

Depression Scale Recognition over Fusion of Visual and Vocal Expression using Artificial Intellectual Method



Pratibha Gayke, Rohini S Kulkarni

Abstract: Now a day's supported visual and audio cues for automatic depression assessment may be a fast emerging research subject. This comprehensive evaluation of existing methodologies focuses on machine learning (ML) algorithm and image processing (IP) algorithm, as documented in over sixty articles over the last ten years. There is a visual indicator of depression, several data collection procedures are used, and finally examined the previous year or existing datasets. In this article describes techniques and algorithms as well as methods for dimensionality reduction, visual feature extraction, regression approaches, and classification decision procedures, and also various fusion tactics. A significant meta-analysis of published data is given, based on performance indicators that are robust to chance, to identify general trends and important pressing concerns for further research using visual and verbal cues alone or in combination with signals for automated depression evaluation. The suggested work also used deep learning and natural language processing to estimate depression levels based on current video data.

Keywords: Image Processing, Artificial Intelligence (AI), Machine Learning, Classification Rule, Convolution Neural Networks (CNN), and Natural Language Processing (NLP), etc.

I. INTRODUCTION

In many cases, people who are sad are completely unaware of their mental illness. They are unable to detect the source of even minor dissatisfaction in themselves, and as a result, such users/people develop suicidal impulses. People are aware that they are depressed in certain conditions, but they are unwilling to seek help from anyone, owing to the erroneous sense of 'humiliation' connected with depression. Recognizing the indications of depression in the early stages of depression is more beneficial [1]. Depression is diagnosed as early as the primary grades, and a simple one-hour conversation with a counsellor can be extremely beneficial.

This has the potential to radically change that student's negative opinion into a good one. Such an how to deal with mental stress of students can benefit from appropriate counselling and can be guided down the right route to success. Facial expressions are the most valuable form of nonverbal communication. Number of researches has been conducted to determine which facial emotions as well as expressions are linked to stress and depression [2]. The popular work is mostly aimed at determining the appearance of depression in college students through the examination of their facial features. This system primarily employs facial detection techniques, natural language processing for speech recognition of communication, feature extraction of various emotions and classification of depressed or non-depressed features. The technology will be programmed to recognize depressive symptoms. Then, utilizing a web camera, footage of various students/people with a frontal face will be taken. Then, in order to forecast depression, the facial traits of those faces will be retrieved. The classification of user with depression or no depression based on the severity of the symptoms of depression [3]

II. RELATED WORK

- F. Zhang, Yona F. Binti A. Gaus, and Jan, H. Meng Automated depression evaluation on the basis of cues of visual and vocal; a fast emerging study subject, as discussed in this article. The current comprehensive assessment of existing methodologies focuses on Machine Learning (ML) algorithm and Image Processing (IP) algorithm, as documented in over sixty articles over the last ten years. Depression's visual symptoms, various data gathering methodologies, and existing datasets are summarized [4].
- J. Crowe, Richard K. Morriss, Michel F. Valstar, and C. Solomon, this study intends to discover not only the differences in auditory aspects between non-depressed and depressed people as well as users, but also how they alter when people or user with depression try to hide their genuine emotions. Our goal, based on a set of optimal criteria, was to create an automatic depression identification system that could correctly label a person as depressed even if they tried to disguise it. This study was not interested in healthy people who changed their behaviour to appear depressed [5].

Manuscript received on August 25, 2021.

Revised Manuscript received on August 31, 2021.

Manuscript published on September 30, 2021.

*Correspondence Author

Pratibha Gayke Shinde*, Assistant Professor, Department of Information Technology, Dr. Vithalrao Vikhe Patil College of Engineering, Ahmednagar, India.

Rohini S Kulkarni, Department of Computer Science and Engineering, K.L.E. Institute of Technology, Hubli, Visvesvaraya Technological University, Karnataka, India.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Depression Scale Recognition over Fusion of Visual and Vocal Expression using Artificial Intellectual Method

- E. M. Goldner and E. Jenkins, The goal of this research is to compile existing information on current approaches to understanding and treating this illness. The findings are believed to be useful in informing practitioners and guiding future study. The findings of a scoping examination of the scientific research article were classify and charted by the critical research model. The great bulk of current research is based on a biological model (81 percent). A broader field of study for treatment-resistant depression research might be beneficial. "Given that various etiological routes are likely to contribute to treatment resistant depression, and existing efforts at prevention and therapy have significant room for improvement, an expanded research agenda could more effectively address this vital public health issue." [6].
- J. Crowe, Richard K. Morriss, Michel F. Valstar, and C. Solomon, this research work includes a detailed survey and meta-analysis of automated emotions of facial expression in persons with non-psycho diseases set side by side to strong control groups. The studies looked at the automatic facial expression in response to corresponding stimuli and used an emotionally meaningful visual induction approach. In sadness, falling off in facial expression is particularly obvious for positive change. In a meta-analysis of eating disorders, researchers discovered that people with eating disorders have falling off of facial emotions in reaction to both positive and negative motive. According to autism

research, people with autism show fewer facial expressions [7].

