International Journal of Education, Learning and Development, 14 (2),1-10, 2026
Print ISSN: 2054-6297(Print)
Online ISSN: 2054-6300 (Online)
Website: https://www.eajournals.org/
Publication of the European Centre for Research Training and Development-UK

A Review of AI Techniques for Emotion Recognition in Communication with Applications in Child Safety and Education

Anxhela Kosta [1], Ilma Lili [1], Arber Ceni [1], Ana Ktona [1], Endrit Xhina [1], Alda Kika [1], Emiranda Loka [1]

[1]University of Tirana, Faculty of Natural Sciences, Department of Informatics, Bulevardi "Zogu I ", Tiranë, Albania
Email: anxhela.kosta@fshn.edu.al

doi: https://doi.org/10.37745/ijeld.2013/vol14n2110 Published April 08, 2026

Citation: Kosta A., Lili I., Ceni A., Ktona A., Xhina E., Kika A., Loka E. (2026) A Review of AI Techniques for Emotion Recognition in Communication with Applications in Child Safety and Education, *International Journal of Education, Learning and Development*, 14 (2),1-10

Abstract: *Emotions have a major role to play in the development of human communication, especially in the digital world, where children increasingly communicate through texts and voices. The development of artificial intelligence (AI) to automatically recognize human emotions has tremendous potential to contribute to the safety, education, and mental well-being of children. This paper aims to provide a comprehensive review of various artificial intelligence-based emotion recognition techniques for both textual and vocal communications. The paper discusses various techniques such as Natural Language Processing (NLP) and Speech Emotion Recognition (SER), as well as more advanced techniques such as BERT, RoBERTa, convolutional neural networks (CNN), and self-supervised approaches such as wav2vec. The paper discusses the advantages of developing systems that can integrate both textual and vocal communications to develop a more comprehensive system of emotion recognition. Apart from the development of such systems, the paper also discusses the potential applications of such systems in the development of parental monitoring systems, educational support systems, and mental well-being assessment systems for children. Finally, it outlines future directions for the development of such systems with a focus on benefiting children's well-being.*

Keywords: emotion recognition, artificial intelligence, speech emotion recognition (SER), natural language processing (NLP), BERT, parental control

INTRODUCTION

How people currently communicate, express themselves, interpret, and form social ties is strongly influenced by their emotional state, as emotions play a key role in shaping relationships with others. Especially in recent years, we are increasingly surrounded by online communication, whether through text messages, voice chats, and similar platforms. This development offers many opportunities for social interaction and learning, but it also introduces challenges such as stress, emotional isolation, and cyber threats (Praveena et al., 2020).

In the effort to better understand the emotional state of children, and above all what they feel, increasing research has been conducted, supported by artificial intelligence (AI). A number of AI-based methods for emotion recognition have achieved significant progress, including approaches in Natural Language Processing (NLP) for textual data and Speech Emotion Recognition (SER) for audio signals (Khare et al., 2023). Models such as BERT and RoBERTa, along with convolutional neural networks (CNNs) and self-supervised models like wav2vec, have been widely used to detect subtle emotional cues in communication (Devlin et al., 2019; Liu et al., 2019).

In contrast, multimodal systems, which combine text and audio, provide a more comprehensive understanding of emotional expression (Narimisaei et al., 2024). In this paper, we systematically review and analyze existing academic work to investigate the theory, methodologies, and applications of AI for emotion recognition, with a particular focus on child protection and education. Through reviewing the literature, we explore how these methods can be adapted to assist parents, teachers, and healthcare professionals in recognizing emotional cues, while also addressing ethical concerns and limitations (Barker et al., 2025). This paper emphasizes both the potential and the responsibility associated with using AI-based emotion recognition in child-related contexts.

Artificial intelligence systems that detect, interpret, and respond to human emotions have advanced significantly through deep learning approaches applied to both written and spoken communication. Higher performance has been achieved when deep learning models are applied to categorical and dimensional emotion representations, often reaching accuracy levels above 60% depending on the dataset and task (Khare et al., 2023).

