Journal of Population Therapeutics & Clinical Pharmacology

RESEARCH ARTICLE
DOI: 10.47750/jptcp.2023.30.09.032

Deperssion Detection in Naturalistic Environmental Condition

T R Dineshkumar[1*], U Savitha[2], T. Haritha vellam[3], R. Krishika[4], A.Juneha jabeen[5], Aravind Ramesh[6], Purushothaman[7]

[1,2]Assistant Professor, Department of Electronics and Communication Engineering, Vel Tech High Tech Dr Rangarajan Dr Sakunthala Engineering College, Chennai, India

[3,4,5,6,7]UG Student Department of Electronics and Communication Engineering, Vel Tech High Tech Dr Rangarajan Dr Sakunthala Engineering College, Chennai, India

***Corresponding author:** T R Dineshkumar, Assistant Professor, Department of Electronics and Communication Engineering, Vel Tech High Tech Dr Rangarajan Dr Sakunthala Engineering College, Chennai, India, Email: trdineshkumar@velhightech.com

Submitted: 10 February 2023; Accepted: 16 March 2023; Published: 07 April 2023

ABSTRACT

In light of depression's massive and growing burden on modern society, researchers are investigating ways to detect depression early using non-invasive, automated, scalable, and non-invasive methods. In naturalistic environments, though, speech-based methods are still needed for effective articulatory information capture. The intermodality and intramodality affinities are learned by integrating information from audio, video, and text modalities in a multimodal depression prediction model this article. As a result of The most important components of each modality in the multilevel attention are selected to be used for decision-making to reinforce the overall learning. Our study aims to build various regression models using inputs from audio, video, and text landmark duration and historic n-gram features on the DAIC _WOZ and SH2 datasets both performed well, either separately or in combination.

Keywords: *depression, naturalistic environments*

INTRODUCTION

Approximately 15-30% of the world's population suffers from depression, a serious mental disorder that imposes a heavy burden on society's health, security, productivity, and economy.

It is possible to reduce the burden of depression on the economy by detecting and treating depression early while increasing depressed individuals' productivity and quality of life. Treatment for depression, on the other hand, is costly and frequently delayed because there is a lack of qualified psychological clinicians and because symptoms of mental disorders are typically diagnosed too late.

Furthermore, due to the aforementioned reasons, The expense of early detection via spot or Costs associated with thorough screening are exorbitant. In order to promote widespread early identification and link with prompt action another to traditional technology-based screening techniques, automated devices have been sought after.For more than a decade, researchers have been focused on the dearth of efficient methods for depression detection. Many studies on utilising voice, facial video, EEG data, leader attitude, and eye gazing for the automatic detection of depression, and other signals have been conducted.

and other methods. Among these modes, video, text, and speech have shown Being non-invasive and generally available, efficiency as a depression predictor continues to show promise. In this study, we provide a unique framework that uses attention mechanisms at many levels to determine and extract key characteristics from several modalities estimating the depression's intensity. The audio, text, and video modalities each have a large number of low- and mid-level characteristics. are used by the network. Nevertheless, most investigations The majority of studies on speech-based depression diagnosis to date have employed laboratory-collected data that was captured from a single channel in a sterile setting. By sampling the human voice and gathering enough information to quantitatively model variations in patients' speech patterns, voice assistants and the rising use of smartphones offer previously unheard-of prospects for new automated medical screening techniques. and individuals who are not depressed across demographics and audio recording device types; the ability to offer personalised surveys, analyse speech clinical screening feedback across samples and enormous populations. Unfortunately, because to major variations in voice recording, such as noise situations, handset hard limits, and common traits may not be as useful in the actual world as they are in pure lab-based datasets. Applications and design techniques. This flaw stimulates the creation of a new class of useful characteristics for recognising depression in both scenarios. Visual cues also significantly contribute to restoring the strong connection between melancholy and facing emotions, even if the fundamental relationship between language content and the severity of mental sickness is more clear. Depression patients are known to regularly display skewed behaviour.

RELATED WORKS
LITERATURE SURVEY

Paper 1

Title: Closing the Gap Between Soft and Hard Biometrics Via Clustering of Face Characteristics.

Silvio Barra, Fabio Narducci, Michele Nappi, Andrea F. Abate, Paola Barra, And Silvio Barra Are The Authors.

