Section A-Research paper
ISSN 2063-5346

Integrating Bibliometric and Co-occurrence Analysis in the Literature of Multilingual and Multimodal Speech Emotion Recognition: A Preliminary Review

Sudipta Bhattacharya[1], Brojo Kishore Mishra[1], Samarjeet Borah[2], Raghvendra Kumar[1]

[1]Department of Computer Science and Engineering, GIET University, Gunupur, India

[2]Department of Computer Applications, SMIT, Sikkim Manipal University, Rangpo, Sikkim, India

sudipta.bhattacharya@giet.edu,

bk.mishra@giet.edu,

samarjeet.b@smit.smu.edu.in,

raghvendra@giet.edu

ABSTRACT

The study of how to identify and translate feelings between different languages is expanding rapidly. As more people become interested in the study of speech signals, numerous methods have been created to decipher the tone of a speaker's words. Traditional speech analysis and classification methodologies are just two of the many tools used in speech emotion identification research (SER). This report provides a comprehensive overview of the research on multilingual and multimodal speech emotion recognition. Limitations of speech emotion detection, speech emotion extraction methods, speech emotion databases, and contributions from speech emotion extraction are discussed. The goal of these articles was to assess the state of multilingual and multimodal speech emotion detection in Scopus relative to current trends and leading countries in this field of study. The search syntax is based on language and medium-agnostic voice emotion recognition. In our essay, we summarise the research conducted over the past decade. Voice emotion detection and multilingual and multimodal were used to locate 706 records in Scopus between January 1, 2013, and October 23, 2022. The VOS viewer programme was then used to evaluate these files.

Keywords Speech emotion detection; Articles; Multimodal; Multilingual; Bibliometric analysis.

1. INTRODUCTION

Humans have an innate preference for verbal exchange. Speech processing is a significant area of study within the larger science of signal processing. Voice recognition, interactive voice communication, medical advancements, emotion detection, contact centre automation, virtual assistants, robots, and many more applications rely on this signal processing technique. Modern developments in artificial intelligence, machine learning, and signal processing have made this a practical possibility. Because of its wide range of applications, emotional analysis of speech has proven challenging for researchers. Lack of sufficient emotional databases, along with determining the appropriate feature vector and classifiers, is a significant obstacle for spoken emotion identification systems. The study of emotional recognition has a long history. Emotion detection research began with the hypothesis that emotions may be read from a person's facial expressions [1]. In recent years, a large body of research has focused on the problem of identifying feelings from audio recordings. The importance of consumers feeling connected to their technology is underlined [2]. There has been a recent uptick in research into a field called speech emotion recognition, or SER for short. The purpose of this study is to identify feelings by analysing vocalisations. To efficiently extract useful emotional qualities from SER, however, is one of the most daunting obstacles. [16-20]. A proposal for a multi-modal approach to SER that makes use of both text and speech is presented in [3]. In this technique, the sound mode is modelled using the cepstral coefficients of Mel frequencies. [14,15] (Music and Film Content Classification System; MFCC: s) (MFCC:s, common feature-set for representing sound). The audio and textual data are processed by RNN-based architectures to identify the most frequent feelings discussed in the media. There are a wide range of human emotions, including anger, despair, apathy, joy, and satisfaction (mixed with excited for the purpose of balancing out the unequal labels). Weighted Average Precision (WAP) scores might range from 68.8 to 71.8% accuracy. Another source is [6], which details a network design that uses input features derived from the 3-dimensional Log Mel-spectrum and runs on convolutional neural networks (CNNs), attentional networks (ANNs), and bidirectional long short-term memories (LSTMs). Emo-DB (UA 85.39%) and IEMOCAP (UA 69.32% on Angry, Glad, Sad, and Neutral on the improvised sessions) are used to evaluate the subject's emotional state. The study also includes cross-corpus evaluations for four emotions, with training on the IEMOCAP dataset and evaluation on the Emo-DB dataset (happiness, sadness, anger, and neutrality). The results, which were accurate to within 6.38%, demonstrated that there are commonalities in the way both languages describe and express feelings. [16,21]. Emo-DB [7], eNTERFACE [8], and AFEW 4.0 [9] are all put through their paces in [6]'s experimental section, which covers a wide range of language options. The model is trained on one (the source dataset) and evaluated on the others; two experiments are run on Emo-DB and four on the other databases (the target dataset). Weighted Average Recall (WAR) (five emotional labels) performance on Emo-DB is 52.27 percent using eNTERFACE as the source set, while it reduces to 47.80 percent with AFEW 4.0. There are six primary categories of education available. Their plan centres on a tailored version of the least-squares regression for their specific application (DaLSR). In [6], a battery of cross-corpus tests is run using Emo-DB, eNTERFACE, and FAU Aibo in addition to a method termed non-negative matrix factorization. Several parallels can be drawn between the tests described here and those in [7]. When tested on Emo-DB, their results using eNTERFACE as a training set show an accuracy of 52.10 percent across all five emotions (Anger, Disgust, Fear, Happiness, and Sadness). Data for the cross-corpus studies mentioned in reference [8] came primarily from Emo-DB, eNTERFACE, and FAU Aibo; the authors also propose using a transfer linear subspace learning system. Using Emo-DB as the assessment dataset and eNTERFACE for training, they achieve 53.85% accuracy (five emotions). [11], [12,[13]

