Bridging Linguistic Divides: A Novel Transformer-Based Neural Machine Translation Framework for Gondi-Hindi Translation

Rahul Shukla, Bhavesh Ajwani, Santosh Kumar

Abstract—In an era where cultural preservation is paramount, bridging linguistic gaps between endangered and low-resource language groups is crucial. This paper introduces a groundbreaking approach for translating the endangered tribal Gondi language into Hindi, fostering dynamic interaction within tribal communities. Our innovative methodology employs a Transformer-based Neural Machine Translation (NMT) model, enriched with learnable positional encodings and vertically weighted attention mechanisms, to facilitate seamless bidirectional translation between Gondi and Hindi. Our comprehensive framework begins with a meticulous preprocessing pipeline, incorporating advanced noise filtering and a unique dialect concatenation step, addressing Gondi dialects such as Dorla, Koya, Madiya, Muria, and Raj Gond. The Transformer-based NMT model, featuring weighted multi-head attention and learnable positional encodings, effectively processes and translates text inputs with high accuracy. To enhance accessibility, we developed an Android application that functions offline, ensuring usability in remote areas with limited internet connectivity. This application is a crucial language bridge, leveraging cutting-edge shapeshifter technology to support Gondi-speaking tribal communities. Empirical results are compelling: our proposed framework achieves a remarkable BLEU score of 98.99%, with a minimum latency of 87 milliseconds, marking a substantial improvement over existing models. Correlation scores between the English-to-Gondi (E2G) and English-to-Hindi (E2H)/Hindi-to-Gondi (H2G) Machine Translation (MT) systems are 0.99.98% and 96.89%, respectively. These findings underscore the effectiveness of our approach, offering large-scale Gondi language data collection and advancing the field of language technology. The proposed framework bridges the linguistic divide and empowers tribal communities to preserve their cultural heritage and engage more effectively in diverse settings. This work sets a new benchmark in language translation, paving the way for future innovations in low-resource language processing.

Index Terms—Neural Machine Translation, Transformer, Deep Learning, Attention Mechanism, Weighted Multi-head Attention, Learnable Positional Encodings, Noise Filtering, Dialect Identification

1 Introduction

LANGUAGE the essence of human connection and cultural identity serves as a conduit for transmitting knowledge, traditions, and heritage [1]Amidst the vast array of languages spoken worldwide, a stark reality emerges: between 6,000 to 7,000 languages currently exist, yet between 50% and 90% of them face severe endangerment or extinction by the dawn of the 22nd century [4]. The forthcoming extinction of languages portends the loss of linguistic diversity and the erosion of invaluable cultural traditions and historical narratives. Indeed, when a language falls silent, entire cultures and ways of life fade into obscurity, underscoring the urgent need for preservation efforts [5] [6]. Recognizing this imperative, UNESCO has spearheaded initiatives such as the Convention on the Protection and Promotion of the Diversity of Cultural Expression since 2005 [7]

Amidst the urgency to prevent linguistic extinction, a notable focus emerges on preserving indigenous dialects, with particular attention drawn to Gondi. The Gondi people, known as "Koitur" in their self-reference, intricately weave their cultural narrative into India's rich tapestry [8]. As native speakers of the Dravidian language, Gondi, their presence spans regions across Madhya Pradesh, Maharashtra, Chhattisgarh, and beyond. Their historical legacy, encapsulated in realms such as Gondwana and the Garha Kingdom, underscores their profound cultural significance [5]- [8] However, the Gondi language, entwined with Telugu roots, confronts the dual challenges of modernization and the overshadowing dominance of more prominent regional languages like Hindi and Marathi. Despite a substantial population, as recorded in the 2001 census with approximately 11 million Gond [9]- [10] individuals, linguistic assimilation and sociopolitical complexities, notably exacerbated by the Naxalite–Maoist insurgency, have accelerated the imperative to safeguard and promote their distinctive linguistic heritage.

Addressing the issue of language endangerment, our work focuses on Gondi-Hindi translation, aiming to foster social upliftment by enabling seamless communication between these distinct linguistic groups. Our proposed method utilizes a comprehensive yet efficient framework with a Transformer-based Neural Machine Translation (NMT) model. The process commences with an extensive pre-processing pipeline that ensures data de-noising and

Rahul Shukla is pursuing a B.Tech in Electronics and Communication Engineering (ECE) from IIIT Naya Raipur, Chhattisgarh (e-mail: rahuls20101@iiitnr.edu.in).
Bhavesh Ajwani is pursuing his B.Tech in Data Science and Artificial Intelligence (DSAI) from IIIT Naya Raipur, Chhattisgarh (e-mail: bhavesh20102@iiitnr.edu.in).
Santosh Kumar is working as an Assistant Professor with the Department of Computer Science and Engineering at IIIT-Naya Raipur, Chhattisgarh (e-mail:santosh@iiitnr.edu.in).

Table 1: Literature survey of the related works

Ref.	Method	Dataset Used	Model used	BLEU score	Remark
Sarode et al. [12]	Neural-based MT	Multiple parallel texts	RNN encoder-decoder with GRU, attention	0.83%	Focused on popular languages
Baliyan et al. [13]	Sentiment Analysis with NMT Labelling	Labeled English dataset	GloVe embeddings with RNN-LSTM	0.76%	Struggles with scalability
A. P. et al. [14]	Comparison of SMT and NMT	Tamil-Sinhala parallel texts	LSTM vs. Transformer NMT	0.79%	Limited efficiency for low-resource
Jha et al. [16]	Multilingual NMT for Indian Languages	Asian Language Treebank	mT5 transformer	0.81%	No audio translation
Our Work	Gondi-Hindi NMT	Gondi-Hindi parallel texts	Transformer NMT with weighted attention	0.59%	Efficient, 87 ms latency

dialect concatenation, thus paving the way for accurate translations across different Gondi dialects, including Dorla, Koya, Madiya, Muria, or Raj Gond [12]- [13] . The next stage of our proposed methodology incorporates implementing a Transformer-based NMT model equipped with innovative features such as weighted multi-head attention and learnable positional encoding. These elements work synergistically to process the text inputs effectively, while visualizations of attention weights across the model's four layers and six heads facilitate hyperparameter tuning and model optimization. The quality and effectiveness of the model are demonstrated by high-performance BLEU scores and perplexity evaluations, alongside latency measurements that assess the translation speed, further solidifying its potential as a tool to bridge the communication divide among these diverse language groups.

1.1 Motivations and contributions

More than 476 million indigenous and tribal people live worldwide [9]– [15]. More than 104 million tribal people, including the scheduled tribe's population in India, is 10.43 core. i.e., more than 8.6% of the national population. Approximately 50% of the population is estimated to reside in forested areas in India with low resource availability [15]. The tribal community relies on the Minor Forest Produce (MFP) for sustainable livelihood and income generation [16]. MFP collection and marketing hold critical importance for tribal people as they spend a significant portion of their valuable time on it, and they derive a significant portion of their income from forest produce such as honey, dry fruits, leaves, and others [17]. Due to the language barrier, tribal communities cannot communicate with modern society, and their growth is accelerating with the rapid growth of disruptive technologies worldwide. The preliminary source availability for educating tribal communities with modern education and providing them with sustainable development in tribal regions has been demolished due to huge gaps in efficient communication for tribal people [17]. Because Gondi is a primary language for communication between tribal people, tribal communities do not have other language diversity to transform their cultural values for social upliftment through the establishment of better communication [1] [4].

