Gaana Sampradaya: Preserving Folk Wisdom of Kannada Culture in The Digital Age

Daya Naik[1], Nivin K S[2], Ameer Sawad[3], Arafath Ali[4], Badarinadh A P[5], Shivshankar Pawar[6]
[1,2,3,4,5,6]*Department of Artificial Intelligence and Machine Learning, Srinivas Institute of Technology, Mangaluru*
doi.org/10.64643/IJIRTV12I12-206673-459

***Abstract*—Preserving folk literature is essential for safeguarding cultural identity, especially in the context of low-resource languages such as Kannada. With the increasing use of artificial intelligence for language translation, it becomes important to evaluate how accurately AI systems represent culturally rich and metaphorical content. Kannada folk songs, in particular, contain deep symbolic meanings, traditional expressions, and emotional context that are difficult for general-purpose translation models to capture accurately. This project focuses on evaluating the semantic accuracy of AI-generated English translations of Kannada folk songs.**

Many studies in neural machine translation have focused on improving translation for low-resource languages. However, most of this work is centered on generating translations rather than evaluating how well these translations preserve cultural meaning and semantic accuracy. In particular, there is a lack of reliable methods to compare AI-generated translations with expert interpretations, especially in the case of folk literature. This gap highlights the importance of developing an evaluation-based approach that considers multiple aspects of translation quality. In this work, a structured dataset was created containing original Kannada folk lyrics, AI-generated English translations, and expert-provided interpretations. A similarity-based evaluation model was then designed using TF-IDF cosine similarity, chrF proxy score, and Jaccard word set similarity to examine semantic, structural, and vocabulary-level alignment. The system also includes a manual comparison feature that allows individual translation pairs to be evaluated. The findings indicate that while AI translations are able to convey the general meaning, they often fail to capture cultural nuances and metaphorical expressions. Overall, the proposed framework offers a practical and scalable method for assessing translation quality and contributes to ongoing efforts in cultural preservation and low-resource language research.

***Index Terms*—Kannada Folk Songs, Machine Translation, Natural Language Processing, Semantic Evaluation, TF-IDF, Cultural Preservation.**

I. INTRODUCTION

In today's interconnected world, language barriers still make it difficult for people to access regional culture and traditional knowledge. Kannada, one of the classical languages of India, has a rich collection of folk songs that reflect emotions, traditions, and social values. However, because the language is region-specific, much of this cultural content remains inaccessible to those who do not understand Kannada. Most available translation tools are designed for general or conversational use and are not well suited for handling poetic or culturally rich text. As a result, they tend to produce direct, word-by-word translations that fail to convey the original emotion and deeper meaning of folk songs. This makes it challenging for a wider audience to truly appreciate their cultural significance. In this project, we focus on developing a semantic-aware approach to evaluate AI-based translations of Kannada folk songs using Natural Language Processing (NLP) techniques. Instead of simply generating translations, the system emphasizes how well the meaning, context, and cultural elements are preserved. One of the major challenges in this domain is the limited availability of Kannada datasets, along with the complex and figurative nature of folk song language. To address this, we construct a parallel dataset of Kannada lyrics and their corresponding English interpretations and apply transformer-based approaches that can better understand contextual meaning.

The overall goal of this work is to reduce the cultural gap by using AI to make Kannada folk literature more accessible to a global audience, while also preserving its original essence. This report presents the motivation behind the project, the methods used, and the insights gained from the study.

II. LITERATURE REVIEW

[1] Vaswani et al. introduced the Transformer architecture for sequence-to-sequence tasks such as machine translation, replacing recurrent and convolutional structures with self-attention mechanisms. The model employs multi-head attention and positional encoding to capture long-range dependencies and contextual semantics effectively. While the Transformer significantly improves translation quality, it does not explicitly address cultural or poetic nuances present in traditional folk literature.

[2] The CompMusic Project focused on AI-driven digitization and preservation of traditional music traditions, including Indian classical music. It created structured datasets containing audio, lyrics, and metadata for cultural analysis. Although the project provides a strong framework for cultural preservation, Kannada folk songs and semantic translation evaluation were not directly addressed.

