

Research Article

Audio Segmentation Techniques and Applications Based on Deep Learning

Shruti Aggarwal ¹, Vasukidevi G,² S. Selvakanmani,³ Bhaskar Pant,⁴ Kiranjeet Kaur,⁵ Amit Verma,⁵ and Geleta Negasa Binegde ⁶

¹Department of Computer Science and Engineering, Thapar University, Patiala 147004, Punjab, India

²Department of Science and Humanities, R.M.K. College of Engineering and Technology, R.S.M. Nagar, Pudukkottai, Tamil Nadu, India

³Department of Artificial Intelligence and Data Science, Velammal Institute of Technology, Velammal Knowledge Park, Chennai, Tamil Nadu, India

⁴Department of Computer Science and Engineering, Graphic Era Deemed to be University, Bell Road, Clement Town 248002, Dehradun, Uttarakhand, India

⁵University Center for Research and Development, Chandigarh University, Ajitgarh, Punjab, India

⁶Department of Computer Science, College of Engineering and Technology, Mettu University, Metu, Ethiopia

Correspondence should be addressed to Shruti Aggarwal; shruti.aggarwal@thapar.edu and Geleta Negasa Binegde; geleta.negasa@meu.edu.et

Received 22 May 2022; Revised 12 July 2022; Accepted 18 July 2022; Published 19 August 2022

Academic Editor: Punit Gupta

Copyright © 2022 Shruti Aggarwal et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Audio processing has become an inseparable part of modern applications in domains ranging from health care to speech-controlled devices. In automated audio segmentation, deep learning plays a vital role. In this article, we are discussing audio segmentation based on deep learning. Audio segmentation divides the digital audio signal into a sequence of segments or frames and then classifies these into various classes such as speech recognition, music, or noise. Segmentation plays an important role in audio signal processing. The most important aspect is to secure a large amount of high-quality data when training a deep learning network. In this study, various application areas, citation records, documents published year-wise, and source-wise analysis are computed using Scopus and Web of Science (WoS) databases. The analysis presented in this paper supports and establishes the significance of the deep learning techniques in audio segmentation.

1. Introduction

The fundamental goal of audio segmentation is to divide an audio signal into small segments so that the entities may be easily identified. Each segment contains audio data from a certain acoustic category, such as speech, animal voices, music, human activity sounds, environmental sounds, and so on [1]. The level of abstraction in audio class analysis varies depending on the deployment. For example, the radio broadcast audio signal segmentation has focused on detecting speech, silence, and other noise disturbances [2]. The general concept and process of audio segmentation are given in Figure 1. The audio stream is fed into an audio

segmentation architecture, which is an open form of architecture that can take many various kinds [3]. Traditionally, to produce meaningful and quality segments, audio signal is passed through several stages of processing, as shown in Figure 1. The output stream, which contains an adjunct data series of segment-level labels, is then transmitted to a routing switch, where each audio form is routed to the proper type of post-processing [4]. Speech parts are driven into automatic speech recognizers for linguistic or speaker role processing in broadcast transmissions, while music parts are driven into a sound effect collection library [5]. For each audio sequence, segmentation is performed by computing multiple features [6]. These characteristics are

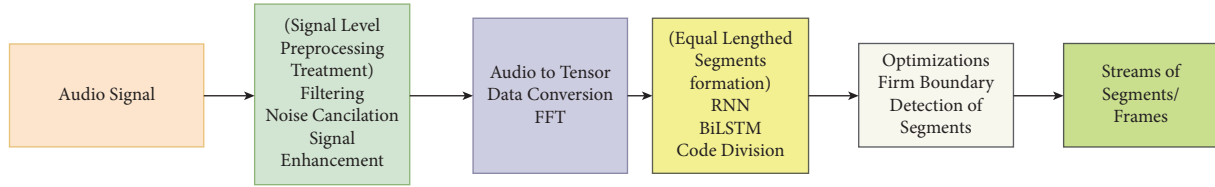


FIGURE 1: Audio segmentation process.

calculated on an audio segment, a frame, or a set of samples that is a subset of the audio segment.

In recent years, audio segmentation and deep learning have received widespread attention for research focus. Several countries and researchers have successfully applied audio segmentation techniques in various fields like speech recognition, music, or noise removal with different deep learning algorithms [7]. A study analysis literature review method was employed by analyzing articles and conferences published from 2005 to 2021 using the VOS viewer software. One hundred seventy documents were downloaded in (.CSV) file format using two keywords, “Audio Segmentation” and “Deep Learning,” from the Scopus database.

1.1. Audio Data Analysis. A sound is represented as an audio signal in which the frequency, bandwidth, decibel, and so on are the parameters. A typical audio signal can be represented as a function of Amplitude and Time [8]. Several digital devices help in the audio recording and then represent these sounds in a computer-readable way. These are some instances of these formats:

- (i) mp3 (MPEG-1 Audio Layer 3) format
- (ii) wav (Waveform Audio File) format
- (iii) WMA (Windows Media Audio) format

The extraction of acoustics features relevant to the task at hand is involved a typical audio data processing procedure followed by decision-making techniques, including detection and classification. As a result, audio data analysis is used to analyze and comprehend audio signals captured by digital equipment, with various applications in healthcare, production, and enterprise [9]. Among these, applications are customer intelligence analysis from user service calls, social-media content analysis, medical aids, patient-care systems, and public safety.

