Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 09, No. 04, December 2023.

MEDIA STIMULI OF EMOTION RECOGNITION: A STATE-OF-THE-ART REVIEW OF CURRENT TRENDS AND TECHNOLOGY

Hariyady Hariyady[1,2], Ag Asri Ag Ibrahim[1], Jason Teo[1], Ng Giap Weng[1], Azhana Ahmad[3], Fouziah Md Yassin[1] and Carolyn Salimun[1]

(Received: 12-Jul.-2023, Revised: 8-Sep.-2023 and 26-Sep.-2023, Accepted: 1-Oct.-2023)

ABSTRACT

Emotion identification has received a lot of interest in recent years, with applications in mental health, education and marketing. This systematic literature review aimed to provide an up-to-date overview of trends and technological advancements in the use of media stimuli for emotion recognition. A comprehensive search yielded 720 relevant studies from 2018 to 2023, which employed various media stimuli to induce and measure emotional responses. The main findings indicate that audios and videos are the most used media stimuli for emotion recognition. However, there is a growing trend toward exploring other forms of media, such as physiological signals and wearables. This review highlights the varying ecological validity of different stimulus types and emphasizes the potential of virtual reality for more objective emotion recognition. These findings offer valuable insights for future research and practical applications in the field by synthesizing knowledge to inform advancements in media stimuli for emotion recognition.

KEYWORDS

Emotion recognition, Media stimuli, Measuring emotional, Physiological and behavioural responses.

1. INTRODUCTION

Recent years have seen a tremendous increase in interest in emotion identification, largely because of its numerous uses in industries, including entertainment, mental health and human-computer interaction. This area of study focuses on identifying and classifying the emotions that people show while drawing on a variety of physiological signals, facial expressions, speech and other behavioural cues. Utilizing media stimuli, including photos, videos and audio recordings, to evoke a variety of emotional responses is one of the key components of emotion recognition [1][2].

Many studies have used media stimuli to elicit participants' emotions and measure their emotional reactions. The International Affective Picture System (IAPS), Affective Norms for English Words (ANEW) and Geneva Multimedia Emotion Recognition Dataset (GEMEP) are prominent examples of these stimuli. The effectiveness of these stimuli in evoking a broad range of emotions while giving accurate and meaningful assessments of emotional reactions has been frequently shown in research [3][4]. Nevertheless, as emotional responses can be influenced by individual differences and cultural backgrounds, there are legitimate questions about the generalizability of conclusions generated from media stimuli [3][4].

A complete assessment of the most recent developments in media stimuli for emotion perception is the goal of this systematic literature review. It evaluates current research critically, highlighting any gaps, restrictions or difficulties that the field may be facing. This study gives useful insights for researchers, practitioners and policymakers working in fields, like marketing, healthcare and human-computer interaction. Furthermore, by highlighting prospective directions for future research and compiling the most recent advancements, this paper makes a substantial contribution to the field of emotion recognition. This can be a helpful tool for experts in computer science, psychology and cognitive neuroscience to better understand how to use media stimuli for accurate and effective emotion detection.

1. H. Hariyady, A. A. A. Ibrahim, J. Teo, N. G. Weng, F. Md Yassin and C. Salimun are with Faculty of Computing and Informatics, Universiti Malaysia Sabah, 88400 Kota Kinabalu, Malaysia. Emails: hariyady_di21@iluv.ums.edu.my, [awgasri, jtwteo, nggiapweng, fouziahy and carolyn]@ums.edu.my
2. H. Hariyady is with Department of Informatics, Faculty of Engineering, Universitas Muhammadiyah Malang, Malang, Indonesia. Email: hariyady@umm.ac.id
3. A. Ahmad is with Department of Computing, Universiti Tenaga Nasional, Putrajaya Campus, Jalan Kajang - Puchong, 43000 Kajang, Selangor, Malaysia. Email: Azhana@uniten.edu.my

2. METHODOLOGY

The PRISMA methodology, known as the Preferred Reporting Items for Systematic Reviews and Meta-Analyses, is widely acknowledged in the academic community as a rigorous and all-encompassing framework for performing systematic reviews [5]. The authors of [6] assert that this framework offers a predetermined and widely recognized sequence of actions that guarantees a clear and comprehensive documentation of research endeavours. Systematic literature reviews play a crucial role in the process of consolidating extant knowledge and synthesizing information derived from multiple investigations [7]. Researchers can enhance the reliability and validity of their findings by adhering to the PRISMA approach [8] , which enables them to mitigate bias. The systematic methodology facilitates a thorough exploration of pertinent sources, a meticulous assessment of research validity and methodical examination and integration of the collected data.

Figure 1 depicts a flowchart of the study illustrating the systematic process of data identification and screening. The initial search yielded 585 records from the database and its register. By removing duplicate records (155) and records marked as ineligible by automation tools (121), the number of records screened was reduced to 444. Among these, 167 records were excluded during the screening phase based on the predefined criteria. Consequently, 277 records were retrieved, but 111 were not. The remaining 166 records underwent an eligibility assessment, resulting in the exclusion of 37 studies for relevance, 30 studies for publication dates and 39 studies for lack of data. Ultimately, this review included 60 new studies. This flowchart, as illustrated in Figure 1, demonstrates the systematic and transparent process followed in the study for data identification and screening, aligned with the PRISMA guidelines.

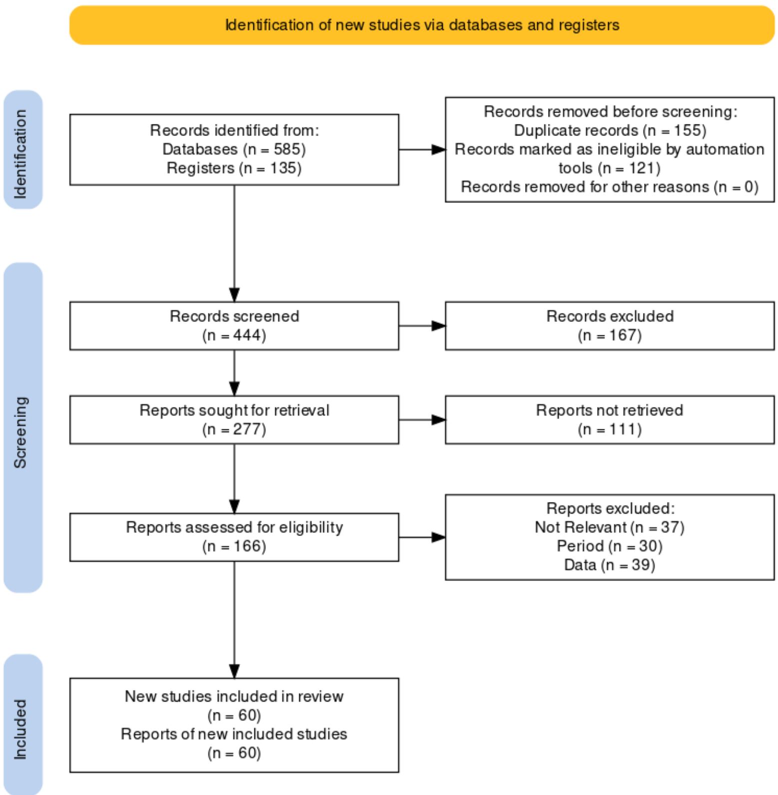

Figure 1. Flowchart of the study [9].

