Speech Emotion Recognition with Deep Neural Networks

Mahesh Anap[1], Late Sakshi[2], Sarode Kanchan[3], Varule Mayuri[4]
Department of Electronic and Telecommunication[1,2,3,4]
Pravara Rural Engineering College, Loni, Maharashtra, India

Abstract*: The recognition and categorization of emotions is becoming increasingly important in the field of Human-Computer Interaction (HCI). Body language, including tone of voice and facial expression, can be used to identify emotions.*

Speech Emotion Recognition (SER) is one of the most popular methods for identifying emotions, according to the current study. The EMO-DB dataset is utilized in this study, while the SER dataset comprises four distinct datasets. This technique is employed because it has a high temporal resolution at a low cost and no hazards. Many academics have used SER signals in sequence during the past few decades to deal with Human-Computer Interface (HCI) and detect emotions. It entails filtering out background noise from audio signals, obtaining temporal or spectral properties from them, analyzing the signals in the time or frequency domains, and finally creating a multi-class classification plan. The method of recognizing and categorizing human emotions using auditory cues is covered in the paper.

Artificial Neural Network (ANN), KNN Classifier, Support Vector Machine (SVM), Random Forest Classifier, Convolution Network (CNN), and Decision Tree (DT) Models were among the machine learning techniques employed in the method. The experimental result that was achieved is promising, exhibiting good accuracy in the classification of emotions.

Keywords: Emotions, Audio Signal, Random Forest (RF), Support Vector Machine (SVM) Convolution Network (CNN), Decision Tree (DT), EMO-DB Dataset

REFFRENCES

[1]. Sajjad, M., & Kwon, S. (2020). Clustering-based speech emotion recognition by incorporating learned features and deep BiLSTM. IEEE Access, 8, 79861-79875.

[2]. Mitra, Ayushi. ‖ Sentiment Analysis Using Machine Learning Approaches (Lexicon based on movie review dataset). ‖ Journal of Ubiquitous Computing and Communication Technologies (UCCT) 2, no. 03 (2020): 145-152.

[3]. M. Li, X. Qiu, S. Peng et al., ─Multimodal Emotion Recognition Model Based on a Deep Neural Network with Multiobjective Optimization, ‖ Wireless Communications and Mobile Computing, vol. 2021, Article ID 6971100, 10 pages, 2021.

[4]. M. Abdel-Basset, R. Mohamed, and S. Mirjalili, A novel whale optimization algorithm integrated with Nelder-Mead simplex for multi- objective optimization problems, ‖ Knowledge- Based Systems, vol. 212, p. 106619, 2021.

[5]. W. Rahman, M. K. Hasan, and A. Zadeh, ─M- BERT: injecting multimodal information in the BERT structure, ‖ in Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 2359‒2369, 2020.

[6]. M. D. Ding and L. Li, ─CNN and HOG dual-path feature fusion for face expression recognition, ‖ Information and Control, vol. 49, no. 1, pp. 47‒54,2020.

[7]. Ullah, R.; Asif, M.; Shah, W.A.; Anjam, F.; Ullah, I.; Khurshaid, T.; Wuttisittikulkij, L.; Shah, S.; Ali, S.M.; Alibakhshikenari, M. Speech Emotion Recognition Using Convolution Neural Networks and Multi-Head Convolutional Transformer. Sensors 2023

ISSN
2581-9429
IJARSCT

[8]. Latif, S.; Rana, R.; Khalifa, S.; Jurdak, R.; Qadir, J.; Schuller, B.W. Survey of deep representation learning for speech emotion recognition. IEEE Trans. Affect. Comput. 2021, 14, 1634–1654.

[9]. Singh, P.; Srivastava, R.; Rana, K.P.S.; Kumar, A multimodal hierarchical approach to speech emotion recognition from audio and text. Knowl.-Based Syst. 2021.