1.2. Related Work. In the task of audio segmentation, several authors have devised a segmentation approach by a classification system based on neural networks. A multilayer perceptron trained using genetic algorithms to achieve multiclass audio segmentation is an example of a feed-forward network. Many data are needed to train deep neural networks for reliable predictions [10]. Some studies have used data augmentation approaches to expand the quantity of data to overcome these problems. To tackle the data shortage problem, Raza used two approaches to enhance the amount of the dataset.

The authors suggested that the noise injection method effectively covers data shortage. Normally, data is audio data that has augmented to prevent overfitting by deliberately injecting noise; it adds random noise to the audio signal and performs audio transformation that slightly deforms the pitch and tempo [11]. When data augmentation is performed, the standard quality of the source data has a vital influence. High-quality data mean a clear signal without any other type of noise in the audio signal. However, noise is unavoidable during recordings, and every sound recording has a different length [12]. For effective analysis, noise removal is very important, and normalizing and generalizing the raw dataset is also required.

A study showed an improvement in performance by conducting denoising in the preprocessing step [13]. The authors mentioned the importance and effect of data generalization [14]. It is important to extract appropriate features for each label to classify the data according to the class. There are several methods for extracting the data, like MFCC, spectrogram, and using a deep learning network.

2. Research Trends in Web of Science Database for Audio Segmentation Based on Deep Learning

In Figure 2, we have presented the Source-Wise Analysis of audio segmentation and deep learning research trends using the Web of Science Database [15]. The experiment was conducted based on the data collected and analyzed from the Web of Science database using two keywords, “Audio Segmentation” and “Deep Learning,” from 1999–2021. Seventy-five publications selected from the Web of Science Core Collection are shown in Figure 2. Figure 2 represents the sources or the fields where the audio segmentation is used with deep learning. In Engineering Electrical Electronic, the audio segmentation has 36 publication records, the maximum values in the Web of Science Database. The second highest records are 22 documents in Computer Science Artificial Intelligence. Also, in Acoustics, the audio segmentation has 16 documents and 14 in Computer Science Information Systems [16].

In Figure 3, we represent the research work related to audio segmentation that started majorly in 2005. Until a decade, there was a very slow increase in this type of research, but post-2016, there was a sharp rise in this area of research [17]. The exponential increase in research trends can be seen in audio segmentation-related research since 2017.



FIGURE 2: Source-wise analysis of audio segmentation research trends.

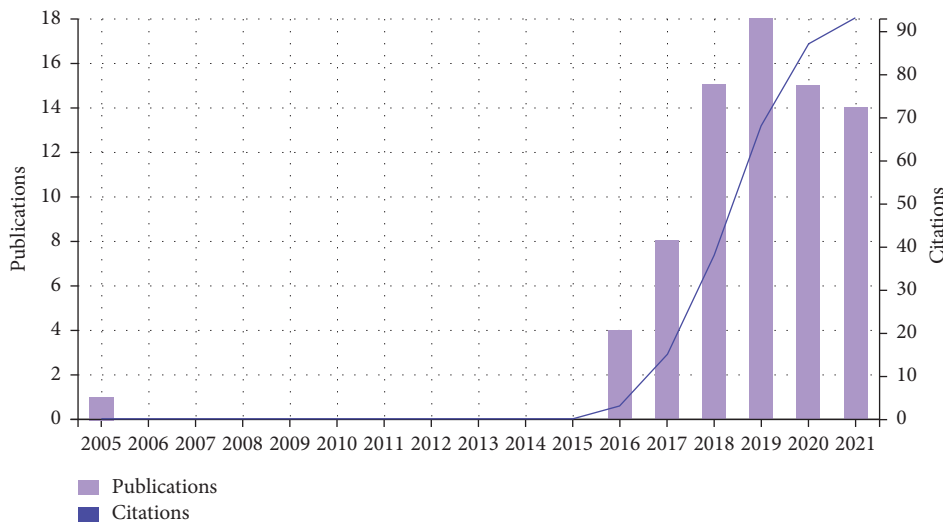


FIGURE 3: Trend analysis of audio segmentation.

It replicates that audio segmentation and deep learning field is becoming an attractive research area in the steady research zone progressively as citation number is growing previous five years [18]. It is undoubtedly realized that there is the maximum number of publications accomplished in 2019 out of the last fifteen years (2005 to 2021) in this area of research. In the current year, 2021 has witnessed considerable confidence amongst researchers regarding its application in this domain for research, that is why most publications related to this field have been published [19]. As per the trend analysis, there is a very high potential for research in this domain, as shown by the cumulative rising pattern of research for audio segmentation methods [20].

2.1. *Keywords Related to Audio Segmentation and Deep Learning.* There are six different clusters of keywords in the

co-occurrence network, which can be created using the VOS viewer software. In this network visualization, each cluster has a different colour [21]. This analysis considers the keywords that appeared in at least three collected documents. From 1310 keywords, only 119 have met the threshold represented in co-occurrence network visualization to compose the critical areas of audio segmentation and deep learning, as shown in Figure 4.

The colours red, blue, and green are shown in Figure 4 to represent co-occurrence in the related keywords. The shades of purple, orange, and so on show that the co-occurrence is a combination of two or more two domains.

