ECOLINGUISTIC STUDY WITH AI-BASED FRAMEWORK ON NIAS MEDICINAL PLANTS

Made Sani Damayanthi Muliawan[1], I Nyoman Muliana[2] I.G.A.A Dian Susanthi[3]
[1,2,3] Master of Linguistics Program
Postgraduate Warmadewa University, Denpasar, Bali, Indonesia
(muliawansanidama@gmail.com[1], inyomanmulianabali@gmail.com[2], gungdian03@gmail.com[3])

Abstract

Indigenous medicinal plant knowledge constitutes a crucial component of ecolinguistic systems, as it is embedded in linguistic expressions that reflect ecological relationships, healing practices, and cultural values. However, this knowledge is increasingly threatened by language shift and insufficient documentation, particularly within low-resource indigenous communities. This study develops an AI-based ecolinguistic framework to systematically document and represent Nias ethnomedicinal knowledge by integrating ethnobotanical field data with culturally grounded artificial intelligence approaches. Qualitative data were obtained through semi-structured interviews with traditional Nias healers, resulting in the identification of fifteen commonly used medicinal plant species. To assess cultural salience and communal consensus, the study applied the Relative Frequency of Citation (RFC) index. The quantitative findings reveal an uneven distribution of cultural prominence among the documented species. Notably, *Gundre* and *Mbulu Nazalöu* emerged as the most frequently cited plants (FC = 14; RFC = 0.93 each), indicating their central role within the Nias ethnomedical knowledge system. The documented knowledge was subsequently structured using a Knowledge Graph model and enhanced through a Retrieval-Augmented Generation (RAG) architecture to enable contextualized, culturally sensitive knowledge representation. The proposed framework demonstrates how artificial intelligence can support the preservation, organization, and revitalization of endangered indigenous medicinal knowledge while maintaining its ecolinguistic integrity.

Keywords: *Ecolinguistics; Indigenous Knowledge; Nias Medicinal Plants; Knowledge Graph; Retrieval-Augmented Generation*

A. Introduction

The Nias people constitute a prominent indigenous ethnic group in Indonesia, inhabiting an archipelago situated off the western coast of Sumatra. The geographical isolation of Nias from the mainland has contributed to the preservation of a distinctive ecological environment marked by high biodiversity and rich genetic resources, including a wide range of grasses, herbs, shrubs, and endemic tree species. Over generations, these natural resources have been sustainably managed and transmitted through indigenous ecological knowledge systems, particularly in relation to the use of traditional medicinal plants for treating illnesses and maintaining community health. In geographically isolated regions such as Nias, local communities tend to develop a strong reliance on their surrounding environment, as limited access to modern medical facilities reinforces the importance of traditional healing practices. Consequently, medicinal plant knowledge becomes deeply embedded in everyday cultural practices, serving not only as a means of healthcare but also as a reflection of the community's ecological awareness and adaptive relationship with their natural environment. In rural Nias communities, flora constitutes an indispensable component of daily life, serving as primary sources of food, nutrition, and medicine (Suwardi et al., 2022).

Indigenous knowledge systems constitute a vital repository of cultural, linguistic, and ecological information that has been developed and transmitted across generations. Among these systems, traditional medicinal plant knowledge plays a crucial role in sustaining community health, environmental awareness, and cultural identity. Recent studies emphasize that such knowledge is inherently ecolinguistic, as it is encoded and preserved through language practices that reflect human–environment interactions (Fill & Penz, 2020; Maffi, 2021). However, indigenous medicinal knowledge worldwide is increasingly threatened by language shift, ecological degradation, and the marginalization of traditional epistemologies. UNESCO (2021) warns that the loss of indigenous languages entails the erosion of irreplaceable environmental knowledge embedded within them. In this context, scholars argue that documenting and structurally representing indigenous knowledge is not merely a cultural task but an urgent interdisciplinary challenge spanning linguistics, anthropology,

ecology, and digital technologies (Harrison, 2021; Rahman et al., 2021).

