International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 2s
DOI: https://doi.org/10.17762/ijritcc.v10i2s.5924
***Article Received**: 18 October 2022 **Revised**: 25 November 2022 **Accepted**: 20 December 2022*

Affect Recognition in Human Emotional Speech using Probabilistic Support Vector Machines

Ratna Kanth Nelapati[1*], Saraswathi Selvarajan[2]
[1]Department of CSE, Koneru Lakshmaiah Education Foundation, Green Fields, Vaddeswaram, Guntur, Andhra Pradesh, India
[2]Department of Information Technology, Puducherry Technological University, Puducherry, India
* Corresponding author's Email: nratnakanth@yahoo.co.in

Abstract: The problem of inferring human emotional state automatically from speech has become one of the central problems in Man Machine Interaction (MMI). Though Support Vector Machines (SVMs) were used in several worksfor emotion recognition from speech, the potential of using probabilistic SVMs for this task is not explored. The emphasis of the current work is on how to use probabilistic SVMs for the efficient recognition of emotions from speech. Emotional speech corpuses for two Dravidian languages- Telugu & Tamil- were constructed for assessing the recognition accuracy of Probabilistic SVMs. Recognition accuracy of the proposed model is analyzed using both Telugu and Tamil emotional speech corpuses and compared with three of the existing works. Experimental results indicated that the proposed model is significantly better compared with the existing methods.

Keywords: Human Emotional Speech, Vector Machines, human-computer interaction (HCI) etc.

I. Introduction

In human-computer interaction (HCI) applications, role of the spoken language interfaces is increasing. Consequently, the prominence of recognizing emotions automatically from human speech is growing. In psychology and linguistics understanding human emotions and modeling them is one of the predominant research areas and it is gaining an increasing attraction in the engineering community. The quest to develop man machine interfaces which are very responsive and adaptive to the behavior of users serves as the major motivational force behind automatic emotion recognition from speech.

There is a growing need to understand what the information is conveyed by a user but also how that information is conveyed [1]. The significance of vocal characteristics in emotion expression and the prevailing effects of vocallyexpressed emotion on interpersonal communications is revealed by the pioneering work of Fairbank. During an interaction, knowing the speaker's emotional state helps the listeners to elicit extra information beyond what is conveyed by the lexical content of the dialogues. Particularly this can help to perceive the actualsense of speech concealed among words [2]. A study of recognizing emotions from speech was first done in 1972 and is aimed at finding the general qualitative acoustic correlations of emotions in speech. Using statistical properties of definite acoustic characteristics in the studies of emotion recognition has begun in the mid 1980's.

Lot of researchwas carried out in recent years for recognizing emotions automatically from human speech. Generally, emotion recognition systems are categorized into speaker independent emotion recognition systems and speaker dependent emotion recognition systems. But in majority of the works, speaker dependent recognition is used rather than speaker independent recognition as it is a more difficult task. Using speaker dependent recognition systems, a recognition accuracy of 70% to 90% is achieved while speaker independent systems achieved lower recognition rates [3].

To recognize emotions from speech two key phases are there. First one is determining effective speech emotion features and the second one is creating appropriate mathematical model. A good feature set can distinguish the speech emotional content efficiently. Hence, one crucial issue in recognizing emotion from speech is the elicitation of appropriate speech featureswhichare independent of the lexical content or the speaker[4].

There are two channels in human speech communication. First channel carries the lexical or verbalcontent of the conversation ("what was said"). This channel is the explicit channel. Second one is the implicit channel that contains paralinguistic information of the speech ("how it was said"). A great deal of effort has been put forth in the field of automatic speech recognitionto obtainlexical content from the speech, but more research is required for a reliable interpretation of the implicit channel.

