Hindawi
Advances in Human-Computer Interaction
Volume 2022, Article ID 9602429, 11 pages
https://doi.org/10.1155/2022/9602429

Review Article

A Review of the Advancement in Speech Emotion Recognition for Indo-Aryan and Dravidian Languages

Syeda Tamanna Alam Monisha and Sadia Sultana

Department of Computer Science and Engineering, Shahjalal University of Science and Technology, Sylhet, Bangladesh

Correspondence should be addressed to Syeda Tamanna Alam Monisha; alammonisha@gmail.com and Sadia Sultana; sadia-cse@sust.edu

Received 3 October 2022; Revised 10 November 2022; Accepted 21 November 2022; Published 1 December 2022

Academic Editor: Francesco Bellotti

Speech emotion recognition (SER) has grown to be one of the most trending research topics in computational linguistics in the last two decades. Speech being the primary communication medium, understanding the emotional state of humans from speech and responding accordingly have made the speech emotion recognition system an essential part of the human-computer interaction (HCI) field. Although there are a few review works carried out for SER, none of them discusses the development of SER system for the Indo-Aryan or Dravidian language families. This paper focuses on some studies carried out for the development of an automatic SER system for Indo-Aryan and Dravidian languages. Besides, it presents a brief study of the prominent databases available for SER experiments. Some remarkable research works on the identification of emotion from the speech signal in the last two decades have also been discussed in this paper.

1. Introduction

Living in a society, we humans communicate with each other to share our thoughts, feelings, ideas, and different types of information. We use different communication mediums, like text messages, emails, audio, and videos, to express ourselves to others. Besides this, people nowadays use a variety of emojis along with text messages to represent their feelings more precisely. However, without any doubt, among all the communication forms, speech is the most natural and easiest one to express ourselves.

In recent years, contact with computing devices has become more chatty. Dialogue systems like Siri, Alexa, Cortana, and many more have more widely penetrated the consumer market than before [1]. Thus, to make them more like a human conversational partner, it is important to recognize human emotions from the user's voice signals. Understanding the emotional state of a speaker is important for perceiving the exact meaning of what he or she says. Therefore, research studies on automatic speech emotion recognition, which is the task of predicting the emotional state of humans from speech signals have emerged in recent years as it enhances human-computer interaction (HCI) systems and makes them more natural. Moreover, with the world being digitized day by day, speech emotion recognition has found increasing applications in our daily lives. Call centers, e-tutoring, surveillance systems, psychological treatments, robotics, and online marketing are just some of them.

It has been seen from cross-lingual speech emotion recognition studies that models trained with a language corpus do not perform well when tested on a different language corpus compared to the monolingual recognition rate [2, 3]. However, it will be interesting to find out whether those models perform better for different languages from the same groups. To carry out this study, the first thing is to find out the available resources of the target language group. Hence, in this study, we aimed to investigate recent advancements in SER for the Indo-Aryan and Dravidian language families. Indo-Aryan and Dravidian languages are spoken by 800 million and 250 million people, all over the world, respectively [4, 5]. The speakers of Indo-Aryan languages are mostly from Bangladesh, India, Nepal, Sri Lanka, and Pakistan, and speakers of Dravidian languages

are mainly from southern India. Having a large number of speakers, yet most of them are low-resource languages. So far, there is no review work that highlights the SER experiments for the Indo-Aryan or Dravidian language groups. Therefore, this study presents a brief review of some work done for the development of SER for languages of the Indo-Aryan and Dravidian families.

The remaining part of the paper is organized as follows: Section 2 gives a brief overview of a speech emotion recognition system with different types of emotional speech corpora, features, and classification algorithms utilized for the development of an SER system. Trends in speech emotion recognition research have been discussed in Section 3. Section 4 discusses some research works on SER in different languages in the last two decades. In Section 5, the advancement of SER works in Indo-Aryan and Dravidian languages is shown, and lastly, the study is concluded in Section 6.

2. Overview of Speech Emotion Recognition System

A speech emotion recognition (SER) system analyzes human speech and makes predictions about the emotion reflected by the speech. The system that recognizes the emotion from speech may be dependent or independent of the speaker and gender. Comparatively, the recognition accuracy of a speaker-dependent system is higher than that of a speaker-independent system, but the disadvantage of this strategy is that the system only responds appropriately to the person who trained the system. As reflected in Figure 1, the first requirement for building an SER system is a suitable speech dataset having different emotional states. For this purpose, raw speech data are collected from speakers in a variety of ways. Based on the generation of the corpus, emotional speech databases may be natural [6–10], acted [11–13], or elicited [14, 15]. Table 1 summarizes some prominent databases used for SER.

Once the data are collected, the raw speech data go through some preprocessing techniques such as noise reduction, silence removal, framing, windowing, and normalization for enhancing the speech signal [34]. After the preprocessing of raw data, the system opts for the feature extraction phase, which analyzes speech signals and obtains different speech characteristics. Any machine learning model's success is largely dependent on its features. Selecting the right features could result in a more effective trained model, whereas choosing the wrong ones would significantly impede training. The selection of the proper signal features is crucial for better performance in recognizing the emotion of speech. From the beginning of SER research, various arrangements of speech features known as acoustic features like Mel Frequency Cepstral Coefficient (MFCC), pitch, zero crossing rate (ZCR), energy, and linear predictive cepstral coefficients (LPCC) have been used [35]. In various studies, nonspeech characteristics called nonacoustic features have also been integrated with the acoustic ones for the identification of emotion [36, 37]. Gestures, facial images, videos, and linguistic features are some of them.

After the feature selection process, a classifying algorithm is implemented to recognize the speech emotion. For the recognition of emotion from voice signals, many classifying algorithms have been used by researchers. A variety of supervised and unsupervised machine learning models have been employed for this purpose. Hidden Markov model (HMM), support vector machine (SVM), Gaussian mixture model (GMM), *K*-nearest neighbor (KNN), artificial neural network (ANN), and decision tree (DT) are some of them. In recent years, along with the traditional classification methods, several deep learning techniques are also being utilized for the classification process and have shown promising results. Convolutional neural network (CNN), long short-term memory (LSTM), deep CNN, and recurrent neural network (RNN) are the commonly used ones. In many SER studies, multiple classifiers are integrated to enhance the recognition rate. Authors Zhu et al. [38] combined two classifiers, deep belief network (DBN) and support vector machine (SVM), to classify the emotions of anger, fear, happiness, sadness, neutrality, and surprise in the Chinese Academy of Sciences emotional speech database. They used MFCC, pitch, formant, short-term ZCR, and short-term energy as features and achieved a mean accuracy of 95.8%, which is better than using SVM or DBN individually.