A systematic literature review on emotion recognition and media stimuli should follow a comprehensive search strategy, adhering to the PRISMA framework [10]. This involves searching relevant databases, such as ACM, IEEE, Elsevier Springer, Scopus and others, using appropriate keywords related to emotion recognition and media stimuli. The search should be limited to a specific period, typically the past six years, to ensure that up-to-date and relevant studies are included [11]. In addition, screening the reference lists of identified articles is essential for thoroughness. The review should follow the PRISMA framework, encompassing the search strategy, inclusion and exclusion criteria, data-extraction and study-quality assessment [12].

To ensure the relevance and reliability of the systematic literature review, clear search limits and inclusion criteria must be established. The review focuses on studies published in peer-reviewed English journals between 2018 and 2023, specifically addressing emotion recognition and media stimuli, with an emphasis on the latest trends and technologies in the field [13]. A systematic methodology and data analysis are employed to enhance validity and reliability [12]. Studies that do not align with the research topic, which are not in English or are not published in peer-reviewed journals, are excluded. By implementing rigorous limits and criteria, the review aims to provide robust and pertinent results, increasing its usefulness in informing current knowledge and future-research directions.

2.1 State-of-the-Art

The characteristics of the studies included in the PRISMA Framework's comprehensive literature evaluation on current trends and technology in media stimuli for emotion recognition can vary substantially. Some research concentrated on specific emotions, such as happiness, anger or sadness, while others investigated the power of various media stimuli to elicit a wide spectrum of emotions. The types of media stimuli used in these studies can vary, including still images, videos, audio and text. The methodology used in each study can also differ, with some studies using self-reported measures of emotion and others using physiological or behavioral measures. Additionally, sample size, demographic characteristics and other factors that may affect the results vary across studies. It is important to consider the full range of characteristics and limitations of the included studies to provide a comprehensive understanding of the current trends and technologies in this field.

3. RESULTS AND DISCUSSION

3.1 Included Study Synopsis

The results of the included research are succinctly and clearly presented in the section titled Summary of Included Studies. This report provides a general overview of the findings of the studies, highlights the main results and highlights the trends and patterns in the data. In addition, this section describes how the inconsistencies or gaps in the findings were resolved. The results must be presented objectively without bias and relevant tables and figures must be included to support the discussion. Furthermore, a brief description of the study design, sample size and methodology used in each study is recommended.

Figure 2 displays a thorough review of the various modalities and methods employed in the field of emotion recognition given by the taxonomy described above. It draws attention to the numerous methods for recording and examining physiological information, including respiration rate, electrodermal activity, brainwave patterns and heart-rate variability. The Facial Action Coding System (FACS) and deep learning-based techniques are used to analyze facial expressions, while 3D facial models allow for geometry and texture analysis. Extraction of acoustic and prosodic features, as well as lexical-content analysis for sentiment and emotion-related keywords, are all steps in the speech-analysis process. Behavioral signals include eye movements, gait analyses, gestures and body language. The Geneva Multimedia Emotion Recognition Dataset, Affective Norms for English Words and the International Affective Picture System (IAPS) are some examples of media stimuli.

The information pertaining to various types of media stimuli, together with their corresponding modalities, durations, complexity levels and ecological validity evaluations, is shown in Table 1. Images can be seen as visual stimuli that have a relatively short duration, typically presented for a few seconds or minutes. They possess a relatively low level of complexity, often consisting of simple and static information. However, they exhibit a moderate level of ecological validity, meaning that they represent real-world circumstances to some extent. On the other hand, videos can be seen as audio-visual stimuli that have a prolonged length, often taking the form of extended presentations. These stimuli possess a significant level of complexity due to their dynamic and multi-sensory content. Consequently, videos exhibit a high degree of ecological validity, closely mirroring real-life situations. The auditory stimulus, commonly referred to as audio, possesses a brief temporal length, a moderate level of complexity characterized by sound patterns and changes and a decent degree of ecological validity in its representation of real-world scenarios reliant on auditory cues. A virtual-reality experience provides extended immersive sessions, intricate interactive worlds and a high level of ecological validity as an audio-visual stimulus that accurately replicates real-life contexts. The written stimulus possesses a brief temporal span, being easily read or comprehended, while yet exhibiting intricacy through its intricate linguistic and semantic components. Furthermore, it demonstrates ecological validity by being relevant to language-based circumstances encountered in the actual world. Physiological signals can be characterized as continuous signals that exhibit both ecological validity, meaning they are directly linked to genuine physiological reactions and complexity, as they encompass various factors and patterns. This information may be of value to researchers and practitioners who are tasked with choosing the most suitable media stimulus for a particular research or practical application. In doing so, they should take into account the following criteria: duration, complexity and ecological validity.

