IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 5 June 2024, accepted 27 June 2024, date of publication 3 July 2024, date of current version 22 July 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3422480

SURVEY

A Comprehensive Survey on Affective Computing: Challenges, Trends, Applications, and Future Directions

SITARA AFZAL[1], HASEEB ALI KHAN[1], MD JALIL PIRAN[2], (Senior Member, IEEE), AND JONG WEON LEE[1]

[1]Department of Software, Sejong University, Seoul 05006, South Korea
[2]Department of Computer Engineering, Sejong University, Seoul 05006, South Korea

Corresponding authors: Md Jalil Piran (piran@sejong.ac.kr) and Jong Weon Lee (jwlee@sejong.ac.kr)

This work was supported in part by the Ministry of Science and ICT (MSIT), South Korea, under the Information Technology Research Center (ITRC) Support Program supervised by the Institute for Information & Communications Technology Planning & Evaluation (IITP) under Grant IITP-2023-RS-2022-00156354; and in part by the Faculty Research Fund of Sejong University, in 2024.

ABSTRACT Affective computing, as its name implies, focuses on the recognition of human emotions, sentiments, and feelings. This interdisciplinary field encompasses diverse areas such as languages, sociology, psychology, computer science, and physiology. However, there is a notable absence of research exploring the interaction between machine learning (ML) and mixed reality (XR) for affective computing. This paper aims to address this gap by discussing the importance of affective computing and delving into its concepts, methods, and outcomes. Drawing upon ML and XR approaches, we conduct a comprehensive survey of recent methodologies employed in affective computing. Additionally, we examine state-of-the-art approaches and available affective data resources. Furthermore, we explore various applications where affective computing has a significant impact, providing valuable insights for future scholars seeking to deepen their understanding of its relevance and practical implications.

INDEX TERMS Affective computing, emotional analysis, machine learning, mixed reality.

I. INTRODUCTION

Emotions, sentiments, and feelings, along with emotion recognition, constitute what is referred to as "affective computing" [1]. In 1997, Prof. Picard introduced the idea of affective computing, which has helped computers recognize and communicate their moods, as well as effectively respond to humans' moods [2], [3]. In many practical applications, it is desirable to develop a cognitive, intelligent system that can detect and comprehend people's feelings while also providing sensitive and cordial responses [4]. Emotion is a cultural and psychobiological adaptation mechanism that enables people to adapt dynamically to environmental changes. Emotions give meaning to our lives, deepen our relationships with others, alert us to our wants and sentiments, and help us make changes [5].

The associate editor coordinating the review of this manuscript and approving it for publication was Wei-Wen Hu.

There are five components to a single sentiment: a cognitive evaluation, a physical sensation, an intention, a subjective feeling, a motor reaction, and, in most situations, an interpersonal component. "Emotion Regulation" (ER) is a group of mental processes that affect our emotions, as well as the timing of our emotions [6]. A dynamic process intended to down-regulate or up-regulate positive or negative thoughts, it is essential to human mental function. In order to identify expressed emotions (anger, disgust, fear, happiness, sorrow, or surprise) and predict their sentiment (positive or negative), a framework based on the frame of reference of intermodal attention [7] was developed.

It is called emotion dysregulation when a person is unable to effectively control or process their emotions, resulting in unintended intensification or deactivation [3]. Managing emotionally intense experiences is therefore a component of self-regulation. A specific method of emotion management is not necessarily beneficial or harmful [8]. In doing so, it avoids making generalized judgments about which coping mechanisms are more or less adaptive depending on the situation [9], [10]. In various social media platforms, affective computing can greatly help in understanding the thoughts expressed [7]. In this way, emotional computing is seen by many researchers as a way to advance the creation of human-centric artificial intelligence (AI) and human intelligence [11]. There is a difference between sentiment research and emotion recognition in affective computing [12]. A diagram showing the different kinds of emotions can be found in figure 1.

FIGURE 1. Types of emotions.

To understand and calculate human feelings, psychologists presented two common theories: the discrete emotion model (or categorical emotion model) [13] and the dimensional emotion model [14]. In order to identify people's emotional state, three main areas of focus for emotion recognition have been investigated [15]: visual emotion recognition (VER), audio/speech emotion recognition (AER/SER) and physiological emotion recognition (PER). It has been shown that virtual reality (VR) systems can elicit emotional responses that lead to psychologically positive changes when experiencing increased presence in a virtual environment [16]. These widely used datasets have inspired the development of emotional computing, which is Machine Learning (ML). Study [17] reports that in everyday human interaction, human emotions are primarily expressed through facial expressions (55%) voice (38%) and words (7%) Textual, auditory, and visual signals are generally referred to as physical data.

People's propensity to openly express their ideas and thoughts on social media platforms and websites makes it easy to gather a wide range of bodily affect information. As a result of these data, many academics attempt to identify minor emotions conveyed either explicitly or implicitly. As a result, [18] and [19] present an innovative approach to multimodal music emotion analysis in affective computing using LSTM networks. With the proposed model, a dual-channel LSTM replicates the human auditory and visual processing pathways, enabling the emotional information of music and facial expressions to be effectively processed. As a result, tangible affect recognition may not be successful, since people may intentionally or unintentionally cover up their feelings (called social masking) [20].

A. MOTIVATION

Increasingly, scientists in the fields of ergonomics and intelligent systems are working to improve the effectiveness and adaptability of human-computer interaction (HCI) in various situations. Computers' ability to comprehend human emotions and behavior is a crucial component of their flexibility. The performance of the state-of-the-art methods and their implications for recognition have not been fully addressed in previous surveys. The most important contribution of our review is that it addresses all areas of emotional computers through a variety of research methods and findings, as well as discussions and future efforts. Currently, most HCI technologies cannot recognize human emotion, but the development of advanced HCI systems depends on automatic computer recognition of human emotion.

B. CONTRIBUTIONS

The performance of state-of-the-art methods and their implications for their recognition ability have not been fully addressed in previous reviews. Our review offers a variety of research methods and findings, as well as discussions and future efforts, to address all aspects of emotional computers. The major contributions of this paper are as follows.

- To the best of our knowledge, this is the only survey of its kind to use ML to categorize emotional processing into broad categories of emotion identification.
- By examining how well the various affective modality is utilized to study and identify affect, we provide a comprehensive classification of the state-of-the-art emotional computational tools.
- To the best of our knowledge, this is the only survey that classifies emotional processing using mixed reality (MR) machines and even further taxonomies them based on augmented reality (AR) and VR.
- In order to categorize benchmark databases for emotional computing, audio, speech, text, and visual modalities, are used. Those resources' salient features and accessibility are outlined.
- Current approaches and tools are discussed along with their limitations.
- Finally, we outline several open research problems in the field of AR affective computing and ML.

The remainder of the paper is organized as follows. The scope of the paper is presented in Section II. The existing survey papers are discussed in Section III. In Section IV, a number of important techniques are compared. The dataset is covered in detail in Section V. In Section VI, the main challenges of the paper are summarized, and several

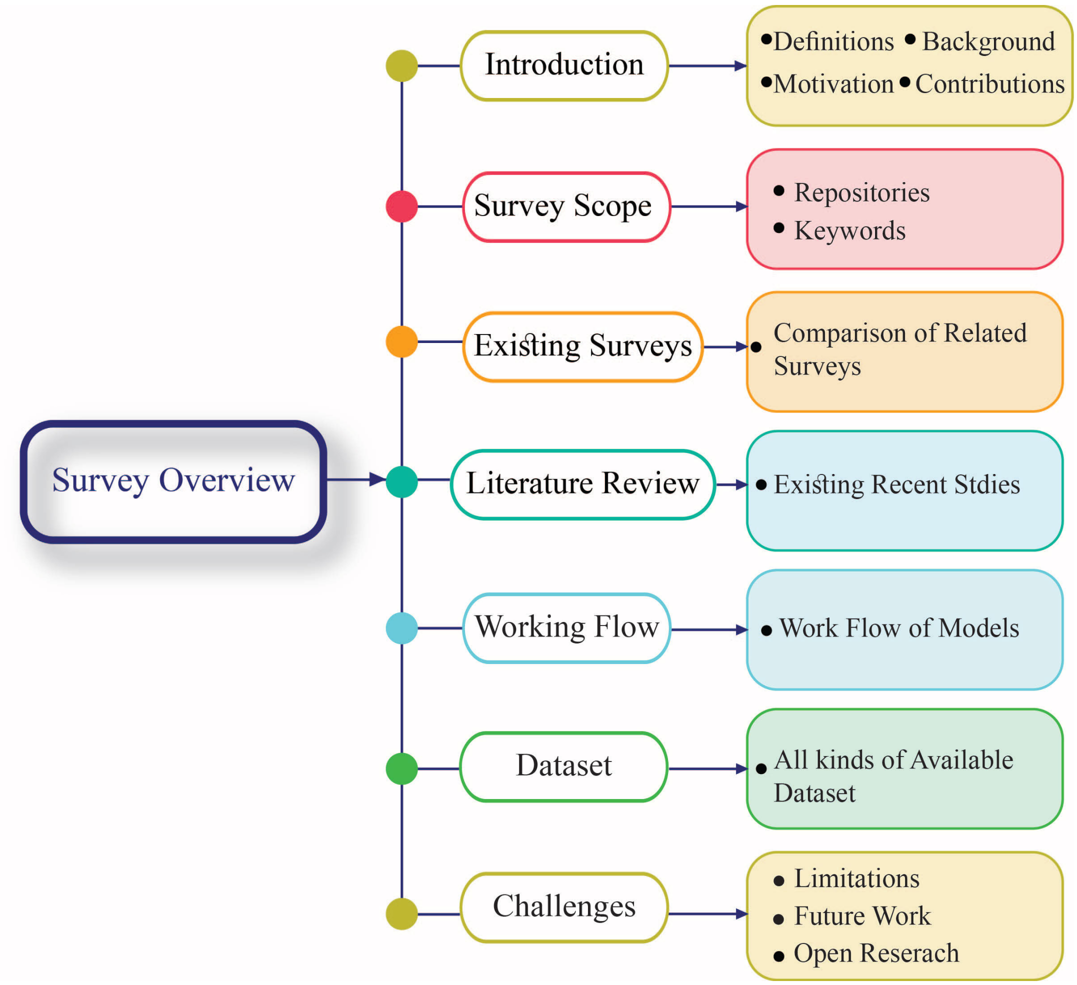

FIGURE 2. Organization of this survey paper.

TABLE 1. List of acronyms.

Acronym	Definition
VER	Visual Emotion Recognition
SER	Speech emotion recognition
AER	Audio emotion recognition
AI	Artificial Intelligence
VR	Virtual Reality
MR	Mixed Reality
AR	Augmented Reality
XR	Extended Reality
ML	Machine learning
DL	Deep learning
SVM	Support vector machine
KNN	K-Nearest Neighbor
RF	Random Forest
ANN	Artificial Neural network
DT	Decision Tree
ER	Emotion Regulation
HCI	Human Computer Interface
FER	Facial Emotion Recognition

research directions are identified. Finally, Section V draws the conclusions. An overview framework is shown in Figure 2 to improve the general readability of the paper.

II. SCOPE OF THIS SURVEY

We will discuss the approach used in this literature review in this section. Affective computing as it pertains to ML, deep learning (DL), and MR is the focus of this survey, to the best of our knowledge. As far as we know, no other study has taken into account affective computing in terms of both ML and MR. The relevant articles between 2014 and 2021 are considered. There are three main models in this survey: identification, screening, and inclusion.

As the first step of this survey, 'Identification' is used to classify the data from all the databases. To conduct this survey, we consider some research questions and keywords. VR, AR, affective computing, ML, supervised learning, and emotion recognition are explored and acquired from various publishers including the Institute of Electrical and Electronics Engineers (IEEE), Springer, Hindawi, and the Association for Computing Machinery (ACM).

We consider only the relevant articles from all the selected articles during this step. The duplicate record is removed using EndNote X9 for relevant articles that are not redundant. To determine whether the study articles' titles and abstracts were relevant to a developed research topic, the titles and abstracts were manually reviewed. We evaluated research papers that contributed a significant amount but were highly relevant.

Data selected following the qualifying stage are subjected to quantitative and qualitative evaluations at this stage of the research process. Adding records is the first step in the quality check analysis. During the last stage, known as qualitative analysis, some of these records are analyzed using meta-analysis. As a final step, data is extracted from such data that were included in the survey's final analysis.

