



e-JURNAL PENYELIDIKAN DAN INOVASI
ISLAMIC INNOVATION CENTER FOR
TECHNOLOGY, UIS
<https://ejpi.uis.edu.my>
e-ISSN 2289-7909

VOL. 13 NO. 2 (AUGUST, 2026) Special Edition AIDI
Page 56-75
DOI: <https://doi.org/10.53840/e-jpi.v13i2.405>

Received: 8 July 2026

Accepted: 18 August 2026

Published: 31 August 2026

A Bibliometric Analysis of Artificial Intelligence in Arabic Speech and Quranic Recitation Recognition

Analisis Bibliometrik Kecerdasan Buatan dalam Pengecaman Pertuturan Arab dan Bacaan Al-Quran

**Khairulnizam Abd Rahman, Nuh Akhtar Mohamad Khazani, Che Wan Shamsul Bahri
Che Wan Ahmad, Nur Muizz Mohamed Salleh, Rafiza Kasbun**

Faculty of Creative Multimedia and Computing, Universiti Islam Selangor (UIS), Malaysia

Email: khairulnizam@uis.edu.my, cwshamsul@uis.edu.my

Muharman Lubis

Telkom University, Indonesia

Email: muharmanlubis@telkomuniversity.ac.id

ABSTRACT

This study aims to explore the development and application of artificial intelligence techniques in Arabic speech recognition, with a specific focus on recitation accuracy, pronunciation analysis, and language learning support. It seeks to identify trends, methods, and challenges in AI-based Arabic recitation recognition systems. A bibliometric and systematic review approach was employed using the Scopus database. Relevant publications from 2016 to 2026 were retrieved using a structured query combining keywords related to speech recognition, machine learning, and Arabic language processing. The selected studies were analyzed based on research trends, methodologies, and application domains. The results indicate a growing interest in deep learning approaches such as neural networks, recurrent neural networks (RNN), convolutional neural networks (CNN), and transformer-based models for Arabic speech and recitation recognition. Applications are primarily focused on pronunciation assessment, computer-assisted language learning, and Quranic recitation systems. However, challenges remain in handling dialectal variations, limited annotated datasets, and the complexity of Arabic phonetics. This study is limited to publications indexed in Scopus and may exclude relevant works from other databases. Future research should focus on improving dataset availability, incorporating tajweed rules, and enhancing model robustness for diverse Arabic dialects and recitation styles. This paper provides a comprehensive overview of AI-driven Arabic recitation recognition, highlighting current trends and gaps while offering insights for future research in educational technology and speech processing.

Keywords: Arabic speech recognition, recitation recognition, artificial intelligence, deep learning, pronunciation assessment, computer-assisted language learning, bibliometric review.



INTRODUCTION

The advancement of artificial intelligence (AI), particularly in speech recognition and deep learning, has significantly enhanced the ability of machines to process and interpret human language (Chan et al., 2016; Chiu et al., 2018). These developments have enabled more accurate and efficient speech-based systems across various domains, including education and language learning.

Arabic speech recognition, however, remains a challenging area due to its rich phonetic structure, morphological complexity and reliance on diacritics for accurate pronunciation. These challenges are further amplified in Quranic recitation, where precise articulation and adherence to tajweed rules are essential. Recent studies highlight that deep learning architectures such as sequence-to-sequence models and hybrid frameworks have significantly improved recognition performance (Chiu et al., 2018; Watanabe et al., 2017).

Despite these advancements, research in Arabic recitation recognition remains fragmented across disciplines, making it difficult to obtain a comprehensive understanding of trends, methodologies and research gaps. The increasing complexity of speech recognition systems and architectures further emphasizes the need for systematic analysis (Rao et al., 2017).

This study aims to systematically examine the body of literature on Arabic AI-based recitation recognition through a bibliometric approach. It focuses on identifying relevant publications using a structured search strategy and analyzing their distribution over time to reveal research growth patterns. The study also seeks to explore keyword occurrences to uncover major themes and emerging topics, as well as to identify influential authors, institutions, and countries contributing to the field. Furthermore, it aims to classify the commonly used artificial intelligence techniques and application areas in recitation recognition, while highlighting key challenges and gaps to inform future research directions.

This study is guided by several key questions. It seeks to determine how research on Arabic AI-based recitation recognition has evolved over the period from 2016 to 2026 in terms of publication trends. It also aims to identify the main research themes and keyword patterns that characterize this field. Another important question concerns the types of artificial intelligence techniques and models that are most frequently applied. In addition, the study investigates who the leading contributors are, including authors, institutions, and countries. Finally, it explores the major challenges and research gaps that emerge from the existing body of literature.

LITERATURE REVIEW

Bibliometric Analysis

The bibliometric analysis provides a structured examination of the scientific output related to Arabic AI-based recitation recognition. It focuses on evaluating publication trends, keyword occurrences, authorship patterns, and citation impact to uncover the intellectual structure of the field. The analysis reveals how research activity has evolved over time, highlighting periods of growth that correspond with advancements in artificial intelligence, particularly deep learning techniques. Keyword analysis further illustrates the dominant themes, with frequent emphasis on speech recognition, pronunciation assessment and language learning applications, alongside emerging interests in Quranic recitation and phoneme-level analysis. In addition, the distribution of publications across countries and institutions reflects the global contribution to this domain, while citation patterns help identify influential studies and foundational works. Overall, the bibliometric approach enables a comprehensive understanding of research development, key contributors and thematic directions within Arabic recitation recognition.

Past Studies

Previous studies in Arabic speech and recitation recognition have evolved from traditional statistical approaches to modern deep learning techniques. Earlier works primarily relied on conventional models

such as Hidden Markov Models, whereas recent research has shifted toward deep neural networks, including recurrent and convolutional architectures (Miao et al., 2016; Watanabe et al., 2017).

Sequence-to-sequence models have emerged as a dominant approach due to their ability to model complex temporal dependencies in speech data (Prabhavalkar et al., 2017; Chiu et al., 2018). In addition, the development of Arabic-specific natural language processing resources, such as word embedding models, has further supported advancements in this field (Soliman et al., 2017).

Artificial intelligence has also been increasingly applied in educational contexts, particularly in pronunciation assessment and computer-assisted language learning systems (Huang et al., 2023). Furthermore, recent innovations in self-supervised learning have demonstrated the potential to improve speech representation without requiring large labeled datasets (Baevski et al., 2020).

Despite these advancements, several challenges remain, including limited annotated datasets, dialectal variation, and the complexity of Arabic phonetics. These limitations continue to hinder the development of robust and generalizable recitation recognition systems.

METHODOLOGY

The data for this study were obtained from the Scopus database using a structured search query designed to capture publications related to Arabic speech recognition, artificial intelligence, and recitation or pronunciation. The search was limited to documents published between 2016 and 2026 to ensure relevance to recent technological developments. Only documents with complete bibliographic information were considered to enable accurate analysis. The initial results were refined by applying inclusion criteria such as relevance to Arabic language processing and the use of AI-based techniques, while unrelated subject areas and incomplete records were excluded. The final dataset represents a curated collection of publications that reflect the current state of research in Arabic AI-based recitation recognition and serves as the basis for subsequent bibliometric analysis.

This study adopts a bibliometric approach to analyze the selected dataset and uncover patterns within the research domain. Quantitative analysis is conducted to examine publication trends over time, authorship distribution, and contributions by countries and institutions. In addition, keyword analysis is used to identify dominant themes and emerging topics, providing insight into the focus areas of existing studies. Citation analysis further highlights influential publications and the overall impact of research within the field. The findings are interpreted to reveal the intellectual structure, research evolution, and key directions in Arabic AI-based recitation recognition, offering a comprehensive overview that supports future research development.

