantee greater equivalence of outcomes with different specialists by using Interspeech gathering [12]. Moreover, there exits further developed alternatives which can change over sound accounts into pictures that could enhance the grouping standard further by separating additional highlights by a similar size sound examples. As referenced on top of, there's conjointly the decision of utilizing our strategy as a structure block for a great deal of complicated, hybrid, or multimodal depression detection framework that would target the score of PHQ-8. We will likewise lead all future examinations on a fixed, adjusted test and training dataset.

REFERENCES

[1] World Health Organization. Depression and Other Common Mental Disorders: Global Health Estimates; Technical Report; World Health Organization: Geneva, Switzerland, 2017. [Google Scholar]

[2] Bachmann, S. Epidemiology of suicide and the psychiatric perspective. *Int. J. Environ. Res. Public Health* **2018**, *15*, 1425. [Google Scholar] [CrossRef]

[3] Y. Bengio, A. Courville, P. Vincent **Representation learning: a review and new perspectives** IEEE Trans. Pattern Anal. Mach. Intell. (2013)

[4] L.G. Hafemann, L.S. Oliveira, P. Cavalin, Forest species recognition using deep convolutional neural networks, in: 2014 22nd International Conference on Pattern Recognition (ICPR), IEEE, 2014

[5] M. Valstar, B. Schuller, K. Smith, F. Eyben, B. Jiang, S. Bilakhia, S. Schnieder, R. Cowie, M. Pantic **Avec 2013: the continuous audio/visual emotion and depression recognition challenge** Proceedings of the 3rd ACM International Workshop on Audio/Visual Emotion Challenge, ACM (2013).

[6]M. Valstar, B. Schuller, K. Smith, T. Almaev, F. Eyben, J. Krajewski, R. Cowie, M. Pantic **Avec 2014: 3d dimensional affect and depression recognition challenge** Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge, ACM (2014).

[7] M. Valstar, J. Gratch, B. Schuller, F. Ringeval, D. Lalanne, M. Torres, S. Scherer, G. Stratou, R. Cowie, M. Pantic **Avec 2016: Depression, mood, and emotion recognition workshop and challenge** Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge, ACM (2016).

[8] F. Ringeval, B. Schuller, M. Valstar, *et al.* **AVEC 2017: Real-life depression, and affect recognition workshop and challenge** Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge, ACM (2017).

[9] J.R. Williamson, T.F. Quatieri, B.S. Helfer, R. Horwitz, B. Yu, D.D. Mehta **Vocal biomarkers of depression based on motor incoordination** Proceedings of the 3rd ACM International Workshop on Audio/Visual Emotion Challenge, ACM (2013).

[10] I. Moore, Elliot, M.A. Clements, J.W. Peifer, L. Weisser **Critical analysis of the impact of glottal features in the classification of clinical depression in speech** IEEE Trans. Bio-Med. Eng., 55.

[11] N. Cummins, S. Scherer, J. Krajewski, S. Schnieder, J. Epps, T.F. Quatieri **A review of depression and suicide risk assessment using speech analysis** Speech Commun., 71.

[12] J.F. Cohn, T.S. Kruez, I. Matthews, Y. Yang, M.H. Nguyen, M.T. Padilla, F. Zhou, F. De, la Torre, Detecting depression from facial actions and vocal prosody, in: International Conference on Affective Computing and Intelligent Interaction and Workshops, 2009.

[13] Y. Yang, C. Fairbairn, J.F. Cohn **Detecting depression severity from vocal prosody** IEEE Trans. Affect. Comput., 4 (2) (2013).

[14] J.R. Williamson, T.F. Quatieri, B.S. Helfer, G. Ciccarelli, D.D. Mehta **Vocal and facial biomarkers of depression based on motor incoordination and timing** Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge, ACM (2014).