III. RELATED WORK

The key findings of this study are summarized in Table 2 in comparison with earlier research in the area of affective computing. In no previous study for emotion recognition, ML-based and MR approaches have been examined concurrently. Prior research also ignored the numerous disciplines that contribute to the field of affective computing, as well as their ideas, models, and methodologies. This study discusses the classification and recognition of emotions in terms of their categories and methods for detecting them. We have also examined the challenges, remedies, and possible trends of research studies.

Researchers provide a systematic review of emotion models, databases, and recent advances in [19], which includes both physical and psychological studies. In addition, they examined the topologies and capabilities of cutting-edge multimodal emotional analysis and unimodal affect identification. A mixed reality or ML approach to emotion recognition was not considered by the authors of [21]. This paper [22] surveys recent efforts in automated analysis of human affective behavior, addressing the challenge of processing naturally occurring expressions beyond deliberate displays. It examines psychological perspectives on emotion perception and explores current approaches in machine understanding of real-world affective behavior, highlighting scientific and engineering challenges in advancing affect sensing technology.Affective computing is only considered in the context of interdisciplinary fields like psychology and computer science, along with their theories, concepts, models, and implications. Additionally, they present some existing affective databases.

In [23], authors surveyed the state-of the art studies affective computing. Their main focus was on the types of features in emotions datasets as they reviewed affective image contents. In affective computing, they discuss approaches to identifying emotions using features extraction. The authors of [24] described the state-of-the-art affective computing technologies for large-scale heterogeneous multimedia data in a comprehensive manner. Using handcrafted features-based features, they compare relevant techniques on AC of many multimedia forms, including photos, music, movies, and multimodal data.

Additionally, this study examines the different uses, problems, and upcoming challenges of emotion detection. In order to achieve the study's objective, the following research questions were developed:

- (RQ1) What different interconnected domains are a part of emotion detection?
- Answer (1): Emotion detection integrates psychology, computer science, linguistics, neuroscience, and related fields, analyzing behavior, facial expressions, vocal cues, and physiological signals.
- (RQ2) What are the most common fields wherein affective computing is used?
- Answer (2): Affective computing finds applications in healthcare, education, human-computer interaction, virtual reality, gaming, marketing, social media analysis, and robotics.
- (RQ3) Which recent research questions might have an impact on future work in emotion detection?
- Answer (3): Recent research focuses on multimodal data integration, personalized emotion detection, ethical concerns, and real-time algorithms, shaping future emotion detection endeavors.
- (RQ4) How ML and Mixed Reality recognize the emotions.
- Answer (4): Machine learning and Mixed Reality utilize deep learning, NLP, computer vision, and sensors to recognize emotions, enabling effective training on labeled datasets and integration with virtual and real-world environments.

A thorough search on the Microsoft academic research portal was conducted to obtain a comprehensive understanding of the published literature on affective computing.

IV. EMOTION MODELS

The definition of the emotion or affect is crucial to building an emotional computing standard. Ekman articulated the fundamental idea of emotions for the first time in the 1970s. Despite psychologists' efforts to categorize emotions in neuroscience, philosophy, and computer science, there is no universally accepted emotion model. Generic emotion models used in affective computing include continuous emotion models and multidimensional emotion models (sometimes called continuous emotion models).

V. WORK FLOW OF EMOTION RECOGNITION

In this section, we discuss the workflow and contributions of all forms of affective computing, including ML, deep learning, and virtual reality.

A. TEXTUAL BASED EMOTION RECOGNITION

The majority of methods for textual emotion recognition, or TSA, are based on statistical or knowledge-based approaches. In the former case, an extensive emotional vocabulary is required to model a thesaurus, while in the latter case, a large database with emotional labels is required. Online social media and e-commerce systems are rapidly expanding, allowing users to freely express their opinions, which generates a significant amount of text data. A text

TABLE 2. Summary of existing related survey articles.

Research	Year	AR/VR	ML	All four categories A/V/T/S	DB	Wearables	Accuracy	Applications
[19]	2022	✗	✓	✓	✓	✗	✓	✓
[21]	2021	✗	✗	✓	✗	✗	✗	✓
[23]	2018	✗	✓	✗	✓	✗	✗	✓
[24]	2019	✗	✓	✓	✓	✗	✗	✗
[25]	2018	✗	✓	✗	✓	✗	✓	✗
[26]	2018	✗	✓	✗	✓	✗	✓	✓
[27]	2018	✗	✓	✗	✓	✗	✓	✓
[28]	2020	✗	✓	✗	✗	✗	✗	
[29]	2019	✗	✓	✗	✓	✗	✓	✗
[30]	2022	✓	✗	✗	✗	✓	✗	✓
Our Survey	2023	✓	✓	✓	✓	✓	✓	✓

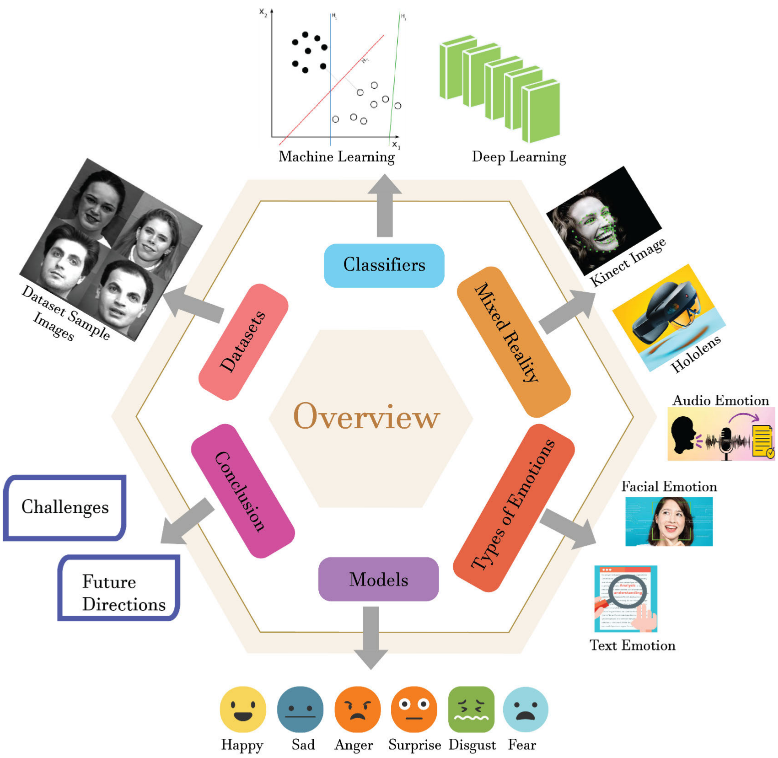

FIGURE 3. The overall framework of this survey.

sentiment classification method was developed to detect subtle sentiments or emotions expressed directly or implicitly from user-generated material [31].

The term ''feature engineering'' is often used to identify characteristics that are associated with emotion by using traditional approaches [32]. Through the use of DL-based models, it is possible to create an end-to-end sentiment analysis of text data. Statistics and knowledge-based methods are mainly used in conventional ML-based TSAs [34]. It is common for semantically polarized terms to be mixed up in various lexicons, including Word-Net, WordNet Affect, Sentic-Net, and Senti-WordNet [33].

Table 3 presents the articles and their contribution to text-based emotion detection using deep learning and ML techniques. Table 3 provides an overview of the main contributions along with datasets and evaluation parameters. Emotion classification in text has garnered significant attention in recent years, with researchers focusing on various aspects to enhance accuracy and effectiveness. [34] conducted a comprehensive comparative analysis of existing annotated corpora for emotion classification in text. By evaluating datasets such as Affective Text, Blogs, ISEAR, SSEC, and TEC, they identified the strengths, weaknesses, and practical implications for improving emotion detection models. Their findings underscored the importance of dataset selection and highlighted the significance of accuracy as a performance measure in this domain.

Reference [35] made significant strides in multilingual emotion analysis by creating a dataset of over 22,000 annotated tweets. This dataset enabled fine-grained emotion intensity analysis and facilitated the evaluation of bias in emotion classification systems. By leveraging tweets as their primary data source, they provided valuable insights into cross-cultural emotion expression and emphasized the importance of accuracy in evaluating model performance.

In another study [36], researchers focused on sentiment analysis of social media text, developing and evaluating a semantic analysis task to create a comprehensive polarity corpus. By utilizing tweets and LiveJournal messages, they provided researchers with a standardized benchmark for comparing different sentiment analysis approaches. Their work contributed to the advancement of sentiment analysis in social media data. In another study [37], authors extended the scope of emotion analysis to art with the creation of the WikiArt Emotions Dataset. This dataset facilitated exploration of emotional responses in art, providing researchers with a unique perspective on the intersection of emotion and creativity.

Similarly [38] proposed a novel approach to sentiment analysis by leveraging distant supervision with a diverse set of noisy labels from 1246 million tweets. Their method led to richer representations and state-of-the-art performance on multiple sentiment, emotion, and sarcasm detection benchmarks, showcasing the effectiveness of their approach on large-scale social media data. In another study [39], author addressed the limitations of existing word representation learning methods in capturing emotional similarity by proposing a novel method for obtaining emotion-enriched word representations. Their work opened new avenues for incorporating emotion into natural language processing tasks, contributing to a deeper understanding of emotional semantics in text data.

Reference [40] introduced MEDA, a Multi-label Emotion Detection Architecture, designed to capture and correlate multiple emotions expressed in text. By using datasets such as RenCECps and NLPCC2018, they demonstrated the effectiveness of their approach in multi-label emotion classification tasks. Similarly [41], presented "SpanEmo," a novel model for multi-label emotion recognition that employed span-prediction and a new loss function to improve performance. By utilizing the SemEval2018 dataset, they showcased the potential of their approach in capturing associations between emotion classes and words in a sentence, contributing to the advancement of emotion recognition in text data.

B. AUDIO EMOTION RECOGNITION

A process of processing and understanding speech signals is used in audio emotion recognition (also known as SER) [59]. There are a number of ML- and DL-based SER systems that have been implemented to enhance analysis [60], [61]. In ML-based SER, the extraction of acoustic features and the selection of classifiers are the primary focuses. A DL-based SER constructs a CNN architecture without taking feature engineering and selection into account, however, in order to forecast the final feeling [62].

In ML-based SER systems, two crucial processes take place: learning representations of emotional speech and selecting an appropriate classification for the final emotion prediction. A well-known audio feature extraction tool called OpenSMILE [63] extracts all the essential components of speech. For SER systems, HMMs, GMMs, SVMs, RFs, and ANNs are frequently employed as classifiers. The SER also uses maximum classification and enhanced traditional easy-to-understand classifiers in addition to these learners. The following Table 4 shows an overview of the main contributions, along with the dataset and evaluation parameters.

C. VISUAL EMOTION RECOGNITION

Visual-based emotion recognition, a prominent area in affective computing, focuses on deciphering emotional states from facial expressions and other visual cues. This field leverages advancements in computer vision and machine learning to develop algorithms capable of accurately identifying and classifying emotions based on visual stimuli. Researchers have explored various methodologies and techniques to enhance the accuracy and efficiency of emotion recognition systems, leading to significant progress in understanding human emotions through visual data.Images or films with face emotional clues are used to execute FER [74].

In face expression identification research, several methodologies have been explored to enhance accuracy and efficiency. One approach, as documented in [71], employs subsequent frames based on elastic bunch graph, a technique that dynamically analyzes facial movements over time. By capturing the subtle changes in facial expressions across frames, this approach can effectively identify and classify various emotional states. It utilizes advanced algorithms to track and analyze facial features, enabling accurate recognition even in complex scenarios. The utilization of datasets like CK+ allows for extensive validation and fine-tuning of the algorithm, resulting in robust performance.

In another study [72], fuzzification was explained, It is a process that introduces uncertainty into the classification process, allowing for more flexible and nuanced identification of facial expressions. Combined with tracking Active Appearance Model (AAM) characteristics, this approach enables the model to adapt to variations in facial expressions

TABLE 3. An overview of text based emotion recognitions studies.