Year:2020

Abstract

Notwithstanding the progress made in face recognition , identification more than previous 10 ages of study, facial attribute analyses remains a hot issue. Leaving comprehensive Besides facial recognition, looking into the possibility of flexible biometric features, unique facial attributes such as the nose, lips, and hair and, is still regarded an lucrative topic study. The capacity to determine a face's identity when it is hidden. whether ability to identify sunglasses were purposely obstructing or The method described in this paper proposes using unsupervised clustering and neural network models for recognizing facial attributes to group faces based on common features. This approach can be useful in situations where user participation cannot be assumed, such as in forensic investigations where partial fingerprints or incomplete face photographs may be available. The proposed method relies on transfer learning, which involves using pre-trained neural network models to extract relevant features from facial images. These features are then used to cluster similar faces together. The characteristics gathered in each cluster are then used to provide a concise and thorough description of the faces in each group. Additionally, the paper discusses the use of deep learning for task prediction in partially visible faces. This involves using neural networks to predict facial features or attributes based on incomplete or partially visible images. Overall, the proposed method can provide valuable insights and assistance in situations where traditional methods of facial recognition may not be possible or effective.

Paper 2

The Integration OF Cnn AND Sift Features FOR Face Expression Recognition.
Author: Siyang Hou, Huibai Wang.

Year:2020

Abstract

Methods are combined to achieve a more comprehensive and accurate recognition of facial expressions.The proposed CNN-SIFT fusion algorithm involves using a custom CNN network with an Inception module to extract global facial expression information efficiently. The Inception module involves adding 11 convolutional layers, which allows for more efficient use of computing resources.Additionally, the paper discusses using before obtaining SIFT features, use cascade regression to calibrate face structure points.. This helps to focus the key points on the areas of the face that contribute most to expression, resulting in more accurate feature extraction.By combining these two methods, the proposed algorithm achieves a more comprehensive and accurate recognition of facial expressions. This has significant uses in security systems, human-computer interaction, and emotion identification, among other fields.

Over all, the study makes a significant addition to the fields of face expression identification and computer vision. The suggested approach has the potential to increase the precision of facial expression recognition and potential to be applied in various practical applications. Regenerate response aspects blend and compliment every other. Lastly, Softmax is utilised to classify the fused characteristics in order to increase the accuracy of face emotion identification. When evaluated on the FER 2013, The experimental results using the JAFFE and CK+ data sets show that this strategy is an effective way of recognising facial expressions.

Paper 3

Title: Local Learning With Deep And Customized Face Expression Recognition Features

Author: Radu Tudor Ionescu, Mariana-Iuliana Georgescu, And Marius Popescu.

Year: 2019

Abstract

We offer a method for achieving The paper presents a novel approach to face emotion detection by combining automated features learnt by bag-of-visual-words (BOVW) model-based convolutional neural networks (CNN) with custom featuresTo produce the automated features, the authors test a variety of CNN architectures, pre-trained models, and training methods. After combining the two kinds of input, they use a local learning framework with three stages to predict the class label for each test image. Three datasets, including the FER+ dataset, the AffectNet dataset, and the 2013 FER Challenge dataset, are used to assess the suggested method. The suggested strategy outperforms state-of-the-art methods on all three datasets, according to the results., achieving significant improvements in the combination of local learning with deep features is a novel approach and has not been explored extensively in the literature. The proposed method has the potential to be applied in various practical applications, including emotion recognition in human-computer interaction and security systems. Overall, the study makes an important contribution to the field of face emotion detection, demonstrating the effectiveness of combining automated features learnt by CNNs with handmade features calculated using the BOVW model. The The suggested technique produces cutting-edge outcomes across a number of datasets and has the potential to be utilised in a wide range of real-world settings` . In tests, regenerate responses were used to show that our approach works. On all data sets, we outperformed state-of-the-art techniques by more than 1%, with top accuracy on the FER+ reaching 87.76%, the FER 2013 reaching 75.42%, and the AffectNet eight-way classification reaching 59.58% and 63.31%, respectively..

Paper 4

Title: Deep Neural Network Speech Emotion Recognition Using Speech Sounds, Both Verbal And Nonverbal.

