Investigating Graph-based Features for Speech Emotion Recognition

Anastasia Pentari[1], George Kafentzis[2] and Manolis Tsiknakis[3,4]
[1]*Postdoctoral Researcher*, *Computational BioMedicine Laboratory, Foundation for Research and Technology-Hellas*
[2]*Postdoctoral Researcher*, *Computer Science Department, University of Crete*
[3]*Professor of Biomedical Informatics and eHealth*, *Department of Electrical and Computer Engineering, Hellenic Mediterranean University*
[4]*Affiliated Researcher*, *Institute of Computer Science*
Heraklion, Greece
emails: anpentari@ics.forth.gr, kafentz@csd.uoc.gr, tsiknaki@ics.forth.gr

2022 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI) | 978-1-6654-8791-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/BHI56158.2022.9926795

***Abstract*—During the last decades, automatic speech emotion recognition (SER) has gained an increased interest by the research community. Specifically, SER aims to recognize the emotional state of a speaker directly from a speech recording. The most prominent approaches in the literature include feature extraction of speech signals in time and/or frequency domain that are successively applied as input into a classification scheme. In this paper, we propose to exploit graph theory and structures as alternative forms of speech representations. We suggest applying the so-called Visibility Graph (VG) theory to represent speech data using an adjacency matrix and extract well-known graph-based features from the latter. Finally, these features are fed into a Support Vector Machine (SVM) classifier in a leave-one-speaker-out, multi-class fashion. Our proposed feature set is compared with a well-known acoustic feature set named the Geneva Minimalistic Acoustic Parameter Set (GeMAPS). We test both approaches on two publicly available speech datasets: SAVEE and EMOVO. The experimental results show that the proposed graph-based features provide better results, namely a classification accuracy of 70% and 98%, respectively, yielding an increase by 29.2% and 60.6%, respectively, when compared to GeMAPS.**

***Index Terms*—Affective Computing, Emotion Recognition, Speech Analysis, Visibility Graph Theory, Graph-based Features**

I. Introduction

In the last few decades, speech-based analyses contribute to the recognition of cognitive, affect, and emotional states of speakers [1]. Affective states expressed through voice have gained the research interest of various fields, including healthcare [2]–[6], engineering [7], [8], human-computer interaction [9], [10], and emergency situations [11], [12]. Among many affect-related computing tasks, speech emotion recognition (SER) is the art and science of automatically, robustly, and effectively detecting the emotional state of a speaker via appropriate engineering methods and tools [13], [14].

Although speech is a natural means of communication, it still remains very complex to be analyzed, due to the heterogeneity in languages and speakers, among others, leading to methods to appropriately encode speech in a set of meaningful parameters, the so-called features. However, even though there exist a variety of speech-based features which derive from the analysis of speech signals mostly in the time and frequency domains, as well as many classification methods which aim to recognize the emotional states in a precise manner, the SER problem is yet to be solved [15]. Reasons for this include language (tonal vs non-tonal), emotional space (discrete vs continuous), speakers (acted vs non-acted), and recording conditions (in-the-wild vs controlled). In most approaches, three main characteristics of speech have been considered: phonation, articulation, and prosody, thus, leading to a variety of speech-based features [16].

Recently, the graph theory has emerged as a powerful tool to process and analyze signals derived from neuroscience [17] to music [18] sources. Many graph-based theories have been proposed, based on the application, with the so-called *Visibility Graph* (VG) theory to constitute a simple and computationally fast method of converting time series into graph representations. The main idea behind the VG procedure is to map a time series into a network, by exploiting the geometric structure of the signal's amplitudes [19]. This is a way of constructing the most important quantity of graph theory, i.e., the so-called *adjacency matrix*. The adjacency matrix shows the interrelations between pairs of objects. The interrelations are denoted by its edges whilst the objects are described as nodes of a graph. On the other hand, there exists a variety of graph-based characteristics, including the most well-known, which are the *degree of connectivity* (DoC), the *density* (D), the *modularity* (M), the *clustering coefficient* (CC), the *shortest path length* (L) and the *small-world coefficient* (S). In total, these graph-based characteristics have been proved effective in differentiating signals, such as in music genre classification [18].