Research	Contributions	Dataset	Evaluation
[42]	Lexicon-based strategy, external evidences system, Context dependent approach	Review Dataset	Accuracy
[43]	Baseline lexicons, Train word-class correlations,	Weblogs	Accuracy
[44]	Concept-level sentiment analysis, logic computing, ML	VER	Visual Emotion Recognition
[45]	Multidisciplinary sentiment classification, Twitter ad hoc conversations of significant conflicts	Twitter	Accuracy
[46]	Asses sentimenatl analysis, context for lexicons methodologies	Twitter	Accuracy
[47]	Sentiment classification using SVMs, combination of phrases and adjective for favorability measures .	Music review	Accuracy
[48]	The opinion-level framework, intra-opinion features and inter-opinion, Bayesian model	Twitter	Accuracy
[49]	Strengthen the impartial, boundary between positive and negative reviews	Amazon, cinema, movies,	Accuracy
[50]	Hybrid strategy, combination of deep learning and a vocabulary method	Amazon, IMBD and Yelp	Accuracy
[51]	Identification of Internet slang and emoticons, k-Nearest Neighbors, Decision Tree, Random Forest, Logistic Regression, Naive Bayes, and Support Vector Machine	Weibo	Accuracy
[52]	New word embedding technique, Dynamically adjust vector representation	IMDB Yelp 13, 14 and 15	Accuracy
[53]	Novel text processing design (VDCNN), character level, tiny convolutions and pooling procedures	Amazon	Accuracy
[54]	Unique parameterized convolutional neural network, Categorization of sentiments, Incorporate aspect data into CNN	SemEval 2014	Accuracy
[55]	LM employing contextual BLSTM (cBLSTM), Modified version of bidirectional LSTM (BLSTM)	IMDB	Accuracy
[56]	Joint framework for explicit aspect, Opinion terms co-extraction unifies recursive neural networks	Laptop	Accuracy
[57]	Examination of two distinct types of tweets obtained during the COVID19 pandemic	Twitter	Accuracy
[34]	The paper contributes by providing a comprehensive analysis of emotion classification in text	Affective Text , Blogs, ISEAR, SSEC,TEC	Accuracy
[35]	This paper contributes by creating and providing a multilingual dataset	Tweets	Accuracy
[36]	The paper contributes by developing and evaluating a semantic analysis task	Tweets	Accuracy
[37]	The paper enable exploration of emotional responses in art.	WikiArt Emotion	Accuracy
[58]	The paper introduces GoEmotions,facilitating advancement in emotion understanding applications.	GoEmotion	Accuracy
[38]	The paper demonstrates the efficacy of extending distant supervision in sentiment analysis	Tweets	Accuracy
[39]	The paper proposes a novel method for obtaining emotion-enriched word representations	Own Data	Accuracy
[40]	The paper introduces MEDA, adept at capturing emotions expressed in text	RenCECps, NLPCC2018	Accuracy
[41]	The paper presents "SpanEmo," a novel model for multi-label emotion recognition, employing span-prediction	SemEval2018	Accuracy

and environmental conditions. By dynamically adjusting the classification criteria based on the observed characteristics, the model can achieve higher accuracy rates, particularly in challenging scenarios where facial expressions may be ambiguous or obscured.

In another study [75], researchers explores the challenges in developing automated systems for detecting and interpreting facial expressions, aiming for human-like interaction between humans and machines. It surveys past research on face detection, expression extraction, and classification, while considering the capabilities of the human visual system, providing insights for the development of automatic facial expression analyzers.

Linear discriminant analysis (LDA) is a statistical technique used to identify patterns and relationships within datasets. In the context of emotion recognition, LDA is applied to extract discriminative features from facial expressions, allowing for effective differentiation between different emotional states [73]. By analyzing datasets like JAFFE, MMI, and SFEW, this approach can uncover subtle variations in facial expressions and map them to specific emotional categories, leading to more accurate emotion recognition.

TABLE 4. An summary of audio based emotion recognitions studies.

Research	Main Contributions	Dataset	Evaluation
[64]	Examination of two types of tweets obtained during the COVID-19	Speaker language	Accuracy
[65]	Utilization of prosodic features, Utilization of voice quality parameters, Investigation of the interactions between prosodic and voice quality, Utilization of the bayesian classifier to recognize emotions	Speaker	Accuracy
[66]	Fuse the features, multiple kernels learning-based approach	Berlin database	Accuracy
[67]	Deep Neural Network based identification of emotions	Speaker	Accuracy
[68]	Validate the use of a corpus of semantically impartial text, Transcription of both neutral and emotional (acted) speech	TTS	Accuracy
[69]	Automated speech emotion detection algorithm, Computer model of the human auditory system	Speaker	Accuracy
[70]	Classification on three corpora, namely the RAVDESS, IITKGP-SEHSC, and the Berlin EmoDB using SVM as classifier	Speaker	Accuracy
[71]	Merging of personalized and non-personalized features for speech emotion, Utilze fuzzy C-means clustering algorithm, Employe several trees to identify various emotional states	CASIA	Accuracy
[72]	Extraction of keypoints from spectrogram, Utilizion of CNN, Predictions of seven emotions	Speaker	Accuracy
[73]	Performs auto encoders, Utilzioon of Recurrent neural networks, Classify four fundamental moods, Recognize emotions from spectrogram features	Speaker	Accuracy

Auto-encoders, as explored in [74] and [76], are a type of artificial neural network used for unsupervised learning of efficient codings of input data. In emotion recognition, auto-encoders are employed to extract relevant features from facial images, which are then classified using Self-Organizing Maps (SOMs). By combining geometric characteristics with Local Binary Patterns (LBP) features, this approach captures both spatial and textural information in facial expressions, resulting in robust emotion classification performance. Datasets like MMI and CASME II provide valuable training and validation data for optimizing the model's parameters.

This research [77] focuses on detecting subtle facial deformations associated with different emotional states, acknowledging that emotional expressions are not always overtly pronounced. By analyzing datasets like BU-4DFE, which contain a diverse range of facial expressions, this approach aims to uncover hidden patterns and cues that signify specific emotional states. Techniques such as feature extraction and pattern recognition are used to identify and classify these subtle deformations, leading to improved accuracy in emotion classification tasks.

Micro-expressions, as studied in [78] and [79], are brief, involuntary facial expressions that occur within a fraction of a second, often revealing true emotions that individuals may try to conceal. Recognizing and interpreting these micro-expressions is crucial for understanding genuine emotional responses. This research leverages innovative techniques such as weighting schemes and spatiotemporal descriptors to detect and analyze micro-expressions accurately. By utilizing datasets like HOG3D, which contain high-speed video recordings of facial expressions, this approach can capture the subtle dynamics of facial movements and infer underlying emotional states with high precision.

Neutrality facial images serve as important reference points for emotion recognition algorithms, providing a baseline for comparison against other emotional expressions. By creating standardized neutrality images, researchers can ensure consistency and accuracy in emotion recognition tasks [80], [81]. Additionally, the development of identification systems incorporating novel techniques such as hard negative generation (HNG) networks and radial metric learning (RML) networks enhances the robustness and reliability of emotion recognition systems. These systems are trained on datasets like CK+ and MMI, enabling them to accurately classify a wide range of emotional expressions.

Attention-based modules, as investigated in [82], [83], and [84], allow emotion recognition systems to focus on the most salient regions of the face, enhancing their ability to extract relevant features and classify emotional expressions accurately. Uncertainty analysis techniques help quantify the confidence level of emotion predictions, allowing for more informed decision-making in ambiguous scenarios. Dynamic-temporal stream modules capture the temporal dynamics of facial expressions, providing valuable context information for emotion recognition. By integrating these modules into emotion recognition systems and evaluating their performance on datasets like FER, SemEval 2014, CASME II, SMIC-HS, and SAMM, researchers can improve the robustness and accuracy of emotion recognition algorithms.

Deep learning techniques, as explored in [85], [86], and [87], have revolutionized emotion recognition by enabling automatic feature learning from raw data. By training deep neural networks on large-scale datasets like CK+, JAFFE, Twitter, CASME-II, and Oulu, researchers can extract hierarchical representations of facial expressions, capturing both low-level visual features and high-level semantic information. Transfer learning mechanisms further assist in leveraging pre-trained models and adapting them to specific emotion recognition tasks, enhancing generalization and performance across diverse datasets. FER techniques are summarized in Table 5 as representative examples.

D. MR BASED EMOTION DETECTION

Emotional representation can be explained by several models (Section II). The Geneva Emotion Wheel [94], Differential Emotions Scale [95], and Appraisal Models are some of the models that use specific items to identify emotions depending on the events as they are framed in dimensional models (valence, arousal, and dominance). Analyzing emotional reactions utilizing multiple entities is crucial for grasping complicated emotions and capturing disparities in emotional responses. In light of this, it's crucial to examine how VR is able to evoke different types of emotions.

Consequently, this section examines the potential use of various VR media formats in the induction of various behaviors [96]. The research presents a reliable method for detecting emotion based on brain signals and data interfacing with the virtual world, providing a secure view of the information. The design of a freely available and easy-to-use application with augmented reality and uploading content to the App boosts portability and simplifies the user's perception of the virtual information added to the actual environment is the most noticeable element of this work. Happiness and sadness are practically detected, and other emotions are detected using the DEAP dataset.

With the help of this system, a depressed person can recover from his or her depression. Support Vector Machine is used to classify excited, angry, pleasant, and sad feelings, and calming videos for each mood are played in the virtual environment [97]. There has been limited research on multimodal inputs in a mobile AR learning environment that combines emotion, markers, and speech.

The purpose of this research is to offer a framework for a mobile AR learning system that comprises the combination of multimodal inputs, including emotion, image-based markers, and speech, in order to establish how such a combination can improve the learning experience. Based on the decision tree, the suggested framework integrates multimodal inputs and evolves into a four-phase learning system that uses Kolb's experiential learning model. To examine this learning technique, 38 kids were chosen and separated into two groups for a vocabulary experiment at a primary school. Combining the three multimodal inputs–speech, marker, and emotion–produced better results quantitatively in terms of learning effectiveness, mental load, engagement, challenge, and competency. As a result, by merging multimodal inputs with Kolb's experiential learning model, the suggested multimodal framework can be utilized as a guideline to design a multimodal-based AR application for learning environments [98]. The proposed work addresses the difficulties in identifying people and assessing head posture. The study of the features generated at intermediate levels by reducing the amount of kernals is done, which improves the person detection performance.

Furthermore, the learned attributes are passed to forest trees in order to identify the detected person's exact head attitude. On the Pointing 04 and Facepix benchmark datasets, the proposed Forest CNN (FCNN) architecture is evaluated for the head pose estimation methods [99]. This Human-Computer Interaction discipline relies on the algorithmic robustness and sensitivity of the sensor to improve recognition. Sensors play an important part in precise detection by delivering very high-quality input, boosting the system's efficiency and reliability. Automatic identification of human emotions would aid in teaching machines social intelligence. This paper provides a brief overview of numerous approaches and techniques used for emotion recognition. The article provides an overview of the databases that are used as data sets for algorithms that recognize emotions through facial expressions.

Later, Microsoft HoloLens (MHL), a mixed reality device, is introduced for viewing emotion recognition in Augmented Reality (AR). A brief overview of its sensors, their application in emotion identification, and some preliminary findings of emotion recognition utilizing MHL are presented. The study then compares the outcomes of emotion recognition by the MHL and a regular webcamera [100]. The authors in [101],investigated the effects of social cognition training (SCT) delivered through VR. They analyzed data from task logs in a pilot study and randomized controlled trials on VR-SCT. The study explored the impact of treatment sessions on VR accuracy and response time, as well as associations between baseline performance and VR accuracy.

While emotion recognition improved during VR-SCT, the authors noted that improvements in VR may not necessarily generalize to non-VR tasks and daily life. In the study [102], the authors estimated emotions during MR games. They used EEG data and self-assessments to label emotions based on game type. The proposed method achieved high accuracy (99.80% for SVM with STD and 99.00 for MLR with the mean method) and contributes to understanding human emotion in VR and MR contexts, potentially improving future experiences and content.

The authors in [103], developed a real-time technique for detecting the affective states of VR users. They conducted an experiment with 18 participants who observed 16 emotionally charged videos in a VR home theater while their EEG signals were recorded. The technique was evaluated using two variants: one based on Linear Mixed-Effects (LME) models for feature selection, and the other using Recursive Feature Elimination with Cross Validation (RFECV). Both variants demonstrated reliable performance, with classification model accuracies ranging from 87% to 93%.

In the research [104], the authors address the need for care robots to consider emotional well-being when assisting elderly and disabled individuals. They developed an emotion detection and recognition platform that combines a virtual-reality-interact immersive environment with physiological

TABLE 5. An overview of visual based emotion recognitions studies.

