International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 05 | May - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

A Comprehensive Review of Speech Emotion Recognition Systems

Dr. Rupinder Kaur, Assistant Professor, ADGIPS, er.rupinderkaur.cse@gmail.com

Devangi Bedi, B.Tech (IT), 4th Year, ADGIPS devangibedi2904@gmail.com
Akshita Panwar, B.Tech (IT), 4th Year, ADGIPS akshitapython@gmail.com
Aarav Bamba, B.Tech (IT), 4th Year, ADGIPS aaravbmb@gmail.com

ABSTRACT

Speech Emotion Recognition (SER) has emerged as a pivotal technology in the realm of Human-Computer Interaction (HCI) and sophisticated speech processing applications. By extracting and classifying key features from speech signals, SER systems aim to identify a speaker's emotional state. Despite significant progress, the fundamental differences in how humans and machines perceive and interpret emotional cues continue to pose challenges. These challenges often stem from the interdisciplinary nature of the field, which intersects speech processing, psychology, and user interface design. This review paper provides a structured and in-depth analysis of recent advancements in SER system architectures, methodologies, and feature extraction techniques. It consolidates key findings from the latest research to present a holistic understanding of the current landscape. Furthermore, the paper highlights existing research gaps and technical limitations, offering a foundation for future studies and the development of more robust and context-aware SER systems.

Keywords: Speech emotion recognition, database, pre-processing, feature extraction, classifier.

1. INTRODUCTION

We humans have a unique ability to convey ourselves through speech. These days alternative communication methods like text messages and emails are available. However, speech is still the most significant part of human culture and is data rich. Both paralinguistic and linguistic information is contained in the speech. Classical automatic speech recognition systems focused less on some of the essential paralinguistic information passed on by speech like gender, personality, emotion, aim, and state of mind [1]. The human mind utilizes all phonetic and paralinguistic data to comprehend the utterances' hidden importance and has efficacious correspondence [2]. The superiority of communication gets badly affected if there is any meagerness in the cognizance of paralinguistic features. There have been some arguments regarding children who cannot comprehend the speaker's emotional conditions evolve substandard social skills. In certain instances, they manifest psychopathological manifestations [3], which accentuates the significance of perceiving speech's emotional conditions leading to ineffective communication.

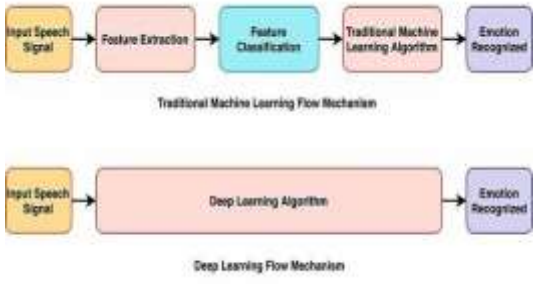

Figure1: Flowchart comparing traditional machine learning and deep learning approaches

Therefore, creating coherent and human-like communication machines that comprehend paralinguistic data, for example, emotion, is essential [4]. Emotion recognition has been the subject of exploration for quite a long time. The fundamental structure of research in emotion recognition was formed by detecting emotions from facial expressions [5]. Emotion recognition from speech signals has been studied to a great extent during recent times. In human-computer interaction, emotions play an essential role [6]. In recent times, speech emotion recognition (SER), which expects to investigate the emotion states through speech signals, has been drawing increasing consideration. Nevertheless, SER remains a challenging task, with the question of how to extract effective emotional features.