In this study, we conducted a comprehensive literature assessment of multilingual and multimodal speech emotion recognition from a linguistic perspective by using bibliometric and co-occurrence analyses. Bibliometric analysis is a method for quantifying the relationships and influence of publications in a certain field of study, while co-occurrence analysis is simply the counting of matched data within a collection unit. Around 706 instances of "voice emotion recognition" data were examined in this study, spanning from January 1, 2013, to October 23, 2022. Using the search terms "multilingual and multimodal speech emotion recognition" and "analysed with VOS-viewer software," we were able to retrieve these citations from the Scopus database.
In the context of these knowledge gaps, we propose to research the following research questions

- What are the publication trends of studies on peer assessment in online language courses, including the number of publications, study fields, and distribution of journals?
- Who are the most prominent researchers in the field of peer assessment in online language courses? What are the most often used terms and which countries and institutions have produced the most studies on this topic?
- What cross-disciplinary subjects typically include peer review in Speech emotion detection?

The ethics statement, study design, data collection, and visualisation are all included in part 2 of this article, which also discusses multilingual and multimodal speech emotion detection analysis. Section 3 of this analysis discusses the results. Excel and the VOSviewer tool were used to analyse the information that was taken from the Scopus database. Based on these numbers, we may deduce that researchers in the fields of voice processing in Japan, China, India, the United Kingdom, and the United States are the most prolific. Our data shows that conference presentations account for 73% of all voice-processing research papers published in the last five years. Section 4 describes the topic and the driving force behind the study. Section 5 provides evidence for the conclusion.

2. METHODS

The bibliometric methodology includes the use of quantitative techniques on bibliometric data and a summary of the bibliometric and intellectual structure of a field by analysing the relationships between different research components [23,24]. This data can be used to show the contributions of different disciplines, find connections and silos, and spot trends and possible gaps [23,24]. As a result, it gives both a map of the science and an analysis of how well it works, which helps figure out how a field of research has changed over time [22,25]. Figure 1 represents methodological scheme for the bibliometric analysis

Figure 1 Methodological scheme for the bibliometric analysis

2.1 Identify Study field and Time period

The field of multimodal and multilingual speech emotion detection has grown and changed a lot in recent years, thanks to improvements in machine learning, natural language processing, and affective computing. If the analysis is limited to the last 10 years, it can focus on the most recent and important research in the field while still giving a full picture of the research landscape. This time period also lets us look at how recent technological advances have changed the field of multimodal and multilingual speech emotion detection. This research was founded on a descriptive and bibliometric examination of a literature database [9].

2.2 Select Information sources

In our analysis we have chosen Scopus database. Scopus is a bibliographic database that lists scientific, technical, medical, and social science literature. It has information about citations, author profiles, and bibliometric indicators for more than 24,000 journals, conference proceedings, and books.

2.3 Search criteria

The time range covered by the Scopus data we pulled begins on January 1, 2013, and ends on October 23, 2022. Multilingual and Multimodal Speech Emotion Detection was the title that came up when the search terms "TITLE-ABS-KEY ((speech AND emotion AND detection) AND multilingual AND multimodal)" were used.

2.4 Identification

Records identify 707 documents including documents, articles, review articles, book chapters, and proceedings.

2.5 Selecting of software

Microsoft excel for table visualisation for documents, countries and affiliations and Vos-viewer for bibliometric network analysis. Analyses were performed using the CSV file format and VOS viewer. When preparing data for bibliometric analysis, we looked for red flags like inconsistent or incorrect information (such the use of abbreviations or acronyms in place of full names of countries) [10].

2.6 Records Analysis

The data analysis was done in two parts. The first was a performance study that tracked how many articles were published, which nations and institutions contributed the most, and which journals were the most widely read in the field of social cohesiveness. The second part of the analysis focused on science mapping, which used bibliometric maps to look at the intellectual structure of the field [23,25]. In this study, we looked at how often author keywords, countries, and authors all happened together.

3. RESULT

The search for "Multilingual and Multimodal Speech Emotion Detection" yielded 198 archival results between January 1, 2013, and October 23, 2022. (Figure 2). It's remarkable to see how far research into emotion recognition in speech has come in recent years.

