Speech Emotion Recognition Using Hierarchical Approach Based on Bhattacharyya Distance

Abrar Adil[*], Sana Ul Haq[†], Muhammad Saeed Shah[‡], Imtiaz Rasool[§], Muhammad Kamran[**]

Abstract

In human-human interaction emotions play an essential role in conveying the information apart from the verbal communication. It is quite challenging for machines to recognize human emotions and respond accordingly. Most research in emotion recognition has focused on using the flat approach in which emotions are classified in a single step using a single best set of features. This paper presents a hierarchical approach based on Bhattacharyya distance for human emotion recognition from speech. The basic aim is to improve the emotion classification performance. The analysis is performed using the Surrey Audio-Visual Expressed Emotion (SAVEE) database, while various attribute selectors and classification techniques are implemented to obtain the best results. The experimental results showed better performance for the proposed hierarchical approach as compared to flat approach and other state-of-the-art techniques. The best classification accuracy of 78.12% is achieved for the flat approach, while the best performance of 91.66% is obtained for the hierarchical approach using seven emotions of the SAVEE database.

Keywords: Speech Emotion Recognition; Feature Selection; Hierarchical Classification; Info Gain; Gain Ratio; Support Vector Machine.

Introduction

Recognition of emotions from various modalities, e.g., audio, visual or physiological signals, is an emerging field of research (Khare et al., 2024). The reason behind this is the necessity of systems capable of automatically recognizing human emotions and respond accordingly (Ghazi et al., 2010). In recent years, the requirement for automatic emotion recognition systems has acquired a great interest (Tarasov & Delany, 2011). These systems have applications in customer care centers, security agencies, smart phones and other electronic gadgets. Due to the vast demand of these technologies, research is being carried out in designing

[*] Department of Electronics, University of Peshawar, Peshawar 25120, Pakistan, abrar.adil@yahoo.com

[†]Corresponding Author: Department of Electronics, University of Peshawar, Peshawar 25120, Pakistan, sanaulhaq@uop.edu.pk

[‡] Department of Electronics, University of Peshawar, Peshawar 25120, Pakistan, saeedshah@uop.edu.pk

[§] Department of Electronics, University of Peshawar, Peshawar 25120, Pakistan, imtiazrasoolkhan@uop.edu.pk

[**] Department of Electronics, University of Peshawar, Peshawar 25120, Pakistan, kamranmu@uop.edu.pk

better human-machine interaction systems, by either experimenting with new techniques or modifying the existing systems (Xiao et al., 2007).

A considerable amount of research is being carried out in the field of automatic recognition of emotions from speech to design systems that are capable of interpreting human voice signals to respective emotions, thus helping the machines to understand and respond to customer demands (Haag et al., 2004). Araño et al. (2021) used a combination of Mel Frequency Cepstral Coefficients (MFCCs) and image features extracted from spectrograms for speech emotion classification. The Long Short-Term Memory (LSTM) network with Mel-Frequency Cepstral Coefficients (MFCCs) features provided better results as compared to Support Vector Machine (SVM) classifier. Mannepalli et al. (2022) proposed multiples support vector neural network classifier for emotion classification using speech. The proposed model outperformed the Adaptive Fractional Deep Belief Network (AFDBN), Fractional Deep Belief Network (FDBN), and Deep Belief Network (DBN). Alluhaidan et al. (2023) achieved better classification performance by combining MFCCs and time-domain features (MFCCT). The proposed Convolutional Neural Network (CNN) performed better in comparison to other machine learning models. Er (2020) performed speech emotion recognition using SVM classifier with acoustic and deep features. The average classification results of 79.4%, 90.2% and 85.4% were obtained for the RAVDESS, EMO-DB, and IEMOCAP datasets, respectively. A brain emotional learning model was proposed by Liu et al. (2018) for emotion recognition from speech. The classification was performed using MFCC related features. The proposed model achieved classification performance of 76.4% on SAVEE, 90.3% on CASIA, and 71.1% on FAU Aibo datasets.