[3] Aharoni et al. proposed a massively multilingual neural machine translation model capable of translating over 100 languages using a unified architecture. The model supports parameter sharing and zero-shot translation, improving performance for low-resource languages. However, the system focuses on general-purpose translation and does not consider poetic or culturally sensitive text such as folk songs.

[4] Rapp demonstrated that incorporating Semantic Role Labeling (SRL) into neural machine translation improves translation quality by preserving "who did what to whom" relationships. The study showed improvements in BLEU scores across multiple language pairs. While effective, the approach was evaluated on structured sentences rather than poetic or lyrical content.

[5] Voita et al. introduced a context-aware neural machine translation model that resolves anaphora by incorporating previous sentence context into the Transformer decoder. This improves coherence and referential accuracy across sentences. The work is relevant to lyrical text but does not explicitly address low-resource Indian languages or folk literature.

[6] Artetxe et al. proposed an unsupervised neural machine translation framework that eliminates the need for parallel corpora by using monolingual data and back-translation. This approach is suitable for low-resource languages; however, its performance depends heavily on the quality and size of monolingual corpora.

[7] Ranathunga et al. presented a comprehensive survey of neural machine translation techniques for low-resource languages. The study highlighted strategies such as multilingual training, transfer learning, and data augmentation. While the survey provides valuable guidance, it does not focus on domain-specific translation such as folk or cultural texts.

[8] Wang et al. reviewed methods for low-resource neural machine translation, categorizing solutions into monolingual data exploitation, auxiliary language usage, and multimodal learning. Although these methods address data scarcity, the study does not evaluate semantic fidelity in culturally rich translations.

[9] Chandra and Kulkarni applied BERT-based models to analyze semantic and sentiment consistency across multiple translations of the Bhagavad Gita. Their results demonstrated the effectiveness of transformer-based models in preserving emotional and philosophical meaning. However, the study focused on evaluation rather than automated translation generation.

[10] Koneru et al. explored unsupervised machine translation for Dravidian languages, including Kannada, using back-translation and shared encoder-decoder architectures. The study showed improved performance by leveraging linguistic similarities across Dravidian languages. Nevertheless, the approach was not tested on poetic or folk song domains.

[11] Gala et al. introduced IndicTrans2, a multilingual machine translation system covering all 22 scheduled Indian languages, supported by the Bharat Parallel Corpus. The model provides strong baseline performance for Indian language translation. However, its effectiveness in preserving cultural

symbolism and poetic expressions in folk songs remains unexplored.

[12] Shanker et al. developed an emotion-annotated dataset for Telugu song lyrics and evaluated transformer-based models for emotion classification. The study highlights the importance of emotional preservation in lyrical content but does not integrate translation or cross-lingual semantic evaluation.

[13] Das et al. studied attention-based neural machine translation models for Indian languages such as Hindi and Bengali, demonstrating improved performance in morphologically rich and low-resource settings. While applicable to Kannada, the work did not address lyrical or folk-text translation.

[14] Choudhary et al. proposed an attention-based neural machine translation model combined with pre-trained word embeddings for low-resourced Indian languages. The approach improved handling of out-of-vocabulary words and contextual meaning. However, it was evaluated on standard text corpora rather than culturally rich folk literature.

[15] Deshpande et al. reviewed various machine translation approaches for Sanskrit, analyzing rule-based, statistical, and neural methods. The study highlighted challenges posed by rich morphology and free word order, which are also characteristics of Kannada. While informative, the work did not propose a semantic-aware evaluation framework for translated content. Address data scarcity but depend heavily on the availability of high-quality monolingual data. Overall, existing research lacks a comprehensive evaluation framework that measures how effectively AI preserves semantic and cultural meaning in translations. This gap motivates the development of the GaanaSampradaya system.

