International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

A Study of Multimodal Structure for Stress Recognition in IT Experts: Deep Analysis of Facial Expression Recognition with Deep Speech and Tone Analyzer

Mr. Shailesh Kurzadkar [1], Dr. Vijay Bhandari [2], Dr. Anup Bhange[3]

Submitted: 15/05/2024 **Revised**: 28/06/2024 **Accepted**: 10/07/2024

Abstract: In today's world, mental stress is growing more pervasive and progressively more severe, endangering people's physical and mental well-being. Early stress detection is essential to preventing the negative consequences of stress on individuals. The usefulness of use objective indicators to identify stress has been shown in numerous studies. An increasing number of researchers have been attempting to identify stress using deep learning technology in recent years. In this paper, FER model is proposed.The computer vision problem known as Facial Expression Recognition (FER) aims to recognize and classify the various emotional expressions that are displayed on a human face. DeepSpeech can be used to train a model using gathering of voice data. The trained model can then be used for recognition or inference. Several pre-trained models are included in DeepSpeech. Digital audio is fed into DeepSpeech, which then outputs a "most likely" text transcript of the audio.IBM Tone Analyzer service employs language analysis to identify analytical, confident, tentative, fearful, angry, and joyful tones in user input (text).

Keywords: *Mental Stress, FER model, Deep Speech,Tone Analyzer*

1.1 Introduction

In recent years, the field of stress analysis and sentiment analysis of posts on microblogging platforms has flourished. High levels of mental stress have been shown to cause detrimental behavioral changes in both men and women. Stress is a person's reaction to threats, whether they come from within or beyond. It may have an impact on a person's memory, decision-making skills, and everyday performance. A person's physical and mental health may deteriorate, and they may even develop immune system problems, cardiovascular disease, depression, or other illnesses, if acute stress persists in their lives. Stress has become more common and intense in today's culture. Stress can be assessed by filling out a questionnaire or by consulting a psychologist. Because psychological evaluation is subjective and instantaneous, it frequently results in inaccurate or misleading stress detection and cannot satisfy real-time detection criteria. People's voices, facial expressions, and involuntary bodily functions shift as a result. Due to its well-known uses in security, education, medical rehabilitation, FER in the wild, and safe driving, facial expression recognition (FER) techniques utilizing computer vision and artificial intelligence have seen an exponential increase in recent years. The movement of facial muscles produces facial expressions, which are incredibly important in human communication. These expressions convey a variety of information, from minor ones like raising an eyebrow during a conversation to states of deep survival. A source image's lighting, posture, background, and camera angle, in addition to any occlusion or misalignment, have a big impact on FER. Both extrapolation data from the perceptual system and calculations made in the visual-perceptual system, which are aided by perceptual processes, are necessary for efficient FER.

However, working flow and feature extraction have been the main topics of the works done so far.Both academic and industrial viewpoints have examined FER, which can offer insight into a person's personality, temperament, cognitive capacity, and psychopathology. For instance, it has been demonstrated that FER technology can be used to conduct quantitative study on the effects of neuropsychiatric illnesses on perception and expression.

A quick reaction is evoked by facial expressions, which frequently mimic emotions.When facial muscles flex in reaction to a specific action or query, expressions can convey a great deal of information quickly in management or social human interactions.[33] Therefore, in order for computational systems to effectively assess an individual's mood, automatic FER techniques are required.

DeepSpeech receives an audio stream and converts it into a string of letters in the designated alphabet.Two fundamental phases enable this conversion: The audio is first transformed into a series of probabilities over alphabetic characters. Second, a series of characters is created from this set of probabilities.

[1]*PhD Scholar, Department of CSE, Madhyanchal Professional University, Bhopal,India*
[2]*Assosiate Professor, Department of CSE, Madhyanchal Professional University, Bhopal,India*
[3] *Assistant Professor KDK College of Engineering, Nagpur, India*

The Tone Analyzer service employs language analysis to identify analytical, confident, tentative, fearful, angry, and joyful tones in user input (text). We can infer the significance of such a service from the knowledge that communication is insufficient if tone is ignored.

1.2 Related Work

In order to detect acute stress, Zhang et al. proposed a deep learning system that used voice, facial expressions, and ECG data in real time [34]. Additionally, we created the temporal attention module (TAM) to identify keyframes associated with facial expression stress detection. The suggested methodology just needs basic preprocessing and avoids complex feature extraction. Our work's contributions can be summed up as a deep learning framework for severe stress detection incorporating voice, facial expressions, and ECG. The suggested framework makes use of TAM, and the fusion approach is based on the matrix eigenvector, which provides 85.1% detection accuracy. To highlight the unique temporal representation for stress-related facial expressions, the TAM gives various learnable weights to various facial expression frames.