2.2. *Year-Wise Publications and Research Trends.* Figure 5 shows the publication variations published from 2005 to 2021, showing the gradual increase in the number of

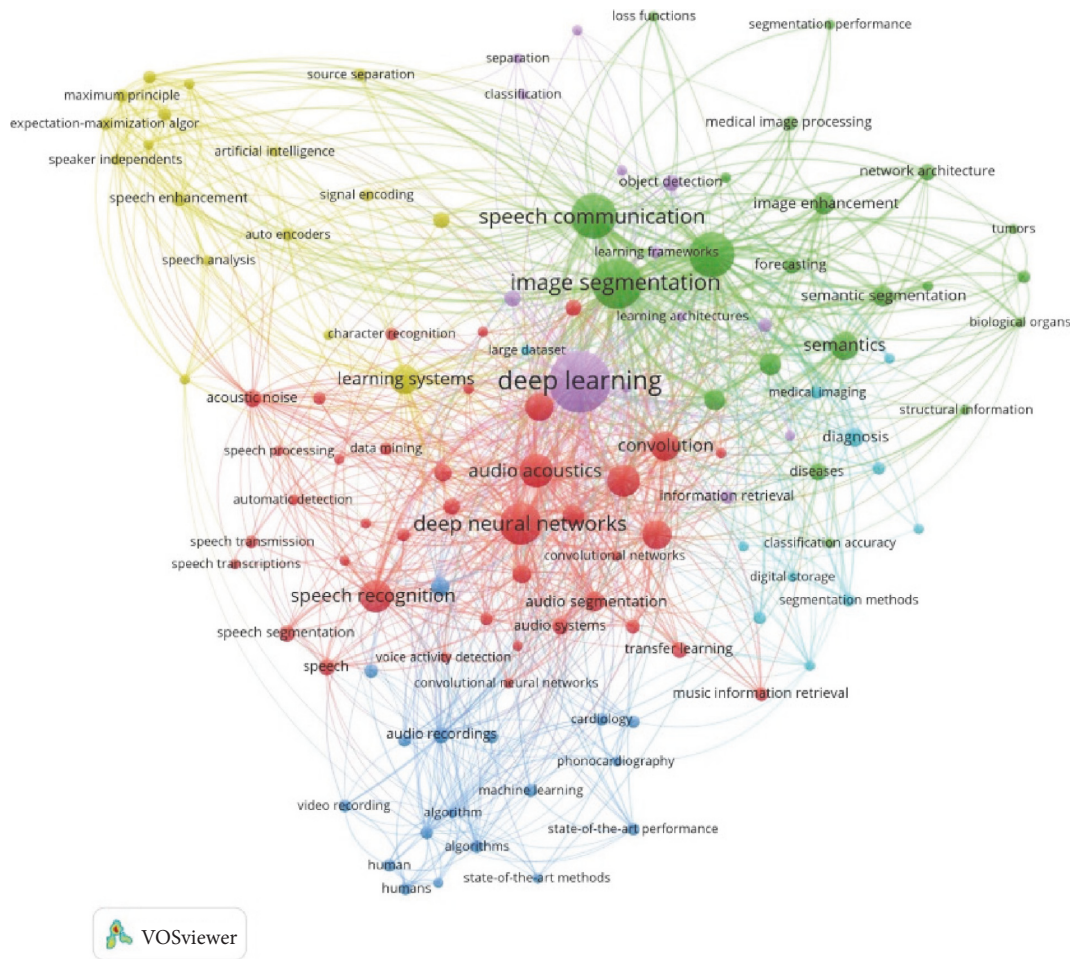


FIGURE 4: Co-occurrence keyword network on audio segmentation and deep learning.

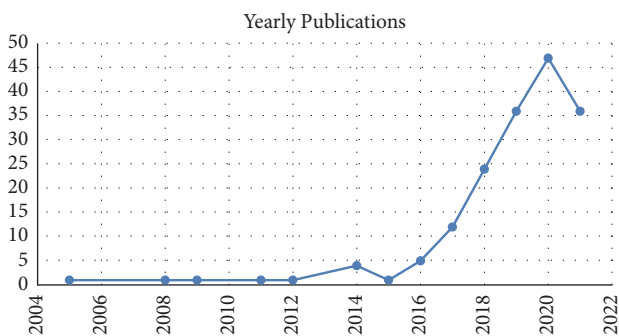


FIGURE 5: A graphical representation of the publication’s year-wise.

publications in audio segmentation based on deep learning techniques [22].

Figure 5 shows the distribution of documents published by year related to the application of audio segmentation with deep learning techniques [23]. The research related to this domain started from 2005 to 2012 gradually. Furthermore, since 2016, the number of publications dealing with audio segmentation methods has continuously increased trends in research. In this sense, it is evident that many documents have been published during the last two years (2019 and 2020), with

36 and 47 documents, respectively. At the same time, 73.52% (125) of the total publications of audio segmentation with deep learning were published in the last five years (2015–2020).

2.3. Country-Wise Research Trends for Audio Segmentation.

Many researchers globally implement audio segmentation methods. The documents published in the last 20 years come from various countries [24]. Neural networks are also used with audio segmentation, where the sum of squared errors is used to evaluate efficiency [25]. Researchers have worked globally in implementing various audio segmentation applications, as shown in Figure 6.

Table 1 shows the top 20 countries, the number of research documents related to audio segmentation, and citations. Figure 6 shows the Density Visualization for global research analysis of audio segmentation researchers based on the Scopus database [26]. As we can see, those two countries the United States and China have strong links by comparing to other countries [27]. So, their clustering is highest among the other countries [28].