From an ecolinguistic perspective, language is not a neutral medium but a system that organizes, prioritizes, and transmits ecological meaning. As Stibbe (2021) asserts, *"languages shape the stories we live by, including how humans conceptualize and interact with the natural world"* (p. 3). This insight underscores the importance of preserving indigenous plant lexicons, which often encode detailed information about medicinal properties, preparation methods, and ecological habitats. In recent years, artificial intelligence (AI) has been increasingly explored as a tool for cultural knowledge representation and preservation. Advances in natural language processing and knowledge-based systems have enabled the development of AI frameworks capable of organizing complex, domain-specific information (Hogan et al., 2022; Ji et al., 2022). However, critical AI scholars caution that many AI systems remain culturally insensitive, often privileging high-resource languages and decontextualized data while marginalizing indigenous epistemologies (Bender et al., 2021; Birhane et al., 2021).

To address these limitations, researchers have called for human-centered and culturally grounded AI approaches that prioritize contextual integrity, transparency, and epistemic justice. Birhane et al. (2021) explicitly argue that *"AI systems are not value-neutral artifacts but socio-technical systems embedded within specific cultural and political contexts"* (p. 611). This perspective highlights the necessity of designing AI frameworks that respect indigenous knowledge structures rather than reducing them to abstract data points. One promising approach for culturally grounded AI is the use of Knowledge Graphs (KGs) combined with Retrieval-Augmented Generation (RAG) architectures. Knowledge Graphs enable the structured representation of entities and their relationships, making them particularly suitable for modeling interconnected linguistic, medicinal, and ecological knowledge (Hogan et al., 2022). Meanwhile, RAG architectures integrate knowledge retrieval with generative language models, allowing AI systems to generate outputs that remain anchored in verified domain-specific information (Lewis et al., 2020; Gao et al., 2023). Rather than optimizing computational performance, such architectures can be employed to support explainable and context-sensitive knowledge mediation, especially in low-resource language settings.

Indonesia, as one of the world's most linguistically diverse countries, provides a critical context for examining the intersection of language, ecology, and technology. The Nias language, spoken by an indigenous community in North Sumatra, embodies rich ethnobotanical knowledge related to traditional medicinal practices. Yet, like many indigenous languages, Nias faces challenges related to language maintenance and intergenerational knowledge transmission (Magueresse et al., 2021). Despite this, systematic AI-based representations of Nias medicinal plant knowledge remain scarce. Against this backdrop, the present study proposes an AI-powered ecolinguistic framework for representing Nias medicinal plant knowledge. By integrating ethnobotanical documentation with Knowledge Graph-based representation and Retrieval-Augmented Generation, this study aims to demonstrate how indigenous medicinal knowledge can be structurally represented in a culturally grounded AI system. Rather than evaluating algorithmic performance or educational outcomes, the study focuses on knowledge representation, cultural salience, and contextual integrity, contributing to ongoing discussions on ethical AI, ecolinguistics, and indigenous knowledge preservation.

Ecolinguistics examines the interrelationships between language, ecology, and culture, emphasizing how linguistic practices encode environmental knowledge and shape human interactions with the natural world. Within this perspective, indigenous languages function not merely as communication tools but as repositories of ecological wisdom, transmitting culturally grounded knowledge across generations (Fill & Penz, 2020; Stibbe, 2021). Recent ecolinguistic scholarship underscores that the loss of indigenous languages often entails the erosion of ecological knowledge embedded within them. As Maffi (2021) argues, linguistic diversity and biological diversity are mutually reinforcing, such that *"the erosion of one accelerates the erosion of the other"*. This insight provides a theoretical foundation for examining medicinal plant lexicons as key components of ecolinguistic systems. In indigenous contexts, medicinal plant knowledge is linguistically encoded through naming practices, metaphors, and culturally specific classifications that reflect ecological relationships. These linguistic structures enable communities to organize environmental knowledge in ways that are locally meaningful and practically applicable. From an ecolinguistic standpoint, documenting and representing indigenous medicinal

plant knowledge is therefore essential for both cultural sustainability and ecological understanding.

Cultural salience refers to the degree to which particular knowledge items are collectively recognized and valued within a community (Phillips & Gentry, 2021; Yaseen et al., 2022). Ethnobotanical theory suggests that plants with higher cultural salience are more frequently mentioned in discourse, more consistently taught across generations, and more deeply embedded in ritual and everyday practices. Quantitative indices such as the Relative Frequency of Citation (RFC) are commonly employed to operationalize this concept by capturing communal consensus rather than individual preference (Ullah et al., 2022). From a theoretical perspective, cultural salience is not merely a statistical construct but a reflection of sociocultural prioritization. Linguistically salient plant names often occupy central positions in narrative practices, healing discourse, and environmental knowledge systems. This framework allows the present study to interpret RFC values as indicators of sociocultural embeddedness rather than as measures of objective medicinal efficacy.