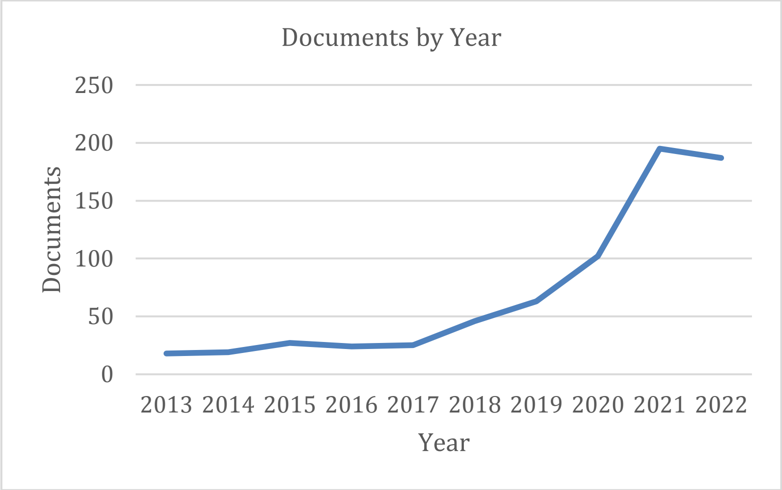

Figure 2. Documents in Scopus Arranged by Year

China, the US, India, the UK, Spain, Italy, Germany, Singapore, Canada, and Australia have reported significant work (66% of the total reported activity). In addition, the research on Multilingual and Multimodal Speech Emotion Detection received significant contributions from Japan, Ireland, and France (Figure 3).

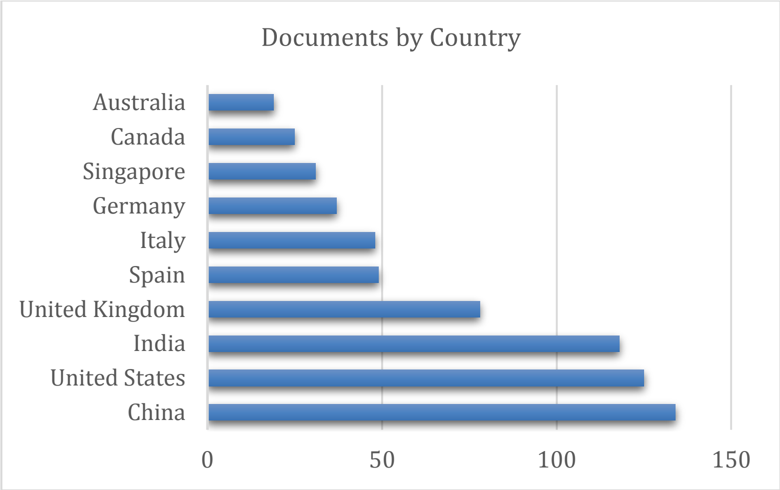

Figure 3. Scopus Database Documents, Segmented by Country/Territory (2013- October 2022)

Fifty-three percent of the sources we analysed were outcome articles, thirty percent were conference papers, and the remaining six and eleven percent were book chapters and reviews, respectively. (Figure. 4). To our knowledge, no comprehensive study on the topic of multilingual and multimodal speech emotion detection has been published in book form. The fields of engineering (13%), computer science (40%), the social sciences, mathematics, the arts and humanities, decision sciences, and similar disciplines have dominated the research (Figure. 5).

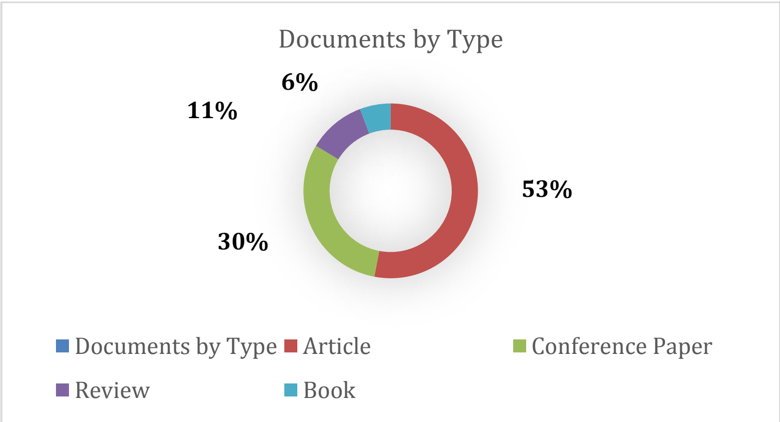

Figure 4: Scopus Database Documents, Classified (2013- October 2022)

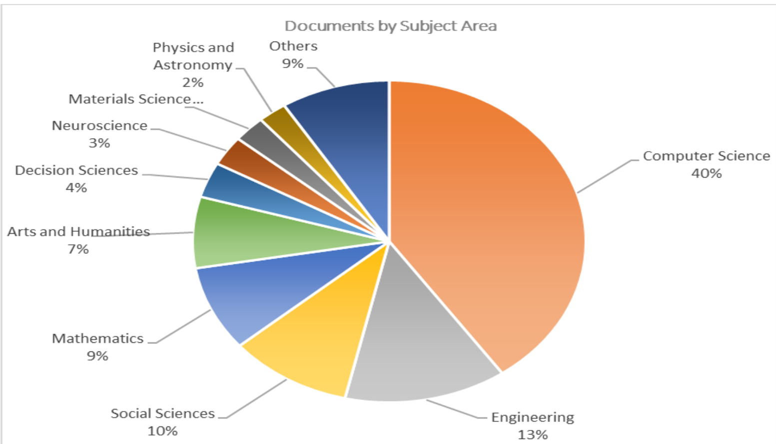

Figure 5.: Scopus Database Documents, Grouped by Subject (2013- October 2022)

There was a total of 706 peer-reviewed articles published on the topic of multilingual and multimodal speech emotion recognition between January 1, 2013 and October 23, 2022. (Figure 6).