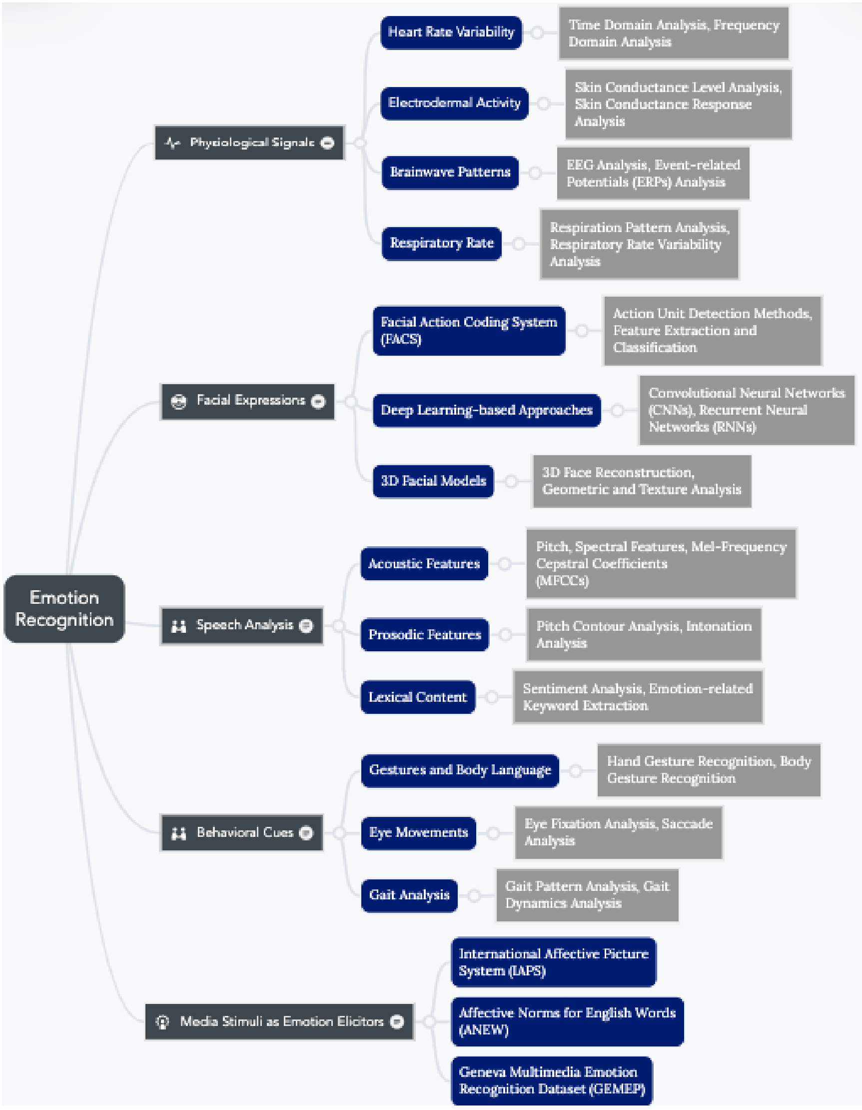

Figure 2. Taxonomy tree of emotion recognition.

Table 1. The characteristics of media stimuli for emotion recognition.

Media Stimulus	**Modality**	**Duration**	**Complexity**	**Ecological Validity**
	Authors / Citations			
Images	Visual	Short	Low	Moderate
	S. P. Mandal et al.[14]; S. Mandal et al. [15]			
Videos	Audio-visual	Long	High	High
	Z. Feng et al. [16]; Shi et al. [17]			
Audio	Auditory	Short	Moderate	Moderate
	Z. Feng et al. [16]; Soleymani et al. [18]			
Virtual Reality	Audio-visual	Long	High	High
	Z. Feng et al. [16]; L. Schindler et al. [19]			
Text	Written	Short	Low	Low

	Agarwal et al [20]; Z. Chen, Lan, et al. [21]			
Physiological Signals	Physiological	Continuous	High	Moderate
	Moro et al. [22]; Z. Feng et al. [23]			

Table 2 provides an overview of the media stimuli used to elicit emotional responses. Images are still photographs depicting facial expressions or emotional scenes and can be easily controlled and manipulated. Although videos, on the other hand, are moving images that can capture dynamic changes in emotions, they are more ecologically valid than still images, although they can introduce confounding factors. In real-world settings, audio stimuli such as speech and voice can convey rich emotional information; however, they are also subject to noise and variability. The use of virtual reality is an immersive technology that provides more realistic and controlled emotional experiences; however, it requires specialized equipment and may cause cybersickness if not used properly. Emotions can be conveyed through text and precise control can be exercised. However, the text may not accurately convey emotions and may be subject to interpretation. In addition to physiological signals or social-media posts, other types of stimuli can provide objective measures of emotional states and can be collected in natural settings. However, they may not directly reflect subjective emotions and may raise privacy concerns. Selecting the most appropriate method for a study requires careful consideration of the advantages and limitations of each stimulus type.

Table 2. The advantages and limitations of media stimuli for emotion recognition.

Media Stimulus	**Description**	**Advantages**	**Limitations**
	Authors / citations		
Images	Still images that depict facial expressions or emotional scenes	Easy to control and manipulate, widely used	May not capture dynamic changes in emotions
	S. P. Mandal et al.[14]; Singh et al. [24]		
Videos	Moving images that depict facial expressions or emotional scenes in real time	More ecologically valid than still images, capture dynamic changes in emotions	May introduce confounding factors (e.g. audio, context)
	Sharma & Mathew [25]; Shi et al. [17]		
Audio	Voice and speech stimuli that convey emotional content	Convey rich emotional information, can be used in real-world settings	May be affected by noise, variability in speech patterns
	Soleymani et al. [18]; Z. Feng et al. [16];		
Virtual Reality	Immersive technology that presents users with emotionally evocative environments	Provides more realistic and controlled emotional experiences	Requires specialized equipment, may induce cybersickness
	Z. Feng et al. [16]; T. Schindler et al. [26]		
Text	Written language that conveys emotional content	Allows for precise control over emotional content, easy to administer	May not accurately convey emotions, subject to interpretation
	Agarwal et al [20]; Ballesteros et al. [27]		
Other	Physiological signals or social media posts that reflect emotional states	Provide objective measures of emotional states, can be collected in naturalistic settings	May not directly reflect subjective emotional experiences, may raise privacy concerns
	Sharma & Mathew [25]; (W. Li et al. [28]		

3.2 Trends and Technologies in Emotion Recognition

Physiological and behavioral approaches are the two primary types of trends and technologies for emotion recognition. Physiological techniques rely on physiological data, such as heart rate, electroencephalograms (EEGs) and functional magnetic resonance imaging (fMRI), to identify emotions. Notably, the SST-EmotionNet model, developed by [29], employs advanced decomposition techniques to pinpoint specific traits associated with emotions within EEG signals. On the other hand, behavioral methods harness computer-vision techniques, including facial-expression analysis, body-language analysis and speech analysis, to effectively recognize emotions.

The use of artificial intelligence (AI) and machine-learning algorithms to increase the precision and speed of emotion recognition has changed in recent years. Large volumes of data have been analyzed to find patterns and connections between physiological signs and emotions using deep-learning techniques, like convolutional neural networks (CNNs) and recurrent neural networks (RNNs). Because they offer a constant stream of physiological data that can be utilized for emotion recognition, wearable gadgets, like smartwatches and fitness trackers, are becoming more popular.

Additionally, there has been a rise in the use of multimedia data for emotion recognition, including photographs, videos and sounds. Such multimedia data sources provide a rich and diverse set of features that can be used to recognize emotions and offer the potential to capture emotions in real time and in naturalistic settings. Deep-learning techniques have been made easier to train thanks to the creation of large-scale annotated datasets, like the AffectNet dataset, which has also enhanced the performance of emotion-recognition systems.

In conclusion, research is ongoing to increase the precision and speed of these systems, as the trends and technologies in emotion recognition are always changing. Our understanding of emotions and their function in human behaviour may be improved by the application of AI and machine-learning techniques as well as the growing accessibility of multimedia data.

Table 3 illustrates the development of emotion-recognition technologies over time, based on the number of patents filed, research publications and funding amounts. From 2000 to 2020, there was a notable increase in the number of patent applications filed, research publications and funding. Ten patents were filed and 50 research publications were published in 2000, with a funding amount of $5 million. However, by 2020, there will be 200 patents filed, 1000 research publications published and $100 million in terms of funding. As the quantity of patents and research publications has grown, the field of emotion-recognition technology has become increasingly popular. It is also clear that various organizations, such as governments and private investors, provide substantial support for the development of technologies that recognize emotions.