THEORETICAL BACKGROUND OF EMOTION RECOGNITION

Emotion recognition is an interdisciplinary field that integrates concepts from psychology, linguistics, and, more recently, computer science. In essence, it focuses on identifying human emotions from observable signals such as speech, text, body language, and facial expressions. The main goal is to develop systems that not only recognize emotions but can also be applied constructively in social, educational, and healthcare contexts.

Early research in this field was strongly influenced by psychological theories of emotion. One of the most well-known frameworks is Ekman's model of six basic emotions: happiness, sadness, anger, fear, disgust, and surprise, which are considered universal in human expression (Ekman, 1992). Dimensional models such as the valence–arousal framework suggest that emotions should not only be viewed as discrete categories, but also as continuous variations in intensity and polarity (Praveena et al., 2020). These theories laid the foundation for early work in affective computing and enabled the integration of human sciences with artificial intelligence.

Since the emergence of affective computing in the late 1990s, researchers have developed systems capable of detecting and responding to emotional signals (Picard, 1997). Early approaches relied on handcrafted features, such as speech rhythm, intonation, and lexical sentiment indicators, combined with traditional machine learning classifiers like Support Vector Machines (SVM) and Naive Bayes.

However, these methods were limited in their ability to capture complex language patterns and subtle emotional variations. As illustrated in Figure 1, emotion recognition systems operate as a cycle that includes emotion elicitation, detection, and system response.

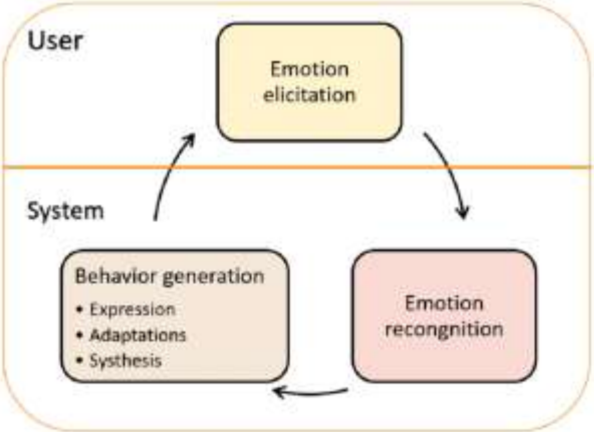

Figure 1. General framework of emotion elicitation and recognition adapted from Somarathna et al. (2021).

Similarly, in speech processing, Speech Emotion Recognition (SER) initially focused on prosodic features such as pitch, energy, and tempo. These features were analyzed using conventional classifiers. With the advancement of deep learning, models such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory networks (LSTMs) enabled automatic feature extraction directly from audio spectrograms, significantly improving performance (Khare et al., 2023).

More recently, self-supervised learning approaches, including models such as wav2vec 2.0 and HuBERT, have further advanced the field. These models are trained on large amounts of unlabeled data and then fine-tuned for specific emotion recognition tasks, achieving state-of-the-art results on conversational datasets (Alhussein et al., 2025).

Despite these advancements, emotion recognition remains a complex and subjective problem. A single sentence or tone of voice may be interpreted differently depending on factors such as age, culture, language, and personal experience. This challenge is particularly significant in applications involving children, whose emotional expressions are often more variable and less standardized than those of adults. As highlighted by Barker et al. (2025), ethical concerns such as subjectivity, bias, privacy, and consent must be considered as fundamental aspects of the field.

Overall, emotion recognition has evolved from classical psychological models to advanced deep learning and self-supervised techniques, enabling more accurate detection of emotions in both text and speech. However, important challenges remain, particularly in developing child-centered, bias-aware, and privacy-preserving systems suitable for real-world applications.