A classification of methodologies that process and at the same time characterize speech signals to identify emotions embedded in them is an SER system. An SER system needs a classifier, a supervised learning construct, programmed to perceive any emotions in new speech signals. [7]. A supervised system like that introduces the need for labeled data with emotions embedded in it. Before any processing can be done on the data to extract the features, it needs preprocessing. For this reason, the sampling rate across all the databases should be consistent. The classification process essentially requires features. They help reduce raw data into the most critical characteristics only, regardless of whether it suffices to utilize acoustic features for displaying emotions or if it is mandatory to cooperate with different kinds of features like linguistic, facial features, or speech information. Classifiers' performance can be said to depend mainly on the techniques of feature extraction and those features that are viewed as salient for a particular emotion [8]. If additional features can be consolidated from different modalities, for example, linguistic and visual, it can strengthen the classifiers. However, this relies on the significance and accessibility. These features are then permitted to pass to the classification system with a broad scope of classifiers at its disposal. All have been analyzed to classify emotions according to their acoustic correlation in speech utterances from numerous machine learning algorithms. Linear discriminant classifiers, Gaussian Mixture Models (GMM), Hidden Markov Models (HMM), k-nearest neighborhood (KNN) classifiers, Support Vector Machines (SVM), decision tree, and artificial neural networks (ANN) are a few models that have been generally used to classify emotions dependent on their acoustic features of intrigue [9]. In recent times, deep learning classifiers have

become common such as Deep Belief Networks, Deep Neural Network, Deep Boltzmann Machine, Convolution Neural Network, Recurrent Neural Network, and Long Short-Term Memory.

2. LITERATURE REVIEW:

Deep Learning Approaches to Speech Emotion Recognition: A Survey by Y. Zhang et al. (2021) This paper reviews deep learning methods applied to SER, including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and hybrid models. By evaluating these architectures, the study highlights that CNNs are effective for extracting spatial features, while RNNs capture temporal dynamics in audio signals. The authors concluded that hybrid CNN-RNN models improve performance across diverse emotional datasets by leveraging both spatial and temporal features. "Feature Extraction Techniques in Speech Emotion Recognition: An Analysis" by J. Liu and M. Wang (2019) This research focuses on comparing traditional and contemporary feature extraction methods in SER, such as Mel-frequency cepstral coefficients (MFCC), prosodic features, and deep feature representations. The study concluded that while traditional features perform well in constrained environments, deep learning methods that autonomously extract features yield higher accuracy in complex, real-world datasets. "The Role of Prosody in Emotion Detection from Speech: A Comparative Review" by R. Lee et al. (2018) This study evaluates the use of prosodic features pitch, rhythm, and loudness for emotion detection, comparing them with other acoustic features. Prosodic features were found to be highly effective for distinguishing between high-arousal and low-arousal emotions. The following table summarizes the literature review of researchers who have explored Speech Emotion Recognition system as follow:

No.	Title	Authors	Year	Focus	Key Findings
1	*Cross-Corpus Analys for Speech Emotio Recognition*	A. Smith al.	2022	Cross-corpus SE using domai adaptation and transf learning	Combining datasets with transf learning reduces overfitting an improves generalizability acro corpora
2	*A Survey on Multilingu Speech Emotio Recognition Systems*	H. Patel et a	2020	Multilingual SE systems and linguist variation impact	English-based models underperform i non-English corpora; recommen language-specific models an multilingual datasets
3	*Performance Evaluatio of Classical vs. Dee Learning Methods fo Speech Emotio Recognition*	M. Kumar, Verma	2019	Comparison classical ML and dee learning for SER	Deep learning models (especiall LSTM) outperform classical ones due t better handling of speech sequences
4	*Speech Emotio Recognition Using Dat Augmentation Techniques*	L. Nguyen al.	2021	Data augmentation SER	Techniques like pitch shifting and nois addition improve robustness an generalizability, especially for small imbalanced datasets
5	*Multi-Modal Emotio Recognition Using Audi and Textual Cues*	R. Das and Mehta	2021	Integration of speec and text features emotion recognition	Combining acoustic signals wit corresponding textual transcrip significantly improves emotio classification accuracy, especially i ambiguous speech segments
6	*Lightweight CN Architectures for Rea Time Speech Emotio Detection*	T. Zhao et a	2023	Development efficient CNN mode for SER	Proposed lightweight CNN mode achieve competitive accuracy wit reduced computational cost, enablin real-time deployment on mobile an embedded systems