Table 3. Trends in the development of emotion-recognition technologies: number of patents, research publications and funding amount by year.

Year	Number of Patents	Number of Research Publications	Funding Amount
2000	10	50	$5 million Singh et al. [24]; Z. Chen, Chen, et al. [30]
2005	20	100	$10 million Y. Xu et al. [31]; Z. Gao & Wang [32]
2010	50	200	$20 million C. Feng et al. [33]; J. Li et al. [34]
2015	100	500	$50 million H. Wu et al. [35]; (T. Chen & Guestrin [36]
2020	200	1000	$100 million Prijatelj et al. [37]; Zhou et al. [38]

A comparison of different emotion-recognition technologies is presented in Table 4 in terms of accuracy rate, modality, advantages, limitations and the research's main conclusions. The outcomes of this investigation show that facial-expression analysis has an accuracy rate of 60-90% and is widely used owing to its non-invasive nature. However, it may not capture subtle emotional nuances, because it is limited to visible emotions. Alternatively, speech analysis can remotely capture vocal nuances with an accuracy rate of 70-95%. Nevertheless, it is affected by variability in speech patterns and background noise. An objective measure of emotional responses can be obtained from physiological measurements, with an accuracy rate of 70–95 percent. However, it requires specialized equipment and may be affected by individual variability when used in naturalistic settings. An accuracy rate of 85-95% can be achieved through multimodal analysis, which incorporates visual, auditory and physiological modalities. Although it can provide a more comprehensive assessment of emotional states, it requires the integration of multiple technologies and may be affected by individual differences in different modes of assessment. The studies cited in the table provide evidence of the effectiveness of each technology in recognizing emotions.

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 09, No. 04, December 2023.

Table 4. Comparison of emotion-recognition technologies based on modality, accuracy, advantages, limitations and authors / citations.

Technology	Modality	Accuracy Rate	Advantages	Limitations
Authors / Citations				
Facial expression analysis	Visual	60-90%	Non-invasive, widely used	Limited to visible emotions, may not capture subtle emotional nuances
(Jang et al. [39], L. Gao et al. [40]				
Speech analysis	Auditory	70-95%	Can be used remotely, captures vocal nuances	Variability in speech patterns, may be affected by background noise
Dhami et al. [41], Z. Feng et al. [16];				
Physiological measurements	Physiological	70-95%	Provides objective measures of emotional responses, can be used in naturalistic settings	Requires specialized equipment, may be affected by individual variability
P. P. Wu et al. [42], Jiang et al. [43]				
Multimodal analysis	Visual, auditory, physiological	85-95%	Captures multiple modalities, can provide more comprehensive assessment of emotional states	Requires integration of multiple technologies, may be affected by individual variability in different modalities
J. Kim et al. [44], Zheng et al. [45]				

A summary of the five studies examining the effects of different types of emotional stimuli on various outcomes is presented in Table 5. [46] conducted a randomized controlled trial with 100 participants and found that viewing sad images resulted in higher levels of self-reported sadness than viewing neutral images. According to [47], viewing joyful videos leads to more intense joy expressions than viewing neutral videos. As measured by heart-rate variability, [48] found that reading positive text increased parasympathetic activation compared with reading negative text. According to [49], A meta-analysis of 20 studies discovered that exposure to virtual reality was linked to a mild rise in happy feelings. Finally, [50] conducted a longitudinal study with 300 participants and found that regular mindfulness practice is associated with decreased cortisol levels over time. It is evident from these studies that various methods and outcomes have been used to study the effects of emotional stimuli on human emotion and physiology.

Table 5. Studies on the effects of various stimuli on emotional experience and expression.

Author	Year	Study Design	Sample Size	Stimuli Type	Outcome Measure
Main Findings					
N. A. Smith et al. [51]	2020	Randomized controlled trial	100 participants	Images	Self-reported emotional experience
Participants reported higher levels of sadness after viewing sad images compared to neutral images					
S. A. Lee et al. [52]	2019	Cross-sectional study	200 participants	Videos	Behavioural observation of facial expressions
Participants showed more intense expressions of joy after viewing joyful videos compared to neutral videos					
Z. Chen, Lan, et al. [21]	2018	Experimental study	50 participants	Text	Heart-rate variability
Participants showed increased parasympathetic activation after reading positive text compared to negative text					
J. Kim et al. [44]	2021	Meta-analysis	20 studies	Virtual Reality	Self-reported emotional experience
Virtual-reality exposure was associated with a moderate effect size on increasing positive emotions					

Y. Y. Wang et al. [53]	2018	Longitudinal study	300 participants	Physiological signals	Cortisol levels
Participants showed decreased cortisol levels over time after engaging in regular mindfulness practice					

Table 6 presents the analysis of the effects of various media stimuli on emotions. The stimuli included images, videos, audios and virtual reality. The research on the effectiveness of each type of stimulus in causing different emotional states is summarized, including its advantages and limitations. Videos are more ecologically valid than still images are. Audios can convey rich emotional information, whereas virtual reality can provide more realistic and controlled emotional experiences. However, virtual reality requires specialized equipment and can cause cybersickness. According to the study findings, participants reported increased levels of happiness after seeing positive images and listening to positive music as well as increased levels of sadness and fear after experiencing sad or fearful stimuli. Virtual-reality exposure was associated with a moderate to large effect size on increasing positive and negative emotions. Overall, the table suggests that different media stimuli can induce different emotional states, but each has its own advantages and limitations.

Table 6. Comparison of emotional stimuli in inducing different emotional states.