III. REQUIREMENTS

The requirements for the proposed semantic-aware AI-based translation system include both hardware and software components. On the hardware side, a standard computer or laptop with adequate processing capability is sufficient for running the current model. While a GPU can be useful for faster training and handling advanced neural models, it is not strictly necessary at this stage. Sufficient RAM is required to efficiently process the dataset, perform TF-IDF vectorization, and compute similarity scores without performance issues. In addition, basic storage space is needed to manage datasets, output files, and project-related documents. On the software side, the project is implemented using Python due to its strong support for machine learning and natural language processing tasks. Frameworks such as TensorFlow or PyTorch can be used for developing neural models, while NLP libraries assist in tasks like text preprocessing, tokenization, and semantic analysis. The system also relies on a well-structured dataset containing Kannada folk songs along with their corresponding English translations, which is essential for both evaluation and analysis. Furthermore, standard evaluation metrics such as BLEU score may be used to assess translation quality, along with other similarity-based methods. Overall, the system is designed to run efficiently on commonly available computing resources, with scope for scaling using more advanced hardware in future developments.

I. Hardware Requirements

The current evaluation model can be executed efficiently on a standard personal computer or laptop equipped with a modern processor. Adequate RAM is important to ensure smooth handling of tasks such as loading datasets, performing TF-IDF vectorization, and computing similarity measures without affecting performance. In terms of storage, basic capacity is sufficient to store the dataset, generated outputs, and related project files. At this stage, the system does not require specialized hardware like GPUs, as the computations involved are relatively lightweight. However, GPU support may be considered in future developments if more advanced deep learning models or large-scale language processing techniques are incorporated.

II. Software Requirements

The proposed system was implemented using python on windows and Linux operating systems, ensuring cross-platform compatibility. deep learning models were developed using TensorFlow and pytorch, while NLTK and the hugging face transformers library were used for text preprocessing and transformer-based neural machine translation. development and experimentation were carried out using vs code and

jupyter notebook, with google colab optionally utilized for GPU-accelerated training. version control was managed using git. the dataset comprises a parallel kannada–English folk song corpus, containing original kannada lyrics along with expert-verified English translations, curated to preserve semantic meaning, emotional depth, and cultural context for effective evaluation of the proposed translation and comparison model.

IV. METHODOLOGY

This chapter explains the step-by-step process used in the GaanaSampradaya project. The methodology includes data collection, dataset preparation, preprocessing, AI translation com- parison, evaluation metrics, and generation of the final accuracy report. Each step is designed to ensure systematic preservation and analysis of Kannada folk songs using Artificial Intelligence and Natural Language Processing (NLP) techniques.

I. Dataset Collection and Preprocessing

The initial phase of the project focused on collecting Kannada folk songs from reliable online and offline sources, representing cultural narratives, traditional practices, and symbolic expressions rooted in rural Karnataka. For each song, the dataset includes original Kannada lyrics, expert-generated English translations providing human interpretation, AI-generated English translations, optional cleaned Kannada transcriptions, and relevant metadata such as song title or theme. All collected and processed data were systematically organized into a structured Excel file named FolkDatasetUpdated.xlsx, which serves as the foundational dataset for training, evaluation, and comparison in the proposed system.

II. Dataset Preparation and Organization

After collecting the folk songs, the dataset was preprocessed and standardized into a consistent, machine-readable format. This involved organizing the data into clearly defined columns, namely Song Name, Original Kannada Lyrics, AI Translation, and Expert Meaning. All text entries were encoded in UTF-8 to support Kannada script, unnecessary special characters and formatting inconsistencies were removed, and missing or incomplete entries were carefully handled. The final dataset was structured such that each row corresponds to a single folk song, enabling systematic and reliable comparison between AI-generated and expert translations.

III. Preprocessing of Text

After collecting the folk songs, the dataset was preprocessed and standardized into a consistent, machine-readable format. The data was organized into clearly defined columns including Song Name, Original Kannada Lyrics, AI Translation, and Expert Meaning. All text entries were encoded using UTF-8 to ensure proper representation of Kannada script, while unnecessary special characters and formatting inconsistencies were removed. Missing or incomplete entries were carefully handled to maintain data quality. The final dataset was structured such that each row represents a single folk song, enabling systematic and reliable comparison between AI-generated and expert translations.