The proposed framework uses a transformer-based NMT model to translate the Gondi language into Hindi for better communication with tribal people. The proposed framework takes Gondi data for pre-processing and dialect identification to ensure accurate translations across Gondi dialects. The proposed framework-based NMT model employs innovations like customized weighted multi-head attention and learnable positional encoding to process Gondi inputs effectively, as evidenced by a high Bilingual Evaluation Understudy (BLEU) score for classification. Moreover, in this work, we develop a working prototype *"Gondi Gaatha"* Android application that embodies the aspiration to empower tribal members to actively participate in linguistic preservation, contributing to a repository of Gondi language data and promoting digital inclusion through offline functionality. Ultimately, the motivations resonate beyond translation, embracing cultural heritage, social cohesion, and technological progress, converging towards a more inclusive and interconnected world. The research contributes to scholarly discourse and preserving languages like Gondi, fostering communication among diverse linguistic groups. The significant contributions of this work are summarized as follows:

1) We proposed a novel framework for Hindi-Gondi neural machine translation, incorporating transformer models renowned for Gondi's language understanding capabilities. The paper presents an innovative solution that not only translates text but also respects the contextual nuances of the Gondi language, thereby improving translation accuracy and cultural relevance.
2) We developed a working prototype *"Gondi Gaatha"* Android application, a significant contribution; a user-friendly platform empowers tribal communities, enabling them to participate in the translation process. The application is a repository of Gondi language resources, offering real-time translation, access to Gondi literature, a comprehensive dictionary, and culturally relevant quotes.

3) The paper uniquely focuses on community engagement as a cornerstone for linguistic preservation. By involving tribal members in translation efforts, the study contributes to accumulating valuable Gondi language data, which can be further utilized to advance language technologies and cultural documentation.
4) The developed application's offline functionality addresses the connectivity challenges in tribal regions, ensuring access to linguistic resources even in areas with limited internet connectivity. The technological inclusion aligns with the broader objective of reducing digital disparities and providing equitable access to language tools for tribal communities for better sustainable livelihood. The performance of the proposed framework is validated with different current state-of-the-art methods based on various benchmark settings.

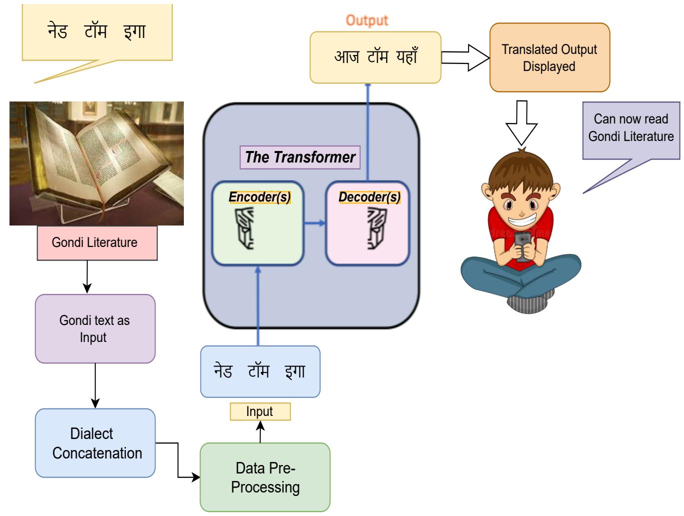

Figure 1: Working of the proposed framework for Gondi to Hindi

1.2 Paper structure

The rest of the paper is organized as follows. Section **??** discusses the proposed framework, and results and discussion are discussed in Section III. Section IV describes the case study, and the paper concludes in Section V.

2 Proposed Framework

Our proposed framework, as depicted in Fig. 5 presents a cutting-edge approach for Hindi-Gondi translation, integrating a Transformer-based Neural Machine Translation (NMT) model into a comprehensive system. This system commences with an elaborate pre-processing pipeline that involves noise filtering and dialect concatenation to process the raw Gondi-Hindi parallel dataset. It concatenates the specific Gondi dialect—Dorla, Koya, Madiya, Muria, or Raj Gond—before translation, enhancing translation accuracy. Once preprocessed, the transformer-based NMT model processes the textual inputs with its weighted multi-head attention and learnable positional encodings. Rigorous evaluation validates our approach's efficacy in delivering precise translations for the Hindi-Gondi language pair, reinforcing its potential to bridge the communication gap between these diverse linguistic communities.

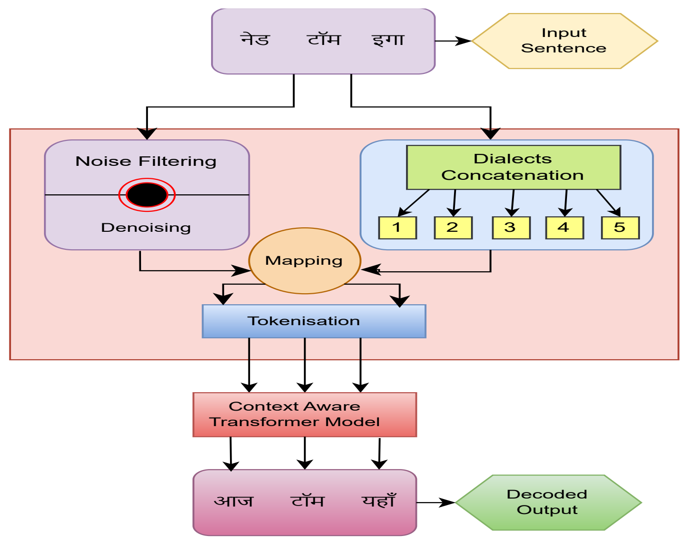

Figure 2: Conceptual View of Proposed Method

2.1 Data Preparation and Description

The Gondi-speaking community, spread across six Indian states, has encountered notable dialectical variations influenced by dominant regional languages. This linguistic diversity, though reflective of cultural dynamics, has posed occasional challenges in inter-Gond communication. In response, we conducted a series of workshops that convened Gondi representatives from each state, culminating in the creation of a comprehensive Gondi language dataset, comprising an impressive 30,000-row sample (shown in Fig. 3). This dataset, derived from regions dense in Gondi speakers like Bastar in Chhattisgarh, serves as a foundational cornerstone for future initiatives in Hindi-Gondi translations.

Amidst this cultural renaissance, an in-depth analysis of the dataset as shown in Fig. 3 reveals fascinating insights into the structural characteristics of Gondi and Hindi sentences. On average, Gondi sentences contain around 12.31 words, with a median of 7 words and a mode of 5 words. The range of Gondi sentence lengths spans from 1 to 286 words, with a standard deviation of approximately 16.33 and an interquartile range (IQR) of 9 words. Whereas, Hindi sentences range with an average consisting around 14.24 words, and they have a median of 9 words with mode of exactly 6 words. The span that hindi sentences lengths coverage from 1 to 246 words, and standard deviation of 18.21 and an IQR with 12 words.

