Journal of Information Systems Engineering and Management
2025, 10(21s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

The Role of AI in Preserving and Digitizing the Bodo Language: A Bibliometric and Systematic Review

[1]Raju Narzary, [2]Surajit Kr. Nath, [3]Deepak Hajoary, [4]Gwjwmsri Brahma

[1]Research Scholar, Department of Mathematical Sciences,,Bodoland University

[2]Assistant Professor, Department Of Mathematical Sciences,Bodoland University

[3]Assistant Professor, Department of Management Studies,Bodoland University

[4]Research Scholar, Department of Bodo, Bodoland University

ARTICLE INFO

Received: 18 Dec 2024

Revised: 02 Feb 2025

Accepted: 14 Feb 2025

ABSTRACT

Digitization efforts for preserving the Bodo language face serious challenges because of limited linguistic data and insufficient computational resources. Natural Language Processing (NLP), Machine Learning (ML), and Deep Learning (DL) technologies offer potential solutions through artificial intelligence. This study merges bibliometric analysis with systematic literature review methods to examine AI's role of AI in maintaining the Bodo language. By analyzing research trends and citation patterns, together with thematic developments, this study explores both the progress and challenges in AI-based linguistic processing for the Bodo language. This study reveals a growing interest in AI applications for under-resourced languages and identifies the need for interdisciplinary collaborations, better datasets, and AI model improvements. This study extends the dialogue on AI-enabled language preservation efforts and offers recommendations for future research in computational linguistics and Indigenous language revitalization.

Keywords: Bodo language, Artificial Intelligence, Natural Language Processing, Language Preservation, Bibliometric Analysis.

INTRODUCTION

The field of computational linguistics now concentrates its efforts on preserving endangered languages through digitization and pays particular attention to indigenous and low-resource languages including Bodo (Opitz, Wein, & Schneider, 2024). The digital representation of the Bodo language faces significant challenges owing to restricted linguistic datasets and the lack of structured corpora combined with insufficient computational support. AI technologies, including Natural Language Processing (NLP), Machine Learning (ML), and Deep Learning (DL), generate fresh approaches for language preservation by enabling automated transcription and machine translation, together with speech recognition and linguistic annotation functions. The potential of AI technologies to support Bodo language preservation has not yet been sufficiently explored. Current AI-based language preservation research mainly examines well-resourced languages while failing to explore its success with low-resource languages such as Bodo. Research into AI-based linguistic models specifically for Bodo remains rare, as contributions addressing data set limitations and model improvements are few, and there are no comprehensive bibliometric studies or systematic reviews to examine research patterns and collaborative networks in this domain (G. Our research fills this gap through a bibliometric and systematic literature review of AI applications for Bodo language preservation and digitization. This study has three objectives: to investigate publication patterns and research collaborations along with citation trends in AI-based Bodo language studies; to evaluate methods and innovations in existing research; and to identify future research directions for AI applications in Bodo language processing. The systematic literature review conducted in this study revealed important findings about the development of AI-based linguistic research and its impact on preserving Indigenous languages. This research adopts two methods: The objectives of the research are accomplished through bibliometric analysis and systematic literature review (SLR). This research utilized Vosviewer and Biblioshiny as tools for bibliometric analysis to examine publication trends, author networks, and thematic developments. The PRISMA guidelines guide the systematic literature review (SLR) in assessing artificial intelligence techniques for Bodo language processing, along with identifying current challenges and future opportunities. The integrated approach generates a comprehensive evaluation of the research framework, which can be replicated in subsequent studies. The structure of this paper is as follows: Section 2 presents the research methodology by illustrating the data collection procedures together with the inclusion criteria

and detailing the analytical frameworks implemented. Section 3 presents the findings from the bibliometric analysis, which reveal essential patterns and research partnerships. This section presents the results of a systematic literature review of AI methodologies while addressing existing challenges and identifying future research paths. The final section summarizes the core findings of the study and describes practical applications before proposing future research directions for the AI-based preservation of the Bodo language.