IV. AI Translation Evaluation and Comparison

The AI translation for each Kannada folk song was generated using an existing neural machine translation system commonly used for Kannada-to-English translation. These systems are based on modern neural architectures capable of learning contextual and semantic relationships within text. The generated English translations represent how contemporary AI models interpret Kannada folk literature and were directly compared with expert-provided English meanings of the same songs. Since folk songs often contain cultural metaphors, symbolic expressions, and poetic elements, the AI-generated translations may differ significantly from human interpretations. This process establishes a foundation for assessing the ability of AI systems to preserve the semantic and cultural depth of traditional folk literature.

To evaluate translation quality, a pairwise comparison strategy was adopted. For each song, the AI-generated translation and the expert translation were independently preprocessed using identical text-cleaning steps to ensure fair evaluation. Similarity scores were then computed for each pair using selected NLP-based metrics, and the results were stored in a structured report along with the corresponding song name. This approach ensures consistent, song-wise evaluation and enables identification of patterns where AI translations align closely with expert meanings or fail to capture contextual nuances.

The evaluation employs three complementary NLP-based similarity metrics to provide a comprehensive assessment of translation accuracy. TF-IDF Cosine Similarity measures semantic similarity by converting both translations into TF-IDF vectors and computing the cosine similarity between them, where higher values indicate stronger semantic alignment. The chrF Proxy Score, a character-level n-gram metric, captures partial matches and morphological variations, making it effective for linguistically rich and poetic text. Jaccard Wordset Similarity evaluates lexical overlap by measuring the proportion of common unique words between AI and expert translations, offering insight into vocabulary-level similarity. Together, these metrics provide a balanced evaluation of semantic, structural, and lexical correspondence between AI-generated and expert translations.

V. SYSTEM DESIGN

The GaanaSampradaya project focuses on evaluating the semantic accuracy of AI-generated English translations of Kannada folk songs by comparing them with expert interpretations. Since folk songs carry deep cultural meaning, metaphors, and symbolic expressions, traditional translation systems often struggle to capture their true essence. This project addresses that gap by building a structured evaluation pipeline that preprocesses the data, applies similarity metrics such as TF-IDF cosine similarity, chrF proxy score, and Jaccard wordset similarity, and generates a detailed accuracy report. The system is designed to be modular, efficient, and capable of handling diverse textual inputs, making it a valuable tool for cultural preservation and low-resource language research.

I. Architecture Design

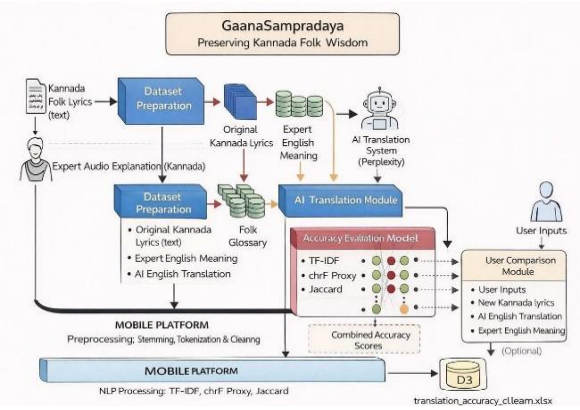

Figure 1: Architecture Design

The GaanaSampradaya system follows a modular architecture designed to evaluate AI-generated English translations of Kannada folk songs by comparing them with expert interpretations. Kannada lyrics, expert explanations, and AI translations are organized into a structured dataset, with expert meanings serving as the reference standard. A folk glossary is maintained to preserve culturally significant terms and contextual consistency.

All text data undergo preprocessing, including cleaning and tokenization, before evaluation. The accuracy evaluation module computes similarity using TF-IDF cosine similarity, chrF proxy score, and Jaccard wordset similarity, which together assess semantic, structural, and lexical alignment between AI and expert translations.