Furthermore, we prepared the Gondi language dataset which led to the The Gondi-Hindi dataset that we created resulted in the creation of "Gondi Manak Shabdkosh" dictionary that takes care of the influence that dominant state imposes on the language so as to preserve Gondi's core essence This dictionary serves as a representation of the transformational power of collaborative language projects by facilitating accurate translations between Hindi

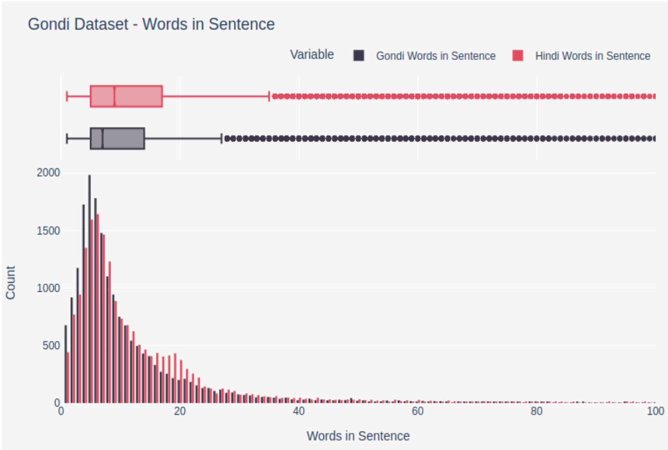

Figure 3: Words Distribution in Gondi Dataset

and Gondi and enhancing communication within the Gond community. Through enriched dialectic exchanges in these workshops, the dataset encompasses diverse textual content from various regions, as detailed in Table **??**. This table illustrates the dialects spoken in different regions, emphasizing the importance of preserving these vulnerable and endangered dialects.

Gondi Dialects and their Distribution			
Region	Dialect	Lipi	Status
Bastar	Dorla	Devanagari	Endangered
Gadchiroli	Koya	Telugu	Vulnerable
Dantewada	Madiya	Devanagari	Endangered
Adilabad	Muria	Odia	Vulnerable
Mandla	Raj Gond	Devanagari	Endangered

2.2 Data preprocessing and Augmentation

In our research, we developed a sophisticated pipeline for pre-processing of the Gondi languages, noise filtering, and dialect identification as the pipeline for the raw Gondi-Hindi parallel dataset, a crucial step in optimizing the effectiveness of our transformer-based neural machine translation model.

2.2.1 Efficient Text Processing

We used the text-pre-processing method known as tokenization. We segmented the raw Gondi text data into individual words or tokens in this step. The proposed framework used ensemble machine learning models to interpret the Gondi text data by converting the permit or otherwise unstructured data into a unique format that models learn from the processed text data for training purposes. Furthermore, we used tokenization of the Gondi text data that is further amplified in our prepared Gondi dataset, primarily because of Gondi's rich morphological structure, which resembles other Dravidian languages. Proper tokenization ensures that the model effectively captures and utilizes these intricate morphological characteristics.

2.2.2 Noise Filtering

Arguably, the most critical aspect of our preprocessing pipeline was noise filtering. 'Noise' in this context refers to any words or tokens in the dataset that do not contribute to effective learning or could hinder it. This could include irrelevant, inconsistent, too infrequent, or overly ubiquitous words. Given that Gondi has various dialectal variations and the possibility of extraneous lexical items in the dataset, noise filtering was necessary and indispensable. To implement noise filtering, we devised a mathematical function that measured the information entropy of the words in the corpus. Entropy is a concept borrowed from Information Theory (IT) that measures the uncertainty or randomness of information. We hypothesized that 'noisy' words would have high entropy due to their inconsistency or randomness. Hence, by measuring the entropy of words, we could identify and filter out the 'noisy' words.

$$E(w) = -\sum_i p_i \log(p_i) \quad (1)$$

Where $E(w)$ depicts the entropy of the word $[w]$ and $[p_i]$ denotes the probability of the i^{th} context word given $[w]$. We calculated the context word probabilities within a defined window size around each word in the Gondi sentences. This method allowed us to significantly reduce the noise in our dataset

2.3 Dialect Concatenation

Incorporating dialect identification into the translation pipeline, the primary objective of the proposed framework is to leverage to enrich the input Gondi language representation for the customized transformer model. By concatenating the dialect label with the original Gondi text which is denoted as

$$\text{Input}_{\text{concat}} = \textbf{G}\text{ondi} \oplus \textbf{D}\text{ialect} \quad (2)$$

We provide the supplied data a layer of added complexity. By using this method, the input becomes more contextually rich and gives the model more subtle linguistic and dialectal information. The transformer model is then tasked with translating the concatenated input into Hindi after being pre-trained on a sizable corpus of parallel Gondi-Hindi literature. By addressing the nuances of Gondi dialects during the translation process, we hope to advance the output's complexity and fidelity through this methodological integration.

2.4 Context-Aware Transformer NMT model

The chosen model for our translation task is based on the Transformer architecture, a type of neural network introduced by Vaswani et al. [21]. The Transformer model stands out due to its non-sequential attention mechanism. Traditional recurrent neural networks (RNNs) handle sequence data by looping through each element in the sequence, which can be computationally expensive and challenging to parallelize. In contrast, the Transformer model simultaneously applies self-attention to the entire sequence, enabling it to process sequences in parallel and capture dependencies regardless of their distance. This quality is

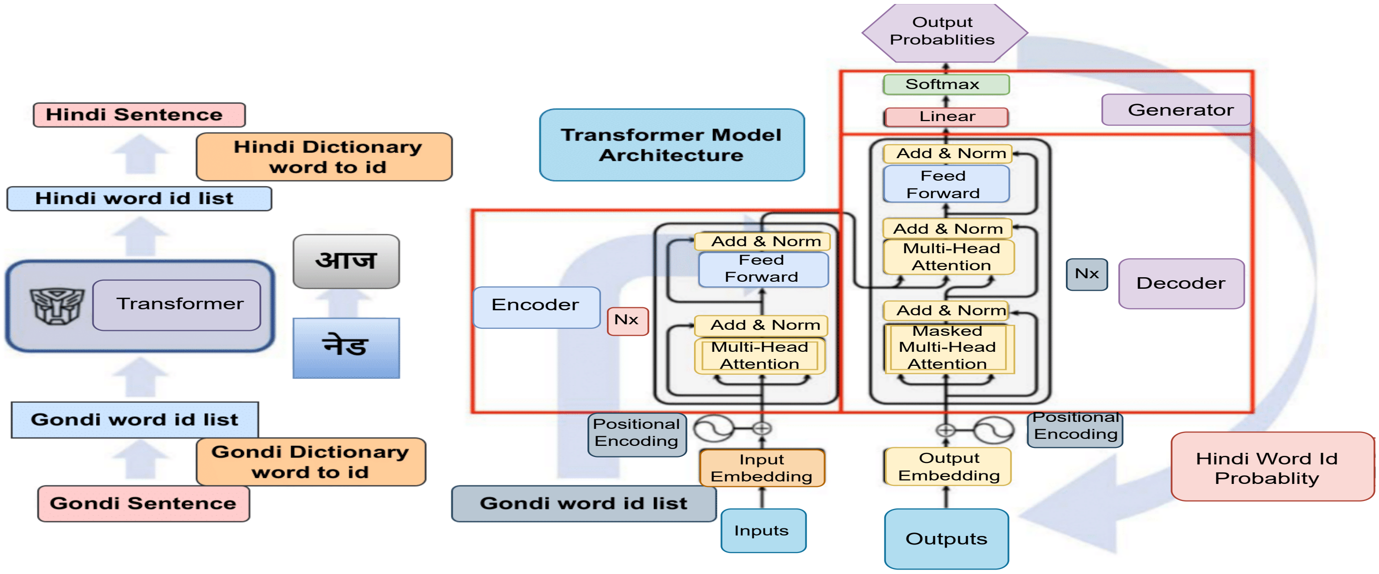

Figure 4: Layered architecture of proposed model

especially beneficial in tasks such as translation, where long-distance relationships between words can carry crucial meaning. The core components of the Transformer model include an encoder stack, a decoder stack, and connections between them. The encoder and decoder are composed of a series of identical layers stacked on each other, facilitating efficient parallelization and effective handling of complex dependencies in sequence data.