III. PROPOSED SYSTEM

A. System Architecture:

Fig.1 is our system model or architecture diagram, in which user will sign up and the details get store in the local database, the user will login and provide a video, which is then separated in two parts a visual and audio[8]. In image processing CNN will be used for detecting face and to find emotions on face like happy, sad, angry, fear, neutral, disgust. In this process images is taken from video and different frames are generated.

In audio extraction, when we upload a video then audio is separated from video and wave generation takes place. For speech recognition we use bag of words logic, we have some positive word dataset as well as some negative word dataset, with the help of this we can find the probability of how many words person speaks positive or negative, at last we apply Naive Bayes algorithm for final result [9].

After getting final result the user has a facility to see his report. In a method to quantify the on social media platforms variations between stressed and non-stressed peoples, three modules have been primarily defined in this project to handle the difficulty of stress detection: Categorization of System Framework, Social Interactions, and Attributes [10].

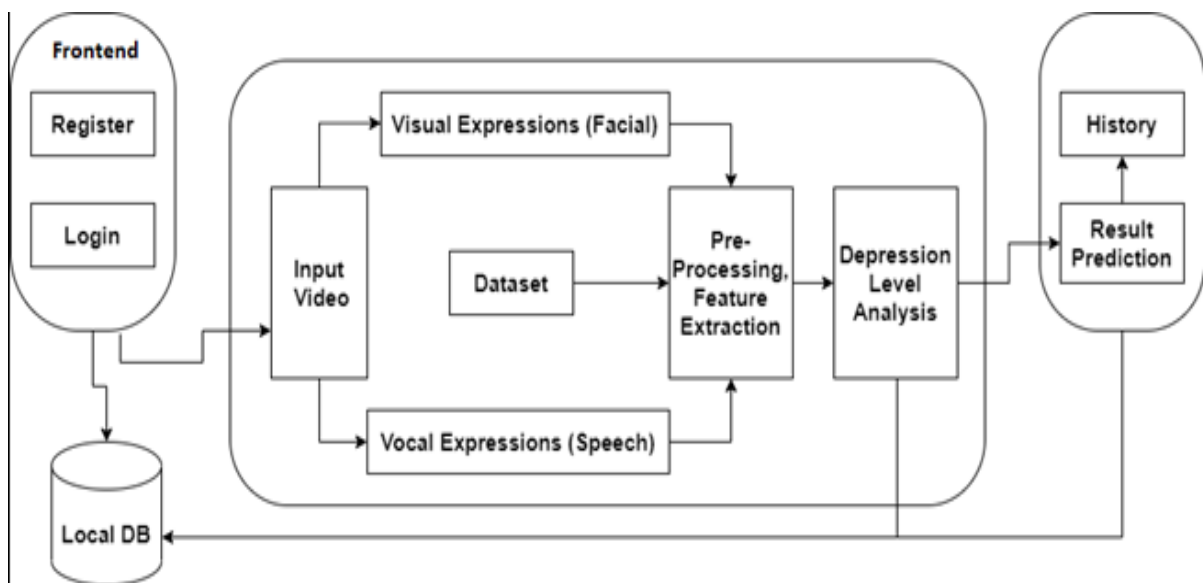


Fig.1: System Architecture

B. System Framework:

Here present a unique hybrid model in this framework that combines a determinant graph representation with a CNN to use social & content interaction message for stress exposure. The suggested model can increase detection review over 6-9 percent in F1-score, according to experimental data. By delving deeper into the data on social interactions, the number of social network system with poor connections is over 14 percent greater among stressed users than among non-stressed users, showing that stressed peoples friends are less linked and difficult than

non-stressed people's friends [11].

C. Social Interaction:

To investigate the relationship between peoples stress levels and their social cooperation on social networks, looking at the problem from two perspectives:

(1) social interaction content between stressed and non-stressed peoples, looking at the differences in content of social interaction; (2) social cooperation network, looking at the differences in structure in terms of structural variance, Strong/weak ties, as well as societal influence. In this research uncovered some fascinating social trends. For e.g., we discovered that the maximum social structures with sparse connections among stressed peoples is 14 percent greater than among non-stressed peoples, implying that the social structure of stressed peoples friends is minimum of connected and difficult, compared to that of non-stressed peoples [12].