Knowledge representation theory provides a foundation for understanding how complex, relational knowledge can be structured in computational systems. Knowledge Graphs (KGs) represent knowledge as networks of entities and relationships, enabling the modeling of interconnected concepts across multiple domains (Hogan et al., 2022; Ji et al., 2022). From a theoretical standpoint, Knowledge Graphs align with ecolinguistic principles by emphasizing relational meaning rather than isolated data points. In indigenous knowledge systems, meaning is often derived from relationships between language, ecology, and practice. KGs therefore offer a theoretically compatible structure for representing medicinal plant knowledge in a way that preserves contextual integrity. Importantly, recent scholarship cautions against treating Knowledge Graphs as purely technical artifacts. Instead, they should be understood as epistemic models that reflect particular worldviews and classification systems (Hogan et al., 2022). This theoretical position supports the use of Knowledge Graphs in the present study as representational tools designed to respect indigenous epistemologies rather than to impose external taxonomies.

Human-centered AI theory emphasizes that AI systems are socio-technical constructs embedded within cultural, ethical, and political contexts.

Bender et al. (2021) argue that language technologies are shaped by the data and assumptions underlying their design, cautioning that *"scale alone does not guarantee meaning or understanding"*. Similarly, Birhane et al. (2021) assert that AI systems must be critically examined for whose knowledge they represent and whose voices they marginalize. Within this theoretical framework, culturally grounded AI prioritizes transparency, explainability, and contextual fidelity over raw computational performance. This approach is particularly relevant for low-resource indigenous languages, where data scarcity amplifies the risks of misrepresentation and cultural erasure (Magueresse et al., 2021). The present study adopts this theoretical stance by conceptualizing AI not as an autonomous knowledge producer but as a mediating tool that supports the representation and retrieval of culturally validated knowledge. This aligns with recent calls for ethical AI practices that foreground epistemic justice and community-centered design.

Retrieval-Augmented Generation (RAG) offers a theoretical bridge between symbolic knowledge representation and neural language generation. Unlike standalone language models, RAG architectures retrieve relevant external knowledge before generating responses, thereby grounding outputs in verified information sources (Lewis et al., 2020; Gao et al., 2023). Theoretically, RAG aligns with ecolinguistic and human-centered AI principles by reducing the risk of decontextualized or hallucinatory outputs. Rather than generating meaning in isolation, the model engages in knowledge mediation, drawing from structured representations that encode cultural and ecological relationships. In the context of indigenous knowledge representation, RAG serves as a mechanism that supports contextual grounding, enables culturally coherent reasoning, and facilitates transparency in AI-generated outputs. This theoretical positioning underscores the role of RAG in the present study as a supporting architecture rather than an object of algorithmic evaluation.

Taken together, ecolinguistics, cultural salience theory, knowledge representation, and human-centered AI form an integrative theoretical framework for this study. Ecolinguistics provides the linguistic–ecological lens, cultural salience explains patterns of knowledge prioritization, Knowledge Graph theory offers a representational structure, and RAG enables knowledge-grounded mediation. This integrated framework supports the study's objective of representing Nias medicinal plant

knowledge in a culturally grounded AI system, contributing to broader discussions on indigenous knowledge preservation, ethical AI, and the role of language in ecological sustainability.

B. Research Method

1). Research Design

This study employed a qualitative–quantitative descriptive research design grounded in ecolinguistics and ethnobotany. The methodological approach was structured to document, represent, and analyze indigenous medicinal plant knowledge in the Nias language through a culturally grounded AI framework. Rather than evaluating pedagogical outcomes or algorithmic performance, the study focused on knowledge representation and cultural salience analysis. The research design integrated three complementary components: (1) qualitative ethnobotanical data collection, (2) quantitative cultural salience analysis using the Relative Frequency of Citation (RFC) index, and (3) AI system design employing a Retrieval-Augmented Generation (RAG) architecture supported by a Knowledge Graph. This integrative design aligns with recent methodological recommendations in ecolinguistic and indigenous knowledge research, which emphasize representational fidelity over experimental validation (Stibbe, 2021; Maffi, 2021).