In our implementation, each encoder and decoder stack has N layers (the optimal number of layers can be determined experimentally). Each layer in the encoder comprises two sub-layers: the first is a multi-head self-attention mechanism, and the second is a position-wise, fully connected feed-forward network. In the decoder, there is an additional sub-layer that performs multi-head attention on the outputs from the encoder stack. Residual connections are present around each of these sub-layers, and they are followed by layer normalization. As depicted in Figure 4, the multi-head attention mechanism enables the model to simultaneously focus on various positions within the input.

In this setup, the model uses multiple sets of learned linear transformations. Suppose we have h heads; each of these heads has its own learned linear transformations represented by weight matrices W_i^Q, W_i^K, W_i^V (for i=1$\cdots$ h).

For each head i, the input vectors are independently transformed into queries, keys, and values using these weight matrices:

$$Q_i = XW_Q^i, \quad K_i = XW_K^i, \quad V_i = XW_V^i \tag{3}$$

The attention output for each head is then:

$$\text{Attention}_i(Q_i, K_i, V_i) = \text{softmax}\left(\frac{Q_i K_i^\top}{\sqrt{d_k}}\right) V_i \tag{4}$$

The Transformer model extends this idea to multi-head attention. This mechanism runs the attention process in parallel (h heads), using different learned linear projections of the original q, k, and v vectors. The output of each chair is concatenated and linearly transformed to result in the final production. This allows the model to capture various types of information in the input sequence. We introduce weights $[\alpha_i]$ for each attention head $[i]$ in the multi-head attention mechanism. The final output of the attention mechanism becomes a weighted sum of the work of each chair:

$$\text{Attention\_output} = \sum_{i=1}^{h} \alpha_i \times \text{Attention}_i(Q_i, K_i, V_i) \tag{5}$$

The weights $[\alpha_i]$ are used as learned during the training of the model. This allows the model to understand the relative importance of each attention head. The Transformer uses positional encoding to inject information about the position of the words in the sequence. For each position p in the sequence and each dimension i of the input representation, the positional encoding is calculated as follows:

$$PE(p, 2i) = \sin\left(\frac{p}{10000^{2i/d}}\right) \tag{6}$$

$$PE(p, 2i+1) = \cos\left(\frac{p}{10000^{2i/d}}\right) \tag{7}$$

The learnable positional encoding matrix PE_{learn}, which is added to the original positional encodings PE_{orig}. The new positional encodings PE_{new} are given by:

$$PE_{\text{new}} = PE_{\text{orig}} + PE_{\text{learn}} \tag{8}$$

The PE_{learn} matrix is initialized with small random values and updated during training, allowing the model to adjust the positional encodings based on the data.

Figure 5 vividly showcases the positional encoding with a dynamic spectrum of sinusoidal waves, each representing the unique fingerprint of sequence positions and depths within the transformer model.0

After passing through the decoder, the output is transformed using a final linear layer and a softmax operation to generate the output probabilities for each word in the vocabulary:

$$P = \text{softmax}(Y'''W + b) \tag{9}$$

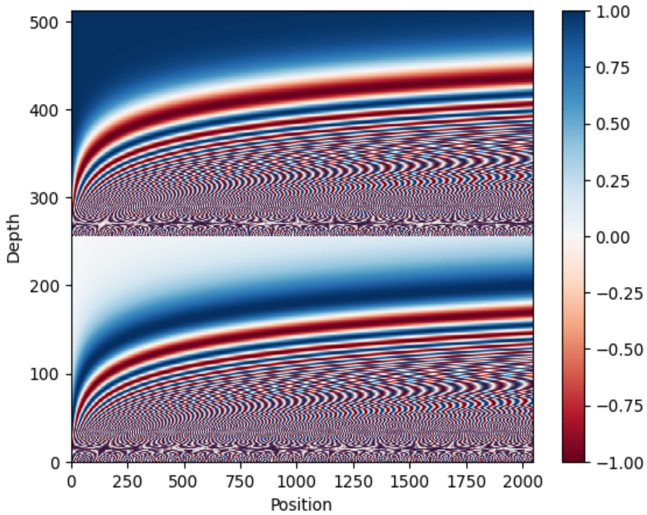

Figure 5: Spectral Visualization of Positional Encoder

Where W and b are the final linear layer's weight matrix and bias, respectively, the softmax () operation is applied over the vocabulary dimension. The proposed framework uses the Transformer model that offers several advantages over recurrent and convolutional models for NLP tasks. The proposed framework performs the following steps: at first, it eliminates the need for recurrence, allowing for much greater parallelization during training and leading to substantial speedups. In the second step, it replaces the sequential processing of RNNs with a constant number of operations, allowing it to handle longer sequences effectively. This allows the model to better capture long-range dependencies common in language. Further, using self-attention, the model can focus on different parts of the input sequence to better understand each word's context. This provides a more nuanced understanding of the sequence, unlike models like LSTM or GRU, which may have difficulty maintaining context over long sequences. The Transformer model's performance is highly competitive, frequently achieving state-of-the-art results on various NLP tasks, including translation, summarization, and sentiment analysis.

3 Results and discussion

In this study, we proposed a *Hindi-Gondi* machine translation model using a modified Transformer architecture equipped with an encoder-decoder structure and enhanced by a weighted multi-head attention mechanism and learnable positional encodings. This proposed model is trained on a large corpus of parallel Hindi and Gondi language sentences prepared from a workshop conducted at IIIT-Naya Raipur, Chhattisgarh. Following the training process, we evaluated the performance of our proposed model using a set of Gondi sentences uttered by Gondi speakers to gauge its translation quality. The performance of the proposed model exhibited a spectrum of results. There were instances where the model demonstrated an impressive comprehension of the context and grammatical structure, generating accurate Hindi translations. However, specific translations are needed to fully capture the intricacies of the Gondi language, providing inaccurate translations or occasionally returning blank outputs. This implies potential issues in either the training process or the dataset.

3.1 Attention Weights Speculisation

The visualization of attention weights of the proposed framework is shown in Fig. 5, which offers critical insights into how different parts of the input sequence influence the output predictions of our transformer model. By examining the intensity and distribution of attention across various heads and layers of the proposed framework, the proposed framework identifies patterns and areas where the model focuses most, revealing dependencies and potential biases in the model's learning process. Such detailed scrutiny aids in hyperparameter tuning by pinpointing layers and heads that might require adjustments in terms of depth, number of heads, or other architectural details. Ultimately, this leads to more informed decisions in model optimization, potentially improving performance and generalization capabilities on unseen data.