Table 1. Literature Review

3. SPEECH PROCESSING:

The recorded audio signals contain the target speaker's speech and background noise, non-target speakers' voices, involves manipulating signals to change the signal's essential characteristics or extract vital information from it. Speech processing consists of the following steps:

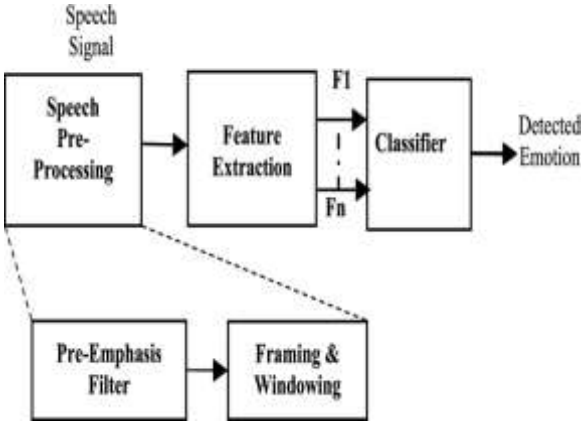

Figure 2: Block diagram of a traditional speech emotion recognition system

1) PREPROCESSING: The first step after collecting the data is preprocessing. The collected data would be utilized to prepare the classifier in an SER system. While few of these preprocessing procedures are utilized for feature extraction, others take care of the normalization of the features so that the variations in the recordings of the speakers do not affect the recognition process [17].

2) FRAMING: The next step is known as signal framing. It is also alluded to as speech segmentation and is the way toward apportioning constant speech signals into fixed length sections to surpass a few SER difficulties. Emotions often tend to vary during a speech as a result of the signals being non-stationary. Despite this fact, the speech remains invariant even though it is for a very short period, such as 20 to 30 milliseconds. Speech signal, when framed, helps to estimate the semi-fixed and local features [33]. We can also retain the connection and data between the frames by intentionally covering 30% to 40% of these segments. The utilization of processing methods, for example, Discrete Fourier Transform (DFT) for feature extraction, SER can be controlled by persistent speech signals. Accordingly, fixed size frames are appropriate for classifiers, for example, ANNs, while holding the emotion data in speech.

3) WINDOWING: Once the framing in a speech signal is conducted, the frame is subject to the window function. During Fast Fourier Transform (FFT) of information, leakages occur due to discontinuities at the edge of the signals, henceforth reduced by the windowing function [34]. Generally, one of the sorts of the windowing function is Hamming window as defined in Eq. (1), $w(n) = 0.54 - 0.46 \cos 2\pi n\, M - 1$ (1) where the frame is w(n), the window size is M, and $0 \le n \le M - 1$.

4) VOICE ACTIVITY DETECTION: Three sections are included in utterance: unvoiced speech, voiced speech, and silence. If vocal cords play an active role in sound production, voiced speech is produced [10]. On the contrary, the speech is unvoiced if vocal cords are inactive. Voiced speech can be distinguished and extricated because of its periodic behavior. A voice activity detector could be used to detect voiced/unvoiced speech and silence in a speech signal.

5) NORMALIZATION: It is a methodology for adjusting the volume of sound to a standard level [17]. For normalization, the maximum value of the signal is obtained, and then the whole signal sequence is divided by the calculated maximum to estimate that every sentence has a similar level of volume. Z-normalization is generally used for normalization and is calculated as $z = x - \mu\ \sigma$ (2) where μ is the mean, and σ is the standard deviation of the given speech signal.

6) NOISE REDUCTION: The environment is full of noises, and these noises are also encapsulated with every speech signal. Critically, the accuracy will be affected by the presence of noise in the speech signal. Therefore, for reducing this noise, several noise reduction algorithms can be utilized, like minimum mean square error (MMSE) and log-spectral amplitude MMSE (LogMMSE) [35]. The crucial phases in emotion recognition are feature selection and dimension reduction. Speech consists of numerous emotions and features, and one cannot state with certainty which set of features must be modeled and thus making a requirement for the utilization of feature selection techniques [36]. It is essential to do as such to preclude that the classifiers are not confronted with the scourge of dimensionality, incremented training time, and over-fitting that profoundly influence the prediction rate.