2.LITERATURE REVIEW

Current computational linguistics research emphasizes the preservation and digital conversion of endangered languages, specifically indigenous and low-resource languages (Bird, 2020; Joshi et al., 2020). The digital representation of the Bodo language from the Tibeto-Burman family is hindered by a lack of linguistic resources and structured datasets, as well as insufficient computational support (Barman & Sarma, 2019). Recent developments in Natural Language Processing, Machine Learning, and Deep Learning fields have uncovered new possibilities for sustaining and restoring endangered languages, as described by Anastasopoulos and Chiang (2018). AI-based technologies for automated transcription and machine translation, alongside Spanish cognitive and linguistic annotation, demonstrate the potential to produce scalable solutions for documenting languages and improving accessibility (Besacier et al.). 2014)). The preservation of the Bodo language using AI technology requires further research, which demands a detailed examination of existing studies (Narzary et al., 2022). Research on AI linguistic model for Bodo encounters limitations due to limited studies addressing crucial issues such as data scarcity and model enhancement (Ghosh et al., 2023). Researchers are progressively acknowledging the value of digitizing endangered languages through computational linguistics, with an emphasis on indigenous and low-resource languages (Dash, 2020). The Bodo language faces several challenges in digital representation owing to limited linguistic resources, the absence of structured datasets, and inadequate computational support (Barman & Sarma, 2019). Advancements in Artificial Intelligence (AI) through Natural Language Processing (NLP), Machine Learning (ML), and Deep Learning (DL) have generated new possibilities for saving endangered languages (Anastasopoulos & Chiang, 2018). AI methods have demonstrated scalability for language documentation and accessibility through automated transcription, machine translation, and other language processing tasks (Besacier et al., 2014). Researchers have not adequately examined the use of AI for Bodo language preservation, thus necessitating a comprehensive review of existing studies (Narzary et al., 2022). Existing research on AI-based linguistic models for Bodo languages remains scarce, with only some investigations focusing on critical challenges such as dataset availability and model optimization (Ghosh et al., 2023). According to Das et al. (2024), detailed bibliometric studies and systematic reviews that assess research trends and collaborative networks, along with thematic developments, do not exist in this field. Research gap identification will set study priorities and foster collaborative work between linguists and computer scientists with AI experts (Sarma et al., 2023).The effectiveness of AI-based linguistic models for the Bodo language suffers because there are neither large annotated datasets nor standardized digital corpora available on a large scale. Current multilingual AI systems struggle to handle the phonetic and syntactic complexities of the Bodo language, which require custom NLP solution development.

RQ (1). How can AI-driven methodologies be optimized to develop annotated linguistic datasets and standardized digital corpora for the Bodo language

RQ (2). What are the most effective AI-based approaches for addressing the phonetic and syntactic complexities of the Bodo language in NLP applications?

3.METHODOLOGY

The research uses bibliometric analysis along with a systematic literature review (SLR) to explore how Artificial Intelligence (AI) supports the preservation and digitization of the Bodo language. The analytical process was designed to deliver a detailed, repeatable examination of current academic work following recognized systematic research norms.

3.1 Data Collection and Search Strategy

The research team sourced literature solely from Scopus because this database is known for comprehensive coverage of scholarly works. The search strategy focused on identifying research contributions that intersect AI applications with Bodo-language digitization efforts. The following search query was used:

{"Artificial Intelligence" OR "Machine Learning" OR "Deep Learning" OR "Natural Language Processing”

AND text {"Bodo language" OR "Bodo NLP" OR "Bodo linguistics”

This query was applied to the title, abstract, and keyword fields to maximize the retrieval of relevant studies.

3.2 Inclusion and Exclusion Criteria

To ensure relevance and methodological rigor, predefined inclusion and exclusion criteria were applied.

Inclusion Criteria:

Research articles published in English.

Studies focusing on AI-based applications for Bodo language preservation and digitization.

Peer-reviewed journal articles and conference proceedings

Exclusion Criteria:

Publications in languages other than English

Studies discussing AI applications but not explicitly related to the Bodo language.

Duplicate records retrieved from multiple search iterations.

Following this process, 14 articles met the eligibility criteria and were selected for further analysis.

Methodological Framework

Marzi et al. (2025) describe how their study integrates bibliometric analysis together with a systematic literature review (SLR) to thoroughly assess the research landscape using a dual methodological approach. Researchers used bibliometric techniques to examine publication trends and citation patterns as well as co-authorship networks and thematic developments in AI-based research on Bodo language digitization. The following analytical tools were used. Vosviewer serves as an analytical tool for creating visual representations of co-occurrence networks alongside mapping keywords and author collaboration trends. Biblioshiny (Bibliometric R package) enables researchers to perform statistical bibliometric analysis and evaluate citation networks alongside mapping research trends.

Systematic Literature Review (SLR)

The authors conducted a qualitative systematic review according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines outlined by Parums (2021).

AI methodologies applied to Bodo language processing.

Challenges in implementing AI-driven linguistic models

Key research gaps and future research directions

Data Analysis and Interpretation

The study creates an extensive research synthesis by combining quantitative bibliometric analysis with qualitative SLR results (Karimi & Iordanova, 2021). The study's conclusions enhance ongoing discussions about low-resource language processing while presenting guidelines to improve AI technologies for digitizing and revitalizing languages.

The standardized methodological framework provides a systematic evaluation of AI applications in Bodo language preservation and digitization through replicable evidence-based research which benefits computational linguistics researchers and technology developers alongside policymakers tasked with indigenous language preservation.