As described in the given table, maximum number of publications was published in the United States [29]. These data are extracted from the Scopus database having a

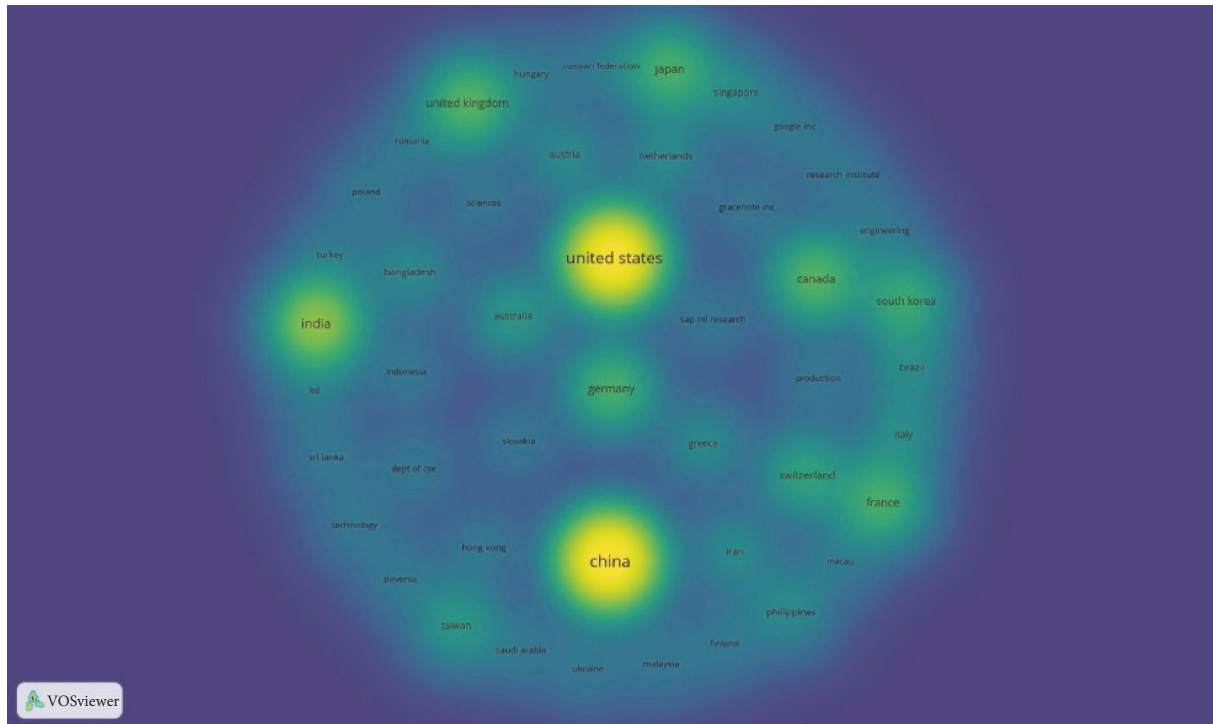


FIGURE 6: Global research trends for audio segmentation.

TABLE 1: Publications and citations (country-wise) and global occurrence analysis.

S. no.	Country	Documents	Citations
1	United States	30	220
2	China	28	247
3	India	11	26
4	Canada	7	58
5	United Kingdom	7	19
6	France	6	74
7	Germany	6	12
8	Japan	6	87
9	South Korea	5	24
10	Spain	4	6
11	Switzerland	4	11
12	Australia	3	15
13	Taiwan	3	29
14	Austria	2	42
15	Bangladesh	2	12
16	Brazil	2	0
17	Greece	2	4
18	Iran	2	15
19	Italy	2	35
20	Netherlands	2	7

minimum of three documents [30]. United States, China, and India are the top three countries, respectively, where research on audio segmentation is highest, and the entire documents and citations are as high. It clearly shows how the related research is co-related [31].

In this study, India has 11 documents and 26 citations, indicating that Indian authors are more actively involved in research based on the audio segmentation field [32]. So, most researchers are from the United States, China, and

TABLE 2: Author-wise citations for audio segmentation research trends along with year and references.

S. no.	Document	Citations	Reference
1	Zhang s. (2018)	94	[17]
2	Huang h. (2020)	63	[32]
3	Wang z. (2018)	43	[19]
4	Messner e. (2018)	42	[20]
5	Leglaive s. (2018)	34	[18]
6	Baraldi l. (2017)	33	[13]
7	Gwardys g. (2014)	25	[11]
8	Akbari m. (2019)	24	[24]
9	Lim m. (2018)	21	[21]
10	Leglaive s. (2019)	20	[25]
11	Lu w.-t. (2018)	20	[22]
12	Deng j. (2016)	16	[12]
13	Wu y. (2019)	15	[26]
14	Rahmani m.h. (2017)	15	[14]
15	Min x. (2020)	14	[33]
16	Laporte c. (2018)	14	[23]
17	Valliappan c.a. (2019)	13	[28]
18	Jati a. (2017)	13	[15]
19	Guo j. (2019)	12	[27]
20	Hossain s. (2019)	12	[29]
21	Baby a. (2017)	11	[16]
22	Leglaive s. (2020)	10	[34]
23	Hesamian m.h. (2019)	10	[30]
24	Li h. (2019)	10	[31]

India, and a lot of research potential lies in countries like Canada and the United Kingdom [33].