To achieve better classification performance, some researchers has suggested the ensemble technique which aggregates the outputs of different weak classifiers. Mohan et al. (2023) used MFCC features for emotion classification. The 2D CNN and eXtreme Grading Boosting (XG-Boost) were combined to obtain classification accuracy of 96.5% using 16 emotions of RAVDESS dataset. The proposed ensemble model outperformed the CNN-LSTM and Random Forest classifiers. Bhanusree et al. (2023) extracted features using time-distributed attention-layered CNN and performed classification using Random Forest. The proposed technique obtained classification accuracies of 90.3% and 92.2% using the IEMOCAP and RAVDESS datasets, respectively. A hybrid model consisting of Convolutional and LSTM (ConvLSTM) networks was proposed by Badr et al. (2021). The proposed technique achieved a classification performance of 91.0% on the RAVDESS dataset using MFCC features. Novais et al. (2022) performed emotion recognition from speech using Adaptive Boosting (AdaBoost), Random Forest and Neural Network. An ensemble technique of majority vote was also investigated. The classification performance of 75.6% was obtained on the RAVDESS dataset using Random Forest, and 86.4% was achieved on a group of RAVDESS, SAVEE, and TESS datasets using Neural Network. The individual classifiers outperformed the ensemble method using majority vote. Chalapathi et al. (2022) used high-dimensional acoustic features with AdaBoost classifier for speech emotion recognition. The classification accuracy of 94.8% was achieved on the RAVDESS database using seven emotions.

The emotion recognition techniques can be broadly categorized into flat and hierarchical approaches. The flat approach performs single-stage classification of all emotion classes using the best set of features. In the hierarchical approach, all emotion classes are arranged in binary groups on the basis of similarity or difference between them. Thus, a downward-branching hierarchical tree is formed. At each level of binary classification, a different set of features is used that can effectively separate the two classes. The hierarchical approach is expected to efficiently classify the more confusing classes due to usage of distinct set of features at each binary level of classification. On the other hand, the flat approach uses a single best set of features to classify all emotions. This research is focused on investigating the advantage of hierarchical approach over the flat approach for speech emotion classification. The following sections present the methodology, experimental results, discussion, and conclusion.

Methodology

The emotion classification was performed using the following steps: acquisition of SAVEE database, feature extraction, feature normalization, feature selection, and classification.

SAVEE Database

In this research, the SAVEE database (Haq & Jackson, 2011) is used for the analysis. It was recorded at the University of Surrey, UK. The database was captured from four native British male speakers in six basic emotions plus neutral state. The data was recorded using high quality equipment. The evaluation of recordings was performed by 10 subjects. The dataset has phonetically balanced sentences for each emotion class. It contains 15 sentences per each emotion, and 30 sentences for the neutral state. The number of recorded sentences per speaker is 120, and thus the dataset contains 480 instances in total. The average human classification accuracy for the audio data is 66.5% for seven emotions.

Feature Extraction

The emotional state of a speaker is reflected in the speech mainly in the form of its spectrum and prosody. The spectrum indicates the characteristics of a vowel sound and prosody indicates the rhythm of the speech. The acoustic features that are commonly used for emotion classification include prosody, voice quality and spectral features (Zhang et al., 2012). In this research, the openSMILE toolkit (Eyben et al., 2009) was used to extract 6669 audio features. These features included signal energy, Mel spectrum, cepstral, pitch, spectral, raw signal, and voice quality related low-level descriptors (LLDs), their delta (Δ) and delta-delta (ΔΔ) coefficients. A total of 6669 features were extracted for each speech signal by applying 39 statistical functions to these features.

Feature Normalization

The extracted features normally have different range of values because of different nature. It is therefore necessary to normalize them to a specific range for equal weighting of these features, and to avoid bias due to high range of values of some features. The two prominent normalization techniques are the Min-Max normalization (Jain & Bhandare, 2011) and Z-Score normalization (Jain et al., 2005).