Baheti Reshma Radheshamjee, Kinariwala Supriya[12] In-person interviews, discussions, or other activities are used to identify stress that has been proposed for the current system. when two or more people are analyzed by someone else. Based on users' weekly social media data, this proposed a system framework that uses their social interactions and tweet content to determine their psychological stress states. Each term in this dictionary has a grade between -5 and +5. The both NB and SVM algorithms were used to classify and predict the data. To increase the accuracy of the results, Word Sense Disambiguation was implemented using the ngram and Skip-gram models. When paired with SVM, Ngram and WSD produce 65% precision.

Stress levels among office workers are rising due to work-related issues and an overwhelming amount of work. Therefore, in order to detect stress in its early stages and prevent any detrimental impacts on well-being, it is essential to routinely track and control employees' stress levels. After evaluating the literature regarding identifying stress in office settings using multimodal measurements,[23] examined the features and parameters comprising physical, social, and associated data.

In [10] The study makes use of secondary as well as primary information sets. This study examines how social and emotional factors affect social media data and usage using a cross-sectional methodology and a combination of qualitative and quantitative methodologies. The aim of effort is to develop models for sentiment and emotion identification that may be applied to stress management. Additionally, original data is used to evaluate the models. The study aims to determine a user's sentiments or emotions for different themes or areas using Latent Dirichlet Allocation (LDA). Hybrid machine learning and deep learning models that are developed and applied using data that comprises a diverse range of tweets provide sentiment analysis.

[35] work offered a novel approach to face emotion recognition. Enhanced Stress CNN (ESCNN) is the name of the suggested technique. The ESCNN uses Tensor Flow and MobileNet V2 to execute pretrained models on the FER2013 dataset. A person is categorized as stressed or unstressed based on the examination of their facial expressions.Relu6, Max-pooling layer, Fully Connected Layer, and SoftMax Probabilities are additional components of MobileNet V2 in ESCNN.Haar To find faces in the picture, cascade face detection is also being used.

By presenting segments of genuine stimuli—in this case, speech—in two distinct random sequences so that each segment occurs in two distinct contexts—that is, surrounded by various stimuli—we have recently developed a technique to estimate the effective integration window of an arbitrary response.[36]

[37] suggested a system to categorize song tones into suitable classes and a supervised learning methodology to analyze the tones of specific song lyrics. The suggested methodology section might be broken down into four sections: data collection, data processing, IBM Watson Tone Analyzer tone class extraction, and classifier-based tone classification.

1.3 Working flow of FER

Figure 3 illustrates the generic pipeline of FER's stepwise working flow, which is described in this part.

Figure 1

Preprocessing and data collecting via visual sensors are crucial phases. Usually, the information is obtained from a variety of sources, including surveillance, mobile phones, and Pi Cam devices.,cameras.

Facial detection is another name for region of interest (ROI) detection, which in this case refers to the face. AI-based methods are used for ROI detection, which finds and identifies faces in pictures.

These techniques have been widely used in a number of applications that entail tracking or surveillance, including security, law enforcement, entertainment and personal safety. A system called facial emotion recognition analyzes

emotions from a variety of sources, including images and videos. It is a part of the a multidisciplinary branch of study on computers' ability to recognize and analyze human emotions and affective states, "affective computing" refers to a family of technologies that frequently builds upon artificial intelligence technology.

Human emotions can be inferred from facial expressions, which are nonverbal communication methods. Researchers in the fields of psychology (Ekman and Friesen 2003; Lang et al. 1993) and human-computer interaction (Cowie et al. 2001; Abdat et al. 2011) have been interested in decoding such emotion displays for decades. The development of FER technology has recently been significantly influenced by the widespread use of cameras as well as advancements in machine learning, biometrics analysis, and pattern recognition.

Face detection, facial expression detection, and expression classification to an emotional state are the three phases that make up FER analysis (Figure 2). The examination of facial landmark positions—such as the end of the nose and the eyebrows—is the foundation for emotion recognition. According to the algorithm, facial expressions can be categorized as either compound emotions (such as happily sad, happily surprised, happily disgusted, sadly scared, sadly angry, sadly surprised) or basic emotions (such as anger, disgust, fear, joy, sorrow, and surprise) (Du et al. 2014). In other situations, facial expressions may be associated with a person's physical or mental state (e.g., boredom or fatigue).

Fig 2

The photos or videos that are used as input to FER algorithms come from a variety of sources, including surveillance cameras, cameras positioned near storefront advertising screens, social media, streaming services, and individual devices.