EMOTION RECOGNITION TECHNIQUES

Text-Based Approaches (NLP)

Emotion detection from text involves analyzing linguistic signals to infer the emotional state of the user. This area has evolved alongside the development of Natural Language Processing (NLP). Early methods relied on simple representations such as bag-of-words and TF-IDF, combined with classifiers like Support Vector Machines (SVM) and Naive Bayes. While these approaches were efficient, they often failed to capture semantic nuances, irony, contextual ambiguity, and long-range dependencies in language (Praveena et al., 2020).

With the advancement of deep learning, models such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) were introduced to automatically extract more complex emotional features from text. On the other hand, these architectures still had limitations when dealing with long contextual relationships. A major breakthrough came with attention-based models such as BERT and RoBERTa, which generate contextual embeddings and significantly improve performance in capturing subtle emotional meanings (Devlin et al., 2019; Liu et al., 2019; Narimisaei et al., 2024).
More recently, instruction-tuned large language models (LLMs) have been explored for emotion-related tasks. These models show promising results in few-shot and zero-shot settings, but they also introduce challenges such as overgeneration, hallucination, and limited reliability in sensitive contexts like children's communication (Khare et al., 2023).

Speech-Based Approaches (SER)
Speech signals contain rich acoustic and prosodic features that reflect the emotional state of the speaker. Traditional approaches focused on handcrafted features such as Mel-Frequency Cepstral Coefficients (MFCCs), pitch, intensity, and energy, which were then used with classifiers like SVM or Random Forest (Praveena et al., 2020). As shown in Figure 2, SER systems typically involve preprocessing, feature extraction, and classification stages for detecting emotional states from speech signals (Kakuba et al., 2022).

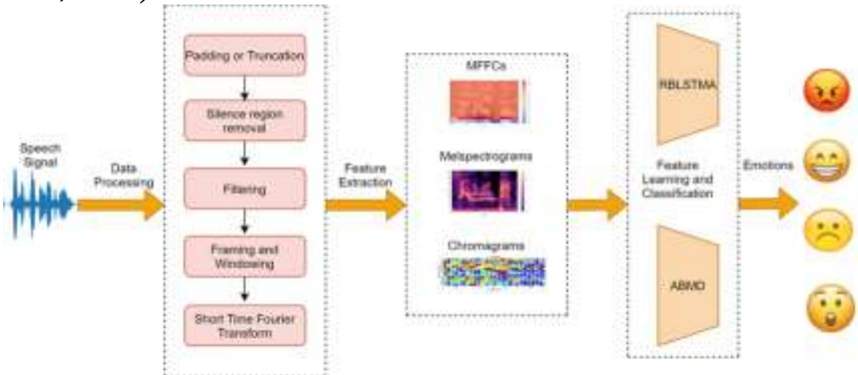

Figure 2. Overview of a Speech Emotion Recognition (SER) system, including preprocessing, feature extraction, and classification (adapted from Kakuba et al., 2022).

With the rise of deep learning, CNN and LSTM-based models enabled automatic feature extraction from audio spectrograms, leading to significant improvements in emotion recognition performance (Khare et al., 2023).

More recently, self-supervised learning has transformed this field. Models such as wav2vec 2.0 and HuBERT are trained on large amounts of unlabeled speech data and then fine-tuned for emotion recognition tasks. These approaches have achieved state-of-the-art results on conversational datasets and represent a promising direction for future research (Alhussein et al., 2025).

However, recognizing emotions in children's speech remains particularly challenging. Variations in tone, speaking style, and vocal development make it difficult for models trained on adult data to generalize effectively.

Multimodal Fusion

Since emotions are not expressed through a single modality, combining text and speech provides a more comprehensive understanding of emotional states. Multimodal approaches aim to integrate information from multiple sources to improve accuracy and robustness.

Early fusion techniques combine features from text and audio before classification, while late fusion methods merge predictions from separate models. More advanced approaches, such as attention-based or hybrid fusion, dynamically weigh different modalities depending on their relevance in a given context (Khare et al., 2023).