2.4. *Prominent Researchers for Audio Segmentation.* The publication searched from the Scopus database using two

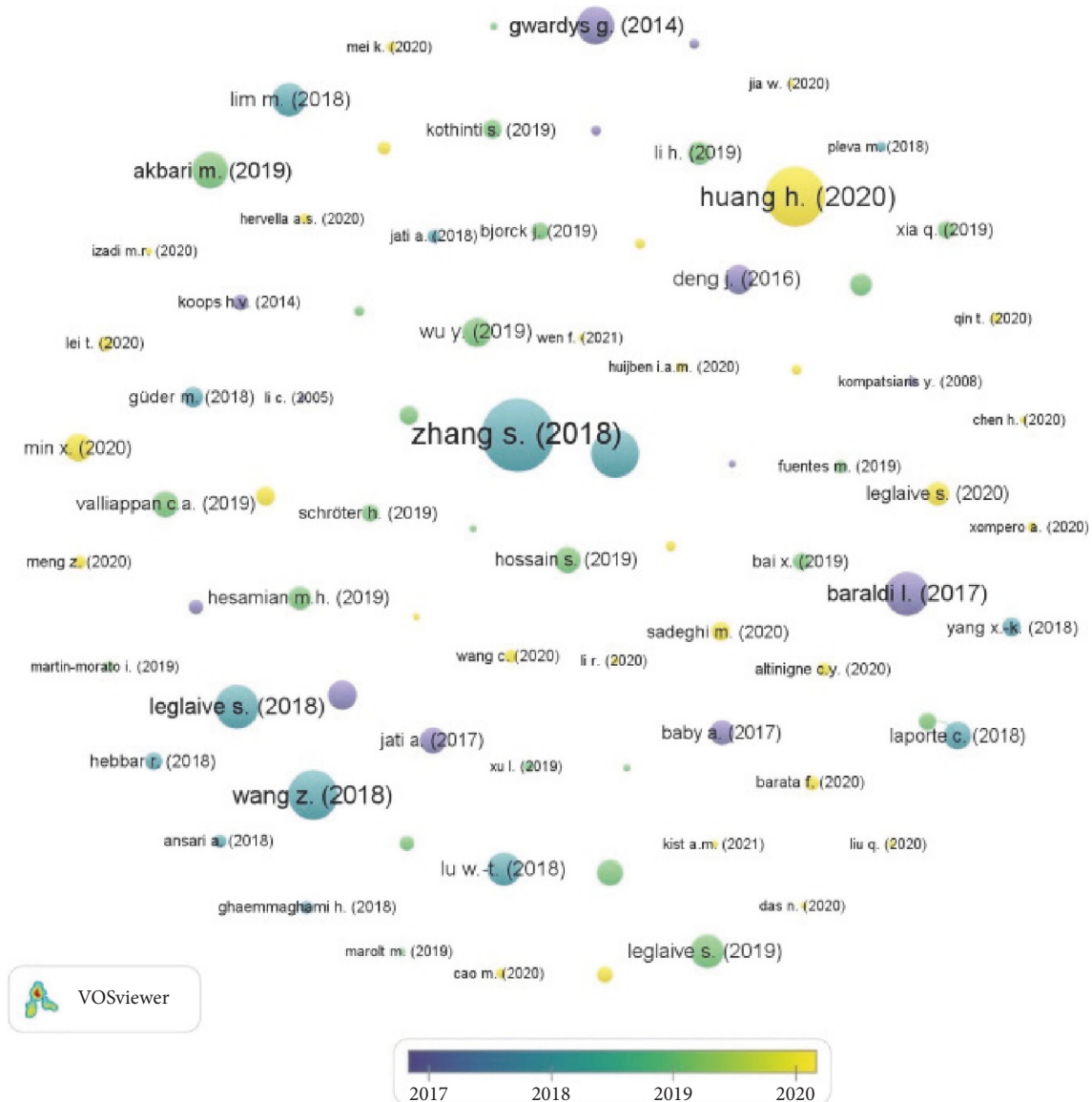


FIGURE 7: Author-wise analysis of research trends in audio segmentation.

search keywords, “Audio Segmentation” and “Deep Learning,” has been cited several times as described in Table 2. By applying the filter of a minimum of 10 citations for each document, we got 24 publications [34]. Table 2 represents the citations for 24 publications identified using the VOS viewer software package.

Figure 7 shows the author-wise analysis of audio segmentation research content based on the Scopus database. The number mentioned in author’s citations is ninety-four, the highest number in this research area. From Figure 7, we can say that Zhang s. (2018) is cited most. Huang h. (2020) is the second-largest because of the number of citations. The number of citations of Huang h. (2020) is sixty-three.

3. Application Areas of Audio Segmentation

Audio segmentation is often utilized in various applications, like Automatic Speech Recognition [35], Automatic

Language Identification, and Automatic Emotion Recognition systems [36]. The audio signal is segmented into a sequence of frames and classified into several classes like music [37], speech [38], noise [39], and so on. The noise is filtered out of the sound signal in this approach because audio recordings are significant variations, like ratio [40], audio encoding [41], bandwidth [42], language [43], speaking styles [44], gender [45], and sound pitch [46], which are the challenges.

Segmentation provides the most effective method for splitting multimedia data into digital data by extracting diverse aspects of multimedia data [47]. This segmentation yields useful information such as speaker signal and identity division, as well as automatic indexing and data retrieval of all instances of a certain speaker [48]. We can do automatic online speech recognition acoustic models to improve overall system performance by collecting all segments produced by the same speaker [49]. Typically, a certain set of

properties is extracted from each audio frame [50]. Features are used two ways: with the extracted value or changes over time [51]. It is feasible to calculate statistical features such as variance and variance using changes over time [52].