However, even if the graph theory provides robust tools for analyzing signals, few works focus on speech signal processing [20]–[22]. In this work, we investigate whether: (i) the VG theory, as a means of representing the speech signals, can be proved suitable for constructing the adjacency matrix and (ii) the most prominent mentioned graph-based characteristics, derived from this adjacency matrix, can provide an effective feature set for recognizing primary emotional states, such as anger, happiness, sadness, surprise and neutral.

Authorized licensed use limited to: Hellenic Mediterranean University. Downloaded on April 01,2025 at 22:22:09 UTC from IEEE Xplore. Restrictions apply.

To the best of our knowledge, this is the first application of graph-based feature sets on SER. Our proposed methodology is inspired by [18], where the authors aim to analyze music songs using graph-based theory.

The rest of the paper is organized as follows: Section II is a description of the previous work. In Section III we introduce the two building blocks of our proposed methodology, i.e., how we exploited the VG theory on the speech-based analysis and the most well-known graph-based characteristics used in terms of the emotional states classification. In Section IV we present the experimental evaluations made on the two publicly available datasets, i.e., EMOVO and SAVEE [23], [24]. In conclusion, Section V summarizes the main outcomes and gives directions for further work.

II. Related Work

Throughout the last two decades, the research community has shown an increased interest in the automatic identification of emotion, through mathematical approaches and computational analyses of speech signals and facial expressions [25], emphasising on the application of machine learning techniques [26]. However, a main issue of the emotion recognition task is to extract the most important information from speech signals. To this end, feature extraction has proved to be an effective way to derive the significant characteristics from the signals. Under this perspective, researchers have created a variety of voice features sets, by exploiting the speech signals' representations in time and frequency domain [1].

Specifically, spectrogram-based speech features were used in [27], [28], which were proved promising in recognizing the emotional states of actor-based datasets analysis. However, a main limitation is that no further investigation was made in free context speech signals. On the other hand, a group of features, the so-called *Mel frequency Cepstral coefficients* (MFCCs) have been shown to be effective in the analysis and automatic detection of emotional states [29] and they probably constitute one of the most widely used features for this task. Not surprisingly, deep learning methods have been recently introduced that utilize either feature sets, spectral representations, or even raw waveform data [38]. Such models show significant performance improvement but require resources for training, large datasets, and technical expertise in the deep learning area. Moreover, additional and well-known feature sets were introduced in the Interspeech Paralinguistics Challenges [30], [31]. Although these feature sets exploit significant speech information in various domains, the large number of the characteristics and their combinations increase the process complexity. Recently, in [32], the authors analyzed speech signals using a combination of various features, thus creating the well-established set of features known as the *Geneva Minimalistic Acoustic Parameter Set* (GeMAPS), as well as its extended version, eGeMAPS. The eGeMAPS is one of the most widely used speech feature set and the one that we compare with in this work.

III. Methodology

This section introduces the main graph-based theories employed in terms of our analysis, namely, the visibility graph theory and the extraction of the graph-based characteristics.

A. Preprocessing

Before the description of our main building blocks, the speech signals passed through a preprocessing stage. Firstly, since our evaluation was based on acted speech datasets, no denoising, enhancement, or filtering processes were applied. Finally, natural pauses that exist in these recordings constitute significant speech information in many mental diseases, such as Parkinson's disease, and thus, they must be taken into consideration.

An important preprocessing step is to transform the speech signal samples into a more compact time series representation. Since the visibility graph theory processes each time series' sample, utilizing the original speech waveform would result into a computationally demanding procedure. As a consequence, instead of taking into account the large number of signals' samples, we split each one of them into overlapping rectangular frames, of length 9 ms (which correspond to 2000 samples in our datasets) and overlap equal to 4.5 ms (that is, 1000 samples). From each frame, we compute the zero-mean Root Mean Square energy (RMS-energy), as a non-negative measure of the amplitudes values inside each frame. This sliding windows process led to the creation of new, compact time series $\mathbf{x} \in \mathrm{R}^{1\times N}$, of length N for each speech signal. Hereafter, the rest of our analysis would be on such compact time series.

