IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

# AN AUTOMATED EMOTION RECOGNITION FROM ARABIC SPEECH USING MACHINE LEARNING TECHNIQUE

[1]Wegdan Gima Salim Al Fadahli, [2]Reham Khalifa Sulaiman Al Hinai, [3]Sherimon P.C.,

[4]Vinu Sherimon, and [5]Remya Revi K

[1,2,3] Faculty of Computer Studies, [4] Information Technology Department, [5] Computer Science and Engineering Department

[1,2,3]Arab Open University, Muscat, Sultanate of Oman

[4]University of Technology and Applied Sciences, Muscat, Sultanate of Oman

[5]Saintgits College of Engineering, Kottayam, India

***Abstract:*** An intelligence system is developed to recognize emotions of customers from Arabic speech with help of supervised machine learning method. The proposed system uses Cubic-Support Vector Machine classifier to recognize three emotions such as Happy, Angry, and Surprised from an Arabic speech. Publicly available benchmarked Arabic Natural Audio Dataset is employed for the experimental analysis. The classification model performed well with a good value of precision, recall and F1-score.

***Index Terms*** **- Intelligent system, Speech recognition, Arabic speech, customer service, machine learning.**

## I. INTRODUCTION

Remote communication method is very inevitable in today's world, especially in the field of customer service. The customer care service centers of many companies make phone call to customers to discuss about the services and products as well as to get the feedbacks. Also, the customers can call to the customer services centers to make complaints regarding the services and products and to discuss other service-related matters. The need of understanding the emotions of a customer is significant for any business growth. An intelligent speech emotion recognition method can be implemented in the customer service call centers to identify the emotions of customers [1]. This will help to recognize the unhappy customers, satisfied customers, etc.

Currently there are very little emotion recognition systems available in the Sultanate of Oman to identify the customer emotions. This project focus on the development of intelligent emotion recognition system from Arabic speech. The system can be implemented on public call centre facilities of Oman, which can be extensively used for identifying the satisfaction level and emotion of the caller and thereby to improve the services rendered to public. The objective here is to develop a speech emotion recognition system by using Machine Learning (ML) technique. In the proposed method, features extracted from the audio signal are used to train a classification model to identify the emotion of the customers.

The upcoming sections discuss about the related works, proposed method, experimental result analysis, and conclusion.

## II. RELATED WORKS

The initial activity in this research would be to identify and study similar research undertaken on speech and emotion recognition. Given below are references to similar works relating to research on analysis of speech and emotion recognition. Han et al. extracted high level features from raw audio signal using deep learning technique to identify emotions from the speech [2]. Nithya et al. proposed a deep learning-based method to classify emotions from the audio signals [3]. Kumar and Mahajan describe the usage of different machine learning techniques for Speech emotions recognition accuracy in different languages [4]. Kumar and Iqbal proposed various classification and audio signal analysis techniques to identify emotions from speech [5].

## III. PROPOSED METHOD

A supervised machine learning model is proposed to identify three emotions- Happy, Angry, and Surprised- from Arabic speech. Here, we use Arabic Natural Audio Dataset (ANAD), which contains the audio signals of the three emotions. The dataset is prepared from videos of Arabic TV Shows. The dataset contains a total number of 1,383 audios, where 505 audios in "Happy" class, 137 audios in "Surprised" class, and 741 audios in the "Angry" class [6]. Twenty-five acoustic and spectral Low Level Descriptors (LLDs) were extracted from every audio signal. The LLDs extracted are: "Fundamental frequency and its envelope, Probability of voicing, Zero-crossing rates, Intensity, Mel frequency cepstral coefficients, and Line spectral frequency" [7]. The following statical values were calculated from each extracted descriptor: "standard deviation, mean, maximum (max), minimum (min), range of max and min, absolute position of max and min, linear regression (difference and quadratic error between contour and its linear approximation, and slope and offset of the linear approximation), kurtosis, skewness, quartiles, and inter-quartile ranges". Also, the delta co-efficient of

each LLD was calculated and a total number of 950 features were obtained. A feature reduction technique called "Kruskal Wallis" test was used to decrease the number of features to 844 [8].

A Cubic-Support Vector Machine (C-SVM) classifier is used to train the extracted features for the multi-class classification problem using "one-vs-one" method [9]. The classification problem consists of three classes of emotions: Happy, Angry, and Surprised as mentioned above. The trained classification model is able to recognize the emotions of customers from their voice, and this automated system assists the call centre personals to easily deal and understand the customers.

## IV. EXPERIMENTAL ANALYSIS

The experiments are carried out on GPU based system with NVIDIA GTX 1060 card, 6.0 GB RAM using MATLAB platform. The validation criteria used to evaluate the performance of the model is a 10-fold cross validation. "Precision, Recall, and F1-score" are three evaluation metrics used for assessing the model performance, and are calculated from the confusion matrix of the model [10]. Also, we used Area Under the Receiver Operating Characteristic (AUROC) curve to evaluate the performance of the model.