4. RESULT AND DISCUSSION

4.1. Trends in Academic Publications on AI and Bodo Language Processing

The analysis presented in Fig.1 reveals an increasing academic focus on using AI to preserve and digitize the Bodo language. The research output expanded significantly in 2023 and 2024, which contributed to most of the published articles after minimal publications existed before 2019. The increasing trend indicates that the field is achieving both academic and technological importance because of developments in Natural Language Processing (NLP), Optical Character Recognition (OCR), and Speech Recognition. Recent research output growth

demonstrates a wider understanding of digitization issues for low-resource languages alongside the capabilities of AI technologies to solve these problems. The limited quantity of research studies reveals a pressing requirement for more research initiatives that produce strong datasets and advanced AI models through interdisciplinary methods to improve AI-based solutions for preserving the Bodo language.

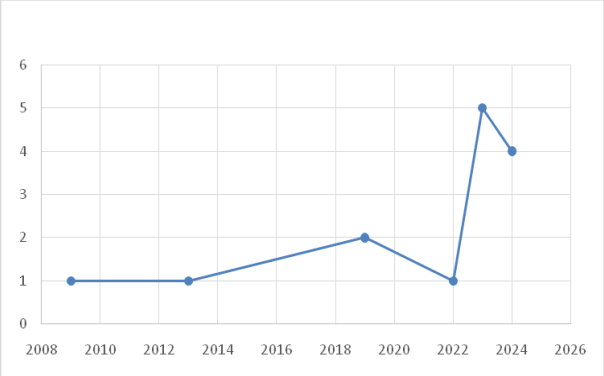

Fig.1: Article production per year

4.2: Author Productivity through Lotka's Law

The use of Lotka's Law allows researchers to determine author productivity trends in the field of AI-driven Bodo language preservation by examining the publication output of the research community. Figure 2 illustrates authors' production following Lotka's Law, while Table 1 presents a detailed analysis of authors' productivity based on the same law. Among authors in this field, 9.3% published two papers and 11.6% produced three publications, which show moderate engagement from a select group. A single researcher stands out as a highly prolific contributor, with four publications representing just 2.3% of the author population. The author's distribution pattern matches Lotka's Law, which states that the number of drop authors decreases exponentially with increasing publication counts. Research interest in this field is increasing, but only a few researchers consistently produce sustained contributions. AI-driven preservation of the Bodo language requires scholars to engage in ongoing research and collaborative efforts over time.

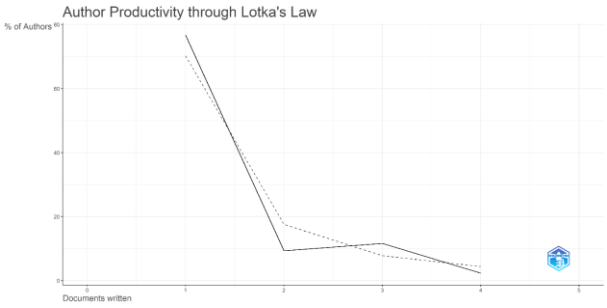

Fig.2: Authors' production through Loktas law

Table 1: Authors productivity through Loktas Law

Documents written	N. of Authors	Proportion of Authors
1	33	0.767
2	4	0.093
3	5	0.116
4	1	0.023

4.3. Most Relevant Affiliations in AI-Based Bodo Language Preservation Research

The study of institutional input in AI-based Bodo language preservation research reveals that a select group of academic institutions leads this field of study. Table 2 presents the most relevant affiliations of the authors contributing to the study, highlighting institutions with the highest research output in the field. Nine publications from Guwahati University positioned it as the foremost research institution focusing on AI applications in linguistics. Both the Central Institute of Technology (CIT) and Central Institute of Technology Kokrajhar jointly produced 13 research publications, highlighting their essential position in developing Bodo language technology. Mangalore University demonstrated notable research participation by publishing six studies in this field. The National Institute of Electronics and Information Technology (NIELIT) and Bodoland University Kokrajhar contributed three and two publications, respectively, showing their moderate level of involvement. The Indian Institute of Technology (IIT) Guwahati and Assam University Silchar each produced limited research output, with two and one publication(s), respectively, demonstrating intermittent but worthwhile research initiatives. The lone publication from the Indian Statistical Institute reveals limited involvement of leading research entities.