2). Data Sources and Participants

Data were obtained from ten traditional Nias healers, who served as key informants due to their recognized expertise in indigenous medicinal practices. Purposeful sampling was applied to ensure epistemic authority and cultural validity. Through semi-structured interviews, the informants identified fifteen medicinal plant species, which formed the core ecolinguistic dataset of the study. The medicinal plants were documented using indigenous lexicons, descriptions of traditional uses, and ecological contexts. The research population and data sources are summarized in Table 1.

Table 1

Research Participants and Data Sources

Category	Quantity	Role
Traditional Nias Healers	10 informants	Ethnobotanical corpus development
Medicinal Plants	15 species	Core ecolinguistic dataset

3). Qualitative Data Collections

Qualitative data were collected through semi-structured ethnobotanical interviews, a method widely adopted in

indigenous knowledge documentation to preserve cultural authenticity and contextual depth (Ullah et al., 2022). Interview questions focused on local plant names, medicinal functions, preparation methods, and ecological habitats. All interviews were audio-recorded with participants' consent and transcribed verbatim. Given the low-resource nature of the Nias language, the transcriptions were subsequently normalized to address dialectal variation while maintaining semantic integrity. This process ensured consistency in lexical representation without erasing culturally meaningful linguistic variation.

4). Quantitative Cultural Salience Analysis

To assess the relative cultural importance of each documented medicinal plant, the study employed the Relative Frequency of Citation (RFC) index. RFC is a widely recognized ethnobotanical metric for quantifying communal consensus regarding plant use and cultural salience (Yaseen et al., 2022). The RFC value was calculated using the following formula:

$$\text{RFC (\%)} = \left(\frac{N_p}{N}\right) \times 100\%$$

where (N_p) represents the number of informants who mentioned a specific plant species, and (N) denotes the total number of informants (N = 10). Higher RFC values indicate greater cultural salience and stronger communal recognition. This quantitative measure was used solely to describe patterns of knowledge distribution within the community and was not intended to assess medicinal efficacy or biomedical effectiveness.

5). AI System Design RAG Framework

The AI-based ecolinguistic system was designed using a Retrieval-Augmented Generation (RAG) framework, which integrates structured knowledge retrieval with neural language generation. In this study, the RAG architecture functioned as a knowledge mediation tool, enabling the retrieval and presentation of culturally grounded medicinal plant information. A Knowledge Graph (KG) was constructed to represent relational links between indigenous plant names, medicinal functions, and ecological habitats. This relational structure supports contextual coherence and facilitates transparent knowledge retrieval. Table 2 outlines the core components of the RAG-based system architecture.

Table 2

System Architecture of the Retrieval-Augmented Generation (RAG)

Framework for Nias Ecolinguistic Learning

Component	Function	Data Input
Data Acquisition	Corpus curation	10 healers
Pre-processing	Dialect normalization	15 plant transcriptions
Knowledge Graph	Ecolinguistic mapping	15 plants
Inference Layer	Response generation	Transformer fine-tuned on Nias corpus

The AI system was not evaluated using performance metrics such as accuracy or F1-score. Instead, its role was limited to supporting the structured representation and retrieval of indigenous knowledge in alignment with human-centered and culturally grounded AI principles (Bender et al., 2021; Birhane et al., 2021).

C. Results and Discussion

1). Qualitative Ethnobotanical Findings

Based on fieldwork conducted through direct observation and semi-structured interviews with traditional healers on Nias Island, a total of 15 medicinal plant species were identified as being actively used in indigenous healing practices. These plants represent a wide range of botanical families and reflect the depth of ethnobotanical knowledge embedded within Nias cultural traditions The documented medicinal plants are predominantly used to treat common health conditions such as fever, wounds, digestive disorders, respiratory problems, inflammation, and infectious diseases. Several species were reported to have multifunctional therapeutic roles, indicating a holistic medical system developed through long-term interaction with the local environment.

For example, *Imperata cylindrica* (Alang-alang/Go'o) was widely used for treating fever, wounds, stomachache, hypertension, and toothache, while its roots were also employed as an alternative remedy for diabetes. Similarly, *Curcuma longa* (Kunyit/Gundre) and *Piper betle* (Sirih/Tawuo) were frequently cited for digestive health and respiratory or oral conditions. Other species, such as *Orthosiphon aristatus* (Kumis Kucing/Sogambi Mao), were specifically associated with kidney-related illnesses, whereas *Etlingera elatior* (Kecombrang/Silimo) was recognized for its broad therapeutic functions, including anemia treatment, wound healing, and circulation improvement

In addition to general medicinal applications, certain plants were linked to culturally specific practices. For instance, *Vitex pinnata* (Daun Laban/Mbulu Mali-Mali Mao) was traditionally consumed by pregnant women to maintain uterine health, while *Melastoma malabathricum* (Senggani/Ndruru-ndruru) was used as

an antidote for poisoning. These findings illustrate that ethnobotanical knowledge on Nias Island is not only extensive but also functionally structured and culturally embedded.