Research	Main Contributions	Dataset	Evaluation
[76]	Fully automatic face expression identification, Utilize subsequent frames based on elastic bunch graph,	CK+	Accuracy
[77]	Fuzzification, Track AAM characteristics	custom	Accuracy
[78]	Linear discriminant analysis to recognize emotions	JAFFEE MMI SFEW	Visual Emotion Recognition
[79]	Auto–encoders, SOM-based classification, Use of deep neural network, Combine geometric characteristics with LBP features	MMI	Accuracy
[80]	Auto–encoders, SOM-based classification, deep neural network, Combine geometric characteristics with LBP features	CASME II	Accuracy
[81]	Subtle deformations in emotion classification	BU-4DFE	Accuracy
[82]	FER employing multi-model 2D and 3D films, convolution network, Encode both static and dynamic information	BU-4DFE	Accuracy
[88]	Recognition of face micro-expressions, weighting scheme as well as the excellent spatiotemporal descriptor HOG3D for action recognition	HOG3D	Accuracy
[83]	TCreation of the neutrality facial images	BU-4DFE CK+ MMI	Accuracy
[84]	The "identification by generating" system, brand-new hard negative generation (HNG) network and a generalized radial metric learning (RML) network, is the main contribution of this research.	CK+, MMI	Accuracy
[85]	CNN to operate as distinct streams within a bi-stream identity-aware net	MMI CK+	Accuracy
[86]	Attention-based Salient Expressional Region Descriptor (SERD) and the Multi-Path Variation-Suppressing Network modules	FER	Accuracy
[87]	Uncertainty using a straightforward but effective Self-Cure Network (SCN)	SemEval 2014	Accuracy
[89]	Dynamic-temporal stream, static-spatial stream, and local-spatial stream module for the TSCNN, Aims to learn and incorporate time, complete facial area, and nearest neighbor signals again for face	CASME II, SMIC-HS, SAMM	Accuracy
[90]	Deep emotion-conditional adaptation net (ECAN) that can learn domain-invariant and discriminative feature representations	CK+ MMI JAFFE	Accuracy
[91]	Transfer knowledge of Emotional Education Mechanism (EEM), Self-taught student network and a competent teacher network (STSN)	Twitter	Accuracy
[92]	Convolutional networks to encode the spatial properties of microexpressions at several expression-states	CASME-II	Accuracy
[93]	Siamese-cascaded metric learning framework that teaches fine-grained distinctions between expressions in video-based tasks	CK+, Oulu, MMI	Accuracy

signal detection. By synchronously recording signals such as Electrocardiograph (ECG), Electromyogram (EMG), Electrodermal activity (EDA), and functional near-infrared spectroscopy (fNIRS), they explored how objective environmental factors impact subjective feelings.

The verification experiments demonstrated the system's effectiveness in detecting emotions The authors in [105], address the importance of emotion recognition in human-computer interaction (HCI), neuroscience, and psychology. They focus on EEG-based emotion recognition using a publicly available dataset called VREED. By extracting differential entropy (DE) features from different wavebands (theta, alpha, beta, and gamma), they classify emotional states (positive/negative). The best average accuracy achieved was 76.22% Â± 2.06 using the Support Vector Machine (SVM) classifier. Interestingly, the gamma band showed the highest average accuracy, consistent with previous EEG-based emotion recognition studies.

Certainly! In this research [99], the authors explore emotion recognition in computer science. They discuss the importance of non-verbal cues (such as gestures, body movement, and facial expressions) for conveying feelings and feedback to users. The study covers various approaches and techniques for emotion recognition, including a focus on using the Microsoft HoloLens (MHL) mixed reality device and comparing its results with a regular webcam.

In the research [106], the authors propose a hybrid approach for emotion detection that combines image-based and sound-based features. They acquire audio and video samples, extract the corresponding image and sound samples, and then fuse the features obtained from both modalities. The fused features are used to train an emotion detection hybrid model, which is applied to recognize emotions in processed audio and video data. Overall, this method aims to improve prediction accuracy and enhance model robustness for real-time emotion recognition in various applications, including VR and MR experiences.

The authors in [107], address the challenging task of multimodal emotion recognition in real-world environments. They emphasize the difficulty of recognizing affective or physiological states in individuals, both for humans and computer systems. The study focuses on finding discriminative features, as this approach holds promise for improving multimodal emotion recognition. The authors propose an enhanced autocorrelation feature for multi-pitch detection and compare its performance to other state-of-the-art features in signal and speech processing. Notably, the enhanced autocorrelation outperforms other features on a challenging dataset, which lies between naturalistic emotional utterances and well-understood acted emotional datasets. This work contributes to advancing emotion recognition techniques in practical settings.

The authors in [108], addressed the challenge of detecting emotions in VR games, where the Head-Mounted Display (HMD) obscures crucial facial features like eyes and eyebrows. They trained a Convolutional Neural Network (CNN) on modified FER2013 dataset images to predict emotions in full-face images. The model accurately recognized seven emotions (anger, happiness, disgust, fear, impartiality, sadness, and surprise). By testing it on VR games and collecting self-reported emotion data, they demonstrated the potential for enhancing gameplay analysis and creating more engaging experiences.

Researchers introduced in [109], a novel framework for emotion recognition using frontal EEG (electroencephalogram) in VR affective scenes. Unlike traditional wet electrodes, they employed textile dry electrodes for data collection. Key steps included feature extraction from time, frequency, and space domains, and model stacking using GBDT, RF, and SVM. The framework achieved an impressive mean accuracy of approximately 81.30%, outperforming previous studies. Its practical application lies in wearable devices for EEG-based emotion recognition in VR environments. This paper describes a new approach to designing and developing an AR and haptics-based STEM product that uses Vuforia, Unity 3D, and Open-Haptics to improve student engagement.

A computer vision-based system is also built and implemented in the real-time website to detect students' engagement in online learning, which uses features such as facial emotions, position estimate, and head rotation. Marker-based augmentation is used and with the use of the haptic Touch Omni device a sensation of touch is provided for an immersive experience for the students. As a result, the proposed method can boost classroom learning tasks by utilizing Augmented Reality and Haptics-based STEM products and evaluating them with a computer vision system. An extensive user survey is conducted and analysis is done on the suggested setup in order to verify the increase in engagement.

Emotion plays an important part in the learning process since it is linked to motivation, interest, and attention. Affective states are manifested in the brain and in general bodily activity. Biosignals, like heart rate (HR), electroencephalography (EEG), and electrodermal activity (EDA) are physiological expressions that are altered by emotional state. Analyzing these biosignal records can reveal a person's emotional state. Modern medical education has made significant strides toward more diversified learning resources through the use of virtual reality (VR) and mixed reality (MR) apps.

The purpose of this research is to investigate the usefulness of wearable biosensors for affect detection in a learning process involving a serious game on the Microsoft HoloLens VR/MR platform. Methods: During two educational sessions, a wearable array of sensors recording HR, EDA, and EEG signals was deployed by 11 participants with varying educational levels (undergraduate, postgraduate, and specialist neurosurgeon doctors). The first scenario was a standard virtual patient case used to establish the participant's personal biosignal baselines. The second was a neuroanatomy case in a VR/MR setting. We recorded EEG (theta/beta ratio and alpha rhythm), HR, and EDA as affective markers [110].

In [111], the authors investigate the development of approaches for identifying the affective states of virtual reality (VR) users in real time utilizing EEG data and virtual reality visual stimuli. They extract features using an emotion classification algorithm based on convolutional neural networks. They employ a Deep Network of Reinforcement Learning with Q-Learning. In [112], the authors examine the use of EEG data to detect individual emotional states in virtual reality scenarios, as well as how to increase computing efficiency and emotional valence recognition accuracy. It makes no explicit mention of other papers on emotion recognition in virtual reality. They use EEG signals to determine individual emotional states in virtual reality environments. Their findings showed an improvement in computing efficiency and emotional valence recognition accuracy.

This paper [113] discusses applying machine learning methods to detect EEG-based emotion states via virtual video inputs. It does not expressly reference any other publications on emotion recognition in virtual reality. Three prominent machine learning models are implemented on an obtained database. Researchers EEG signals were extracted into 34 temporal and frequency domain characteristics. Researchers in [103] deploy an EEG signal-based approach to detect the affective states of VR users in real time. They use Recursive Feature Elimination with Cross Validation (RFECV) to choose features. Their findings include the development of a technique for inferring users' affective states based on electrophysiological responses, as well as the creation of user-dependent models for real-time affect detection in virtual reality.

In this work, the authors [114] look into the use of Heart Rate and Electro Demography signals for emotion classification in Virtual Reality. For their findings, they combine Heart Rate and Electro demography signals for emotion identification in VR and achieved good accuracy results with the use of a KNN classifier. In this paper [105] is about EEG-based emotion recognition in a virtual reality environment. For their findings they Used differential entropy (DE) features for EEG-based emotion recognition. They Employed five classifiers for automated classification of two emotional states.

TABLE 6. An overview of MR based emotion recognitions studies.

Research	Main Contributions	Approach	Model
[90]	Presents various models including the Geneva Emotion Wheel, Differential Emotions Scale, and Appraisal Models for emotion identification	Emotional representation	Dimensional models
[92]	Examines the potential of various VR media formats in inducing different emotions	VR media formats	—
[93]	Introduces a method for detecting emotions based on brain signals interfacing with the virtual world	Emotion detection based on brain signals	SVM
[94]	Proposes a framework for a mobile AR learning system integrating emotion, image-based markers, and speech inputs	Multimodal inputs in AR learning	Decision tree-based framework
[95]	Addresses difficulties in identifying people and assessing head posture using Forest CNN (FCNN) architecture	Forest CNN architecture	Forest CNN
[115]	Discusses the importance of sensors in Human-Computer Interaction (HCI) for precise detection	Sensors in HCI	Sensor-based detection
[116]	Provides an overview of various approaches and techniques used for emotion recognition	Emotion recognition approaches	–
[117]	Describes an AR and haptics-based STEM product for improving student engagement	AR and haptics-based STEM product	Vuforia, Unity 3D, OpenHaptics
[118]	Investigates the development of approaches for identifying affective states of VR users using EEG data	EEG-based emotion recognition in VR	EEG data, convolutional neural networks
[119]	Applies machine learning methods to detect EEG-based emotion states in virtual video inputs	EEG emotional state	ML models
[120]	Deploys an EEG signal-based approach to detect affective states of VR users in real time	EEG signal-based approach for affect detection	Recursive Feature Elimination with Cross Validation
[121]	Utilizes Heart Rate and Electro Demography signals for emotion identification in VR	Heart Rate and Electro Demography signals	KNN classifier
[122]	Utilizes differential entropy (DE) features for EEG-based emotion recognition in a VR environment	EEG-based emotion recognition in VR	ML classifier

VI. DATASETS

Our purpose in this section is to describe the datasets we use to detect emotions in text, audio, and visual images. There are three types of affective computing databases: textual, speech, audio, and visual. Models for emotional computing and network architectures are greatly influenced by these databases' characteristics.

A. TEXT BASED DATASET

The TSA database contains text data at various granularities, like words, sentences, and documents. The Multi-domain Sentiment (MDS) [115] database contains more than 100,000 phrases derived from Amazon.com reviews. These sentences are divided into two emotion types (positive and negative) and five sentiment categories. IMDB [116] is another significant resource that is frequently utilized for categorizing binary sentiment. There are 25,000 reviews of extremely divisive movies available for testing and 25,000 for training. The Stanford University-annotated semantic lexical database is called Stanford Sentiment Treebank (SST) [117]. It has 215,154 phrases with fine-grained emotional descriptors.

B. AUDIO BASED DATASET

There are two types of speech databases: spontaneous and non-spontaneous (simulated and induced). The first non-spontaneous voice collection was created primarily from professional actors' performances. Due to their ability to professionally mimic well-known emotional traits, these performance-based datasets are considered trustworthy. There are over 500 words in the Berlin Database of Emotional Speech (Emo-DB) [118] spoken by 10 actors (five men and five women) in various emotional states, including rage, anxiety, fear, boredom, and contempt. Artificial emotions, however, are more likely to be exaggerated than actual feelings. In an effort to close this gap, databases have been created for spontaneous speech.

C. VISUAL BASE DATASET

To create the initial FER datasets, lab volunteers voluntarily expressed their emotions (in-the-lab). A 1998 publication entitled JAFFE [76] contains 213 photographs of 10 Japanese female models expressing 7 different facial expressions. In order to develop the Cohn-Kanade+ (CK) [119], subjects were instructed to make 7 different facial expressions. In order to offer methods and benchmark findings for tracking features of face, action units (AUs), and identifying emotions, photos of face expressions were captured and evaluated. MMI [121], in contrast to CK, contains onset apex offset orders. Multiple views and multiple poses are supported by many 3D or 4D datasets made for FER. For the Binghamton University 3D Facial Expression (BU-3DFE) [122] database, 606 facial expression segments were taken from 100 people using six different facial expressions.