Multimodal systems are particularly important in children's communication, where emotional expression may vary significantly across different channels. As highlighted by Narimisaei et al. (2024), these approaches can reduce ambiguity and improve the reliability of applications in sensitive domains such as education and well-being monitoring.

DATA SETS AND EVALUATION METRICS

Commonly used datasets

The work on the development of emotion recognition systems has been greatly facilitated by standardized datasets that provide a benchmark for comparative purposes. Most of these corpora are collected in adults and in laboratory conditions, which hinders their applicability to children and natural contexts (Livingstone & Russo, 2018).

- IEMOCAP (Interactive Emotional Dyadic Motion Capture): One of the most widely used datasets comprising audio, video, and transcripts of both improvised and scripted conversations annotated for categorical and dimensional emotions (Busso et al., 2008).
- RAVDESS (Ryerson Audio-Visual Database of Emotional Speech and Song): Features voice recordings from professional actors, reading monologues and speaking in a neutral tone as well as singing. The quality of the recordings and labeling for eight emotions render the RAVDESS dataset as a benchmark dataset in SER and multimodality (Livingstone & Russo, 2018).
- CREMA-D (Crowd-sourced Emotional Multimodal Actors Database): It has 91 actors with wide ranges of different ethnicities and ages, toward a more diverse source for SER comparison. The articles have been rated by many raters to assure reliability (Cao et al., 2014).
- EMO-DB (Berlin Emotional Speech Database): A historical German dataset of basic emotions produced by actors. Commonly used for early method comparisons but is not applicable for widespread multilingual use (Burkhardt et al., 2005).
- MSP-Improv: Short for MSP-Improv, includes time-coded multi-modal dialogues from improv and neural sketches, a collection of naturalistic dyadic conversation data, providing more spontaneous and less predictable environments that are more similar to naturally occurring conversations (Martínez et al., 2020).

- DAIC-WOZ (Distress Analysis Interview Corpus – Wizard-of-Oz): Comprises structured interviews to recognize instances of anxiety and depression. This dataset is especially related to mental well-being (Gratch et al., 2014).
- SEMAINE: Multimodal dataset reflecting extended interactions between users and virtual agents that can be used to explore the dynamics of emotions (McKeown et al., 2012).

It is important for understanding additional funders' contributions, although relatively few resources for most of these do not cater to children. While this grouping has proponents (Barker et al., 2025), obtaining emotional data in this group will necessitate new ways to manage consent, privacy, and ethics, which is an area of critical need for future investigation.

Evaluation Metrics

Evaluating emotion recognition systems is as challenging as developing them, and this comes from the sensitive factors of class imbalance, label subjectivity, and different nature of expressed emotions. There is a need to employ discharge and beyond discharge levels to address this issue.

- Model Accuracy: Typically the most common measure, though often misleading in imbalanced datasets where classes of "neutral" or "positive" drastically outnumber the "negative" class.
- F1-Score: A trade-off between precision and recall, thus better suitable for multi-class classification (Sokolova & Lapalme, 2009).
- Weighted Accuracy (WA): Uses as well the number of instances of each class in the computation, which offers a fairer vision of unbalanced datasets.
- UAR (Unweighted Average Recall): All classes receive equal weight even if data distribution is unbalanced; this metric has been acknowledged as one of the fairest measures for SER evaluation (Schuller et al., 2011).
- Cross-dataset cross-validation: This is an increasingly recommended approach to test model generalization and overfitting in a single dataset (Stuhlsatz et al., 2011).

For applications involving children, evaluation should go beyond these traditional metrics. Factors such as fairness, cultural diversity, and age differences need to be considered, since emotional expression varies depending on context, language, and developmental stage (Narimisaei et al., 2024).

APPLICATIONS IN CHILD SAFETY AND EDUCATION

Emotion recognition using artificial intelligence is becoming increasingly relevant in digital safety, educational support, and the psychological well-being of children. While much of the existing research focuses on technical model development, several studies emphasize the importance of evaluating these technologies from educational, psychological, and ethical perspectives (Praveena et al., 2020; Narimisaei et al., 2024). These applications can generally be grouped into two main areas: monitoring and protection in digital environments, and educational and psychological support.