3. Speech Emotion Recognition Trends

The very first approach for determining the emotional state of a person from his/her speech was made in the late 1970s by Williamson [39]. Williamson provided a speech analyzer for the determination of an individual's underlying emotion by analyzing pitch or frequency changes in the speech pattern. Later on, in 1996, Dellaert et al. [40] published the first research paper on the topic and introduced statistical pattern recognition techniques in speech emotion recognition. Authors Dellaert et al. [40] implemented *K*-nearest neighbors (KNN), Kernel regression (KR), and Maximum Likelihood Bayes' (MLB) classifier using pitch characteristic of the utterances for the recognition of four different emotions, happiness, fear, anger, and sadness. Along with MLB and nearest neighbor (NN), Kang et al. [41] implemented the hidden Markov model (HMM), where HMM performs the best with 89.1% accuracy for recognizing happiness, sadness, anger, fear, boredom, and neutral emotions utilizing pith and energy features. Onward, HMM has been largely used by researchers for speech emotion recognition showing satisfactory results [42–45]. SVM, GMM, and decision tree (DT) are some more traditional machine learning models which have been reliably used over the years for the same purpose [45–50]. In the 2000s, neural network (NN) has also been widely used for speech emotion recognition studies [51–54]. Indeed, in the earlier approaches, the use of conventional machine learning algorithms was widespread for recognizing the underlying emotion in human speech.

However, in the last decade, the trend of using conventional machine learning models for the recognition of emotion from human speech has moved towards deep learning models. Therefore, deep learning approaches have become more

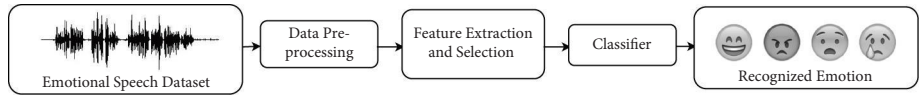

Figure 1: Speech emotion recognition system.

Table 1: Summary of some commonly used databases for speech emotion recognition.

S/N	Database	Language	Year	Size of database	Data format	Emotions
1	RAVDESS [16]	English	2018	7356 recordings by 24 actors	Audio-visual	Happy, angry, calm, sad, fearful, disgust, surprise, neutral
2	EmoDB [11]	German	2005	800 recordings by 10 actors	Audio	Joy, boredom, fear, sadness, anger, disgust, neutral
3	IEMOCAP [14]	English	2008	12 hours of data by 10 actors	Audio-visual	Happiness, anger, sadness, frustration, neutral
4	SUBESCO [12]	Bangla	2021	7 hours of recordings containing 7,000 utterances by 20 native speakers	Audio	Happiness, anger, sadness, fear, surprise, disgust, neutral
5	FAU AIBO [6]	German	2008	9.2 hours of speech by 51 children talking to robot Aibo	Audio	Angry, emphatic, neutral, positive, rest
6	BanglaSER [17]	Bangla	2022	1,467 recording by 34 actors	Audio	Angry, happy, neutral, sad, surprise
7	MELD [7]	English	2018	13,000 utterances from the TV-series friends by multiple speakers	Audio-visual and textual	Anger, fear, joy, surprise, sadness, disgust, neutral
8	AESDD [18]	Greek	2018	500 utterances from 5 different actors	Audio	Anger, disgust, fear, happiness, sadness
9	SAVEE [13]	English	2014	480 British English utterances by 4 male actors	Audio-visual	Disgust, fear, anger, happiness, sadness, surprise, neutral
10	RECOLA [19]	French	2013	3.8 hours of recordings by 46 participants	Audio-visual	Social behaviors (agreement, dominance, engagement, performance, rapport)
11	CHEAVD [8]	Chinese	2017	140 min emotional segments extracted from talk shows, TV plays, and films by 238 speakers	Audio-visual	Angry, fear, happy, neutral, sad, surprise
12	LSSED [20]	English	2020	147,025 utterances from 820 subjects	Audio	Happiness, fear, anger, excitement, sadness, boredom, disappointment, disgust, surprise, normal and other
13	Urdu [2]	Urdu	2018	400 utterances by 38 speakers from Urdu talk shows	Audio	Angry, happy, neutral, sad
14	Urdu-Sindhi speech emotion corpus [21]	Urdu, Sindhi	2020	1435 speech recordings	Audio	Happiness, anger, sadness, disgust, surprise, sarcasm, neutral
15	IITKGP-SEHSC [22]	Hindi	2011	12000 utterances by 10 professionals	Audio	Happy, anger, fear, disgust, surprise, sad, sarcastic, neutral
16	KSUEmotions [23]	Arabic	2017	5 hours and 10 minutes of recordings by 23 speakers	Audio	Neutral, happiness, sadness, surprise, anger
17	Oriya emotional speech database [15]	Oriya	2010	900 emotional utterances for text fragments from various drama scripts of Oriya language by 35 Oriya speakers	Audio	Happiness, sadness, astonish, anger, fear, neutral
18	Mandarin Chinese emotional speech database [24]	Mandarin	2008	3,400 emotional speech utterances by 18 males and 16 females	Audio	Anger, happiness, sadness, boredom, neutral
19	CASIA natural emotional audio-visual database [9]	Chinese	2014	2 hours spontaneous emotional segments extracted from 219 speakers from films, TV plays and talk shows	Audio-visual	Happy, angry, disgust, surprise, worried, sad, fear
20	Egyptian Arabic speech emotion (EYASE) database [25]	Arabic	2020	579 utterances by 3 male and 3 female professional actors from Egyptian TV series	Audio-visual	Angry, happy, neutral, sad

Table 1: Continued.

S/N	Database	Language	Year	Size of database	Data format	Emotions
21	Interface multilingual emotional speech database [26]	English, French, Spanish and Slovenian	2002	In English interface database contains 8,928 utterances, Slovenian 6,080, French 5,600 and Spanish 5,520 by 2 actors	Audio	Joy, disgust, anger, fear, sadness, surprise, neutral
22	Toronto emotional speech set (TESS) [27]	English	2010	2,800 utterances by 2 actresses	Audio	Happiness, anger, sadness, disgust, pleasant surprise, fear, neutral
23	ANAD [28]	Arabic	2018	1,384 recordings from Arabic talk shows	Audio	Happy, angry, surprised
24	Multilingual emotional speech database of north east India (MESDNEI) [29]	Assamese	2009	4,200 utterances of 5 native languages of Assam by 30 speakers	Audio	Sadness, anger, fear, disgust, happiness, surprise, neutral
25	ShEMO [10]	Persian	2019	Speech data of 3 hours and 25 minutes from online radio plays	Audio	Anger, fear, happiness, sadness, surprise, neutral
26	SEMOUR^{+}: a scripted emotional speech repository for Urdu [30]	Urdu	2021	27,640 instances recorded by 24 actors	Audio	Anger, happiness, surprise, disgust, sadness boredom, fearful, neutral
27	Arabic natural corpus [28]	Arabic	2018	1,384 recordings from online Arabic talk shows	Audio	Angry, surprised, happy
28	EMOVO [31]	Italian	2014	588 utterances of 14 sentences by 6 actors	Audio	Joy, sadness, anger, disgust, fear, surprise, neutral
29	Punjabi emotional speech database [32]	Punjabi	2021	900 utterances by 15 speakers	Audio	Happy, angry, sad, fear, surprise, neutral
30	IITKGP-SESC [33]	Telegu	2009	12,000 utterances by 10 professionals	Audio	Happy, anger, sad, disgust, fear, sarcastic, surprise, neutral

popular, showing promising results. Deep learning algorithms are neural networks with multiple layers. CNN, DCNN, LSTM, BLSTM, and RNN are some widely implemented deep learning techniques for SER [55–57].

