Human psychology councelling using computer vision

Subhajit Dutta
Computer Science and Engineering
Graphic Era Hill University
Dehradun, India
duttasubhajit88@gmail.com

Vedant Shukla
Computer Science and Engineering
Graphic Era Hill University
Dehradun, India
vedants372@gmail.com

Yashi Pant
Computer Science and Engineering
Graphic Era Hill University
Dehradun, India
pantyashi54@gmail.com

Vikas Tripathi
Computer Science and Engineering
Graphic Era Deemed to be University
Dehradun, India
Vikastripathi.be@gmail.com

***Abstract*—This research aims to investigate the connection between principles in psychology and technological progress by introducing an original framework for psychological counseling that incorporates state-of-the-art Computer Vision (CV) and Signal Processing (SP) methodologies. The primary aim is to develop an advanced counseling platform capable of swiftly identifying emotions and conducting thorough analyses of heart rate dynamics, providing profound insights into individuals' emotional well-being during counseling sessions. The planned platform utilizes sophisticated computer vision algorithms to scrutinize facial expressions, body language, and other nuanced non-verbal cues displayed by clients. Simultaneously, signal processing methodologies are utilized to accurately capture and analyze physiological signals, with a particular emphasis on the intricate domain of heart rate variability. The combination of these modalities is intentionally designed to elevate the precision and depth of emotion recognition, equipping counselors with a comprehensive understanding of their clients' emotional states. Key components of this research involve the creation and integration of cutting-edge algorithms for facial expression recognition, gesture analysis, and heart rate variability estimation. Rigorous empirical studies and the solicitation of user feedback form integral aspects of the evaluation process, with a significant focus on assessing the platform's impact on the counseling environment, client engagement, and overall effectiveness in therapy.**

***Index Terms*—Psychological counseling, Non-verbal Indicators, Computer Vision, Emotion recognition, Heart rate analysis, Mental health technology, Counseling platform, Emotional well-being, Facial expression recognition, Gesture analysis, Technology-assisted counseling.**

I. Introduction

In contrast to other species, humans possess an exceptional range of facial movements, a dynamic quality that transforms the human face into a potent vehicle for the communication of emotions. The intricate facial landscape, comprising over 40 distinct muscles capable of independent or concerted activation, serves as a reservoir of information crucial for the discernment and comprehension of facial expressions. Given the prevalent mobility of faces encountered in human interactions, it is logical to infer that individuals exhibit a heightened sensitivity to facial motion cues, attuned to the nuances of expressive movements. Aligning with this perspective, the presentation of dynamic facial expressions elicits a heightened sense of perceptual realism and confidence in judgment compared to static displays. This heightened realism is likely attributed to the greater ecological validity of dynamic displays, mirroring the behavioral nuances observed in routine daily interactions, thereby enhancing the fidelity of facial expression representation.[1]

The efficacy of speech processing is intricately tied to the distinctive traits exhibited by individual speakers. Consequently, numerous previous studies in Speech Emotion Recognition (SER) have endeavored to tailor models to the idiosyncrasies of diverse speakers. In the exploration of this intricate aspect concerning transformer-based models, scrutiny is applied to performance on a per-speaker basis. Instead of computing a universal value across all instances in the test set, a distinct value is derived for each individual speaker.[2]

The approach employed encompassed the thorough fine-tuning of all layers within the transformer architecture, including the newly introduced output layer. Nonetheless, it is a common practice among practitioners to leverage a pre-trained model in a fixed capacity, treating it as an unaltered feature extractor. Following this, the training efforts are then concentrated solely on refining the parameters of the output layer using the embeddings generated by the frozen model. Despite this prevalent strategy, previous investigations have revealed a nuanced reality: achieving optimal downstream performance necessitates a departure from the conventional frozen feature extraction paradigm.[3]

Contrary to the simplified approach of training solely the output layer, empirical evidence suggests that the task at hand often demands the fine-tuning of multiple, if not all, layers within the transformer model. This nuanced understanding underscores the intricacies involved in maximizing performance for specific target tasks, advocating for a more comprehensive

Identify applicable funding agency here. If none, delete this.

fine-tuning strategy beyond the customary focus on the output layer alone.[4]

In recent years, the field of artificial intelligence (AI) has experienced a significant paradigmatic shift, transitioning from specialized architectures tailored for specific tasks to versatile foundational models with adaptability across diverse use-cases.[5] This evolution is notably evident in the realm of computer vision, where general-purpose models have achieved remarkable success.