Monitoring and Protecting Children in Digital Environments

Parental control and online harm prevention are among the leading areas of application of emotion recognition. Since children are more likely to suffer from cyberbullying, peer pressure, and other types of digital stress, NLP and SER models can be deployed in parental monitoring tools to detect emotional distress early. For instance, NLP can use chat messages to identify patterns of hostile language, while SER can determine when a voice call indicates fear or stress (Khare et al., 2023). Systematic reviews in the literature warn that these should not be viewed as a substitute for parental judgment but rather as supportive devices. There is the danger of false positives or misinterpretations that, in turn, can lead to over-monitoring, which undermines children's autonomy and privacy (Barker et al., 2025).

Educational Support and Psychological Well-Being

A second critical dimension is the exploitation of emotion recognition technologies to enhance education and the welfare of children at school. Teachers in educational environments are challenged to detect the instantaneous emotional status of students. Integrated SER systems can learn to detect when students are bored, frustrated, or anxious during learning, and this can inform teachers to change pedagogy in real time (Khare et al., 2023).

In online learning environments, text analysis of quizzes and homework can be directly linked with audio signal processing analysis from virtual interactions that can form a more complete multimodal profile of student engagement and emotions (Narimisaei et al., 2024). Outside of the classroom, the uses can reach even into psychological care. School counselors and mental health professionals may use systems that follow long-term emotional trends based on data children voluntarily share. Such systems have the potential to detect early stages of psychological disorders like depression, anxiety, and social isolation, thus may intervene early on to provide personalized support (Narimisaei et al., 2024). However, limitations remain numerous. The absence of databases specifically tailored to children is one of the biggest challenges, since current models are trained using adult data and therefore lack reliability. Also, fear of similar judgment occurs based on cultural, linguistic, and age biases, which increase misclassification and thus require more diversity-sensitive models (Barker et al., 2025). Ultimately, the ethical issues of children's privacy and consent remain unresolved and need to be considered at every stage of the application in schools and other educational settings.

ETHICAL CONSIDERATIONS AND CHALLENGES

The use of AI-based emotion recognition systems in activities with children gives rise to several ethical and social dilemmas. The technology has the capacity to support online protection, learning, mental health care (to name a few), yet it is being warned against for its potential to infringe on privacy, bias outcomes, and misuse (Barker et al., 2025). Two decades into its usage, these agonizing paradoxes have broken researchers into two camps: those who are optimistic about the technology's transformative potential, and those who worry that by not pairing it with clear ethical safeguards, we are setting ourselves up to play out such dystopian dramas in the real world.

Privacy, Justice and the Risk of Abuse

Emotional information is thought to be one of the most privacy-sensitive types of information a computer system can gather, as it can reveal the internal psychological state of the individual. In the context of children, any kind of data collection from text messages, voicemails, or online interactions should involve informed consent from parents and secure the safety of the data for its storage, usage, and deletion (Praveena et al., 2020). Persisting or surreptitious surveillance, particularly when consent is not obtained, is commonly considered to be unethical, and it risks trust between children and adults who care for them.

A third problem is biases in the data with which models are trained. The majority of the emotion corpora are based on adults, quite often from a specific culture, making the realized models oblivious to potential differences in language, age, or culture, as in the case of children (Khare et al., 2023). This introduces the potential for misclassification, which threatens to perpetuate, rather than remediate, existing inequities.

It is the abuse of technology that is the greatest threat. If applied to education and home-based situations, experts fear the use of emotion recognition technologies can turn into an instrument of over-surveillance, a climate of distrust, the disempowerment of children, and the labeling of childhood emotions (Barker et al., 2025). For this reason, writing on these systems stresses the idea that these systems are devices that are supportive, not instruments of discipline or control.