3.2 Evaluating Model Performance

To assess the performance of our model in translating Gondi to Hindi, we employed a comprehensive evaluation strategy that includes BLEU scores, Perplexity, and traditional Loss and Accuracy metrics. These measures together provide a holistic view of the model's capabilities in terms of translation quality, linguistic understanding, and overall efficiency. Furthermore, we evaluated the decoding speed by calculating the latency per sentence. This was determined by averaging the time required to translate the entire test set, which comprises 1,000 individual sentences

3.2.1 BLEU Score

The BLEU score is a metric for quantifying the quality of machine-translated text against reference translations. It calculates the precision of n-grams (consecutive word sequences) in the translated text compared to the reference text, adjusted by a brevity penalty to discourage overly short translations. Our model achieved a high BLEU score, indicating it closely matches human reference translations in accuracy and context. This is attributed to our model's ability to capture linguistic nuances effectively through advanced attention mechanisms. The formula for the BLEU score involves calculating the geometric mean of the n-gram precision scores p_n across different n-gram lengths, weighted by w_n, and applying a brevity penalty BP:

$$\text{BLEU} = \text{BP} \cdot \exp\left(\sum_{n=1}^{N} w_n \log p_n\right) \tag{10}$$

The brevity penalty BP addresses the length discrepancy between the candidate and reference translations, penalizing overly short translations:

$$\text{BP} = \begin{cases} 1 & \text{if} c > r \\ e^{(1-r/c)} & \text{if} c \leq r \end{cases} \tag{11}$$

Where c denotes the length of the candidate translation, and r is the effective reference length. This formula ensures

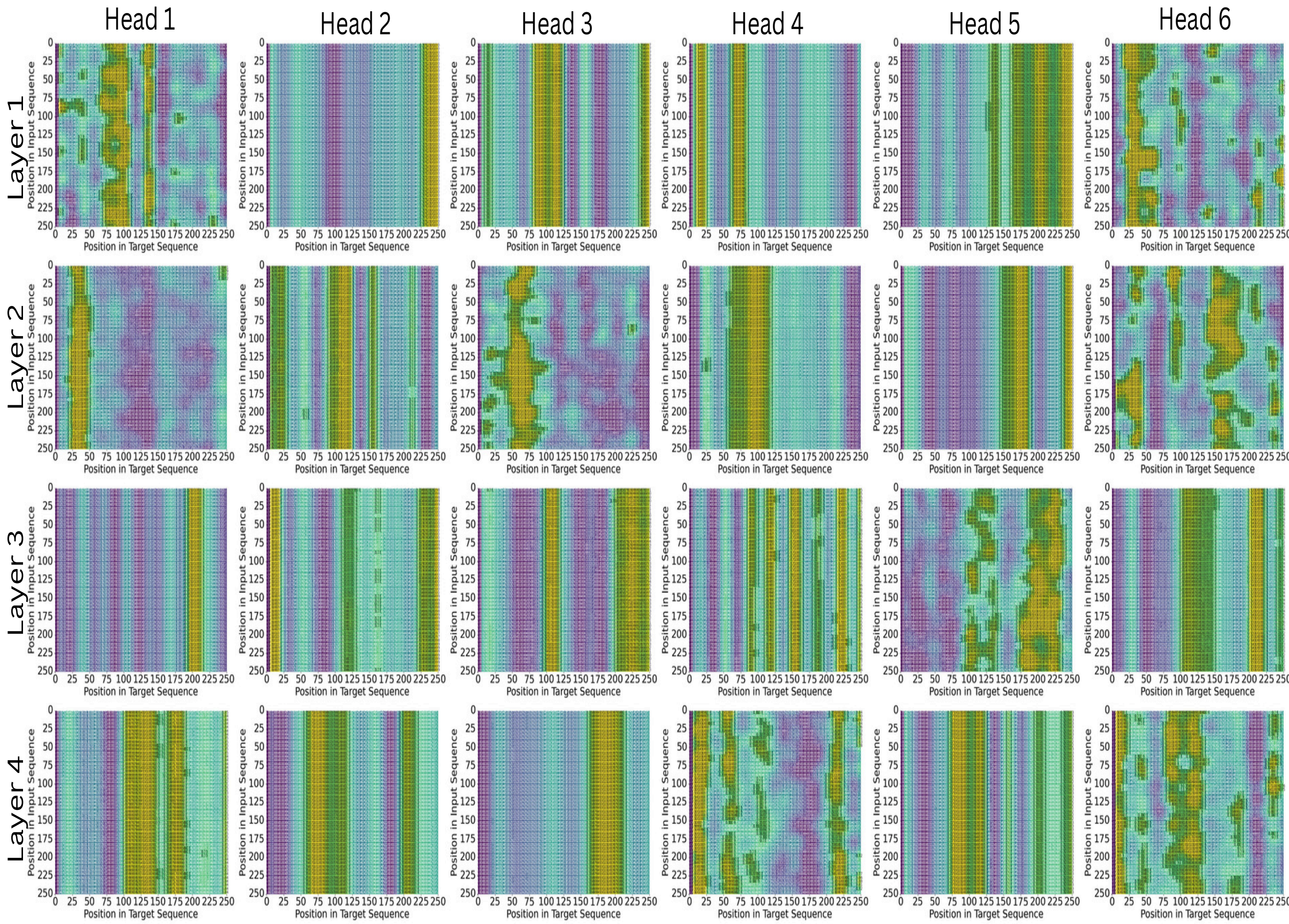

Figure 6: Attention Weights Visualised

Table 2: Training Parameters and Results of Different Multi-class Segmentation Models

Model	Train, Val, Test %	Epochs	BLEU Score	Perplexity	Latency(ms)
LSTM (Seq2Seq)	70%, 20%, 10%	80	0.78	10.0	76
BiLSTM-Attention	70%, 20%, 10%	59	0.84	8.5	91
BERT-base	70%, 20%, 10%	78	0.72	7.0	117
DeBERTa-v3	70%, 20%, 10%	70	0.85	5.5	135
Our Model (Transformer)	70%, 20%, 10%	50	98.99	2.0	87

that translations contain correct phrases and are appropriately concise or expansive.

3.2.2 Perplexity

Perplexity assesses the model's predictive power, with lower scores indicating superior language understanding and predictive accuracy. Our model's low Perplexity score demonstrates its enhanced ability to grasp and predict the complexities of Gondi and Hindi languages. The Perplexity formula is presented as:

$$\text{Perplexity} = 2^{-\frac{1}{N}\sum_{i=1}^{N}\log_2 P(w_i|w_{i-1},\ldots,w_{i-n+1})} \tag{12}$$

In the given expression, N represents the total number of words in the test set, and $P(w_i|w_{i-1},\ldots,w_{i-n+1})$ denotes the model's estimated probability of the word w_i given the preceding $n-1$ words.

3.2.3 Loss and Accuracy

Our model employs Sparse Categorical Cross-entropy as the loss function, adept for multi-class classification tasks like ours, where each Gondi word is mapped to one of many Hindi translations. This choice suits our scenario well, allowing for a direct handling of integers for class labels without the need for one-hot encoding, thus optimizing memory usage and computational efficiency.

The formula for Sparse Categorical Cross-entropy is:

$$\text{Sparse Categorical Crossentropy} = -\sum_{c=1}^{M} y_{o,c} \log(p_{o,c}) \tag{13}$$

Here, M is the number of classes, $y_{o,c}$ is a binary indicator of whether class c is the correct classification for observation o, and $p_{o,c}$ is the predicted probability of observation o being of class c.

Our model's low loss score coupled with high accuracy demonstrates its effectiveness in accurately classifying and translating Gondi to Hindi, reflecting its robust understanding of the language and superior performance in handling the complexities of translation tasks.