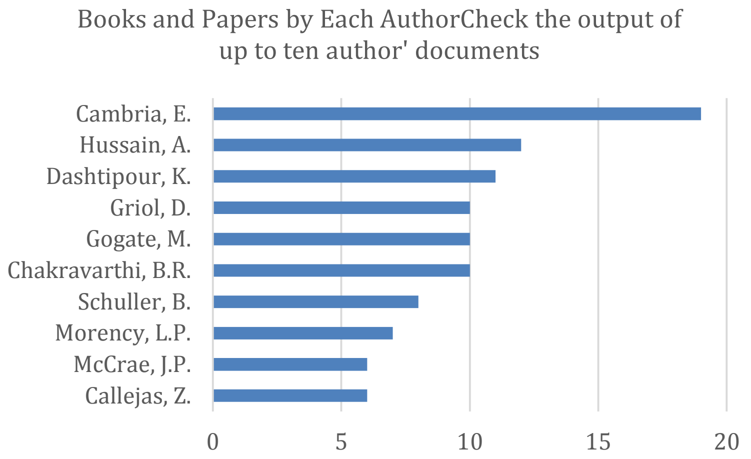

Figure 6: a total of 706 scientific studies on multilingual and multimodal speech emotion recognition trends were published.

3.1 Network analysis

Multilingual and multimodal speech emotion detection [3] publications are increasingly being evaluated using network analysis, an essential aspect of bibliometric analysis. You can choose between two alternatives. The first examines the frequency with which certain keywords appear together, while the second looks at the extent to which authors contributed equally to a given piece of work. The research trend in speech emotion recognition was determined by analysing 706 articles published between January 1, 2013, and October 23, 2022. Figure 7 represents the Network visualisation of title co-occurrences

.

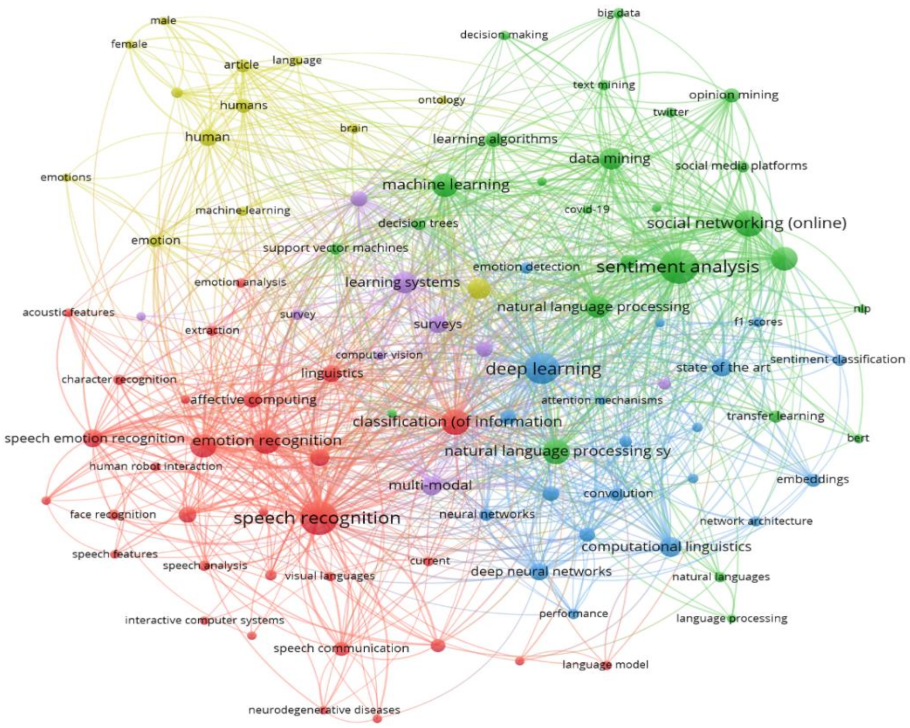

Figure 7. Network visualisation of title co-occurrences

3.1.1 Co-occurrence network analysis

Multilingual and multimodal speech emotion detection [3] publications are increasingly being evaluated using network analysis, an essential aspect of bibliometric analysis. You can choose between two alternatives. The first examines the frequency with which certain keywords appear together, while the second looks at the extent to which particular authors contributed equally to a given piece of work. The research trend in speech emotion recognition was determined by analysing 706 articles published between January 1, 2013, and October 23, 2022.

3.1.2 Co-authorship network analysis

The publication-based author-collaboration network is shown in Figure 8. The names of the writers are represented by the nodes, the relationships between authors are represented by the links between the nodes, and the sizes of the nodes represent the quantity of works by each author. The authors having the most ties to each other, as determined by the co-authorship network analysis, are A. Hussain and K. Dashtipour. If the needed textual data is available, we can create authorship networks for this study. So, the choice of the cutoff value has to be taken into mind. Just 23 of the 908 words were suitable for further study. Any given sentence must appear at least (5). 60 percent of the most relevant phrases are selected automatically. The VOSviewer programme eliminates 20 names from the full list of authors in order to better display the co-authorship structure of this paper (items). The 20 terms can be broken down into four categories, with a total of 36 links and a mean connection strength of 76.