B. Graph Construction: Visibility Graph Theory

A graph $G = (V, E)$ is a comprehensive representation of a signal, consisting of $|V \times V|$ number of nodes and $|E|$ number of edges. As mentioned, the sliding windows method is applied on each speech signal and the zero-mean RMS-energy is computed from each window. The main reason is that the visibility graph theory requires that the input time series' values should be positive. The use of RMS-energy is motivated by its relation to speech intensity, which is successively related to pitch and duration, according to psychoacoustic models [1]. Having computed these new time series, we applied the visibility graph theory.

The idea behind the VG theory is to represent each new time series sample by a node. The interrelations between two nodes describes the geometric structure of two samples. In more detail, The VG procedure is iterative, where at each iteration the geometric interrelations among $2k$ samples of the new time series are estimated. Geometric interrelation denotes the capability of a sample $i \in 1, \ldots, N$, in the range (if such exists) of $[i-k, i+k]$ with $1 \leq k \leq N$, to detect all its visible neighboring samples, i.e., to "see" all the samples it can, unless an obstacle appears. For each sample that it can observe, i.e., the i-th amplitude is greater than the j-th, where $j \in [i-k, i+k]$, the corresponding position (i, j) of the extracted visibility graph equals one, otherwise, if the j-th

Authorized licensed use limited to: Hellenic Mediterranean University. Downloaded on April 01,2025 at 22:22:09 UTC from IEEE Xplore. Restrictions apply.

element has greater value than the i-th, the latter's visible capability ends and this corresponds to zero value at the (i, j) position. The extracted visibility graph is denoted as $\mathbf{A} \in \mathrm{R}^{N \times N}$. It is worth mentioning that the right and left neighbors of the i-th element are always visible.

For instance, suppose that our time series consists of the elements $[3, 2, 4, 1]$. This implies that the extracted VG adjacency matrix would be of size (4×4). At the first iteration we have to apply the VG theory for the first element of this time series. Specifically, the first row of the extracted matrix would be equal to $[0, 1, 1, 0]$. Repeating the VG process for all elements, the final adjacency matrix would be the following:

$$\begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

It should be noted that the diagonal of the adjacency matrix consists of zero elements. Moreover, it can represent the input time series either in binary or in weighted form. However, in our experiments, the adjacency matrix was in binary version, and the shift value k was equal to 2.

C. Graph-based Characteristics

The graph-based structures are totally different representations from the signals in time or frequency domain. Therefore, the graph-based characteristics describe the signal properties in a manner which mainly depends on how the adjacency matrix is constructed. The most standard graph-based features, and the ones proposed in our analysis, are the following:

- **Degree of connectivity**

The degree of connectivity (DoC) denotes the number of edges that are immediately connected to a specific node. Equation (1) computes the degree of connectivity measure [34]:

$$DoC = \sum_{i=1}^{N} \sum_{j=1}^{N} \mathbf{A}_{i,j} \ , \tag{1}$$

where $\mathbf{A}$ is the adjacency matrix derived from the visibility graph procedure.

- **Density**

Density (D) is a graph-based measure which implies how sparse or dense a graph is [35]. A dense graph is a graph which has a number of edges close to the maximal number of edges. Its computation derives from the following equation:

$$D = \frac{2|E|}{N(N-1)} \ , \tag{2}$$

where $|E|$ denotes the number of the edges included in the graph.

- **Modularity**

Modularity (M) is a graph measure which describes the graph's strength of division, i.e., its tendency to be split into clusters, the so-called modules. The connectivity in each module is dense whilst between two modules is sparse [18]. For instance, a high modularity value can characterize a graph with a tied structure, i.e., a closely connected community of nodes. Its computation is described in more detail in [18].