The Min-Max normalization method normalizes data to the range $[v_{min} \;\; v_{max}]$ using the following relation:

$$\bar{x} = \frac{x - x_{min}}{x_{max} - x_{min}} \times (v_{max} - v_{min}) \; + v_{min} \tag{1}$$

where $\bar{x}$, x_{min} , and x_{max} represent the normalized, minimum and maximum values of attribute x.

The Z-Score normalization scales the data to zero mean and unit variance. It is defined by the relation:

$$z = \frac{x - \mu}{\sigma} \tag{2}$$

where z, μ and σ are the normalized value, mean and standard deviation of feature x. The Z-Score normalization technique is used in this research.

Feature Selection

Feature selection is an important element of a classification task. The redundant and unrelated features need to be removed in order to reduce the computational complexity and to boost the classification performance. In some feature selection techniques the best set of features are selected based on some criterion, while in others individual features are ranked based on certain measure. In this research, both the subset feature selection and ranker methods are used to achieve higher classification performance. The Correlation-based Feature Selection

(CFS) was used as a feature subset evaluator, while the Info Gain and Gain Ratio were used as ranker methods. The CFS method creates subsets of different features for stability. The feature subsets can be selected using different search methods including Best First, Greedy Stepwise, and Linear Forward Selection. The CFS method provides higher grade to the feature subset whose attributes are highly correlated to the class but are weakly correlated to each other (Hall, 1999).

The Info Gain attribute evaluator provides higher rank to an attribute by measuring the information gain with respect to the class (Azhagusundari & Thanamani, 2013). Info Gain is an important attributes evaluation technique due to its methodology. In decision tree, Info Gain decides which of the features are more relevant, thus it keeps only those attributes and discards the remaining at an early stage. The info Gain of an attribute with respect to a class is defined by:

$$Info\ Gain(Class, Attribute) = H(Class) - H\left(\frac{Class}{Attribute}\right) \quad (3)$$

where H denotes the information entropy.

Gain Ratio is a modified form of the Info Gain which reduces its bias (Witten et al., 2016). During feature selection, Gain Ratio checks the size and number of branches. It corrects the Info Gain by looking at the intrinsic information. The Gain Ratio is defined by the following relation:

$$Gain\ Ratio\ (S, A) = \frac{Gain\ (S,A)}{Intrinsic\ Info\ (S,A)} \quad (4)$$

The Intrinsic information is defined as the entropy of distribution of instances into branches. The Gain Ratio ranks the feature by computing the gain ratio with respect to the class.

Classification

In the final step of a classification task the classes need to be correctly separated from each other. In general, the flat approach is used for multi-class classification problems. The flat approach utilizes the best set of features to classify all classes in a single step. The flat approach results in lower classification performance for highly correlated classes and high dimensional data, e.g., protein classification. For this reason, when the classes are large in number and are closely related to each other, the hierarchical classification technique may be useful. The hierarchical approach breaks down the multi-class classification problem to binary classification tasks. In this technique a branched tree is constructed through binary classification. The hierarchical technique is believed to perform better than the traditional flat approach as a different set of features is used at each level of binary classification. This research utilizes the hierarchical approach for emotion classification due to its better capability of solving the multi-class classification problems. The classification was performed using four different classification methods, i.e., Bayes Net (Liu et al., 2016), Support Vector Machine (SVM) (Platt, 1999), Meta Bagging (Büchlmann & Yu, 2002) and Functional Tree (Gama, 2004), using the Weka toolkit (Witten et al., 2016). These classifiers utilize different approaches for the classification. The Bayesian classifiers are based on the Bayes' Theorem. These classifiers compute the probabilities of different classes using the given features. SVM transform the data from low to high dimensional space where it can be easily separated. It is faster and performs better for less training data and high dimensional space. Bagging is a machine learning ensemble technique aimed to enhance the stability and accuracy of classification algorithms. It reduces overfitting by decreasing the variance. Functional tree is a multivariate learning algorithm that utilizes constructive induction to aggregate a univariate decision tree. It has better generalization ability.