Table 3
Traditional Medicinal Plants Identified in the Field Study in Nias Island

Plant Name	Scientific Name	Traditional Medicinal Uses
Alang-alang (Go'o)	*Imperata cylindrica*	Used to treat fever, wounds, stomachache, hypertension, and toothache. The roots are also used as an alternative treatment for diabetes.
Bunglu (Mbulu Mboli)	*Zingiber zerumbet*	Used to relieve fever and cough, particularly in children.
Cirik Babi (Söfö-Söfö)	*Ageratum conyzoides*	Used for treating cuts and abrasions, diarrhea, and as a natural insecticide.
Cocor Bebek (Mbulu Zini-zini)	*Kalanchoe pinnata*	Used for healing wounds and bruises, relieving stomach pain, treating hemorrhoids, and reducing headaches.
Daun Jarak (Lafandru)	*Ricinus communis*	Used to relieve respiratory problems such as cough and shortness of breath.
Daun Ungu (Mbulu Nazalöu)	*Graptophyllum pictum*	Used to treat hemorrhoids, constipation, boils, inflammation, and skin diseases, as well as to maintain kidney health.
Dringo (Sarango)	*Acorus calamus*	Used to treat fever and swelling.
Gandaruso (Lio-Lio)	*Justicia gendarussa*	Used to improve blood circulation, relieve nausea, and treat rheumatism.
Daun Laban (Mbulu Mali Mali Mao)	*Vitex pinnata*	Used for treating burns; also traditionally consumed by pregnant women to maintain uterine health.
Kecombrang (Silimo)	*Etlingera elatior*	Used to treat measles, increase energy, eliminate body odor, treat anemia, improve blood circulation, heal fresh wounds, and strengthen bones.
Kumis Kucing (Sogambi Mao)	*Orthosiphon aristatus*	Used to treat kidney problems, including kidney stones and urinary tract infections.
Kunyit, (Gundre)	*Curcuma longa*	Used to treat digestive disorders (gastritis), cough, and appendicitis.
Sirih (Tawuo)	*Piper betle*	Used to maintain oral health and digestive system health.
Bandotan (Mbangu-Mbangu)	*Ageratum conyzoides*	Used to treat fever, malaria, sore throat, pneumonia, middle ear infection, bleeding wounds, nosebleeds, dysentery, and sprains.
Senggani (Ndruru-ndruru)	*Melastoma malabathricum*	Traditionally used as an antidote for poisoning.

2). Ethnobotanical Utilizations

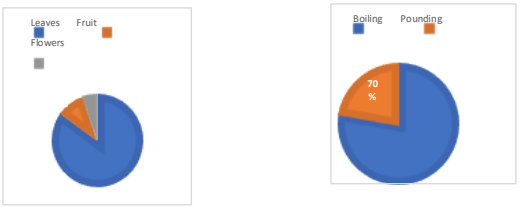
Beyond plant identification, the study documented traditional processing

methods, application routes, and plant parts used in medicinal preparation. Analysis of ethnobotanical utilization patterns shows that boiling is the most prevalent processing method (70%), followed by pounding (20%), while soaking, burning, and grating collectively account for the remaining 10% . In terms of application, the majority of remedies are administered through oral consumption (80%), whereas topical applications including smearing, bathing, and ointment use constitute 20%. Regarding plant parts, leaves are the most frequently utilized component (85%), followed by roots, stems, and fruits (10%), and flowers (5%).

These utilization patterns indicate that indigenous medicinal practices in Nias prioritize accessible plant parts and simple preparation techniques, reflecting pragmatic ecological knowledge. Figure 2 visualizes these ethnobotanical utilization patterns. The figure illustrates the distribution of processing methods, application routes, and plant parts used in traditional Nias medicine, highlighting dominant practices such as boiling, oral administration, and leaf utilization.