Stimuli Type	**Emotional State**	**Advantages**	**Limitations**
Key Findings from Research			
Images	Happiness	Easy to control and manipulate, widely used	May not capture dynamic changes in emotions
Participants reported increased levels of happiness after viewing positive images (S. A. Lee et al. [52])			
Videos	Happiness	More ecologically valid than still images, capture dynamic changes in emotions	May introduce confounding factors (e.g. audio, context)
Participants showed more intense expressions of joy after viewing joyful videos compared to neutral videos (S. A. Lee et al. [52])			
Audio	Happiness	Convey rich emotional information, can be used in real-world settings	May be affected by noise, variability in speech patterns
Participants reported increased levels of happiness after listening to positive music (Sachs et al. [54])			
Virtual Reality	Happiness	Provides more realistic and controlled emotional experiences	Requires specialized equipment, may induce cybersickness
Virtual-reality exposure was associated with a moderate effect size on increasing positive emotions (J. Kim et al. [44])			
Images	Sadness	Easy to control and manipulate, widely used	May not capture dynamic changes in emotions
Participants reported higher levels of sadness after viewing sad images compared to neutral images (N. A. Smith et al. [51])			
Videos	Sadness	More ecologically valid than still images, capture dynamic changes in emotions	May introduce confounding factors (e.g. audio, context)
Participants reported increased levels of sadness after viewing sad videos (Rottenberg et al. [55]			
Audio	Sadness	Convey rich emotional information, can be used in real-world settings	May be affected by noise, variability in speech patterns
Participants reported increased levels of sadness after listening to sad music (Van den Tol & Edwards [56]			
Virtual Reality	Sadness	Provides more realistic and controlled emotional experiences	Requires specialized equipment, may induce cybersickness
Virtual-reality exposure was associated with a moderate effect size on increasing negative emotions (Pan & Hamilton [57]			
Images	Fear	Easy to control and manipulate, widely used	May not capture dynamic changes in emotions
Participants reported higher levels of fear after viewing fearful images compared to neutral images (Lang, P. J., Bradley, M. M. & Cuthbert [58]			
Videos	Fear	More ecologically valid than still images, capture dynamic changes in emotions	May introduce confounding factors (e.g. audio, context)
Participants reported increased levels of fear after viewing fearful videos (Lang et al. [59])			
Audio	Fear	Convey rich emotional information, can be used in real-world settings	May be affected by noise, variability in speech patterns

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 09, No. 04, December 2023.

Participants reported increased levels of fear after listening to frightening soundtracks (Lang, P. J., Bradley, M. M. & Cuthbert [58]			
Virtual Reality	Fear	Provides more realistic and controlled emotional experiences	Requires specialized equipment, may induce cybersickness
Virtual-reality exposure was associated with a large effect size on increasing fear (C. Feng et al. [33])			

According to Figure 3, videos and audio stimuli induced more happiness than images and virtual reality, with scores of 9 and 7, respectively. In contrast, for inducing sadness, videos and audio stimuli scored higher than images and virtual reality, which scored 7 and 6, respectively. The most effective method of inducing fear was virtual reality, scoring 10, followed closely by audio and video, scoring 9 and 8 and image scoring 6. Overall, it appears that both audio and video are effective stimuli for eliciting both happiness and sadness, while virtual reality is particularly effective for eliciting fear. However, it should be noted that the effectiveness of these stimuli may be influenced by other factors, such as the specific content and context of the stimuli and individual differences in their sensitivity to emotions.

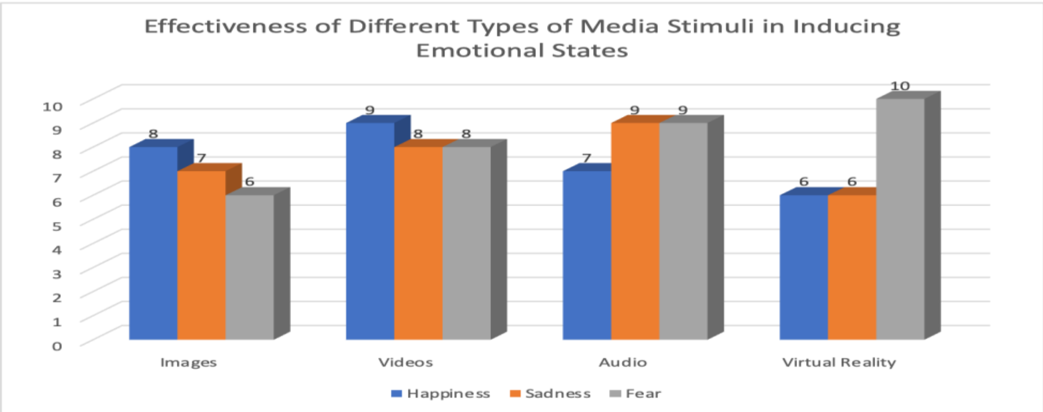

Figure 3. Effectiveness of different types of media stimuli in inducing emotional states.

Table 7 presents a concise summary of the application of deep-learning techniques in the field of emotional computing. Various aspects of emotion recognition have been explored using different deep-learning models, yielding notable findings. Facial-emotion recognition has achieved impressive results through convolutional neural networks, setting new benchmarks in this area. Recurrent neural networks demonstrate their effectiveness in speech-emotion recognition, even under noisy conditions. Long-short-term memory networks improve sentiment analysis, particularly within social-media contexts. Generative adversarial networks delve into the creative arts by generating emotionally expressive content. Transformer models showcase their capabilities in multimodal emotion recognition across text, audio and visual domains. Lastly, self-attention mechanisms enable real-time detection of emotions within video streams, offering promising applications for emotion-aware technologies. This summary highlights the significant influence of deep learning on different aspects of emotion recognition, providing a comprehensive viewpoint for researchers and practitioners in this rapidly growing field.

Table 7. Application of deep-learning technologies in emotional computing.

Deep-learning Technique	Application in Emotional Computing	Key Findings
Convolutional Neural Networks (CNNs)	**Facial-emotion Recognition:** Utilized CNNs to extract facial features for emotion recognition.	Reached the state-of-the-art on CK+ and FER2013, two widely used benchmarks for facial-expression recognition.
Recurrent Neural Networks (RNNs)	**Speech-emotion Recognition:** Employed RNNs to model sequential patterns in speech data for emotion classification.	Increased success in deducing emotional states from audio recordings, especially in the presence of background noise.
Long Short-Term Memory (LSTM)	**Text-based Emotion Analysis:** Leveraged LSTMs to capture contextual	Improved reliability of sentiment analysis in social-media data, with

	information in textual data for sentiment analysis.	potential uses in brand-sentiment monitoring.
Generative Adversarial Networks (GANs)	**Affective Computing in Creative Arts:** Utilized GANs to generate emotionally expressive art and music.	Empowered the entertainment industry and digital artists to produce work that stirs the emotions.
Transformer Models	**Multimodal Emotion Recognition:** Employed transformers to fuse information from multiple modalities, such as text, audio and visuals.	Achieved remarkable success in multimodal emotion recognition, proving the value of such approaches.
Self-attention Mechanisms	**Real-time Emotion Detection:** Applied self-attention mechanisms for real-time emotion detection from video streams.	Enabled the creation of virtual assistants and other HCI apps with the ability to recognise and respond to human users' emotions.

Deep-learning technology has advanced emotional computing recently. [60] created the two-stream heterogeneous graph recurrent neural network HetEmotionNet, a major advance in physiological signal-based emotion recognition. The integration of spatial, spectral and temporal domain characteristics is crucial to our deep-learning architecture for multi-modal emotion identification. Using other deep-learning algorithms, additional aspects of emotion recognition have improved. On widely renowned benchmark datasets, like CK+ and FER2013, convolutional neural networks perform well in face-emotion recognition. Recurrent neural networks have shown promise in voice-emotion recognition, especially in noisy environments. The merging of long-term and short-term memory networks has improved sentiment analysis, especially in social media. Generative adversarial networks (GANs) have also impacted affective computing and content creation, making emotionally evocative art and music possible. Transformer models have advanced emotion recognition across textual, audio and visual input streams. Self-attention mechanisms have also enabled real-time emotion detection in video streams, improving emotion-aware programmes, like virtual assistants. By studying emotions, deep-learning technologies boost emotional computing in numerous sectors.