In recent times, multitask learning and attention mechanism are also being used for improved performance [58, 59]. For cross-corpus and cross-lingual speech emotion recognition, the transfer learning technique is being widely used [3, 60, 61].

Figure 2 depicts an analysis that shows that the use of deep learning techniques like CNN, RNN, LSTM, and DBN has increased over the years, along with traditional machine learning algorithms like SVM, DT, KNN, HMM, and GMM.

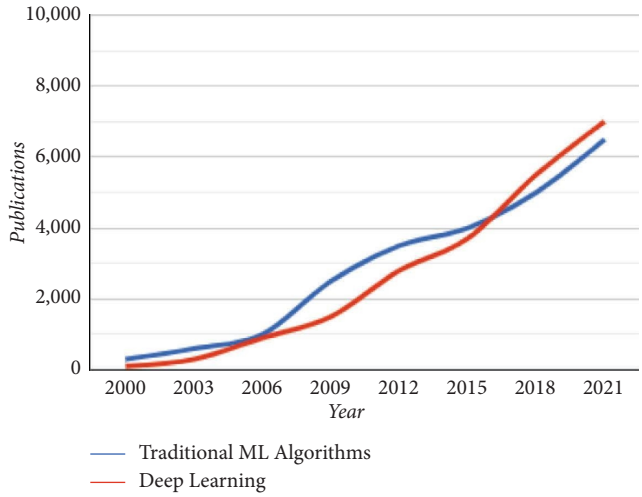

Figure 2: Growth of using deep learning techniques for SER. The graph displays the publication number in the last two decades, having machine learning or deep learning terms in the title of SER papers, according to Google Scholar around the time this article was submitted.

4. Survey on Speech Emotion Recognition Research Studies

After the first published research work on speech emotion recognition in 1996, the field of SER has received a great deal of attention over the past 20 years. Moderate progress has been made to create an automatic SER system. Several acoustic and nonacoustic features have been utilized along with different classifying models. Comparatively, the number of SER experiments conducted in English, German, and French languages is higher than that of the research conducted in other languages. One main reason is the availability of established and publicly accessible databases for the mentioned languages. RAVDESS, IEMOCAP, and SAVEE are some prominent emotional speech databases for the English language, EmoDB for Berlin, FAU AIBO for German, and RECOLA for French. The IEMOCAP database was used by researchers in [56, 58, 59, 62–64] for speech emotion recognition. Fayek et al. [56] evaluated deep learning techniques with CNN and LSTM-RNN using the

IEMOCAP database and achieved 64.78% and 61.71% test accuracy for CNN and LSTM-RNN, respectively. Implementing spectrogram-based self-attentional CNN-BLSTM, Li et al. [58] gained a weighted accuracy of 81.6% and unweighted accuracy of 82.8% for the IEMOCAP dataset for classifying angry, happy, neutral, and sad emotions. Using BLSTM with an attention mechanism, Yu and Kim [59] got a weighted accuracy of 73% and an unweighted accuracy of 68% for the IEMOCAP corpus. Meng et al. [62] used attention mechanism-based dilated CNN with residual block and BiLSTM for both IEMOCAP and Berlin EmoDB and got 74.96% speaker-dependent and 69.32% speaker-independent accuracy for IEMOCAP and 90.78% speaker-dependent and 85.39% speaker-independent for Berlin EmoDB.

A combination of prosodic and modulation spectral features (MSFs) with an SVM classifier was implemented by Wu et al. [65] for the Berlin EmoDB database and the recognition rate was 91.6% for recognizing the emotions in the Berlin EmoDB database. An improved recognition rate of 96.97% was observed by deep convolutional neural network (DCNN) for the Berlin EmoDB database for the recognition of angry, neutral, and sad emotions [66]. For Chinese language authors, Zhang et al. [67] employed SVM and a deep belief network (DBN) with MFCC, pitch, and formant features and got 84.54% mean accuracy by SVM and 94.6% by DBN for the Chinese Academy of Sciences emotional speech database. A higher mean accuracy of 95.8% was achieved for the same Chinese dataset in [38] by combining deep belief network (DBN) with support vector machine (SVM).

Experiments have also been conducted for cross-lingual speech emotion recognition. Sultana et al. [3] showed a cross-lingual study for English and Bangla languages using RAVDESS and SUBESCO datasets, respectively, where the proposed system integrates a deep CNN and a BLSTM network with a TDF layer. Transfer learning was used for the cross-lingual experiment, achieving weighted accuracy of 86.9% for SUBESCO and 82.7% for RAVDESS. Latif et al. [2] used an SVM classifier for cross-lingual emotion recognition for Urdu, German, English, and Italian languages. The authors used SAVEE, EmoDB, EMOVO, and URDU databases for English, German, Italian, and Urdu languages, respectively, for the evaluation of the cross-corpus study. Xiao et al. [68] investigated the cross-lingual study of emotion recognition from speech using the databases EmoDB, DES, and CDESD for German, Danish, and Mandarin languages, respectively. Using CDESD as the training dataset and EmoDB as the testing dataset, the authors achieved the best accuracy of 71.62% for the cross-corpus study with a sequential minimal optimization (SMO) classifier. The IEMOCAP and Recola databases were used for cross-lingual study by Neumann [69] for English and French languages, respectively, where an attentive convolutional neural network (ACNN) was used. 59.32% unweighted average recall was achieved for the IEMOCAP test database while trained on Recola and 61.27% for Recola while training was carried out on the IEMOCAP database. A cross-lingual cross-corpus study was carried out for four languages, German, Italian, English, and Mandarin by Goel and Beigi [70]. Transfer learning and multitask learning techniques were used, providing accuracy of 32%, 51%, and 65% for EMOVO, SAVEE, and EmoDB databases, respectively, using IEMOCAP as the training database.

Apart from using available prominent databases, researchers are also creating emotional speech corpus using acted, elicited, or natural recordings and experimenting with various classification models for the identification of speech emotion. A multilingual database containing 720 utterances by 12 native Burmese and Mandarin speakers was built by Nwe et al. [43]. Using the short-time log frequency power coefficients (LFPC) feature, the authors implemented the HMM classifier, which classifies six emotions, namely, anger, disgust, fear, joy, sadness, and surprise, with an average accuracy of 78% and the best accuracy of 96%.

5. Advancement of Speech Emotion Recognition in Indo-Aryan and Dravidian Languages

Indo-Aryan languages, also known as Indic languages, are the native languages of the Indo-Aryan people, which are a branch of the Indo-Iranian languages in the Indo-European language family. An estimation made at the beginning of the 21st century shows that more than 800 million people, mostly in India, Bangladesh, Sri Lanka, Nepal, and Pakistan, speak Indo-Aryan languages [4]. Hindi, Bangla, Sinhala, Urdu, Punjabi, Assamese, Nepali, Marathi, Odia, Gujarati, Sindhi, Rajasthani, and Chhattisgarhi are some prominent Indo-Aryan languages. Besides, Dravidian or Dravidic languages are spoken by 250 million people primarily in southern India, southwest Pakistan, and north-eastern Sri Lanka [5]. Tamil, Malayalam, Telugu, and Kannada are the most spoken Dravidian languages. Although a lot of work on speech emotion recognition in English, German, Chinese, Mandarin, and French languages has been conducted by researchers, compared to that, the number of experiments in the Indo-Aryan and Dravidian languages is not much. Inadequacy of available resources and variation in the nature of the languages are some reasons for that. However, in the last decade, improvement has been seen in speech emotion recognition research for both language families. Figure 3 shows an analysis of research works done for some of the languages.