Transparency, Accountability and Safeguards

In order to make sure usage is fair and ethical, it is critical for emotion recognition systems to be explainable and transparent. Parents, teachers, and the children themselves must understand how a system arrives at a particular decision. As a consequence, "black-box" models are less interpretable, which can diminish the accountability of the application and may result in trust issues toward the technology (Narimisaei et al., 2024). This renders the need to integrate Explainable AI (XAI), particularly in high-risk scenarios such as education and the safety of children.

Another ethical foundation is baking in protections to the technology that will preclude it from being used improperly. Researchers suggest several key mechanisms:

- Minimization of data and use of on-device processing to minimize the amount of sensitive information exposed.
- Differential privacy methods to preserve private attributes.
- Unbiased Overview: third-party agencies monitoring usage and constraining abuse (Alhussein et al., 2025).
- Introduction of a "Do Not Use" list that will prevent these systems from being used for activities, such as disciplinary decisions, ongoing covert monitoring, and commercial profiling of children.

Together, the core principles of transparency, accountability, and technical safeguards offer a basis for ethics that will help to ensure the use of technology is for the benefit of children's safety, development, and well-being.

FUTURE DIRECTIONS

Although AI-based emotion recognition has seen major breakthroughs, its development for child protection and education is still in its early stages. Several challenges and research gaps remain, which must be addressed to ensure that these systems are trustworthy, ethical, and effective in sensitive environments.

One of the most significant challenges is the lack of datasets specifically designed for children. Most existing datasets, such as IEMOCAP, RAVDESS, and CREMA-D, are based on adult data collected in controlled environments, which do not fully capture the complexity of children's emotional expressions. Future research should therefore focus on developing datasets that are ethically collected and representative of different ages, languages, and cultural contexts. Such efforts will improve model reliability and reduce bias, leading to more inclusive systems (Khare et al., 2023; Barker et al., 2025).

Another important direction is privacy-preserving technology. Since emotional data reflects internal psychological states, it is highly sensitive. Techniques such as on-device processing, federated learning, and differential privacy can help minimize risks by reducing the need to store or transmit personal data (Alhussein et al., 2025).

Ensuring fairness and inclusivity is also essential. Models trained on limited or homogeneous datasets may not perform well in multilingual or multicultural environments. Future work should focus on cross-linguistic learning and fairness-aware approaches to ensure that systems work equally well across different populations (Narimisaei et al., 2024).

Explainability and human involvement will play a key role in the future of emotion recognition systems. In sensitive environments such as education and home settings, it is not enough for systems to provide predictions—they must also explain how those decisions are made. Explainable AI (XAI) can help users better understand model outputs, while human-in-the-loop approaches ensure that final decisions remain under human control (Praveena et al., 2020; Barker et al., 2025).

Finally, the future of emotion recognition lies in multimodal and context-aware systems. By combining multiple sources of information—such as text, speech, and interaction context—these systems can achieve a more accurate and comprehensive understanding of emotions (Khare et al., 2023). In conclusion, advancing emotion recognition for children will require improved datasets, stronger privacy protections, fair and inclusive models, greater transparency, and more sophisticated multimodal systems. Addressing these challenges will be essential for developing AI tools that truly support children's safety, education, and well-being.

CONCLUSION AND RECOMMENDATION

Emotion recognition has emerged as a central concern at the intersection of artificial intelligence, communication, and child development. This paper has provided an overview of the theoretical foundations of the field, the transition from classical feature-based methods to deep learning and self-supervised approaches, and the rise of multimodal systems. It has also highlighted the importance of datasets and evaluation metrics, while identifying the ongoing lack of child-specific data.

Applications of AI-based emotion recognition show strong potential in areas such as parental control, classroom support, and psychological development monitoring. However, as the literature suggests, these opportunities must be carefully balanced with risks such as bias, over-monitoring, and unethical use. Ethical implementation requires privacy-preserving techniques, explainable models, and the continued involvement of human judgment to ensure that AI serves as a supportive tool rather than a controlling mechanism.