A huge, unrestricted collection called FER2013 [98] contains 35,887 grey images with a resolution of 48 48 pixels that were autonomously gathered using the image search API 950,000 of the one million photos in EmotionNet [123] were generated automatically. 91,793 facial photos from Emotion in the Wild (ExpW) [124] have each been individually labeled using one of the seven fundamental facial expressions. The annotating technique eliminated non-face photos. Over one million facial photos are included in AffectNet [125], of which 4 lac 50k images have been manually classified as one of eight distinct expressions (neutral, disdain, the six fundamental emotions, and more). AffectNet also includes information on the dimensional intensity of valence and arousal. The Real-world Affective Face Database (RAF-DB) [126] is a collection of 29,672 extremely different facial photos annotated by the public and acquired from the Internet (seven basic and eleven compound emotion labels). Over 16,000 video snippets that were cut from hundreds of films with different topics make up Dynamic-Facial-Expressions in the Wild (DFEW) [127].

VII. CHALLENGES

ML-based techniques have been used in affective computing's early literature [15]. The ML pipeline [140]entails the pre-processing of unprocessed signals, the creation of expertly constructed feature extractors (including selecting features when practical), and masterfully crafted classifiers. ML-based approaches for affective analysis are difficult to reuse across similar issues because to their task-specific and domain-specific feature descriptors, despite the fact that many types of hand-crafted features have been generated for diverse modality.

The most popular ML-based classifier are the SVM classifier [141], GMM classifier [68], RF classifier [142],

TABLE 7. Available databases for research purposes.

Database	Number of samples	Emotions
Oulu [128]	80	SBE
FER2013 [129]	32	SBE
SFEW [130]	32	SBE
EmotionNEt [123]	32	Compound
Oulu [128]	32	SBE
ExpW [124]	32	SBE+
AffectNEt [125]	32	SBE+
CASME II [131]	35s	SBE +
SMIC [132]	16	3, positive, negative, surprise
GoEmotion [58]	5800	27 emotions of neutral
4DFAB [133]	180	SBE
3DFE [122]	100	SBE
MMI [121]	25	SBE
CK+ [119]	123	SBE+
CK [120]	210	SBE+
MAHNOB-HCI [134]	540	Facial, Audio and EEG
NVIF [135]	100	SBE+
KTFEv2 [136]	5260	SBE+
VAD [137]	19267	SBE+
MMED [138]	22248	SBE+
JAFFE [139]	10	SBE

KNN classifier [143], and ANN classifier [144], with the SVM classifier being employed in the majority of ML-based affective computing tasks.

Due to their superior feature representation learning capabilities, DL-based models have recently gained popularity and outperformed ML-based models in the majority of affective computing applications [145]. CNNs and their derivatives are made to extract significant and discrete characteristics from static information (such as face and spectral analysis images) [146]. RNNs and its variants are made for capturing temporal dynamics for sequence information (such as physiological signals and movies) [147].

The deep spatial-temporal feature extraction is a task that can be handled by CNN-LSTM models. By enhancing data and using cross-domain learning, adversarial learning is frequently employed to increase the robustness of models [147]. Additionally, to enhance overall performance, various attention methods and auto encoders are used with DL-based techniques. The most discriminative traits appear to be automatically learned by DL-based algorithms, which appears to be an advantage. When compared to ML-based models, DL-based techniques have not yet had a significant influence on physiological emotion recognition.

In the field of emotion research, VR is being used more and more, but there are some limitations. Therefore, standards must be developed and implemented. Research can be conducted more effectively and restrictions can be reduced by applying these principles. It is important for researchers to assess the likelihood of different kinetic settings on the typical population in order to decrease motion sickness. By filtering participants at an early stage, it becomes easier to collect

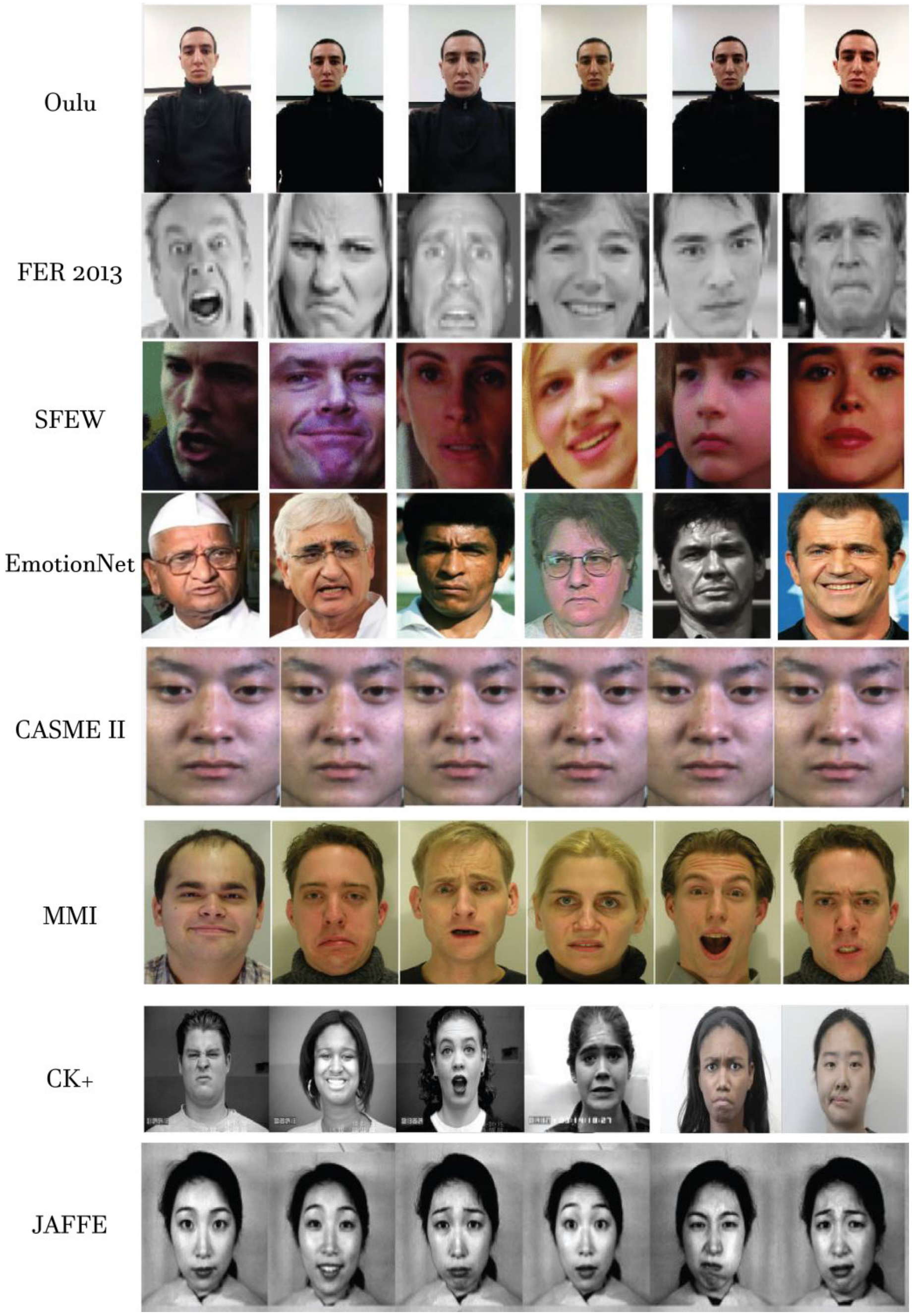

FIGURE 4. Sample images of openly available databases.

data. A questionnaire could be used as the primary method of contacting the individual.

A few of the questionnaires that can be found in the literature are the Motion Sickness Susceptibility Questionnaire (MSSQ) and the Virtual Reality Sickness Questionnaire (VRSQ) [78]. It is possible for the reader to consult. Participants must have prior exposure to immersive VR environments, as with the first. Some research findings may be influenced by the novelty bias caused by using VR for the first time. Further explanations can be found in the differences between predicted and actual feelings.

People with previous VR experience or training sessions with selected participants before data collection can lessen the effects of the first-time VR experience. As a result of understanding the perspective of a VR world and the operation of hardware, especially when using VR controllers, participants receive additional benefits. For credible research, it is necessary to confirm the applicability of

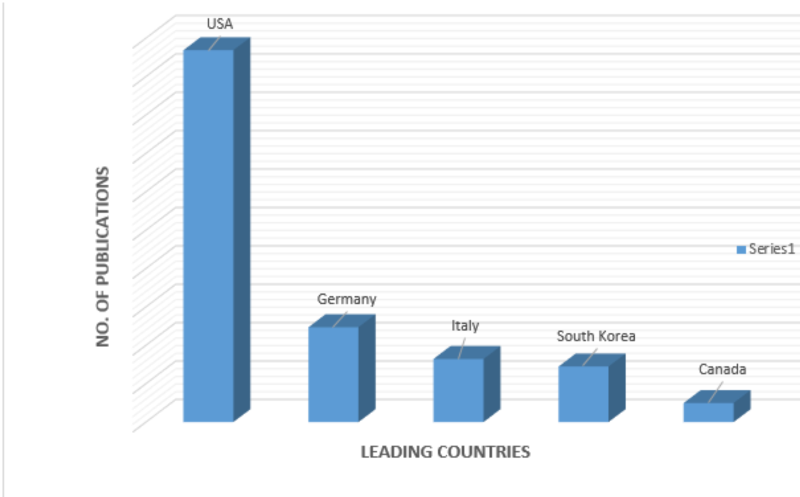

FIGURE 5. Leading countries working on affective computing with ML-AR applications based on Web of Science.

using those hardware interfaces integrated with different facial structures. Figure 5 shows the leading countries that publish articles related to affective computing only, they separately consider machine learning and augmented reality.

VIII. OPEN RESEARCH PROBLEMS

The field of affective computing focuses on developing machines and systems that can recognize, interpret, and respond to human emotions.

1) How to recognize and interpret complex emotional states more accurately and reliably is one of the open research problems in affective computing.
2) The lack of large and diverse labeled datasets to train emotion recognition models is one of the open research problems in affective computing. For ML algorithms to recognize and interpret emotions accurately, large amounts of labeled data are required. Existing emotion recognition datasets, however, are often limited in size, scope, and diversity, which can limit their performance and generalizability.
3) There is also a lack of consistency in the labeling of emotions. A standard labeling scheme for emotions does not exist, since emotions are interpreted differently by different cultures and individuals. It may be difficult to compare results across studies and develop emotion recognition models that are effective across diverse populations because of this.
4) Another open research problem is how to develop more robust and interpretable ML models for affective computing. ML models used in affective computing, such as deep neural networks, are known as "black box" models, which make understanding how they work difficult. As a result, emotion recognition models are difficult to trust and fine-tune for specific applications due to their lack of interpretability.
5) A more personalized and adaptive emotion recognition model is needed that can take into account individual differences and changes in emotional states over time. Developing models that adapt to individual users' emotional patterns requires learning from their emotional expressions.
6) The development of ML approaches for affective computing and the development of more accurate and effective emotion recognition systems will be made possible by addressing these research problems.

IX. CONCLUSION

This survey examined research works based on current studies on affective computing and an illustrative taxonomy of affective computing. Various corresponding measures were used to assess the recognition outcomes obtained by classification or regression. Emotional computing develops computational models for DL-based or ML-based affective understanding, as well as benchmark databases for training. A survey of the commonly used baseline databases for affective computing was also presented in this article, including textual, audio, and visual databases. Most affective analysis techniques can be applied to these publicly available databases. ML-based methods, DL-based approaches, and mixed reality-based approaches are some of the recent developments in affective computing. A survey of text sentiment classification, voice emotion recognition, and visual emotion recognition (FER and EBGR) is presented.

Even though affective computing systems that use either unimodal or multimodal data have made substantial progress, there are only a few reliable and potent algorithms to predict emotion and distinguish feelings. Consequently, affective computing's important future research directions are summarized as follows.

New and expanded baseline databases, particularly multi-modal effect databases (textual, audio, visual, physiological), would be crucial. Affective analysis problems such as FER under occlusions and fake emotion expression need to be solved. Fusion methodologies, particularly those based on regulation or statistics, need to be improved.

In the presence of constrained or biased databases, zero/few-shot learning and unsupervised learning techniques, such as self-supervised learning, can improve the stability and robustness of affective analysis. The use of emotive analysis in robotics is well known. As discussed in this paper, emotionally intelligent robots are capable of accurately mimicking and reacting to emotions.