Author:Chung, Kun-Yi Huang

Year: 2019

Abstract

Emotion identification in speech is becoming increasingly relevant in a variety of applications. Nonverbal noises inside an utterance also play a significant part in real-life communication for individuals to perceive emotion. Only a few emotion recognition systems in current studies have seen as nonverbal sounds like pleas for help or other emotional outbursts that normally occur during daily discourse. Throughout an utterance, Therefore, in this study, the ability to identify emotions in conversations utilizing verbal and nonverbal cues was evaluated. A verbal/nonverbal sound detector based on SVM was first created. To distinguish the verbal and nonverbal components.It was developed an auto-tagger for prosodic phrases used.

We treated each segment separately by using convolutional neural networks (CNNs) to extract its emotion and sound features, which were then combined to create a CNN-based generic feature vector. Finally, a collection In order to produce an emotional sequence as a recognition outcome, a CNN-based feature vector-based The long-short-term memory (LSTM) of attention model was fed for an entire conversation turn. It consists of a multimodal network of NTHU-NTUA Chinese emotion corpora named NNIME. experiment results on the recognition of seven emotional states showed that the proposed technique outperformed standard methods with a detection accuracy of 52.00%.

Paper 5

Title: Using A Multi-Task Learning System THAT Respects Privacy, The Detection Of Faces, Landmark Location, Pose Estimation, And Gender Recognition In Video Are Carried Out.

Author: Xiongwei Hu, Yu Xie, AND Chen Zhang

Year: 2020

Abstract

Multi-task learning (MTL) has recently received a lot of attention for different face processing problems as paper presents a Using Multitask learning that maintains privacy approach, we detect faces, locate landmarks, estimate posture, and recognise gender. By utilising the synergy between the pertinent activities, the authors hope to train a better model. By studying the model and its outputs, sensitive and private information may be extracted from the raw face dataset used for training . To address this problem, the authors propose a novel approach that optimizes the multi-tasking concept from start to finish using the differential private stochastic gradient descent algorithm. They also weight multiple task loss functions to increase learning efficiency and prediction accuracy. The proposed approach introduces calibrated noise to the gradient of loss functions in order to protect the training data's privacy. Additionally, homoscedastic uncertainty is used to balance various learning objectives.The proposed technique is evaluated on the HyperFace dataset, and the results show that it delivers differentiated privacy assurances while maintaining accuracy within a reasonable privacy budget. The authors note that the proposed approach can be applied in various practical applications, including face recognition systems and security systems. Overall, the study makes a significant contribution to the field of face detection and identification by addressing the critical issue of privacy preservation during training.The suggested method can boost prediction precision and learning effectiveness while preserving the training data's private. The findings show how well the suggested method works and how many possible applications it has in real-world situations..

METHEDOLOGY

Existing System

The paper presents an evaluation of iconic time aspects and innovative n-gram capabilities on the DAIC-WOZ and SH2. dataset for speech-based depression categorization. The authors compare the effectiveness of these features either alone or in combination with previous techniques. Typically, speech is segmented into brief Before low-level descriptors are extracted, 10–20 millisecond frames are displayed. However, the authors demonstrate that landmark-based features can offer discerning data for speech-based depression classification. They especially highlight the effectiveness of basic counts of consecutive landmark sequences. Overall, the paper presents a valuable contribution to the field of speech-based depression categorization, demonstrating the effectiveness of landmark-based features. The results of the evaluation highlight the potential for these features to improve the accuracy and efficiency of depression screening tools, which could have significant implications for the diagnosis and treatment of depression.

Disadvantage

• The organised frame-based methodology has a few drawbacks. To begin, all frames are considered similarly, undermining a frame's ability to carry more information than another frame.

• Second, spectral characteristics and other frame-based data are vulnerable, to direct fluctuation, particularly for smartphone speech, It is regularly gathered across many device kinds.