Commonly emotion is described as the feeling of subjects over small durations of time. They may be related to persons, objects, or events. The emotional experience of humans is highly subjective in nature and is not universal.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 2s
DOI: https://doi.org/10.17762/ijritcc.v10i2s.5924
***Article Received**: 18 October 2022 **Revised**: 25 November 2022 **Accepted**: 20 December 2022*

Hence, universal and objective definitions are required for emotion. There is a long list of paralinguistic properties. A few to mention are age, voice quality, gender, emotion, dialect, stress & nervousness, charisma, alcohol or drug consumption, pathological state etc. Among the above-mentioned properties, emotion has a significant role in several applications such as identifying angry customers in call centers, get emotional feedbacks from the users in entertainment electronics, resolve linguistic ambiguities in automatic speech recognition, synthesize more natural emotional speech in text-to-speech systems [5].

Robots were once used in the industrial sector only but now they are finding their way in a wide variety of application areas. One such area is service environment, where robots are used to take care of elderly or disabled people. As these service robots are in continuous interaction with humans, Human Robot Interaction (HRI) became an important research area where the focus is on emotional interaction [3].

II. Literature Survey

Sreenivasa Rao et al [6] used prosodic features for emotion recognition. Energy, pitch, and duration were used to represent prosodic information. Prosodic features were computed for the whole sentence, words and syllables which are called global and local prosodic features respectively. Experiments were done with a Telugu emotion speech corpus (IITKGP-SESC) using SVMs for classification task.

Agnes Jacob [7] used decision tree and logistic regression models for predicting emotions from a Malayalam speech corpus. Vocal tract features were used by taking the first four formants with their band widths.Decision tree model achieved significantly better accuracy than the model based on logistic regression.

Rajisha et al [8] used short-term energy, pitch, and Mel Frequency Cepstral Coefficients (MFCCs) to recognize emotions from Malayalam speech using SVM and Artificial Neural Network (ANN) as the classifiers.

Pravena and Govind [9] performed an exclusive analysis of the excitation source. Tamil and Malayalam emotional speech corpuses were developed for the analysis. Emotionally biased utterances were used for recording, instead of emotionally neutral utterances. Experimental results indicated that emotionally biased utterances have distinguished emotions more effectively when compared with emotionally neutral utterances.

Milton and Selvi [10] used several classifiers such as k-Nearest Neighbor (KNN), Gaussian Mixture Model (GMM), Artificial Neural Network (ANN) and Support Vector Machine (SVM).The performance of Linear Prediction Coefficients (LPC) and Autoregressive (AR) parameter was analyzed and observed thatreflection coefficient features recognized emotions better when compared with LPC features.

Linhui Sun et al [11] used DNN-decision tree SVM model to infer emotions from the emotional corpus of Chinese Academy of Sciences. Some emotions are easily confused with each other. From such emotions more distinctive features are extracted by deep mining of the emotional information from speech signal. For emotion classification, a Decision Tree SVM model is created and each SVM in this model is trained by features extracted for diverse emotion groups using different DNNs.

Luefeng Chen et al [12] used speaker dependent and speaker independent features to recognize emotion from speech. A high dimensional feature set is formed by fusing these personalized and non-personalized features. Using a fuzzy C-means clustering algorithm, the high dimensional feature set is partitioned into different sub-classes. A two-layer fuzzy multiple random forest (TLFMRF) is constructed to infer emotions.

Anjali Bhavan et al [13] used spectral features and constructed a bagged ensemble of SVMs using a Gaussian kernel.

Zhen-Tao Liu [14] et al used MFCC features along with their first order δ coefficients for emotion recognition. By the inspiration of how the limbic system in brain processes the emotions, a brain emotional learning (BEL) model is deployed where the weights in the model are adjusted using a genetic algorithm. The performance of the model is tested using FAU AIBO, SAVEE and CASIA Chinese emotion databases.

An important aspect in speech emotion recognition is finding proper feature representation of speech data. Instead of using a manual feature encoding technique, Diana et al [15] adopted sparse encoding framework that can represent features automatically which has the advantage of generalizing well to the new data. Useful properties of speech are captured with the help of hierarchical sparse coding scheme which is good at differentiating between emotions.