D. Attribute Categorization:

Here initially develop two different sets of aspects to count the differences between stressed and non-stressed peoples on social media platforms: image and speech-level aspects from a single user and user-level aspects averaged from a people's weekly activity.

IV. SYSTEM MODELS

A. Emotion and Face Detection:

Humans have traditionally had an easy time detecting emotions from facial expression, but accomplishing the same task with a computer algorithm is a much more difficult task. Current systems largely focus on face investigation while keeping the background intact, resulting in a plethora of superfluous and deceptive features that muddle CNN training [13]. This is categorized into three different parts:

- **Facial Detection:**-- Capability to locate or capture users face as a input image as well as frame. To calculate the output is the detected faces coordinates of border or bounding box.
- **Facial Recognition:**-- In that image processing algorithm compare multiple persons faces and its emotions. This is accomplished by contrasting face embedding vectors.
- **Emotion Detection:**-- The expression on the face is classified as joyful, terror, sad, surprise, angry, disgust, and neutral. Humans have a proclivity for picking up nonverbal cues from facial expressions. Computers are now recovering the ability to read emotions. So, how to recognize emotions in a photograph? Here, used an open source data set from Kaggle called Facial Emotion Recognition (FER) to train a CNN to recognize emotions. The emotions can be categorized into seven different classes' i.e. joyful, terror, sad, surprise, angry, disgust, and neutral.

B. Speech to Text Conversion:

Speech recognition is a crucial function in a variety of applications, including home automation and artificial intelligence. During this procedure, captured audio was transmitted to Google, which turned the audio file into a wave file in text format

V. ALGORITHM DETAILS

A. CNN Algorithm:

CNN is one of the most common image recognition and classification algorithms. CNNs are frequently employed in

domains like emotion recognition, facial recognition, object detection, and so on. Then image classification algorithm i.e. CNN can takes an input image, to pre-process that image, and categorizes it into different groups of emotions (joyful, terror, sad, surprise, angry, disgust, and neutral). A CNN is a neural network algorithm with more than one convolutional and pooling layer [14].

- **Step 1:** The machine is fed a data set containing images and reference emotions. Face Emotion Recognition (FER) is the name of the dataset, and that is publicly-available data collection that was formed publically freely accessible on Kaggle Dataset.
- **Step 2:** Here you may build the model by importing the necessary libraries.
- **Step 3:** The image features are extracted pixel by pixel using a convolution neural network.
- **Step 4:** The retrieved pixels are subjected to matrix factorization. The matrix is $m \times n$ in size.
- **Step 5:** This matrix is subjected to maximum pooling, which involves selecting the highest value and re-inserting it into the matrix.
- **Step 6:** Every negative value is turned to zero as part of the normalization process.
- **Step 7:** Rectified linear units are used to convert values to zero, with each value filtered and the negative value converted to zero.
- **Step 8:** After computing maximum probability, the hidden layers apply weights to the input values from the visible layers.

B. Speech to Text Conversion:

• Text Mining:

The process of extracting useful information from natural language text is known as text mining. The technique of extracting high-quality information from text is referred to as text mining. The ultimate purpose is to use Natural Language Processing (NLP) to convert text into data for study [15].

• Natural Language Processing (NLP):

NLP is a branch of AI in that computers intelligently examine, comprehend, and infer effect from human language. We can use natural language processing (NLP) to organize and arrange knowledge for tasks like automatic summarization, topic segmentation, translation, sentiment analysis, relationship extraction i.e. feature extraction, named entity recognition and speech recognition. NLP primarily acts as an important aspect called as speech reorganization in which system analyze primary source of audio data in the form of spoken words. Syntactic analysis is used in NLP to evaluate how natural language conforms to grammatical norms. Here are some syntax techniques that can be used:

1. **Tokenization:** This is a crucial step in natural language processing that divides a string of words into semantically relevant tokens. Word tokens are usually separated by blank spaces, while sentence tokens are separated by stops.

Depression Scale Recognition over Fusion of Visual and Vocal Expression using Artificial Intellectual Method

2. **Part-of-Speech Tagging:** It entails determining which portion of speech each word belongs to. It indicates whether the word is a noun, a pronoun, an adjective, a verb, an adverb, a preparation or a conjunction.
3. **Bag of Words:** It divides each string into words, lists them in a dictionary, and turns each data word into its root word.