Here, a relationship is defined as the co-occurrence of two terms (keywords). The VOSviewer user manual states that a higher positive number indicates a stronger connection. If it's high, the association is significant. The number of times two substances are mentioned in the same research study is proportional to the strength of the linkages between them.

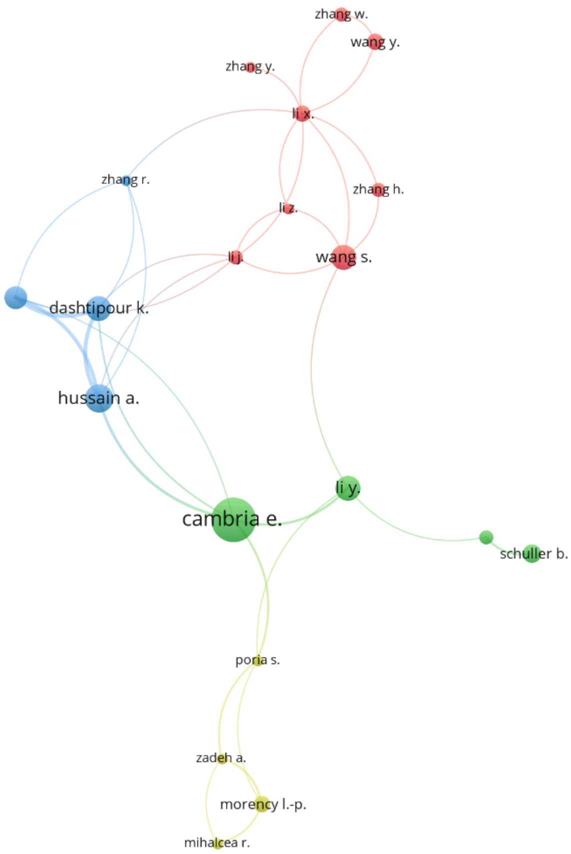

Figure 8. The author co-authorship network.
Note: VOS-viewer presents a network view. There are 20 nodes, 4 clusters, and 36 links in the entire network. The value of the total link strength is 76.

The study found that the total link strength was correlated with the number of publications in which two or more items appeared (co-incidence connection between objects or keywords).

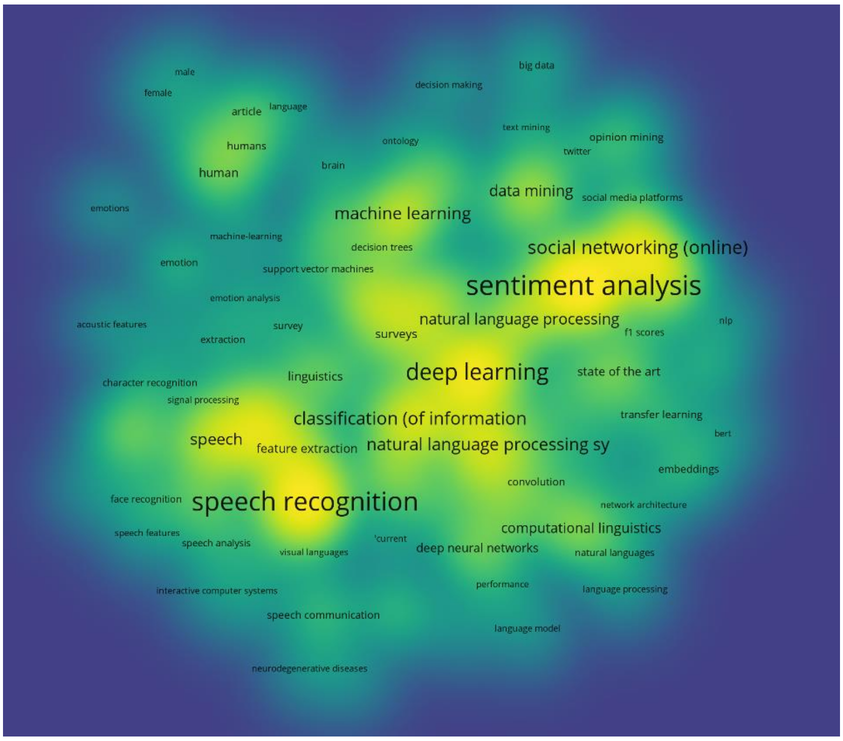

Figure 9. The network of title co-occurrences visualised by density.

simultaneously. In our experiment, we found that deep learning, sentiment analysis, and speech recognition had the strongest connections. Figure 9 shows a density visualisation calculated from the total strength of the links.