- **Clustering coefficient**

The clustering coefficient (CC) denotes the tendency of a node to create cliques [35]. In more detail, CC characterizes the ability of a node to cluster together with other nodes. The higher the CC, the higher the nodes tendency to create groups of nodes. In our implementation, the CC's computation was based on the following equation.

$$CC_i = \frac{1}{k_i(k_i - 1)} \sum_{j=1,l=1}^{N} \mathbf{A}_{i,j}\mathbf{A}_{j,l}\mathbf{A}_{l,i} \ . \tag{3}$$

where $k_i = \sum_{j=1}^{N} \mathbf{A}_{i,j}$ is the number of edges that are connected to node i. Simply put, CC is the number of triangles that a node i is involved in. Notice that equation (3) computes the *local* clustering coefficients for an undirected graph. The *global* clustering coefficient is the mean value over the CC_i.

- **Shortest path length**

The shortest path length (L) is a measure which can be computed in a simple manner, as described in [35]. L characterizes the minimum number of edges we have to pass through in order to be transferred from a node i to a j, in the binary adjacency matrix case. The higher the L, the more fully connected is the graph, i.e., it consists of a lot of edges. Notice again that, there exists the *local* shortest path, which concerns each node examination and the *global*, which is the mean value over all the local shortest paths.

- **Small-world coefficient**

The small-world phenomenon is characterized by high clustering coefficient and shortest path length and essentially it describes the ability of a node to be different from the other nodes' behaviour, in a satisfying level. Theoretically, the small-world phenomenon appears when $S > 1$. Practically, S is just a graph-based feature which helps classify emotions. The small-world phenomenon is analogous to the "hub" phenomenon, for larger values of the coefficient. Motivated by [35], [36], we computed its global value, i.e., the value which concerns the whole graph, as follows:

$$S_i = \frac{\frac{CC_i}{CC_{i,random}}}{\frac{L_i}{L_{i,random}}} \ . \tag{4}$$

The global small-world coefficient is the mean value over all local small-world coefficients. It should be noted that in our implementation, instead of using the Watts's and Strogatz's algorithm, as presented in [36], for the creation of the random graph, we selected the well-known procedure of Erdős and Rènyi [37].

IV. Experimental Evaluation

In this section, we present the results of our approach on two well-known public datasets, namely, the EMOVO and the SAVEE datasets [23], [24]. Our proposed methodology is compared with the well-known eGeMAPS feature set for

Authorized licensed use limited to: Hellenic Mediterranean University. Downloaded on April 01,2025 at 22:22:09 UTC from IEEE Xplore. Restrictions apply.

emotion recognition. It should be noted that we propose 6 graph-based features while eGeMAPS consists of 88 features per speech utterance.

A. Data Description

The two aforementioned public datasets are characterized by different languages and specific sentences. Both datasets have been sampled at $F_s = 44100$ Hz. Specifically:

- EMOVO: It is an Italian acted-speech emotional corpus which contains 7 emotional states from 6 actors. The emotions which have been recorded concern anger, disgust, fear, joy, sadness, surprise and neutral states. The recordings were created by 2 different groups. 588 samples are included in the dataset.
- SAVEE: This dataset is made by 4 English male actors. It is a well-known British acted-speech multimodal corpus, consisting of 480 utterances which express 7 different emotions, i.e., the anger, disgust, fear, happiness, sadness, surprise and neutral emotional state.

B. Experimental Results

The main focus of our experimental analysis is to apply graph theory-based features on speech signals via the visibility graph-based representation to successively extract specific graph-based features. As a comparative method, we choose the feature extraction based on the eGeMAPS [32]. Both feature sets were fed to a *Support Vector Machine* (SVM) classifier, which is able to categorize emotional states in a multi-class fashion. Since SVMs are in principle binary classifiers, a one-vs-rest scheme is adopted. It is worth mentioning that the choice of the SVM classifier is a common choice in the literature, as this categorization algorithm has been proved to be of high effectiveness, among a variety of machine learning and deep learning algorithms, in recognizing emotional states [27]. The experimental evaluations were carried out by computing the *Unweighted Average Recall* (UAR) accuracy (%), a measure which takes into consideration datasets which consist of balanced classes. The UAR accuracy, presented in the following Table I, is the mean value over the accuracies, derived from a leave-one-speaker-out (LOSO) method classification approach.