4. SPEECH CLASSIFIERS:

For any utterance, the underlying emotions are classified using speech emotion recognition. Classification of SER can be carried out in two ways: (a) traditional classifiers and (b) deep learning classifiers. Numerous classifiers have been utilized for the SER system, but determining which works best is difficult. Therefore, the ongoing research is widely pragmatic. SER systems generally utilize several traditional classification algorithms. The learning algorithm predicts a new class input, which requires the labeled data that recognizes the respective classes and samples by approximating the mapping function. After the training process, the remaining data is utilized for testing the classifier performance. Examples of traditional classifiers include Gaussian Mixture Model, Hidden Markov Model, Artificial Neural Network, and Support Vector Machines. Some other traditional classification techniques involve k-Nearest Neighbor, Decision Trees, Naïve Bayes Classifiers, and k-means are preferred. Additionally, an ensemble technique is used for emotion recognition, which combines various classifiers to acquire more acceptable results.

GAUSSIAN MIXTURE MODEL (GMM) GMM is a probabilistic methodology that is a prodigious instance of consistent HMM, consisting of just one state. The main aim of using mixture models is to template the data in a mixture of various segments, where every segment has an elementary parametric structure, like a Gaussian. It is presumed that every information guide alludes toward one of the segments, and it is endeavored to infer the allocation for each portion freely. GMM was contemplated for determining the emotion classification on two different speech databases, English and Swedish. The outcome stipulated that GMM is an expedient method on the frame level. The two MFCC methods show similar performance, and MFCC low features outperformed the pitch features. A semi-natural database GEU-SNESC (GEU Semi Natural Emotion Speech Corpus), was proposed. Five emotions: happy,

sad, anger, surprise, and neutral, were considered for the classification using the GMM classifier. For the characterization of emotions, the linear prediction residual of the speech signal was incorporated. The recognition percentage was discerned to be 50–60%.

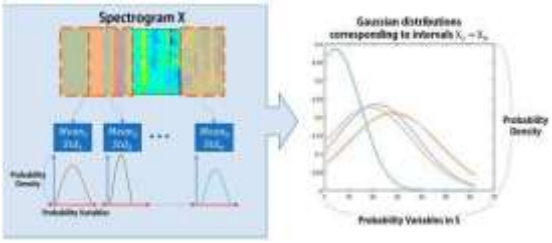

Figure 3: Statistical feature extraction from spectrogram using Gaussian distributions.

HIDDEN MARKOV MODEL (HMM) HMM is a usually utilized technique for recognizing speech and has been effectively expanded to perceive emotions. HMM is a statistical Markov model in which the system is assumed to be a Markov process with an unobserved state. The term ‘‘hidden’’ indicates the ineptitude of seeing the procedure that creates the state at an instant of time. It is then possible to use a likelihood to foresee the accompanying state by referencing the current situation’s target realities with the framework. In, the authors demonstrated that HMM performs better on log frequency power coefficient features than LPCC and MFCC. The emotion classification was done based on text-independent methods. They attained a recognition rate of 89.2% for emotion classification and human recognition of 65.8%. Hidden semi-continuous Markov models were utilized to construct a real-time multilingual speaker-independent emotion recognizer. A higher than 70% recognition rate was obtained for the six emotions comprising anger, sadness, fear, joy, happiness, and disgust. The INTERFACE emotional speech database was considered for the experiment.

SUPPORT VECTOR MACHINE (SVM) An SVM classifier is supervised and preferential. The classifier is generally described for linearly separable patterns by splitting hyperplanes. SVM makes use of the kernel trick to model nonlinear decision boundaries. The SVM classifier aims to detect that hyperplane having a maximum margin between two classes’ data points. The original data points are mapped to a new space if the given patterns are not linearly separable by utilizing a kernel function.

