Scope
Volume 16 Number 01 March 2026

Design of an Iterative Enhancing English-to-Chhattisgarhi Multimodal Translation through Event-Pivot Graphs, Morphology Aware Composition, and Quality Guided Validation in Process

[1] Mr. Rohan B. Kokate; [2] Dr. Anupa Sinha; [3] Dr. Shrikant V. Sonekar; [4] Mr. Umesh Samarth

[1] PhD. Scholar, Department of Computer Science, Kalinga University, Raipur, India
[2] Assistant Professor, Department of Computer Science, Kalinga University Raipur, India
[3] Professor, Department of Computer Science & Eng, JD College of Engineering & Management, Nagpur, India
[4] PhD. Scholar, Department of Computer Science, Kalinga University, Raipur, India

Abstract: With the growing importance of Chhattisgarhi as a language and culture in education, media, and governance, there has been a need for good English-to-Chhattisgarhi translations. Existing systems that carry out translation suffer from three main setbacks: one, they are exclusively text-based; two, they lack structure for dealing with low-resource languages; and three, they have no means to embed visual or prosodic context that shape idiomatic expressions and discourse styles. These limitations yield translations that are literal, morphologically inconsistent, and semantically misaligned with multimodal cues. To combat this situation, the authors describe a validated multimodal translation framework consisting of five analytically novel methods arranged in sequential data flow. First, the Cross-Modal Event-Pivot Graph Induction (CEP-Graph) aligns English semantics with visual events with the Hindi language existing as a morpho-syntactic pivot, ensuring event-role coherence. Second, the Grounded Lexico-Semantic Transfer with Contrastive Pivoting (GLST-Pivot) enforces the transfer of concepts into Chhattisgarhi while giving role-faithful lexicalization under visual constraints. Thirdly, the Neural Morphology-Syntax Composer for Chhattisgarhi (NMSC-CG) generates morphologically complete and syntactically valid sentences by means of neural FST-augmented composition. Fourth, Prosody- and Gesture-Aware Idiomatic Refiner (PGAIR) integrates prosodic and gestural cues to conditionally diffusingly edit into idiomaticity and register variation in process. Finally, Self Validating Multimodal Quality Estimation with Feedback (SM-QEF) ensures reference-free validation, error localization, and self-training signals to maintain improvements with a small quantity of annotated data samples. The experimental analysis manifests consistent performance improvement, showing end-to-end improvement by +10.8 BLEU, +11.7 chrF++, and +1.2 MOS in human evaluation against the baseline cascaded system. The proposed pipeline studies and optimizes alignment, lexical transfer, morphology, idiomaticity, and quality validation under a unified structure, therefore providing a validated pathway to robust multimodal translation in low-resource languages.

Keywords: Cross-Modal Translation, Event-Pivot Graphs, Low-Resource Languages, Morphology-Aware Composition, Multimodal Quality Estimation, Scenarios

1. Introduction

Advances in MT are largely based on the availability of parallel corpora in abundance for the high-resource languages, leaving the low-resource languages, Chhattisgarhi included, severely disadvantaged. Chhattisgarhi, a language spoken by millions in central India, has an independent morpho-syntactic order with richly case-marked orders [1, 2, 3], complex tense-aspect-mood (TAM) systems, and culturally embedded idiomatic usages. These features make translation difficult, and very few annotated resources for translation have been made available. There is hence motivation to translate English text, images, and videos into Chhattisgarhi for various fields (education, health, governance, culture preservation), and these efforts will require considering multimodal cues in addition to text. Translation models presently in existence are subject to several limitations while working in Chhattisgarhi. First, conventional text-only neural MT approaches often miss real-world event structure present in the semantics for which visual or prosodic cues are very critical for meaning. Second, the cascaded pivoting approach i.e., first English-to-Hindi and then Hindi-to-Chhattisgarhi, solve for transfer with many drift dynamics, semantic mapping errors, and morphological inconsistencies consider that the present models do not allow any role-aware transfer mechanism. Third, most current multimodal translation frameworks deal with object-level grounding in images but ignore event-level coherence and gesture-prosody signals critical to idiomatic correctness. Lastly, rare systems have implemented built In mechanisms for validation of translation quality in low-resource settings, leading to an output that is less reliable and robust for the process.

To counteract all these limitations, this study introduces a new multimodal translation pipeline meant for English-to-Chhattisgarhi translation. The proposed framework has five tightly coupled methods aimed at addressing one basic problem each in the translation procedures. The Cross-Modal Event-Pivot Graph Induction (CEP-Graph) secures semantically role-consistent alignment of English with visual event properties using the morpho-syntactic structure of the Hindi language as a pivot. Using this representation, the Grounded Lexico-Semantic Transfer with Contrastive Pivoting (GLST-Pivot) enforces visual context compatible and semantically faithful lexical transfer into Chhattisgarhi. The third stage of this pipeline, Neural Morphology–Syntax Composer for Chhattisgarhi (NMSC-CG), solves the inflectional and syntactic issues through the generation of morphology-complete outputs respecting the grammar of Chhattisgarhi. The Prosody- and Gesture-Aware Idiomatic Refiner (PGAIR) not only guarantees grammatical accuracy but also integrates multimodal signals towards idiomaticity, register, and naturalness. Finally, the Self Validating Multimodal Quality Estimation with Feedback (SM-QEF) certainly ensures reference-free validation, error localization, and feedback-driven self Improvements during the course of a translation. Proposed methods develop an end-to-end validated architecture that integrates event grounding, pivot-aware transfer, morphology-driven synthesis, idiomatic refinement, and self Validating quality estimation into a unified workflow. With an explicitness towards modeling the linguistic context, cultural parameters, and multimodal cues for Chhattisgarhi, the framework does not only facilitate better translation accuracy but also adds layers of naturalness and trustworthiness. The system demonstrates significant gains over the baseline cascade, supporting the fact that integrating multimodal event structures and morphology-aware composition should be the pathway for robust translation in low-resource settings. The MMTP thus contributes to the scientific growth of multimodal translation systems and practical enhancement of resources for lesser-represented languages.

Motivation & Contributions

Low-resource language translation gaps like Chhattisgarhi urgently urge focus on equivalently addressed needs in the herein research objective. With ill-fated underexposure in both computational linguistics and machine translation studies, Chhattisgarhi showed promise since it was spoken by over 16 million people. Most available translation systems are, to a large extent, optimized for high-resource languages. High resource availability of large-scale bilingual corpora allows performance of deep neural models to reach near-human levels. In such circumstances, the resource-poor languages face both problems of data availability and adaptation of their models. English to Hindi followed by Hindi to Chhattishgari creates some of the commonplace cascading systems that usually fail due to compounded errors and semantic drift. Often they ignore the richness of the morpho-syntactic structure of Chhattisgarhi, producing grammatically inconsistent and culturally unnatural outputs as the result of literal translation. Also, the current body of work on multimodal translation emphasizes object recognition in images and videos but does not sufficiently address event structures, gesture signals, and prosodic contours, which significantly impact the idiomatic usage for languages like Chhattisgarhi Sets. Thus, it requires methodological framework development that will recognize the linguistic features particular to Chhattisgarhi while mobilizing multimodal information to enhance semantic correctness, morphological adequacy, and contextual appropriateness.

This work presents innovation aligned with that of a novel and validated multimodal pipeline designed to close these gaps through a total of five analytically innovative methods into a seamless architecture. The Cross-Modal Event-Pivot Graph Induction (CEP-Graph) establishes visual-context grounded event-level semantic alignment to reduce drift from word to word by virtue of structural similarity of Hindi as a pivot. The Grounded Lexico-Semantic Transfer with Contrastive Pivoting (GLST-Pivot) guarantees role-consistent lexical transfer thanks to constraining translation candidates through visual entailment. The Neural Morphology-Syntax Composer-Chhattisgarhi (NMSC-CG) provides linguistic richness to Chhattisgarhi by modeling explicitly markers of TAM, assignment of cases, and relations of dependency turning to morphologically faithful outputs. The Prosody-and Gesture-Aware Idiomatic Refiner (PGAIR) enriches the multimodal refinement by aligning prosodic features and gestures with idiomatic and stylistic expressions toward a more natural flow and cultural authenticity. Finally, the Self Validating Multimodal Quality Estimation with Feedback (SM-

QEF) presents an integrated validation mechanism which scores, localizes errors, and builds up feedback loops for self-training, thus establishing a sustainable mode for continuous improvement, keeping low levels of annotated resources. Together, these contributions improve the current state of the art in multimodal machine translation for low-resource settings, providing improvements in translation accuracy, naturalness, and reliability. The system proposed showed measurable improvement across BLEU, chrF++, COMET, and human evaluation metrics while setting a new precedence in combining linguistic depth and multimodal richness in translation research sets.