TABLE I
AVERAGE UAR [%] MULTI-CLASS CASE, LEAVE-ONE-SPEAKER-OUT

	EMOVO	**SAVEE**
eGeMAPS [33]	37.4	40.8
Graph-features	**98**	**70**

Table I suggests that the proposed graph-based feature set outperforms a prominent and widely used feature set in speech emotion recognition field, i.e., the extended GeMAPS set of speech features. Please note that the graph-based features were normalized based on their maximum value.

In Fig. 1, we present a visualization of the distribution of EMOVO's graph-based features for all emotions and all utterances, by reducing the features' dimensionality from 6D to 2D, through the *t-Distributed Stochastic Neighbor Embedding* (t-SNE) method [39].

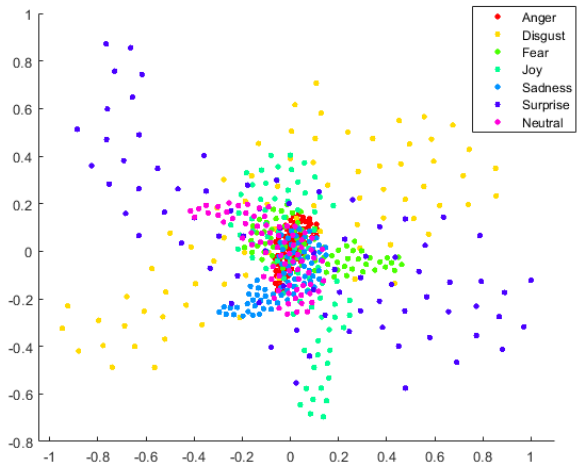

Fig. 1. t-SNE dimensionality reduction (6D to 2D) for the samples of the EMOVO dataset.

Overall, the graph-based characteristics are shown to be very promising in addressing the problem of emotional states categorization from speech in a variety of acted speech datasets. It is importance to note that these datasets concern different languages and contain both speaker genders.

V. CONCLUSIONS AND FUTURE WORK

This study introduced a graph-based feature set for speech emotion recognition and it constitutes the first application of the graph-based features on emotional speech analysis. Specifically, our proposed pipeline consists of the following steps: at first we applied the sliding windows method on the speech signals. From each window we computed the zero-mean RMS value of each window, as a non-negative measure which helped us to estimate a new, appropriate time series that was successively presented as input to the visibility graph. Next, the adjacency matrix is constructed, i.e., the graph representation of the speech signals. To this end, conventional graph-based characteristics were extracted from each speech-based utterance, and the procedure is finalized with a multi-class SVM categorization of the emotional states. Our proposed methodology proved to be more effective against a well-established competitor, the well-known eGeMAPS feature set in two publicly available acted emotional speech datasets.

As an extension of our work, we could evaluate the performance of our proposed methodology on different datasets and different languages. Another challenging task would be the application of our features on cross-corpora classification and furthermore on context-free datasets. In addition, we plan to propose a more detailed interpretation of these graph-based features in terms of speech signals. Finally, different classification schemes can be tested and more features can be incorporated or devised and included in the proposed set.

Authorized licensed use limited to: Hellenic Mediterranean University. Downloaded on April 01,2025 at 22:22:09 UTC from IEEE Xplore. Restrictions apply.

ACKNOWLEDGEMENTS

The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 101017331 (ODIN).

REFERENCES

[1] B. Schuller A. Batliner, "Computational paralinguistics: emotion, affect and personality in speech and language processing,", John Wiley & Sons, 2013.

[2] DJ France, RG Shiavi, S. Silverman, M. Silverman, DM Wilkes,"Acoustical properties of speech as indicators of depression and suicidal risk, ", in IEEE Transactions on Biomedical Engineering, vol. 47, no. 7, pp. 829-837, 2000.