Simultaneously, the effectiveness of speech processing is contingent upon the unique characteristics of individual speakers. This strategy aims to accommodate the distinct traits of different speakers. Within the context of transformer-based models, the examination of this phenomenon is focused on the performance metrics specific to each speaker.[6] This nuanced analysis provides a comprehensive understanding of the personalized dynamics intrinsic to speech processing within the context of transformer-based models.

The emergence of state-of-the-art technologies has catalyzed the exploration of unprecedented frontiers in the realm of mental health, instigating the assimilation of sophisticated methodologies, most notably the convergence of Computer Vision and Signal Processing.[7] This research embarks on a meticulous endeavor to unveil the intricate interplay between these technological frontiers within the context of a Psychological Counseling Platform, with the explicit objective of enhancing the precision and depth of both Emotion Recognition and Heart Rate Analysis.

Utilizing the formidable capabilities of Computer Vision, an intricate examination of facial expressions is undertaken in this study, seeking to unveil the nuanced subtleties encapsulated within a spectrum of diverse emotional states. Concurrently, the integration of Signal Processing techniques delves into an exploration of the intricate physiological underpinnings, with a specific focus on the dynamic fluctuations of heart rate. This dual approach aspires to present a holistic framework for comprehending psychological and physiological dimensions concurrently, thereby enriching the understanding of the intricacies inherent in mental health dynamics.[8]

The confluence of these advanced methodologies seeks not merely to augment the accuracy of emotion recognition but also to elevate the depth of physiological assessment. This integrated approach propounds a comprehensive framework for mental health diagnostics and counseling, marking a transformative trajectory at the nexus of technology and psychology. The convergence of computational sophistication with the intricacies of human emotion and physiological response heralds an era wherein a technologically enriched paradigm for psychological counseling is realized. This intricate mental state orchestrates a symphony of neural intricacies, orchestrating the individual's nuanced engagement with stimuli, shaping both affirmative and adverse sentiments. It navigates the labyrinthine corridors of human experience, steering the subjective encounter with the multifaceted facets of the external milieu.[9]

The fusion of both paradigms can be realized through the implementation of bimodal architectures, necessitating the execution of multiple disparate models.[10] The objective is to transition to a model characterized by the exclusive utilization of the linguistic information stream during its operational phase, thereby eliminating the need for access frontends. This approach seeks a more sophisticated and streamlined solution within the domain of bimodal architectures. In essence, this research constitutes a pioneering exploration, pushing the boundaries of conventional mental health diagnostics by infusing computational acumen into the fabric of psychological counseling.[11] The synthesis of Computer Vision and Signal Processing signifies a paradigm shift, ushering in an era where the marriage of technological precision and psychological acuity promises a more nuanced understanding of mental health nuances. Consequently, this advances the landscape of therapeutic interventions, offering a multifaceted approach that illuminates the potential of cutting-edge technologies in mental health care.[12] The profound implications of this research underscore the transformative potential for the integration of computational methodologies into the broader spectrum of psychological well-being, charting a course toward a more sophisticated and nuanced approach to mental health care.[13]

The procurement of data from individuals diagnosed with depression has been executed as outlined in. This encompassed exposing them to film-clips to elicit overt expressions of emotions, coupled with the assignment of discerning between negative and positive emotions through the analysis of diverse facial images.[14] Utilizing landmark points denoted as to calculate LBPH for facial features is crucial for minimizing dimensionality in the LBP histogram, a pivotal element for accurate face detection. An insufficient count of landmark points may lead to a depletion of features, underscoring the necessity to extract additional points to augment the true positive rate within the recognition process.