Due to the fundamental differences in vocal tract function and vocal excitation between whispered speech and normal speech, emotion recognition systems which are exclusively modeled for normal speech produce poor results when used on whispered speech. Jun Deng et al [16] tried to alleviate this problem using three feature transfer learning

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 2s
DOI: https://doi.org/10.17762/ijritcc.v10i2s.5924
***Article Received**: 18 October 2022 **Revised**: 25 November 2022 **Accepted**: 20 December 2022*

techniques which are based on denoising auto encoders, extreme learning machines auto encoders and shared hidden layer auto encoders. Performance of the proposed model is analyzed using Berlin emotional speech corpus and Geneve whispered emotion corpus.

Human annotation of emotional speech samples is a laborious task. By combining semi-supervised learning and active learning, Zixing Zhang et al [17] proposed cooperative learning for exploiting of unlabeled data for minimizing the costly effects of human annotation. The labeling task is shared between human and machine in an efficient way. When instances are predicted with high confidence they are labeled by the machine. On the other hand, instances are labeled by human when they are predicted with low confidence value. Experiments proved that cooperative learning is superior to the individual semi-supervised learning and active learning techniques.

Dias Issa [18] employed one dimensional convolutional neural network for emotion recognition by extracting spectral contrast features, Tonnetz representation, chromagram, mel-scale spectrogram and mel-frequency cepstral coefficients from speech files. Interactive Emotional Dyadic Motion Capture (IEMOCAP), Berlin (EMO-DB), and Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) were used for the emotion recognition task by deploying an incremental model which can improve the accuracy by modifying the initial model.

III. Probabilistic SVM for Speech Emotion Recognition

Generally, information or data fusion methods can benefit any kind of parameter estimation done using multiple sources. Using fusion techniques, data from several models and correlated information that is associated with databases is combined to get improvement in accuracy and to yield more precise inferences than those could be achieved by using just one classifier. There are various levels such as feature level, decision level etc. at which fusion can be applied. When features are elicited from multiple modalities feature level fusion is applied. For example, in emotion recognition, features can be extracted from multiple modalities like face expressions, speech, gestures and postures. In such cases applying fusion techniques on the features gives better results.

More classification accuracy can be achieved by the fusion of multiple classifiers than that is achieved by the individual classifiers. Classifier fusion is also known as decision level fusion. Output of multiple classifiers is combined in decision level fusion to attain improvement in accuracy. If classifiers produce labels of the emotional classes as output, then plain majority voting or weighted majority voting can be used to predict the output class. If a classifier produces probabilistic outputs that indicate the probabilities a test instance belongs to an emotional class, then the probabilistic outputs from multiple classifiers can be combined in an efficient manner to accomplish improved accuracy in predicting the target classes.

Support Vector Machines were used in several works for emotion recognition from speech, [6], [8], [9], [10], [19], [20], [21]. But the SVMs are not used with probabilistic outputs. If a classifier produces class labels as output, then very little can be achieved with classifier fusion. On the other hand, if a classifier can produce posterior probability as the output, then much improvement in recognition accuracy is possible with classifier fusion. Consider a scenario where a classifier makes only a small part of a final decision and partial decisions of several classifiers should be merged to get the final decision. In such a scenario, posterior probabilities are required but SVMs output an uncalibrated value which is not a probabilistic value. Multiple approaches were used to produce probabilistic outputs from SVMs [22], [23], [24], [25], [26]. Wahba proposed an approach to generate probabilistic outputs from SVM [22], [23]. Vapnik suggested another method which maps the SVM outputs to probabilities [24]. Hastie and Tibshirani proposed one more method to fit probabilities to SVM outputs [25]. The authors of [27-29] have worked in deep learning and that approach could be employed for speech recognition with effective posterior probability.

One more method of fitting probabilities to the output of SVM was proposed by Platt [26] and this method is used in the present work to produce probabilistic outputs from SVMs. In this method class conditional densities $p(f \mid y)$ are not estimated. Instead, a parametric model which directly fits the posterior $P(y = 1 \mid f)$ is used. Parametric form of the sigmoid is

$$P(y = 1 \mid f) = \frac{1}{1+\exp(Af+B)} \qquad (1)$$

The above sigmoid model has two parameters A and B.