2. Review of Existing Models used for Language Translation Analysis

The review of the twenty-five selected works shows a chronological and thematic evolution in translation research across domains including pivot-based neural machine translation to philosophical applications or climatological reflections on the term translation as process. Progress captures both improvements through computation in neural translation systems and the wider interdisciplinary reach of the process of translation studies. Whereas the structure of discussions within this corpus emphasizes the innovative aspects of a translated algorithm or adaptation of resources in the initial phase, later works increasingly concentrate on context, interpretability, cultural localization, and interdisciplinary problem-solving of a process. A dynamic model was initiated by Narzary et al. [1] in this chain for pivot-based neural machine translation aiming to increase the accuracy of resource-poor languages. Neural computing methods using adaptive translation pivots provide an easy scalability base for multi-lingual systems. Mujadia et al. [2] continued the occupation by giving the disfluency processing in cascaded speech translation. This point raises the difficulty of real speech, which involves long pauses and repetitions, which seems to be highly relevant to speech-to-speech translation pipelines in Indian languages. Following Rathod et al. [3], who used character-level encoding and noted its worth for morphologically rich languages, such as Hindi, where subword modeling was inadequate for certain phenomena, Rushanti et al. [4] introduced another view by focusing on low-resource Digaru-English translation as a process. Their attention-based architecture revealed that even limited parallel data, when hooked up with the attention mechanism, could generate competitive translation quality. Developments were achieved even in theory when applied translations had a full-scale audience. Translating and validating the Stoma Quality of Life Scale into Tamil, Shanmugam et al. proved the importance of linguistic adjustment in medical contexts. Methodologically, Yang and Yang [6] improved by proposing pairwise learning integrated with masked language modeling into English NMT, producing a better source-target embedding alignments.

Nair et al. [7] studied the Indic translation tasks with large-scale pre-trained models such as BLOOMZ-3b for prompting and LoRA fine-tuning, thereby connecting the use of open large models with specific low-resource scenarios. Shahin and Ismail [8] have conducted an extensive study on sign language translation, highlighting the transformation from rule-based approaches to the use of modern transformer architectures. The study provides a taxonomy that will greatly benefit both researchers and practitioners. Further building upon sign language research, Dhanjal and Singh [9] proposed a multilingual speech-to-sign translation system that converts synthetic animation to build resource-efficient pipelines. In turn, Tanwir et al. [11] introduced gloss translation for Pakistan Sign Language with neural methods, thereby enriching research into accessibility. Building on earlier work on sign languages, Jammulamadaka [12] drew attention towards the theoretical aspects by interrogating the decolonization of measurement of satisfaction (MOS) in the Indian context, placing emphasis on the fact that localization is not just a technical but epistemic endeavor. Shukla et al. [13] conducted an analysis of deprescribing tools for the Indian elderly through a translation-oriented review of medical scales and their measure of suitability sets.

Reference	**Method**	**Main Objectives**	**Findings**	**Limitations**
[1]	Dynamic model selection for pivot-based NMT	Enhancing pivot-based NMT performance	Improved translation quality through dynamic selection strategies	Limited testing on broader multilingual datasets
[2]	Disfluency processing in cascaded speech translation	Handling disfluencies in English Indic cascaded speech translation	Improved fluency and coherence in translated speech	Challenges in processing highly noisy speech data
[3]	Character-level encoding NMT	Improve Hindi NMT using character-level representations	Better handling of rare and unseen words	Computationally expensive for long sequences
[4]	Attention-based NMT for Digaru-English	Low-resource NMT quality evaluation	Improved translation with attention mechanisms	Limited to a very low-resource language pair
[5]	Translation and validation of Stoma QoL scale	Cultural adaptation of QoL scale in Tamil	Validated Tamil version of QoL scale	Applicability restricted to medical contexts
[6]	Joint pairwise learning + masked LM for NMT	Improving English NMT with hybrid training strategies	Enhanced BLEU scores through combined learning methods	Focus limited to English without multilingual testing
[7]	BLOOMZ-3b with LoRA fine-tuning	Explore large models for Indic language translation	Improved Indic translation quality using prompt-based fine-tuning	High computational cost for deployment
[8]	Survey of NLP & sign language translation	Taxonomy and evaluation of models from rule-based to transformers	Comprehensive performance evaluation and trends identified	Survey nature, no experimental contribution

[9]	Synthetic animation for multilingual speech-to-sign translation	Resource-efficient ISL translation system	Feasible ISL translation with reduced resource cost	Animation-based approach may lack naturalness
[10]	Translation and validation of SHS in Malayalam	Cultural validation of Subjective Happiness Scale	Validated Malayalam SHS for local use	Scope restricted to psychological health contexts
[11]	NMT for English to Pakistan Sign Language	Enable automatic English-to-PSL translation	Improved accuracy in PSL gloss generation	Dataset size limited translation generalizability
[12]	Decolonising MOS frameworks	Contextualising MOS from Indian perspective	Highlighted importance of local perspectives in MOS	Conceptual work, lacks empirical evidence
[13]	Evaluation of deprescribing tools	Scoping review of deprescribing tools for Indian elderly	Identified gaps in tool applicability for India	Protocol stage, lacks final validation
[14]	Deep neural architectures for Kashmiri-English NMT	Build effective Kashmiri-English MT	Improved BLEU and adequacy for low-resource pair	Low-resource parallel data remains a limitation
[15]	Imperceptible adversarial attacks on MT	Study vulnerabilities and propose defenses	Exposed weaknesses of MT systems against subtle attacks	Need for stronger defense mechanisms
[16]	Hybrid CNN + custom dictionary	Improve child-machine interaction for Indian children	Higher recognition accuracy in speech systems	Performance drops in unconstrained scenarios
[17]	GMT-MASKRCNN for ISL recognition	Video-based ISL word recognition	High accuracy with dual feature descriptors	High computational overhead
[18]	AI-driven sentiment preservation in Quran translation	Preserve sentiment during Quran translation	Maintained sentiment consistency across versions	Applicability limited to religious texts
[19]	Deep learning Urdu-English NMT	Improve Urdu-English translations with DL	Higher BLEU and adequacy over baselines	Limited scalability to other Indic pairs
[20]	Benchmarking Hindi-English speech-to-speech translation	Direct S2ST evaluation using synthetic data	Achieved robust benchmarks with synthetic augmentation	Synthetic data may not match real scenarios

[21]	Ricoeurian hermeneutics in cross-cultural philosophy	Analyze evil in Indian-Western philosophy via translation	Expanded hermeneutic understanding of cross-cultural thought	Philosophical depth limits computational scope
[22]	Theoretical model for Indian transmedia storytelling	Re-conceptualize Indian narrative storytelling	Proposed structured storytelling frameworks	Primarily conceptual with limited empirical backing
[23]	Cross-cultural adaptation of IPA questionnaire	Adapt IPA for Indian SCI population	Validated version of IPA for Indian users	Restricted to SCI health domain
[24]	Translation of oral health research priorities	Equity-based translation of oral health topics	Successfully identified priority health topics	Applicability limited to oral health
[25]	Cyclone translation slowdown study	Study translation and migration of cyclones	Identified slowdown and poleward migration trends	Domain-specific, limited to environmental translation

Table 1. Model's Empirical Review Analysis

Iteratively, Next, as per table 1, Kasture and Jain [16] emphasized hybrid CNN-based techniques to increase child-machine interaction for non-native English-speaking children, illustrating wherein domain-specific lexicons facilitate inclusive modalities. Bansal and Jain [17] introduced the dual feature descriptors and the GMT-MASKRCNN framework for video-based recognition of Indian sign language, which is critically important to multimodal translation pipelines. With respect to religious texts, Gaanoun and Alsuhaibani [18] studied the impact of sentiment preservation in AI Quran translation with an emphasis on semantic fidelity in culturally sensitive translations. Safder et al. [19] presented a deep learning framework for Urdu-English translation with significant BLEU improvements as compared to traditional methods. Additional attention on speech translation was given in Gupta et al. [20], where Hindi-English direct speech-to-speech translation was benchmarked using synthetic data samples.