Looking ahead, progress in this field will depend on the development of child-centered datasets, fairness-aware models, and explainable AI frameworks informed by interdisciplinary research. By addressing these challenges, emotion recognition systems can evolve into more inclusive, trustworthy, and ethically responsible tools that support children's safety, education, and overall well-being.

REFERENCES

Barker, D., Tippireddy, M. K. R., Farhan, A., & Ahmed, B. (2025). Ethical considerations in emotion recognition research. Psychology International, 7(2), 43. https://doi.org/10.3390/psycholint7020043

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. https://doi.org/10.48550/arXiv.1810.04805

Khare, S. K., Blanes-Vidal, V., Nadimi, E. S., & Acharya, U. R. (2023). Emotion recognition and artificial intelligence: A systematic review (2014–2023) and research recommendations. Information Fusion, 96, 102019. https://doi.org/10.1016/j.inffus.2023.102019

Liu, Y., et al. (2019). RoBERTa: A robustly optimized BERT pretraining approach. https://doi.org/10.48550/arXiv.1907.11692

Narimisaei, J., Naeim, M., Imannezhad, S., Samian, P., & Sobhani, M. (2024). Exploring emotional intelligence in artificial intelligence systems: A comprehensive analysis of emotion recognition and response mechanisms. Annals of Medicine & Surgery, 86(8), 4657–4663. https://doi.org/10.1097/MS9.0000000000002315

Praveena, K. B., Suresh, B., & Patrer, D. (2020). Emotion recognition with AI: Techniques and applications. World Journal of Advanced Research and Reviews, 8(2), 344–352. https://doi.org/10.30574/wjarr.2020.8.2.0362

Somarathna, R., Bednarz, T., & Mohammadi, G. (2021). Virtual reality for emotion elicitation: A review. Preprint https://doi.org/10.1109/TAFFC.2022.3181053

Kakuba, S., Poulose, A., & Han, D. S. (2022). Attention-based multi-learning approach for speech emotion recognition with dilated convolution. https://doi.org/10.1109/ACCESS.2022.3223705

Busso, C., Bulut, M., Lee, C. C., Kazemzadeh, A., Mower, E., Kim, S., Chang, J. N., Lee, S., & Narayanan, S. S. (2008). IEMOCAP: Interactive emotional dyadic motion capture database. Language Resources and Evaluation, 42(4), 335–359.https://doi.org/10.1007/s10579-008-9076-6

Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W. F., & Weiss, B. (2005). A database of German emotional speech. https://doi.org/10.21437/Interspeech.2005-446

Cao, H., Cooper, D. G., Keutmann, M. K., Gur, R. C., Nenkova, A., & Verma, R. (2014). CREMA-D: Crowd-sourced emotional multimodal actors dataset. IEEE Transactions on Affective Computing. https://doi.org/10.1109/TAFFC.2014.2336244

Ekman, P. (1992). An argument for basic emotions. Cognition and Emotion, 6(3–4), 169–200.

Gratch, J., et al. (2014). The distress analysis interview corpus of human and computer interviews.

Livingstone, S. R., & Russo, F. A. (2018). The Ryerson audio-visual database of emotional speech and song (RAVDESS). https://doi.org/10.1371/journal.pone.019639

Martínez, H. P., et al. (2020). MSP-Improv: An acted corpus of dyadic interactions to study emotion perception. https://doi.org/10.21437/Interspeech.2020-2444

Picard, R. W. (1997). Affective computing. MIT Press.

Schuller, B., et al. (2011). Recognizing realistic emotions and affect in speech: State of the art and lessons learnt. https://doi.org/10.1016/j.specom.2011.01.011

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. https://doi.org/10.1016/j.ipm.2009.03.002

Stuhlsatz, A., et al. (2011). Deep neural networks for acoustic emotion recognition. https://doi.org/10.1109/ICASSP.2011.5947651