Fitting the Sigmoid

Above Eq. (1) has two parameters A and B. Maximum likelihood is estimated from the training set (f_i, y_i) and used to fit the parameters A and B. A new training set (f_i, t_i) is formed, where each t_i is a target probability defined as

$$t_i = \frac{y_i+1}{2} \qquad (2)$$

Consider the following Eq. (3).

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 2s
DOI: https://doi.org/10.17762/ijritcc.v10i2s.5924
Article Received: 18 October 2022 Revised: 25 November 2022 Accepted: 20 December 2022

$$\min - \sum_i t_i \log(p_i) + (1 - t_i) \log(1 - p_i) \quad (3)$$

where, $p_i = \frac{1}{1+\exp(Af_i+B)}$

The parameters A and B in Eq. (3) are computed by minimizing the negative log likelihood of the training data. This minimization involves minimization of two parameters. Any optimization algorithm can be used for this minimization. The parameters A and B are estimated using threefold cross validation done as follows:

a. Divide the training set into three parts.
b. Construct three SVMs where each SVM is trained using permutations of two out of three parts.
c. Evaluate the f_i 's over the leftover third part.
d. The training set for the sigmoid is the union of all three sets of f_i 's.

3.1 ONE-VS-ONE (OVO) Approach using Probabilistic SVM

Hsu et al. described One-versus-One (OVO) classifier [30]. OVO classifier is one of the standard models which is used in several works for classification task. It serves as a benchmark model to compare the performance of a new model. A One-vs-One (OVO) classifier for 7 emotions is shown in Fig. 1. OVO uses majority voting algorithm for predicting the classification of a given test instance. In this approach, r*(r-1)/2 binary classifiers are built for r emotional classes. Data from the i^{th} and j^{th} classes is used to train the binary classifier C_{ij}. When a test sample x_i is given for classification, the votes for class i are increased by one if classifier C_{ij} infers that the given sample belongs to i^{th} class. But, if the classifier C_{ij} predicts that the given sample belongs to class j, then its votes are increased by one. Finally, the test instance is given the class that got maximum number of votes.

SVMs are a good choice for the binary classification tasks in the OVO method. If r emotional classes are there in the classification task, then r*(r-1)/2 pairwise SVMs are created. Pairwise SVM $S_{i,j}$ is constructed using the training data that belongs to classes i and j. For example, the binary SVM for 1v2 is constructed using the training data of classes 1 and 2. When the test instance is given to this binary SVM it gives either 1 or 2 as output. Likewise other binary SVMs also produce their outputs.

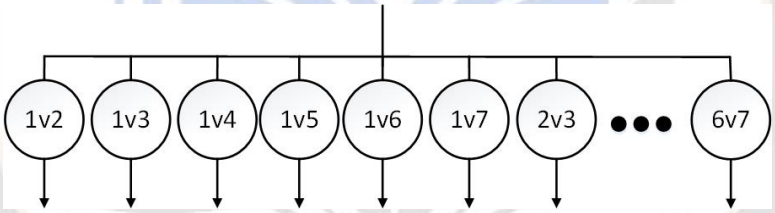

Fig. 1. One-vs-One (OVO)

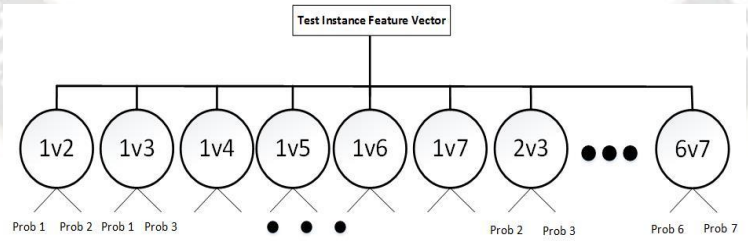

Fig. 2. Probabilistic One-vs-One model

Probabilities were fit to the SVM using the approach discussed in the previous section. This probabilistic SVM is used in place of the standard SVM for pairwise classification task in the OVO model. When a test instance is given to a probabilistic SVM constructed for pairwise classification, it produces two probabilistic values. First value indicates with what probability the test instance belongs tofirst class and second value indicates with what probability it belongs to second class. In the present work 7 emotions were considered, hence 21 pairwise probabilistic SVMs were constructed. From the outputs of these 21 SVMs a vector of length seven is formed giving the probabilities of each of the seven emotions for the given test instance. Finally, given instance is labeled with the emotion which got highest probability value.