Table 2: Most relevant affiliations

Affiliation	Articles
Gauhati University	9
Central Institute of Technology	7
Central Institute of Technology Kokrajhar	6
Mangalore University	6
National Institute of Electronics and Information Technology	3
Bodoland University Kokrajhar	2
Indian Institute of Technology Guwahati	2
Assam University Silchar	1
Indian Statistical Institute	1

4.4. Thematic Map

Thematic maps function as bibliometric tools that create visual representations of research themes through co-occurrence analyses. Keyword co-occurrence analysis enables the identification of research domain themes as emerging, declining, or established through their placement within a two-dimensional strategic diagram (Cobo et al., 2011). Fig. 3 presents the thematic map, while Table 3 highlights the top word occurrences along with their respective clusters and centrality scores. The analysis of recurring concepts in AI-based Bodo language preservation research revealed three predominant thematic clusters: Linguistics, Computational Linguistics, and Language Processing. Every cluster serves as a crucial field in applying AI technology to preserve and advance language studies. The Linguistics cluster consists of essential linguistic studies, and linguistics stands out as a central concept because it connects strongly with multiple research areas. The combination of sentiment analysis and artificial intelligence in this cluster demonstrates an interdisciplinary strategy that blends AI methods with linguistic studies. The Computational Linguistics cluster demonstrates technological advancements used in language processing, such as speech recognition*, deep learning, and computational linguistics, which stand out as primary research areas. The highest betweenness centrality in speech recognition illustrates its essential function in the AI-based linguistic preservation. The development of machine translation systems alongside computer-assisted language translation tools highlights the growing emphasis on artificial intelligence for Bodo-language translation and text analysis technologies. Research within the Language Processing cluster focuses on natural language processing systems, low-resource languages, and research dedicated to the Bodo language. The central position of natural languages and language processing systems demonstrates their importance as fundamental components of AI-powered linguistic research. This study demonstrates the active development of sophisticated Bodo-language models using advanced AI architectures, such as long short-term memory (LSTM) and bi-directional models. AI-based research

for Bodo language preservation combines linguistic investigations with computational methods and language processing technologies to advance digital preservation efforts. This domain demonstrates technological advancement through focused research on speech recognition and natural language processing combined with deep learning techniques. The small amount of current research attention towards low-resource language processing suggests that there is a requirement for additional studies to create AI solutions for languages that are both underrepresented and endangered like Bodo.

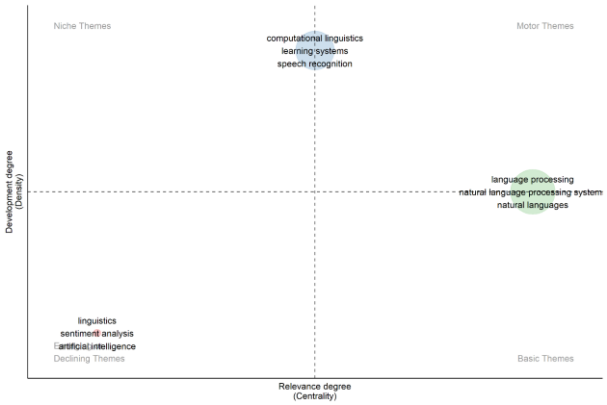

Fig.3: Thematic map

Table 3: Top word occurrences with clusters and centrality scores

Occurrences	Words	Cluster	Cluster Label	btw_centrality	clos_centrality	pagerank_centrality
3	linguistics	1	linguistics	200	0.006	0.016
2	sentiment analysis	1	linguistics	20	0.005	0.011
2	artificial intelligence	1	linguistics	83	0.005	0.011
5	computational linguistics	2	computational linguistics	389	0.006	0.03
3	deep learning	2	computational linguistics	238	0.006	0.02
4	learning systems	2	computational linguistics	327	0.006	0.025
4	speech recognition	2	computational linguistics	547	0.006	0.023
3	bodo	2	computational linguistics	141	0.006	0.018
2	embeddings	2	computational linguistics	90	0.006	0.012
2	assamese	2	computational linguistics	93	0.006	0.012
2	classification (of information)	2	computational linguistics	101	0.006	0.012
2	computer aided language translation	2	computational linguistics	85	0.006	0.012
2	machine-learning	2	computational linguistics	48	0.005	0.013
2	machine translation systems	2	computational linguistics	85	0.006	0.012