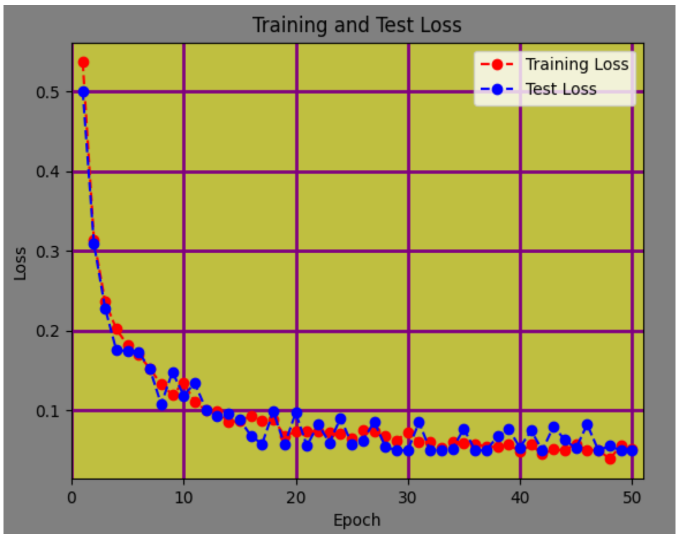

Figure 7: Spectral Visualization of Positional Encoder

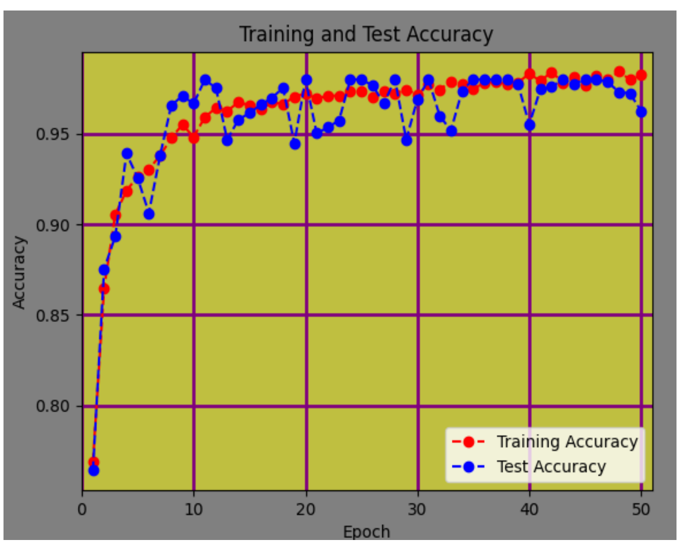

Figure 8: Spectral Visualization of Positional Encoder

3.2.4 Sentence-Level Metric Analysis

Using the earlier methodology for sentence-level correlation analysis, Table. 3 provides the Pearson correlation coefficients between the metrics and human scores using gondi samples. The proposed framework shows high-level metric scores for combining Gondi with other languages (shown in Table. 3).

Table 3: Pearson Correlation Based Metric Scores With Human Scores.

Method	E-T	H-E	G-H	G-U	G-B
black	0.2697	0.2753	0.2956	0.295	0.2695
ChrF++	0.2697	0.2753	0.2956	0.295	0.2695
BERT+SVM	0.4697	0.3653	0.4676	0.4985	0.43695
BERT+para	0.6998	0.6873	0.6979	0.6795	0.69698
Proposed	0.8398	0.8879	0.9598	0.9676	0.9889

Abbreviation: E2T MT= English to Tamil Machine Translation, H2E=Hindi to English Machine Translation, G2H MT= G2H MT, G-U MT= Gondi to Urdu MTn, G2B MT= Gondi-Bengali MT

1) We used sentence-level correlation analysis to evaluate correlations between segmented word tokens. Table 3 provides the Pearson correlation values of the metrics concerning the human scores of tribal communities (personalized scores) and non-tribal communities (population). We have the following inferences from the analysis based on overall observations.
2) BLEU and ChrF++ are string-based metrics that measure correlations among different H2E, G2H, Gondi-Bengali, and English-Tamil words. It is observed that these techniques do not account for the good translations from the MT systems It has been noticed that these methods fail to consider the accurate translations produced by the MT systems, and and depicted a correlation score of 0.2697 (E2T), 0.2753(H2E), 0.2956(G2H), 0.295(Gondi-Urdu) and 0.2695 (Gondi-Bengali). Similarly, for the transcribed from English-Hindi as testing the performance of the proposed framework, the correlation of these metrics is 0.2668 and 0.5894, respectively, using the BLUE method. For E2G, it is reported at 0.0596 and 0.4895, respectively.
3) Based on the overall evaluation (shown in Fig. 10), the proposed framework provides accuracy for the classification (personalized and Generalized) of spoken Gondi and Hindi languages (spoken by tribal people). It is reported that the type of words spoken by tribal people (personalized) provides strong statistical correlations with those of non-tribal people (generalized). It has been observed that the proposed framework yields higher accuracy for the classification of spoken words by tribal people.

3.2.5 Average Metric Analysis

Using the sentence-level score-matching technique, we measured the average similarity matrices based on the evaluation. Moreover, we used BERTScore and SBSim metrics, measured using multiBERT and paraBERT models. The SacreBLEU method measures average BLEU scores for Gondi spoken words.The average score for each metric is calculated by taking the mean of the scores assigned to individual sentences, as shown in Table 4. The metrics Yisi-1, BERT-based Score, and SBSim are assessed using the multiBERT and paraBERT models. For the SacreBLEU method, average BLEU scores are computed for spoken Gondi words.

1) From Table 4, we observe that the BLEU score shows obtained minimum correlation values of 0.0595 and 0.043 to assist for both (E2G) and E2H/H2G MT systems, respectively. It becomes evident that there is a minimal correlation between the BLEU score and the human scores. Specifically, the correlations for the E2G and E2H/H2G MT systems are 0.596 and 0.4556, respectively.
2) The ChrF++ score depicts an improved correlation of 0.2788 for E2G and E2H MT with 0.6890 and 0.6644 respectively.
3) We used para-BERT, in place of the multi-BERT MT model to evaluate the Yisi-1 score-based matching scores. is observed to improve the correlation of the metric by a maximum of 0.4463% (0.4296 to 0.8571) and 0.1699% (0.9289 to 0.7569) for E2G and E2H model, respectively. When considering similarity-based metrics, it is noticed that employing para-BERT for evaluating the Yisi-1 score rather than a multi-BERT mT system results in an enhancement of the metric's correlation by up 0.8571, 0.7569, and 0.9289 for E2G systems (see Table 4).

Table 4: Relationships between human and metrics scores based on string similarity

Evaluation metrics	E2G		E2H	
	SM	CM	SM	CM
HM	0.6593	1.000	0.8994	1.000
BLEU	0.0595	0.043	0.2498	0.5780
BLEU+Multimodal	0.0595	0.4895	0.2668	0.5894
ChrF++	0.2788	0.6797	0.6890	0.6644
Yisi-1-multimodal	0.48946	0.8759	0.6893	0.8969
Yisi-1	0.4296	0.8571	0.7569	0.9289
BERTScore-multimodal	0.7590	0.8957	0.8946	0.9289
BERT Score-single	0.7798	0.8885	0.8189	0.9290
Proposed-multi	0.8990	0.9296	0.9378	0.9689
Proposed-para	0.7493	0.9998	0.9888	0.9899

Abbreviation: E-T MT= English to Tamil MT, H2E MT, G-H MT= G2H MT, SM=Score $_{metric}$, CM=CoRR metric, HS=Human scores

3.3 Comparison with existing methods

We compared the translation performance of our model with existing methods. We used existing frameworks or models such as Transformer, LSTM, and GRU to compare the performance of the proposed framework (shown in Table 5). The performance of the proposed framework is evaluated on the prepared Hindi and Gondi language datasets in different benchmark settings. The semantic information is more accurately expressed in texts for long sequence spoken Gondi languages or egocentric videos. Hence, natural language description-based evaluation of language summaries is used in the evaluation measure. We converted the predicted summary to a Hindi text using the text description provided for the prepared Gondi language datasets and then utilized the BLEU score. For a fair comparison, all models were trained on the same dataset and evaluated using the same metric: the BLEU score. In addition to the BLEU scores, we also present the number of training epochs and parameters of each model to provide insight into the complexity and computational cost of each model. We used machine-learning models to classify gondi-Hindi languages. These models are the M1-support vector machine classifier, M2-decision tree, M3-incremental learning methods, M4-LSTM model, and M5-customized CNN. Figure 9 shows the average accuracy for Hindi-Gondi language classification using different models (M1-M5). Based on overall observations, the model M5 depicts higher accuracy than other models on a 5-fold cross-validation setting.