The system also supports direct user-level comparison for individual inputs.
Final accuracy scores are stored for analysis, enabling systematic evaluation and future extensibility.

II. Data Flow Design

The Data Flow Diagram (DFD) of the GaanaSampradaya system illustrates the movement of data from input collection to translation evaluation. The process begins with the collection of Kannada folk lyrics and expert explanations provided by users or researchers.
These inputs are transcribed and translated to generate expert English meanings and AI-based English translations.

The translated content is organized during dataset preparation into a structured FolkDataset containing song details, original lyrics, AI translations, and expert interpretations. This dataset is then processed by the accuracy evaluation model, which computes similarity scores using TF-IDF cosine similarity, chrF proxy score, and Jaccard wordset similarity.

The resulting accuracy scores are stored for analysis. Additionally, the system includes a user comparison module that enables direct evaluation of manually provided translations. This DFD clearly represents the structured and systematic flow of data throughout the system.

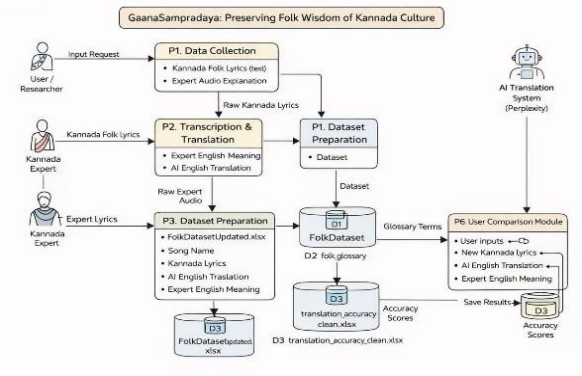

Figure 2: Data Flow Design

III. Use Case Design

The Use Case Diagram of the GaanaSampradaya system illustrates the interaction between users, Kannada experts, and the AI translation system within the translation evaluation framework. Users or researchers provide Kannada folk lyrics and related data, while Kannada experts contribute expert interpretations that serve as reference translations. The AI translation system generates English translations for the given lyrics.

The system supports key functionalities such as dataset organization, folk glossary generation, translation accuracy evaluation, and comparison of new translations using similarity metrics. Users can also view evaluation reports produced by the system. This use case diagram highlights the coordinated interaction of all actors to enable structured, user-driven analysis of Kannada folk song translations.

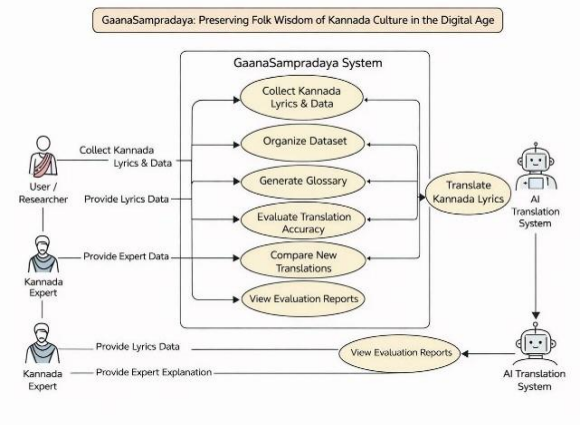

Figure 3: Use Case Design

IV. Model Integration Design

The model design of the GaanaSampradaya system, shown in Figure, illustrates the workflow used to evaluate the accuracy of AI-generated English translations of Kannada folk songs. The process begins with the input of Kannada folk lyrics along with expert explanations and expert lyric data. These inputs are organized during the dataset preparation stage, where linguistic and cultural information is structured into a folk dataset. From this dataset, a glossary generation module extracts frequently occurring culturally significant Kannada terms and assigns contextual importance to them, ensuring preservation of cultural meaning during evaluation. The AI translation module then generates English translations for the Kannada lyrics, which are compared against expert English interpretations.