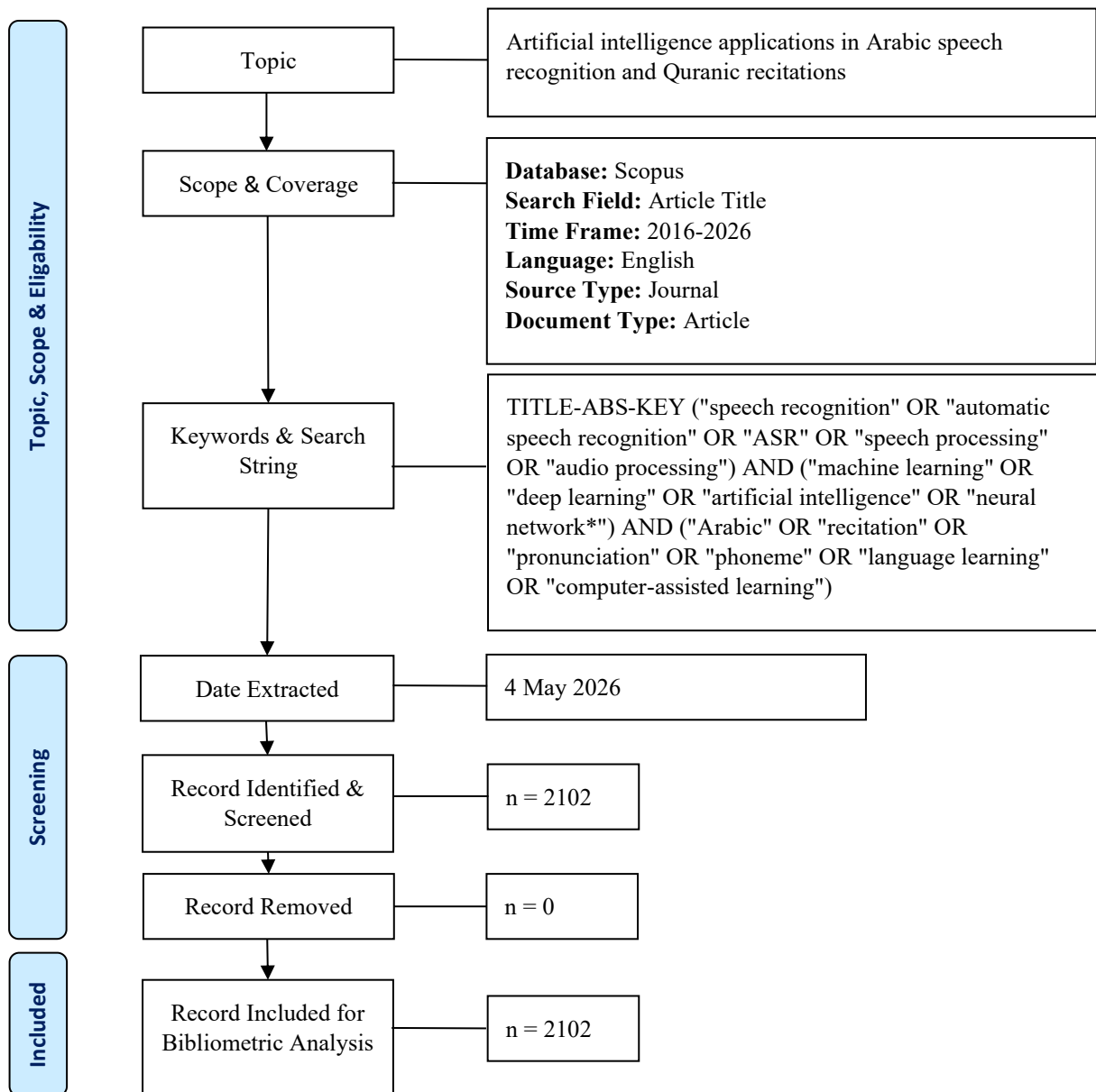


Figure 1: Flow diagram of the search strategy.

RESULTS AND FINDINGS

The results section presents the findings of the bibliometric analysis based on the selected dataset of publications. It provides a descriptive overview of the publication patterns, including document types, authorship trends, keyword occurrences, and other relevant bibliometric indicators. These results are intended to illustrate the structure and development of the research field, highlighting dominant publication channels and key contributions within the literature. Overall, this section serves as a foundation for understanding the dissemination trends and scholarly characteristics of the studied domain.

Table 1 shows the distribution of document types within the dataset, with a total of 2,102 publications. Conference papers dominate the output, accounting for 1,242 publications (59.09%), indicating that the field is highly driven by conference-based dissemination. Journal articles represent the second largest category with 756 publications (35.97%), suggesting a substantial but secondary emphasis on peer-reviewed journal publishing.

Table 1: Document Type

Document Type	Total Publications (TP)	Percentage (%)
Conference Paper	1242	59.09
Article	756	35.97
Book Chapter	49	2.33
Review	32	1.52
Retracted	13	0.62
Data Paper	6	0.29
Book	3	0.14
Editorial	1	0.05
Total	2102	100.00

The remaining document types collectively form a small proportion of the literature. Book chapters contribute 49 publications (2.33%), while review articles account for 32 (1.52%), reflecting a relatively limited focus on synthesis studies. Other types such as retracted papers (0.62%), data papers (0.29%), books (0.14%), and editorials (0.05%) are minimal in comparison, highlighting their marginal role in the overall scholarly output. Overall, the distribution indicates a strong preference for conference dissemination, with journal articles serving as the primary complementary publication channel.

Table 2 shows that nearly half of the publications come from conference proceedings (49.29%), making it the dominant source type in the dataset. Journals contribute 39.11%, indicating a strong but secondary role in scholarly dissemination. Book series account for 10.04%, while books represent only a small fraction at 1.57%. Overall, the findings highlight that conference proceedings and journals are the primary sources of research output in this field.

Table 2: Source Type

Source Type	Total Publications (TP)	Percentage (%)
Conference Proceedings	1036	49.29
Journal	822	39.11
Book Series	211	10.04
Books	33	1.57
Total	2102	100.00

Table 3 shows that English is overwhelmingly the dominant language of publication, accounting for 2,072 publications (98.6%). This indicates that the field is highly internationalized with a strong reliance on English as the primary medium of scholarly communication. Other languages contribute only marginally, with Chinese (0.67%), Korean (0.24%), Spanish (0.19%), and Arabic (0.10%) representing very small proportions of the dataset. Several additional languages, including Malay, Russian, Latvian each account for 0.05–0.10% of publications.

Table 3: Languages

Language	Total Publications (TP)*	Percentage (%)
English	2072	98.57
Chinese	14	0.67
Korean	5	0.24
Spanish	4	0.19
Arabic	2	0.10
Malay	2	0.10
Russian	2	0.10
Latvian	1	0.05
Total	2102	100.00

Overall, the results highlight a clear concentration of research output in English, while non-English publications remain limited and scattered across a few languages.

Based on Table 4, the distribution of publications by subject area reveals a strong dominance of Computer Science with 1,823 publications (86.73%), followed by Engineering (39.82%) and Mathematics (29.73%). This pattern reflects that the research corpus is predominantly anchored in technical and technology-driven disciplines. The presence of Arts and Humanities (15.60%) and Social Sciences (14.75%) within the mid-tier cluster suggests that non-technical dimensions likely encompassing ethics, user experience, and human-computer interaction also feature meaningfully within the body of literature examined.

Nevertheless, a notable gap exists in the representation of life sciences and medical-related fields. Disciplines such as Medicine (4.80%), Neuroscience (2.38%), Biochemistry & Genetics (1.19%), and Immunology & Microbiology (0.05%) recorded considerably low publication counts, indicating that the subject under investigation remains largely underexplored among researchers from these communities. Furthermore, given that the cumulative percentage exceeds 100%, it can be inferred that a portion of the publications were simultaneously indexed under multiple subject categories, reflecting the inherently multidisciplinary nature of the overall publication corpus.