Figure 2

Ethnobotanical Utilizations Patterns in Nias Traditional Medicine



3). Quantitative Cultural Salience Analysis

Quantitative cultural salience was measured using the *Relative Frequency of Citation* (RFC) index to assess the relative cultural importance of each medicinal plant within the local knowledge system of the Nias community. The RFC value was calculated by dividing the number of informants who mentioned a particular plant (FC) by the total number of informants (N = 15). Higher RFC values indicate stronger cultural embeddedness and more frequent practical usage in traditional healing practices.

The results show that *Curcuma longa* (Kunyit/Gundre) and *Graptophyllum pictum* (Daun Ungu/Mbulu Nazalöu) exhibited the highest cultural salience, each with an RFC value of 0.93 (FC = 14). This finding indicates that these two species are the most widely recognized and consistently utilized medicinal plants among the Nias informants. The second-

highest salience group consisted of *Ageratum conyzoides* (Cirik Babi/Söfö-Söfö), *Piper betle* (Sirih/Tawuo), and *Kalanchoe pinnata* (Cocor Bebek/Mbulu Zini-zini), all of which obtained an RFC value of 0.86 (FC = 13). These species are strongly embedded in daily ethnomedical practices, particularly for treating wounds, digestive disorders, and maintaining oral health. Plants with moderate salience included *Ricinus communis* (Daun Jarak/Lafandru) and *Acorus calamus* (Dringo/Sarango), each with an RFC value of 0.80, followed by *Zingiber zerumbet* (Bunglu/Mbulu Mboli) and *Melastoma malabathricum* (Senggani/Ndruru-ndruru), both with an RFC value of 0.73. Although not mentioned by all informants, these species still play a substantial role in local therapeutic practices. In contrast, *Etlingera elatior* (Kecombrang/Silimo), *Imperata cylindrica* (Alang-alang/Go'o), and *Ageratum conyzoides* (Bandotan/Mbangu-Mbangu) showed an RFC value of 0.67, indicating relatively high recognition but lower universality across informants. The lowest salience values were found for *Vitex pinnata* (Daun Laban/Mbulu Mali-Mali Mao), *Orthosiphon aristatus* (Kumis Kucing/Sogambi Mao), and *Justicia gendarussa* (Gandaruso/Lio-Lio), each with an RFC value of 0.60.

Overall, the distribution of RFC values suggests that ethnobotanical knowledge in the Nias community is not evenly distributed across plant species, but rather concentrated on a limited number of core medicinal plants that are considered most effective, accessible, and culturally trusted. This pattern reflects a selective cultural transmission process, whereby plant species with strong empirical efficacy are more deeply embedded in collective memory and everyday medical practice. Table 6 presents the complete results of the frequency of citation analysis and RFC values for all medicinal plants identified in this study.

Table 4

Quantitative Cultural Salience of Traditional Medicinal Plants in Nias Island

Local Name	FC	RFC (%)
Gundre	14	0.93
Mbulu Nazalöu	14	0.93
Söfö-Söfö	13	0.86
Tawuo	13	0.86
Mbulu Zini-zini	13	0.86
Lafandru	12	0.80
Sarango	12	0.80
Mbulu Mboli	11	0.73
Ndruru-ndruru	11	0.73
Silimo	10	0.67
Mbangu-Mbangu	10	0.67
Go'o	10	0.67

Mbulu Mali-Mali Mao	9	0.60
Sogambi Mao	9	0.60
Lio-Lio	9	0.60

Overall, the results demonstrate that ethnobotanical knowledge in the Nias community is unevenly distributed, with a limited number of highly salient plants forming the core of traditional medicinal practices. This pattern reflects selective cultural transmission, where plants perceived as effective, accessible, and culturally trusted are more deeply embedded in collective memory and everyday use. These findings provide an empirical foundation for the AI-based ecolinguistic framework proposed in this study, ensuring that knowledge representation is grounded in documented cultural salience and authentic field data, rather than abstract or externally imposed classifications. These empirically grounded findings provide a structured and culturally validated foundation for the proposed AI-powered ecolinguistic framework, ensuring that knowledge representation reflects indigenous priorities rather than externally imposed classifications.