REFERENCES

[1] K. S. Fleckenstein, "Defining affect in relation to cognition: A response to Susan McLeod," *J. Adv. Composition*, vol. 11, no. 2, pp. 447–453, 1991.
[2] R. W. Picard, "Affective computing," *J. Adv. Composition*, vol. 1, pp. 447–453, Aug. 2000.
[3] R. W. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: Analysis of affective physiological state," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 10, pp. 1175–1191, Jul. 2001.
[4] J.-S. Park, J.-H. Kim, and Y.-H. Oh, "Feature vector classification based speech emotion recognition for service robots," *IEEE Trans. Consum. Electron.*, vol. 55, no. 3, pp. 1590–1596, Aug. 2009.
[5] K. Anderson and P. W. McOwan, "A real-time automated system for the recognition of human facial expressions," *IEEE Trans. Syst., Man Cybern., B Cybern.*, vol. 36, no. 1, pp. 96–105, Feb. 2006.
[6] Z. Yin, M. Zhao, Y. Wang, J. Yang, and J. Zhang, "Recognition of emotions using multimodal physiological signals and an ensemble deep learning model," *Comput. Methods Programs Biomed.*, vol. 140, pp. 93–110, Mar. 2017.
[7] M. Shad Akhtar, D. Singh Chauhan, D. Ghosal, S. Poria, A. Ekbal, and P. Bhattacharyya, "Multi-task learning for multi-modal emotion recognition and sentiment analysis," 2019, *arXiv:1905.05812*.
[8] G. Chanel, J. Kronegg, D. Grandjean, and T. Pun, "Emotion assessment: Arousal evaluation using EEG's and peripheral physiological signals," in *Proc. Multimedia Content Represent., Classification Security: Int. Workshop MRCS*, vol. 1, 2006, pp. 530–537.
[9] L. A. Schmidt and L. J. Trainor, "Frontal brain electrical activity (EEG) distinguishes valence and intensity of musical emotions," *Cognition Emotion*, vol. 15, no. 4, pp. 487–500, Jul. 2001.
[10] L.-P. Morency, J. Whitehill, and J. Movellan, "Generalized adaptive view-based appearance model: Integrated framework for monocular head pose estimation," in *Proc. 8th IEEE Int. Conf. Autom. Face Gesture Recognit.*, Sep. 2008, pp. 1–8.
[11] E. Cambria, D. Das, S. Bandyopadhyay, and A. Feraco, "Affective computing and sentiment analysis," in *A Practical Guide to Sentiment Analysis*, vol. 1. Heidelberg, Germany: Springer, 2017, pp. 1–10.
[12] M. Munezero, C. S. Montero, E. Sutinen, and J. Pajunen, "Are they different? Affect, feeling, emotion, sentiment, and opinion detection in text," *IEEE Trans. Affect. Comput.*, vol. 5, no. 2, pp. 101–111, Apr. 2014.
[13] P. Ekman, T. Dalgleish, and M. Power, "Basic emotions," in *Handbook of Cognition and Emotion*, vol. 98, nos. 45–60. Chichester, U.K.: Wiley, 1999, p. 16.
[14] A. Mehrabian, *Basic Dimensions for a General Psychological Theory: Implications for Personality, Social, Environmental, and Developmental Studies*. Cambridge, MA, USA: Oelgeschlager, Gunn & Hain, 1980.
[15] C. A. Corneanu, M. O. Simón, J. F. Cohn, and S. E. Guerrero, "Survey on RGB, 3D, thermal, and multimodal approaches for facial expression recognition: History, trends, and affect-related applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 8, pp. 1548–1568, Aug. 2016.
[16] Z. Yin and J. Zhang, "Task-generic mental fatigue recognition based on neurophysiological signals and dynamical deep extreme learning machine," *Neurocomputing*, vol. 283, pp. 266–281, Mar. 2018.
[17] A. J. Holte and F. R. Ferraro, "True colors: Grayscale setting reduces screen time in college students," *Social Sci. J.*, vol. 60, no. 2, pp. 274–290, Apr. 2023.
[18] W. Chen, "A novel long short-term memory network model for multimodal music emotion analysis in affective computing," *J. Appl. Sci. Eng.*, vol. 26, no. 3, pp. 367–376, 2022.
[19] Y. Wang, W. Song, W. Tao, A. Liotta, D. Yang, X. Li, S. Gao, Y. Sun, W. Ge, W. Zhang, and W. Zhang, "A systematic review on affective computing: Emotion models, databases, and recent advances," *Inf. Fusion*, vols. 83–84, pp. 19–52, Jul. 2022.
[20] J. Zhang, Z. Yin, P. Chen, and S. Nichele, "Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review," *Inf. Fusion*, vol. 59, pp. 103–126, Jul. 2020.
[21] R. Arya, J. Singh, and A. Kumar, "A survey of multidisciplinary domains contributing to affective computing," *Comput. Sci. Rev.*, vol. 40, May 2021, Art. no. 100399.
[22] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 1, pp. 39–58, Jan. 2009.
[23] S. Zhao, G. Ding, Q. Huang, T.-S. Chua, B. Schuller, and K. Keutzer, "Affective image content analysis: A comprehensive survey," in *Proc. 27th Int. Joint Conf. Artif. Intell. Survey Track*, 2018, pp. 5534–5541.
[24] S. Zhao, S. Wang, M. Soleymani, D. Joshi, and Q. Ji, "Affective computing for large-scale heterogeneous multimedia data: A survey," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 15, no. 3s, pp. 1–32, Nov. 2019.
[25] B. Ko, "A brief review of facial emotion recognition based on visual information," *Sensors*, vol. 18, no. 2, p. 401, Jan. 2018.
[26] S. Li and W. Deng, "Deep facial expression recognition: A survey," *IEEE Trans. Affect. Comput.*, vol. 13, no. 3, pp. 1195–1215, Jul. 2022.
[27] W. Merghani, A. K. Davison, and M. Hoon Yap, "A review on facial micro-expressions analysis: Datasets, features and metrics," 2018, *arXiv:1805.02397*.
[28] G. R. Alexandre, J. M. Soares, and G. A. P. Thé, "Systematic review of 3D facial expression recognition methods," *Pattern Recognit.*, vol. 100, Apr. 2020, Art. no. 107108.
[29] S. Poria, N. Majumder, R. Mihalcea, and E. Hovy, "Emotion recognition in conversation: Research challenges, datasets, and recent advances," *IEEE Access*, vol. 7, pp. 100943–100953, 2019.
[30] R. Somarathna, T. Bednarz, and G. Mohammadi, "Virtual reality for emotion elicitation—A review," *IEEE Trans. Affect. Comput.*, vol. 14, no. 4, pp. 2626–2645, Oct. 2023.
[31] F. Pozzi, E. Fersini, E. Messina, and B. Liu, *Sentiment Analysis in Social Networks*. Cambridge, MA, USA: Morgan Kaufmann, 2016.
[32] P. J. Stone and E. B. Hunt, "A computer approach to content analysis: Studies using the general inquirer system," in *Proc. Spring Joint Comput. Conf.*, 1963, pp. 241–256.
[33] J. Suttles and N. Ide, "Distant supervision for emotion classification with discrete binary values," in *Proc. 14th Int. Conf. Comput. Linguistics Intell. Text Process. (CICLing)*, 2013, pp. 121–136.
[34] L. A. M. Oberländer and R. Klinger, "An analysis of annotated corpora for emotion classification in text," in *Proc. 27th Int. Conf. Comput. Linguistics (COLING)*, 2018, pp. 2104–2119.
[35] S. Mohammad, F. Bravo-Marquez, M. Salameh, and S. Kiritchenko, "SemEval-2018 task 1: Affect in tweets," in *Proc. 12th Int. Workshop Semantic Eval.*, 2018, pp. 1–17.
[36] P. Nakov, S. Rosenthal, S. Kiritchenko, S. M. Mohammad, Z. Kozareva, A. Ritter, V. Stoyanov, and X. Zhu, "Developing a successful SemEval task in sentiment analysis of Twitter and other social media texts," *Lang. Resour. Eval.*, vol. 50, no. 1, pp. 35–65, Mar. 2016.
[37] S. Mohammad and S. Kiritchenko, "WikiArt emotions: An annotated dataset of emotions evoked by art," in *Proc. 11th Int. Conf. Lang. Resour. Eval. (LREC)*, 2018, pp. 1–14.
[38] B. Felbo, A. Mislove, A. Søgaard, I. Rahwan, and S. Lehmann, "Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm," 2017, *arXiv:1708.00524*.
[39] A. Agrawal, A. An, and M. Papagelis, "Learning emotion-enriched word representations," in *Proc. 27th Int. Conf. Comput. Linguistics*, 2018, pp. 950–961.
[40] J. Deng and F. Ren, "Multi-label emotion detection via emotion-specified feature extraction and emotion correlation learning," *IEEE Trans. Affect. Comput.*, vol. 14, no. 1, pp. 475–486, Jan. 2023.
[41] H. Alhuzali and S. Ananiadou, "SpanEmo: Casting multi-label emotion classification as span-prediction," 2021, *arXiv:2101.10038*.
[42] X. Ding, B. Liu, and P. S. Yu, "A holistic lexicon-based approach to opinion mining," in *Proc. Int. Conf. Web Search Data Mining*, 2008, pp. 231–240.
[43] P. Melville, W. Gryc, and R. D. Lawrence, "Sentiment analysis of blogs by combining lexical knowledge with text classification," in *Proc. 15th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jun. 2009, pp. 1275–1284.

[44] S. Poria, E. Cambria, G. Winterstein, and G.-B. Huang, "Sentic patterns: Dependency-based rules for concept-level sentiment analysis," *Knowledge-Based Syst.*, vol. 69, pp. 45–63, Oct. 2014.

[45] I. Blekanov, M. Kukarkin, A. Maksimov, and S. Bodrunova, "Sentiment analysis for ad hoc discussions using multilingual knowledge-based approach," in *Proc. 3rd Int. Conf. Appl. Inf. Technol.*, Nov. 2018, pp. 117–121.

[46] E. Aydogan and M. A. Akcayol, "A comprehensive survey for sentiment analysis tasks using machine learning techniques," in *Proc. Int. Symp. Innov. Intell. Syst. Appl. (INISTA)*, Aug. 2016, pp. 1–7.

[47] T. Mullen and N. Collier, "Sentiment analysis using support vector machines with diverse information sources," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2004, pp. 412–418.

[48] Y. Xia, E. Cambria, A. Hussain, and H. Zhao, "Word polarity disambiguation using Bayesian model and opinion-level features," *Cognit. Comput.*, vol. 7, no. 3, pp. 369–380, Jun. 2015.

[49] A. Valdivia, M. V. Luzón, E. Cambria, and F. Herrera, "Consensus vote models for detecting and filtering neutrality in sentiment analysis," *Inf. Fusion*, vol. 44, pp. 126–135, Nov. 2018.

[50] T. Le, "A hybrid method for text-based sentiment analysis," in *Proc. Int. Conf. Comput. Sci. Comput. Intell. (CSCI)*, Dec. 2019, pp. 1392–1397.

[51] D. Li, R. Rzepka, M. Ptaszynski, and K. Araki, "A novel machine learning-based sentiment analysis method for Chinese social media considering Chinese slang lexicon and emoticons," in *Proc. AffCon@ AAAI*, 2019, p. 2328.

[52] F. Liu, L. Zheng, and J. Zheng, "HieNN-DWE: A hierarchical neural network with dynamic word embeddings for document level sentiment classification," *Neurocomputing*, vol. 403, pp. 21–32, Aug. 2020.

[53] A. Conneau, H. Schwenk, L. Barrault, and Y. Lecun, "Very deep convolutional networks for text classification," 2016, *arXiv:1606.01781*.

[54] B. Huang and K. M. Carley, "Parameterized convolutional neural networks for aspect level sentiment classification," 2019, *arXiv:1909.06276*.

[55] A. Mousa and B. Schuller, "Contextual bidirectional long short-term memory recurrent neural network language models: A generative approach to sentiment analysis," in *Proc. 15th Conf. Eur. Chapter Assoc. Comput. Linguistics*, 2017, pp. 1023–1032.

[56] W. Wang, S. Jialin Pan, D. Dahlmeier, and X. Xiao, "Recursive neural conditional random fields for aspect-based sentiment analysis," 2016, *arXiv:1603.06679*.