IV. Feature Extraction from Speech Samples

4.1 Telugu Emotional Speech Corpus (TESC)

For experimental analysis of the proposed method an emotional speech corpus for Telugu language is constructed with 5 female and 5 male native Telugu speakers.Ten emotionally neutral sentences spoken in day-to-day life were

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 2s
DOI: https://doi.org/10.17762/ijritcc.v10i2s.5924
***Article Received**: 18 October 2022 **Revised**: 25 November 2022 **Accepted**: 20 December 2022*

selected for studying the emotions. Seven emotions were considered: happy, surprise, fear, sad, disgust, anger and neutral. Three recording sessions were used to build the corpus.The speech corpus consisted of a total of 2100 speech samples with each emotion having 300 samples (10 sentences x 7 emotions x 10 speakers x 3 sessions). Out of 2100 samples 700 samples were selected for the present study with 100 samples for each of the seven emotions.

4.2 Tamil Emotional Speech Corpus (TEC)

One more acted emotional speech corpus for Tamil language was also prepared using 8 (4 female 4 male) native Tamil speakers. Fifteen emotionally neutral sentences spoken in day-to-day life were selected for studying the emotions. Seven emotions were considered: happy, surprise, fear, sad, disgust, anger and neutral. Two recording sessions were used to build the corpus. The speech corpus consisted of a total of 1680 speech samples with each emotion having 240 samples (10 sentences x 7 emotions x 10 speakers x 3 sessions). Out of 1680 samples 840 samples were selected for the present study with 120 samples for each of the seven emotions.

Table 1: Feature groups, short-term features and phrase level features

Group of Features (5)	Short Term Features (23)	Phrase Level Features (15)
Cepstrum	MFCC 0-12	Minimum, Maximum, Mean, Median, Variance, Standard Deviation, minimum~maximum, minimum~mean, minimum~median, maximum~mean, maximum~median, quartile 1, quartile 3, SD/Mean, SD/Median
Pitch	Fundamental Frequency (f_0)	
Energy	Log energy Energy Entropy	
Spectral	Spectral centroid Spectral energy Spectral flux Spectral rolloff	
Zero Crossing Rate	ZCR	

4.3 Feature Extraction Process

The list of feature groups, short-term features and phrase level features used in the present work are shown in Table 1. From each speech sample, short term features listed in the second column of Table 1 are extracted. To extract the short-term features, a frame length of 20 ms and frame shift of 10 ms is used. From these short-term features, phrase level features likemaximum, minimum, mean, median, variance, standard deviation, standard deviation/median etc. are computed.

4.4 Metrics for Performance Evaluation

Precision and Recall are the two metrics used for performance evaluation of the proposed model. They are calculated using the following expressions:

$$\text{Precision} = t_p/(t_p+f_p),$$

$$\text{Recall} = t_p/(t_p+f_n),$$

where t_p is true positives, f_p is false positives and f_n is false negatives.

V. Performance Analysis of Probabilistic OVO Method

Precision and recall for speaker independent cross validation on Telugu corpus using the proposed Probabilistic One-vs-One model is shown in Table 2. Three existing approaches considered for comparison are Sreenivasa Rao et al [6], Praveena et al[9],and Milton et al [10]. Among the three existing approaches that are considered, hierarchical model by Sreenivasa Rao et al [6] is performing better than the other two approaches with a precision of 56.42% and recall of 56.25%. The proposed probabilistic OVO model performs better than the existing three approaches with a precision and recall of 58.35% and 58.02% respectively.