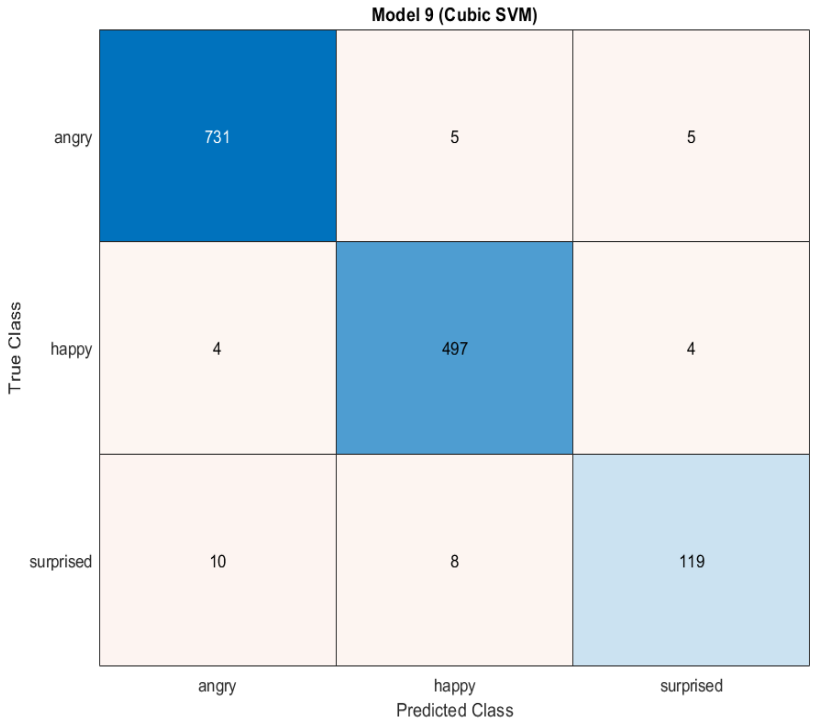
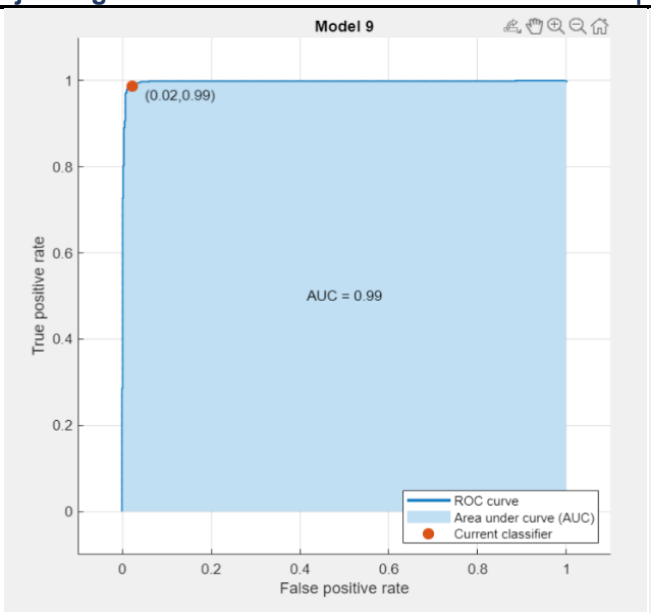



**Fig. 1 Confusion matrix**

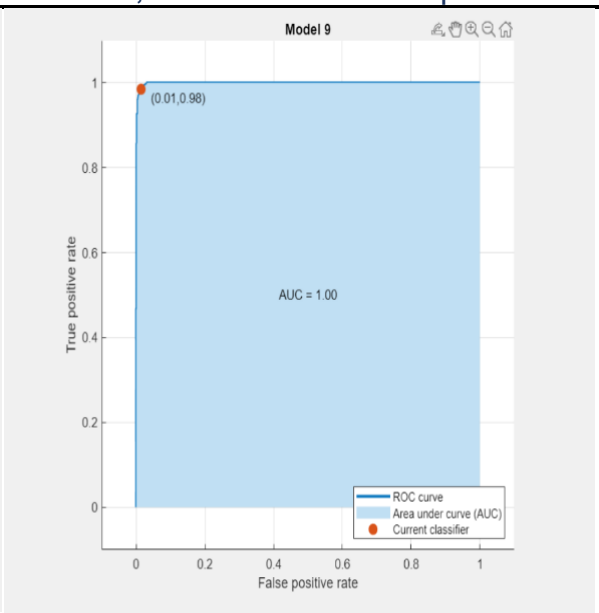
The confusion matrix of the model is given in Fig. 1, and we calculated the values of the three-evaluation metrics of each class from this matrix. The values are tabulated as given in Table 1. From the table it is clear that the class "Happy" achieved the highest F1-score of 0.983. The class "surprised" achieved only a comparatively low F1-score of 0.898, since the number of audios in this class is less compared to other two classes. Also, the Receiver Operating Characteristic for each class is obtained as shown in Figure 2. It can be observed that for every class, the value of AUC ROC is above 0.98 that shows the high distinguishing capability of the classification model to discriminate each class. Thus, these analysis shows that the proposed model achieved a good classification performance.