[57] K. Chakraborty, S. Bhatia, S. Bhattacharyya, J. Platos, R. Bag, and A. E. Hassanien, "Sentiment analysis of COVID-19 tweets by deep learning classifiers—A study to show how popularity is affecting accuracy in social media," *Appl. Soft Comput.*, vol. 97, Dec. 2020, Art. no. 106754.

[58] D. Demszky, D. Movshovitz-Attias, J. Ko, A. Cowen, G. Nemade, and S. Ravi, "GoEmotions: A dataset of fine-grained emotions," 2020, *arXiv:2005.00547*.

[59] C. Min Lee and S. S. Narayanan, "Toward detecting emotions in spoken dialogs," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 2, pp. 293–303, Mar. 2005.

[60] J. Pohjalainen, F. F. Ringeval, Z. Zhang, and B. Schuller, "Spectral and cepstral audio noise reduction techniques in speech emotion recognition," in *Proc. 24th ACM Int. Conf. Multimedia*, 2016, pp. 670–674.

[61] Z. Huang, M. Dong, Q. Mao, and Y. Zhan, "Speech emotion recognition using CNN," in *Proc. 22nd ACM Int. Conf. Multimedia*, Nov. 2014, pp. 801–804.

[62] H. M. Fayek, M. Lech, and L. Cavedon, "Evaluating deep learning architectures for speech emotion recognition," *Neural Netw.*, vol. 92, pp. 60–68, Aug. 2017.

[63] F. Eyben, M. Wöllmer, and B. Schuller, "OpenEAR—Introducing the Munich open-source emotion and affect recognition toolkit," in *Proc. 3rd Int. Conf. Affect. Comput. Intell. Interact. Workshops*, Sep. 2009, pp. 1–6.

[64] C. Busso, S. Lee, and S. Narayanan, "Analysis of emotionally salient aspects of fundamental frequency for emotion detection," *IEEE Trans. Audio, Speech, Lang., Process.*, vol. 17, no. 4, pp. 582–596, May 2009.

[65] M. Lugger and B. Yang, "The relevance of voice quality features in speaker independent emotion recognition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. ICASSP*, Apr. 2007, pp. 4–17.

[66] Y. Jin, P. Song, W. Zheng, and L. Zhao, "A feature selection and feature fusion combination method for speaker-independent speech emotion recognition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2014, pp. 4808–4812.

[67] D. Bitouk, R. Verma, and A. Nenkova, "Class-level spectral features for emotion recognition," *Speech Commun.*, vol. 52, nos. 7–8, pp. 613–625, Jul. 2010.

[68] E. Navas, I. Hernaez, and I. Luengo, "An objective and subjective study of the role of semantics and prosodic features in building corpora for emotional TTS," *IEEE Trans. Audio, Speech Lang., Process.*, vol. 14, no. 4, pp. 1117–1127, Jul. 2006.

[69] E. Yüncü, H. Hacihabiboglu, and C. Bozsahin, "Automatic speech emotion recognition using auditory models with binary decision tree and SVM," in *Proc. 22nd Int. Conf. Pattern Recognit.*, Aug. 2014, pp. 773–778.

[70] A. Bhavan, P. Chauhan, Hitkul, and R. R. Shah, "Bagged support vector machines for emotion recognition from speech," *Knowl.-Based Syst.*, vol. 184, Nov. 2019, Art. no. 104886.

[71] L. Chen, W. Su, Y. Feng, M. Wu, J. She, and K. Hirota, "Two-layer fuzzy multiple random forest for speech emotion recognition in human-robot interaction," *Inf. Sci.*, vol. 509, pp. 150–163, Jan. 2020.

[72] A. M. Badshah, J. Ahmad, N. Rahim, and S. W. Baik, "Speech emotion recognition from spectrograms with deep convolutional neural network," in *Proc. Int. Conf. Platform Technol. Service (PlatCon)*, 2017, pp. 1–5.

[73] S. Ghosh, E. Laksana, L.-P. Morency, and S. Scherer, "Representation learning for speech emotion recognition," in *Proc. Interspeech*, Sep. 2016, pp. 3603–3607.

[74] Y.-L. Tian, T. Kanade, and J. F. Colin, "Recognizing action units for facial expression analysis," in *Proc. Multimodal Interface Hum.-Mach. Commun.*, 2002, pp. 32–66.

[75] M. Pantie and L. J. M. Rothkrantz, "Automatic analysis of facial expressions: The state of the art," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 12, pp. 1424–1445, Jul. 2000.

[76] D. Ghimire and J. Lee, "Geometric feature-based facial expression recognition in image sequences using multi-class AdaBoost and support vector machines," *Sensors*, vol. 13, no. 6, pp. 7714–7734, Jun. 2013.

[77] Sujono and A. A. S. Gunawan, "Face expression detection on Kinect using active appearance model and fuzzy logic," *Proc. Comput. Sci.*, vol. 59, pp. 268–274, 2015.

[78] Y. Yan, Z. Zhang, S. Chen, and H. Wang, "Low-resolution facial expression recognition: A filter learning perspective," *Signal Process.*, vol. 169, Apr. 2020, Art. no. 107370.

[79] A. Majumder, L. Behera, and V. K. Subramanian, "Automatic facial expression recognition system using deep network-based data fusion," *IEEE Trans. Cybern.*, vol. 48, no. 1, pp. 103–114, Jan. 2018.

[80] Y. Wang, J. See, R. C.-W. Phan, and Y.-H. Oh, "Efficient spatio-temporal local binary patterns for spontaneous facial micro-expression recognition," *PLoS ONE*, vol. 10, no. 5, May 2015, Art. no. e0124674.

[81] Q. Zhen, D. Huang, H. Drira, B. B. Amor, Y. Wang, and M. Daoudi, "Magnifying subtle facial motions for effective 4D expression recognition," *IEEE Trans. Affect. Comput.*, vol. 10, no. 4, pp. 524–536, Oct. 2019.

[82] Y. Yao, D. Huang, X. Yang, Y. Wang, and L. Chen, "Texture and geometry scattering representation-based facial expression recognition in 2D+3D videos," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 14, no. 1s, pp. 1–23, Mar. 2018.

[83] H. Yang, U. Ciftci, and L. Yin, "Facial expression recognition by de-expression residue learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2168–2177.

[84] X. Liu, B. V. K. Vijaya Kumar, P. Jia, and J. You, "Hard negative generation for identity-disentangled facial expression recognition," *Pattern Recognit.*, vol. 88, pp. 1–12, Apr. 2019.

[85] Z. Meng, P. Liu, J. Cai, S. Han, and Y. Tong, "Identity-aware convolutional neural network for facial expression recognition," in *Proc. 12th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, Jul. 2017, pp. 558–565.

[86] S. Xie, H. Hu, and Y. Wu, "Deep multi-path convolutional neural network joint with salient region attention for facial expression recognition," *Pattern Recognit.*, vol. 92, pp. 177–191, Aug. 2019.

[87] K. Wang, X. Peng, J. Yang, S. Lu, and Y. Qiao, "Suppressing uncertainties for large-scale facial expression recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 6897–6906.

[88] M. Chen, H. T. Ma, J. Li, and H. Wang, "Emotion recognition using fixed length micro-expressions sequence and weighting method," in *Proc. IEEE Int. Conf. Real-Time Comput. Robot. (RCAR)*, 2016, pp. 427–430.

[89] B. Song, K. Li, Y. Zong, J. Zhu, W. Zheng, J. Shi, and L. Zhao, "Recognizing spontaneous micro-expression using a three-stream convolutional neural network," *IEEE Access*, vol. 7, pp. 184537–184551, 2019.

[90] S. Li and W. Deng, "A deeper look at facial expression dataset bias," *IEEE Trans. Affect. Comput.*, vol. 13, no. 2, pp. 881–893, Apr. 2022.

[91] H. Li, N. Wang, X. Ding, X. Yang, and X. Gao, "Adaptively learning facial expression representation via C-F labels and distillation," *IEEE Trans. Image Process.*, vol. 30, pp. 2016–2028, 2021.