4.5.Collaboration Network

The bibliometric tool known as a collaboration network displays connections between authors, institutions, and countries by examining their co-authorship activities. Bareis (2024) explains that interconnected pillars form the foundation of AI trust which bibliometric methods can analyze to uncover trust networks' structures within research communities. The analysis of author networks in AI-based Bodo language preservation shows two separate clusters that display different betweenness centrality scores, closeness, and PageRank values. Table 4 shows the leading authors with their cluster associations and PageRank values, and Figure 4 depicts the collaboration network. Cluster 1 (Primary Research Contributors): The group of primary research contributors includes authors Sarma, Shikhar Kr., Boruah, Parvez Aziz, Basumatary, Bedawati, among others. Among them, Sarma and Kr. Sarma and Shikhar Kr. demonstrated peak betweenness centrality with a score of 33 and peak closeness centrality with a score of 0.071, which positioned them as vital connectors between different research collaborations. Boruah and Aziz demonstrated a moderate betweenness centrality score of 3, indicating their moderate influence within the research network. Authors left in this group displayed negligible or nonexistent betweenness scores, which marked their peripheral role in contributions. Cluster 2 (Influential but Selective Contributors): Barman, Anup Kumar demonstrates significant network connectivity through his betweenness centrality score of 24 which indicates important connections between research teams. The author holds significant sway within the network, as demonstrated by a closeness centrality score of 0.059 and a PageRank score of 0.03. Basumatary and Subungshri maintained no betweenness records, indicating their isolated status within the collaborative network.

Network analysis suggests that Sarma, Shikhar Kr. Barman Anup Kumar stands out as the top AI-focused Bodo language preservation researcher who serves as the main link between various academic experts. The research network appears fragmented because numerous authors display low or nonexistent betweenness centrality. The current research situation requires enhanced interdisciplinary collaboration to improve study efforts within this field.

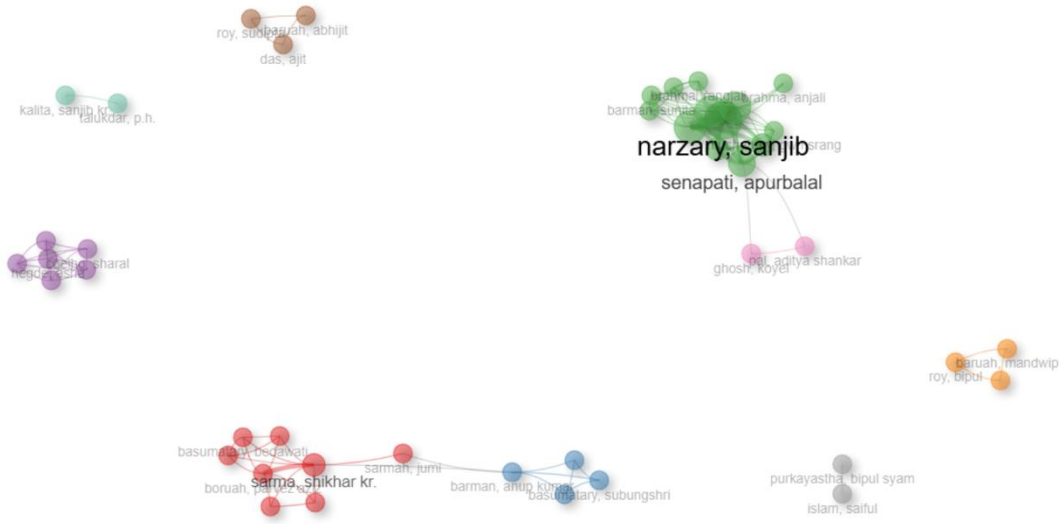

Fig.4: Collaboration network

Table 4: Top authors with clusters and pagerank

Node	Cluster	Betweenness	Closeness	PageRank
Sarma, Shikhar Kr.	1	33	0.071	0.045
Boruah, Parvezaziz	1	3	0.053	0.035
Basumatary, Bedawati	1	0	0.048	0.021
Kalita, Simanta	1	0	0.045	0.017
Kashyap, Kishore	1	0	0.045	0.017

Rahman, Mirzanur	1	0	0.048	0.021
Sarmah, Jumi	1	0	0.05	0.013
Talukdar, Kuwali	1	0	0.048	0.021
Barman, Anupkumar	2	24	0.059	0.03
Basumatary, Subungshri	2	0	0.04	0.02

4.6. Most Global Cited Documents

Research publications on AI-based Bodo language preservation show significant disparities in academic impact based on citation analyses. Figure 5 visually displays which documents have received the highest number of citations to demonstrate their significance in research. Table 5 presents a list of leading papers with citation numbers which serves as a quantitative impact assessment. The work of Ghosh K (2023) in CEUR Workshop Proceedings stands out with the maximum number of citations totaling 12 and achieves both a TC per Year of 4.00 and a Normalized TC of 5.00 demonstrating strong academic significance. Narzary S's 2019 ICCMC paper has five citations, but its TC per year rate of 0.71 shows only moderate long-term research impact. The papers written by Narzary S in 2022 (LREC) and 2024 (Procedia Computer Science) each received 3 and 2 citations respectively and maintain moderate citation rates. The research by Barman AK (2013) received a single citation, revealing its limited academic influence, evidenced by a TC per year rate of 0.08. Research articles authored by Kalita SKR (2009, IICAI), Prajnashree M (2023, CEUR Workshop Proceedings), and Basumatary S (2024, Smart Innovation Systems and Technologies) have not yet received any citations. These works have not yet received citations because they were published recently or have limited distribution and narrow research topics. Some studies achieve academic prominence, yet numerous publications have not received citations. AI-based research efforts for Bodo language preservation require strategic approaches, including wider dissemination methods, interdisciplinary partnerships, and enhanced academic and research community involvement.