In this research, the hierarchical approach was used for audio emotion classification based on Bhattacharyya distance (Bhattacharyya, 1943). The similarity between two multinomial distributed classes can be measured using the Bhattacharyya distance. The Bhattacharyya distance between two normally distributed classes is given by:

$$d_{Bhat} = \frac{1}{8}\left(\mu_i - \mu_j\right)^T \left(\frac{\sum_i + \sum_j}{2}\right)^{-1} \left(\mu_i - \mu_j\right) + \frac{1}{2}\ln\left(\frac{\left|\frac{\sum_i + \sum_j}{2}\right|}{\sqrt{|\sum_i|\left|\sum_j\right|}}\right) \quad (5)$$

where μ_i and μ_j denote the means, while $\sum_i$ and $\sum_j$ are the covariance of two classes.

All of the emotion classes were mapped in a tree structure based on Bhattacharyya distance, as shown in Figure 1. First, all seven emotions were grouped together in a master class. In the next step, two groups class A and class B were made. The emotions close to each other, i.e., anger, fear, happy, and surprise were placed in class A, while the rest of emotions, i.e., disgust, neutral, and sad were placed in class B. Furthermore, class A was divided into two subclasses A_1 and A_2, while class B was partitioned into subclasses B_1 and B_2. The anger and happy emotions were grouped together in subclass A_1 as they showed more closeness to each other. Similarly the remaining subgroups were made, except the subgroup B_2 which contains only sad emotion. In the last step, all the binary emotions were further classified to separate all the seven emotion classes. At each level of this branched structure, feature selection and classification were performed using different techniques. For each binary classification, different set of features were selected that can effectively separate the two classes.

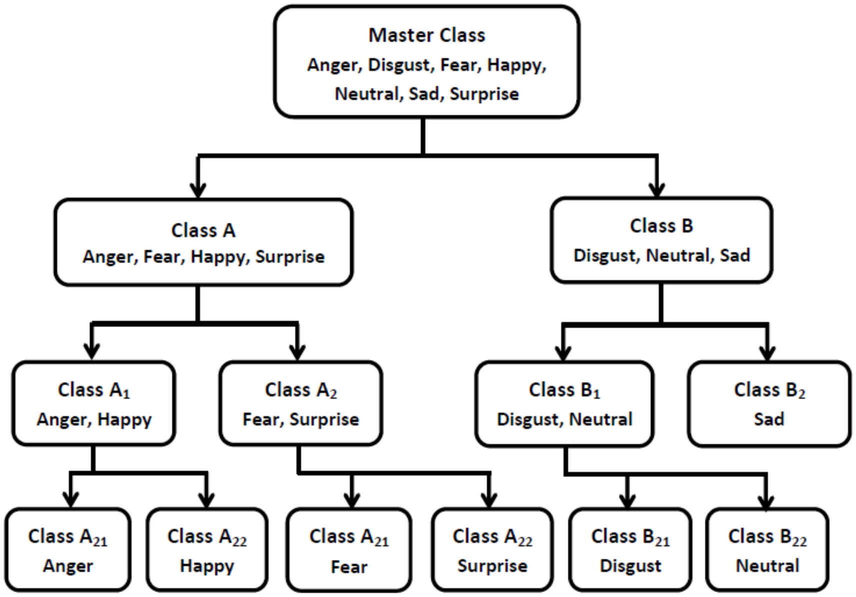

Figure 1: Hierarchical approach of classification.

Experimental Results

The classification experiments were performed using both the flat and hierarchical approaches for comparison purpose. The results were averaged over 10-fold cross validation.