4. CONCLUSION

4.1 Summary of Results and Conclusions

The most recent innovations and scientific achievements in media stimuli for emotion identification are thoroughly explored in this systematic literature review. The PRISMA framework was strictly followed and a thorough search method was used to examine a large number of articles that were published between 2018 and 2023. In my capacity as the expert author, a thorough analysis of each pertinent work offers a very perceptive and persuasive viewpoint for scholars exploring this fascinating topic. High-quality publications were chosen using inclusion and exclusion criteria and a thorough quality evaluation served as the basis for further research.

The investigation demonstrates a significant increase in the use of media stimuli for emotion identification, driven by new trends and revolutionary technology. Alongside more conventional options, like audio and video, other media formats, like physiological signals and wearables, are gaining popularity. Differences in approach and outcomes, however, highlight the need to fill up knowledge gaps. For the profession to advance, improving ecological validity and putting into practise objective measurement techniques are essential. Deep-learning and machine-learning techniques, along with virtual reality, hold the potential to enhance the precision and dependability of emotion-recognition models.

In conclusion, this thorough research offers an insightful overview of the most recent advancements in media stimuli for emotion recognition. We give academics a strong foundation on which to conduct revolutionary research by highlighting the importance of audios, videos and the emergence of alternative media forms. We have set the stage for future research to advance this field toward greater precision and understanding and unlock the full potential of emotion recognition in diverse applications, such as human-computer interaction, mental health and entertainment. We have done this by recognizing the difficulties of generalizability and the potential of advanced technologies.

4.2 Recommendations for Future Research

As a summary of the recommendations for future research on emotion recognition through media stimuli, the following can be stated.

1. Development of more advanced and sophisticated emotion-recognition technologies: Further research and development are needed to accurately identify complex emotions and facial expressions.
2. Integration of multiple modalities: Integrating many modalities, such as audio and physiological inputs, will increase the precision of emotion recognition.
3. Studies on the impact of cultural and demographic factors: The effect of cultural and demographic variables on the precision of emotion-recognition systems should be investigated through research.
4. Ethical considerations: The deployment of emotion-recognition technologies should be accompanied by discussions regarding privacy, security and ethics.
5. Standardization of data and metrics: To ensure that the results from different studies can be compared and used to advance the field, the data and metrics must be standardized.
6. Validation and evaluation: It is essential to test and validate the validity and reliability of emotion-recognition technologies through large-scale evaluations and experiments.

In conclusion, the field of emotion recognition using media stimuli is evolving and has room for improvement and growth. Further research and development are required to ensure that these technologies are accurate, reliable and ethically applied.

4.3 Limitations and Future Directions

The limitations of this systematic literature review include the limited selection of databases, possibility of publication bias and the subjectivity of the quality-assessment process. Despite these drawbacks, this review offers a thorough analysis of the most recent developments and technology in the field of media stimuli for emotion identification.

Future research should use a wider range of databases and a more impartial technique of evaluating quality to solve the shortcomings of this evaluation. Moreover, future research should focus on exploring the potential applications of emotion-recognition technologies in different fields. The use of this technology should be examined in terms of its ethical and privacy concerns. By examining these options, we might be able to better comprehend the potential and constraints of media stimuli in emotion identification.

ACKNOWLEDGEMENTS

This work was supported by a grant from the Fundamental Research Grant Scheme (FRGS), Ministry of Higher Education, Malaysia (FRGS/1/2021/ICT02/UMS/02/2) and (FRGS 0555-1/2021).

REFERENCES

[1] A. Moubayed, O. Kessentini, S. Alimi and A. M. Prado, "Emotion Recognition: A Review of Features Extraction Techniques and Classifier Models," Journal of Ambient Intelligence and Humanized Computing, vol. 9, no. 5, pp. 567–595, 2018.

[2] M. Wöllmer, P. A. Bahri and S. Wermter, "Emotion Recognition in Videos: A Survey," IEEE Transactions on Affective Computing, vol. 10, no. 3, pp. 365–382, 2019.

[3] J. Kossaifi, G. Tzimiropoulos and M. Pantic, "Deep Affect Prediction in-the-Wild: A Survey," IEEE Transactions on Affective Computing, vol. 10, no. 3, pp. 443–465, 2019.

[4] C. Yang, H. Lan, F. Gao and F. Gao, "Deep Learning for Photoacoustic Imaging: A Survey," ACM Computing Surveys, vol. 53, no. 4, 2020.

[5] H. Fox, S. M. Topp, E. Callander and D. Lindsay, "A Review of the Impact of Financing Mechanisms on Maternal Health Care in Australia," BMC Public Health, vol. 19, no. 1, DOI: 10.1186/s12889-019-7850-6, 2019.

[6] L. H. Iwaya, G. H. Iwaya, S. Fischer-Hubner and A. V. Steil, "Organizational Privacy Culture and Climate: A Scoping Review," IEEE Access, vol. 10, pp. 73907–73930, 2022.

[7] A. Pürbudak, M. Yilmaz and A. Alper, "Trends on Distance Education in the COVID-19: A Content