ARTIFICIAL NEURAL NETWORKS (ANN) ANNs have been typically used for several kinds of issues linked with classification. It essentially consists of an input layer, at least one hidden layer, and an output layer. Since the layers consist of several nodes, the nodes present in an input and output layer depend upon the characterization of labeled class and data, while a similar number of nodes can be present in the hidden layer as per the requirement. The weights are arbitrarily chosen and are related to each layer. The qualities of a picked sample from training data are staked to the information layer and later forwarded to the next layer. The backpropagation algorithm is used for updating the weights at the output layer. The weights are foreseen to be able to classify the new data once the training has finished. Two models are formulated to recognize emotions from speech based on ANN and SVM in [56], where the effect of feature dimensionality reduction to accuracy was evaluated. The features are extracted from CASIA Chinese Emotional Corpus. Initially, the ANN classifier showed 45.83% accuracy, but after the principal component analysis (PCA) over the features, ANN resulted in 75% improvement while SVM showed slightly better results, i.e., 76.67% of accuracy.

DECISION TREE: A decision tree is a nonlinear classification technique based on the divide and conquers algorithm. Roots indicate tests for the particular value of a specific attribute, and from where decision alternative branches originate, edges/branches represent the output of the text and connect to the next leaf/ node, and leaf nodes represent the terminal nodes that predict the output and assign class distribution or class labels. For regression problems, continuous values, which are generally real numbers, are taken as input. In classification problems, a Decision Tree takes discrete or categorical values based on binary recursive partitioning involving the fragmentation of data into subsets, further fragmented into smaller subsets. This process continues until the subset data is sufficiently homogenous, and after all the criteria have been efficiently met, the algorithm stops the process. A binary decision tree consisting of SVM classifiers was utilized to classify seven emotions in [58]. Three databases were used, including EmoDB, SAVEE, and Polish Emotion Speech Database. The classification done was based on subjective and objective classes. The highest recognition rate of 82.9% was obtained for EmoDB and least for Polish Emotional Speech Database with 56.25%.

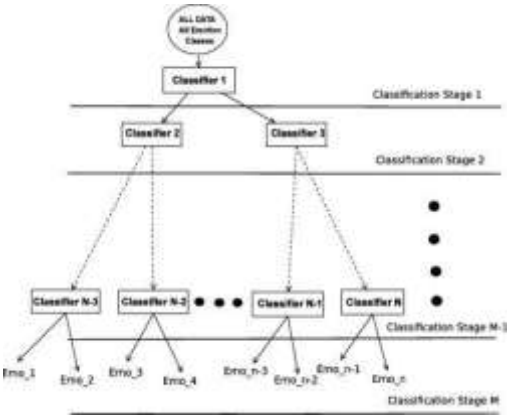

Figure 4: Hierarchical multistage classifier architecture for emotion recognition.

DEEP NEURAL NETWORKS: Deep Neural Networks (DNN) is a neural network with multiple layers and multifaceted nature to process data in complex ways. It can be described as networks with a data layer, an output layer, and one hidden layer in the center. Each layer performs precise types of organizing and requisites in a method that some suggest as ‘‘feature hierarchy.’’ One of the key implementations of these refined neural networks is overseeing unlabeled or unstructured data. A custom-made database was proposed in [61]. For the recognition of emotions, DNN was utilized. First, the network was optimized for four emotions, giving the recognition rate of 97.1% and then for three emotions, resulting in a 96.4% recognition rate. Only the MFCC feature was considered for the experiment. An amalgam of the traditional classification approach – GMM with the neural network was utilized to recognize emotions [62]. A total of four

distinct algorithms were used for the classification process: DNN, GMM, and two different variations of Extreme Machine Learning (EML). It was found that the DNN-EML approach outshined the GMM-based algorithms in terms of accuracy.