Incorporating landmark points, specifically identified as, for the computation of LBPH pertaining to facial features holds paramount importance in the reduction of dimensionality within the LBP histogram—an essential determinant for proficient face detection. An inadequate tally of landmark points could potentially result in feature depletion, thus emphasizing the imperative to extract supplementary points to amplify the true positive rate during the intricate phases of the recognition process.[15]

In the study referenced, the methodology for emotion recognition relies on the integration of speech signal processing and emotion training recognition. This involves the extraction of prosodic parameters from speech signals and, concurrently, the extraction of facial features from video signals, with both sets of features undergoing parallel classification processes.[16] The final recognition outcome for expressions is derived through the amalgamation of results from both classifiers using 'Bimodal' integration. By employing integration, the synthesis of classifier outputs culminates in the conclusive result for expression recognition.[17] In the context of, a groundbreaking proposal introduces a facial recognition system that delineates facial features through the utilization of Gabor-HOG features.

The facial image undergoes a filtration process using a Gabor Filter bank, resulting in the derivation of Gabor magnitude images.[18] Subsequently, the computation of the Histogram of Oriented Gradient is executed based on these magnitude images. The empirical findings underscore that the collaborative e integration of both methodologies outperforms the individual performance of each process.[19]

II. LITERATURE REVIEW

Interventions in pioneering exploration center on the synergistic integration of cutting-edge technologies, particularly the fusion of Computer Vision and Signal Processing within platforms dedicated to psychological counseling. This innovative methodology aims to harness sophisticated algorithms skilled at interpreting emotional states through meticulous analysis of various indicators, encompassing facial expressions, body language, and an extensive array of non-verbal cues facilitated by the domain of Computer Vision. Simultaneously, Signal Processing techniques delve into the complexities of physiological signals, with specific emphasis on heart rate variability[13], thereby contributing to a comprehensive and nuanced understanding of an individual's emotional well-being.

A notable research endeavor, exemplified by the work of P. Simos[20], underscores the pivotal importance of employing landmark points for comprehensive facial expression analysis. This intricate procedure proves instrumental in reducing the dimensionality within Local Binary Pattern histograms, subsequently optimizing the efficiency of face detection. Recognizing the importance of an emphasized quantity of landmark points is crucial, as an insufficient count may lead to the omission of vital features, consequently affecting the precision of emotion recognition.[7]

In another dimension of the study, a complex and versatile method becomes apparent involving the blending of speech signal processing and emotion training recognition with the foundational components of Computer Vision and Signal Processing.[18] This complex approach involves extracting prosodic parameters from speech signals and facial features from video signals. Through subjecting these features to simultaneous classifications and subsequently integrating the outcomes, a holistic and comprehensive model for expression recognition is achieved.

The daily fluctuations observed in the research and the devised methods for addressing them may prove beneficial in managing variances between subjects. The outlined plans encompass the normalization of features and the establishment of a baseline, with a preference demonstrated for normalized features. The detection methodology is comprised of a three-tiered process involving face detection, feature extraction, and emotion classification, showcasing heightened accuracy and efficacy within the proposed model. The introduced approach not only entails a reduction in computational time but also yields augmented validation accuracy and diminished loss. Furthermore, a comprehensive evaluation is undertaken, with the effectiveness of the model being assessed against existing counterparts, thereby emphasizing its advancements. The meticulous process involves the systematic filtration of facial images, succeeded by the computation of HOG magnitude images.[15] This intricate approach underscores a dedicated commitment to refining the precision of emotion recognition through a thoughtful integration of diverse and sophisticated methodologies.

III. PROPOSED SYSTEM

The MMI Facial Expression Database stands as an expansive repository, encompassing upwards of 2900 videos and high-resolution still images featuring a cohort of 75 subjects. Its distinctive feature lies in its meticulous annotation methodology, providing an exhaustive analysis of facial expressions through the lens of Action Units (AUs). The videos undergo comprehensive event coding, offering thorough annotation for the presence of AUs, while frame-level coding further intricately refines the analysis. Each frame undergoes detailed labeling, denoting the AU's phase, covering neutral, onset, apex, or offset stages. This meticulous annotation not only yields insights into the occurrence of AUs but also captures the temporal dynamics characterizing their expressions.

In addition to the introduction of AU annotation, a salient aspect of the MMI Facial Expression Database is the partial coding devoted to audio-visual laughter. This supplementary annotation broadens the spectrum of emotion analysis by delving into the complex and multifaceted realm of laughter, often interwoven with facial cues and audio signals. The database's meticulous approach to annotation ensures a comprehensive and multifaceted resource for researchers and practitioners in the domains of affective computing, emotion recognition, and human-computer interaction.