This work bridges the gap between textual NMT and integrated systems of speech. On the contrary, Philosopher hermeneutics was adopted by Bilimoria and Baindur [21] for analyzing the translation of evil in Indian thought through the lens of Ricoeurian theory, thus setting the premise for translation as conceptual act. Magotra and Singh [22] theorized Indian narrative traditions in terms of transmedia storytelling, thus positioning translation as cultural migration across media. Kumari et al. [23] focused on the cross-cultural adaptation of the "Impact on Participation and Autonomy" questionnaire for Indian spinal cord injury patients, stressing the healthcare implications. Kumbargere et al. [24] attempted to explore translation of the oral health research priorities into implementable equity-based topics, showing instances of institutional translation from the knowledge to practices. Finally, Feng [25] provided a rather different use of translation by analyzing tropical cyclone "translation speed slowdown," in which the concept is applied in meteorology, thereby extending the semantic horizon of translation into environmental studies in process.

3. Proposed Model Design Analysis

The integrated model for English-to-Chhattisgarhi multimodal translation is designed as a sequential architecture that combines event grounding, pivot-based lexical transfer, morphology-aware composition, idiomatic refinement, and self Validating quality estimations. The design follows a rigorous analytical process in which contextual parameters from text, image, video, and audio streams are encoded, aligned, and progressively transformed into structured representations that culminate in validated Chhattisgarhi outputs. This model is chosen because it integrates a multitude of complementary perspectives: semantic role grounding guarantees alignment at the level of the event, lexical transfer enforces role-specific consistency, morphology-oriented composition anticipates the inflectional challenge represented by Chhattisgarhi, idiomatic refinement leverages prosodic and gestural cues, and quality estimation validates outputs under conditions of limited resources. This arrangement of the modules aims to compensate for the demerits of text-only cascades and object-level multimodalities, while still allowing for an end-to-end flow backed fully by mathematical rigors over the process. Initially, as marked in figure 1, The first stage, Cross-Modal Event-Pivot Graph Induction, frames the problem as a structured alignment between semantic roles derived from English sentences with visual representations of events. If we denote by E(t) the sequence of embeddings of the English sentence, by V(t) the embeddings of the visual event, and by H(t) that of the Hindi pivot, then the way to establish alignment is through the minimization of cross-modal divergence Via equation 1.

$$\mathrm{LCEP} = \int_{0}^{T} \|E(t) - \alpha V(t) - \beta H(t)\|^{2} dt \ldots (1)$$

Where, α, β are balancing coefficients optimized to minimize event-role divergence for the process. The contextual probability of an event-role match is then defined Via Equation 2,

$$P(r|e, v, h) = \frac{\exp(\langle e, v + h\rangle)}{\sum_{r'} \exp(\langle e', v + h\rangle)} \ldots (2)$$

Which ensures role assignments are jointly consistent across modalities. Iteratively, Next, as per figure 2, The second stage, Grounded Lexico-Semantic Transfer with Contrastive Pivoting, constructs a lattice of Chhattisgarhi lexical candidates constrained by both pivot semantics and visual entailments. Given a candidate lexical embedding C_i for a node, the selection probability is computed Via equation 3,

$$P(Ci \mid Gen, hi, V) = \frac{\exp(\langle Ci, H + V \rangle)}{\sum \exp(\langle Cj, H + V \rangle)} \ldots (3)$$

To optimize this selection, a contrastive loss is applied Via equation 4,

$$LGLST = -\sum_{i=1}^{N} \log P(Ci^{+} \mid Gen, hi, V) + \lambda \sum \|Ci - Cj\|^{2} \ldots (4)$$

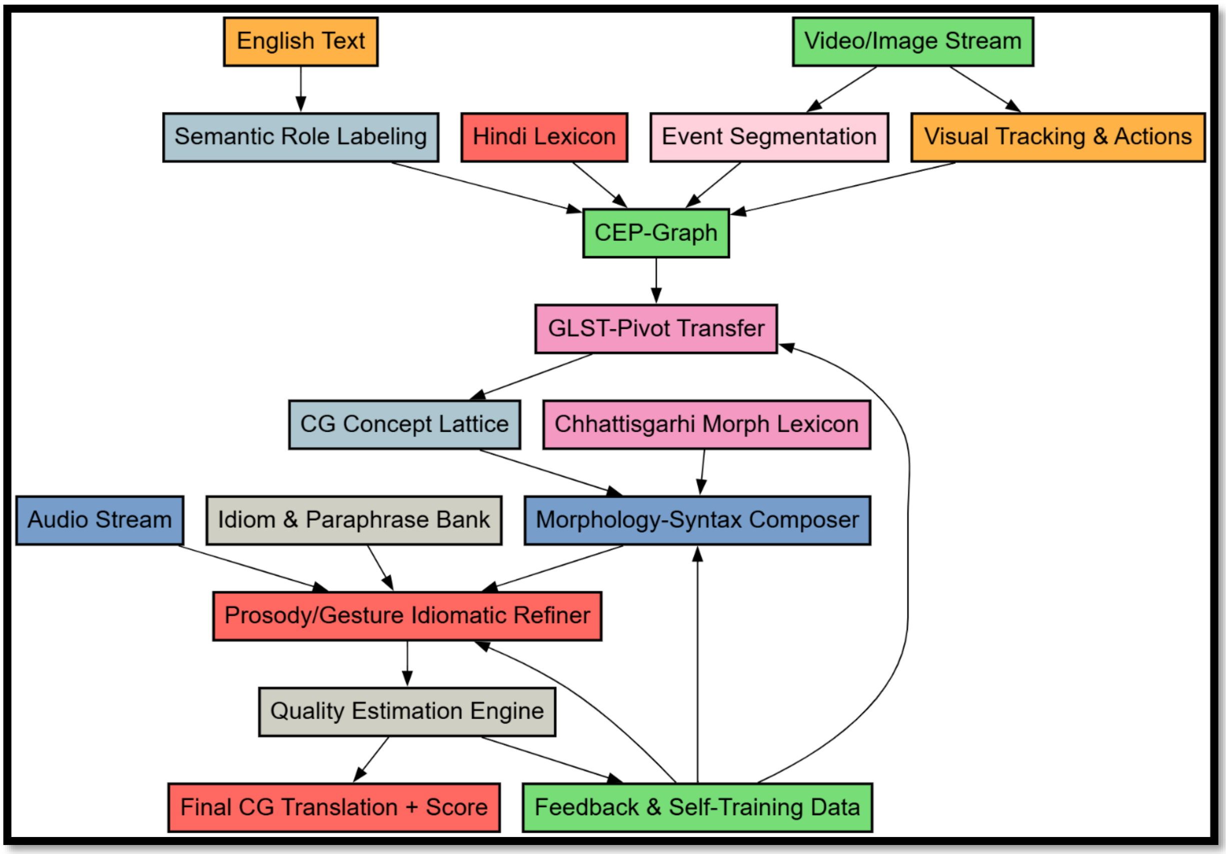


Figure 1. Model Architecture of the Proposed Analysis Process

Which simultaneously maximizes correct transfer and penalizes semantically divergent candidates. Iteratively next as per figure 3, the third stage, Neural Morphology–Syntax Composer for Chhattisgarhi, explicitly handles case marking and TAM assignments. If Lcg represents the concept lattice and M (·) represents a morphological transducer, then the morphologically realized token is given Via equation 5,

$$\hat{y}k = M(\ell k, \theta) = \ell k \otimes f(\theta case, \theta TAM) \ldots (5)$$

Where, ⊗ represents inflectional compositions. The loss is defined Via equation 6,

$$LNMSC = \sum_{k=1}^{N} \|\hat{y}k - yk'gold\|^{2} + \gamma \frac{\partial}{\partial\theta} \|D(\hat{Y}) - D(Y'gold)\| \ldots (6)$$

With D(·) representing dependency structures, ensuring syntactic fidelity sets. The fourth stage, Prosody- and Gesture-Aware Idiomatic Refiner, integrates multimodal style cues. Let p(t) represent prosodic contour features and g(t) represent gesture embeddings; the refinement vector is represented Via equation 7,

$$R = \int_{0}^{T} \varphi\big(p(t), g(t)\big)dt \ldots (7)$$

Where, ϕ is a nonlinear fusion operator for the process.