**Table 1 Performance evaluation**

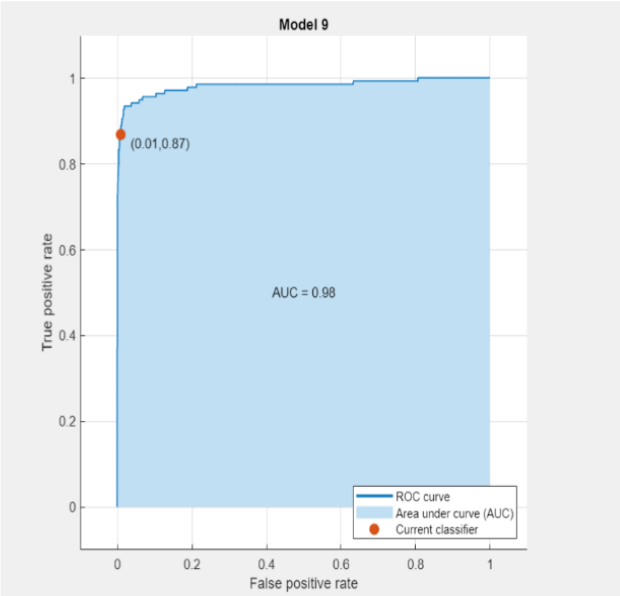
| Class | Precision | Recall | F1-score |
|---|---|---|---|
| Angry | 0.981 | 0.987 | 0.983 |
| Happy | 0.975 | 0.984 | 0.979 |
| Surprised | 0.930 | 0.869 | 0.898 |




(a) Class "Angry"




(b) Class "Happy"




(c) Class "surprised"

**Figure 2 ROC curves of the model**

## IV. CONCLUSION

In this work, a supervised machine learning classification model, C-SVM is developed to recognize the emotions from Arabic speech. The proposed model is developed to assist the customer care service centre to works efficiently by identifying the emotions of customers. Happy, Angry, and Surprised are the three emotions which are recognized using the proposed model. The model is evaluated using benchmarked Arabic Natural Audio dataset. The values of evaluation metrics shows that the model achieved a good classification performance.

## V. ACKNOWLEDGMENT

The research leading to these results has received funding from the Research Council (TRC) of the Sultanate of Oman under the Block Funding Program "BFP/URG/ICT/21/153".

## REFERENCES

**[1]** Khalil RA, Jones E, Babar MI, Jan T, Zafar MH, Alhussain T. Speech Emotion Recognition Using Deep Learning Techniques: A Review. IEEE Access. IEEE; 2019; 7:117327–45.
**[2]** Han K, Yu D, Tashev I. Speech emotion recognition using deep neural network and extreme learning machine. Proc. Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH. 2014; September (September):223–7.
**[3]** Nithya RS, Prabhakaran M, Betty P. Speech emotion recognition using deep learning. Int. J. Recent Technol. Eng. 2018;7(4):247–50.
**[4]** Kumar Y, Mahajan M. Machine learning based speech emotions recognition system. Int. J. Sci. Technol. Res. 2019;8(7):722–9.
**[5]** Kumar A, Iqbal JLM. Machine Learning Based Emotion Recognition using Speech Signal. Int. J. Eng. Adv. Technol. 2019;9(1S5):295–301.
**[6]** Klaylat S, Osman Z, Hamandi L. Enhancement of an Arabic Speech Emotion Recognition System. Int. J. Appl. Eng. Res. [Internet]. 2018a;13(5):2380–9. Available from: http://www.ripublication.com
**[7]** Klaylat S, Osman Z, Hamandi L, Zantout R. Emotion recognition in Arabic speech. Analog Integr. Circuits Signal Process. [Internet]. Springer US; 2018b;96(2):337–51. Available from: https://doi.org/10.1007/s10470-018-1142-4
**[8]** Zantout R, Klaylat S, Hamandi L, Osman Z. Ensemble models for enhancement of an arabic speech emotion recognition system [Internet]. Lect. Notes Networks Syst. Springer International Publishing; 2020. Available from: http://dx.doi.org/10.1007/978-3-030-12385-7_15
**[9]** Pradhan A. Support vector machine: A survey. Int. J. Emerg. Technol. Adv. Eng. 2012;2(8):82–5.
**[10]** Heydarian M, Doyle TE, Samavi R. MLCM: Multi-Label Confusion Matrix. IEEE Access. IEEE; 2022; 10:19083–95.