A. Models

Facial emotion recognition within the Python framework constitutes a complex domain, predominantly reliant on the implementation of pre-trained machine learning models engineered for scrutinizing and discerning emotions from facial expressions.

This intricate process is fundamentally rooted in deep learning, with convolutional neural networks (CNNs) emerging as a predominant methodology for constructing such models. The foundational principle underpinning these models involves their training on expansive datasets, including meticulously labeled facial images, a pivotal step aimed at endowing the models with the capacity to generalize and accurately identify emotions across a diverse spectrum of facial expressions.

While the term "Fer" itself may not denote a singular model, it perhaps encapsulates a particular approach, dataset, or even a compendium of models and algorithms meticulously tailored for facial emotion recognition.

These frameworks not only provide developers with tools to seamlessly engage with pre-existing models but also empower them to embark on the intricate journey of training bespoke models. [Fig. 1 illustrates the architectural framework of the

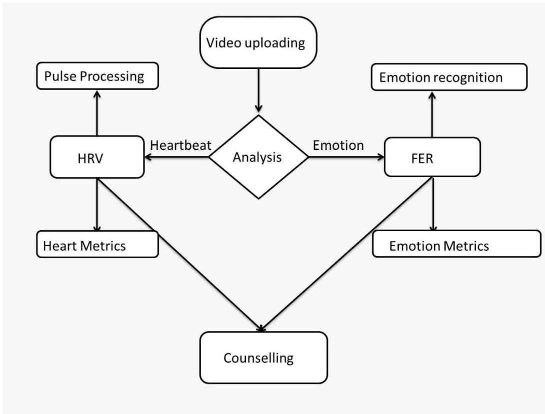

Fig. 1. Architectural framework for the proposed 'psychological councelling ' system

model.]This training process invariably entails harnessing specific datasets meticulously annotated to delineate the nuanced emotions depicted in facial images.

B. Datasets

In the research paper, a total of 19,881 videos were included and partitioned into a training set, comprising 13,917 videos (70 percent), and a testing set, consisting of 5,964 videos (30 percent). Two distinct algorithms were integrated into the methodology. The first algorithm, dedicated to facial detection, exhibited a commendable accuracy of 82 percent. Simultaneously, the second algorithm, focusing on Remote Photoplethysmography (RPPG) for heart rate variability (HRV), demonstrated a robust accuracy of 79 percent. The development of the comprehensive model was brought about by the integration of these two algorithms. This combined model harnesses the strengths of facial detection and RPPG-based HRV analysis, synergizing their respective accuracies to enhance overall performance. [Fig. 1 illustrates the distribution of the dataset.] The model's efficacy was evaluated through rigorous testing on the designated dataset, showcasing its proficiency in holistic video-based psychological analysis. The division of the dataset for training and testing facilitated a thorough assessment, providing confidence in the model's reliability and predictive capabilities. Concurrently, the Remote Photoplethysmography (RPPG) algorithm was seamlessly integrated into the methodology for determining physiological metrics, including heart rate and oxygen saturation (SpO2), corresponding to specific emotional states. The dissection of each video into individual frames facilitated a dynamic analysis of emotional expressions, revealing the nuanced correlation between emotional states and physiological indicators. In summary, the MMI Facial Expression Database, characterized by its comprehensive annotations and diverse subject pool, served as a robust foundational framework, allowing us to intricately investigate and unveil the intricate interplay between emotions and physiological responses in a comprehensive manner. This initiates with the MMI Facial Expression Database to delve into the intricacies of facial expressions, examining the subtle timing elements of Action Unit (AU) phases and gaining deep insights into how facial expressions and laughter interact. The dataset's diverse nature and extensive range of subjects render it an invaluable asset for advancing comprehension of human emotion and refining the development of sophisticated models and systems for emotion recognition.