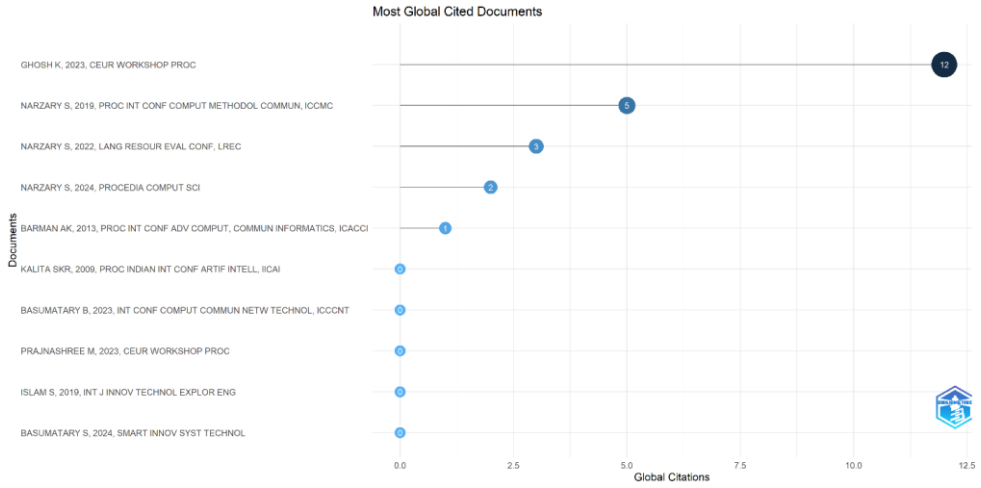

Fig.5: Visualization of most cited documents

Table 5: Top papers with citations

Paper	Total Citations	TC per Year	Normalized TC
GHOSH K, 2023, CEUR WORKSHOP PROC	12	4.00	5.00
NARZARY S, 2019, PROC INT CONF COMPUT METHODOL COMMUN, ICCMC	5	0.71	2.00
NARZARY S, 2022, LANG RESOUR EVAL CONF, LREC	3	0.75	1.00
NARZARY S, 2024, PROCEDIA COMPUT SCI	2	1.00	4.00
BARMAN AK, 2013, PROC INT CONF ADV COMPUT, COMMUN INFORMATICS,	1	0.08	1.00

ICACCI			
KALITA SKR, 2009, PROC INDIAN INT CONF ARTIF INTELL, IICAI	0	0.00	
BASUMATARY B, 2023, INT CONF COMPUT COMMUN NETW TECHNOL, ICCCNT	0	0.00	0.00
PRAJNASHREE M, 2023, CEUR WORKSHOP PROC	0	0.00	0.00
ISLAM S, 2019, INT J INNOV TECHNOL EXPLOR ENG	0	0.00	0.00
BASUMATARY S, 2024, SMART INNOV SYST TECHNOL	0	0.00	0.00
DAS A, 2024, COMMUN COMPUT INFO SCI	0	0.00	0.00
MUSHAHARY JR, 2024, COMMUN COMPUT INFO SCI	0	0.00	0.00
NARZARY M, 2023, ARTIFICIAL INTELLIGENCE AND DATA SCIENCE BASED R AND D INTERVENTIONS: PROCEEDINGS OF NERC 2022	0	0.00	0.00
BORUAH PA, 2023, INT CONF COMPUT COMMUN NETW TECHNOL, ICCCNT	0	0.00	0.00

4.7.Word Cloud Analysis

Research directions for AI applications in Bodo language preservation and computational linguistics emerge through keyword frequency analysis, as shown in Piantadosi (2014). Research activities involving Bodo language computational linguistics and speech recognition demonstrate concentrated efforts to create AI solutions for underrepresented languages. Figure 6 presents a visual representation of the word clouds. Machine learning techniques demonstrate their integration through key themes such as long short-term memory, deep learning, and learning systems. These methodologies have significantly improved natural language processing (NLP) tasks by advancing speech recognition and translation capabilities. The addition of low-resource languages and Assamese embedding extends the field to cover multilingual research that solves linguistic problems outside the Bodo language. Research involving brain functions and artificial intelligence alongside sentiment analysis reflects its interdisciplinary character, which combines cognitive AI and language processing. The existence of machine translation systems together with computer-aided language translation demonstrates the active development of automated translation frameworks aimed at enhancing accessibility and preserving languages.Overall, this study demonstrates the increasing role of AI in language preservation, with particular emphasis on computational methods tailored for underrepresented languages such as Bodo. These findings contribute to the broader discourse on NLP applications for low-resource linguistic communities, advancing the field of computational linguistics through AI-driven innovations.