5.1. Emotional Speech Databases for Indo-Aryan and Dravidian Languages. Some established and validated emotional speech corpora are available for some of the languages. Hindi is the most spoken language among the Indo-Aryan languages in terms of native speakers. The IITKGP-SESC, Indian Institute of Technology Kharagpur Simulated Emotion Speech Corpus, developed by a team of Indian Institute of Technology Kharagpur in 2009, is the first corpus in Telugu, an Indian language [33]. The corpus contains 12,000 emotional speech utterances in Telugu, with happiness, surprise, anger, disgust, sadness, fear, sarcasm, and neutral emotions expressed by ten speakers.

Afterward, emotions being language independent, Koolagudi et al. [22] felt the necessity of a speech corpus in other Indian languages and created the Indian Institute of Technology Kharagpur Simulated Emotion Hindi Speech Corpus

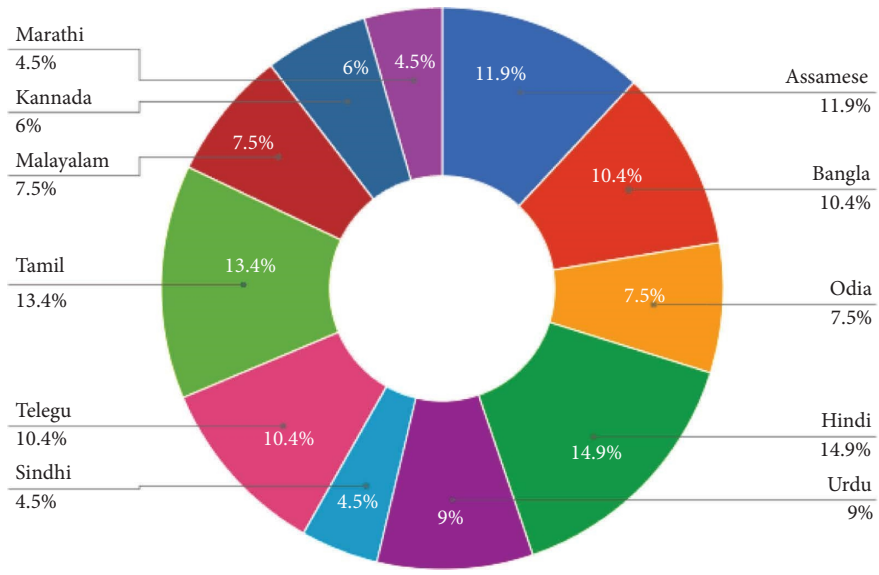

Figure 3: Analysis of SER experiments done in Indo-Aryan and Dravidian languages based on publication number in the last two decades according to Google Scholar around the time this article was submitted.

(IITKGP-SEHSC) developed in the Hindi language. The database contains 12,000 utterances of Hindi speech recorded by ten professional FM radio artists in India. Eight emotions, namely, happiness, sadness, surprise, anger, sarcasm, fear, disgust, and neutral, are present in the database.

A publicly available speech emotion corpus is available in the Urdu language containing 400 utterances by 38 speakers from different Urdu talk shows annotated with emotions of anger, happiness, neutral, and sadness [2]. Asghar et al. [71] built a corpus comprising 2,500 emotional speech utterances by 20 speakers with sadness, anger, disgust, happiness, and neutrality emotions.

SUBESCO, SUST Bangla Emotional Speech Corpus, is the largest available emotional speech corpus for the Bangla language consisting of more than 7 hours of speech with 7,000 utterances [12]. Happiness, surprise, anger, sadness, disgust, fear, surprise, and neutrality are the emotional states present in the database.

Mohanty and Swain [15] developed an Oriya emotional speech corpus for the Oriya language having six emotion classes, namely, happiness, anger, fear, sadness, astonishment, and neutrality.

For the Assamese language, there exists an emotional speech corpus containing utterances in five native Assamese languages, namely, Assamese, Karbi, Bodo (or Boro), Missing (or Mishing), and Dimasa [29].

A Punjabi speech database was created by Kaur and Singh [32] consisting of 900 emotional speech utterances by 15 speakers. Happiness, fearful, angry, surprised, sad, and neutral are the six emotions present in the database.

Kannada emotional speech (KES) database developed by Geethashree and Ravi [72] contains acted emotional utterances in the local languages of Karnataka. The database includes the basic emotions of happiness, sadness, anger, and fear, with a neutral state by four native Kannada actors.

A Malayalam elicited emotional speech corpus for recognizing human emotion from the speech was built by Jacob [73]. The database consists of 2,800 speech recordings in the six basic emotions and neutral by ten native educated and urban female Malayalam speakers.

Apart from these corpora, there are many more small speech databases created for emotion recognition purposes in Indo-Aryan and Dravidian languages [74–77].

5.2. Speech Emotion Recognition for Indo-Aryan and Dravidian Languages. Over the last fifteen years, there has been a moderate progress in SER research for languages of the Indo-Aryan and Dravidian families. Although the earlier approaches were traditional machine learning-based, in recent times, state-of-the-art models are being used by researchers with good performance. After the first large Telugu (IITKGP-SESC) [33] and Hindi (IITKGP-SEHSC) [22] emotional speech databases were published in 2009 and 2011, respectively, many experiments have been done for the languages. In 2021, Agarwal and Om [78] used deep neural network with deer hunting optimization algorithm and got the highest accuracy of 93.75% for the IITKGP-SEHSC dataset. The model implemented for the RAVDESS database outperforms the state-of-the-art accuracy giving 97.14% highest recognition rate [78]. Combining DCNN and BLSTM, the model proposed by Sultana et al. [3] obtained state-of-the-art efficiency with 82.7% and 86.9% accuracy for the RAVDESS and SUBESCO databases for English and Bangla languages, respectively.

Swain et al. [93] in 2022 implemented a deep convolutional recurrent neural network-based ensemble classifier for Odia and RAVDESS databases, which provides better results than some state-of-the-art models for the mentioned databases, giving an accuracy rate of 85.31% and 77.54%, respectively. Likewise, conventional approaches, along with deep learning techniques, are also showing good performance for the language families. Table 2 summarizes some experiments on speech emotion recognition for Indo-Aryan and Dravidian languages.

Table 2: Review of some speech emotion recognition experiments for Indo-Aryan and Dravidian languages.