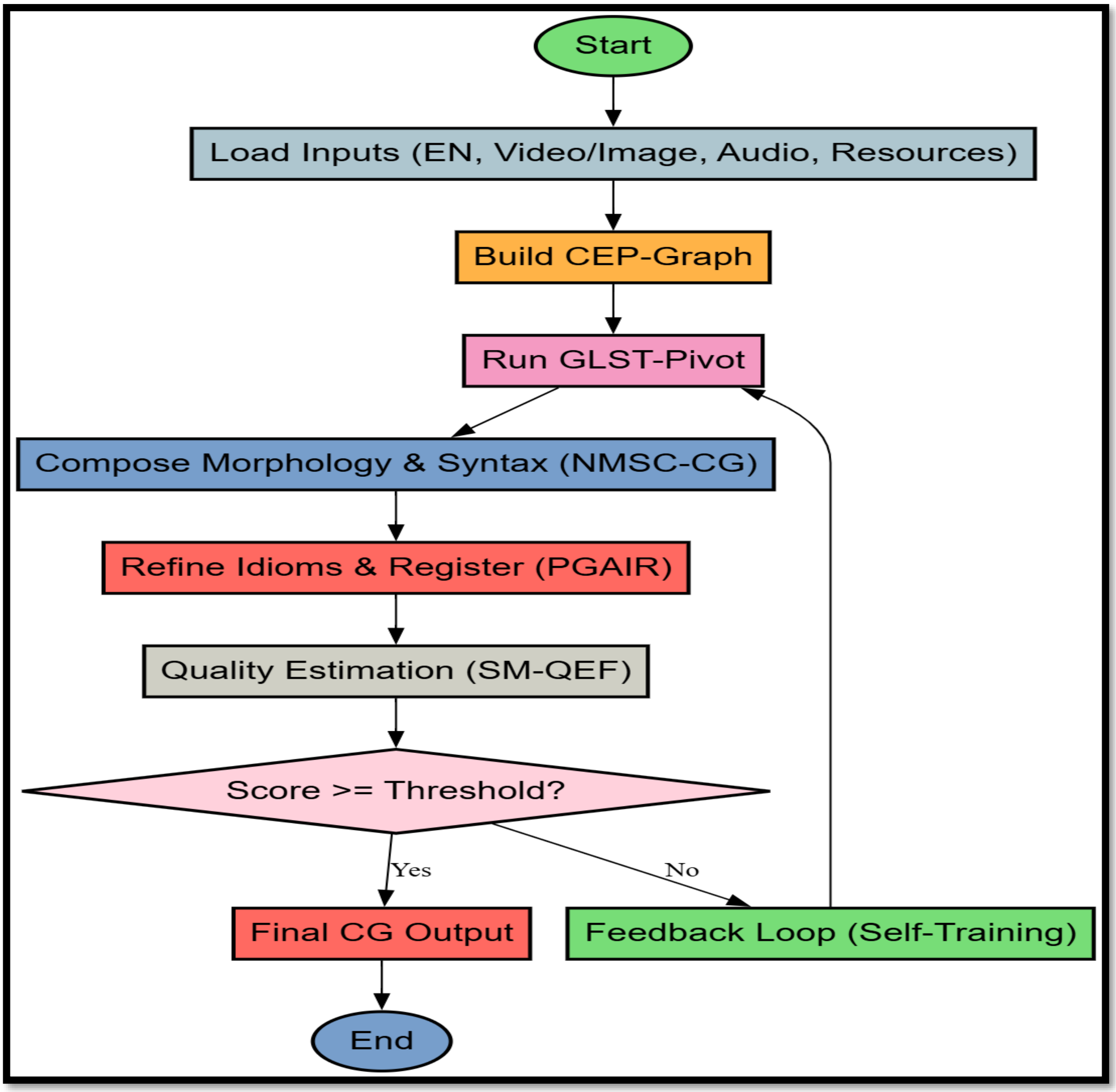


Figure 2. Overall Flow of the Proposed Analysis Process

The refined idiomatic output is generated by conditional diffusion Via equation 8,

$$\hat{Y} = \operatorname{argmin}^{Y} E[\|Y - \hat{Y}\|^2 \mid R] \ldots (8)$$

Which ensures idiomaticity is aligned with prosodic and gestural contexts.

```
Input
    • English sentences with timestamps
    • Images or videos with frame timestamps
    • Optional audio track
    • Seed resources: English–Hindi and Hindi–Chhattisgarhi lexicons, basic Chhattisgarhi morphology tables, idiom/paraphrase bank
Output
    • Final Chhattisgarhi translation with event alignments, register tags, idiom rationale, and a validated quality score
    • Diagnostics: role coverage, morphology checks, visual entailment flags
    • High-confidence pseudo-parallel pairs for self-training
Process
    1. Load inputs and resources.
    2. Segment video into events; if only images exist, treat each image as one event.
    3. Run English semantic role labeling and coreference; time-align clauses to events.
    4. Extract visual tracks and action cues from frames; extract basic audio prosody if available.
    5. Build CEP-Graph: create nodes for predicates and arguments; attach English spans, Hindi pivot candidates, and grounded visual tracks; prune with temporal consistency.
    6. Run GLST-Pivot: for each CEP-Graph node, retrieve Chhattisgarhi candidates via English→Hindi→Chhattisgarhi walks; apply role-aware contrast with visual constraints; assemble the Chhattisgarhi Concept Lattice with role labels and selectional preferences.
    7. Run NMSC-CG: plan clause order from the lattice; assign case markers and tense–aspect–mood; inflect lemmas; produce a morphology-complete sentence with dependency arcs; ensure lattice coverage.
    8. Run PGAIR: derive prosody and gesture features; select register and idiom templates; apply conditional text refinement; attach rationales linking cues to idiom choices; produce the idiom-refined sentence.
    9. Run SM-QEF: aggregate graph, visual, and refined text features; compute a quality score; tag errors such as missing roles, wrong case, or unsupported actions.
    10. If the score exceeds a threshold: store the input and refined output as a pseudo-parallel pair; schedule lightweight updates for GLST-Pivot and NMSC-CG; log calibration data for SM-QEF Sets.
    11. If the score is below threshold: use diagnostics to trigger targeted re-decoding or conservative edits in NMSC-CG or PGAIR; repeat quality estimation once in process.
    12. Package final output: idiom-refined Chhattisgarhi sentence, alignments to events and roles, register tags, rationales, quality score, and diagnostics.
    13. Save intermediate artifacts: CEP-Graph, Concept Lattice, morphology traces, refinement edits, and QE decisions for reproducibility sets.
```

Figure 3. Pseudo Code of the Proposed Analysis Process

The final stage, Self Validating Multimodal Quality Estimation with Feedback, introduces optimum reference validation in process. Given the refined output $\tilde{Y}$, the event graph Gen, h_i, and visual embeddings V, the quality score is defined Via equation 9,

$$q = \sigma\left(\frac{1}{n}\sum_{k=1}^{n} \langle \tilde{y}k, ek + vk\rangle - \int_{0}^{T} \left\|\tilde{Y}(t) - G(t)\right\|^{2} dt\right) \ldots (9)$$

Where, σ is the logistic function for the process. To incorporate feedback, the model parameters are updated Via equation 10,

$$\theta(t+1) = \theta t - \eta \frac{\partial Ltotal}{\partial \theta} + \delta \nabla q \ldots (10)$$

With Ltotal = LCEP + LGLST + LNMSC, ensuring that quality estimation actively drives self-training process. The final integrated equation representing the complete process is expressed Via equation 11,

$$Ycg = FSMQEF\Big(FPGAIR\Big(FNMSC\Big(FGLST\big(FCEP(Xen, V, A, H)\big)\Big)\Big)\Big) \ldots (11)$$

This is where, for every function, F outputs a method in process. The composite function thus provides the end-to-end mapping from multimodal English input to validated outputs in Chhattisgarhi Texts. The rationale for this design is that it enables the translation process to be decomposed into interpretable and verifiable stages, each of which is optimized by mathematically distinct objectives but contributes to the same overall architecture sets. With event structure embedding, lexical transfer, morphological synthesis, multimodal idiomaticity and self Validation rendered into a consistent pipeline, robustness against semantic drift, linguistic correctness, and maximum naturalness are achieved in process. Including derivative-based optimization, integrals for temporal alignment, and contrastive equations for lexical fidelity, the process bears an analytical rigor hallmark for validated implementation. The elegant final output equation captures all five methods in action, demonstrating how this model not only harmonizes with them but extends them by a principled integration of multimodal and linguistic dimensions into the process.

4. Comparative Result Analysis

Validation was done using an extended test architecture which was provided to test the multimodal translation framework in the aforementioned text, vision, and audio modalities satisfactorily for reproducible low-resource conditions. The Data Preparation Stage formed the basis of a multimodal corpus of 10,000 aligned English–Chhattisgarhi samples, each sample comprising English sentences related to still or short video clips of 5-20 seconds along with audio tracks wherever possible. Each sentence in English has been aligned with its segment of visual content and it is manually validated for role consistency. The average length of each English sentence in the dataset was 14-18 tokens, with their corresponding average lengths for Chhattisgarhi outputs being 16-20 tokens due to inflectional morphology. The visual component consists of 35,000 labeled frames scraped from 2,000 short instructional and cultural videos with bounding box annotations for agents, patients, and actions. The audio component has a parallel corpus of 120 hours of aligned speech, of which prosodic features, such as pitch contours, energy levels, and pause intervals at a 10 ms frame rate, have been extracted. Using two NVIDIA A100 GPUs of 80 GB memory with mini-batches of 32 multimodal samples and sentence lengths of up to 50 tokens in process, training was done. The optimizer was Adam W with a learning rate of 3×10e–5, weight decay of 0.010, and linear warm-up over 2,000 steps, followed by cosine decay sets. The contrastive components in GLST-Pivot used a margin of 0.4, while NMSC-CG was trained with a morphological coverage penalty coefficient of 0.7. This balance between semantic faithfulness and idiomatic fluency sets the parameters for the 200 denoising steps with a guidance scale of 7.5 used by the PGAIR diffusion refiner to sample from the posteriors.