5. CHALLENGES:

As we might have thought lately, SER is no longer a peripheral issue. In the last decade, the research in SER has become a significant endeavor in HCI and speech processing. The demand for this technology can be reflected by the enormous research being carried out in SER. Human and machine speech recognition have had large differences since, which presents tremendous difficulty in this subject, primarily the blend of knowledge from interdisciplinary fields, especially in SER, applied psychology, and human-computer interface. One of the main issues is the difficulty of defining the meaning of emotions precisely. Emotions are usually blended and less comprehendible. The collection of databases is a clear reflection of the lack of agreement on the definition of emotions. However, if we consider the everyday interaction between humans and computers, we may see that emotions are voluntary. Those variations are significantly intense as these might be concealed, blended, or feeble and barely recognizable instead of being more prototypical features. Discussing the above facts, we may conclude that additional acoustic features need to be scrutinized to simplify emotion recognition. One more challenge is handling the regularly co-occurring additive noise involving convolute distortion (emerging from a more affordable receiver or other information obtaining devices) and meddling speakers (emerging from background). The various methodology utilized to record elicited emotional speech, enacted emotional speech, and authentic, spontaneous emotional speech must be unique to each other. Recording certified emotion raises a moral issue, just as challenges control recording circumstance and emotional labeling. A broadly acknowledged recording convention is a deficit for the recording of elicited emotion. Another challenge is in applying a reduction in dimensionality and feature selection. Feature selection is costlier and unfeasible because of the enhancement's intricacy that focuses on an appropriate feature subset between the large set of features, particularly when utilizing the wrapper techniques. There is an elective strategy that can be utilized, known as filter-based component determination techniques. They are not founded on classification decisions however consider different qualities like entropy and correlation. The filter has been recently proved to be more helpful for high-resolution data. It comes with a setback; however, these are not appropriate for a wide range of classifiers. Likewise, the feature selection cut-off points may prompt ignoring some ‘‘significant’’ data involved in un-selected features like in CNN.

The problems arise at various stages, including at the time of labeling the utterances. After the utterances are recorded, the speech data is labeled by human annotators. However, there is no doubt that the speaker's actual emotion might vary from the one perceived by the human annotator. Even for human annotators, the recognition rates stay lower than 90%. It is believed that it also depends on both context and content of speech, what the human annotators can infer. SER is affected by culture and language also. Various works have been put forward on cross-language SER that show the ongoing systems and features' insufficiency. Classification is one of the crucial processes in the SER system as it depends on the classifier's ability to interpret the results accurately generated by the respective algorithm. There are various challenges related to the classifiers, like the deep learning classifier CNN is significantly slower due to max-pooling and thus takes a lot of time for the training process. Traditional classifiers such as KNN, Decision Tree, and SVM take a larger amount of time to process the larger datasets. notorious for overfitting problems. We have already discussed various challenges, but not the most ignored challenge, of multi-speech signals. The SER system itself must choose the signal on which the focus should be done. Despite that, this could be controlled by another algorithm, which is the speech separation algorithm at the preprocessing stage itself. The ongoing frameworks nevertheless fail to recognize this issue.

6. CONCLUSION

The capability to drive speech communication using programmable devices is currently in research progress, even if human beings could systematically achieve this errand. The focus of SER research is to design proficient and robust methods to recognize emotions. In this paper, we have offered a precise analysis of SER systems. It makes use of speech databases that provide the data for the training process. Feature extraction is done after the speech signal has undergone preprocessing. The SER system commonly utilizes prosodic and spectral acoustic features such as formant frequencies, spectral energy of speech, speech rate and fundamental frequencies, and some feature extraction techniques like MFCC, LPCC, and TEO features. Two classification algorithms are used to recognize emotions, traditional classifiers, and deep learning classifiers, after the extraction of features. Even if there is much work done using traditional techniques, the turning point in SER is deep learning techniques. Although SER has come far ahead than it was a decade ago, there are still several challenges to work on. Some of them are highlighted in this paper. The system needs more robust algorithms to improve the performance so that the accuracy rates increase and thrive on finding an appropriate set of features and efficient classification techniques to enhance the HCI to a greater extent.