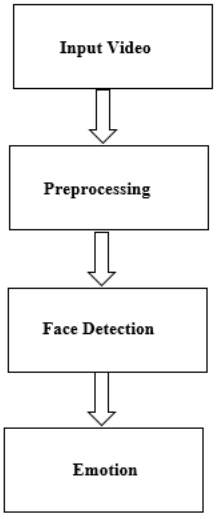

Fig 3

The overall process of face emotion recognition consists of three phases: Pre-processing, face detection, and sentiment classification. In the pre-processing phase, the dataset is prepared to work with generalized algorithms and generate efficient results. The face detection phase involves detecting faces in real-time captured images

1.4 Working flow of Deep Speech

Training and inference are two important voice recognition tasks that DeepSpeech can be utilized for. A trained model is necessary for speech recognition inference, which is the process of turning spoken audio into written text. DeepSpeech can be used to train a model using a collection

of voice data. The trained model can then be used for recognition or inference. Several pre-trained models are included in DeepSpeech.

DeepSpeech receives an audio stream and converts it into a string of letters in the designated alphabet.Two fundamental phases enable this conversion: The audio is first transformed into a series of probabilities over alphabetic characters. Second, a series of characters is created from this set of probabilities.

The process of assessing a text's sentiment in relation to a particular element is called Aspect Based Sentiment Analysis (ABSA), sometimes referred to as fine-grained opinion mining. The limitations of conventional sentiment analysis techniques have given rise to a relatively new area of study called aspect-based sentiment analysis.

Conventional techniques for sentiment analysis give a text a single sentiment label (such as positive, negative, or neutral) and treat it as a whole. This is sufficient for a lot of tasks, but there are also a lot of circumstances in which knowing a text's sentiment regarding a certain feature would be helpful.

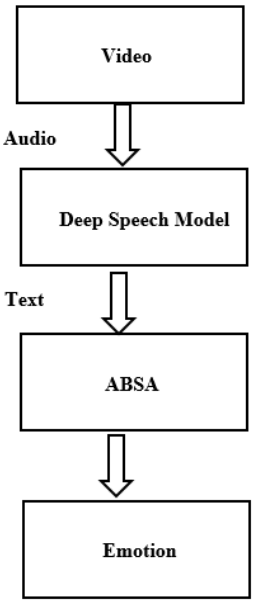

Fig-4

1.5 IBM Tone Analyzer

To identify emotional and linguistic tones in written text, the IBMTone Analyzer service employs linguistic analysis. The service is capable of analyzing tone at the sentence and document levels. By using the service, you can enhance the tone of your written communications by learning how they are interpreted. Companies can use the service to comprehend and enhance their customer chats, or to learn the tone of their customers' communications and reply to each one properly.

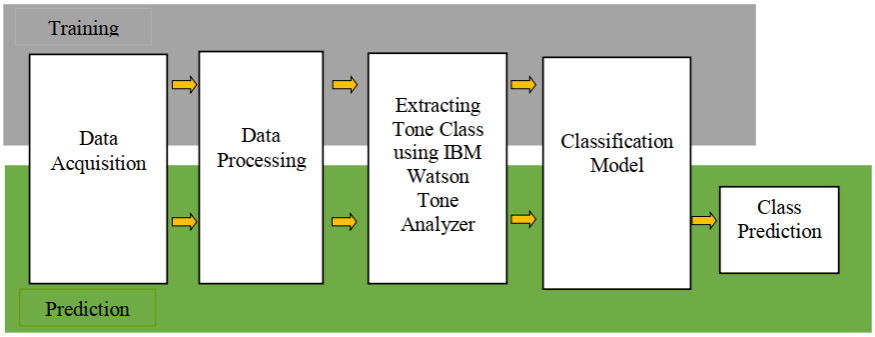

Fig-5

The Tone Analyzer service employs language analysis to identify analytical, confident, tentative, fearful, angry, and joyful tones in user input (text). We can infer the significance of such a service from the knowledge that communication is insufficient if tone is ignored.

1.6 Conclusion

In this article, FER model is proposed.The computer vision problem known as Facial Expression Recognition (FER) aims to recognize and classify the various emotional expressions that are displayed on a human face.

Also DeepSpeech can be used to train a model using a corpus, or collection of voice data. The trained model can then be used for recognition or inference. Several pre-trained models are included in DeepSpeechA "most likely" text transcript of the audio is produced by DeepSpeech once it receives digital audio.

And finally IBM Tone Analyzer service employs language analysis to identify analytical, confident, tentative, fearful, angry, and joyful tones in user input (text).