[3] H. Wang, Y. Liu, X. Zhen, X. Tu, "Depression Speech Recognition With a Three-Dimensional Convolutional Network," in Frontiers in human neuroscience, vol. 15 713823, 30 Sep. 2021.

[4] L. Hansen, YP Zhang, D. Wolf, K. Sechidis, N. Ladegaard, R. Fusaroli, "A generalizable speech emotion recognition model reveals depression and remission," in Acta Psychiatr Scand., vol. 145, no.2, pp 186-199, 2021.

[5] S. Tokuno et al., "Usage of emotion recognition in military health care," in Defense Science Research Conference and Expo (DSR), pp. 1-5, 2011.

[6] N. Azam, T. Ahmad, N. Ul Haq, Automatic emotion recognition in healthcare data using supervised machine learning. PeerJ. Computer science, 7, e751. https://doi.org/10.7717/peerj-cs.751, "Automatic emotion recognition in healthcare data using supervised machine learning," in PeerJ. Computer science, v. 7, e751, 2021.

[7] C. Xiong, D. Dongyang, W. Zhiyong, "Emotion controllable speech synthesis using emotion-unlabeled dataset with the assistance of cross-domain speech emotion recognition," in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 5734-5738, 2021.

[8] J. Gao, D. Chakraborty, H. Tembine, O. Olaleye, "Nonparallel emotional speech conversion," in arXiv preprint arXiv:1811.01174, 2018.

[9] C. Luefeng, S. Wanjuan, F. Yu, W. Min, S. Jinhua, H. Kaoru , "Two-layer fuzzy multiple random forest for speech emotion recognition in human-robot interaction," in Information Sciences, vol. 509, pp. 150-163, 2020.

[10] S. Ramakrishnan, I.M.M El Emary, "Speech emotion recognition approaches in human computer interaction," in Telecommunication Systems, vol. 52, pp. 1467–1478, 2013.

[11] T. Deschamps-Berger, L. Lamel, L. Devillers, "End-to-End Speech Emotion Recognition: Challenges of Real-Life Emergency Call Centers Data Recordings," in IEEE International Conference on Affective Computing and Intelligent Interaction (ACII), pp. 1-8, 2021.

[12] M. Bojanić et al., "Call redistribution for a call center based on speech emotion recognition," in Applied Sciences, 10(13), 4653, 2020.

[13] B. W. Schuller, "Speech emotion recognition: Two decades in a nutshell, benchmarks, and ongoing trends," Communications of the ACM, vol. 61, no. 5, 90-99, 2018.

[14] M. B. Akçay, K. Oğuz, "Speech emotion recognition: Emotional models, databases, features, preprocessing methods, supporting modalities, and classifiers," in Speech Communication, vol. 116, pp. 56-76, 2020.

[15] J. Kacur, B. Puterka, PJ. avlovicova, M. Oravec, "On the Speech Properties and Feature Extraction Methods in Speech Emotion Recognition," in Sensors, 21(5):1888, 2021.

[16] M. El Ayadi, M. S. Kamel, F. Karray, "Survey on speech emotion recognition: Features, classification schemes, and databases," in Pattern recognition, vol. 44, no. 3, pp. 572-587, 2011.

[17] F. Hirsch, A. Wohlschlaeger, "A Graph analysis of nonlinear fMRI connectivity dynamics reveals distinct brain network configurations for integrative and segregated information processing," in Nonlinear Dynamics, April 2022.

[18] DFP Melo, IS Fadigas, HBB Pereira, "Graph-based feature extraction: A new proposal to study the classification of music signals outside the time-frequency domain," in PLoS One, 12;15(11):e0240915, Nov. 2020.

[19] L. Lacasa, B Luque, F. Ballesteros, J. Luque, J.C. Nuño, "From time series to complex networks: The visibility graph," in Proceedings of the National Academy of Sciences, 105(13):4972-4975, 2008.