Table 4: Subject Area

Subject Area	Total Publications (TP)	Percentage (%)
Computer Science	1823	86.73
Engineering	837	39.82
Mathematics	625	29.73
Arts and Humanities	328	15.60
Social Sciences	310	14.75
Physics and Astronomies	211	10.04
Decision Sciences	207	9.85
Material Sciences	120	5.71
Medicine	101	4.80
Energy	82	3.90
Neuroscience	50	2.38
Business, Management and Accounting	42	2.00
Chemical Engineering	33	1.57
Health Professors	27	1.28
Biochemistry, Genetics and Molecular Biology	25	1.19
Multidisciplinary	25	1.19
Chemistry	21	1.00
Environmental Science	18	0.86
Psychology	18	0.86
Agricultural and Biological Science	7	0.33
Earth and Planetary Sciences	6	0.29
Economics , Econometrics and Finance	6	0.29
Dentistry	2	0.10
Immunology and Microbiology	1	0.05
Nursing	1	0.05

Research Trends

Based on Table 5 and Figure 2, the annual distribution of publications demonstrates a consistent upward trend in scholarly output from 2016 (98 publications) to 2026 (124 publications as of the time of data collection), with a notable peak in 2025 recording the highest total of 411 publications. This steady growth in total publications (TP) reflects an increasing academic interest in the subject area over the decade. However, the citation metrics reveal an inverse relationship between publication year and

citation impact, earlier publications from 2016 and 2017 recorded substantially higher citation per paper (C/P) values of 42.83 and 32.23 respectively, compared to more recent years such as 2024 (1.31) and 2025 (1.31). This pattern is consistent with the natural bibliometric phenomenon whereby older publications accumulate citations over time, while recent works have yet to reach their full citation potential.

The h-index and g-index values further reinforce this observation, with 2018 achieving the highest h-index (32) and 2016 recording the highest g-index (64), both attributable to the longer citation accumulation period of earlier works. The number of cited papers (NCP) relative to total publications also indicates that a proportion of recent publications particularly in 2025 and 2026 remain uncited, which is expected given their recency. Nonetheless, the sustained growth in publication volume across the years, particularly the sharp increase from 2022 onwards, signals a rapidly expanding research community and growing scholarly attention toward this field, suggesting its emergence as an increasingly prominent area of academic inquiry.

Table 5: Year of Publication

Year	TP	NCP	TC	C/P	C/CP	<i>h</i>	<i>g</i>
2016	98	92	4197	42.83	45.62	20	64
2017	138	130	4448	32.23	34.22	28	64
2018	157	141	4297	27.37	30.48	32	61
2019	163	140	2458	15.08	17.56	25	44
2020	145	130	3159	21.79	24.30	27	51
2021	177	155	3198	18.07	20.63	29	50
2022	216	196	2306	10.68	11.77	27	36
2023	223	176	2735	12.26	15.54	22	46
2024	250	179	1078	4.31	6.02	14	24
2025	411	131	540	1.31	4.12	9	17
2026	124	13	20	0.16	1.54	2	3
Total	2102						

Notes: TP=total number of publications; NCP=number of cited publications; TC=total citations; C/P=average citations per publication; C/CP=average citations per cited publication; *h*=h-index; and *g*=g-index.

Documents by year

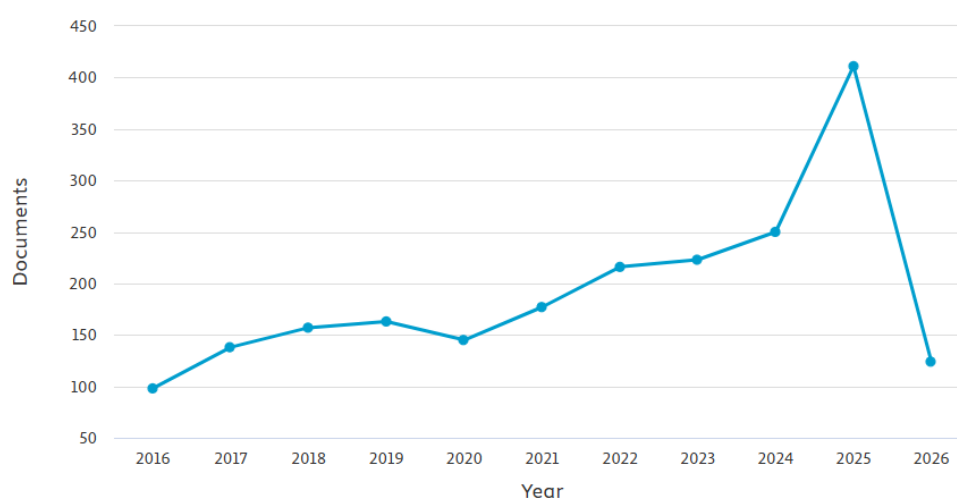


Figure 2: Total Document by Year

Based on Table 6, the country-wise analysis shows that China leads in publication output with 477 papers (22.69%), followed by India (307; 14.61%) and the United States (250; 11.89%). Although China and India contribute the highest number of publications, the United States demonstrates the strongest

research impact with the highest citation count (10,735) and citation-based indicators. Saudi Arabia (150; 7.14%) also shows a significant contribution, reflecting strong regional interest in the field.

Table 6: Top 20 Countries contributed to the publications

Country	TP	NCP	TC	C/P	C/CP	<i>h</i>	<i>g</i>
China	477	22.69%	289	4052	8.49	14.02	32
India	307	14.61%	191	1797	5.85	9.41	18
United States	250	11.89%	212	10735	42.94	50.64	44
Saudi Arabia	150	7.14%	124	2292	15.28	18.48	25
United Kingdom	82	3.90%	70	1133	13.82	16.19	18
Japan	79	3.76%	64	1573	19.91	24.58	16
Germany	75	3.57%	64	1302	17.36	20.34	18
Egypt	66	3.14%	52	1175	17.80	22.60	14
South Korea	61	2.90%	45	1110	18.20	24.67	16
Algeria	55	2.62%	41	527	9.58	12.85	14
UAE	49	2.33%	37	822	16.78	22.22	14
Malaysia	48	2.28%	29	226	4.71	7.79	10
France	46	2.19%	43	1049	22.80	24.40	15
Iraq	46	2.19%	27	208	4.52	7.70	8
Tunisia	46	2.19%	34	348	7.57	10.24	12
Canada	42	2.00%	36	1070	25.48	29.72	15
Jordan	40	1.90%	27	671	16.78	24.85	12
Indonesia	37	1.76%	22	214	5.78	9.73	8
Taiwan	37	1.76%	22	250	6.76	11.36	8
Morocco	36	1.71%	26	295	8.19	11.35	8

Notes: TP=total number of publications; NCP=number of cited publications; TC=total citations; C/P=average citations per publication; C/CP=average citations per cited publication; *h*=*h*-index; and *g*=*g*-index.

Other countries such as the United Kingdom, Japan, Germany, Egypt, and South Korea contribute moderately, each representing between 2% to 4% of total publications. Meanwhile, countries like Malaysia, Indonesia, Iraq and Morocco show smaller but notable participation. Overall, the results indicate a globally distributed research landscape, with strong contributions from Asia, the Middle East, Europe, and North America, highlighting the international nature of AI-based applications in Arabic speech recognition and Quranic recitations research.

Based on Table 7, the author analysis highlights a small group of highly productive and influential researchers in the field of AI-based applications in Arabic speech recognition and Quranic recitations. Meyer, B.T. (Germany) is the most prolific author with 17 publications, while Shahin, I. and Nassif, A.B. (UAE) also show strong productivity and impact with high citation averages. Among all authors, Rao, K. (United States, Google Inc.) stands out significantly with the highest citation count (2,118), indicating major influence despite a relatively moderate number of publications.