The accuracy evaluation module forms the core of the model, computing similarity scores using TF-IDF cosine similarity, chrF proxy score, and Jaccard wordset similarity. These metrics collectively measure semantic, structural, and lexical similarity, and are combined to produce an overall accuracy score. The final evaluation results are stored for further analysis and reporting.

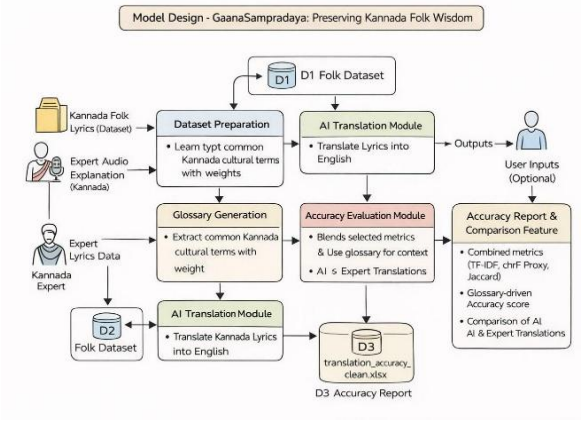

Figure 4: Model Integration Design

VI. RESULT AND ANALYSIS

The GaanaSampradaya project evaluates the semantic accuracy of AI-generated English translations of Kannada folk songs using three similarity metrics. The results obtained from the dataset, transcription outputs, evaluation pipeline, and manual comparison feature together pro- vide a clear understanding of how effectively the AI captures cultural meaning, vocabulary, and structure. This chapter discusses the findings observed during experimentation and interprets what these results reveal about the capabilities and limitations of AI translation systems in low-resource cultural contexts.

I. Dataset Analysis

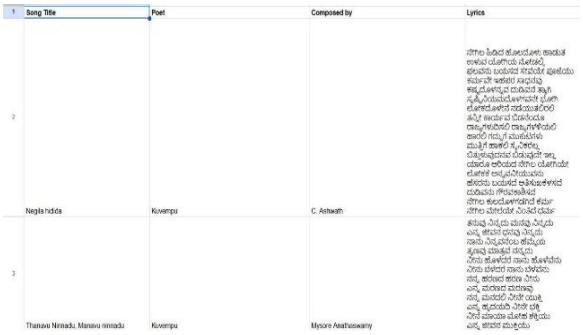

Figure 5: Dataset Analysis

The dataset used for evaluation, illustrated in Figure, consists of Kannada folk song lines along with their AI-generated English translations and expert-provided meanings. Analysis of the dataset revealed significant linguistic diversity, ranging from simple descriptive lines to highly poetic, symbolic, and culturally rich expressions. While straightforward lines related to nature or daily life were generally translated more accurately by AI systems, metaphorical and culturally dependent expressions posed notable challenges. This variation resulted in observable performance differences across songs in the evaluation results. The analysis further highlighted that Kannada folk songs rely heavily on metaphor, rhythm, and oral tradition, which makes them particularly difficult for generic translation models. Recognizing these characteristics helped explain the variation in similarity scores across different translations.

II. Expert Transcription and Interpretation Analysis

The expert interpretations, shown in Figure, serve as the reference or "gold standard" for evaluating AI-generated translations. These interpretations are derived from carefully transcribed folk song audio or authoritative written sources and often include necessary cultural explanations. Analysis reveals that expert meanings go beyond literal translation by incorporating cultural context, emotional depth, and symbolic significance. Such richness is crucial, as many Kannada folk songs convey layered meanings that cannot be captured through direct word-for-word translation. This highlights the gap AI systems must bridge and underscores the importance of culturally sensitive evaluation frameworks like GaanaSampradaya. The expert interpretations, shown in Figure, serve as the reference or "gold standard" for evaluating AI-generated translations. These interpretations are derived from carefully transcribed folk song audio or authoritative written sources and often include necessary cultural explanations. Analysis reveals that expert meanings go beyond literal translation by incorporating cultural context, emotional depth, and symbolic significance. Such richness is crucial, as many Kannada folk songs convey layered meanings that cannot be captured through direct word-for-word translation. This highlights the gap AI systems must bridge and underscores the importance of culturally sensitive evaluation frameworks like GaanaSampradaya.