REFERENCE

1. Mekruksavanich, S.; Jitpattanakul, A. Sensor-based Complex Human Activity Recognition from Smartwatch Data Using Hybrid Deep Learning Network. Proceedings of the 36th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC), Jeju, Republic of Korea, 27–30 June 2021; pp. 1–4.
2. Nassif, A.B.; Shahin, I.; Attili, I.; Azzeh, M.; Shaalan, K. Speech recognition using deep neural networks: A systematic review. IEEE Access 2019, 7, 19143–19165. [Google Scholar] [CrossRef]
3. Latif, S.; Qadir, J.; Qayyum, A.; Usama, M.; Younis, S. Speech technology for healthcare: Opportunities, challenges, and state of the art. IEEE Rev. Biomed. Eng. 2020, 14, 342–356.

4. Cho, J.; Kim, B. Performance analysis of speech recognition model based on neuromorphic architecture of speech data preprocessing technique. J. Inst. Internet Broadcast Commun. 2022, 22, 69–74.
5. Lee, S.; Park, H. Deep-learning-based Gender Recognition Using Various Voice Features. In Proceedings of the Symposium of the Korean Institute of Communications and Information Sciences, Seoul, Republic of Korea, 17–19 November 2021; pp. 18–19.
6. Fonseca, A.H.; Santana, G.M.; Bosque Ortiz, G.M.; Bampi, S.; Dietrich, M.O. Analysis of ultrasonic vocalizations from mice using computer vision and machine learning.
7. Lee, Y.; Lim, S.; Kwak, I.Y. CNN-based acoustic scene classification system. Electronics 2021, 10, 371.
8. Ma, X.; Wu, Z.; Jia, J.; Xu, M.; Meng, H.; Cai, L. Emotion recognition from variable-length speech segments using deep learning on spectrograms. Proc. Interspeech 2018, 2018, 3683–3687.
9. Badshah, A.M.; Ahmad, J.; Rahim, N.; Baik, S.W. Speech Emotion Recognition from Spectrograms with Deep Convolutional Neural Network. In Proceedings of the 2017 International Conference on Platform Technology and Service (PlatCon), Busan, Republic of Korea, 13–15 February 2017; pp. 1–5.
10. Zhang, S.; Li, C. Research on feature fusion speech emotion recognition technology for smart teaching. Mob. Inf. Syst. 2022, 2022, 7785929.
11. Subramanian, R.R.; Sireesha, Y.; Reddy, Y.S.P.K.; Bindamrutha, T.; Harika, M.; Sudharsan, R.R. Audio Emotion Recognition by Deep Neural Networks and Machine Learning Algorithms. In Proceedings of the 2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA), Virtual Conference, 8–9 October 2021; pp. 1–6.
12. Zheng, L.; Li, Q.; Ban, H.; Liu, S. Speech Emotion Recognition Based on Convolution Neural Network Combined with Random Forest. In Proceedings of the 2018 Chinese Control and Decision Conference (CCDC), Shenyang, China, 9–11 June 2018; pp. 4143–4147.
13. Li, H.; Zhang, X.; Wang, M.J. Research on speech Emotion Recognition Based on Deep Neural Network. In Proceedings of the 2021 IEEE 6th International Conference on Signal and Image Processing (ICSIP), Nanjing, China, 22–24 October 2021; pp. 795–799.
14. Zhang, Y.; Du, J.; Wang, Z.; Zhang, J.; Tu, Y. Attention-based Fully Convolutional Network for Speech Emotion Recognition. In Proceedings of the 2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Honolulu, HI, USA, 12–15 November 2018; pp. 1771–1775.
15. Carofilis, A.; Alegre, E.; Fidalgo, E.; Fernández-Robles, L. Improvement of accent classification models through grad-transfer from spectrograms and gradient-weighted class activation mapping. IEEE/ACM Trans. Audio Speech Lang. Process. 2023, 31, 2859–2871.