References

[1] MS.N.Pavani, P.Supriya, A.SiriChandana3, B.Trinetra4, S.V.N.S.S.Supriya, "STRESS DETECTIONUSING IMAGE PROCESSING AND MACHINE LEARNING", Dogo Rangsang Research Journal, Vol-08 Issue-14 No. 02: 2021

[2] U SRINIVASULU REDDY, ADITYA VIVEK THOTA, A DHARUN, "Machine Learning Techniques for Stress Prediction in Working Employees", 2018 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)

[3] D Y Liliana, "Emotion recognition from facial expression using deep convolutional neural network", 2018 International Conference of Computer and Informatics Engineering (IC2IE), Conf. Ser. 1193 012004

[4] Kusuma H R, Devika Rani B S, Anjali K, Ashwini N, "Stress Detection in It Professionals Using Image Processing and MachineLearning", International Journal of Advances in Engineering and Management (IJAEM), Volume 5, Issue 7 July 2023,pp. 255-259

[5] Dr. S. Vaikole, S. Mulajkar, A. More, P. Jayaswal, S. Dhas, "Stress Detection through Speech Analysis using Machine Learning", IJCRT | Volume 8, Issue 5 May 2020,pp. 2239-2244

[6] B. Padmaja, V. V. Rama Prasad and K. V. N. Sunitha, "A Machine Learning Approach for Stress Detection using a Wireless Physical Activity Tracker", International Journal of Machine Learning and Computing, Vol. 8, No. 1, February 2018, pp. 33-38

[7] Adnan Ghaderi, Javad Frounchi, Alireza Farnam, "Machine Learning-based Signal Processing Using Physiological Signals for Stress Detection", 22nd Iranian Conference on Biomedical Engineering(ICBME 2015), Iranian Research Organization for Science and Technology (IROST), Tehran, Iran, 25-27 November 2015

[8] Virginia Sandulescu, Sally Andrews, Nicola Bellotto, Oscar Martinez Mozos, "Stress Detection Using Wearable Physiological Sensors", International Work-Conference on the Interplay Between Natural and Artificial Computation,June 2015, DOI:10.1007/978-3-319-18914-7_55

[9] Xingxing Zhang , Chao Xu 1, Wanli Xue 1, Jing Hu , Yongchuan He and Mengxin Gao, "Emotion Recognition Based on Multichannel Physiological Signals with Comprehensivenonlinear Processing", www.mdpi.com/journal/sensors, Sensors 2018, 18, 3886; doi:10.3390/s18113886

[10] Tanya Nijhawan, Girija Attigeri and T. Ananthakrishna, "Stress detection using natural language processing and machine learning over social interactions",springer open access journal, https://doi.org/10.1186/s40537-022-00575-6,2022

[11] Dorota Kami´nska, "Recognition of Human Mental Stress Using Machine Learning: A Case Study on Refugees", *Electronics* 2023, *12*, 3468.

[12] Reshma Radheshamjee Baheti, Supriya Kinariwala, "Detection and Analysis of Stress using Machine Learning Techniques", International Journal of Engineering and Advanced Technology (IJEAT), Volume-9 Issue-1, October 2019

[13] Manjunath R., Shivashankar, Shivakumar Swamy N, Erappa G, Manohar Koli, Nandeeswar S. B., Niranjan R. Chougala, "A Smart Biomedical Healthcare System to Detect Stress using Internet of Medical Things, Machine Learning and Artificial Intelligence", International Journal of Intelligent Systems and Applications in Engineering,Vol 11,Issue 4,2023,pp. 335-343

[14] Rohini Hanchate, Harshal Narute, Siddharam Shavage, Karan Tiwari, "Stress Detection Using Machine Learning", International Journal of Science and Healthcare Research, Vol. 8; Issue: 2; April-June 2023,pp-307-311

[15] Kavitha S Patil, Pranav Sivaprasad, Udhay Kiran K, Sujay G S, "Projective exploration on individual stress levels using machine learning", International Research Journal of Engineering and Technology (IRJET), Volume: 10 Issue: 04 | Apr 2023,pp-1449-1454

[16] Mohamed Razeed Mohamed Nowfeek, "A Review of Machine Learning Approach for Mental Stress Detection", Journal of Information Systems & Information Technology (JISIT), Vol. 6 No.2, 2021; pp-72 – 83

[17] Nisha Raichur, Nidhi Lonakadi, Priyanka Mural, "Detection of Stress Using Image Processing and Machine Learning Techniques", International Journal of Engineering and Technology (IJET), Vol 9 No 3S July 2017

[18] Z. Zainudin, S. Hasan, S.M. Shamsuddin and S. Argawal, "Stress Detection using Machine Learning and Deep Learning", Asian Conference on Intelligent Computing and Data Sciences (ACIDS) 2021

[19] S. A. Singh, P. K. Gupta, M. Rajeshwari, and T. Janumala, ''Detection of stress using biosensors,'' *Mater. Today*, vol. 5, no. 10, pp. 21003–21010, 2018

A. Alberdi, A. Aztiria, and A. Basarab, ''Towards an automatic early stress recognition system for office

environments based on multimodal measurements: A review,'' *J. Biomed. Informat.*, vol. 59, pp. 49–75, Feb. 2016.