C. Naive Bayes (NB) Classifier:

It's a probabilistic algorithm that's commonly used in text classification. The NB classification algorithm is simple to build and works well with large data entries as well as datasets. As a result of its clarity, NB is known to better even the most sophisticated classification systems. Despite the fact that the features are linked, the Naive Bayes method examines each one separately.

VI. RESULTS AND DISCUSSION

In this section results are formed as a implementation of ML and IP based application and results is generated on any web browsers for social community users to prevent various health and mental stress issues of patients as well as user's in his daily life and social interaction time. The analysis results of the proposed system are carried out according to the following parameters:

- Time consumption = A
- Response Time = B
- Computation Cost = C
- Performance accuracy = D
- Scalable & User Friendly = E

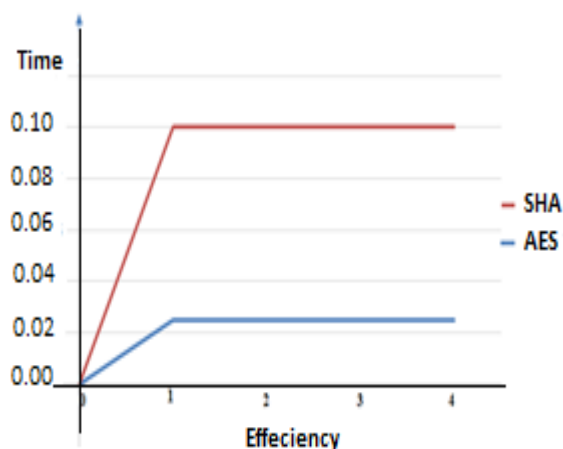


Fig.3: Time and Efficiency Chart

In this context, the whole system has taken many more input attributes, but the main focus is on system performance and time. We will obtain the following analytical results for our proposed system based on some attributes and above mentioned parameters is as follows:

Table 1: Result Table

Parameter	Existing	Proposed
A	10	4
B	10	5
C	8	8
D	10	3
E	8	2

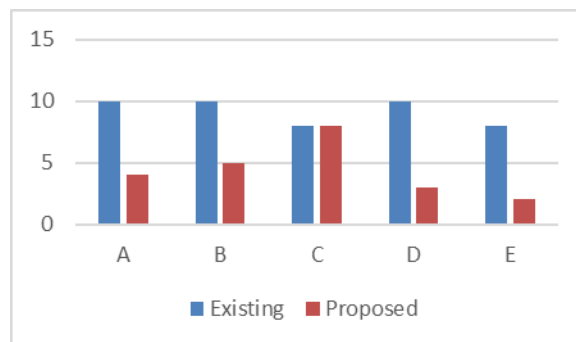


Fig.2: Time line chart of Result Analysis

VII. CONCLUSION

In the present article introduce a unified deep learning-based depression detection framework. Finally, we introduced a novel technique for classification applications of word embedding optimization. On user level, we conducted a similar evaluation of some of the most extensively used deep learning techniques and its algorithms for depression identification from tweets. We ran our tests on datasets that were freely available online. Our tests revealed that our CNN-based models outperformed CNN-based models. Models with improved embeddings were able to keep their generalization capabilities while maintaining performance. We presented the findings of a study that looked at the automatic detection of depression in a human-computer interaction context using audio and video elements. We wanted to see how difficult it would be to deceive or defraud such an automated system. We discovered that depressed peoples seemed to be subsequent to the hypothesized pattern of reduced energy levels in speech in our study of maximally matched healthy and depressed people. Many of the prosodic and spectral elements previously used in emotion recognition were discovered to be significant as well in depression recognition. However, not all of the distinguishing characteristics between depressed and healthy participants were the same as those employed in emotion recognition.

ACKNOWLEDGMENT

I want to thank the scientists and the publishers for their support. I also thank the guide and the reviewer for their useful thoughts and the university administration for the infrastructure and assistance they need.

REFERENCES

1. Girard, Jeffrey M., Jeffrey F. Cohn, Mohammad H. Mahoor, Seyed mohammad Mavadati, and Dean P. Rosenwald "Social risk and depression: Evidence from manual and automatic facial expression analysis" 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), pp. 1- 8. IEEE, 2013.
2. Deepak A Vidhate, Parag Kulkarni "Performance comparison of multiagent cooperative reinforcement learning algorithms for dynamic decision making in retail shop application", International Journal of Computational Systems Engineering, Inderscience Publishers (IEL), Vol 5, Issue 3, pp 169-178, 2019.