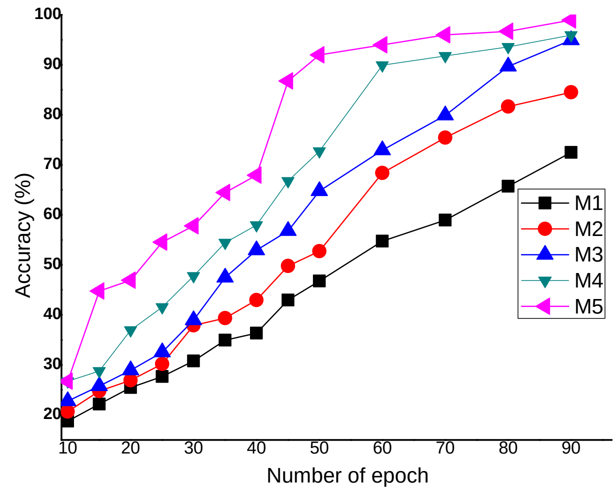

Figure 9: Average accuracy for Hindi-Gondi language classification using different models (M1-M5).

Table 5: Comparison of different models

Model	Epochs	Parameters	BLEU Score
Transformer	50	85M	0.70
LSTM	30	60M	0.62
GRU	30	55M	0.79
Ours	40	80M	0.98

The proposed framework employed the transformer model, known for its self-attention mechanism, which obtained a high BLEU score, indicating its effective translation performance in different settings. However, it also had many parameters [8], reflecting its complexity. The performance of the proposed framework is compared with the LSTM model [11], a type of recurrent neural network based on customized parameters that require fewer training epochs. Still, it shows a slightly lower BLEU score than other models. Similarly, the performance of the proposed framework is compared with the Gated Recurrent Unit (GRU) model, which signifies that the GRU model takes the lower number of epochs with 0.59 black scores. Moreover, another type of recurrent neural network exhibits a trade-off between complexity and performance [14] [15] [17]. The proposed framework integrates the strengths of the other models and achieves the highest BLEU score, demonstrating its superior performance in the Hindi-Gondi translation task. Notably, it reached this high performance with a moderate number of parameters and training epochs, reflecting its efficiency [18] [19].

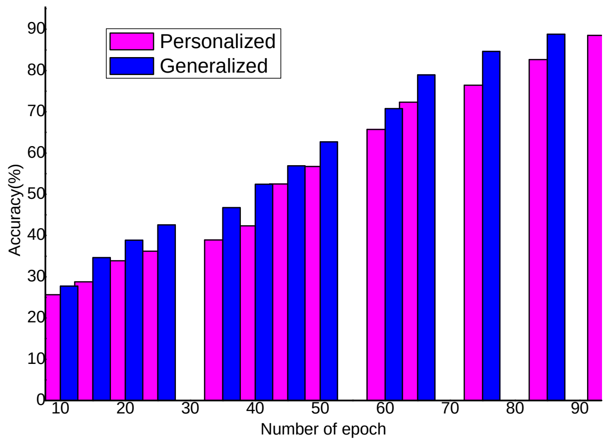

Figure 10: Accuracy based on personally spoken Gondi word

4 Case study

In this case study, we created a groundbreaking Android-based application for tribal committees. We developed a working prototype model aimed at the tribal communities in Chhattisgarh. The primary objective of the use case is to collect a massive amount of Gondi language samples and expand access to information in Gondi through the building of linguistic resources that can be appropriately utilized by the tribal community, such as the creation of the Gondi language dictionary, the development of children's stories, and educating tribal people to learn other languages. Moreover, this application inspires students who want to pursue higher studies, strengthens primary education, makes them more learners with other skills, and leverages them for further learning capabilities.

4.1 Mobilizing Community Engagement

The cornerstone of this paper was community engagement in the translation process of the Gondi language. Therefore, we developed an Android application, *"Gondi Gaatha"* as a platform for the participants to provide Gondi translations for given Hindi words or sentences. The innovation resolves the absence of available Gondi language resources and offers an enriching experience for the tribal communities to contribute to the linguistic heritage. In pursuing advancing language technologies for low-resource languages, we simultaneously aim to fulfil (i) collecting valuable Gondi language data and (ii) expanding access to information in the native language. As the community engaged in the translation process, we managed to amass a trove of Gondi language data that could fuel the development of advanced language technologies, such as speech-to-text (STT) and MT tools. Concurrently, the community's achievement in amplifying the accessibility to information in Gondi served as a milestone in their success.

4.2 Designing and Deploying the Language Translation

A key concern of this use case is to design a user-friendly and intuitive application interface for language translation in tribal communities, given that the target Gondi speakers and audience have varying spoken Gondi languages. Hence, the prepared language translation-based application provides familiarity to the tribal communities with the accessibility of advanced technologies. We thus meticulously curated *"Gondi Gaatha"* as popular and motivational stories for young youth of tribal communities to grab better opportunities for society upliftment and be effortlessly navigable and understandable in another language for all users. The proposed framework enables the application's user interface to encapsulate four innovative features to enrich the Gondi people's and other users' experiences. *Gondi Granth (epic)* can be built by making it a collection of Gondi language works of literature that may contain huge collections of Gondi-translated literature. *"Gondi Shabdkosh"* will provide a collection of the distinct meanings of comprehensive dictionary words of Gondi translations that leverage the Gondi people to learn about communications in other languages for knowledge transfer. Moreover, *"Gondi Geetmala"* has been created by the proposed framework as a collection of well-known Gondi quotes and inspirational stories. Each feature, designed with meticulous attention to detail, offers a unique interactive avenue for Gondi users. We have incorporated a working android prototype to cater to offline language translation functionality into *"Gondi Gaatha"* for tribal communities. The interactive application lets end users download the language models and continue their translation tasks without an active internet connection. Models are deployed in Android architecture to reduce latency, provide high computation based on collected raw gondi datasets for training the models, and make *"Gondi Gaatha"* a truly accessible tool for all tribal community members.

Figure 11: Translate user interface (UI)

4.2.1 *To mitigate the demographic's potential unfamiliarity:*

This case study part depicts this research work's innovative design, development, and deployment elements to leverage the technological platforms for mitigating the demographic potential unfamiliarity among people. It outlines the steps taken to ensure the application's effectiveness and accessibility for the tribal communities of Chhattisgarh, India. Given the target demographic's potential unfamiliarity with the developed Gondi-Hindi translation model's solutions and advanced technology interfaces, creating an intuitive interface for fast access to the Gondi languages and mitigating the language barrier for different tribal populations.

4.2.2 *Socio-economic development*

The use case outcome could enable the governance of the Ministry of Tribal Affairs (MoTA) India to provide a more focused approach toward the integrated socio-economic development of the scheduled tribes and tribal welfare for sustainable development. Furthermore, it also enhances their capability in their planning, project formulation for tribal people, research, evaluation, statistics, and training of tribal people. The proposed solution is deployed even in tribal regions at the early stages of their economic needs with a bare minimum income (cash) for sustainable development of tribal in dense tribal areas. It takes immediate quantitative analysis using mathematical, statistical, or computational techniques to mitigate poverty levels and save the lives of tribal people and other Indian citizens by learning a different skill and seeking the best opportunity across the world.