[20] U. Pluntke, S. Gerke, A. Sridhar, J. Weiss, and B. Michel, ''Evaluation and classification of physical and psychological stress in firefighters using heart rate variability,'' in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2019, pp. 2207–2212

[21] G. Shanmugasundaram, S. Yazhini, E. Hemapratha, and S. Nithya, ''A comprehensive review on stress detection techniques,'' in *Proc. IEEE Int. Conf. Syst., Comput., Automat. Netw. (ICSCAN)*, Mar. 2019, pp. 1–6

[22] J. Wijsman, R. Vullers, S. Polito, C. Agell, J. Penders, and H. Hermens, ''Towards ambulatory mental stress measurement from physiological parameters,'' in *Proc. Humaine Assoc. Conf. Affect. Comput. Intell. Interact.*, Sep. 2013, pp. 564–569

[23] S. Elzeiny and M. Qaraqe, ''Blueprint to workplace stress detection approaches,'' in *Proc. Int. Conf. Comput. Appl. (ICCA)*, Aug. 2018, pp. 407–412.

[24] S. Elzeiny and M. Qaraqe, ''Machine learning approaches to automatic stress detection: A review,'' in *Proc. IEEE/ACS 15th Int. Conf. Comput. Syst. Appl. (AICCSA)*, Oct. 2018, pp. 1–6.

[25] S. S. Panicker and P. Gayathri, ''A survey of machine learning techniques in physiology based mental stress detection systems,'' *Biocybern. Biomed. Eng.*, vol. 39, no. 2, pp. 444–469, Apr. 2019

[26] G. Giannakakis, D. Grigoriadis, K. Giannakaki, O. Simantiraki, A. Roniotis, and M. Tsiknakis, ''Review on psychological stress detection using biosignals,'' *IEEE Trans. Affect. Comput.*, early access, Jul. 9, 2019, doi: 0.1109/TAFFC.2019.2927337

[27] Y. S. Can, N. Chalabianloo, D. Ekiz, J. Fernandez-Alvarez, G. Riva, and C. Ersoy, ''Personal stress-level clustering and decision-level smoothing to enhance the performance of ambulatory stress detection with smartwatches,'' *IEEE Access*, vol. 8, pp. 38146–38163, 2020

[28] Y. S. Can, N. Chalabianloo, D. Ekiz, and C. Ersoy, ''Continuous stress detection using wearable sensors in real life: Algorithmic programming contest case study,'' *Sensors*, vol. 19, no. 8, p. 1849, Apr. 2019

[29] N. Keshan, P. V. Parimi, and I. Bichindaritz, ''Machine learning for stress detection from ECG signals in automobile drivers,'' in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Santa Clara, CA, USA, Oct. 2015, pp. 2661–2669

[30] V. Nasteski, ''An overview of the supervised machine learning methods,'' *Horizons*, vol. 4, pp. 51–62, Dec. 2017, doi: 10.20544/ horizons.b.04.1.17.p05.

A. Cantara and A. Ceniza, ''Stress sensor prototype: Determining the stress level in using a computer through validated self-made heart rate (HR) and galvanic skin response (GSR) sensors and fuzzy logic algorithm,'' *Int. J. Eng. Res. Technol.*, vol. 5, no. 3, pp. 28–37, 2016

[31] G. Mattavelli et al, Facial expressions recognition and discrimination in Parkinson's disease, J. Neuropsychol. 15 (1) (2021) 46–68

[32] Jing Zhang et.al, "Real time mental stress detection using multimodality expressions with deep learningframework",frontiers in neuroscience,doi 10.3389/fnins.2022.947168

[33] Wan-Ting Chew et.al "Facial Expression Recognition Via Enhanced Stress Convolution Neural Network for Stress Detection", IAENG International Journal of Computer Science, 49:3, IJCS_49_3_20, Volume 49, Issue 3: September 2022

[34] Menoua Keshishian et.al, "Understanding Adaptive, Multiscale Temporal Integration In Deep Speech Recognition Systems", 35th Conference on Neural Information Processing Systems (NeurIPS 2021).