3. Alghowinem, Sharifa, Roland Goecke, Jeffrey F. Cohn, Michael Wagner, Gordon Parker, and Michael Breakspear. "Cross-cultural detection of depression from nonverbal behavior" 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), vol. 1, pp. 1-8. IEEE, 2015.
4. Deepak A Vidhate, Parag Kulkarni "A Framework for Dynamic Decision Making by Multi-agent Cooperative Fault Pair Algorithm (MCFPA) in Retail Shop Application", Information and Communication Technology for Intelligent Systems, Springer, Singapore, pp 693-703, 2019.
5. Pampouchidou, A., O. Simantiraki, C-M. Vazakopoulou, C. Chatzaki, M. Pediaditis, K. Marias et al. "Facial geometry and speech analysis for depression detection" 39th Annual International Conference on Engineering in Medicine and Biology Society (EMBC), pp. 1433-1436. IEEE, 2017.
6. Deepak A Vidhate, Parag Kulkarni "A Novel Approach by Cooperative Multiagent Fault Pair Learning (CMFPL)", Communications in Computer and Information Science, Springer, Singapore, Volume 905, pp 352-361, 2018.
7. Harati, Sahar, Andrea Crowell, Helen Mayberg, and Shamim Nemati. "Discriminating clinical phases of recovery from major depressive disorder using the dynamics of facial expression" 38th Annual International Conference of Engineering in Medicine and Biology Society (EMBC), pp. 2254- 2257, IEEE, 2016.
8. Deepak A Vidhate, Parag Kulkarni, "Exploring Cooperative Multi-agent Reinforcement Learning Algorithm (CMRLA) for Intelligent Traffic Signal Control", Smart Trends in Information Technology and Computer Communications. SmartCom 2017, Volume 876, pp 71-81.
9. Cohn, Jeffrey F., Tomas Simon Kruez, Iain Matthews, Ying Yang, Minh Hoai "Detecting depression from facial actions and vocal prosody" 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops. ACII 2009., pp. 1-7. IEEE, 2009.
10. Deepak A Vidhate, Parag Kulkarni, "A Novel Approach for Dynamic Decision Making by Reinforcement Learning-Based Cooperation Methods (RLCM)", International Conference on Intelligent Computing and Applications, Springer, Singapore, pp 401-411, 2018.
11. Tasnim, Mashrura, Rifat Shahriyar, Nowshin Nahar, and Hossain Mahmud. "Intelligent depression detection and support system: Statistical analysis, psychological review and design implication" 18th International Conference on Health Networking, Applications and Services (Healthcom), pp.1-6 IEEE, 2016.
12. Pampouchidou, Anastasia, Kostas Marias, Manolis Tsiknakis, P. Simos and Fabrice Meriaudeau "Designing a framework for assisting depression severity assessment from facial image analysis" International Conference on on Signal and Image Processing Applications (ICSIPA), pp.578-583, IEEE, 2015.
13. Deepak A Vidhate, Parag Kulkarni, "Multiagent Cooperative Reinforcement Learning by Expert Agents (MCRLEA)", International Journal of Intelligent Information Systems, Science Publishing Group, volume 6, issue 6, pp72-84, 2017.
14. Meng, Hongying, Di Huang, Heng Wang, Hongyu Yang, Mohammed Al-Shuraifi, and Yunhong Wang. "Depression recognition based on dynamic facial and vocal expression features using partial least square regression" 3rd ACM international workshop on Audio/visual emotion challenge, pp. 21-30, ACM, 2013.
15. Deepak A Vidhate, Parag Kulkarni, "A novel approach to association rule mining using multilevel relationship algorithm for cooperative learning" 4th International Conference on Advanced Computing & Communication Technologies, pp 230-236, 2014.



Rohini S Kulkarni received a B.E. degree in computer science and engineering from the K.L.E. Institute of Technology, Hubli, Visvesvaraya Technological University, Karnataka, India, in 2017-2021. My progress in research interests allows Cloud Computing, AI, Software Development, and Data Science.

AUTHORS PROFILE



Pratibha Gayke / Shinde received the B.E degree in Information Technology from North Maharashtra University, Jalgaon in 2011. And also received the M. Tech degree in Computer Science and Engineering from Jawaharlal Nehru Technological University, Hyderabad in 2014, where she is currently pursuing the PhD degree. Currently working as an Asst. Prof. with Department of

Information Technology, Dr. Vithalrao Vikhe Patil Foundation's, Dr. Vithalrao Vikhe Patil College of Engineering, Ahmednagar. Also having a membership of the Institution of Engineers (India). My current research interest includes Machine Learning, Image Processing and Artificial Intelligence.