5 Conclusion and Future Direction

The proposed framework this paper has introduced a ground-breaking effort that uses technology to conserve linguistic history, improve community involvement, and remove communication challenges. Driven by the transformer model from "Gondi Gaatha," the proposed framework highlighted the persistence of linguistic variation by demonstrating technical proficiency in understanding and generating human languages. The paper shows how technology revolutionizes language preservation while promoting creativity, diversity, and culture. Empowering linguistic diversity initiatives to promote social inclusion, enhance communication, and improve language technology highlighted a dedication to recognizing and symbiotically integrating linguistic diversity in a world where communication is worldwide. Fundamentally, community involvement is essential because tribal people actively protect their language, add to an extensive database of Gondi language information, and promote the development of language technology. The framework ensured that language empowerment crosses geographic boundaries by addressing the particular issues faced by tribal groups via its offline capabilities and user-friendly design. DL approaches, and transformer architectures are coming together to focus on better translation for low-resource languages, which is expected to usher in a new era of cross-cultural understanding and communication. In the future, we plan to develop hypotheses to support the primary education system in tribal community languages and enlarge the dataset size via intensive engagement with tribal communities for model training. This tendency proposes a future characterized by significant cross-cultural interactions and heightened language inclusiveness worldwide.

References

[1] Mehta, D., Diddee, H., Saxena, A., Shukla, A., Santy, S., Mothilal, R.K., Srivastava, B.M.L., Sharma, A., Prasad, V. and Bali, K., 2022. Learnings from Technological Interventions in a Low Resource Language: Enhancing Information Access in Gondi. arXiv preprint arXiv:2211.16172.

[2] United Nations Office of Legal Affairs. *Convention on the Protection and Promotion of the Diversity of Cultural Expressions - Main Page.* UN.org, 2005. Available at: https://legal.un.org/avl/ha/cppdce/cppdce.html

[3] Z. Man et al., "WDSRL: Multi-Domain Neural Machine Translation With Word-Level Domain-Sensitive Representation Learning," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 32, pp. 577-590, 2024, doi: 10.1109/TASLP.2023.3336515.

[4] R. Lv, J. Guo, R. Wang, X. Tan, T. Qin and Q. Liu, "N-Gram Nearest Neighbor Machine Translation," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, doi: 10.1109/TASLP.2024.3407537.

[5] S. Jiang, Q. Chen, Y. Xiang, Y. Pan and X. Wu, "BaSFormer: A Balanced Sparsity Regularized Attention Network for Transformer," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 32, pp. 2125-2140, 2024, doi: 10.1109/TASLP.2024.3374062.

[6] V. M. Sánchez-Cartagena, M. Esplà-Gomis, J. A. Pérez-Ortiz and F. Sánchez-Martínez, "Non-Fluent Synthetic Target-Language Data Improve Neural Machine Translation," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 46, no. 2, pp. 837-850, Feb. 2024, doi: 10.1109/TPAMI.2023.3333949.

[7] S. Saini and V. Sahula, "A Survey of Machine Translation Techniques and Systems for Indian Languages," 2015 IEEE International Conference on Computational Intelligence & Communication Technology, Ghaziabad, India, 2015, pp. 676-681, doi: 10.1109/CICT.2015.123.

[8] T. Saeki, S. Maiti, X. Li, S. Watanabe, S. Takamichi and H. Saruwatari, "Text-Inductive Graphone-Based Language Adaptation for Low-Resource Speech Synthesis," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 32, pp. 1829-1844, 2024, doi: 10.1109/TASLP.2024.3369537

[9] Chimirala, U.M., 2020. Towards a convivial tool for narrative assessment: Adapting MAIN to Gondi (Dantewada, India), Halbi and Hindi for Gondi-and Halbi-Hindi speaking bilinguals. ZAS Papers in Linguistics, 64, pp.77-99.

[10] India – Census of India 2011 – Language Atlas – India". censusindia.gov.in.

[11] Zhixing Tan, Shuo Wang, Zonghan Yang, Gang Chen, Xuancheng Huang, Maosong Sun, & Yang Liu (2020). Neural machine translation: A review of methods, resources, and tools. AI Open, 1, 5-21.

[12] Pune, India, 2023, pp. 1-5. Chaubey, G., Tamang, R., Pennarun, E., Dubey, P., Rai, N., Upadhyay, R.K., Meena, R.P., Patel, J.R., Van Driem, G., Thangaraj, K. and Metspalu, M., 2017. Reconstructing the population history of the largest tribe of India: the Dravidian speaking Gond. European journal of human genetics, 25(4), pp.493-498.

[13] A. Baliyan, A. Batra and S. P. Singh, "Multilingual Sentiment Analysis using RNN-LSTM and Neural Machine Translation," 2021 8th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2021, pp. 710-713.

[14] A. Pramodya, R. Pushpananda and R. Weerasinghe, "A Comparison of Transformer, Recurrent Neural Networks and SMT in Tamil to Sinhala MT," 2020 20th International Conference on Advances in ICT for Emerging Regions (ICTer), Colombo, Sri Lanka, 2020, pp. 155-160, doi: 10.1109/ICTer51097.2020.9325431.

[15] T. J. Sefara, S. G. Zwane, N. Gama, H. Sibisi, P. N. Senoamadi and V. Marivate, "Transformer-based Machine Translation for Low-resourced Languages embedded with Language Identification," 2021 Conference on Information Communications Technology and Society (ICTAS), Durban, South Africa, 2021, pp. 127-132, doi: 10.1109/ICTAS50802.2021.9394996.

[16] A. Jha, H. Y. Patil, S. K. Jindal and S. M. N. Islam, "Multilingual Indian Language Neural Machine Translation System Using mT5 Transformer," 2023 2nd International Conference on Paradigm Shifts in Communications Embedded Systems, Machine Learning and Signal Processing (PCEMS), Nagpur, India, 2023, pp. 1-5, doi: 10.1109/PCEMS58491.2023.10136051.

[17] S. Duan, H. Zhao, J. Zhou and R. Wang, "Syntax-aware Transformer Encoder for Neural Machine Translation," 2019 International Conference on Asian Language Processing (IALP), Shanghai, China, 2019, pp. 396-401, doi: 10.1109/IALP48816.2019.9037672.

[18] Mehta, D., Santy, S., Mothilal, R., Srivastava, B., Sharma, A., Shukla, A., Prasad, V., U, V., Sharma, A., & Bali, K. (2020).

Learnings from Technological Interventions in a Low Resource Language: A Case-Study on Gondi. In Proceedings of the Twelfth Language Resources and Evaluation Conference (pp. 2832–2838). European Language Resources Association.
[19] Rama, T., Çoltekin, ., & Sofroniev, P. (2017). Computational analysis of Gondi dialects. In Proceedings of the Fourth Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial) (pp. 26–35). Association for Computational Linguistics.
[20] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., & Polosukhin, I. (2017). Attention is All You Need. In Advances in Neural Information Processing Systems (pp. 5998-6008). Available at: https://arxiv.org/abs/1706.03762
[21] W. Wang and Y. Qian, "Universal Cross-Lingual Data Generation for Low Resource ASR," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 32, pp. 973-983, 2024, doi: 10.1109/TASLP.2023.3345150.
[22] Austin, Peter K., and Julia Sallabank. The Cambridge Handbook of Endangered Languages. Cambridge University Press, 2011.