[10]. Huddar, M.G.; Sannakki, S.S.; Rajpurohit, V.S. Attention-based Multi-modal Sentiment Analysis and Emotion Detection in Conversation using RNN. Int. J. Interact. Multimed. Artif. Intell. 2021.

[11]. Akçay, M.B.; Oˇguz, K. Speech emotion recognition: Emotional models, databases, features, preprocessing methods, supporting modalities, and classifiers. Speech Commun.2020

[12]. Tuncer, T.; Dogan, S.; Acharya, U.R. Automated accurate speech emotion recognition system using twine shuffle pattern and iterative neighborhood component analysis techniques. Knowl.-Based Syst. 2021, 211, 106547.

[13]. Saleem, N.; Gao, J.; Khattak, M.I.; Rauf, H.T.; Kadry, S.; Shafi, M. Deepresgru: Residual gated recurrent neural network-augmentedkalman filtering for speech enhancement and recognition.

[14]. Knowl.-Based Syst. 2022

[15]. Shilandari, A.; Marvi, H.; Khosravi, H.; Wang,

[16]. W. Speech emotion recognition using data augmentation method by cyclegenerative adversarial networks. Signal Image VideoProcess. 2022, 16, 1955–1962.

[17]. Chen, Q.; Huang, G. A novel dual attention- based BLSTM with hybrid features in speech emotion recognition. Eng. Appl. Artif.Intell.2021, 102, 104277

[18]. M. Abdel-Basset, R. Mohamed, and S. Mirjalili, A novel whale optimization algorithm integrated with Nelder-Mead simplex for multi- objective optimization problems, ‖ Knowledge- Based Systems, vol. 212, p. 106619, 2021.

[19]. W. Rahman, M. K. Hasan, and A. Zadeh, M- BERT: injecting multimodal information in the BERT structure, ‖ in Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 2359 – 2369, 2020.

[20]. M. D. Ding and L. Li, ―CNN and HOG dual-path feature fusion for face expression recognition, ‖ Information and Control, vol. 49, no. 1, pp. 47 – 54,2020.

[21]. Ullah, R.; Asif, M.; Shah, W.A.; Anjam, F.; Ullah, I.; Khurshaid, T.; Wuttisittikulkij, L.; Shah, S.; Ali, S.M.; Alibakhshikenari, M. Speech Emotion Recognition Using Convolution Neural Networks and Multi-Head Convolutional Transformer. Sensors 2023

[22]. Latif, S.; Rana, R.; Khalifa, S.; Jurdak, R.; Qadir, J.; Schuller, B.W. Survey of deep representation learning for speech emotion recognition. IEEE Trans. Affect. Comput. 2021, 14, 1634–1654.

[23]. Singh, P.; Srivastava, R.; Rana, K.P.S.; Kumar, A multimodal hierarchical approach to speech emotion recognition from audio and text. Knowl.- Based Syst. 2021.

[24]. Huddar, M.G.; Sannakki, S.S.; Rajpurohit, V.S. Attention-based Multi-modal Sentiment Analysis and Emotion Detection in Conversation using RNN. Int. J. Interact. Multimed. Artif. Intell. 2021.

[25]. Akçay, M.B.; Oˇguz, K. Speech emotion recognition: Emotional models, databases, features, preprocessing methods, supporting modalities, and classifiers. Speech Commun.2020

[26]. Tuncer, T.; Dogan, S.; Acharya, U.R. Automated accurate speech emotion recognition systemusing twine shuffle pattern and iterative neighborhood component analysis techniques. Knowl.-Based Syst. 2021, 211, 106547.

[27]. Saleem, N.; Gao, J.; Khattak, M.I.; Rauf, H.T.; Kadry, S.; Shafi, M. Deepresgru: Residual gated recurrent neural network-augmentedkalman filtering for speech enhancement and recognition.

[28]. Knowl.-Based Syst. 2022

[29]. Adversarial networks. Signal Image VideoProcess. 2022, 16, 1955–1962.