Table 7: Most Productive Authors

Author's Name	Affiliation	Country	TP	NCP	TC	C/P	C/CP	<i>h</i>	<i>g</i>
Meyer, B.T.	Medizinische Physik & Cluster of Excellence Hearing4all	Germany	17	16	287	16.88	17.94	10	16
Shahin, I.	Dept. of Electrical Engineering	UAE	13	11	390	30.00	35.45	8	13
Nassif, A.B.	Dept. of Electrical Engineering	UAE	12	11	390	32.50	35.45	8	12
Zhang, J.	School of Informatics	Japan	12	19	275	22.92	14.47	7	16
Rao, K.	Google Inc.	United States	11	11	2118	192.5	192.5	11	11
Alotaibi, Y.A.	College of Computer and Information Sciences	Saudi Arabia	10	10	158	15.80	15.80	8	10

Satori, H.	USMBA University	Morocco	10	6	44	4.40	7.33	4	6
Meftah, A.H.	College of Computer and Information Sciences	Saudi Arabia	9	9	149	16.56	16.56	7	9
Meng, H.	Dept. of Systems Engineering and Engineering Management	Hong Kong	9	9	425	47.22	47.22	8	9
Selouani, S.A.	College of Computer and Information Sciences	Saudi Arabia	9	8	92	10.22	11.50	6	9
Xie, Y.	Beijing Language and Culture University	China	9	10	38	4.22	3.80	4	5
Abbas, M.	Mohamed bin Zayed University of AI	UAE	8	6	161	20.13	26.83	4	8
Ali, A.	MIT	United States	8	9	206	25.75	22.89	7	9

Notes: TP=total number of publications; NCP=number of cited publications; TC=total citations; C/P=average citations per publication; C/CP=average citations per cited publication; h=h-index; and g=g-index.

Other notable contributors include researchers from Japan, Saudi Arabia, Hong Kong, Morocco, China, and the United States, reflecting a geographically diverse authorship pattern. Overall, the results show that while a few key authors dominate both productivity and citation impact, the research field remains internationally distributed with contributions from both academic institutions and industry leaders.

Based on Table 8, the most active source titles in this research domain are predominantly conference proceedings and journals affiliated with major technical publishers, particularly IEEE, Springer, and ACM. Interspeech leads in total publications (152 TP) with a substantial citation count of 2,837, affirming its position as the primary dissemination platform for this field. IEEE Access (58 TP) and ICASSP (57 TP) follow closely, with ICASSP notably recording the highest total citations of 5,051 despite ranking third in publication volume — indicating that conference papers published under this venue carry significant scholarly influence. The dominance of conference proceedings over journals in the top sources reflects a disciplinary norm in computer science and engineering, where conferences are regarded as primary and high-impact publication venues.

In terms of citation quality metrics, IEEE/ACM Transactions on Audio Speech and Language Processing stands out with the highest CiteScore (9.5) and SJR (1.37), as well as the highest SNIP value (2.24), signifying its superior journal impact relative to its publication volume (24 TP). Speech Communication (Elsevier) also demonstrates a commendable SNIP of 1.26, suggesting strong contextual citation influence within the field. Conversely, Springer-published series such as Lecture Notes in Computer Science (SJR: 0.18, SNIP: 0.37) and Lecture Notes in Networks and Systems (SJR: 0.15, SNIP: 0.28) recorded lower impact metric values despite relatively high publication counts, which may indicate that these venues serve more as broad dissemination platforms rather than high-impact outlets. This disparity between publication volume and citation quality underscores the importance of evaluating source influence through multiple bibliometric indicators rather than publication count alone.

Table 8: Most Active Source Title

Source Title	TP	TC	Publisher	Cite Score	SJR 2018	SNIP 2018
Proceedings of the Annual Conference of the International Speech Communication Association Interspeech	152	2837	ISCA	3.5	0.689	0.93
IEEE Access	58	1052	IEEE	4.29	0.587	0.87

ICASSP IEEE International Conference on Acoustics Speech and Signal Processing Proceedings	57	5051	IEEE	2.8	0.5	0.8
Lecture Notes in Computer Science Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics	53	232	Springer	0.7	0.18	0.37
Lecture Notes in Networks and Systems	39	51	Springer	0.5	0.15	0.28
Communications in Computer and Information Science	33	54	Springer	0.5	0.17	0.3
International Journal of Speech Technology	33	342	Springer	2.2	0.31	0.76
ACM International Conference Proceeding Series	31	106	ACM	0.6	0.19	0.4
IEEE ACM Transactions on Audio Speech and Language Processing	24	461	IEEE	9.5	1.37	2.24
Speech Communication	23	516	Elsevier	2.61	0.459	1.26

Notes: TP=total number of publications; TC=total citations;

Citation Analysis

The citation analysis section presents an evaluation of the impact and influence of publications within the field of Artificial intelligence applications in Arabic speech recognition and Quranic recitations. It examines key citation metrics such as total citations, average citations per publication, and author-level and country-level impact indicators to identify the most influential works and contributors. This analysis helps to highlight which studies, researchers, and regions have had the greatest academic and practical influence in shaping the development of the field. Overall, the citation analysis provides insight into the visibility, relevance and scholarly impact of research in this domain.

Table 9: Top 10 Highly cited articles

No.	Authors	Title	Year	Cites	Cites per Year
1	Chan, W., Jaitly, N., Le, Q., & Vinyals, O.	Listen, attend and spell: A neural network for large vocabulary conversational speech recognition	2016	2130	141.67
2	Chiu, C.-C., Sainath, T. N., Wu, Y., Prabhavalkar, R., Nguyen, P., Chen, Z., Kannan, A., Weiss, R. J., Rao, K., Gonina, E., Jaitly, N., Li, B., Chorowski, J., & Bacchiani, M.	State-of-the-Art Speech Recognition with Sequence-to-Sequence Models	2018	965	213
3	Watanabe, S., Hori, T., Kim, S., Hershey, J. R., & Hayashi, T.	Hybrid CTC/Attention Architecture for End-to-End Speech Recognition	2017	828	120.63
4	Miao, Y., Gowayyed, M., & Metze, F.	ESEN: End-to-end speech recognition using deep RNN models and WFST-based decoding	2016	645	92
5	Soliman, A. B., Eissa, K., & El-Beltagy, S. R.	AraVec: A set of Arabic Word Embedding Models for use in Arabic NLP	2017	427	71
6	Huang, X., Zou, D., Cheng, G., Chen, X., & Xie, H.	Trends, Research Issues and Applications of Artificial Intelligence in Language Education	2023	425	64.5

7	Knott, B., Venkataraman, S., Hannun, A., Sengupta, S., Ibrahim, M., & van der Maaten, L.	CRYPTEN: Secure Multi-Party Computation Meets Machine Learning	2021	355	47.44
8	Rao, K., Sak, H., & Prabhavalkar, R.	Exploring architectures, data and units for streaming end-to-end speech recognition with RNN-transducer	2017	314	46.17
9	Prabhavalkar, R., Rao, K., Sainath, T. N., Li, B., Johnson, L., & Jaitly, N.	A Comparison of sequence-to-sequence models for speech recognition	2017	279	34.89
10	Baevski, A., Schneider, S., & Auli, M.	Vq-Wav2vec: Self-Supervised Learning Of Discrete Speech Representations	2020	277	31

Visualisation Map

Based on the keyword co-occurrence network depicted in the Figure 3, speech recognition emerges as the most central and dominant keyword, represented by the largest node, indicating its role as the primary research theme within this body of literature. Closely connected to it are highly prominent keywords such as machine learning, deep learning, artificial intelligence, and convolutional neural network, forming a dense core cluster that reflects the field's strong reliance on advanced computational and neural network-based methodologies. Several distinct thematic clusters are also visible, including an Arabic language processing cluster (featuring keywords such as *arabic language*, *arabic ASR*, and *arabic natural language processing*), an educational technology cluster (*language learning*, *pronunciation assessment*, *adaptive learning*), and a speech pathology-related cluster (*dysarthria*, *Parkinson's disease*, *children speech recognition*), suggesting a diverse yet interconnected research landscape.