Audio segmentation is used to analyze and understand the audio recordings, with several applications in healthcare, production, and enterprise [53]. Among these, applications are customer intelligence analysis from user service calls, social-media content analysis [54], medical aids, patient-care systems, and public safety [55]. In healthcare, with the assistance of audio segmentation, a real-time cardiac arrest detection system monitors and detects any upcoming heart-related diseases [56]. It categorized the heart sound recordings into normal and abnormal heart sounds per perceived health risk. The system can provide the potential to monitor many people at a time and supply fast and effective warnings to doctors for further treatments [57].

The experiments conducted have major relevance and contribute to predicting future study domain trends. The various experiments such as source-wise, author-wise, country-wise, citation-wise, and so on give us knowledge of work conducted in this domain and the possible areas of further research [58]. The author-wise analysis gives us information about various authors working in this domain. Similarly, source-wise information helps us understand the various sources where information relevant to the work can be found.

4. Conclusions

As speech technologies' applications are progressing, audio segmentation techniques' significance is also increasing. The huge surge in the number of research articles on deep learning-based audio segmentation indicates the paramount importance of these techniques. This paper highlights the applications and source-wise research significance of audio segmentation. The analysis presented in this paper also exhibits the clear and strong relationship between audio segmentation and deep learning techniques. This work can further be extended to include the domain-specific and contextual analysis of audio segmentation techniques.

Data Availability

The data used to support the findings of this study are included in the article. Should further data or information be required, these are available from the corresponding author upon request.

Conflicts of Interest

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

Acknowledgments

The authors thank Thapar University, Punjab, for the technical assistance. The authors appreciate the support from Mettu University, Ethiopia.

References

- [1] T. Theodorou, I. Mporas, and N. Fakotakis, "An overview of automatic audio segmentation," *International Journal of Information Technology and Computer Science*, vol. 11, pp. 1–9, 2014.
- [2] E. F. Gomes and E. Pereira, "Classifying heart sounds using peak location for segmentation and feature construction," *Workshop Classifying Heart Sounds*, 2012.
- [3] S. . Lefèvre and N. Vincent, "A two-level strategy for audio segmentation," *Digital Signal Processing*, vol. 21, no. 2, pp. 270–277, 2011.
- [4] J. X. Zhang, J. Whalley, and S. Brooks, "A two-phase method for general audio segmentation," in *Proceedings of the IEEE International Conference on Multimedia and Expo*, IEEE, New York, NY, USA, July 2009.
- [5] X. Shao, C. Xu, and M. S. Kankanhalli, "Applying neural network on the content-based audio classification," in *Proceedings of the Fourth International Conference on Information, Communications and Signal Processing, and the Fourth Pacific Rim Conference on Multimedia*, IEEE, Singapore, December 2003.
- [6] A. Raza, A. Mehmood, S. Ullah, M. Ahmad, G. S. Choi, and B. W. On, "Heartbeat sound signal classification using deep learning," *Sensors*, vol. 19, no. 21, p. 4819, 2019.
- [7] K. B. Bhangale and M. Kothandaraman, "Survey of deep learning paradigms for speech processing," *Wireless Personal Communications*, vol. 125, no. 2, pp. 1913–1949, 2022.
- [8] B. Poole, J. Sohl-Dickstein, and S. Ganguli, "Analyzing noise in auto-encoders and deep networks," 2014, <https://arxiv.org/abs/1406.1831>.
- [9] V. N. Varghees and K. I. Ramachandran, "A novel heart sound activity detection framework for automated heart sound analysis," *Biomedical Signal Processing and Control*, vol. 13, pp. 174–188, 2014.
- [10] A. C. Stasis, E. N. Loukis, S. A. Pavlopoulos, and D. Koutsouris, "Using decision tree algorithms as a basis for a heart sound diagnosis decision support system," in *Proceedings of the 4th International IEEE EMBS Special Topic Conference on Information Technology Applications in Biomedicine*, pp. 354–357, IEEE, Birmingham, UK, April 2003.
- [11] C. Liu, D. Springer, Q. Li et al., "An open access database for the evaluation of heart sound algorithms," *Physiological Measurement*, vol. 37, no. 12, pp. 2181–2213, 2016.
- [12] G. Gwardys and D. M. Grzywczak, "Deep image features in music information retrieval," *International Journal of Electronics and Telecommunications*, vol. 60, no. 4, pp. 321–326, 2014.
- [13] J. Q. Deng and Y. K. Kwok, "A hybrid Gaussian-HMM-Deep learning approach for automatic chord Estimation with very large Vocabulary," in *Proceedings of the 17th International Society for Music Information Retrieval Conference*, pp. 812–818, ISMIR, New York City, United States, August 2016.
- [14] L. Baraldi, C. Grana, and R. Cucchiara, "Recognizing and presenting the storytelling video structure with deep multimodal networks," *IEEE Transactions on Multimedia*, vol. 19, no. 5, pp. 955–968, 2017.
- [15] M. H. Rahmani and F. Almasganj, "Lip-reading via a DNN-HMM hybrid system using combination of the image-based and model-based features," in *Proceedings of the 3rd International Conference on Pattern Recognition and Image Analysis (IPRIA)*, pp. 195–199, IEEE, Shahrekord, Iran, April 2017.