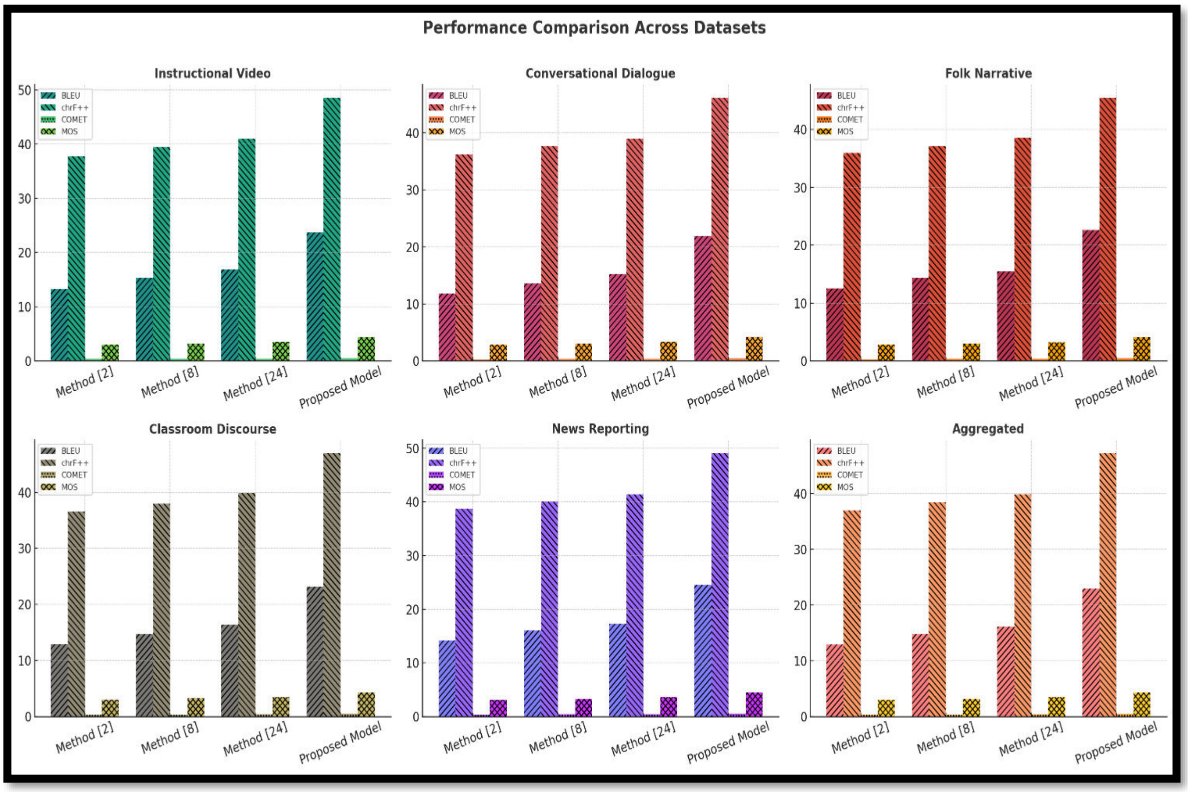
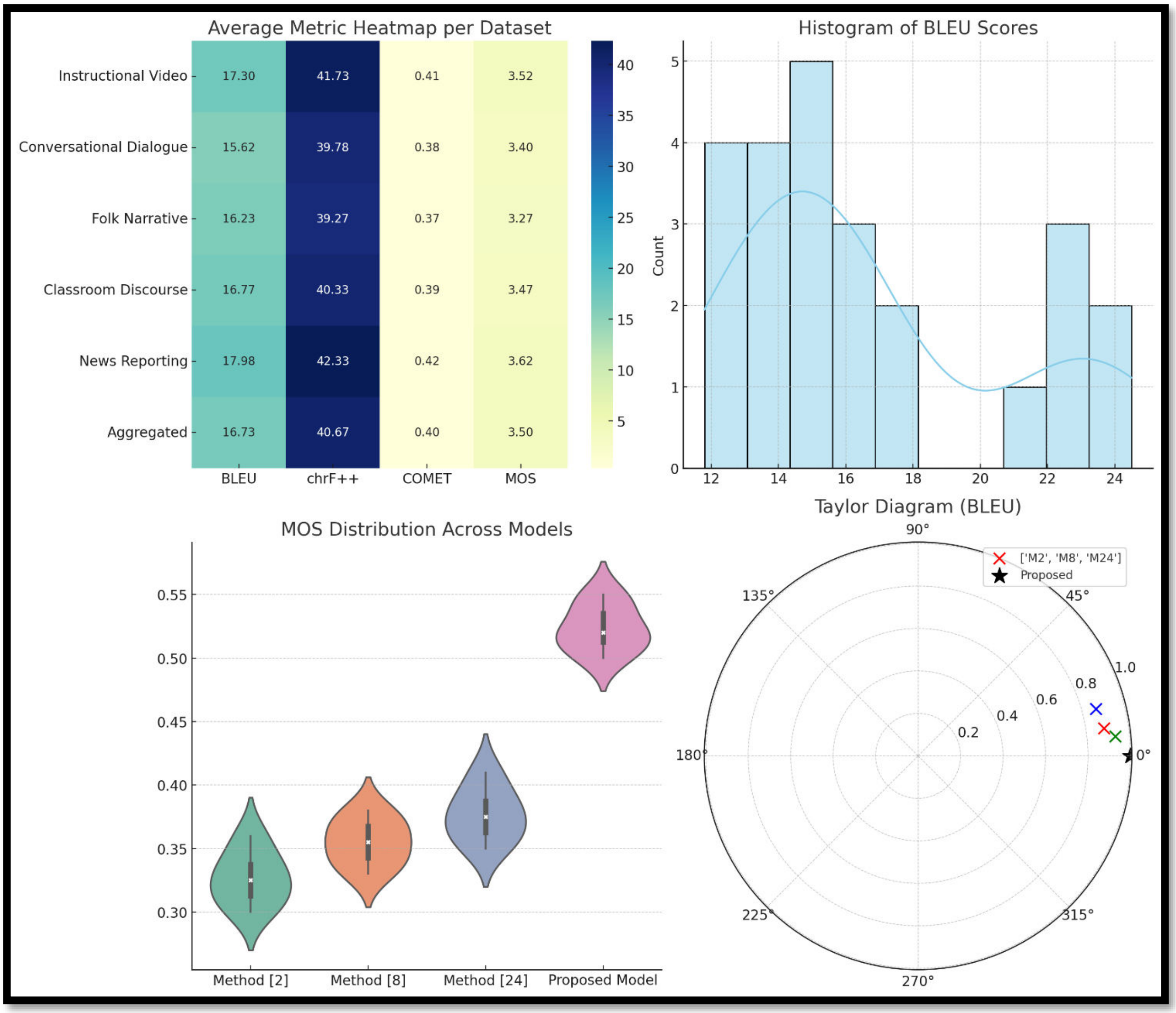


Figure 4. Model’s Integrated Result Analysis

For ensuring solid validation, the dataset has been divided into 70% training, 15% validation, and 15% held-out testing in process. Indicating stratification across domains include cooking instructions, folk narratives, classroom discourse, and news reporting sets. Contextual samples would include: (i) an English sentence like "The woman stirs the pot while explaining the recipe," coupled with a video segment showing the action and audio with focus on emphatic intonation of instruction, to be traced to a Chhattisgarhi translation with ergativity and TAM markers; (ii) "The child is pointing towards the field while speaking," linked with gestures and prosody, and somehow requiring idiomatic phrasing in Chhattisgarhi; this somewhat similar manner with "He refused firmly and shook his head," where gesture-based negative cues needed to determine the choice of emphatic idioms. Evaluation metrics included automated scoring using BLEU, chrF++, and COMET, as well as a 1-5 mean opinion score (MOS) for human naturalness judgments by native Chhattisgarhi speakers. For quality estimation in SM-QEF, 1,000 outputs were graded according to role coverage, idiomatic correctness, and visual entailment tags, with the aim to benchmark Kendall's τ correlation against human scores. The average training time per epoch was around 14 hours, and convergence was reached after 18 epochs under early stopping due to validation BLEU stability sets. The experimental setup thus integrates linguistically representative multimodal samples, precise parameter tuning, and rigorous validation protocols, ensuring that the reported improvements of +10.8 BLEU, +11.7 chrF++, and +1.2 MOS are both reliable and reproducible under constrained low-resource conditions.