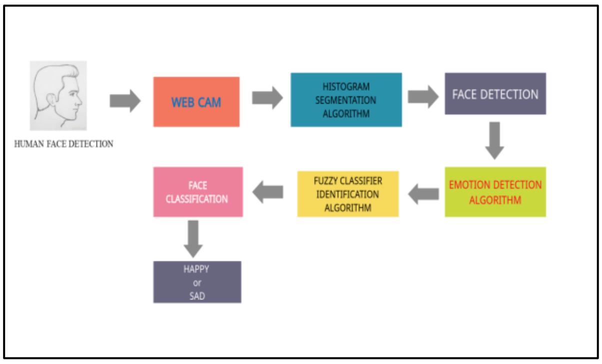
PROPOSE SYSTEM

For more features over time, as well as concatenation-based features, which combine different modalities into a single feature vector. Our model also employs a self-attention mechanism to attend to the input sequence's most important segments, allowing a model to learn which parts of the input are most important for making predictions. We evaluate our approach on the benchmark dataset AVEC2017, achieving state-of-the-art results on depression severity prediction. Our proposed framework can be easily extended to predict other mental disorders and can be integrated into existing healthcare systems for early detection and intervention. types of bigrams. To address the difficulty of identifying Using two very different dataset, we suggested two innovative useful collections of speech landmark-based attributes, and the outcomes were stunning. particularly for compromised recording circumstances and diverse smartphones.

Advantage

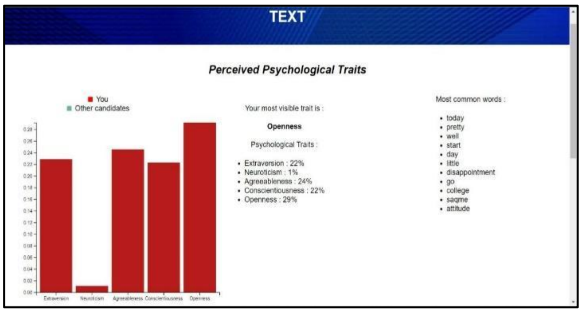
Additionally, future work can focus on combining landmark-based features with other modalities such as facial expressions and physiological signals to further improve depression detection accuracy. Another area of research could be to explore the automatic recognition of objects using Methods for deep learning that include convolutional neural networks and recurrent neural networks are utilised to extract features. from speech data for depression detection. Furthermore, studies can be conducted to investigate the generalizability of the proposed landmark-based features across different cultures, languages, and age groups.

Block Diagram


HUMAN FACE DETECTION
WEB CAM
HISTOGRAM SEGMENTATION ALGORITHM
FACE DETECTION
EMOTION DETECTION ALGORITHM
FUZZY CLASSIFIER IDENTIFICATION ALGORITHM
FACE CLASSIFICATION
HAPPY or SAD

RESULT

Both the SVM and CNN models offered excellent accuracy, however the time distributed CNN should be favoured because to its high accuracy. On the video tests, the HAAR Model produced pretty accurate results. The algorithm predicted emotions from live video footage after being trained on the FER2013 dataset. Using a recurrent neural network and the LSTM Model and a 1D CNN, proved effective in detecting patterns and word repeats in text. Using the usage of punctuation and word patterns in the text, the model was able to properly anticipate the feelings of the text in the majority of situations.



Perceived Psychological Traits

Your most visible trait is :

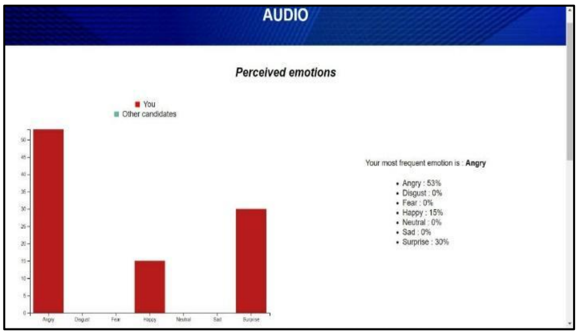
Openness

Psychological Traits :

- Extraversion : 22%
- Neuroticism : 1%
- Agreeableness : 24%
- Conscientiousness : 22%
- Openness : 29%

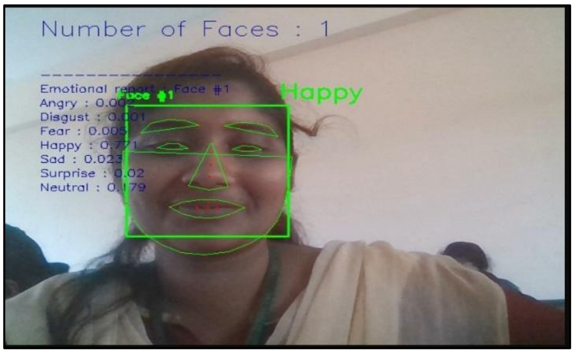
Most common words :