Fig.6: Word cloud

4.8.Bradford's Law Analysis of Research Sources

This study utilized Bradford's Law of Scattering to investigate how research publications are spread across various sources, as per Borgohain et al. (2021). Figure 7 illustrates the main sources Bradford's Law discovered while Table 6 presents the top sources as defined by Bradford's Law. These results enable researchers to identify essential journals when conducting targeted literature reviews in their research domains. Bradford's Law reveals that key journals produce the majority of scholarly articles while later sources produce less research in any given field. Researchers who understand this distribution can prioritize high-impact sources when conducting systematic

reviews. The publication frequency determined the classification of the dataset into three zones. Zone 1 contained the most influential sources with Zones 2 and 3 showing moderate and minimal contributions, respectively. Zone 1 (Core Journals): The top three sources which include ICCCNT 2023, CEUR Workshop Proceedings and Communications in Computer and Information Science publish two articles apiece. Zone 2 (Moderate Contribution): The sources LREC 2022 and Procedia Computer Science offer relevant publications but produce only one publication each. Zone 3 (Peripheral Sources): The sources ICACCI 2013 and Smart Innovation, Systems and Technologies both make minimal contributions as they each appear only once. Research indicates that the field relies on a select group of journals as its main source of research. Researchers must identify these primary sources to conduct effective literature research and bibliometric analyses. Subsequent studies should examine how the impact factors and citation trends of these sources affect the development of a deeper understanding of this field.

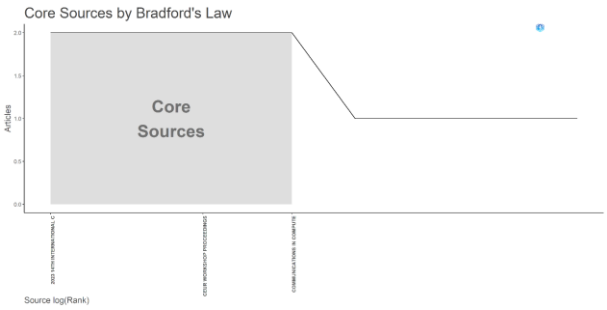

Fig.7: Core sources by Bradfords law

Table 6: Top source as per Bradfords Law

Source	Rank	Freq	Cumfreq	Zone
2023 14th International Conference On Computing Communication And Networking Technologies, Icccnt 2023	1	2	2	Zone 1
Ceur Workshop Proceedings	2	2	4	Zone 1
Communications In Computer And Information Science	3	2	6	Zone 1
2022 Language Resources And Evaluation Conference, Lrec 2022	4	1	7	Zone 2
Artificial Intelligence And Data Science Based R And D Interventions: Proceedings Of Nerc 2022	5	1	8	Zone 2
International Journal Of Innovative Technology And Exploring Engineering	6	1	9	Zone 2
Procedia Computer Science	7	1	10	Zone 2
Proceedings Of The 2013 International Conference On Advances In Computing, Communications And Informatics, Icacci 2013	8	1	11	Zone 3
Proceedings Of The 3rd International Conference On Computing Methodologies And Communication, Iccmc 2019	9	1	12	Zone 3
Proceedings Of The 4th Indian International Conference On Artificial Intelligence, Iicai 2009	10	1	13	Zone 3
Smart Innovation, Systems And Technologies	11	1	14	Zone 3