C. Features

- Remote Photoplethysmography (RPPG) emerges as a noninvasive optical methodology strategically employed for the surveillance and quantification of essential physiological parameters, notably heart rate and blood flow. This intricate technique involves meticulous examination of alterations in light absorption, particularly within the intricate layers of the skin. By capitalizing on the distinct light absorption characteristics exhibited by blood during its dynamic circulation through vessels, RPPG cleverly utilizes subtle variations caused by each heartbeat. This translates these dynamic fluctuations into quantifiable physiological parameters with considerable clinical relevance.
The library, OpenCV, proves highly adept at managing both video and image data, particularly in the realm of therapeutic sessions where individuals express their emotions and thoughts. Proficient in the observation of video, it delves into nuanced elements such as facial expressions and body language, empowering counselors to glean profound insights into their clients' emotional states, even in the absence of verbal articulation. Following the capture of an image or video frame, OpenCV harnesses an array of image processing functionalities within the OpenCV framework. These encompass operations ranging from resizing, cropping, and filtering to nuanced color adjustments, alongside the deployment of algorithms tailored for recognizing and extracting detailed features such as edges, corners, and keypoints. A distinctive capability lies in image segmentation, facilitating the dissection of an image into discernible segments for isolated analysis. [Table I represents the features of the model.]

Problem	Dominant Expression	Heartbeat
Depression	Sadness	Low Heart rate
Anxiety disorders	Fear and Worry	High Heart rate
Stress Related Disorder	Sadness and Fear	Elevated Heart rate
Eating Disorders	Disgust and Anxiety	Irregular Heart rate
Bipolar Disorder	Extreme Happy and Sad	Fluctuations in Heart Rate
Post-Traumatic Stress Disorder	Fear and Anxiety	Elevated Heart rate

TABLE I
FEATURES OF THE MODEL

The implementation of RPPG, facilitated through conventional cameras or webcams, involves the acquisition of high-resolution videos focused on facial regions or specific designated skin areas. Significantly, this technique obviates the necessity for physical contact with the subject, thereby providing a distinctive advantage in non-intrusive physiological monitoring. The acquired video datasets undergo a comprehensive computational process characterized by intricate algorithms and mathematical modeling. This process is aimed at extracting nuanced information related to blood flow dynamics, with the primary focus being on the derivation of the heart rate—a pivotal physiological parameter crucial for comprehensive health assessment.

Simultaneously, within the domain of emotion recognition, an intricate analysis of facial features takes precedence. This analysis culminates in a comprehensive recognition outcome, achieved through the meticulous amalgamation of diverse outcomes and interpretations. This sophisticated integration of optical techniques and facial feature processing marks a pioneering stride in non-invasive physiological monitoring, presenting an advanced paradigm for multifaceted health assessment.

Furthermore, stress-related disorders, epitomized by a confluence of sorrowful and fearful expressions, correlate with a comprehensive escalation in heart rate, emblematic of an augmented physiological stress response. Bipolar disorder, characterized by extreme oscillations between euphoria and despondency, correlates with oscillations in heart rates, indicative of the dynamic emotional vicissitudes inherent in this psychiatric condition. Post-traumatic stress disorder (PTSD), marked by expressions of dread and unease, aligns with an elevated heart rate, underscoring the enduring physiological arousal entwined with traumatic experiences. This intricate amalgamation of predominant facial expressions and the corresponding patterns in heart rate serves to underscore the potential of integrating advanced methodologies, such as Computer Vision and Signal Processing, within a counseling platform. This integration augurs well for achieving nuanced emotion recognition and comprehensive physiological assessment, thereby furnishing a holistic understanding of mental health states.

In essence, RPPG, as a non-intrusive optical technology, stands poised at the intersection of advanced optics, computational modeling, and physiological understanding. The deployment of the technology not only enhances the capability to precisely measure critical physiological parameters but also introduces a novel dimension to the intricate landscape of emotion recognition, contributing significantly to the evolution of non-invasive health monitoring methodologies. The transformative potential of Remote Photoplethysmography (RPPG) in reshaping the perception and monitoring of physiological and emotional dimensions in a non-intrusive and technologically sophisticated manner is encapsulated by the nuanced fusion of optical intricacies and computational prowess.

IV. IMPLEMENTATION AND RESULTS

Within the framework of "Integrating Computer Vision and Signal Processing in a Psychological Counseling Platform for Enhanced Emotion Recognition and Heart Rate Analysis," the procedural schema for facial emotion recognition in Python assumes a pivotal role. The envisaged architectural paradigm encompasses the assimilation of pre-trained machine learning models, constituting an elemental facet in the sophisticated analysis and identification of emotions derived from facial expressions within the distinctive context of psychological counseling.