Flat Approach

The classification results for the flat approach are given in Table 1. For the Gain Ratio ranker method, the best classification accuracy of 78.12% was achieved with SVM classifier using 3500 features. It was followed by other classifiers with classification accuracies of 70.83%, 63.54%, and 48.54% for Function Tree, Meta Bagging and Bayes Net, respectively. For the Info Gain attribute evaluator, the highest classification result of 77.70% was obtained for the SVM classifier using 3500 features. The classification accuracies of Function Tree, Meta Bagging and Bayes Net classifiers were 71.04%, 62.29%, and 52.50%, respectively. For the CFS method, the best classification performance of 71.45% was achieved with the SVM classifier using 103 selected features. The classification results of 70.20%, 68.75%, and 62.91% were obtained using the Function Tree, Bayes Net, and Meta Bagging classifiers.

In general, the SVM classifier provided the best classification performance, which was followed by Function Tree, Meta Bagging and Bayes Net classifiers. The overall performance of Bayes Net was poor for both the Gain Ratio and Info Gain ranker methods, but its performance improved for the CFS technique. For the ranker methods, the SVM, Function Tree, and Meta Bagging classifiers performance improved with increasing number of features, and they performed better for large number of features. On the other hand, the Bayes Net classifier provided better performance for small number of features and its performance deteriorated with the increase in number of features.

The SVM classifier provided better results for the Gain Ratio and Info Gain ranker methods as compared to CFS technique. The classification results of Meta Bagging and Functional Tree were comparable for all the feature selection techniques, i.e., Gain Ratio, Info Gain and CFS. The Bayes Net classifier performed far better for the features selected through CFS as compared to Gain Ratio and Info Gain.

Table 1: Average classification accuracy (%) for the flat approach on SAVEE database.

Attribute Selector	No. of Attributes	Classifier			
		Bayes Net	SVM	Meta Bagging	Functional Tree
Gain Ratio	500	48.54	62.70	58.12	62.70
	3000	44.58	73.33	63.54	70.00
	3500	44.79	78.12	62.70	69.58
	4000	44.79	77.29	63.54	70.83
Info Gain	150	52.50	53.95	58.95	60.00
	3000	44.58	75.00	62.29	70.62
	3500	44.58	77.70	61.66	70.20
	4000	45.00	76.87	61.45	71.04
CFS	103	68.75	71.45	62.91	70.20

Hierarchical Approach

The classification results for the hierarchical approach are given in Table 2. For the Gain Ratio attribute evaluator, the SVM classifier provided the best accuracy of 87.91% using 2500 selected features. The classification accuracies of 82.91%, 78.33% and 77.91% were achieved with Functional Tree, Bayes Net and Meta Bagging classifiers, respectively. In the case of Info Gain ranker method, the best classification performance of 87.50% was obtained with SVM Classifier using 3000 selected features. The classification results for the Functional Tree, Meta Bagging and Bayes Net were 86.45%, 79.16% and 77.91%, respectively. For the CFS technique, the best classification performance of 91.66% was achieved with Bayes Net Classifier using 75 selected features. The classification results of SVM, Functional Tree and Meta Bagging were 87.50% , 82.91% and 79.79%, respectively.

In general, the SVM classifier provided better results, which was followed by Functional Tree, Bayes Net and Meta Bagging. For the Gain Ratio and Info Gain ranker methods, the SVM classifier performed better with large number of features. On the other hand, the other classifiers provided better results with small number of features and their performance deteriorated with increase in the number of selected features. The overall performance of different classifiers improved for the CFS technique with a smaller number of selected features as compared to Gain Ratio and Info Gain methods.

The Bayes Net classifier provided far better results for the features selected through CFS as compared to Gain Ratio and Info Gain. The Functional Tree provided better classification performance for the Info Gain in comparison to Gain Ratio and CFS. The classification performance of SVM was comparable for the three feature selection techniques. The same was observed in the case of Meta Bagging.