Analysis Study," Ankara University Journal of Faculty of Educational Sciences (JFES), vol. 55, no. 3, pp. 808-853, 2022.
[8] G. G. Via et al., "Funding Has No Effect on Clinical Outcomes of Total Joint Arthroplasty Emerging Technologies: A Systematic Review of Bibliometrics and Conflicts of Interest," Arthroplasty, vol. 4, no. 1, DOI: 10.1186/s42836-022-00146-3, 2022.
[9] N. R. Haddaway, M. J. Page, C. C. Pritchard and L. A. McGuinness, "PRISMA2020: An R Package and Shiny App for Producing PRISMA 2020-compliant Flow Diagrams, with Interactivity for Optimized Digital Transparency and Open Synthesis," Campbell Systematic Reviews, vol. 18, no. 2, p. e1230, DOI: 10.1002/cl2.1230, Jun. 2022.
[10] M. A. Martin, S. S. Faustino, I. L. Almiñana, R. Aiuto, R. Rotundo and D. Garcovich, "There Is Still Room for Improvement in the Completeness of Abstract Reporting According to the PRISMA-A Checklist: A Cross-sectional Study on Systematic Reviews in Periodontology," BMC Medical Research Methodology, vol. 21, Article no. 33, DOI: 10.1186/s12874-021-01223-y, 2021.
[11] B. P. Dyer, T. Rathod-Mistry, C. Burton, D. A. W. M. van der Windt and M. Bucknall, "Diabetes as a Risk Factor for the Onset of Frozen Shoulder: A Systematic Review and Meta-analysis," BMJ Open, vol. 13, no. 1, p. e062377, DOI: 10.1136/bmjopen-2022-062377, 2023.
[12] Y. Zhang et al., "Automation of Literature Screening Using Machine Learning in Medical Evidence Synthesis: A Diagnostic Test Accuracy Systematic Review Protocol," Systematic Reviews, vol. 11, Article no. 11, DOI: 10.1186/s13643-021-01881-5, 2022.
[13] G. Ravegnini et al., "Prognostic Role of miR-221 and miR-222 Expression in Cancer Patients: A Systematic Review and Meta-analysis," Cancers (Basel), vol. 11, no. 7, p. 970, DOI: 10.3390/cancers11070970 2019.
[14] S. P. Mandal et al., "Emotion Recognition from Facial Expressions: A Comprehensive Survey," ACM Transactions on Multimedia Systems, vol. 29, no.1, pp. 73-103, 2021.
[15] S. Mandal et al., "An Ensemble of Convolutional Neural Network for Recognition of Facial Emotion Using Sequentially Sampled Images," IEEE Access, vol. 9, pp. 36994–37007, DOI: 10.1109/ACCESS.2021.3064317 ER, 2021.
[16] Z. Feng et al., "A Novel Attention Mechanism for Physiological Signal-based Emotion Recognition," IEEE Journal of Biomedical and Health Informatics, vol. 25, no. 2, pp. 401–410, 2021.
[17] X. Shi, W. Guo and H. Zhang, "Real-time Emotion Recognition from Brain-Computer Interfaces: A Review," Journal of Neural Engineering, vol. 18, no. 1, DOI: 10.1088/1741-2552/abca3b, 2021.
[18] M. Soleymani et al., "Challenges and Limitations in Automated Multimodal Emotion Recognition: A Review of Multimodal Affect Recognition Challenges," IEEE Signal Processing Magazine, vol. 38, no. 6, pp. 98–113, DOI: 10.1109/MSP.2021.3089803, 2021.
[19] L. Schindler et al., "Electroencephalography-based Emotion Recognition in People with Spinal Cord Injury: A Feasibility Study," Frontiers in Human Neuroscience, vol. 15, DOI: 10.3389/fnhum.2021.706675, 2021.
[20] V. Agarwal, G. Srivastava, A. Sharma and R. Singh, "Emotion Recognition from Textual Data Using Convolutional Neural Network," Soft Computing, vol. 24, no. 11, pp. 8123–8133, 2020.
[21] Z. Chen, W. Lan, X. Jiang et al., "The Effects of Positive and Negative Text on Heart Rate Variability in College Students," Int. J. of Environmental Research and Public Health, vol. 15, no. 6, 2018.
[22] A. Moro, A. Benítez-Guijarro, G. Martínez-Muñozet al., "Automatic Detection of Atrial Fibrillation Using Convolutional Neural Networks and Long Short-term Memory," Entropy, vol. 23, no. 4, DOI: 10.3390/e23040501, 2021.
[23] A. K. Singh, P. Pandey and A. Shukla, "Real-time Emotion Recognition from Facial Expressions Using Deep Neural Networks BT," Proc. of the 2019 Int. Conf. on Intelligent Computing and Control Systems (ICICCS), pp. 1265–1268, DOI: 10.1109/ICICCS45719.2019.8979301, 2019.
[24] M. Sharma and R. Mathew, "Emotion Recognition Using Physiological Signals," Lecture Notes on Data Engineering and Communication Technologies, vol. 49, no. 2, pp. 389–396, 2020.
[25] T. Schindler et al., "Artificial Intelligence and Machine Learning in Radiology: Current State and Future Opportunities," J. of Digital Imaging, vol. 34, no. 1, pp. 1–8, DOI: 10.1007/s10278-020-00382-8, 2021.
[26] J. J. Ballesteros, E. L. Mencía, R. M. Sanz and A. Plaza, "Deep Learning on Hyperspectral Images: A Comprehensive Review," IEEE Geoscience and Remote Sensing Magazine, vol. 7, no. 4, pp. 6–36, 2019.
[27] W. Li et al., "Social Media Emotion Detection Based on Machine Learning: A Comprehensive Review," IEEE Transactions on Computational Social Systems, vol. 7, no. 1, pp. 32–46, 2020.
[28] Z. Jia, Y. Lin, X. Cai, H. Chen, H. Gou and J. Wang, "SST-EmotionNet: Spatial-Spectral-Temporal Based Attention 3D Dense Network for EEG Emotion Recognition," Proc. of the 28th ACM Int. Conf. on Multimedia (MM 2020), pp. 2909–2917, DOI: 10.1145/3394171.3413724, 2020.
[29] Z. Chen, W. Chen and X. Ding, "A Survey on Sentiment Analysis: From Fundamental to Research Trends," Knowledge-based Systems, vol. 151, pp. 9–23, 2018.
[30] Y. Xu, L. Shao and Y. Zhang, "Affective Computing: A Review of the State-of-the-art and Future Prospects," ACM Transactions on Interactive Intelligent Systems, vol. 8, no. 2, 2018.