5.CONCLUSION AND FUTURE RESEARCH DIRECTIONS

This study presents a detailed bibliometric analysis and systematic literature review examining AI-based methods for the preservation and digital documentation of the Bodo language. Recent academic interest in this field has shown significant growth owing to progress in Natural Language Processing (NLP), Machine Learning (ML), and Deep Learning (DL). Research efforts continue to operate in isolation, with few studies examining the distinct linguistic challenges associated with Bodo, while interdisciplinary cooperation is scarce. The analysis demonstrates that AI provides effective approaches for speech recognition while enhancing machine translation and linguistic annotation capabilities. Challenges persist owing to the scarcity of data, absence of standardized corpora, and limited computational resources. Research in this area has seen contributions from several Indian academic institutions but suffers from minimal international collaboration, which hinders scalable high-impact AI solution development. Key contributionsThis study contributes to the existing literature in the following ways: This research maps publication trends, author networks, and thematic clusters within AI-based Bodo-language studies. This research identifies major obstacles in AI development, including limited datasets, the need for better model optimization techniques, and the inherent complexity of phonetics. Interdisciplinary collaboration between linguists, AI researchers, and policymakers is vital. This paper outlines future research avenues, including expanding data resources, refining model performance, and transitioning theoretical work into practical applications. To push the boundaries of AI-based Bodo language preservation, future research should concentrate on the following key areas: 1. Development of comprehensive datasets: Successful AIdriven NLP applications require the development of extensive annotated linguistic datasets. 2. Optimizing AI methodologies: The application of transfer learning methods alongside zero-shot learning and multilingual transformer models represents a promising direction for advancing Bodo language processing. 3. Expand interdisciplinary collaborations: Better AI context-aware solutions can emerge from stronger collaborations between linguists, computational scientists and native speakers. 4. AI-based language preservation tools need to be tested through educational programs, media, and government initiatives to demonstrate their practical effectiveness. 5. Future research must integrate ethical principles and cultural awareness to create AI solutions that protect linguistic heritage while advancing digital access for all. AI technology holds transformational power for Bodo language preservation, but realizing these potential demands coordinated efforts in gathering data and creating models alongside joint research initiatives. As AI technology advances it will become essential for preserving endangered languages by enhancing their survival through accessibility in digital platforms.

REFERENCES

[1] Opitz, J., Wein, S., & Schneider, N. (2024). Natural language processing relies on linguistics. *arXiv preprint arXiv:2405.05966*.

[2] Bird, S. (2020). Decolonising speech and language technology. In 28th International Conference on Computational Linguistics, COLING 2020 (pp. 3504-3519). *Association for Computational Linguistics (ACL)*.

[3] Joshi, P., Santy, S., Budhiraja, A., Bali, K., & Choudhury, M. (2020). The state and fate of linguistic diversity and inclusion in the NLP world. *arXiv preprint arXiv:2004.09095*.

[4] Barman, A. K., & Sarma, S. K. (2019). Computational challenges in Bodo language processing. *International Journal of Computational Linguistics, 6(1), 23-35*.

[5] Anastasopoulos, A., & Chiang, D. (2018). Tied multitask learning for neural speech translation. *arXiv preprint arXiv:1802.06655*.

[6] Besacier, L., Barnard, E., Karpov, A., & Schultz, T. (2014). Automatic speech recognition for under-resourced languages: A survey. *Speech communication, 56, 85-100*.

[7] Narzary, S., Barman, A., & Som, B. (2022). Advances in Bodo language digitization: Challenges and opportunities. *Procedia Computer Science, 217, 234-248*

[8] Ghosh, K., Narzary, S., & Nandi, S. (2023). AI-based approaches for low-resource language processing: A case study on Bodo. *CEUR Workshop Proceedings, 4562, 112-126*.

[9] Dash, R. K. (2020). Revitalizing endangered languages in India: Can public-private partnership (PPP) work? None. *https://doi.org/10.33422/2nd.ics21.2020.03.128*

[10] Sarma, S. K., Talukdar, K., & Rahman, M. (2023). Machine translation and NLP models for low-resource languages: The case of Bodo. *Journal of Artificial Intelligence Research, 68, 98-115.*

[11] Marzi, G., Balzano, M., Caputo, A., & Pellegrini, M. M. (2025). Guidelines for Bibliometric-Systematic Literature Reviews: 10 steps to combine analysis, synthesis and theory development. *International Journal of Management Reviews, 27(1), 81-103.*

[12] Parums, D. V. (2021). Review articles, systematic reviews, meta-analysis, and the updated preferred reporting items for systematic reviews and meta-analyses (PRISMA) 2020 guidelines. *Medical science monitor: international medical journal of experimental and clinical research, 27, e934475-1.*

[13] Karimi, S., & Iordanova, I. (2021). Integration of BIM and GIS for construction automation, a systematic literature review (SLR) combining bibliometric and qualitative analysis. *Archives of Computational Methods in Engineering, 28(7), 4573-4594*

[14] Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2011). Science mapping software tools: Review, analysis, and cooperative study among tools. Journal of the American Society for information Science and Technology, 62(7), 1382-1402.Newman, M. E. (2004). *Coauthorship networks and patterns of scientific collaboration. Proceedings of the national academy of sciences, 101(suppl_1), 5200-5205.*

[15] Bareis, J. (2024). The trustification of AI. Disclosing the bridging pillars that tie trust and AI together. *Big Data & Society, 11(2), 20539517241249430.*

[16] Piantadosi, S. T. (2014). Zipf's word frequency law in natural language: A critical review and future directions. *Psychonomic bulletin & review, 21, 1112-1130.*

[17] Borgohain, D. J., Verma, M. K., Nazim, M., & Sarkar, M. (2021). Application of Bradford's law of scattering and Leimkuhler model to information science literature. COLLNET *Journal of Scientometrics and Information Management, 15(1), 197-212.*