The confusion matrix for the best classification performance of hierarchical approach using the Bayes Net classifier with CFS is given in Table 3. The best result was obtained for the neutral state followed by happy and anger emotions. The disgust emotion was confused with neutral state, while fear and surprise emotions were confused with each other. The lowest classification accuracy was obtained for the sad emotion which was confused with disgust and neutral state.

Discussion

The comparison of the best classification results obtained for the flat and hierarchical approaches are given in Table 4. It was observed that the hierarchical technique provided far better results as compared to the flat approach for all the three feature selection techniques, i.e., Gain Ratio, Info Gain and CFS. The SVM classifier performed better as compared to other classifiers for both the flat and hierarchical approaches. The best classification performance of 91.66% was achieved with hierarchical approach using Bayes Net classifier with CFS method, while the best result of 78.12% was obtained for the flat approach using SVM with Gain Ratio ranker method. For the flat approach, SVM, Meta Bagging and Functional Tree provided better classification results for large number of features, while the Bayes Net performed better for a smaller number of selected features. In the case of hierarchical approach, all the classifiers performed better for a smaller number of selected features, except SVM whose performance improved with the increase in number of selected features.

Table 2: Average classification accuracy (%) for the hierarchical approach on SAVEE database.

Attribute Selector	No. of Attributes	Classifier			
		Bayes Net	SVM	Meta Bagging	Functional Tree
Gain Ratio	100	78.33	79.16	77.91	79.37
	400	76.66	85.62	77.08	82.91
	2500	73.95	87.91	75.20	82.29
Info Gain	150	76.04	85.00	77.91	86.45
	200	77.91	86.25	77.91	85.83
	250	77.50	84.58	79.16	84.79
	3000	73.54	87.50	75.83	81.45
CFS	75	91.66	87.50	79.79	82.91

Table 3: The confusion matrix for best accuracy of hierarchical approach.

Actual emotion (Haq & Jackson, 2011)	Recognized emotion							Accuracy (%) per emotion
	A	D	F	H	N	Sa	Su	
A = Anger	55	0	0	5	0	0	0	91.66
D = Disgust	0	51	0	0	9	0	0	85.00
F = Fear	0	0	54	0	0	0	6	90.00
H = Happy	2	0	0	58	0	0	0	96.66
N = Neutral	0	1	0	0	119	0	0	99.16
Sa = Sad	0	5	0	0	5	50	0	83.33
Su = Surprise	0	0	7	0	0	0	53	88.33
Overall classification accuracy (%)								91.66

The comparison between the classification accuracy of proposed flat and hierarchical approaches, human and state-of-the-art techniques are given in Table 5. Yüncü et al. (2014) achieved an average accuracy of 73.8% for six emotions on the SAVEE database. Mao et al. (2014) reported 73.6% accuracy on SAVEE database using convolutional neural network. Liu et al. (2018) proposed a brain emotional learning model with MFCC features for emotion classification. The proposed method obtained average classification accuracy of 76.4% on SAVEE database. The human classification accuracy for the SAVEE database is 66.5%. The proposed technique results in average classification accuracy of 78.1% for the flat approach, and 91.7% for the hierarchical approach. The comparison of these results indicate that the proposed hierarchical approach outperformed the state-of-the-art methods and proposed flat approach by using different best sets of selected features to separate various sets of binary classes.

In this research, the hierarchical classification of human emotions is performed based on Bhattacharyya distance. Other distance measures such as Mahalanobis and KL-divergence also need to be investigated for in depth analysis of the hierarchical approach. The alternative distance measures may provide further improvement in the classification performance of the proposed method.

Table 4: The Comparison of best classification accuracies (%) of flat and hierarchical approaches.