4. DISCUSSION

4.1 Publication trends

The publication patterns and future research directions in multilingual and multimodal speech emotion detection are shown in this preliminary literature review that combines bibliometric and co-occurrence analyses. The expanding interest and demand for emotion identification technology across cultural and linguistic boundaries is reflected in the increasing number of publications in recent years, which is one of the important conclusions of the review. This shift is most likely caused by the expanding role that computerised emotion identification plays in fields like medicine, education, and even entertainment. Another intriguing discovery is that most research on multilingual speech emotion recognition has concentrated on European and East Asian languages, especially English and Mandarin. This could be because these languages are so commonly used and studied that they are perfect for studying how to recognise emotions. This highlights the necessity for additional study of emotion recognition in less-studied languages and cultural contexts to guarantee that the technology is accessible to all. The paper also emphasises the increasing attention paid to multimodal emotion recognition, which incorporates not only spoken language but also other modalities like facial expressions and physiological signs. By integrating data from many sources, this method has the potential to strengthen emotion detection algorithms' reliability and precision. However, this also raises additional problems with regards to data integration and processing, as well as model creation and evaluation. Another interesting development in the studied literature is the increasing use of deep learning-based methods, in particular neural networks. Recent research indicates these approaches are effective, and they are therefore gaining popularity. However, they need a lot of information and processing power, which might be a problem in some situations. Happiness, sorrow, anger, fear, and surprise are the most researched emotions in multilingual speech emotion recognition. Research often employs these feelings since they are thought to be transferable across cultures and languages. Yet, this hints at the necessity for additional study into the identification of other emotions across cultural and social boundaries. It’s important to keep in mind several caveats despite the review's useful information. The review is biased towards English-language articles, for instance, and excludes those written in other languages. Potentially relevant papers published in languages other than English may have been overlooked, resulting in a skewed picture of the state of the art in this field of study. In addition, the review did not evaluate the quality of the articles it contained, which could weaken the reliability of the results. The evaluation identifies numerous avenues for future research that might be beneficial to investigate. For instance, researchers could examine emotion recognition in more complicated multimodal environments that involve more than two modalities, as well as in languages and cultures that have received less attention in the literature. There is also a need for explainable emotion recognition models that shed light on the traits and processes that underpin emotion recognition. When taken as a whole, the preliminary literature review combining bibliometric and co-occurrence analysis on multilingual and multimodal speech emotion detection provides a helpful overview of the current state of the field and the research directions that are likely to shape its future development. Emotion detection technology can help a wide range of people, and academics can help make that technology better by addressing the issues highlighted in this paper.

4.2 Theoretical advancements

Insights into the theoretical developments in this expanding topic are provided via a preliminary literature assessment that combines bibliometric and co-occurrence analysis on multilingual and multimodal speech emotion detection. Many theoretical approaches and concepts, such as cross-cultural psychology, emotional computing, and cognitive neuroscience, are highlighted in this overview of the relevant literature. One of the most significant theoretical contributions to the body of literature under consideration is the incorporation of cross-cultural psychology into the study of emotion. This method highlights the need for culturally sensitive emotion identification models by recognising the impact of cultural context on emotional experiences and expressions. Cultural elements such as language and social conventions have been proven to affect the accuracy of emotion perception in studies of this kind that have investigated cross-cultural differences in emotional manifestations and their recognition. Affective computing is another theoretical method used in the literature; its goal is to create computational models that can detect, analyse, and react to human emotions. This method, which incorporates insights from psychology, neurology, and computer science, has been applied to the creation and assessment of speech and other modalities' emotion recognition models. Machine learning methods, such as neural networks, are frequently used in affective computing models because of their ability to learn from data and steadily improve their performance. Research from the field of cognitive neuroscience is another theoretical perspective that has informed the works we've looked at here. This method has been used to guide the creation of emotion recognition models by illuminating the brain mechanisms underpinning emotional experiences and expressions. Emotions have been demonstrated to be connected with distinctive neural patterns, and studies in this vein have studied the brain regions and networks involved in processing emotional information. Several other theoretical notions pertinent to emotion recognition have also been investigated in the reviewed literature, which complements the aforementioned theoretical methods. The idea of emotional contagion, for instance, has been used to describe the spread of negative feelings from one person to another, which can have a negative effect on emotion recognition models' efficacy. The idea of emotional granularity is used in a similar vein to describe individual variances in the recognition and expression of various emotions. Notwithstanding these theoretical advances, researchers in this subject still face several theoretical hurdles. We need, for instance, theories of emotion that are more sophisticated and culturally sensitive so that we may do justice to the intricate relationship between language, culture, and feelings. In addition, theoretical models that can explain the role of multiple modalities in emotion recognition and how they can be combined to enhance model accuracy and resilience are required. Theoretically, there is also the issue of creating emotion recognition models that are more open and easier to understand. Ethical and legal concerns arise from the fact that it is becoming more difficult to understand how machine learning algorithms arrive at their choices. Scientists attempting to solve this problem should create interpretable models that can shed light on the factors that go into emotion detection. Emotion is experienced and expressed in a wide variety of ways, therefore future theoretical developments should focus on creating more all-encompassing and integrative theories of emotion. Because of this, it may be necessary to draw on concepts from other fields, including but not limited to psychology, neuroscience, linguistics, and cultural studies. In addition, studies should concentrate on creating theories of emotion recognition that can take into consideration the interplay between modalities and their effect on model efficacy. The overall findings of this exploratory literature study combine bibliometric and co-occurrence analyses to shed light on the theoretical developments in the field of multilingual and multimodal speech emotion recognition. More effective and inclusive emotion detection models that can help a wide range of groups can be developed if researchers tackle the theoretical hurdles and limits highlighted in this paper.