[20] T. Wang et al., "Speech signal processing on graphs: The graph frequency analysis and an improved graph Wiener filtering method," in Speech Communication, 127, 82-91, 2021.

[21] Y. Wang, H. Guo, X. Yan, Z. Yang, "Speech Signal Processing on Graphs: Graph Topology, Graph Frequency Analysis and Denoising," Chinese J. Electron., 29, pp. 926-936, 2020.

[22] A. Shirian, T. Guha, "Compact graph architecture for speech emotion recognition," In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 6284-6288, 2021.

[23] G. Costantini, I. Iaderola, A. Paoloni, M. Todisco, "EMOVO Corpus: an Italian Emotional Speech Database, Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)," pp. 3501-3504, May 2014.

[24] https://www.kaggle.com/ejlok1/surrey-audiovisual-expressed-emotion-savee

[25] E. Lieskovská, M. Jakubec, R. Jarina, M. Chmulík, "A Review on Speech Emotion Recognition Using Deep Learning and Attention Mechanism," in Electronics 10, 1163, 2021.

[26] BJ Abbaschian, D. Sierra-Sosa, A. Elmaghraby, "Deep Learning Techniques for Speech Emotion Recognition, from Databases to Models," in Sensors, 21, 1249, 2021.

[27] M. Papakostas, G. Siantikos, T. Giannakopoulos, E. Spyrou, D. Sgouropoulos, "Recognizing Emotional States Using Speech Information," in Adv Exp Med Biol. 989:155-164, 2017.

[28] T. M. Wani, T. S. Gunawan, S. A. A. Qadri, M. Kartiwi, E. Ambikairajah, "A Comprehensive Review of Speech Emotion Recognition Systems," in IEEE Access, vol.9, pp.47795-47814, 2021.

[29] K. Tomba, J. Dumoulin, E. Mugellinm, O. A. Khaled, S. Hawila,"Stress Detection Through Speech Analysis," in ICETE, 2018.

[30] W. Zehra, A.R. Javed, Z. Jalil et al., "Cross corpus multi-lingual speech emotion recognition using ensemble learning," in Complex Intell. Syst. 7, pp. 1845–1854, 2021.

[31] F. Eybe, "Acoustic Features and Modelling. In: Real-time Speech and Music Classification by Large Audio Feature Space Extraction," in Springer Theses (Recognizing Outstanding Ph.D. Research). Springer, Cham, 2016.

[32] F. Eyben et al., "The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for Voice Research and Affective Computing," in IEEE Transactions on Affective Computing, vol. 7, no. 2, pp. 190-202, 1 April-June 2016.

[33] F. Haider, S. Pollak, P. Albert, L. Saturnino, "Emotion recognition in low-resource settings: An evaluation of automatic feature selection methods," in Computer Speech and Language, Volume 65, 2021.

[34] T. A. Song et al., "Graph Convolutional Neural Networks For Alzheimer's Disease Classification," in 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), pp. 414-417, 2019.

[35] DS Bassett, ET Bullmore, "Small-World Brain Networks Revisited," in Neuroscientist, 23(5):499-516, 2017.

[36] D. Tomasi D, ND Volkow, "Mapping Small-World Properties through Development in the Human Brain: Disruption in Schizophrenia," in PLOS ONE, vol. 9, pp. 1-17, 2014.

[37] R. Durrett, "Erdös–Rényi Random Graphs," in Random Graph Dynamics, Cambridge Series in Statistical and Probabilistic Mathematics, pp. 27-69, 2006.

[38] R. A. Khalil, E. Jones, M. I. Babar, T. Jan, M. H. Zafar, and T. Alhussain, "Speech emotion recognition using deep learning techniques: A review", IEEE Access, 7, 117327-117345, 2019.

[39] L.J.P. van der Maaten, G.E. Hinton, "Visualizing High-Dimensional Data Using t-SNE," Journal of Machine Learning Research 9:2579-2605, 2008.

Authorized licensed use limited to: Hellenic Mediterranean University. Downloaded on April 01,2025 at 22:22:09 UTC from IEEE Xplore. Restrictions apply.