Critically, the network reveals a concentration of research activity around a relatively small number of high-frequency keywords, while a large proportion of nodes remain peripheral with sparse connections — indicating that many sub-themes are still nascent and underexplored. The strong co-occurrence ties between *deep learning*, *neural networks*, and *speech recognition* confirm a methodological convergence in the field, where deep learning frameworks have become the predominant approach. However, the limited connectivity of emerging keywords such as *self-supervised learning*, *whisper*, and *grapheme-to-phoneme conversion* suggests these represent frontier research directions with growing but not yet consolidated scholarly attention. This network structure underscores both the maturity of core research themes and the breadth of potential avenues for future investigation within the domain.

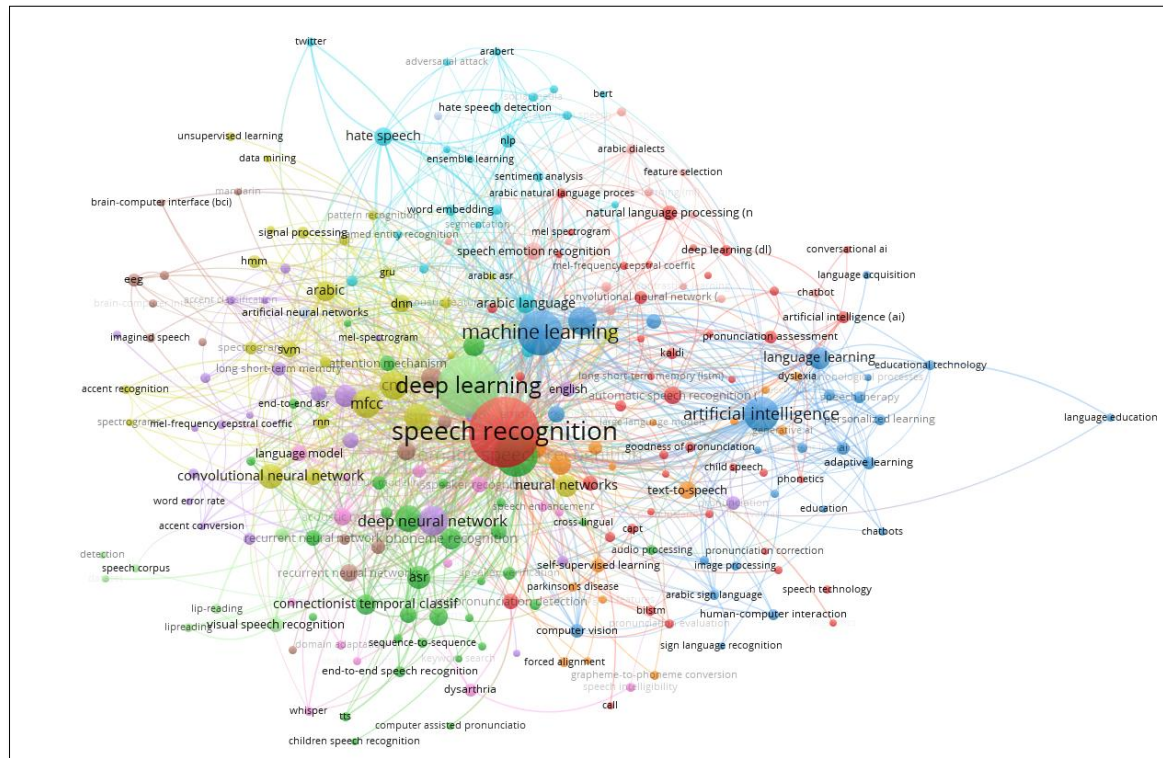


Figure 3: Network visualisation map of the author keywords

Based on Figure 4, the term co-occurrence network derived from title and abstract fields reveals speech recognition as the unequivocal central node, reinforcing its status as the overarching research theme across the analysed corpus. Surrounding this core, several prominent clusters can be identified: a deep learning and neural network cluster (featuring *deep neural networks*, *recurrent neural networks*, *end-to-end models*, and *network architecture*), an educational technology cluster (*language learning*, *computer-aided instruction*, *e-learning*, *natural language processing*), a biomedical and neuroscience cluster (*electroencephalography*, *biomedical signal processing*, *brain*, *electrophysiology*), and a social and affective computing cluster (*emotion recognition*, *social networking*, *hate speech detection*). The diversity of these clusters demonstrates that speech recognition research extends well beyond its technical foundations into interdisciplinary applications spanning education, healthcare, and social media analysis.

Critically, the network highlights a methodological duality within the field — the co-existence of classical machine learning approaches (*support vector machines, nearest neighbour search, Markov processes*) alongside modern deep learning paradigms (*deep neural networks, end-to-end models*), suggesting that the research landscape is in a transitional phase rather than having fully converged on a single methodology. The educational technology cluster on the left periphery, anchored by *language learning, e-learning, and computer-aided language learning*, appears relatively loosely connected to the technical core, indicating that application-driven research in this domain operates somewhat independently from methodological advancements. Furthermore, the presence of biomedical terms such as *electroencephalography* and *speech sound disorder* points to an emerging convergence between speech technology and clinical research — a frontier that remains underdeveloped but holds significant potential for future bibliometric growth.

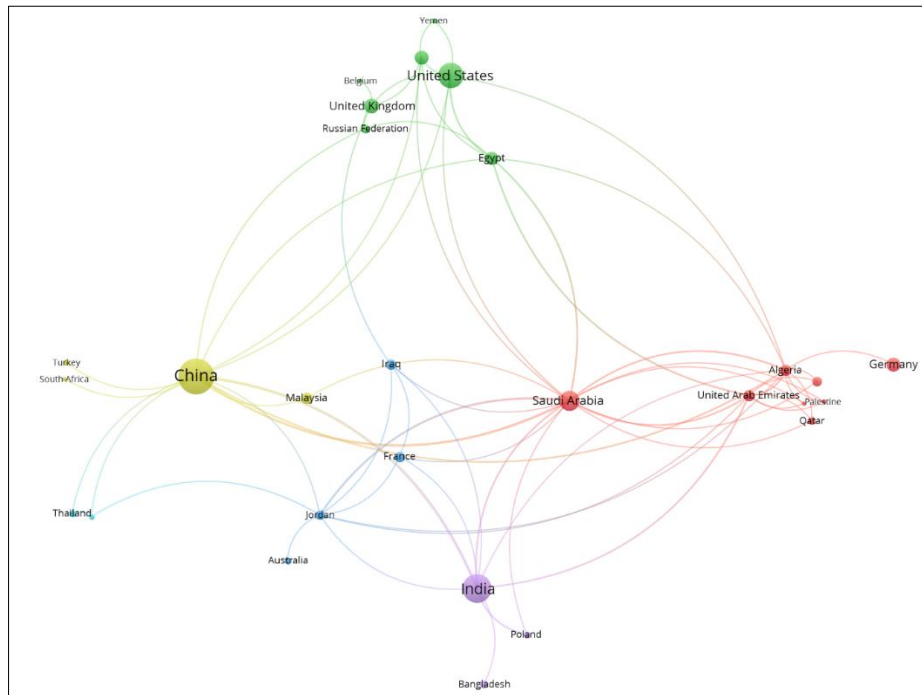


Figure 5: Network visualisation map of the citation by countries

Note: Minimum number of documents of an author = 1; Minimum number of citations of an author = 5

Figure 6 illustrates the document-level citation network, where Chan (2016b), Miao (2016b) and Watanabe (2017b) emerge as the most influential documents, represented by the largest nodes, indicating their foundational role in shaping subsequent research directions within this field. The network is characterised by multiple tightly knit clusters of co-cited documents, reflecting distinct sub-communities of scholarship that share common theoretical or methodological orientations. Notably, a large central cluster dominated by works from 2016 to 2021 suggests that this period represents the most fertile phase of knowledge production and cross-citation activity in the domain.