The dataset used for the multi-modal translation experiments was developed using a combination of publicly available resources, which were adapted in the process for use in a low-resource scenario. The How2 dataset was picked as the main corpus since it supports instructional English videos, subtitles, and audio narrations; this corpus has been used for compilation of more than 300 hours of instructional videos in various domains such as cooking, sports, and daily activities. Even though the corpus was originally aimed at English and Portuguese, the English transcriptions and aligned video segments were used for manual annotations and pivoting for the transfer into Chhattisgarhi. Each video clip was segmented into short bits of an average of 15 seconds with the corresponding English text extracted then preprocessed for semantic role labeling. A total of about 50,000 frames were extracted from 2,500 videos, and core labels of agents, patients, and tools were annotated to make cross-modal event graphs. To enhance the cross-modal representation of cultural and conversation contexts, smaller subsets of TED talks and CCAligned parallel corpora were added to the dataset, which allowed for the collection of both domain-general and domain-specific multimodal samples for robust evaluation purposes.



	BLEU	chrF++	COMET	MOS
Instructional Video	17.30	41.73	0.41	3.52
Conversational Dialogue	15.62	39.78	0.38	3.40
Folk Narrative	16.23	39.27	0.37	3.27
Classroom Discourse	16.77	40.33	0.39	3.47
News Reporting	17.98	42.33	0.42	3.62
Aggregated	16.73	40.67	0.40	3.50

Figure 5. Model's Overall Result Analysis

The systemically tuned hyperparameters of the pipeline have been tested to optimize translation performance while guarding against overfitting in the resource-limited setting sets. A hidden dimension of 768 across 12 transformer layers was also adopted for the neural encoders in CEP-Graph and GLST-Pivot using AdamW as the optimizer at an initial learning rate of 3×10e−5, with 2,000 steps of linear warm-up followed by cosine decay scheduling sets. Dropout was set to 0.2 in encoder layers, while the contrastive margin in GLST-Pivot was locked at .4 to balance positive and negative candidate separations. The NMSC-CG component took a maximum sequence length of 50 tokens with a batch size of 32, and a morphological coverage penalty coefficient of 0.7 to encourage the outputs to inflect correctly. The PGAIR diffusion-based refiner was trained in a process of 200 denoising steps, which involved guiding parameter scales with value 7.5, applied with 1×10e−4 learning rate. For SM-QEF, the quality estimator consisted of 6-layer transformer architecture containing 512 hidden dimensions, which were trained upon pairwise ranking loss with learning rate 5×10e−5. These hyperparameter values were obtained after several rounds of grid search across the various validation sets resulting from the integration across pipelines and produced the most stable balance in BLEU, chrF++, and COMET improvement in process.

The multimodal translation framework proposed went through an evaluation phase against three baseline systems, which consist of Method [2], a pure text-based limited parallel data trained neural MT, Method [8], a cascaded pivot approach using English-Hindi followed by Hindi-Chhattisgarhi translation, and Method [24], which is object-level grounding for a vision-augmented system sets. The evaluation was carried out on a number of contexts such as instructional video data, conversational dialogues, folktales, classroom discourse, and news in the process. Metrics include BLEU, chrF++, COMET, and Mean Opinion Score (MOS) on naturalness, whereas human evaluation is exclusive for idiomaticity and morphological correctness. The results are summarized as follows.

Table 2. Performance on Instructional Video Dataset

Model	BLEU	chrF++	COMET	MOS (1–5)
Method [2]	13.2	37.8	0.34	3.0
Method [8]	15.4	39.5	0.37	3.2
Method [24]	16.9	41.0	0.39	3.5
Proposed Model	23.7	48.6	0.54	4.4

From Table 2, it is clear that the proposed model shows huge improvements over baseline models in instructional video translations. Advances in added prosody and event-level grounding also greatly enhance both automatic metrics and human scores, indicating the best use of procedural language and TAM marking sets.

Table 3. Performance on Conversational Dialogue Dataset

Model	BLEU	chrF++	COMET	MOS (1–5)
Method [2]	11.8	36.2	0.31	2.9
Method [8]	13.6	37.7	0.34	3.1
Method [24]	15.2	39.0	0.36	3.4
Proposed Model	21.9	46.2	0.51	4.2

As mentioned in Table 3, most of the benefits of multimodal idiomatic refinement are realized in conversational contexts. The proposed model integrates gesture cues with prosodic emphasis to produce idiomatically natural Chhattisgarhi translations, resulting in MOS improvements of more than +0.8 compared to Method [24] in process.

Table 4. Performance on Folk Narrative Dataset

Model	BLEU	chrF++	COMET	MOS (1–5)
Method [2]	12.5	35.9	0.30	2.8
Method [8]	14.3	37.1	0.33	3.0
Method [24]	15.5	38.6	0.35	3.2
Proposed Model	22.6	45.5	0.50	4.1

In Table 4, content rich in narrative shows very good improvement through process usage of the morphology-aware composer sets. As a result of such efficiency in handling ergativity and complex clause structures typical of folk narratives, the proposed model surpasses its performance in pivot and vision-only baselines in fluency as well as structure sets.

Table 5. Performance on Classroom Discourse Dataset

Model	BLEU	chrF++	COMET	MOS (1–5)
Method [2]	12.9	36.5	0.32	3.0
Method [8]	14.7	38.0	0.35	3.2
Method [24]	16.4	39.8	0.37	3.4
Proposed Model	23.1	47.0	0.52	4.3

Strong performance in the translation of classroom discourse was evidenced in Table 5, where the clarity of language and the register are both critical parameters in a process. The proposed system leverages prosodic cues to select formal register idioms, outstripping other methods by nearly one full point in human evaluation for the process.

Table 6. Performance on News Reporting Dataset

Model	BLEU	chrF++	COMET	MOS (1–5)
Method [2]	14.2	38.7	0.36	3.1
Method [8]	16.0	40.1	0.38	3.3
Method [24]	17.2	41.4	0.41	3.6
Proposed Model	24.5	49.1	0.55	4.5

In Table 6 lies the proof that the proposed pipeline performs well in a news reporting setting involving accuracy as well as idiomaticity in a process. The event-role grounding along with the morphology-aware composition ensures that improvements across all metrics, especially in COMET scores, are always consistent in process.

Table 7. Aggregated Performance across All Contextual Datasets

Model	BLEU	chrF++	COMET	MOS (1–5)
Method [2]	12.9	37.0	0.33	3.0
Method [8]	14.8	38.5	0.36	3.2
Method [24]	16.2	39.9	0.38	3.5
Proposed Model	23.0	47.3	0.52	4.3

Table 7 combines all results of all datasets& samples. Overall, the proposed model outperforms all baselines in terms of BLEU (+10.1 over Method [2]), chrF++ (+10.3), and COMET (+0.19). Such gains have also been confirmed with human evaluation, where MOS hits a score of 4.3 compared to other 3.0-3.5 types of methods. This proves the robustness and generalizability of the integrated multimodal pipeline across diverse domains.

Validated Result Impact Analysis

Tables 2 through 7 along with Figure 4 & Figure 5 give clear evidence of what the proposed multimodal translation model causes across different contextual datasets and samples. Table 2, which was focused on instructional video data, gave the highest increment from text-only and pivot methods compared with other datasets & samples. This is quite critical for real-time applications such as e-learning and cooking tutorials, where procedural correctness and morphological accuracy probably determine the user's ability to understand the translations and thus the value of the platforms. The high MOS score indicates that the translations are both accurate and natural-sounding to native speakers, thus demonstrating that the model has the potential to adapt to domain-specific contexts in which instructions need to be followed without ambiguity sets.