S/N	Reference	Database		Approach used	Recognized emotions	Results
		Name	Language			
1	Koolagudi et al. [79]	IITKGP-SESC	Telugu	SVM and GMM with energy and pitch parameters	Happy, anger, fear, disgust, sarcastic, sad, neutral, surprise	63.75% average accuracy obtained
2	Sultana et al. [3]	SUBESCO and RAVDESS	Bangla and English	The system integrates a DCNN and a BLSTM network with a TDF layer	Happy, calm, sad, surprise, fearful, disgust, angry, neutral	For the SUBESCO and RAVDESS datasets, the proposed model has achieved weighted accuracies of 86.9% and 82.7%, respectively
3	Kumar and Yadav [80]	IITKGP-SEHSC	Hindi	Deep LSTM with GMFCC and DMFCC features	Happy, fear, angry, sad, neutral	The proposed framework gives average accuracy of 91.2% for male speech and 87.6% for female speech
4	Mohanty and Swain [15]	Oriya emotional speech database	Oriya	Fuzzy *K*-means	Anger, sadness, astonish, fear, happiness, neutral	65.16% recognition rate by incorporating mean pitch, first two formants, jitter, shimmer, and energy as feature vectors
5	Samantaray et al. [48]	MESDNEI	Assamese	SVM with dynamic, quality, derived, and prosodic features	Happy, anger, fear, disgust, surprise, sad, neutral	82.26% average accuracy rate for speaker-independent case
6	Bhavan et al. [81]	EmoDB, RAVDESS and IITKGP-SEHSC	German, English and Hindi	Bagged ensemble of SVM using MFCCs, spectral, and centroids	Happy, sad, calm, angry, surprise, fear, disgust, neutral	Obtained accuracy EmoDB: 92.45%, RAVDESS: 75.69% and IITKGP-SEHSC: 84.11%
7	Swain et al. [82]	Self-created database using utterances from two native languages of Odisha: Cuttacki and Sambalpuri	Oriya	SVM using MFCC as feature vector	Happiness, fear, anger, disgust, sadness, surprise, neutral	82.14% recognition accuracy for SVM classifier
8	Zaheer et al. [30]	SEMOUR$^+$	Urdu	Ensemble classifier, CNN combined with VGG-19 model	Anger, disgust, happiness, surprise, boredom, sadness, fearful, neutral	The proposed model achieved 56% speaker-independent recognition rate
9	Wankhade et al. [47]	Speech emotional database containing dialogues from different bollywood movies	Hindi	SVM classifier with MFCC and MEDC feature set	Angry, happy, sad, neutral	71.66% recognition rate using SVM classifier
10	Ali et al. [83]	Self-created speech emotional corpus recorded in 5 regional languages of Pakistan	Urdu, Sindhi, Pashto, Punjabi, and Balochi	Learning classifiers (adaboostM1, J48, classification via regression, decision stump) with prosodic features	Happiness, sad, anger, neutral	40% classification accuracy with pitch feature
11	Ancilin and Milton [84]	Urdu	Urdu	SVM classifier with mel frequency magnitude coefficient (MFMC)	Happy, sad, anger, neutral	95.25% emotion recognition rate using MFMC
12	Farhad et al. [85]	Urdu	Urdu	Neural network, random forest and meta iterative classifiers with pitch and MFCC features	Happy, sad, angry	With an accuracy of 78.75%, random forest outperforms other classifiers

Table 2: Continued.

S/N	Reference	Database		Approach used	Recognized emotions	Results
		Name	Language			
13	Darekar and Dhande [86]	Marathi database	Marathi	Adaptive ANN combining cepstral, non-negative matrix factorization (NMF) and pitch features	Happy, sad, angry, fear, neutral, surprised	Proposed model obtains 80% accuracy combining the 3 features
14	Koolagudi et al. [87]	IITKGP-SESC	Telugu	SVM and GMM model with epoch parameters were used	Happy, anger, fear, sadness, disgust, neutral	Average recognition rates are 58% and 61% for SVM and GMM, respectively
15	Kandali et al. [49]	Self-created acted emotional speech database by 27 speakers	Assamese	GMM classifier with MFCC features	Happy, sad, disgust, fear, angry, surprise, neutral	Highest mean classification score is 76.5%
16	Dhar and Guha [88]	Abeg: self-collected Bangla emotional speech dataset	Bangla	Logistic regression model with MFCC and LPC features	Happy, angry, neutral	Proposed model achieved 92% accuracy combining MFCC and LPC features
17	Jacob [89]	Hindi emotional speech database containing 2240 wav files collected from 10 speakers	Hindi	ANN model with jitter and shimmer features	Happy, sad, anger, fear, surprise, disgust, neutral	83.3% overall accuracy obtained combining jitter and shimmer features
18	Fernandes and Mannepalli [90]	Acted emotional speech database containing 1400 utterances by 10 actors	Tamil	LSTM and BiLSTM with MFCC, MFCC delta, spectral kurtosis, bark spectrum, and spectral skewness features	Happy, anger, sad, fear, boredom, disgust, neutral	84% accuracy rate obtained using LSTM and BiLSTM with dropout layers
19	Rajisha et al. [91]	Acted emotional dataset created by the authors	Malayalam	ANN and SVM classifier with MFCC, short-time energy, and pitch features	Happy, anger, sad, neutral	88.4% recognition rate obtained using ANN and 78.2% with SVM
20	Kannadaguli and Bhat [92]	Self-created database containing 2800 emotional recordings	Kannada	Bayesian and HMM model with MFCC feature	Happy, excited, angry, sad	Average emotion error rate of 25.5% for Bayesian and 0.2% for HMM approach

6. Conclusion

Speech emotion recognition being an integral part of HCI, a successful SER system with a healthy level of accuracy is essential for the better performance of a human-computer interaction system. This paper presents a survey on speech emotion recognition research for Indo-Aryan and Dravidian languages. A brief review of 31 research studies, including the development of emotional speech corpora and implemented approaches with utilized features for emotion recognition purposes, has been covered for the mentioned language families. Besides, a thorough study of some standard available emotional speech corpora and research works conducted for the identification of emotional states from human speech for different languages has also been presented in this paper. Therefore, researchers working in this field might find helpful insights about speech emotion recognition in this study.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

References

[1] B. W. Schuller, "Speech emotion recognition: two decades in a nutshell, benchmarks, and ongoing trends," *Communications of the ACM*, vol. 61, no. 5, pp. 90–99, 2018.

[2] S. Latif, A. Qayyum, M. Usman, and J. Qadir, "Cross lingual speech emotion recognition: Urdu vs. western languages," in *Proceedings of the 2018 International Conference on Frontiers of Information Technology (FIT)*, pp. 88–93, IEEE, Islamabad, Pakistan, Dec 2018.

[3] S. Sultana, M. Z. Iqbal, M. R. Selim, M. M. Rashid, and M. S. Rahman, "Bangla speech emotion recognition and cross-lingual study using deep cnn and blstm networks," *IEEE Access*, vol. 10, pp. 564–578, 2022.

[4] W. contributors, "Indo-aryan languages — wikipedia, the free encyclopedia," 2022b, https://en.wikipedia.org/w/index.php?title=Indo-%20Aryan_languages&oldid=1107172048.

[5] Wikipedia, "Wikipedia contributors, 2022a. Dravidian languages — wikipedia, the free encyclopedia," https://en.wikipedia.org/w/index.php?title=Dravidian_languages&oldid=11%2009158908.