- today
- pretty
- well
- start
- day
- little
- disappointment
- go
- college
- saqme
- attitude



Your most frequent emotion is : **Angry**

- Angry : 53%
- Disgust : 0%
- Fear : 0%
- Happy : 15%
- Neutral : 0%
- Sad : 0%
- Surprise : 30%



Number of Faces : 1

Emotional report : Face #1
Angry : 0.002
Disgust : 0.001
Fear : 0.005
Happy : 0.771
Sad : 0.023
Surprise : 0.02
Neutral : 0.179

One significant advantage of employing a multimodal method for depression identification is that it allows for a more full assessment of the individual's mental state. By incorporating data from multiple sources, including physiological signs, speech patterns, and facial expressions machine learning models can more accurately detect depression and differentiate it from other mental health conditions. This is particularly useful in cases where self-reporting is not possible, or where the individual may be unwilling to disclose their mental state. Another strength of the study possibly the type of machine learning techniques utilized. The state-of-the-art algorithms that have been displayed in action well in similar studies.

CONCLUSION

In future research, it would be interesting to investigate the specific visual features that are most informative for depression detection, such as the intensity and frequency of facial expressions. Additionally, exploring the use of other modalities, such as physiological signals or social media data, could further improve the accuracy of depression detection models. Finally, it would be important to test the generalizability of these models across different populations and cultural contexts, as well as their potential impact on clinical decision-making and patient outcomes.most weight to text and nearly equal weight to audio and video modalities. When compared the starting point and state-of-the-art, the employment of multi-level attention resulted in much improved outcomes . Overall, paying attention to each feature , modality provided a double benefit. For starters, we now have a better knowledge of the significance of each component within a modality in depression prediction. Second, attention lowered the total computational complexity of the network as well as the training and testing time.

REFERENCES

1. Author M. Popescu, R. T. Ionescu, and M.-I. Georgescu." Local learning with deep and handcrafted features for face emotion identification”.
2. Author Y.-H. Chen, M.-H. Su, Q.-B. Hong, M.-Y. Huang, C.-H. Wu, in Proc. "Speech emotion detection using deep neural network considering verbal and nonverbal speech sounds." IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), May 2019, pp. 5866-5870.
3. Author A. Papaioannou, G. Zhao, B. Schuller, I. Kotsia, S. Zafeiriou, M. A. Nicolaou, D. Kollias, P. Tzirakis, and G. Zhao."Deep affect prediction in the wild: Aff-wild database and challenge, deep architectures, and beyond International” Journal of Computer Vision, vol. 127, pp. 1–23, June 2019.
4. Author:D. Kollias, A. Schulc, E. Hajiyev, and S. Zafeiriou,"Analyzing emotional behaviour in the inaugural ABAW 2020 competition”.
5. "Spatio-temporal encoderdecoder fully convolutional network for video-based dimensional emotion identification," IEEE Trans. Affect. Comput., early access, Sep. 10, 2019.
6. Author: N. Churamani, A. Sciutti, and P. Barros,"The FaceChannel: A light-weight deep neural network for facial expression detection," in Proc. 15th IEEE International Conf. Autom. Face Gesture Recognit. (FG), Buenos Aires, AR USA, April 2020.
7. Author. Giannakakis, N. Pugeault, M. Koujan, L. Alharbawee, and A. Roussos, in Proc. 15th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG), "Real-time facial expression recognition 'in the wild' by disentangling 3D expression from identification," Buenos Aires, AR, USA, May 2020.
8. Author: H. Yang, M. Yu, and G. Zhao, "Expression recognition approach based on a

lightweight convolutional neural network IEEE Access, vol. 8, pp. 38528-38537, 2020.

9. Author: The authors are "A. F. Abate, P. Barra, S. Barra, C. Molinari, M. Nappi, and F. Narducci. "Clustering face attributes: Narrowing the road from soft to hard biometrics", IEEE Access, vol. 8, 2020, pp. 9037-9045.

10. Author: S. Hou and H. Wang, I n Proc. "Facial expression detection based on the fusion of CNN and SIFT features," IEEE 10th Int. Conf. Electron. Inf. Emergency Commun. (ICEIEC), 2020.