Critically, the emergence of Huang (2023) as a distinctly large and isolated node on the right periphery is a significant observation its size indicates high citation impact, yet its peripheral position suggests it represents a new and diverging research trajectory that has not yet been fully integrated into the established citation network. This pattern may signal a paradigm shift or the introduction of a novel methodology that is gaining rapid traction but remains structurally disconnected from earlier foundational works. The presence of numerous smaller, loosely connected nodes across the network also points to a long tail of moderately cited documents, underscoring the breadth and heterogeneity of the research landscape.

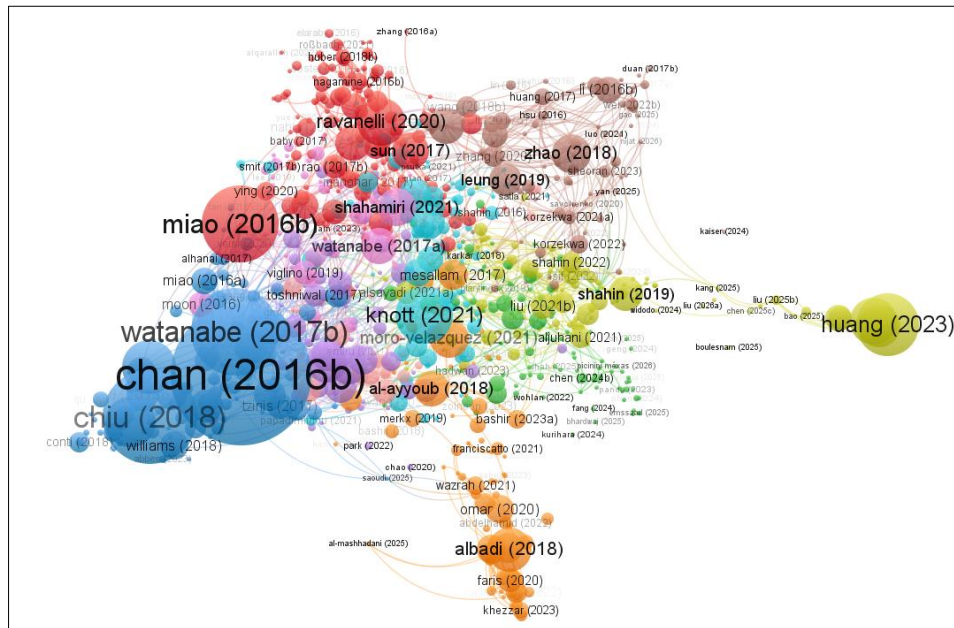


Figure 6: Network visualisation map of the citation by documents

Figure 7 refines the document citation network by applying a minimum threshold of 10 citations, resulting in a cleaner and more interpretable visualisation of the most impactful scholarly works. Chan (2016b) retains its position as the most central document, with Knott (2021), Miao (2016b), Ravanelli (2020), and Shahin (2019) forming the next tier of influential references. The reduced network reveals more clearly defined cluster boundaries, with each colour-coded cluster representing a coherent sub-theme or methodological school within the broader field, such as end-to-end speech recognition, pronunciation assessment, and Arabic speech processing.

Critically, the application of the citation threshold effectively filters out peripheral and low-impact works, allowing the intellectual backbone of the field to become more visible. The persistence of Huang (2023) as a prominent isolated node even under this stricter threshold confirms its rapidly growing influence and suggests it may represent a landmark publication possibly introducing a transformative model or benchmark that is reshaping current research directions. The relatively sparse connections between some clusters, even among highly cited documents, further indicates that certain sub-fields within this domain operate with limited cross-pollination, suggesting opportunities for integrative research that bridges these otherwise siloed bodies of knowledge.

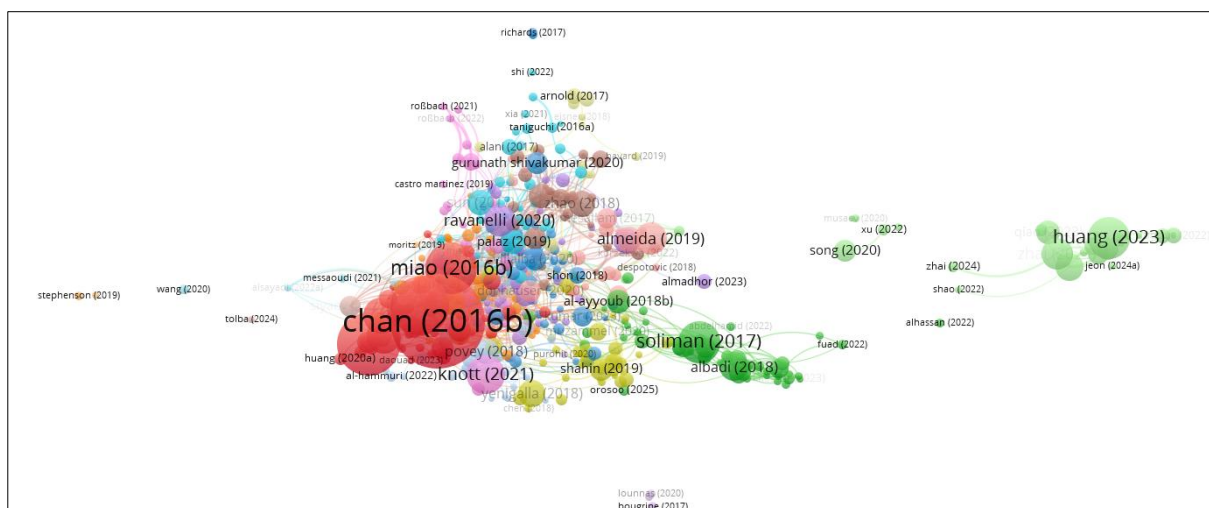


Figure 7: Network visualisation map of the citation by documents

Minimum number of citations of a document = 10

Figure 8 maps the co-authorship network at the country level, revealing a more expansive and densely interconnected global collaboration structure compared to the citation network in Figure 5. China, United States, India, Saudi Arabia, and United Kingdom emerge as the five most prominent hubs, each anchoring distinct yet overlapping clusters. The network demonstrates that Saudi Arabia maintains strong collaborative ties not only within the Arab world connecting with Algeria, Egypt, Morocco, Jordan, and Tunisia but also extends linkages toward Asian nations, reflecting its active role in fostering both regional and international research partnerships in Arabic speech technology.

Critically, the co-authorship network reveals a more globally distributed collaboration pattern relative to the citation network, suggesting that while citation influence remains concentrated among a few dominant nations, collaborative authorship is more inclusive and geographically diverse. The presence of smaller nations such as Ethiopia, Sri Lanka, Mongolia, and Brunei Darussalam on the periphery indicates that emerging research communities are beginning to engage with the global scholarly network, albeit with limited connectivity. However, the continued centrality of China and the United States as primary collaboration hubs reflect a persistent asymmetry in research capacity and influence, where high-income, technologically advanced nations disproportionately shape the direction of international collaborative output in this field.