D. Discussion

This study set out to develop an AI-based ecolinguistic framework for representing indigenous medicinal plant knowledge in the Nias community. The discussion interprets the findings in relation to ecolinguistic theory, cultural salience, and human-centered AI, demonstrating how empirically grounded ethnobotanical data can inform culturally responsible knowledge representation.The findings confirm that medicinal plant knowledge in the Nias community is deeply ecolinguistic in nature, as linguistic labels, medicinal functions, and ecological contexts are inseparably intertwined. Indigenous plant names do not merely denote botanical entities but encode culturally situated knowledge about healing practices, environmental conditions, and social use. This supports ecolinguistic scholarship asserting that language functions as a key medium through which ecological knowledge is structured and transmitted (Fill & Penz, 2020; Stibbe, 2021). The multifunctional use of several plant species, such as *Imperata cylindrica* and *Curcuma longa*, illustrates how indigenous knowledge systems prioritize ecological versatility and practical relevance. These patterns align with Maffi's (2021) argument that linguistic and biological diversity are mutually reinforcing, as culturally salient species are more likely to be linguistically stabilized and intergenerationally preserved.

The Relative Frequency of Citation (RFC) analysis reveals an uneven distribution of medicinal plant knowledge within the community. A limited number of plant species exhibit high cultural salience, while others remain less frequently cited and more specialized. This long-tail distribution reflects selective cultural transmission, where plants perceived as effective, accessible, and socially trusted become central to communal knowledge systems. From a theoretical perspective, RFC values function as indicators of sociocultural prioritization rather than objective medicinal efficacy. This interpretation aligns with recent ethnobotanical studies emphasizing that cultural salience captures collective recognition and frequency of use (Phillips & Gentry, 2021; Yaseen et al., 2022). In the context of this study, plants with higher RFC values represent core ecolinguistic knowledge, whereas low-salience plants reflect more localized or expert-specific practices. Importantly, these patterns provide a principled basis for structuring indigenous knowledge representation, ensuring that culturally significant elements are not treated as equivalent to marginal or rarely transmitted knowledge.

The Knowledge Graph constructed in this study operationalizes ecolinguistic theory by modeling medicinal plant knowledge as a network of relational meanings rather than isolated data points. By linking indigenous lexicons to medicinal functions and ecological habitats, the KG preserves contextual integrity and reflects the relational epistemology characteristic of indigenous knowledge systems. This finding supports recent theoretical work suggesting that Knowledge Graphs are particularly suitable for representing culturally embedded and domain-specific knowledge due to their emphasis on relationships and semantic transparency (Hogan et al., 2022; Ji et al., 2022). In this study, plants with higher RFC values naturally occupy more central positions within the graph, reflecting their stronger cultural embeddedness. Thus, the KG does not merely store information but mirrors the internal structure of the community's knowledge system, reinforcing the alignment between computational representation and indigenous epistemology.

The implementation of Retrieval-Augmented Generation (RAG) further strengthens the framework's capacity for culturally grounded knowledge mediation. Rather than generating responses in isolation, the RAG architecture retrieves relevant nodes from the Knowledge Graph before producing

outputs, thereby ensuring that generated information remains anchored in documented indigenous knowledge. Consistent with recent studies, RAG functions here as a supportive mechanism that facilitates contextual grounding and transparency, particularly in low-resource language settings (Lewis et al., 2020; Gao et al., 2023). Notably, the retrieval process reflects patterns of cultural salience: plants with higher RFC values are more frequently retrieved, while low-frequency plants remain accessible but less prominent. This behavior illustrates how cultural salience can indirectly shape AI-mediated knowledge access without imposing algorithmic optimization or performance-based hierarchies.

From a human-centered AI perspective, the findings underscore the importance of positioning AI systems as mediators rather than producers of knowledge. As argued by Bender et al. (2021) and Birhane et al. (2021), AI technologies are embedded within social and cultural contexts and must be designed to respect epistemic boundaries. By grounding the AI framework in ethnobotanical field data and cultural salience metrics, this study responds to calls for culturally responsible AI that foregrounds indigenous voices and knowledge structures. The framework does not claim to validate medicinal efficacy or replace traditional expertise; instead, it supports the structured representation and retrieval of knowledge already validated within the community.

This study contributes to ecolinguistics and AI research by demonstrating how indigenous medicinal plant knowledge can be represented through an AI-powered framework grounded in cultural salience and relational knowledge structures. It extends existing discussions on ethical AI by providing a concrete methodological example rooted in field-based data. Nevertheless, the study is limited to a single indigenous community and a relatively small number of informants. While this approach prioritizes epistemic depth and cultural validity, future research could explore comparative studies across multiple indigenous languages or incorporate participatory co-design approaches to further enhance community involvement.