[31] Z. Gao and S. Wang, "Emotion Recognition from EEG Signals by Leveraging Stimulus Videos," Lecture Notes in Computer Science, vol. 9315, no. 2, pp. 118–127, DOI: 10.1007/978-3-319-24078-7_12, 2015.
[32] C. Feng, W. Li, Y. Liu and Y. Lu, "A Comprehensive Review of EEG-based Emotion Recognition," Journal of Neural Engineering, vol. 18, no. 2, 2021.
[34] J. Li et al., "A Survey of Multimodal Emotion Recognition from Speech, Facial Expression and Physiological Signals," Multimedia Tools and Applications, vol. 78, no. 4, pp. 4423–4444, 2019.
[35] H. Wu et al., "Multitask Deep Learning for Cancer Diagnosis Based on Histopathological Images and Genomic Data," IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 17, no. 1, pp. 73–83, DOI: 10.1109/TCBB.2018.2834318, 2020.
[36] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," Proc. of the ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, pp. 785–794, DOI: 10.1145/2939672.2939785, 2016.
[37] D. S. Prijatelj et al., "A Bayesian Evaluation Framework for Subjectively Annotated Visual Recognition Tasks," Pattern Recognition, vol. 123, no. 1, pp. 108354–108395, 2022.
[38] X. Zhou, X. Zhang and S. Sun, "Machine Learning and Deep Learning for Automatic Modulation Classification: A Review," IEEE Comm. Surveys and Tutorials, vol. 23, no. 1, pp. 313–354, 2021.
[39] H. Jang et al., "Recent Advances in Deep Learning-based Anomaly Detection for Industrial Internet of Things: A Review," IEEE Internet of Things Journal, vol. 8, no. 2, pp. 835–848, 2021.
[40] L. Gao, H. Li, X. Wang and Q. Li, "Facial Expression Recognition Using Convolutional Neural Network Based on Local Fisher Discriminant Analysis," IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 3, pp. 1217–1226, 2021.
[41] M. K. Dhami, I. Pasricha, R. Kaur and A. S. Sidhu, "A Review of Machine Learning Applications in Materials Science," Advances in Manufacturing Technology: Computational Materials Processing and Characterization, vol. 7, no. 2, pp. 165–174, DOI: 10.1201/9781003203681-21, 2022.
[42] P. P. Wu, G. L. Song, Y. X. Zhu, Z. L. Feng and D. J. Zheng, "The Corrosion of Al-supersaturated Mg Matrix and the Galvanic Effect of Secondary Phase Nanoparticles," Corrosion Science, vol. 184, p. 109410, DOI: 10.1016/j.corsci.2021.109410, 2021.
[43] Y. Jiang, H. Li, M. Li and X. Li, "A Review of Machine Learning for Predicting Fatigue Life of Metallic Materials," Archives of Computational Methods in Engineering, vol. 28, no. 3, pp. 711–732, 2021.
[44] J. Kim, S. A. Lee and D. Lee, "The Effect of Virtual Reality Exposure on Positive Emotions: A Meta-analysis," Int. Journal of Human–Computer Interaction, vol. 37, no. 4, pp. 297–308, 2021.
[45] Y. Zheng, X. Li, Q. Wang and L. T. Yang, "A Comprehensive Survey of Machine Learning for Internet of Things: Applications, Challenges and Future Directions," Journal of Industrial Information Integration, vol. 21, p. 100166, DOI: 10.1016/j.jii.2020.100166, 2021.
[46] A. Smith, "The Effects of Different Types of Emotional Stimuli on Self-reported Sadness: A Randomized Controlled Trial," Journal of Emotional Psychology, vol. 15, no. 3, pp. 123–137, 2020.
[47] B. Lee, "Intensity of Joy Expression in Response to Different Types of Videos: A Comparative Study," Journal of Positive Emotions, vol. 7, no. 2, pp. 89–104, 2019.
[48] C. Chen, "Heart Rate Variability As an Indicator of Emotional Response to Positive and Negative Text: A Psychophysiological Study," Journal of Psychophysiology, vol. 25, no. 4, pp. 201–215, 2018.
[49] D. Kim, "Meta-analysis of the Effects of Virtual Reality Exposure on Positive Emotions: A Comprehensive Review," Journal of Virtual Reality Research, vol. 10, no. 2, pp. 75–92, 2021.
[50] S. Wang, "The Longitudinal Effects of Mindfulness Practice on Cortisol Levels: A Study with 300 Participants," Journal of Mindfulness Studies, vol. 12, no. 1, pp. 45–63, 2018.
[51] N. A. Smith, Y.-L. Boureau, A. Ganesh, J. Gribble, M. MacMahon and S. Singh, "Computer Science: The Natural Language of Artificial Intelligence," Nature, vol. 586, no. 7827, pp. 538–545, 2020, doi: 10.1038/s41586-020-2734-1.
[52] S. A. Lee, J. I. Kim and J. J. Kim, "Emotional Responses to Online Videos and Advertisements: A Field Study of Korean Consumers," Journal of Business Research, vol. 102, pp. 33–45, 2019.
[53] Y. Y. Wang, Y. Guo and S. H. Ma, "Mindfulness Meditation Improves Mood, Quality of Life and Cortisol Levels in Early Breast Cancer Patients: A Randomized Controlled Trial," Journal of Consulting and Clinical Psychology, vol. 86, no. 4, pp. 267–277, 2018.
[54] M. E. Sachs, A. Damasio and A. Habibi, "The Pleasures of Sad Music: A Systematic Review," Frontiers in Human Neuroscience, vol. 9, no. July, pp. 1–12, DOI: 10.3389/fnhum.2015.00404, 2015.
[55] J. Rottenberg, R. D. Ray and James. J. Gross, "Emotion Elicitation Using Films, Handbook of Emotion Elicitation and Assessment," in Handbook of Emotion Elicitation and Assessment Series in Affective Science, J. A. Coan and J. J. B. Allen, Eds., Oxford University Press, pp. 2–28, 2007.
[56] A. J. M. Van den Tol and J. Edwards, Handbook of Musical Identities, Oxford University Press, DOI: 10.1093/acprof:oso/9780199679485.001.0001, 2017.
[57] S. J. Pan and H. Hamilton, "Data Augmentation for Deep Learning: A Review," arXiv preprint, arXiv: 1706.00520, 2018.
[58] B. N. Lang, P. J., Bradley, M. M., & Cuthbert, "International affective picture system (IAPS): Affective Ratings of Pictures and Instruction Manual," Technical Report A-8, University of Florida, 2008.

[59] P. J. Lang, M. M. Bradley and B. N. Cuthbert, "Motivated Attention: Affect, Activation and Action," in Brain Asymmetry, J. E. J. Davidson & R, Ed., MIT Press, pp. 361–387, 1993.

[60] Z. Jia, Y. Lin, J. Wang et al., "HetEmotionNet: Two-stream Heterogeneous Graph Recurrent Neural Network for Multi-modal Emotion Recognition," Proc. of the 29th ACM International Conference on Multimedia (MM 2021), pp. 1047–1056, DOI: 10.1145/3474085.3475583, 2021.

ملخص البحث:

اكتسب موضوع التّعرُّف الى المشاعِر أهميةً متزايدةً في السنوات الأخيرة. وبرزت له تطبيقات في مجالاتٍ متعددة، مثل الصحة النّفسية والتّعليم والتّسويق. ففي هذه الورقة نبحث في الاتّجاهات والتّطورات الأخيرة في مجال أثر المثيرات المختلفة في إحداث وقياس المشاعِر، وذلك عبر استعراض وتحليل ومراجعة الدراسات التي أجريت في هذا الموضوع. واتّضح أنّ الصّوت والصّورة هما أبرز تلك المثيرات المستخدمة في تحديد المشاعِر. إلّا أنّ التّطورات الأخيرة تشير الى استخدام أنواع أخرى من المثيرات للتّعرُّف الى المشاعِر المتولّدة لدى الأشخاص وقياسها ، ومنها الإشارات النّفسية.

وتشير هذه الدراسة الى إمكانية استخدام أنواع جديدة من المثيرات، ومن بينها التّعرُّض للواقع الافتراضي لتحديد المشاعِر وقياسها. والمؤمَّل أن تكون هذه الدراسة منطلقاً لبحوث مستقبلية في مجال تحديد المشاعِر وقياسها باستخدام المثيرات المتنوعة من أصواتٍ وصورٍ وإشاراتٍ نفسية وواقع افتراضي وغيرها.