[6] A. Batliner, S. Steidl, and E. Nöth, "Releasing a Thoroughly Annotated and Processed Spontaneous Emotional Database: The Fau Aibo Emotion Corpus," 2008.

[7] S. Poria, D. Hazarika, N. Majumder, G. Naik, E. Cambria, and R. Mihalcea, "Meld: A Multimodal Multi-Party Dataset for Emotion Recognition in Conversations," 2018, http://arXiv.org/abs/1810.02508.

[8] Y. Li, J. Tao, L. Chao, W. Bao, and Y. Liu, "Cheavd: a Chinese natural emotional audio–visual database," *Journal of Ambient Intelligence and Humanized Computing*, vol. 8, no. 6, pp. 913–924, 2017.

[9] W. Bao, Y. Li, M. Gu et al., "Building a Chinese natural emotional audio-visual database," in *Proceedings of the 2014 12th International Conference on Signal Processing (ICSP)*, pp. 583–587, IEEE, Hangzhou, China, Oct 2014.

[10] O. Mohamad Nezami, P. Jamshid Lou, and M. Karami, "Shemo: a large-scale validated database for Persian speech emotion detection," *Language Resources and Evaluation*, vol. 53, pp. 1–16, 2019.

[11] F. Burkhardt, A. Paeschke, M. Rolfes, W. F. Sendlmeier, and B. Weiss, "A Database of German Emotional Speech," in *Proceedings of the Interspeech 2005 - Eurospeech, 9th European Conference on Speech Communication and Technology*, pp. 1517–1520, Lisbon, Portugal, September 4-8, 2005.

[12] S. Sultana, M. S. Rahman, M. R. Selim, and M. Z. Iqbal, "Sust bangla emotional speech corpus (subesco): an audio-only emotional speech corpus for bangla," *PLoS One*, vol. 16, no. 4, Article ID e0250173, 2021b.

[13] P. Jackson and S. Haq, *Surrey Audio-Visual Expressed Emotion (Savee) Database*, University of Surrey, Guildford, UK, 2014.

[14] C. Busso, M. Bulut, C. C. Lee et al., "Iemocap: interactive emotional dyadic motion capture database," *Language Resources and Evaluation*, vol. 42, no. 4, pp. 335–359, 2008.

[15] S. Mohanty and B. K. Swain, "Emotion recognition using fuzzy k-means from oriya speech," in *Proceedings of the 2010 for International Conference [ACCTA-2010]*, pp. 3–5, 2010.

[16] S. R. Livingstone and F. A. Russo, "The ryerson audio-visual database of emotional speech and song (ravdess): a dynamic, multimodal set of facial and vocal expressions in north american English," *PLoS One*, vol. 13, no. 5, Article ID e0196391, 2018.

[17] R. K. Das, N. Islam, M. R. Ahmed, S. Islam, S. Shatabda, and A. M. Islam, "Banglaser: a speech emotion recognition dataset for the bangla language," *Data in Brief*, vol. 42, Article ID 108091, 2022.

[18] N. Vrysas, R. Kotsakis, A. Liatsou, C. A. Dimoulas, and G. Kalliris, "Speech emotion recognition for performance interaction," *Journal of the Audio Engineering Society*, vol. 66, no. 6, pp. 457–467, 2018.

[19] F. Ringeval, A. Sonderegger, J. Sauer, and D. Lalanne, "Introducing the recola multimodal corpus of remote collaborative and affective interactions," in *Proceedings of the 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, pp. 1–8, IEEE, Shanghai, China, 22-26 April 2013.

[20] W. Fan, X. Xu, X. Xing, W. Chen, and D. Huang, "Lssed: a large-scale dataset and benchmark for speech emotion recognition," in *Proceedings of the ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 641–645, IEEE, Toronto, ON, Canada, June 2021.

[21] Z. S. Syed, S. Ali, M. Shehram, and A. Shah, "Introducing the Urdu-Sindhi speech emotion corpus: a novel dataset of speech recordings for emotion recognition for two low-resource languages," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 4, 2020.

[22] S. G. Koolagudi, R. Reddy, J. Yadav, and K. S. Rao, "Iitkgp-sehsc: Hindi speech corpus for emotion analysis," in *Proceedings of the 2011 International Conference on Devices and Communications (ICDeCom)*, pp. 1–5, IEEE, Mesra, India, Feb 2011.

[23] A. H. Meftah, M. A. Qamhan, Y. Seddiq, Y. A. Alotaibi, and S. A. Selouani, "King saud university emotions corpus: construction, analysis, evaluation, and comparison," *IEEE Access*, vol. 9, pp. 54201–54219, 2021.

[24] T. Pao, Y. Chen, and J. Yeh, "Emotion recognition and evaluation from Mandarin speech signals," *International Journal of Innovative Computing, Information and Control*, vol. 4, pp. 1695–1709, 2008.

[25] L. Abdel-Hamid, "Egyptian Arabic speech emotion recognition using prosodic, spectral and wavelet features," *Speech Communication*, vol. 122, pp. 19–30, 2020.

[26] V. Hozjan, Z. Kacic, A. Moreno, A. Bonafonte, and A. Nogueiras, "Interface Databases: Design and Collection of a Multilingual Emotional Speech Database," in *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC'02)*, European Language Resources Association (ELRA), Las Palmas, Canary Islands, May 2002.

[27] K. Dupuis and M. K. Pichora-Fuller, "Recognition of emotional speech for younger and older talkers: behavioural findings from the toronto emotional speech set," *Canadian Acoustics*, vol. 39, pp. 182-183, 2011.

[28] S. Klaylat, Z. Osman, L. Hamandi, and R. Zantout, "Emotion recognition in Arabic speech," *Analog Integrated Circuits and Signal Processing*, vol. 96, no. 2, pp. 337–351, 2018.

[29] A. B. Kandali, A. Routray, and T. K. Basu, "Vocal emotion recognition in five native languages of Assam using new wavelet features," *International Journal of Speech Technology*, vol. 12, pp. 1–13, 2009.

[30] N. Zaheer, O. U. Ahmad, M. Shabbir, and A. A. Raza, "Speech emotion recognition for the Urdu language," *Language Resources and Evaluation*, pp. 1–30, 2022.

[31] G. Costantini, I. Iaderola, A. Paoloni, and M. Todisco, "Emovo corpus: an Italian emotional speech database," in *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2014)*, pp. 3501–3504, European Language Resources Association (ELRA), Reykjavik, Iceland, May 2014.

[32] K. Kaur and P. Singh, "Punjabi emotional speech database: design, recording and verification," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 9, no. 4, pp. 205–208, 2021.

[33] S. G. Koolagudi, S. Maity, V. A. Kumar, S. Chakrabarti, and K. S. Rao, "Iitkgp-sesc: speech database for emotion analysis," in *Proceedings of the International Conference on Contemporary Computing*, pp. 485–492, Springer, 2009.

[34] T. M. Wani, T. S. Gunawan, S. A. A. Qadri, M. Kartiwi, and E. Ambikairajah, "A comprehensive review of speech emotion recognition systems," *IEEE Access*, vol. 9, pp. 47795–47814, 2021.