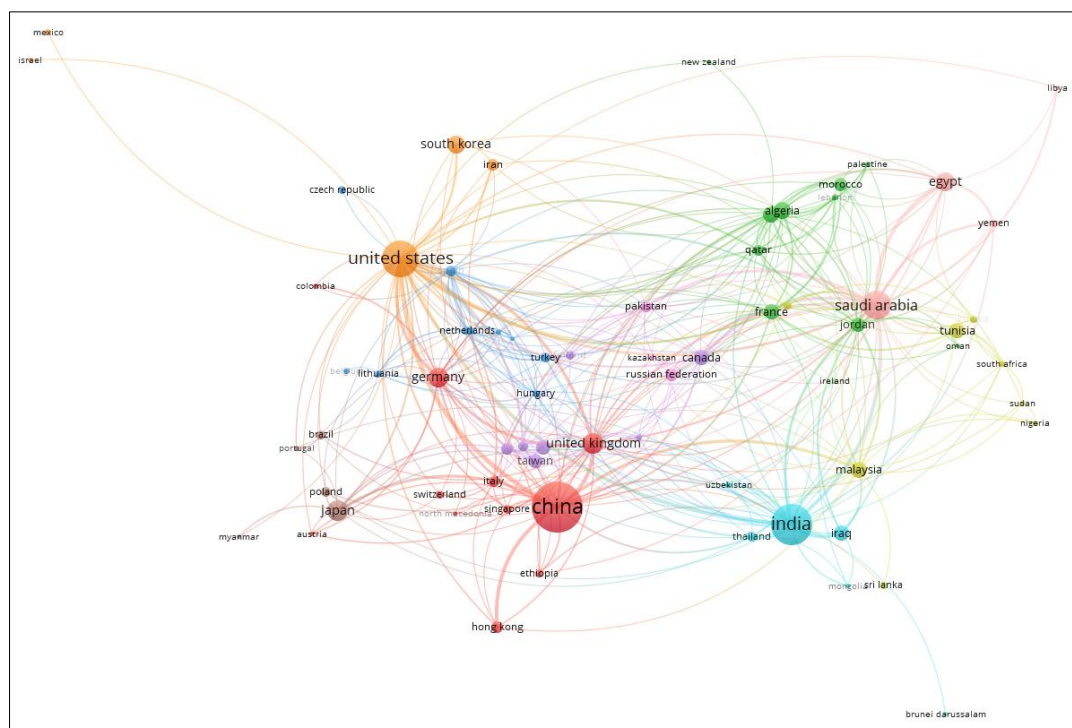


Figure 8: Network visualisation map of the co-authorship

Note: Unit of analysis = Countries; Counting method: Fractional counting; Minimum number of documents of a country = 3; Minimum number of citations of a country = 5

Figure 9 presents the author-level co-authorship network filtered by a minimum of three documents and three citations, revealing a relatively fragmented collaboration structure composed of several small, disconnected clusters. The most visible clusters include a North African group anchored by Laaidi, Naouar and Satori, Hassan (yellow cluster), an Arab-French cluster centred on Lounnac, Khaled, Abbas, Mourad and Lichouri, Mohamed (red cluster), and an English-language cluster around Ali, Ahmed with collaborators such as Glass, James and Hifny, Yasser (green cluster). On the right side, a distinct Asian-Western cluster features Khudanpur, Sanjeev, Watanabe, Shinji and Raj, Dlesh, suggesting cross-institutional collaboration between South Asian and North American researchers.

Critically, the network's most striking characteristic is the absence of a dominant central hub and the lack of inter-cluster connections, indicating that productive author collaborations in this field tend to operate within closed, regionally defined groups rather than across broader international networks. The geographic composition of the clusters North African, Middle Eastern, and Asian-Western mirrors the thematic focus on Arabic and multilingual speech recognition identified in earlier analyses. However, the disconnected nature of these clusters suggests that collaborative synergies between, for instance, the Arab research community and the Asian-Western group remain largely unrealised, representing a structural gap that could be addressed through targeted international research partnerships.

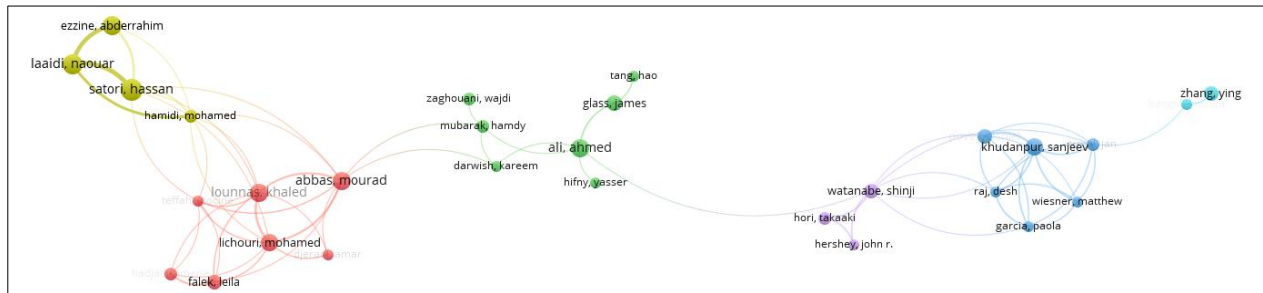


Figure 9: Network visualisation map of the co-authorship

Note: Unit of analysis = Authors; Counting method: Fractional counting; Minimum number of documents of an author = 3; Minimum number of citations of an author = 3

Figure 10 presents a broader author co-authorship network with a lower document threshold but higher citation requirement, surfacing a different set of influential authors. Three prominent clusters dominate the visualisation: a large green cluster anchored by Felden C., Piechocki M., and Debreceny R., a red cluster centred on Liu C. with collaborators including O'farrell G., Sia C.L., and Luo X., and a blue cluster around Wang T. and No W.G.. The node sizes of Felden C. and Liu C. indicate that these authors carry the highest citation weight within their respective clusters, establishing them as the most intellectually influential contributors meeting the citation threshold applied.

Critically, this network presents a starkly different collaboration landscape compared to Figure 9, suggesting that the most cited authors in this corpus are not necessarily the most prolific in terms of document count reflecting a qualitative versus quantitative distinction in research impact. The inter-cluster connections, while present, remain sparse and mediated by only a few bridging authors such as Wang T., who links the green and blue clusters. Notably, the clusters in Figure 10 appear more Western and East Asian in composition, contrasting with the Arab-dominant clusters observed in Figure 9, which may indicate that highly cited works in this field disproportionately originate from researchers in developed academic ecosystems. This disparity underscores the need for greater recognition and integration of contributions from underrepresented research communities within the global citation network.