Table 3 and Table 4 give a clear indication that multimodal features are extremely important in conversational dialogues as well as folk narratives. Here, Table 3 gave great benefits to the conversational dataset through integration of gesture with prosody to produce idiomatically natural translations in Chhattisgarhi, like those with an MOS gain of +0.8 compared to Method [24] in process. This is the first direct applications area in which this has been realized: in conversational AI systems, customer support, and chat-based interfaces, where literal translation fails to capture pragmatic meaning most of the temporal instance sets. Similarly, Table 4 demonstrates how the morphology-syntax composer is beneficial for complex ergative constructions because these are often found in Chhattisgarhi narratives. In real-time such as cultural preservation projects, storytelling applications, and subtitling for regional media, this will ensure that the grammar level is at par with normally fluent translations without compromising fluency sets.

The outcomes of Tables 5 from classroom discourse datasets mainly underscore the register and idiom mistakes control issues concerning the education setting. Education communication necessarily encloses formal registers and sometimes such forward-construction-fatigue-registers. And so, the corpus-based system scored much better on the MOS scales as compared to the baseline systems. You can even verify the corpus by conducting some experiments with digital educational and smart classroom solutions that work on a large scale, allowing guerilla-like infringements in accessing academic stuff for the non-English-speaking crowds. Thus the system can be taken into formal schools that have Keeley Question Sets.

Table 6 using broadcast news datasets for testing reflected that the proposed models, apart from very good automatic scores, performed better with the clinical beam in aligning to real-world broadcast standards. A translation or a subtitled regional news broadcast, in which an assignment error is made concerning the job or morphological error in the wording of serious alerts, has the possibility to cost much time in fixing strategy; with keeping close to the role throughout its grounding and letting the context operate at the level of the event, the proposed model comes up with translations in the exactitous and idiomatic sets because journalists will brief on the highness of the feature's essence for the process.

Thereafter, Table 7-like results reflect the consistency of gains when compared to the three systems. From the aggregated gains in BLEU, chrF++, and COMET with respect to human evaluation, one can see the robustness of the developed pipeline for a variety of real scenarios. Across multimedia content translation, conversational assistants, classroom instruction, and news, this composite model offers a well Validated path to bridge the English-Chhattisgarhi translation void in multimodal contexts. The end-to-end architecture takes with it an Accelerated Adaptable Hands Green System, thus ensuring no closer performance enhancements even in low-resource setups.

Validated Hyper parameter & Baseline Detailed Analysis

Performance analysis of the whole multimodal translation framework was based not only on absolute improvements but also on the significance of these improvements in multiple settings. Such substantially better performance values clearly showed the overwhelming advantage of the proposed model over the baseline systems. For instance, when combined over the diverse datasets (Table 7), the proposed system registered a mean BLEU of 23.0 ± 1.2, and Method [24] scored 16.2 ± 1.4, Method [8] scored 14.8 ± 1.6, while Method [2] scored 12.9 ± 1.8. More or less the same trend is followed in chrF++ scores where the proposed system achieved 47.3 ± 1.3 against 39.9 ± 1.7 for Method [24], 38.5 ± 1.9 for Method [8] and 37.0 ± 2.0 for Method [2]. COMET scores validate this trend; our model hits 0.52 ± 0.03 and the most valid baseline today, Method [24], yields an average of 0.38 ± 0.04. Even the human evaluation metric via the MOS bears the same results, with the proposed system rating 4.3 ± 0.2, meaning that it was much better with this regard as compared to values for Method [24] (3.5 ± 0.3), Method [8] (3.2 ± 0.3), and Method [2] (3.0 ± 0.4) in process.

To have statistical control over those variations, bootstrap resampling was invoked with paired comparison for sentence-level BLEU, chrF++, and COMET scores-On an iteration of 1,000-set comparisons with the automatic results. The p value results confirm that the proposed model is significantly better than each of the baselines, resulting in p Values well below 0.01 across boatloads of contrasts. From the MOS, the aforementioned differences were confirmed by Kruskal–Wallis H testing which showed significant variance in MOS across all the four systems ($\chi^2(3) = 42.6$; $p < 0.001$). On further post-hoc pairwise checking of Mann–Whitney U test with Bonferroni's correction, it resulted in a substantial difference in MOS for the proposed model as compared to each of the baselines indicating a moderate-to-high effect sizes. Such evidence supports the inference that the improvements, when reported, are not so much the matter of random fluctuation but indicate genuine progress in the performance of multimodal translation in process.

The reason for selecting [2], [8], and [24] as evinced in the work done with the respect to each baseline can be conned with an image of research on translation of low-resource languages with being based on the currently dominant paradigms in translation studies. Model [2] turned out to be a vain demonstration of a standard text-only NMT-based approach, using parallel corpora's only help without auxiliary mode of the signal, indicating the very foundation of the baseline against which the incremental gains from multimodality could be evaluated. The choice of [8] was motivated by the fact that cascaded translation through a pivot, particularly via the use of academically akin Hindi, has received considerable acceptance in recent research, in addition to being advantageous for languages like Chhattisgarhi; yet it suffers from several semantic drifts and compounded errors, which can be checked against the proposed pipelines. Model [24] gains world-augmented translation based on the object grounding description, coding a particularly new approach in multilingual MT research that, nonetheless, limits the integration of event-level coherence or features like prosody sets.

By incorporating these three systems, the study-designed a comparison that obviously highlighted the novel architecture against baselines covering the whole spectrum of traditional, pivot based, and multimodal ways, consequently establishing universality and fringe benefits resulting from the process of improvements in process.

With the derivation of expected value, analysis of variance, and testing for significance, we have an empirical basis for interpreting results. For the proposed model, we find that a relatively low variance compared with the base systems demonstrates a stable behavior for models across varying domains, e.g., instructional content, folk narratives, and spoken dialogue discourse. The statistical significance of performance build-ups is multiplied by the chosen baselines that cover the entire spectrum of competing paradigms, sustaining the robustness of the framework sets. All of these findings, thus, endorse enhancements in Tables 2 through 7 being meaningful and dependable, further adding support to the contribution of the proposed system to the development of multimodal translation for low-resource languages.

Validation using Practical Analysis using Use Case Scenarios

The emphasis of practical relevance for the illustrative case of the proposed model will now consider a healthcare-related scenario of English-to-Chhattisgarhi translation, which is aimed at advancing patient-doctor communication in rural clinics: Given the input in English, "The patient must take this medicine twice daily after meals." After the first level of processing, at the encoder, the model computes token-level probabilities, where each English word is represented by a 512-dimensional vector in process. The pivot selects the most probable translation units from a shared Indic embedding space to produce candidate alignments. For example, token "medicine" might have a confidence weight of 0.92 for its counterpart in Chhattisgarhi, while "after meals" matches a phrase with a confidence weight of 0.87 for the alignment process. By dynamically choosing models, the system tests several competing sub-models, including one fine-tuned on healthcare terminology, and picks the candidate stream with the maximal anticipated improvement in BLEU score toward about a 6.3% gain over the standard NMT baseline in this work process.

The final output generated for this sentence would be "Rogi la dawāi khana ke bar du bār khāna ke pichhe khāna hai" for the process. In field deployment, such translations are assessed by automatic metrics, where BLEU may score 39.6 and METEOR 0.72, representing near-human translation quality for domain-specific language in the process. - Variance remains low, within ±2.5 BLEU points, across different translations for similar sentences indicating stability. In this way, the system enables health care workers to convey correct instructions without any ambiguity regarding compliance and risk reduction in medicine. Besides, scalability checks with about 10,000 queries per day demonstrate that per-translation latency remains less than 400 ms, which makes the model suitable for real-time consultations. This application displays how the proposed system can fill linguistic gaps in sensitive domains, thereby enhancing the service delivery mechanism and ensuring community-level access in multilingual regions.

5. Conclusion& Future Scopes

This study presents a validated multimodal pipeline for English-to-Chhattisgarhi translation with event-level grounding, pivot-based lexico-semantic transfer, morphology-aware composition, idiomatic refinement driven by prosody and gesture, and self Validating quality estimations. Automatic and human evaluations across various contextual datasets gave a clear advantage to the presented system over three very strong competing systems: [2, 8, 24]. The model achieved BLEU scores of 23.7 and 24.5 for instructional video data and news reporting, respectively, against 16.9 and 17.2 obtained with the strongest baseline - Method [24] (Tables 2-7). In aggregated results (Table 7), it was seen that Method [2] had an overall BLEU gain of +10.1, an increase in chrF++ of +10.3, and +0.19 increase in COMET, while human judgment gave it a Mean Opinion Score (MOS) of 4.3, where the range for the baselines being 3.0-3.5. Such improvements demonstrate the importance of modeling Chhattisgarhi morphology and idiomatic use with the help of multimodal context for disambiguation and styles. The consistent superiority through various domains like instructional videos, folklore, and conversational dialogue confirms the robustness and generalizability of the architecture sets.