[35] A. Koduru, H. B. Valiveti, and A. K. Budati, "Feature extraction algorithms to improve the speech emotion recognition rate," *International Journal of Speech Technology*, vol. 23, no. 1, pp. 45–55, 2020.

[36] C. Busso, Z. Deng, S. Yildirim et al., "Analysis of emotion recognition using facial expressions, speech and multimodal information," in *Proceedings of the 6th International Conference on Multimodal Interfaces*, pp. 205–211, New York NY United States, 2004.

[37] P. Tzirakis, G. Trigeorgis, M. A. Nicolaou, B. W. Schuller, and S. Zafeiriou, "End-to-end multimodal emotion recognition using deep neural networks," *IEEE Journal of selected topics in signal processing*, vol. 11, no. 8, pp. 1301–1309, 2017.

[38] L. Zhu, L. Chen, D. Zhao, J. Zhou, and W. Zhang, "Emotion recognition from Chinese speech for smart affective services using a combination of svm and dbn," *Sensors*, vol. 17, no. 7, p. 1694, 2017.

[39] J. D. Williamson, "Speech analyzer for analyzing pitch or frequency perturbations in individual speech pattern to determine the emotional state of the person," *US Patent*, vol. 4, no. 093, p. 821, 1978.

[40] F. Dellaert, T. Polzin, and A. Waibel, "Recognizing emotion in speech," in *Proceedings of the Fourth International Conference on Spoken Language Processing. ICSLP*, pp. 1970–1973, IEEE, Philadelphia, PA, USA, 03-06 October 1996.

[41] B. S. Kang, C. H. Han, S. T. Lee, D. H. Youn, and C. Lee, "Speaker dependent emotion recognition using speech signals," in *Proceedings of the Sixth International Conference on Spoken Language Processing*, Beijing, China, October 2000.

[42] B. Schuller, G. Rigoll, and M. Lang, "Hidden Markov model-based speech emotion recognition," in *Proceedings of the 2003 IEEE international conference on acoustics, speech, and signal processing*, IEEE. Baltimore, MD, USA, 06-09 July 2003.

[43] T. L. Nwe, S. W. Foo, and L. C. De Silva, "Speech emotion recognition using hidden Markov models," *Speech Communication*, vol. 41, no. 4, pp. 603–623, 2003.

[44] A. Nogueiras, A. Moreno, A. Bonafonte, and J. B. Mariño, "Speech emotion recognition using hidden Markov models," in *Proceedings of the Seventh European Conference on Speech Communication and Technology*, 2001.

[45] Y. L. Lin and G. Wei, "Speech emotion recognition based on hmm and svm," in *Proceedings of the 2005 International Conference on Machine Learning and Cybernetics*, pp. 4898–4901, IEEE, Guangzhou, China, Aug 2005.

[46] L. Sun, B. Zou, S. Fu, J. Chen, and F. Wang, "Speech emotion recognition based on dnn-decision tree svm model," *Speech Communication*, vol. 115, pp. 29–37, 2019.

[47] S. B. Wankhade, P. Tijare, and Y. Chavhan, "Speech emotion recognition system using svm and libsvm," *International Journal of Computer Science and Applications*, vol. 4, 2011.

[48] A. K. Samantaray, K. Mahapatra, B. Kabi, and A. Routray, "A novel approach of speech emotion recognition with prosody, quality and derived features using svm classifier for a class of north-eastern languages," in *Proceedings of the 2015 IEEE 2nd International Conference on Recent Trends in Information Systems (ReTIS)*, pp. 372–377, IEEE, Kolkata, India, July 2015.

[49] A. B. Kandali, A. Routray, and T. K. Basu, "Emotion recognition from Assamese speeches using mfcc features and gmm classifier," in *Proceedings of the TENCON 2008-2008 IEEE Region 10 Conference*, pp. 1–5, IEEE, Hyderabad, India, Nov 2008.

[50] H. Hu, M. X. Xu, and W. Wu, "Gmm supervector based svm with spectral features for speech emotion recognition," in *Proceedings of the 2007 IEEE International Conference on Acoustics, Speech and Signal Processing-Icassp'07*, pp. IV–413, IEEE, Honolulu, HI, USA, 15-20 April 2007.

[51] J. Nicholson, K. Takahashi, and R. Nakatsu, "Emotion recognition in speech using neural networks," *Neural Computing & Applications*, vol. 9, no. 4, pp. 290–296, 2000.

[52] X. Mao, L. Chen, and L. Fu, "Multi-level speech emotion recognition based on hmm and ann," in *Proceedings of the 2009 WRI World congress on Computer Science and Information Engineering*, pp. 225–229, IEEE, Los Angeles, CA, USA, 2009.

[53] B. Schuller, G. Rigoll, and M. Lang, "Speech emotion recognition combining acoustic features and linguistic information in a hybrid support vector machine-belief network architecture," in *Proceedings of the 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. I–577, IEEE, Montreal, QC, Canada, 17-21 May 2004.

[54] M. W. Bhatti, Y. Wang, and L. Guan, "A neural network approach for human emotion recognition in speech," in *Proceedings of the 2004 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. II–181, IEEE, Vancouver, BC, Canada, May 2004.

[55] B. J. Abbaschian, D. Sierra-Sosa, and A. Elmaghraby, "Deep learning techniques for speech emotion recognition, from databases to models," *Sensors*, vol. 21, no. 4, p. 1249, 2021.

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

[57] S. K. Pandey, H. S. Shekhawat, and S. M. Prasanna, "Deep learning techniques for speech emotion recognition: a review," in *Proceedings of the 2019 29th International Conference Radioelektronika (RADIOELEKTRONIKA)*, pp. 1–6, IEEE, Pardubice, Czech Republic, April 2019.

[58] Y. Li, T. Zhao, and T. Kawahara, "Improved End-To-End Speech Emotion Recognition Using Self Attention Mechanism and Multitask Learning," in *Proceedings of the Interspeech*, pp. 2803–2807, September 2019.

[59] Y. Yu and Y. J. Kim, "Attention-lstm-attention model for speech emotion recognition and analysis of iemocap database," *Electronics*, vol. 9, no. 5, p. 713, 2020.

[60] S. Latif, R. Rana, S. Younis, J. Qadir, and J. Epps, "Transfer Learning for Improving Speech Emotion Classification Accuracy," 2018b, http://arXiv.org/abs/1801.06353.

[61] J. Deng, Z. Zhang, E. Marchi, and B. Schuller, "Sparse autoencoder-based feature transfer learning for speech emotion recognition," in *Proceedings of the 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, pp. 511–516, IEEE, Geneva, Switzerland, Sep 2013.

[62] H. Meng, T. Yan, F. Yuan, and H. Wei, "Speech emotion recognition from 3d log-mel spectrograms with deep learning network," *IEEE Access*, vol. 7, pp. 125868–125881, 2019.

[63] D. Issa, M. Fatih Demirci, and A. Yazici, "Speech emotion recognition with deep convolutional neural networks," *Biomedical Signal Processing and Control*, vol. 59, Article ID 101894, 2020.

[64] M. Neumann and N. T. Vu, "Attentive Convolutional Neural Network Based Speech Emotion Recognition: A Study on the Impact of Input Features, Signal Length, and Acted Speech," 2017, http://arXiv.org/abs/1706.00612.