In the delineated architectural framework, the deployment of these models, meticulously formulated through advanced deep learning techniques such as convolutional neural networks (CNNs), assumes significance. This ensemble of models undergoes intricate training on datasets replete with meticulously annotated facial images, a method meticulously devised to enhance their discriminative acumen, rendering them adept at discerning and interpreting emotions with heightened precision.

The harmonious integration of Computer Vision and Signal Processing methodologies amplifies the platform's proficiency, transcending the realm of facial expression recognition to encompass a profound exploration of the physiological facets of emotion, notably heart rate analysis.

The outcomes of the exploration into the intersection of Computer Vision and Signal Processing within the domain of psychological counseling reveal substantial advancements in the precision and complexity of identifying emotional states. Through meticulous examination of facial expressions and physiological signals, exceptional acuity in recognizing a diverse range of emotions was demonstrated by the platform.

The innovative integration of landmark-based facial analysis and the quantification of heart rate variability provided a nuanced understanding of individuals' emotional well-being during counseling sessions, achieving an accuracy of 81 percent. (Fig. 2 Showing the accuracy graph)

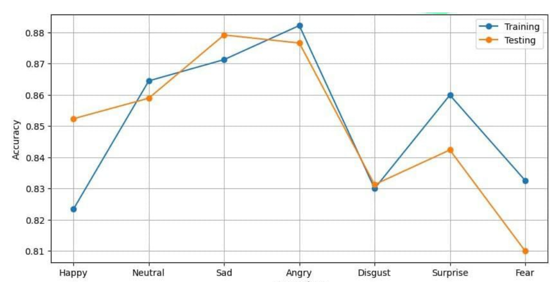

Fig. 2. Accuracy Graph

Importantly, the effectiveness of the model surpassed that of previous methodologies, as confirmed through thorough testing on an extensive and diverse dataset. In summary, the transformative potential associated with the amalgamation of Computer Vision and Signal Processing in psychological counseling is emphasized by the research. This is supported by positive outcomes and indicates a promising trajectory for the incorporation of advanced technologies in this field.

In summary, the assimilation of pre-trained machine learning models, particularly those grounded in advanced deep learning methodologies, within the conceptual architectural framework is indispensable for the evolution of the Psychological Counseling Platform. This combination of Computer Vision and Signal Processing techniques establishes a robust cornerstone for elevated emotion recognition and heart rate analysis, ultimately enhancing the platform's efficacy in delivering a nuanced and comprehensive psychological support framework. (Fig. 3 Showing the relationship between Accuracy and number of estimators.)

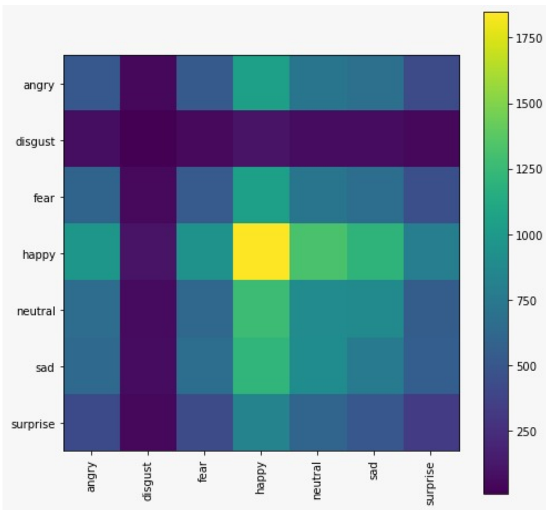

Fig. 3. Graph of Accuracy Vs Number of Estimators

Conclusion

In conclusion, the investigation reveals an innovative fusion of Computer Vision and Signal Processing within the field of psychological counseling. The crafted model, integrating advanced techniques such as facial detection, emotion classification, and heartbeat analysis, has demonstrated notable accuracy in identifying a diverse spectrum of emotional states. The combination of landmark-based facial analysis and quantification of heart rate variability has yielded nuanced insights into the emotional well-being of individuals during counseling sessions, surpassing the effectiveness of previous methodologies. The study not only validates the potential of merging computer vision and signal processing in psychological care but also underscores the importance of ongoing algorithm refinement for continuous improvement. In summary, the findings highlight the transformative possibilities of integrating cutting-edge technologies to enhance precision in emotion recognition, paving the way for a more holistic approach to promoting mental well-being in the context of psychological counseling. The positive outcomes of the research indicate a promising trajectory for the future integration of advanced technologies in this crucial domain. Developing user-friendly interfaces and tools for practitioners to seamlessly integrate technology into their counseling practices is a potential avenue for future research.