- [16] A. Jati and P. G. Georgiou, "Speaker2Vec: Unsupervised learning and Adaptation of a speaker Manifold using deep neural networks with an evaluation on speaker segmentation," in *Proceedings of the Annual Conference of the International Speech Communication Association*, pp. 3567–3571, INTERSPEECH, Stockholm, Sweden, August 2017.
- [17] A. Baby, J. J. Prakash, S. R. Vignesh, and H. A. Murthy, "Deep learning techniques in tandem with signal processing cues for phonetic segmentation for text to speech synthesis in Indian languages," in *Proceedings of the Annual Conference of the International Speech Communication Association*, pp. 3817–3821, INTERSPEECH, August 2017.
- [18] E. Messner, M. Zöhrer, and F. Pernkopf, "Heart sound segmentation -- an event detection approach using deep recurrent neural networks," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 9, pp. 1964–1974, 2018.
- [19] S. Zhang, S. Zhang, T. Huang, W. Gao, and Q. Tian, "Learning affective features with a hybrid deep model for audio-visual emotion recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 10, pp. 3030–3043, 2018.
- [20] Z. Wang and S. Ji, "Smoothed dilated convolutions for improved dense prediction," *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1–27, 2018.
- [21] S. Leglaive, L. Girin, and R. Horaud, "A variance modeling framework based on variational autoencoders for speech enhancement," in *Proceedings of the IEEE 28th International Workshop on Machine Learning for Signal Processing (MLSP)*, IEEE, Aalborg, Denmark, September 2018.
- [22] M. Lim, D. Lee, H. Park et al., "Convolutional neural network based audio event classification," *KSII Transactions on Internet and Information Systems (TIIS)*, vol. 12, no. 6, pp. 2748–2760, 2018.
- [23] W. T. Lu and L. Su, "Vocal Melody extraction with semantic segmentation and audio-symbolic domain transfer learning," in *Proceedings of the 19th International Society for Music Information Retrieval Conference*, pp. 521–528, ISMIR, September 2018.
- [24] C. Laporte and L. Ménard, "Multi-hypothesis tracking of the tongue surface in ultrasound video recordings of normal and impaired speech," *Medical Image Analysis*, vol. 44, pp. 98–114, 2018.
- [25] M. Akbari, J. Liang, and J. Han, "DSSLIC: deep semantic segmentation-based layered image compression," in *Proceedings of the ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2042–2046, IEEE, Brighton, UK, May 2019.
- [26] S. Leglaive, U. Şimşekli, A. Liutkus, L. Girin, and R. Horaud, "Speech enhancement with variational autoencoders and alpha-stable distributions," in *Proceedings of the ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 541–545, IEEE, Brighton, UK, May 2019.
- [27] Y. Wu and W. Li, "Automatic audio chord recognition with MIDI-trained deep feature and BLSTM-CRF sequence decoding model," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 2, pp. 355–366, 2019.
- [28] J. Guo, B. Song, P. Zhang, M. Ma, W. Luo, and J. Lv, "Affective video content analysis based on multimodal data fusion in heterogeneous networks," *Information Fusion*, vol. 51, pp. 224–232, 2019.
- [29] C. A. Valliappan, A. Kumar, R. Mannem, G. R. Karthik, and P. K. Ghosh, "An improved air tissue boundary segmentation technique for real-time magnetic resonance imaging video using segnet," in *Proceedings of the ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, Brighton, UK, May 2019.
- [30] S. Hossain, S. Najeeb, A. Shahriyar, Z. R. Abdullah, and M. A. Haque, "A pipeline for lung tumor detection and segmentation from ct scans using dilated convolutional neural networks," in *Proceedings of the ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, IEEE, Brighton, UK, May 2019.
- [31] M. H. Hesamian, W. Jia, X. He, and P. J. Kennedy, "Atrous convolution for binary semantic segmentation of lung nodule," in *Proceedings of the ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 1015–1019, IEEE, Brighton, UK, May 2019.
- [32] H. Li, D. Chen, W. H. Nailon, M. E. Davies, and D. Laurenson, "A deep dual-path network for improved mammogram image processing," in *Proceedings of the ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 1224–1228, IEEE, Brighton, UK, May 2019.
- [33] H. Huang, L. Lin, R. Tong et al., "Unet 3+: a full-scale connected unet for medical image segmentation," in *Proceedings of the ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 1055–1059, IEEE, Barcelona, Spain, May 2019.
- [34] X. Min, G. Zhai, J. Zhou, X. P. Zhang, X. Yang, and X. Guan, "A multimodal saliency model for videos with high audio-visual correspondence," *IEEE Transactions on Image Processing*, vol. 29, pp. 3805–3819, 2020.
- [35] S. Leglaive, X. Alameda-Pineda, L. Girin, and R. Horaud, "A recurrent variational autoencoder for speech enhancement," in *Proceedings of the ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 371–375, IEEE, Barcelona, Spain, May 2020.
- [36] G. Tzanetakis and P. Cook, "Multi-feature audio segmentation for browsing and annotation," in *Proceedings of the 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. WASPAA'99 (Cat. No. 99TH8452)*, pp. 103–106, IEEE, New Paltz, NY, USA, October 1999.
- [37] S. Venkatesh, D. Moffat, and E. R. Miranda, "Investigating the effects of training set synthesis for audio segmentation of radio broadcast," *Electronics*, vol. 10, no. 7, p. 827, 2021.
- [38] S. A. Deevi, C. P. Kaniraja, V. D. Mani, D. Mishra, S. Ummar, and C. Satheesh, "HeartNetEC: a deep representation learning approach for ECG beat classification," *Biomedical Engineering Letters*, vol. 11, no. 1, pp. 69–84, 2021.
- [39] S. Suyanto, K. N. Ramadhani, S. Mandala, and A. Kurniawan, "Automatic segmented-Syllable and deep learning-based Indonesian Audiovisual speech recognition," in *Proceedings of the 6th International Conference on Interactive Digital Media (ICIDM)*, pp. 1–4, IEEE, Bandung, Indonesia, December 2020.
- [40] F. Barata, P. Tinschert, F. Rassouli et al., "Automatic recognition, segmentation, and sex assignment of nocturnal asthmatic coughs and cough epochs in smartphone audio recordings: observational field study," *Journal of Medical Internet Research*, vol. 22, no. 7, Article ID e18082, 2020.
- [41] O. Stephen and M. Sain, "Deep learning-based Scene image detection and segmentation with speech synthesis in real-time," in *Smart Healthcare Analytics in IoT Enabled Environment*, pp. 163–171, Springer, Cham, 2020.
- [42] C. Park, D. Kim, and H. Ko, "Dilated convolution and gated linear unit based sound event detection and tagging algorithm