[65] S. Wu, T. H. Falk, and W. Y. Chan, "Automatic speech emotion recognition using modulation spectral features," *Speech Communication*, vol. 53, no. 5, pp. 768–785, 2011.

[66] P. Harár, R. Burget, and M. K. Dutta, "Speech emotion recognition with deep learning," in *Proceedings of the 2017 4th International Conference on Signal Processing and Integrated Networks (SPIN)*, pp. 137–140, IEEE, Noida, India, Feb. 2017.

[67] W. Zhang, D. Zhao, Z. Chai et al., "Deep learning and svm-based emotion recognition from Chinese speech for smart affective services," *Software: Practice and Experience*, vol. 47, pp. 1127–1138, 2017.

[68] Z. Xiao, D. Wu, X. Zhang, and Z. Tao, "Speech emotion recognition cross language families: Mandarin vs. western languages," in *Proceedings of the 2016 International Conference on Progress in Informatics and Computing (PIC)*, pp. 253–257, IEEE, Shanghai, China, Dec 2016.

[69] M. Neumann, "Cross-lingual and multilingual speech emotion recognition on English and French," in *Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5769–5773, IEEE, Calgary, AB, Canada, 15-20 April 2018.

[70] S. Goel and H. Beigi, "Cross Lingual Cross Corpus Speech Emotion Recognition," 2020, http://arXiv.org/abs/2003.07996.

[71] A. Asghar, S. Sohaib, S. Iftikhar, M. Shafi, and K. Fatima, "An Urdu speech corpus for emotion recognition," *PeerJ Computer Science*, vol. 8, p. e954, 2022.

[72] A. Geethashree and D. Ravi, "Kannada emotional speech database: design, development and evaluation," in *Proceedings of the International Conference on Cognition and Recognition*, pp. 135–143, Springer, 2018.

[73] A. Jacob, "Modelling speech emotion recognition using logistic regression and decision trees," *International Journal of Speech Technology*, vol. 20, no. 4, pp. 897–905, 2017.

[74] R. Kaushik, M. Sharma, K. K. Sarma, and D. I. Kaplun, "I-vector based emotion recognition in Assamese speech," *International Journal of Engineering and Future Technology*, vol. 1, pp. 111–124, 2016.

[75] V. B. Waghmare, R. R. Deshmukh, P. P. Shrishrimal, G. B. Janvale, and B. Ambedkar, "Emotion recognition system from artificial Marathi speech using mfcc and lda techniques," in *Proceedings of the Fifth International Conference on Advances in Communication, Network, and Computing–CNC*, 2014.

[76] A. Agrawal and A. Jain, "Speech emotion recognition of Hindi speech using statistical and machine learning techniques," *Journal of Interdisciplinary Mathematics*, vol. 23, no. 1, pp. 311–319, 2020.

[77] K. Mannepalli, P. N. Sastry, and M. Suman, "Analysis of emotion recognition system for Telugu using prosodic and formant features," in *Proceedings of the Speech and Language Processing for Human-Machine Communications*, pp. 137–144, Springer, 2018.

[78] G. Agarwal and H. Om, "Performance of deer hunting optimization based deep learning algorithm for speech emotion recognition," *Multimedia Tools and Applications*, vol. 80, no. 7, pp. 9961–9992, 2021.

[79] S. G. Koolagudi, N. Kumar, and K. S. Rao, "Speech emotion recognition using segmental level prosodic analysis," in *Proceedings of the 2011 International Conference on Devices and Communications (ICDeCom)*, pp. 1–5, IEEE, Mesra, India, 2011.

[80] S. Kumar and J. Yadav, "Emotion recognition in Hindi language using gender information, gmfcc, dmfcc and deep lstm," in *Proceedings of the Journal of Physics: Conference Series*, IOP Publishing, Article ID 012049, 2021.

[81] A. Bhavan, P. Chauhan, R. R. Shah, and R. R. Shah, "Bagged support vector machines for emotion recognition from speech," *Knowledge-Based Systems*, vol. 184, Article ID 104886, 2019.

[82] M. Swain, S. Sahoo, A. Routray, P. Kabisatpathy, and J. N. Kundu, "Study of feature combination using hmm and svm for multilingual odiya speech emotion recognition," *International Journal of Speech Technology*, vol. 18, no. 3, pp. 387–393, 2015.

[83] S. A. Ali, A. Khan, and N. Bashir, "Analyzing the impact of prosodic feature (pitch) on learning classifiers for speech emotion corpus," *International Journal of Information Technology and Computer Science*, vol. 7, pp. 54–59, 2015.

[84] J. Ancilin and A. Milton, "Improved speech emotion recognition with mel frequency magnitude coefficient," *Applied Acoustics*, vol. 179, Article ID 108046, 2021.

[85] M. Farhad, H. Ismail, S. Harous, M. M. Masud, and A. Beg, "Analysis of emotion recognition from cross-lingual speech: Arabic, English, and Urdu," in *Proceedings of the 2021 2nd International Conference on Computation, Automation and Knowledge Management (ICCAKM)*, pp. 42–47, IEEE, Dubai, United Arab Emirates, Jan 2021.

[86] R. V. Darekar and A. P. Dhande, "Emotion recognition from Marathi speech database using adaptive artificial neural network," *Biologically inspired cognitive architectures*, vol. 23, pp. 35–42, 2018.

[87] S. G. Koolagudi, R. Reddy, and K. S. Rao, "Emotion recognition from speech signal using epoch parameters," in *Proceedings of the 2010 International Conference on Signal Processing and Communications (SPCOM)*, pp. 1–5, IEEE, Bangalore, India, July 2010.

[88] P. Dhar and S. Guha, "A system to predict emotion from Bengali speech," *International Journal of Mathematics and Soft Computing*, vol. 7, no. 1, pp. 26–35, 2021.

[89] A. Jacob, "Speech emotion recognition based on minimal voice quality features," in *Proceedings of the 2016 International Conference on Communication and Signal Processing (ICCSP)*, pp. 0886–0890, IEEE, Melmaruvathur, India, 6-8 April 2016.

[90] B. Fernandes and K. Mannepalli, "Speech emotion recognition using deep learning lstm for Tamil language," *Pertanika Journal of Science and Technology*, vol. 29, no. 3, 2021.

[91] T. Rajisha, A. Sunija, and K. Riyas, "Performance analysis of Malayalam language speech emotion recognition system using ann/svm," *Procedia Technology*, vol. 24, pp. 1097–1104, 2016.

[92] P. Kannadaguli and V. Bhat, "A comparison of bayesian and hmm based approaches in machine learning for emotion detection in native Kannada speaker," in *Proceedings of the 2018 IEEMA Engineer Infinite Conference (eTechNxT)*, pp. 1–6, IEEE, New Delhi, India, 13-14 March 2018.

[93] M. Swain, B. Maji, P. Kabisatpathy, and A. Routray, "A dcrnn-based ensemble classifier for speech emotion recognition in Odia language," *Complex & Intelligent Systems*, vol. 8, no. 5, pp. 4237–4249, 2022.