Table 2: Precision and Recall for Speaker Independent Testing- Telugu

TELUGU	Existing			Proposed Method
	Praveena et al	Milton et al	Sreenivasa Rao et al	
Precision	53.84	55.82	56.42	58.35
Recall	53.68	55.13	56.25	58.02

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 2s
DOI: https://doi.org/10.17762/ijritcc.v10i2s.5924
***Article Received**: 18 October 2022 **Revised**: 25 November 2022 **Accepted**: 20 December 2022*

Table 3 shows the performance for speaker dependent cross validation on Telugu corpus. From this table it can be observed that for the three existing models OVO, DAG and Hierarchical model by Milton, precision is 88.52%, 89.73% and 90.58% respectively while recall is 88.25%, 89.41% and 90.34% respectively. The proposed Probabilistic OVO model achieved a precision of 91.87% and recall of 91.42% which is higher than the existing three models considered.

Table 3: Precision and Recall for Speaker Dependent Testing- Telugu

TELUGU	Existing			Proposed Method
	Praveena et al	Milton et al	Sreenivasa Rao et al	
Precision	88.52	89.73	90.58	91.87
Recall	88.25	89.41	90.34	91.42

Table 4: Confusion matrix for Telugu Speaker Independent Cross Validation

Emotion	Anger	Disgust	Fear	Happy	Sad	Surprise	Neutral
Anger	**56**	8	0	1	0	1	4
Disgust	13	**47**	0	2	0	2	6
Fear	1	1	**37**	1	18	2	10
Happy	2	1	0	**38**	0	21	8
Sad	1	2	14	0	**40**	1	12
Surprise	4	3	0	23	0	**31**	9
Neutral	5	3	2	6	1	4	**49**

Confusion matrix for Telugu speaker independent cross validation is given in Table 4. For Telugu, out of the seven emotions anger is recognized with highest accuracy of 80%, followed by disgust with 67.14%. Surprise is the least performing emotion with 44.28%. Fear is recognized with an accuracy of 52.85%, happy with 54.28%, sad with 57.14 and neutral with 70%. From the confusion matrix, it can be inferred that that anger is confused with disgust, disgust with anger, fear with sad and neutral while surprise and happy are confusing with each other. While sad is confused with fear and neutral, neutral is partly misclassified as anger, happy and disgust.

Table 5: Precision and Recall for Speaker Independent Testing- Tamil

TAMIL	Existing			Proposed Method
	Praveena et al	Milton et al	Sreenivasa Rao et al	
Precision	58.53	60.75	61.64	62.82
Recall	58.35	60.32	61.20	62.28

Table 5 shows the precision and recall for speaker independent cross validation on Tamil corpus. Precision and recall for OVO is 58.33% and 58.35%, for DAG 60.75% and 60.32% and for Hierarchical model by Milton 61.64% and 61.20% respectively. The proposed probabilistic OVO model performs better than the existing three approaches with a precision and recall of 62.82% and 62.28% respectively.

Table 6: Precision and Recall for Speaker Dependent Testing- Tamil

TAMIL	Existing			Proposed Method
	Praveena et al	Milton et al	Sreenivasa Rao et al	
Precision	90.70	91.20	91.53	93.48
Recall	90.35	90.98	91.42	93.06

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 2s
DOI: https://doi.org/10.17762/ijritcc.v10i2s.5924
***Article Received**: 18 October 2022 **Revised**: 25 November 2022 **Accepted**: 20 December 2022*

From Table 6 we can observe that the proposed probabilistic OVO achieved highest precision and recall for speaker dependent cross validation as well. Precision and recall for proposed model are 93.48% and 93.06% respectively. For the existing models OVO, DAG and Hierarchical model precision is 90.70%, 91.20% and 91.53% respectively and recall is 90.35%, 90.98% and 91.42% respectively

.

Table 7: Confusion matrix for Tamil Speaker Independent Cross Validation

Emotion	Anger	Disgust	Fear	Happy	Sad	Surprise	Neutral
Anger	**72**	28	0	4	0	5	11
Disgust	15	**91**	0	2	0	3	9
Fear	0	1	**87**	2	19	3	8
Happy	10	7	2	**48**	0	32	21
Sad	1	0	17	0	**92**	1	9
Surprise	2	1	0	19	0	**86**	12
Neutral	15	12	5	17	3	13	**55**