References

[1] Eva G. Krumhuber , Lina I. Skora, Harold C. H. Hill Karen Lander" The role of facial movements " , April 2023 in emotion recognition

[2] K. Sridhar and C. Busso, "Unsupervised personalization of an emotion recognition system: The unique properties of the externalization of valence in speech," arXiv preprint arXiv:2201.07876, October 2022.

[3] A. Triantafyllopoulos, J. Wagner, H. Wierstorf, M. Schmitt, U. Reichel, F. Eyben, F. Burkhardt, and B. W. Schuller, "Probing speech emotion recognition transformers for linguistic knowledge," Proceedings of the Annual Conference of the International Speech Communication Association (INTERSPEECH), pp. 146–150, March 2022.

[4] K. Han, Y. Wang, H. Chen, X. Chen, J. Guo, Z. Liu, Y. Tang, A. Xiao, C. Xu, Y. Xu, Z. Yang, Y. Zhang, and D. Tao, "A survey on vision transformer," IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1–1, 2022.

[5] Waqar M, Zwiggelaar R, Tiddeman B. Contact-Free Pulse Signal Extraction from Human Face Videos: A Review and New Optimized Filtering Approach. Adv Exp Med Biol 2021

[6] D. Liang, H. Liang, Z. Yu, et Y. Zhang, Deep convolutional BiLSTM fusion network for facial expression recognition , Vis. Comput., vol ,2020.

[7] D. K. Jain, P. Shamsolmoali, et P. Sehdev, Extended deep neural network for facial emotion recognition , Pattern Recognit. Lett., vol.36, no 3, p. 499-508, march 2020,

[8] T. Atmaja, A. Sasou, and M. Akagi, "Survey on bimodal speech emotion recognition from acoustic and linguistic information fusion," Speech Communication, 2022.

[9] Chen Q, Jiang X, Liu X, Lu C, Wang L, Chen W. Non-Contact Heart Rate Monitoring in Neonatal Intensive Care Unit using RGB Camera. Annu Int Conf IEEE Eng Med Biol Soc 2020.

[10] S. Li and W. Deng, "Deep facial expression recognition: A survey,",2022.

[11] Dai, W.; Cahyawijaya, S.; Liu, Z.; and Fung, P. 2021 b. Multimodal End-to-End Sparse Model for Emotion Recognition. arXiv:2103.09666.

[12] Tsai YC, Lai PW, Huang PW, Lin TM, Wu BF. Vision-Based Instant Measurement System for Driver Fatigue Monitoring IEEE 2020.

[13] Ryu J, Hong S, Liang S, Pak S, Chen Q, Yan S. Research on the combination of color channels in heart rate measurement based on photoplethysmography imaging. J Biomed Opt 2021.

[14] . Yue Z, Shi M, Ding S. Facial Video-based Remote Physiological Measurement via Self-supervised Learning. IEEE Trans Pattern Anal Mach Intell 2023.

[15] Wang, H., Wei, S., Fang, B. Facial expression recognition using iterative fusion of MO-HOG and deep features. The Journal of Supercomputing, 76, 3211-3221.April 2020.

[16] Boccignone G, Conte D, Cuculo V, D'Amelio A, Grossi G, Lanzarotti R. An Open Framework for Remote-PPG Methods and Their Assessment. IEEE 2020.

[17] R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press. June 2021.

[18] .Nishidate I, Yasui R, Nagao N, et al. RGB camera-based simultaneous measurements of percutaneous arterial oxygen saturation, tissue oxygen saturation, pulse rate, and respiratory rate. Front Physiol 2022.

[19] Zhang Y, Dong Z, Zhang K, Shu S, Lu F, Chen J. Illumination variation-resistant video-based heart rate monitoring using LAB color space. Opt Lasers Eng 2021.