4.3 Integration of speech emotion detection and interdisciplinary trends

This emerging topic is highly interdisciplinary, as demonstrated by the preliminary literature assessment that combines bibliometric and co-occurrence analysis on multilingual and multimodal speech emotion detection. Researchers from a wide range of fields, such as psychology, neuroscience, linguistics, and cultural studies, have contributed to the development of emotion recognition models that can accurately capture the complexity and diversity of emotional experiences and expressions, in addition to speech technology and affective computing. The integration of many modalities in emotion recognition models is one of the most important interdisciplinary trends in the reviewed literature. Although though words can convey a wealth of feeling, they are frequently supplemented by non-verbal cues like facial expressions, body language, and even physiological indications. Multimodal models have been demonstrated to outperform unimodal models in many contexts, and related research has investigated how different modalities might be merged to increase emotion identification accuracy and robustness. Emotion recognition with machine learning algorithms is another area where researchers are drawing from several fields. To create predictions or judgements based on data, machine learning is a branch of computer science. Machine learning algorithms can learn to distinguish patterns and features that are indicative of specific emotions when trained on big datasets of voice and other modalities. Deep neural networks, support vector machines, and decision trees are just some of the models that have been developed and evaluated with the help of machine learning to aid in emotion recognition. The literature study also emphasises the value of interdisciplinary cooperation in the development of emotion recognition technology. Emotion recognition models can be made more robust and precise by combining the work of researchers from a variety of fields. Neuroscientists, for instance, can help pinpoint the brain mechanisms behind emotional processing, while psychologists can shed light on the cognitive and social aspects that influence emotional experiences and expressions. Scholars in the field of cultural studies can aid in making sure that models are sensitive to cultural variances in emotional expressions and experiences, while linguists can aid in the construction of language-specific emotion detection models. One other trend seen across disciplines in the surveyed literature is the use of emotion recognition models in non-clinical settings. There are many areas that could benefit from emotion detection technology, including mental health care, classroom instruction, and even the entertainment industry. By providing objective evaluations of emotional states to complement standard clinical assessments, emotion recognition models, for instance, can be used to enhance the precision with which mental health diagnoses are made. An emotion recognition model's potential application in the classroom lies in its ability to tailor instruction to each student's unique emotional requirements and learning styles. By allowing characters to react in real time to the emotional states of users, emotion recognition models can improve the realism and immersion of virtual worlds in the entertainment industry. Despite these cross-disciplinary successes, researchers still face various obstacles that need to be overcome before they can truly progress the area of emotion identification. There is a need for more inclusive and comprehensive data sets, which is a significant obstacle. Most existing emotion identification datasets are centred on the Western hemisphere, which may not accurately reflect the range of emotions experienced by people of different cultures and languages. To overcome this difficulty, scientists need to compile and annotate datasets that are more representative of society at large and able to capture the intricacies and linguistic and cultural differences that characterise people's emotional expressions and experiences.

The requirement for more precise evaluation strategies is another obstacle. Accuracy in identifying just a few common emotions, like joy, sadness, and anger, is frequently used to gauge the success of emotion detection models. Emotions, on the other hand, are nuanced and multifaceted, with varying manifestations and experiences based on time, place, and culture. To overcome this obstacle, scientists should work on creating assessment tools that are sensitive to and capable of capturing a wide variety of emotional expressions and experiences, as well as those that take into consideration individual and cultural differences.

5. CONCLUSIONS

Major findings from the preliminary literature review using bibliometric and co-occurrence analysis: Over the past few years, investigations on multilingual emotion recognition have grown steadily. This is likely due to the increased interest in creating and using such technology to various linguistic and cultural environments. Second, much study on translating emotion has focused on European and East Asian languages like English and Mandarin. Third, multimodal emotion identification research thrives. This field uses speech, facial expressions, and physiological data to comprehend an individual's emotional state. Neural networks are the most prominent deep learning method for recognising emotions in multiple languages and media. . Multilingual speech emotion recognition research focuses on happiness, sadness, anger, fear, and surprise. Emotion recognition in understudied languages and cultures and multimodal situations involving more than two senses needs more research. This first literature review suggests that multilingual and multimodal speech emotion recognition is a promising research subject. Emotion recognition in understudied languages and cultures and advanced multimodal situations needs more research.

The early literature evaluation on multilingual and multimodal speech emotion identification using bibliometric and co-occurrence analysis has limitations. English-only publications may skew the study. In this discipline, relevant publications published in languages other than English may have been overlooked, skewing the state of the art. Second, Scopus-only papers may have constrained the research. . Finally, the articles' quality wasn't assessed, which could cast doubt on the results. Fourth, the analysis only covers articles published until September 2021, lacking the latest research. The study only used bibliometric and co-occurrence analyses, not content or systematic reviews. Hence, it may only show portion of the studied area.