- using weak label,” *The Journal of the Acoustical Society of Korea*, vol. 39, no. 5, pp. 414–423, 2020.
- [43] M. F. M. Esa, N. H. Mustafa, N. H. M. Radzi, and R. Sallehuddin, “Audio Deformation based data augmentation for convolution neural network in Vibration analysis,” *IOP Conference Series: Materials Science and Engineering*, vol. 551, no. 1, Article ID 012066, 2019.
- [44] A. Sendrayaperumal, S. Mahapatra, S. S. Parida et al., “Energy Auditing for efficient planning and Implementation in Commercial and Residential Buildings,” *Advances in Civil Engineering*, vol. 2021, pp. 1–10, 2021.
- [45] L. P. Natrayan, S. S. Sundaram, and J. Elumalai, “Analyzing the Uterine physiological with MMG signals using SVM,” *International journal of pharmaceutical research*, vol. 11, no. 2, pp. 165–170, 2019.
- [46] K. Seeniappan, B. Venkatesan, N. N. Krishnan et al., “A comparative assessment of performance and emission characteristics of a DI diesel engine fuelled with ternary blends of two higher alcohols with lemongrass oil biodiesel and diesel fuel,” *Energy & Environment*, vol. 13, Article ID 0958305X2110513, 2021.
- [47] K. R. Vaishali, S. R. Rammohan, L. Natrayan, D. Usha, and V. R. Niveditha, “Guided container selection for data streaming through neural learning in cloud,” *International Journal of System Assurance Engineering and Management*, vol. 16, pp. 1–7, 2021.
- [48] G. Kanimozhi, L. Natrayan, S. Angalaeswari, and P. Paramasivam, “An effective Charger for plug-in hybrid Electric Vehicles (PHEV) with an enhanced PFC Rectifier and ZVS-ZCS DC/DC high-frequency Converter,” *Journal of Advanced Transportation*, vol. 2022, Article ID 7840102, 2022.
- [49] S. Kaliappan, M. D. Raj Kamal, S. Mohanamurugan, and P. K. Nagarajan, “Analysis of an Innovative Connecting Rod by using finite Element method,” *Taga Journal Of Graphic Technology*, vol. 14, pp. 1147–1152, 2018.
- [50] D. K. Jain, S. K. S. Tyagi, S. Neelakandan, M. Prakash, and L. Natrayan, “Metaheuristic optimization-based Resource Allocation technique for Cybertwin-driven 6G on IoE environment,” *IEEE Transactions on Industrial Informatics*, vol. 18, no. 7, pp. 4884–4892, 2022.
- [51] P. Asha, L. Natrayan, B. T. Geetha et al., “IoT enabled environmental toxicology for air pollution monitoring using AI techniques,” *Environmental Research*, vol. 205, Article ID 112574, 2022.
- [52] A. S. Kaliappan, S. Mohanamurugan, and P. K. Nagarajan, “Numerical Investigation of Sinusoidal and Trapezoidal piston profiles for an IC engine,” *Journal of Applied Fluid Mechanics*, vol. 13, no. 1, pp. 287–298, 2020.
- [53] S. S. Sundaram, N. Hari Basker, and L. Natrayan, “Smart clothes with bio-sensors for ECG monitoring,” *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, no. 4, pp. 298–301, 2019.
- [54] K. Nagarajan, A. Rajagopalan, S. Angalaeswari, L. Natrayan, and W. D. Mammo, “Combined Economic emission Dispatch of Microgrid with the Incorporation of Renewable Energy sources using improved Mayfly optimization algorithm,” *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1–22, 2022.
- [55] S. Magesh, V. R. Niveditha, P. S. Rajakumar, S. Radha Ram Mohan, and L. Natrayan, “Pervasive computing in the context of COVID-19 prediction with AI-based algorithms,” *International Journal of Pervasive Computing and Communications*, vol. 16, no. 5, pp. 477–487, 2020.
- [56] C. S. S. Anupama, L. Natrayan, E. Laxmi Lydia et al., “Deep learning with backtracking search optimization-based skin lesion diagnosis model,” *Computers, Materials & Continua*, vol. 70, no. 1, pp. 1297–1313, 2021.
- [57] S. Raja and A. J. Rajan, “A decision-making model for selection of the Suitable FDM Machine using Fuzzy TOPSIS,” *Mathematical Problems in Engineering*, vol. 2022, Article ID 7653292, 2022.
- [58] S. Aggarwal, M. Suchithra, N. Chandramouli et al., “Rice disease detection using Artificial and Machine learning techniques to Improve Agro-Business,” *Journal of Scientific Programming*, vol. 2022, Article ID 1757888, 2022.