Confusion matrices for Tamil speaker independent cross validation is given in Table 7. For Tamil emotion recognition, anger is recognized with an accuracy of 60%, disgust with 75.83%, fear with 72.5% and happy with 40%. Whereas sad, surprise and neutral are classified with accuracies of 76.66%, 71.66% and 45.83% respectively. Out of the seven emotions sad is recognized with highest accuracy, followed by disgust. Happy is the least identified emotion in Tamil. From the confusion matrix for Tamil given in Table 7, it can be observed that anger and disgust are confused with one another while fear is confused with sad and neutral. Happy is mostly confused with surprise and neutral, while sad is mostly misclassified as fear. Surprise is confused with happy and neutral, whereas neutral is confused with anger, happy, disgust and surprise. In both Telugu and Tamil, neutral is the emotion mostly confused for other emotions.

VI. Conclusion

SVMswere used in several works to recognize emotions from speech, but theyare not used with probabilistic outputs. This paper discussed a way to produce probabilistic outputs from an SVM followed by the discussion of modifying standard One-vs-One (OVO) model to use probabilistic SVM. Classification accuracy of the proposedmethod is analyzed both on Telugu and Tamil emotional speech corpuses. In future work this probabilistic SVMs will be used exploited for efficient speech emotion recognition using decision level fusion technique.

References

[1]. Chul Min Lee and Narayanan S.S., "Toward Detecting Emotions in Spoken Dialogs," IEEE Transactions on Speech and Audio Processing, Vol. 13, No. 2, 2005.

[2]. Jia Rong, Gang Li and Yi-Ping Phoebe Chen, "Acoustic Feature Selection for Automatic Emotion Recognition from Speech," Information Processing and Management, Vol. 45 pp. 315–328, 2009.

[3]. Eun Ho Kim, Kyung Hak Hyun, Soo Hyun Kim, and Yoon Keun Kwak, "Improved Emotion Recognition with a Novel Speaker-Independent Feature," IEEE Transactions on Mechatronics, Vol. 14, No. 3, pp. 317-325, 2009.

[4]. Lijiang Chen, Xia Mao, Yuli Xue and Lee Lung Cheng, "Speech Emotion Recognition: Features and Classification Models," Digital Signal Processing, Vol. 22, pp.1154–1160, 2012.

[5]. B.Yang and M. Lugger "Emotion Recognition from Speech Signals using a New Harmony Features" Signal Process, Vol. 90, pp. 1415-1423, 2010

[6]. K. Sreenivasa Rao, Shashidhar G. Koolagudi, Ramu Reddy Vempada, "Emotion recognition from speech using global and local prosodic features," International Journal of Speech Technology, Vol. 16, PP. 143–160, 2013.

[7]. Agnes Jacob, "Modelling speech emotion recognition using logistic regression and decision trees," International Journal of Speech Technology. Vol. 20, pp. 897–905, 2017.

[8]. Rajisha T.M., Sunija A.P., Riyas K.S., "Performance Analysis of Malayalam Language Speech Emotion Recognition System using ANN/SVM," International Conference on Emerging Trends in Engineering, Science and Technology (ICETEST- 2015).

[9]. D. Pravena, D. Govind, "Development of simulated emotion speech database for excitation source analysis," International Journal of Speech Technology, Vol. 20, pp. 327–338, 2017.

[10]. A. Milton, S. Tamil Selvi, "Class-specific multiple classifiers scheme to recognize emotions from speech

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 2s
DOI: https://doi.org/10.17762/ijritcc.v10i2s.5924
***Article Received**: 18 October 2022 **Revised**: 25 November 2022 **Accepted**: 20 December 2022*

signals," Computer Speech and Language, Vol. 28, pp. 727–742, 2014.

[11]. Linhui Sun, Bo Zou, Sheng Fu, Jia Chen, Fu Wang, "Speech emotion recognition based on DNN-decision tree SVM model," Speech Communication, Vol. 115, 2019.

[12]. Luefeng Chen, Wanjuan Su, Yu Feng, Min Wu, Jinhua She, Kaoru Hirota, "Two-layer fuzzy multiple random forest for speech emotion recognition in human-robot interaction," Information Sciences, Vol. 509, pp. 150-163, 2020.