Technique	Attribute selector	Classifier	No. of attributes	Accuracy (%)
Flat	Gain Ratio	SVM	3500	78.12
	Info Gain	SVM	4000	76.87
	CFS	SVM	103	71.45
Hierarchical	Gain Ratio	SVM	2500	87.91
	Info Gain	SVM	3000	87.50
	CFS	Bayes Net	75	91.66

Table 5: The comparison of classification accuracies (%) of proposed flat and hierarchical approaches, human, and state-of-the-art techniques.

Method	Accuracy (%)
Yüncü et al. (2014)	73.8
Mao et al. (2014)	73.6
Liu et al. (2018)	76.4
Human (Haq & Jackson, 2011)	66.5
Proposed flat approach	78.1
Proposed hierarchical approach	91.7

Conclusion

In this research, a hierarchical approach of emotion classification based on Bhattacharyya distance is investigated. The Bhattacharyya distance was used to find the difference between two classes, and thus enabled us to map different classes in a binary tree structure. A large set of acoustic features were extracted, which was followed by feature normalization using Z-Norm. The useful features for emotion classification were selected using three different techniques, i.e., Gain Ratio, Info Gain and CFS. The emotion classification was performed using four different classifiers, i.e., Bayes Net, SVM, Meta Bagging, and Functional Tree. The best classification performance of 91.66% was achieved with hierarchical approach using Bayes Net classifier with CFS method, while the best result of 78.12% was obtained for the flat approach using SVM with Gain Ratio ranker method. The proposed hierarchical approach outperformed the state-of-the-art methods and proposed flat approach.

The automatic emotion recognition has many potential applications including deepfake detection, assisting the autism-affected subjects, driver's safety, e-learning, and entertainment. The present study is conducted using an English database. It would be interesting to investigate the performance of the proposed technique on other databases in different languages. In addition, other distance measures such as Mahalanobis and KL-divergence also need to be investigated to further improve the classification performance of the hierarchical approach.

References

Alluhaidan, A. S., Saidani, O., Jahangir, R., Nauman, M. A., & Neffati, O. S. (2023). Speech Emotion Recognition through Hybrid Features and Convolutional Neural Network. *Applied Sciences, 13*(8), 4750.

Araño, K. A., Gloor, P., Orsenigo, C., & Vercellis, C. (2021). When old meets new: emotion recognition from speech signals. *Cognitive Computation, 13*, 771-783.

Azhagusundari, B., & Thanamani, A. S. (2013). Feature selection based on information gain. *International Journal of Innovative Technology and Exploring Engineering, 2*(2), 18-21.

Badr, Y., Mukherjee, P., & Thumati, S. (2021). Speech Emotion Recognition using MFCC and Hybrid Neural Networks. *International Joint Conference on Computational Intelligence,* (pp. 366-373).

Bhanusree, Y., Kumar, S. S., & Rao, A. K. (2023). Time-Distributed Attention-Layered Convolution Neural Network with Ensemble Learning using Random Forest Classifier for Speech Emotion Recognition. *Journal of Information and Communication Technology, 22*(1), 49-76.

Bhattacharyya, A. (1943). On a measure of divergence between two statistical populations defined by their probability distributions. *Bulletin of the Calcutta Mathematical Society, 35*(2), 99-109.

Büchlmann, P., & Yu, B. (2002). Analyzing Bagging. *The Annals of Statistics, 30*(4), 927-961.

Chalapathi, M. V., Kumar, M. R., Sharma, N., & Shitharth, S. (2022). Ensemble Learning by High-Dimensional Acoustic Features for Emotion Recognition from Speech Audio Signal. *Security and Communication Networks, 2022*, 1-10.

Er, M. B. (2020). A Novel Approach for Classification of Speech Emotions Based on Deep and Acoustic Features. *IEEE Access, 8*, 221640-221653.

Eyben, F., Wollmer, M., & Schuller, B. (2009). OpenEAR-Introducing the Munich Open-Source Emotion and Affect Recognition Toolkit.

International Conference on Affective Computing and Intelligent Interaction and Workshops (pp. 1-6). IEEE.