[92] D. H. Kim, W. J. Baddar, and Y. M. Ro, "Micro-expression recognition with expression-state constrained spatio-temporal feature representations," in *Proc. 24th ACM Int. Conf. Multimedia*, 2016, pp. 382–386.
[93] D. Liu, X. Ouyang, S. Xu, P. Zhou, K. He, and S. Wen, "SAANet: Siamese action-units attention network for improving dynamic facial expression recognition," *Neurocomputing*, vol. 413, pp. 145–157, Nov. 2020.
[94] Z. Zhao, Y. Zhao, Z. Bao, H. Wang, Z. Zhang, and C. Li, "Deep spectrum feature representations for speech emotion recognition," in *Proc. Joint Workshop 4th Workshop Affect. Social Multimedia Comput. 1st Multi-Modal Affect. Comput. Large-Scale Multimedia Data*, Oct. 2018, pp. 27–33.
[95] J. Han, Z. Zhang, Z. Ren, F. Ringeval, and B. Schuller, "Towards conditional adversarial training for predicting emotions from speech," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Jun. 2018, pp. 6822–6826.
[96] G. Amoolya, K. K. Ashna, G. S. V. Swetha, G. Das, J. Geeta, and S. N. George, "EEG-based emotion recognition and its interface with augmented reality," in *Applied Soft Computing and Communication Networks*. Singapore: Springer, 2021, pp. 41–56.
[97] N. C. Hashim, N. A. A. Majid, H. Arshad, H. Hashim, and Z. Abdi Alkareem Alyasseri, "Mobile augmented reality based on multimodal inputs for experiential learning," *IEEE Access*, vol. 10, pp. 78953–78969, 2022.
[98] D. Anitta and A. Annis Fathima, "CNN—Forest based person identification and head pose estimation for AI based applications," *Int. J. Uncertainty, Fuzziness Knowl.-Based Syst.*, vol. 31, no. Supp01, pp. 47–63, May 2023.
[99] D. Mehta, M. Siddiqui, and A. Javaid, "Facial emotion recognition: A survey and real-world user experiences in mixed reality," *Sensors*, vol. 18, no. 2, p. 416, Feb. 2018.
[100] H. A. Poonja, M. A. Shirazi, M. J. Khan, and K. Javed, "Engagement detection and enhancement for STEM education through computer vision, augmented reality, and haptics," *Image Vis. Comput.*, vol. 136, Aug. 2023, Art. no. 104720.
[101] S. A. Nijman, W. Veling, M. E. Timmerman, and G. H. M. Pijnenborg, "Trajectories of emotion recognition training in virtual reality and predictors of improvement for people with a psychotic disorder," *Cyberpsychology, Behav., Social Netw.*, vol. 26, no. 4, pp. 288–299, Apr. 2023.
[102] A. S. M. Miah, J. Shin, M. M. Islam, and M. K. I. Molla, "Natural human emotion recognition based on various mixed reality (MR) games and electroencephalography (EEG) signals," in *Proc. IEEE 5th Eurasian Conf. Educ. Innov. (ECEI)*, 2022, pp. 408–411.
[103] A. Pinilla, J.-N. Voigt-Antons, J. Garcia, W. Raffe, and S. Möller, "Real-time affect detection in virtual reality: A technique based on a three-dimensional model of affect and EEG signals," *Frontiers Virtual Reality*, vol. 3, Jan. 2023, Art. no. 964754.
[104] B. Yang, B. Sun, Q. Zhang, Y. Wang, Y. Li, J. Yang, and L. Gu, "Emotion detection during human–robots interaction in visual reality environment," in *Proc. Robot. Rehabil. Intell., 1st Int. Conf. (ICRRI)*, Fushun, China. Cham, Switzerland: Springer, Sep. 2020, pp. 79–92.
[105] H. Uyanık, S. T. A. Ozcelik, Z. B. Duranay, A. Sengur, and U. R. Acharya, "Use of differential entropy for automated emotion recognition in a virtual reality environment with EEG signals," *Diagnostics*, vol. 12, no. 10, p. 2508, Oct. 2022.
[106] "Emotion detection method and apparatus, electronic device, and storage medium," Patent U.S. 20 200 256 973 A1, Aug. 13, 2020.
[107] S. Meudt and F. Schwenker, "Enhanced autocorrelation in real world emotion recognition," in *Proc. 16th Int. Conf. Multimodal Interact.*, Nov. 2014, pp. 502–507.
[108] F. Dehghani and L. Zaman, "Facial emotion recognition in VR games," in *Proc. IEEE Conf. Games (CoG)*, Aug. 2023, pp. 1–4.
[109] T. Xu, R. Yin, L. Shu, and X. Xu, "Emotion recognition using frontal EEG in VR affective scenes," in *IEEE MTT-S Int. Microw. Symp. Dig.*, vol. 1, May 2019, pp. 1–4.
[110] P. E. Antoniou, G. Arfaras, N. Pandria, A. Athanasiou, G. Ntakakis, E. Babatsikos, V. Nigdelis, and P. Bamidis, "Biosensor real-time affective analytics in virtual and mixed reality medical education serious games: Cohort study," *JMIR Serious Games*, vol. 8, no. 3, Sep. 2020, Art. no. e17823.
[111] S. A. B. Sierra and C. C. F. Iriarte, "Etiquetado de emociones a partir de videos en ambientes de formación virtual," Universidad Nacional de Colombia, Colombia, 2022, doi: 10.11144/Javeriana.10554.63120.
[112] G. Pei, C. Fan, T. Li, J. Jin, and R. Wang, "EEG-based emotion recognition in immersive virtual reality: Meeting the requirement of accuracy and computational efficiency," in *Proc. 7th Int. Conf. Digit. Signal Process.*, 2023, pp. 8–15.
[113] S. Thejaswini and K. Mamatha, "Machine learning algorithm to detect EEG based emotion states using virtual-video stimuli," in *Proc. Int. Conf. Adv. Electron., Commun., Comput. Intell. Inf. Syst. (ICAECIS)*, Apr. 2023, pp. 151–156.
[114] A. F. Bulagang, J. Mountstephens, and J. Teo, "Investigating electrodermography (EDG) and heart rate (HR) signals for emotion classification in virtual reality," in *Proc. Int. Visualizat., Informat. Technol. Conf. (IVIT)*, 2022, pp. 29–34.
[115] Y. Ding, N. Robinson, Q. Zeng, D. Chen, A. A. Phyo Wai, T.-S. Lee, and C. Guan, "TSception: A deep learning framework for emotion detection using EEG," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2020, pp. 1–7.
[116] J. Blitzer, M. Dredze, and F. C. Pereira, "Biographies, bollywood, boomboxes and blenders: Domain adaptation for sentiment classification," in *Proc. ACL*, 2007, pp. 440–447.
[117] A. Maas, R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng, and C. Potts, "Learning word vectors for sentiment analysis," in *Proc. 49th Annu. Meeting Assoc. Comput. Linguistics, Hum. Lang. Technol.*, 2011, pp. 142–150.
[118] R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Y. Ng, and C. Potts, "Recursive deep models for semantic compositionality over a sentiment treebank," in *Proc. EMNLP*, vol. 1631, Jan. 2013, pp. 1631–1642.
[119] T. Bänziger and K. R. Scherer, "Introducing the Geneva multimodal emotion portrayal (GEMEP) corpus," in *Blueprint for Affective Computing: A Sourcebook*, vol. 2010. Oxford, U.K.: Oxford Univ. Press, 2010, pp. 94–271.
[120] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended Cohn-Kanade dataset (Ck+): A complete dataset for action unit and emotion-specified expression," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.-Workshops*, 2010, pp. 94–101.
[121] T. Kanade, J. F. Cohn, and Y. Tian, "Comprehensive database for facial expression analysis," in *Proc. 4th IEEE Int. Conf. Autom. Face Gesture Recognit.*, 2000, pp. 46–53.
[122] M. Valstar and M. Pantic, "Induced disgust, happiness and surprise: An addition to the mmi facial expression database," in *Proc. 3rd Intern. Workshop Emotion Corpora Res. Emotion Affect (Satellite LREC)*, 2010, p. 65.
[123] I. J. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, and D.-H. Lee, "Challenges in representation learning: A report on three machine learning contests," in *Proc. 20th Int. Conf. (ICONIP)*, Daegu, South Korea, 2013, pp. 117–124.
[124] C. F. Benitez-Quiroz, R. Srinivasan, and A. M. Martinez, "EmotioNet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 5562–5570.
[125] Z. Zhang, P. Luo, C. C. Loy, and X. Tang, "From facial expression recognition to interpersonal relation prediction," *Int. J. Comput. Vis.*, vol. 126, no. 5, pp. 550–569, May 2018.
[126] A. Mollahosseini, B. Hasani, and M. H. Mahoor, "AffectNet: A database for facial expression, valence, and arousal computing in the wild," *IEEE Trans. Affect. Comput.*, vol. 10, no. 1, pp. 18–31, Jan. 2019.
[127] S. Li, W. Deng, and J. Du, "Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2852–2861.
[128] X. Jiang, Y. Zong, W. Zheng, C. Tang, W. Xia, C. Lu, and J. Liu, "DFew: A large-scale database for recognizing dynamic facial expressions in the wild," in *Proc. 28th ACM Int. Conf. Multimedia*, 2020, pp. 2881–2889.
[129] L. Yin, X. Wei, Y. Sun, J. Wang, and M. J. Rosato, "A 3D facial expression database for facial behavior research," in *Proc. 7th Int. Conf. Autom. Face Gesture Recognit. (FGR)*, 2006, pp. 211–216.
[130] G. Zhao, X. Huang, M. Taini, S. Z. Li, and M. Pietikäinen, "Facial expression recognition from near-infrared videos," *Image Vis. Comput.*, vol. 29, no. 9, pp. 607–619, Aug. 2011.
[131] A. Dhall, R. Goecke, S. Lucey, and T. Gedeon, "Static facial expression analysis in tough conditions: Data, evaluation protocol and benchmark," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCV Workshops)*, Nov. 2011, pp. 2106–2112.
[132] W.-J. Yan, X. Li, S.-J. Wang, G. Zhao, Y.-J. Liu, Y.-H. Chen, and X. Fu, "CASME II: An improved spontaneous micro-expression database and the baseline evaluation," *PLoS ONE*, vol. 9, no. 1, Jan. 2014, Art. no. e86041.
[133] X. Li, T. Pfister, X. Huang, G. Zhao, and M. Pietikäinen, "A spontaneous micro-expression database: Inducement, collection and baseline," in *Proc. 10th IEEE Int. Conf. Workshops Autom. Face Gesture Recognit. (FG)*, Apr. 2013, pp. 1–6.

[134] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A multimodal database for affect recognition and implicit tagging," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 42–55, Jan. 2012.
[135] S. Wang, Z. Liu, S. Lv, Y. Lv, G. Wu, P. Peng, F. Chen, and X. Wang, "A natural visible and infrared facial expression database for expression recognition and emotion inference," *IEEE Trans. Multimedia*, vol. 12, no. 7, pp. 682–691, Nov. 2010.
[136] H. Nguyen, N. Tran, H. D. Nguyen, L. Nguyen, and K. Kotani, "KTFEv2: Multimodal facial emotion database and its analysis," *IEEE Access*, vol. 11, pp. 17811–17822, 2023.
[137] S. Wang, X. Li, F. Zheng, J. Pan, X. Li, Y. Chang, Z. Zhu, Q. Li, J. Wang, and Y. Xiao, "VAD: A video affective dataset with danmu," *IEEE Trans. Affect. Comput.*, pp. 1–17, 2024.
[138] G. Xiang, S. Yao, H. Deng, X. Wu, X. Wang, Q. Xu, T. Yu, K. Wang, and Y. Peng, "A multi-modal driver emotion dataset and study: Including facial expressions and synchronized physiological signals," *Eng. Appl. Artif. Intell.*, vol. 130, Apr. 2024, Art. no. 107772.
[139] S. Cheng, I. Kotsia, M. Pantic, and S. Zafeiriou, "4DFAB: A large scale 4D database for facial expression analysis and biometric applications," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 5117–5126.
[140] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with Gabor wavelets," in *Proc. 3rd IEEE Int. Conf. Autom. Face Gesture Recognit.*, 1998, pp. 200–205.
[141] A. Pak and P. Paroubek, "Text representation using dependency tree subgraphs for sentiment analysis," in *Proc. 16th Int. Conf., DASFAA Int. Workshops: GDB, SIM3, FlashDB, SNSMW, DaMEN, DQIS*, Hong Kong, 2011, pp. 323–332.
[142] L. Kerkeni, Y. Serrestou, K. Raoof, M. Mbarki, M. A. Mahjoub, and C. Cleder, "Automatic speech emotion recognition using an optimal combination of features based on EMD-TKEO," *Speech Commun.*, vol. 114, pp. 22–35, Nov. 2019.
[143] R. Santhoshkumar and M. K. Geetha, "Human emotion recognition using body expressive feature," in *Microservices in Big Data Analytics*. Singapore: Springer, 2020, pp. 141–149.
[144] S. Z. Bong, M. Murugappan, and S. Yaacob, "Analysis of electrocardiogram (ECG) signals for human emotional stress classification," in *Proc. Trends Intell. Robot., Automat., Manuf., 1st Int. Conf.*, Kuala Lumpur, Malaysia, 2012, pp. 198–205.
[145] C.-H. Wu and W.-B. Liang, "Emotion recognition of affective speech based on multiple classifiers using acoustic-prosodic information and semantic labels," *IEEE Trans. Affect. Comput.*, vol. 2, no. 1, pp. 10–21, Jan. 2011.
[146] H. H. Do, P. Prasad, A. Maag, and A. Alsadoon, "Deep learning for aspect-based sentiment analysis: A comparative review," *Expert Syst. Appl.*, vol. 118, pp. 272–299, Mar. 2019.
[147] Z. Shen, J. Cheng, X. Hu, and Q. Dong, "Emotion recognition based on multi-view body gestures," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2019, pp. 3317–3321.

SITARA AFZAL received the Bachelor of Science (B.S.) degree in computer science from COMSATS University Islamabad, in 2017, and the Master of Science (M.S.) degree in computer science, in 2019. She is currently pursuing the Ph.D. degree in software convergence with Sejong University, Seoul. Her research interests include data mining, image processing, computer vision, and mixed reality, with a focus on pushing the boundaries of technology, she aim to contribute innovative solutions to real-world problems while advancing the field of computer science through academic excellence and research endeavors.

HASEEB ALI KHAN received the bachelor's degree in electrical engineering from COMSATS University Islamabad, Pakistan, in 2019. He is currently pursuing the master's degree in software convergence with Sejong University, Seoul, South Korea, focusing on AR, virtual reality, mixed reality, and machine learning. His research interests include bridge the gap between virtual and real-world experiences and leveraging the power of artificial intelligence.

MD JALIL PIRAN (Senior Member, IEEE) received the Ph.D. degree in electronics and information engineering from Kyung Hee University, South Korea, in 2016.

He holds a distinguished academic background. He is currently an Associate Professor with the Department of Computer Science and Engineering, Sejong University, Seoul, South Korea. He was a Postdoctoral Fellow with the Networking Laboratory, Kyung Hee University. He has made significant contributions to the field of intelligent information and communication technology (IICT) through his extensive research publications in esteemed international journals and conferences. His areas of expertise encompass machine learning, data science, wireless communications and networking, 5G/6G, the Internet of Things (IoT), and cyber security. In addition to his research endeavors, he actively engages with scholarly journals as an Editor, including IEEE Transactions on Intelligent Transportation Systems, *Journal of Engineering Applications of Artificial Intelligence* (Elsevier), *Journal of Physical Communication* (Elsevier), and *Journal of Computer Communication* (Elsevier). He also serves as the Secretary for the IEEE Consumer Technology Society on Machine Learning, Deep Learning, and AI. Furthermore, he assumes the role of the Track Chair of Machine Learning, Deep Learning, and AI in the CE (MDA) Track for the upcoming 2024 IEEE International Conference on Consumer Electronics (ICCE). In 2022, he chaired the "5G and Beyond Communication" Session at the prestigious IEEE International Conference on Communications (ICC). His expertise as a reviewer extends to leading journals, and he actively participates in various conferences. He represents South Korea as an Active Delegate to the Moving Picture Experts Group (MPEG). His outstanding research contributions have been recognized internationally, as evidenced by the prestigious "Scientist Medal of the Year 2017" awarded by IAAM in Stockholm, Sweden. Moreover, he received accolades from Iranian Ministry of Science, Technology, and Research as an "Outstanding Emerging Researcher," in 2017. His exceptional Ph.D. dissertation was honored as the "Dissertation of the Year 2016" by the Iranian Academic Centre for Education, Culture, and Research in the Engineering Group.

JONG WEON LEE received the M.S. degree in electrical and computer engineering from the University of Wisconsin Madison, in 1991, and the Ph.D. degree from the University of Southern California, in 2002. He is currently a Professor with the Department of Software, Sejong University. His research interests include AR, computer vision, machine learning, human–computer interaction, and virtual reality.