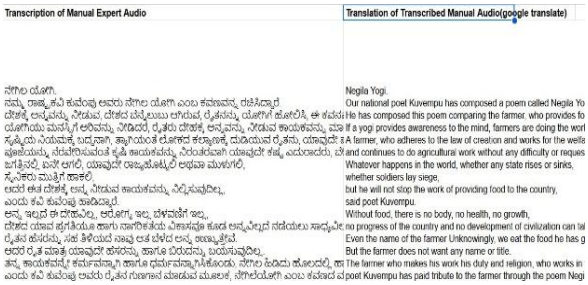

Figure 6: Expert Interpretation Analysis

III. Evaluation Output Analysis

The evaluation model generated three similarity scores—TF-IDF cosine similarity, chrF proxy score, and Jaccard wordset similarity—for each AI–expert translation pair, as shown in Figure 8.3. The analysis revealed that AI translations performed relatively well for simple, literal lines but showed notable deviations when handling metaphorical or culturally rich content. Higher TF-IDF scores indicated partial semantic alignment, while lower chrF and Jaccard scores reflected structural and lexical differences. This pattern suggests that AI systems often simplify or paraphrase translations, preserving readability but losing cultural depth. Overall, the metric-based evaluation effectively highlights both semantic alignment and areas where cultural nuance is diminished.

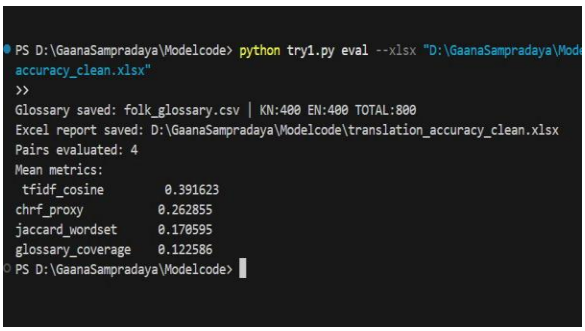

Figure 7: Evaluation Output Analysis

IV. Manual Comparison Output Analysis

The manual comparison output analysis examines results generated when users directly input an AI-generated translation and an expert interpretation into the system, as shown in Figure 8.4. This feature enables evaluation of individual translation pairs without relying on the full dataset, while still applying the same preprocessing and evaluation pipeline to maintain consistency. Similarity scores are computed using TF-IDF cosine similarity, chrF proxy score, and Jaccard wordset similarity.

The analysis revealed trends consistent with dataset-level evaluation: translations with simple structure and literal meaning achieved higher similarity scores, whereas metaphorical or culturally rich expressions resulted in lower scores. This indicates that AI systems are more effective at capturing surface-level meaning than deeper cultural nuance. The manual comparison module proved valuable for focused, line-by-line analysis and validated the reliability of the evaluation model, as its results aligned with those from bulk evaluation.

ai_text	expert_text_en	tfidf_cosine	chrf_proxy	jaccard_wordset
Let the Kannada drumbeat resound, O Shiva, the heart of Karnataka! Let the Kannada drumbeat resound, Awaken the dead-hearted, stir the restless, Unite the quarrelsome and reconcile them, Shed tears for the hunger and hardship, Let harmony flow among those who live together.	The door is finished. This is Kuvempu's composition. The present verse describes the art and architecture of the Somnathpur temple. No worship decorations of any kind are performed here. Yet, there is a situation where a pilgrim should bow down to the wonder of the art there. The poet says that there is no need for a priest/priest between God	0.335673	0.126281	0.029412

Figure 8: Manual Comparison Output Analysis

VII. CONCLUSION AND FUTURE SCOPE

The GaanaSampradaya project presents a systematic and effective framework for evaluating the semantic accuracy of AI-generated English translations of Kannada folk songs. Kannada folk literature represents a rich cultural heritage filled with symbolism, emotion, and traditional wisdom, making accurate translation particularly challenging for AI systems—especially in low-resource language settings. By constructing a parallel dataset and applying multiple evaluation metrics such as TF-IDF cosine similarity, chrF proxy score, and Jaccard wordset similarity, the project provides a reliable and structured method to compare AI translations with expert human interpretations.