The early literature evaluation on multilingual and multimodal speech emotion identification using bibliometric and co-occurrence analysis has limitations. English-only publications may skew the study. In this discipline, relevant publications published in languages other than English may have been overlooked, skewing the state of the art. Second, Scopus-only papers may have constrained the research. . Finally, the articles' quality wasn't assessed, which could cast doubt on the results. Fourth, the analysis only covers articles published until September 2021, lacking the latest research. The study only used bibliometric and co-occurrence analyses, not content or systematic reviews. Hence, it may only show portion of the studied area.

Conflict of interest: NO
Data used: Publicly available data only
Funding: No

References:

[1]F. W. Smith and S. Rossit, ``Identifying and detecting facial expressions of emotion in peripheral vision,'' PLoS ONE, vol. 13, no. 5, May 2018, Art. no. e0197160, doi: 10.1371/journal.pone.0197160.
[2]M. Chen, P. Zhou, and G. Fortino, ``Emotion communication system,'' IEEE Access, vol. 5, pp. 326 337, 2017, doi: 10.1109/ACCESS.2016.2641480.
[3]S. Yoon, S. Byun, and K. Jung, "Multimodal speech emotion recognition using audio and text," in 2018 IEEE Spoken Language Technology Workshop (SLT). IEEE, 2018, pp. 112–118.
[4]H. Meng, T. Yan, F. Yuan, and H. Wei, "Speech emotion recognition from 3d log-mel spectrograms with deep learning network," IEEE access, vol. 7, pp. 125 868–125 881, 2019.
[5]Y. Zong, W. Zheng, T. Zhang, and X. Huang, "Cross-corpus speech emotion recognition based on domain-adaptive least-squares regression," IEEE signal processing letters, vol. 23, no. 5, pp. 585–589, 2016.
[6] O. Martin, I. Kotsia, B. Macq, and I. Pitas, "The enterface'05 audiovisual emotion database," in 22nd International Conference on Data Engineering Workshops (ICDEW'06). IEEE, 2006, pp. 8–8.
[7] A. Dhall, R. Goecke, J. Joshi, K. Sikka, and T. Gedeon, "Emotion recognition in the wild challenge 2014: Baseline, data and protocol," in Proceedings of the 16th international conference on multimodal interaction, 2014, pp. 461–466.
[8] S. Steidl, Automatic classification of emotion related user states in spontaneous children's speech. Logos-Verlag, 2009. [Page 24.]
[9] P. Song, W. Zheng, S. Ou, X. Zhang, Y. Jin, J. Liu, and Y. Yu, "Crosscorpus speech emotion recognition based on transfer non-negative matrix factorization," Speech Communication, vol. 83, pp. 34–41, 2016.
[10] P. Song, "Transfer linear subspace learning for cross-corpus speechemotion recognition." IEEE Trans. Affect. Comput., vol. 10, no. 2, pp. 265–275, 2019.
[11] Bhattacharya, S., Das, N., Sahu, S., Mondal, A., & Borah, S. (2021, March). Deep Classification of Sound: A Concise. In Proceeding of First Doctoral Symposium on Natural Computing Research: DSNCR 2020 (Vol. 169, p. 33). Springer Nature.
[12] Bhattacharya, S., Borah, S., Mishra, B. K., & Das, N. (2022, September). Deep Analysis for Speech Emotion Recognization. In 2022 Second International Conference on Computer Science, Engineering and Applications (ICCSEA) (pp. 1-6). IEEE.
[13] Bhattacharya, S., Borah, S., Mishra, B. K., & Mondal, A. (2022). Emotion detection from multilingual audio using deep analysis. Multimedia Tools and Applications, 1-30.

[14]Mathur, M., Samiulla, S., Bhat, V., & Jenitta, J. (2020, October). Design and Development of Writing Robot Using Speech Processing. In 2020 IEEE Bangalore Humanitarian Technology Conference (B-HTC) (pp. 1-4). IEEE.

[15]Handoko, L. H. (2021). COVID-19 research trends in the fields of economics and business in the Scopus database in November 2020. Science editing, 8(1), 64-71.

[16]Das, N., & Padhy, N. (2021, August). A Bibliometric and Co-occurrence Analysis of Speech Processing Literature Published from the Year 2015 to mid of June 2021. In Proceedings of the International Conference on Data Science, Machine Learning and Artificial Intelligence (pp. 237-243).

[17]Gaikwad, S. K., Gawali, B. W., & Yannawar, P. (2010). A review on speech recognition technique. International Journal of Computer Applications, 10(3), 16-24.

[18]Koolagudi, S. G., & Rao, K. S. (2012). Emotion recognition from speech: a review. International journal of speech technology, 15(2), 99-117.

[19]Thapliyal, N., & Amoli, G. (2012). Speech based emotion recognition with gaussian mixture model. Int. J. Adv. Res. Comput. Eng. Technol, 1(5), 65-69.

[20]Vogt, T., André, E., & Wagner, J. (2008). Automatic recognition of emotions from speech: a review of the literature and recommendations for practical realisation. Affect and emotion in human-computer interaction, 75-91.

[21] Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, *133*, 285-296.

[22] Block, J. H., & Fisch, C. (2020). Eight tips and questions for your bibliographic study in business and management research. *Management Review Quarterly*, *70*, 307-312.

[23] Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2011). An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the Fuzzy Sets Theory field. *Journal of informetrics*, *5*(1), 146-166.