E. References

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623.

https://doi.org/10.1145/3442188.3445922

Birhane, A., Prabhu, V. U., & Kahembwe, E. (2021). Multimodal datasets: Misogyny, porn, and malignant stereotypes. *arXiv preprint arXiv:2110.01963*.

Fill, A., & Penz, H. (2020). *The Routledge handbook of ecolinguistics* (2nd ed.). Routledge. https://doi.org/10.4324/9780429468310

Gao, Y., Liu, Z., Li, Z., & Sun, M. (2023). Retrieval-augmented generation for knowledge-intensive NLP tasks. *ACM Computing Surveys, 55*(6), 1–38. https://doi.org/10.1145/3546258

Gao, Y., Xiong, Y., Gao, L., Liu, K., Pan, J., Bi, Y., Dai, Y., Sun, J., & Wang, H. (2023). Retrieval-augmented generation for large language models: A survey. *arXiv preprint arXiv:2312.10997*.

Harrison, K. D. (2021). *The last speakers: The quest to save the world's most endangered languages*. National Geographic.

Hogan, A., Blomqvist, E., Cochez, M., d'Amato, C., de Melo, G., Gutierrez, C., Kirrane, S., Neumaier, S., Polleres, A., & Navigli, R. (2022). Knowledge graphs. *ACM Computing Surveys, 54*(4), 1–37. https://doi.org/10.1145/3447772

Ji, S., Pan, S., Cambria, E., Marttinen, P., & Yu, P. S. (2022). A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE Transactions on Neural Networks and Learning Systems, 33*(2), 494–514. https://doi.org/10.1109/TNNLS.2021.3070843

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W. T., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems, 33*, 9459–9474.

Maffi, L. (2021). Biocultural diversity and sustainability. *Annual Review of Anthropology, 50*, 25–40. https://doi.org/10.1146/annurev-anthro-101819-110523

Magueresse, A., Caron, G., & Heetderks, E. (2021). Low-resource languages: A review of past work and future challenges. *Language Resources and Evaluation, 55*(1), 1–31. https://doi.org/10.1007/s10579-020-09505-4

Phillips, O., & Gentry, A. H. (2021). The useful plants of Tambopata, Peru:

Statistical hypotheses tests with a new quantitative technique. *Economic Botany, 75*(1), 1–15. https://doi.org/10.1007/s12231-020-09501-9

Rahman, A. H. M. M., Alam, M. S., Khan, S. K., & Naderuzzaman, A. T. M. (2021). Documentation of traditional medicinal plant knowledge: Methods and ethical considerations. *Journal of Ethnobiology and Ethnomedicine, 17*(1), 1–14.

Stibbe, A. (2021). *Ecolinguistics: Language, ecology and the stories we live by* (2nd ed.). Routledge.

Ullah, S., Ahmad, M., Zafar, M., Sultana, S., & Bahadur, S. (2022). Quantitative ethnobotany and traditional knowledge of medicinal plants. *Journal of Ethnobiology and Ethnomedicine, 18*(1), 1–15. https://doi.org/10.1186/s13002-022-00512-8

UNESCO. (2021). *International decade of indigenous languages 2022–2032*. UNESCO Publishing.

Waruwu, O. (2025). Critical Discourse Analysis Of Hate Speech On Instagram: A Politician's Educational Documents In Indonesia, 2025. *Haga : Jurnal Pengabdian Kepada Masyarakat*, *4*(2), 51-63. https://doi.org/10.57094/haga.v4i2.4100

Waruwu, O. (2025). An Analysis Of Proclitics And Enclitics In West Nias Language: A Morphosyntactic Study. *Research on English Language Education*, *7*(2), 53-64. https://doi.org/10.57094/relation.v7i2.4004

Waruwu, O. (2024). Increasing students' reading comprehension ability on descriptive text by using language experience approach at the tenth grade of SMA Negeri 1 Lahomi. *NDRUMI: Jurnal Ilmu Pendidikan dan Humaniora, 7*(2), 16–26. https://jurnal.uniraya.ac.id/index.php/NDRUMI

Yaseen, G., Ahmad, M., Sultana, S., Alharrasi, A. S., Hussain, J., & Zafar, M. (2022). Ethnobotanical indices for quantitative analysis of medicinal plant knowledge. *Plants, 11*(3), 345. https://doi.org/10.3390/plants11030345