[13]. Anjali Bhavan, Pankaj Chauhan, Hitkul, Rajiv Ratn Shah, "Bagged support vector machines for emotion recognition from speech," Knowledge-Based Systems, Vol. 184, 2019.

[14]. Zhen-Tao Liu, Qiao Xie, Min Wu, Wei-Hua Cao, Ying Mei, Jun-Wei Mao, "Speech emotion recognition based on an improved brain emotion learning model," Neurocomputing, Vol. 309, pp. 145-156, 2018.

[15]. Diana Torres-Boza, Meshia Cédric Oveneke, Fengna Wang, Dongmei Jiang, Werner Verhelst, Hichem Sahli, "Hierarchical sparse coding framework for speech emotion recognition," Speech Communication, Vol. 99, pp. 80-89, 2018.

[16]. Jun Deng, Sascha Frühholz, Zixing Zhang, Björn Schuller, "Recognizing Emotions from Whispered Speech Based on Acoustic Feature Transfer Learning," in IEEE Access, Vol. 5, pp. 5235-5246, 2017.

[17]. Zixing Zhang, Eduardo Coutinho, Jun Deng, and Björn Schuller, "Cooperative Learning and its Application to Emotion Recognition from Speech," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, Vol. 23, no. 1, pp. 115-126, Jan. 2015.

[18]. Dias Issa, M. Fatih Demirci, Adnan Yazici, "Speech emotion recognition with deep convolutional neural networks," Biomedical Signal Processing and Control, Vol. 59, 2020.

[19]. N. R. Kanth, S. Saraswathi, "Efficient speech emotion recognition using binary support vector machines & multiclass SVM", IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Madurai, 2015.

[20]. Jain, Manas et al. "Speech Emotion Recognition using Support Vector Machine." ArXiv abs/2002.07590, 2020.

[21]. K. V. Krishna, N. Sainath and A. M. Posonia, "Speech Emotion Recognition using Machine Learning," 2022 6th International Conference on Computing Methodologies and Communication (ICCMC), pp. 1014-1018, 2022.

[22]. G. Wahba, "Multivariate function and operator estimation based on smoothing splines and reproducing kernels," Proc. Nonlinear Model. Forecasting, vol. XII, pp. 95-112, 1992.

[23]. G. Wahba, "Support Vector Machines, Reproducing Kernel Hilbert Spaces, and Randomized GACV," B. Scholkopf, C.J.C. Burges, and A.J. Smola, editor, Advances in Kernel Methods - Support Vector Learning, pp. 69–88. The MIT Press, Cambridge, MA, 1999.

[24]. Vapnik, V., "Statistical Learning Theory," Wiley, New York, 1998.

[25]. Hastie, T., & Tibshirani, R., "Classification by pairwise coupling," The Annals of Statistics, Vol. 26, No. 2, pp. 451–471, 1998.

[26]. J. Platt., "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," In A. J. Smola, P. Bartlett, B. Scholkopf, and D. Schuurmans, editors, Advances in Large Margin Classifers. MIT Press, 2000.

[27]. Sandeep Pande and Manna Sheela Rani Chetty, "Analysis of Capsule Network (Capsnet) Architectures and Applications", Journal of Advanced Research in Dynamical and Control Systems, Vol. 10, No. 10, pp. 2765-2771, 2018.

[28]. Sandeep Pande and Manna Sheela Rani Chetty, "Bezier Curve Based Medicinal Leaf Classification using Capsule Network", International Journal of Advanced Trends in Computer Science and Engineering, Vol. 8, No. 6, pp. 2735-2742, 2019.

[29]. Pande S.D., Chetty M.S.R. (2021) Fast Medicinal Leaf Retrieval Using CapsNet. In: Bhattacharyya S., Nayak J., Prakash K.B., Naik B., Abraham A. (eds) International Conference on Intelligent and Smart Computing in Data Analytics. Advances in Intelligent Systems and Computing, vol 1312.

[30]. Hsu, C., Lin, C., "A comparison of methods for multi-class support vector machines," IEEE Transactions on Neural Networks, Vol. 13, No. 2, pp. 415–425, 2001.