Gama, J. (2004). Functional trees. *Machine learning, 55*, 219-250.

Ghazi, D., Inkpen, D., & Szpakowicz, S. (2010). Hierarchical approach to emotion recognition and classification in texts. *Advances in Artificial Intelligence: 23rd Canadian Conference on Artificial Intelligence*, (pp. 40-50). Ottawa.

Haag, A., Goronzy, S., Schaich, P., & Williams, J. (2004). Emotion recognition using bio-sensors: First steps towards an automatic system. *Tutorial and research workshop on affective dialogue systems*, (pp. 36-48).

Hall, M. A. (1999). *Correlation-based feature selection for machine learning.* The University of Waikato.

Haq, S., & Jackson, P. J. (2011). Multimodal Emotion Recognition. In *Machine Audition: Principles, Algorithms and Systems* (pp. 398-423). IGI Global.

Jain, A., Nandakumar, K., & Ross, A. (2005). Score normalization in multimodal biometric systems. *Pattern Recognition, 38*(12), 2270-2285.

Jain, Y. K., & Bhandare, S. K. (2011). Min max normalization based data perturbation method for privacy protection. *International Journal of Computer & Communication Technology, 2*(8), 45-50.

Khare, S. K., Blanes-Vidal, V., Nadimi, E. S., & Acharya, U. R. (2024). Emotion recognition and artificial intelligence: A systematic review (2014–2023) and research recommendations. *Information Fusion, 102*.

Liu, S., McGree, J., Ge, Z., & Xie, Y. (2016). *Computational and Statistical Methods for Analysing Big Data with Applications.* Elsevier.

Liu, Z.-T., Xie, Q., Wu, M., Cao, W.-H., Mei, Y., & Mao, J.-W. (2018). Speech emotion recognition based on an improved brain emotion learning model. *Neurocomputing, 309*, 145-156.

Mannepalli, K., Sastry, P. N., & Suman, M. (2022). Emotion recognition in speech signals using optimization based multi-SVNN classifier. *Journal of King Saud University-Computer and Information Sciences, 34*(2), 384-397.

Mao, Q., Dong, M., Huang, Z., & Zhan, Y. (2014). Learning salient features for speech emotion recognition using convolutional neural networks. *IEEE Transactions on Multimedia, 16*(8), 2203-2213.

Mohan, M., Dhanalakshmi, P., & Kumar , R. S. (2023). Speech Emotion Classification Using Ensemble Models with MFCC. *Procedia Computer Science, 218*, 1857-1868.

Novais, R. M., Cardoso, P. J., & Rodrigues, J. M. (2022). Emotion Classification from Speech by an Ensemble Strategy. *International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-Exclusion*, (pp. 85-90). Lisbon.

Platt, J. C. (1999). Fast training of support vector machines using sequential minimal optimization. *Advances in kernel methods*, 185-208.

Tarasov, A., & Delany, S. J. (2011). Benchmarking classification models for emotion recognition in natural speech: A multi-corporal study. *IEEE International Conference on Automatic Face & Gesture Recognition*, (pp. 841-846).

Witten, I. H., Frank, E., Hall , M. A., & Pal, C. J. (2016). *Data Mining: Practical Machine Learning Tools and Techniques.* Morgan Kaufmann.

Xiao, Z., Dellandrea, E., Dou, W., & Chen, L. (2007). Hierarchical classification of emotional speech. *IEEE Transactions on Multimedia, 37*.

Yüncü, E., Hacihabiboglu, H., & Bozsahin, C. (2014). Automatic speech emotion recognition using auditory models with binary decision tree and svm. *International conference on pattern recognition* (pp. 773-778). IEEE.

Zhang, S., Li, L., & Zhao, Z. (2012). Audio-visual emotion recognition based on facial expression and affective speech. *International Conference on Multimedia and Signal Processing* , (pp. 46-52). Shanghai.