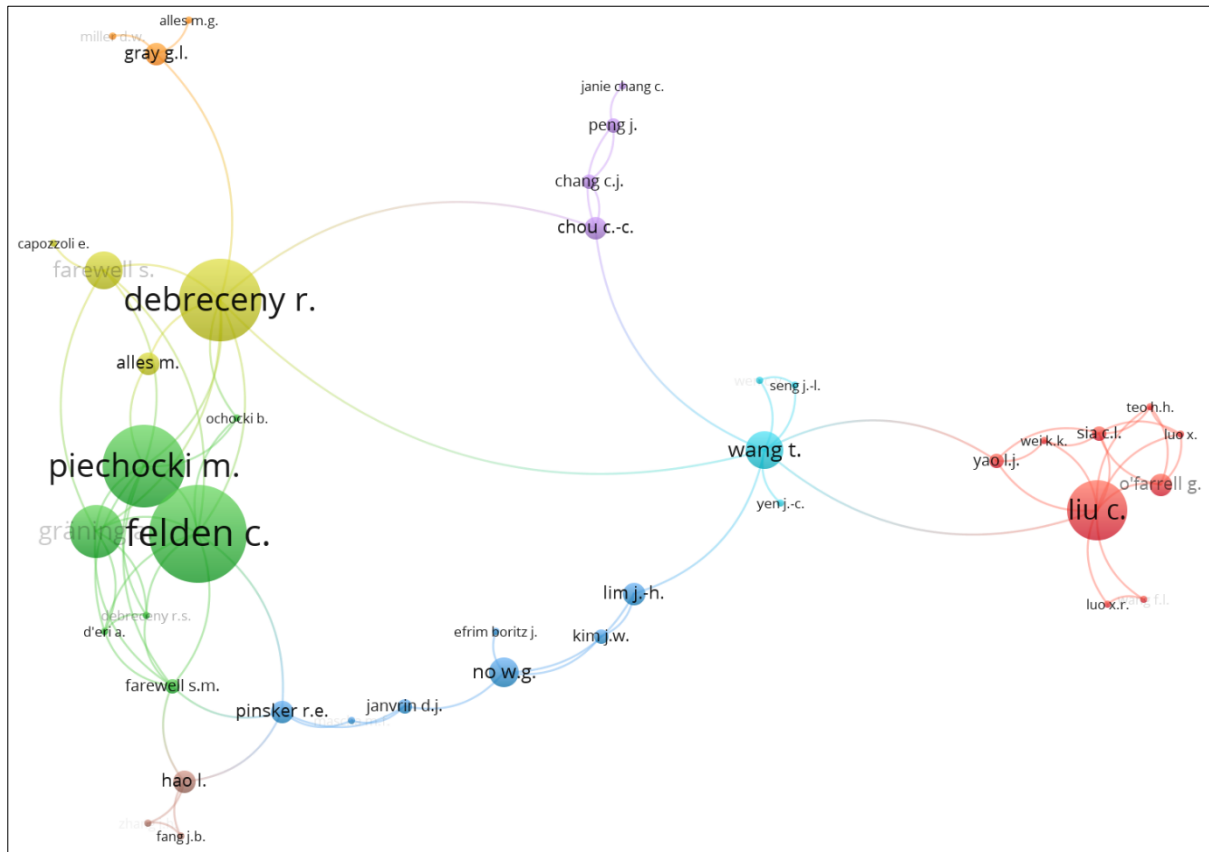


Figure 10: Network visualisation map of the co-authorship

Note: Unit of analysis = Authors; Counting method: Fractional counting; Minimum number of documents of an author = 1; Minimum number of citations of an author = 5

DISCUSSION

In this bibliometric study on AI-based Quranic recitation recognition, the discussion is structured around the research questions (RQs) formulated to examine the development and characteristics of the field. The first research question explores the publication trends and growth of research output over time, highlighting how interest in automated Quranic recitation recognition has evolved within the broader context of artificial intelligence and speech processing. The second research question focuses on the types of documents and publication sources, identifying the dominance of conference proceedings and journal articles, which reflects the field's strong orientation toward rapid technological development and academic dissemination. The third research question examines language distribution, showing a clear dominance of English publications, which indicates the global nature of the research community and limited multilingual scholarly output. The fourth research question investigates the subject area distribution, revealing that Computer Science and Engineering are the primary contributing disciplines, with additional input from Mathematics and other interdisciplinary fields such as Social Sciences and Arts and Humanities. This suggests that the field is inherently multidisciplinary, combining technical AI methods with linguistic and religious context. Overall, these research questions collectively provide a comprehensive understanding of the structure, evolution, and interdisciplinary nature of AI-driven Quranic recitation recognition research.

CONCLUSION

This bibliometric study provides an overview of research trends in AI-based applications in Arabic speech recognition and Quranic recitation recognition. The analysis reveals that the field is highly publication-driven through conference proceedings and journal articles, with English dominating as the

primary language of dissemination. Computer Science and Engineering are the leading subject areas, showing that the research is strongly grounded in technical and computational development, while still incorporating interdisciplinary elements from other fields.

This study contributes by mapping the overall structure of the research landscape in AI-based applications in Arabic speech recognition and Quranic recitations recognition. It identifies key publication patterns, subject area distributions and linguistic trends, offering a clear picture of how the field has developed. The findings can serve as a reference for researchers to understand dominant research directions and potential collaboration areas.

The study is limited by its reliance on indexed databases, which may exclude relevant publications from non-indexed sources or local conferences. Additionally, the analysis is descriptive in nature and does not assess the quality or impact of individual studies in depth beyond bibliometric indicators. Language bias toward English publications may also influence the overall representation of the field.

Future studies can expand the dataset by including multiple databases to achieve broader coverage and reduce publication bias. More advanced analyses, such as citation network analysis, co-authorship networks and thematic evolution, can provide deeper insights into research collaboration and topic development. Additionally, future work may explore domain-specific advancements such as Tajweed-aware AI models and real-time Quranic recitation assessment systems.

ACKNOWLEDGEMENT

This research was sponsored by Universiti Islam Selangor (UIS) for a conference fee and presented at AIDI 2026.

REFERENCES

- Baevski, A., Schneider, S., & Auli, M. (2020). VQ-Wav2Vec: Self-supervised learning of discrete speech representations. *Advances in Neural Information Processing Systems*, 33. <https://doi.org/10.48550/arXiv.1910.05453>
- Chan, W., Jaitly, N., Le, Q. V., & Vinyals, O. (2016). Listen, attend and spell: A neural network for large vocabulary conversational speech recognition. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. <https://doi.org/10.1109/ICASSP.2016.7472621>
- Chiu, C.-C., Sainath, T. N., Wu, Y., Prabhavalkar, R., Nguyen, P., Chen, Z., Kannan, A., Weiss, R. J., Rao, K., Gonina, E., Jaitly, N., Li, B., Chorowski, J., & Bacchiani, M. (2018). State-of-the-art speech recognition with sequence-to-sequence models. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. <https://doi.org/10.1109/ICASSP.2018.8462105>
- Huang, X., Zou, D., Cheng, G., Chen, X., & Xie, H. (2023). Trends, research issues and applications of artificial intelligence in language education. *Educational Technology & Society*, 26(1), 1–15. [https://doi.org/10.30191/ETS.202301_26\(1\).0001](https://doi.org/10.30191/ETS.202301_26(1).0001)
- Knott, B., Venkataraman, S., Hannun, A., Sengupta, S., Ibrahim, M., & van der Maaten, L. (2021). CryptTen: Secure multi-party computation meets machine learning. *Advances in Neural Information Processing Systems*, 34. <https://doi.org/10.48550/arXiv.2109.00984>

- Miao, Y., Gowayyed, M., & Metze, F. (2016). EESSEN: End-to-end speech recognition using deep RNN models and WFST-based decoding. In *Proceedings of the IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*. <https://doi.org/10.1109/ASRU.2015.7404828>
- Prabhavalkar, R., Rao, K., Sainath, T. N., Li, B., Johnson, L., & Jaitly, N. (2017). A comparison of sequence-to-sequence models for speech recognition. In *Proceedings of Interspeech 2017*. <https://doi.org/10.21437/Interspeech.2017-233>
- Rao, K., Sak, H., & Prabhavalkar, R. (2017). Exploring architectures, data and units for streaming end-to-end speech recognition with RNN-transducer. In *Proceedings of the IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*. <https://doi.org/10.1109/ASRU.2017.8268945>
- Soliman, A. B., Eissa, K., & El-Beltagy, S. R. (2017). AraVec: A set of Arabic word embedding models for use in Arabic NLP. *Procedia Computer Science*, 117, 256–265. <https://doi.org/10.1016/j.procs.2017.10.117>
- Watanabe, S., Hori, T., Kim, S., Hershey, J. R., & Hayashi, T. (2017). Hybrid CTC/attention architecture for end-to-end speech recognition. *IEEE Journal of Selected Topics in Signal Processing*, 11(8), 1240–1253. <https://doi.org/10.1109/JSTSP.2017.2763455>