The evaluation results indicate that while AI models can generally capture surface-level meaning and produce fluent translations, they often fail to preserve cultural nuance, metaphorical depth, and poetic structure inherent in folk songs. The multi-metric evaluation approach effectively highlights these limitations, demonstrating the importance of combining complementary similarity measures rather than relying on a single metric. Overall, the project contributes both to AI translation evaluation research and to cultural preservation by offering a measurable way to assess translation quality for traditional folk content.

Future enhancements to the GaanaSampradaya system include fine-tuning translation models on Kannada or broader Dravidian language corpora, expanding the dataset to cover more folk songs, dialects, and audio recordings, and incorporating additional evaluation metrics such as BLEU, BERTScore, ROUGE, or embedding-based similarity. The system also has the potential to evolve into an interactive platform for researchers, educators, and cultural historians, supporting real-time translation analysis and advancing research in low-resource NLP, cultural translation, and digital heritage preservation.

REFERENCES

[1] Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, “Attention Is All You Need,” in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017.

[2] CompMusic Project, “CompMusic: Cultural Music Preservation,” Universitat Pompeu Fabra, Barcelona.

[3] R. Aharoni, M. Johnson, and O. Firat, "Massively Multilingual Neural Machine Translation," arXiv preprint arXiv:1903.00089, 2019.
[4] R. Rapp, "Using Semantic Role Labeling to Improve Neural Machine Translation," in Proc. 13th Int. Conf. Language Resources and Evaluation (LREC), 2022.
[5] E. Voita, R. Sennrich, and I. Titov, "Context-Aware Neural Machine Translation Learns Anaphora Resolution," arXiv preprint arXiv:1805.10163, 2018.
[6] M. Artetxe, G. Labaka, E. Agirre, and K. Cho, "Unsupervised Neural Machine Translation," arXiv preprint arXiv:1710.11041, 2018.
[7] S. Ranathunga et al., "Neural Machine Translation for Low-Resource Languages: A Survey," arXiv preprint arXiv:2104.04906, 2021.
[8] R. Wang, X. Tan, R. Luo, T. Qin, and T.-Y. Liu, "A Survey on Low-Resource Neural Machine Translation," arXiv preprint arXiv:2107.04239, 2021.
[9] R. Chandra and V. Kulkarni, "Semantic and Sentiment Analysis of Selected Bhagavad Gita Translations Using BERT-Based Language Framework," Journal of Cultural Analytics, 2022.
[10] S. Koneru, D. Liu, and J. Niehues, "Unsupervised Machine Translation on Dravidian Languages," in Proc. First Workshop on Natural Language Processing for Indian Languages, 2021.
[11] J. Gala et al., "IndicTrans2: Towards High-Quality and Accessible Machine Translation Models for All 22 Scheduled Indian Languages," arXiv preprint arXiv:2305.16307, 2023.
[12] R. G. R. Shanker et al., "Tollywood Emotions: Annotation of Valence-Arousal in Telugu Song Lyrics," in Proc. 2023 Conf. on Empirical Methods in Natural Language Processing (EMNLP), 2023.
[13] A. Das et al., "A Study of Attention-Based Neural Machine Translation Model on Indian Languages," in Proc. 2016 Conf. on Empirical Methods in Natural Language Processing (EMNLP), 2016.
[14] H. Choudhary, S. Rao, and R. Rohilla, "Neural Machine Translation for Low-Resourced Indian Languages," in Proc. 2020 Conf. on Computational Linguistics, 2020.
[15] S. Deshpande et al., "A Review on Various Approaches in Machine Translation for Sanskrit Language," *International Journal of Advanced Research in Computer Science*, 2020.