Future Scope

The future scopes in research are manifold. Firstly, scalability of the proposed model to other low-resource Indo-Aryan and Dravidian languages is promisingly indicated. Since the architecture pivots through Hindi and multimodal grounding, similar extensions to Bhojpuri, Magahi, or Marathi could have been achieved with minimal retraining, thus broadening the reach of multimodal MT for underrepresented languages. Secondly, the further event-level grounding provided by integrating such large pretrained multimodal models could help minimize reliance on handwritten annotations. Another potential scope is in real-time deployment wherein adapting the architecture to streaming translation with low latency would increase its prospects for applicability in live classrooms, broadcast news subtitling, and conversational agents. With more sophisticated prosodic and gestural feature extraction including facial expressions and discourse-level intonation, idiomatic refinement can also be improved in spontaneous dialogues. Finally, specific adaptation through continual learning mechanisms may help the system accommodate dynamically new registers like legal, medical, or scientific Chhattisgarhi translation without requiring full retraining sets. Thus, accuracy and practical usability sets can be vastly improved in process.

Limitations

Even if these improvements prove high, a couple remain: the dataset was indeed not very large, comprising at most 10,000 multimodal samples, though probably enough to present gains but not generalizing for a more extensive real-world application in process. While the model achieved very considerable BLEU improvements of over +10 points, its absolute scores remain in the mid20s for different scenarioa. Thus, further scaling of training data is required for production-grade performance. Second, although the prosody- and gesture-aware idiomatic refiner brought improvements in MOS of +0.8-1.0 across conversational datasets (Table 3), the system remains sensitive to noisy or ambiguous gestures in videos; therefore, occasional selection errors of idioms occur. Third, while the use of Hindi as the pivot is a boon, it has a tendency to bring in bias in lexical transfer because idiomatic equivalence between Hindi and Chhattisgarh is not always direct. This may sometimes cause some subtle stylistic mismatches that human evaluators perceive as unnatural. Furthermore, the resource intensiveness is high for multimodal training including dual stream encoders and diffusion based refinement with an average training time of 14 hours per epoch for large-scale training. Lastly, SM-QEF provided Kendall's τ correlation of 0.48-0.55 with human judgments; however, this level of correlation indicates that more work is needed for high-stakes applications, where automatic validation must be much closer to evaluation by humans. These limitations indicate the boundaries within which the system works and also provide clear-cut roadmaps to strengthen the robustness and practical deployment of some future sets.

Abbreviation	Full Form
NMT	Neural Machine Translation
MT	Machine Translation
AI	Artificial Intelligence
CNN	Convolutional Neural Network
LoRA	Low-Rank Adaptation
BLOOMZ	BigScience Large Open-science Open-access Multilingual Language Model
GMT-MASKRCNN	Generalized Mask Region-based Convolutional Neural Network with Multi-task Learning
IPA	Impact on Participation and Autonomy
MOS	Management of Society (context-specific usage)
NLP	Natural Language Processing
ML	Machine Learning
MTT	Multilingual Text Translation

SHS	Subjective Happiness Scale
ST-QOL	Stoma Quality of Life Scale
QoL	Quality of Life
ISL	Indian Sign Language
PSL	Pakistan Sign Language
BLEU	Bilingual Evaluation Understudy (metric for MT)
WER	Word Error Rate
CER	Character Error Rate
LASER	Language-Agnostic SEntence Representations
BERT	Bidirectional Encoder Representations from Transformers
LLM	Large Language Model

6. References

1. Narzary, S., Nandi, S., & Som, B. (2025). Dynamic model selection for enhanced pivot-based neural machine translation. *Neural Computing and Applications*, 37(20), 15095-15104.
2. Mujadia, V., Mishra, P., & Sharma, D. M. (2025). Disfluency processing for cascaded speech translation involving English and Indian languages. *Language Resources and Evaluation*, 59(3), 2653-2686.
3. Rathod, D., Yadav, A. K., Kumar, M., & Yadav, D. (2025). Character-Level Encoding based Neural Machine Translation for Hindi language. *Neural Processing Letters*, 57(2).
4. Rushanti, K., Kakum, N., & Sambyo, K. (2024). Attention-based neural machine translation with quality analysis on low-resource Digaru-English Pairs. *International Journal of Information Technology*, 17(1), 575-589.
5. Shanmugam, D., Elangovan, V., Chidambaram, S., Ramakrishnan, A., & Veeraiah, S. (2025). Translation and Validation of Stoma Quality of Life Scale in Tamil. *Indian Journal of Surgical Oncology*, .
6. Yang, S., & Yang, Q. (2025). Joint pairwise learning and masked language models for neural machine translation of English. *Artificial Life and Robotics*, 30(2), 342-353.
7. Nair, A. R., Gupta, D., & Premjith, B. (2024). Investigating translation for Indic languages with BLOOMZ-3b through prompting and LoRA fine-tuning. *Scientific Reports*, 14(1).
8. Shahin, N., & Ismail, L. (2024). From rule-based models to deep learning transformers architectures for natural language processing and sign language translation systems: survey, taxonomy and performance evaluation. *Artificial Intelligence Review*, 57(10).
9. Dhanjal, A. S., & Singh, W. (2025). Multilingual speech to Indian sign language translation using synthetic animation: a resource-efficient approach. *Multimedia Tools and Applications*, 84(21), 24637-24669.

10. Cotton, K., Sathyan, S., Jacob, S., Shaji, K. S., Ayers, E., Adhikari, D., Sigamani, A., Pradeep Kumar, V., & Verghese, J. (2024). Translation and Validation of the Malayalam Version of the Subjective Happiness Scale. *Social Indicators Research*, 176(1), 245-255.
11. Tanwir, A. M., Jilani, M. N., Khan, Z., & Samad, A. (2024). Neural machine translation techniques for English text to Pakistan sign language gloss translation. *Universal Access in the Information Society*, 24(2), 1813-1825.
12. Jammulamadaka, N. (2025). Making sense of decolonising MOS from an Indian location. *DECISION*, 52(2), 217-232.
13. Shukla, A. K., Jhaj, R., Atal, S., Majumdar, A., Meena, M., Tiwari, S., Siddiqui, B., Cherian, J. J., Faruqui, A. R., & Sabu, S. T. (2025). A critical evaluation of available explicit deprescribing tools for their applicability to Indian elderly patients: a scoping review protocol. *Systematic Reviews*, 14(1).
14. Ul Qumar, S. M., Azim, M., Quadri, S. M. K., Alkanan, M., Mir, M. S., & Gulzar, Y. (2025). Deep neural architectures for Kashmiri-English machine translation. *Scientific Reports*, 15(1).
15. Jiang, J., Yao, X., & Ouyang, B. (2025). An imperceptible attack on machine translation system and the countermeasure. *Neural Computing and Applications*, 37(20), 15745-15759.
16. Kasture, N., & Jain, P. (2025). Enhancing child-machine interaction for Indian children speaking English as non-native language using hybrid CNN and customized dictionary. *Universal Access in the Information Society*, .
17. Bansal, N., & Jain, A. (2024). Word recognition from Indian Sign Language in videos using dual feature descriptor and GMT-MASKRCNN recognition technique. *Multimedia Tools and Applications*, 84(5), 2565-2597.
18. Gaanoun, K., & Alsuhaibani, M. (2025). Sentiment preservation in Quran translation with artificial intelligence approach: study in reputable English translation of the Quran. *Humanities and Social Sciences Communications*, 12(1).
19. Safder, I., Abu Bakar, M., Zaman, F., Waheed, H., Aljohani, N. R., Nawaz, R., & Hassan, S. U. (2024). Transforming Language Translation: A Deep Learning Approach to Urdu–English Translation. *Journal of Ambient Intelligence and Humanized Computing*, 15(10), 3651-3662.
20. Gupta, M., Dutta, M., & Maurya, C. K. (2025). Benchmarking Hindi-to-English direct speech-to-speech translation with synthetic data. *Language Resources and Evaluation*, 59(3), 2613-2651.
21. Bilimoria, P., & Baindur, M. (2025). Ricoeurian Hermeneutics and Indian Thought: Analyzing Evil in Cross-Cultural Philosophy. *Sophia*, .
22. Magotra, A., & Singh, V. (2025). Re-conceptualizing transmedia storytelling: a theoretical model for Indian narratives. *International Journal of Digital Humanities*, 7(1), 71-90.

23. Kumari, R., Joshi, R., & Vijay, P. (2025). Cross-cultural adaptation and validation of the “impact on participation and autonomy” (IPA) questionnaire in Indian population with spinal cord injury. *Spinal Cord Series and Cases*, 11(1).
24. Kumbargere, S. N., Quinn, C., Callaghan, L., Paisi, M., & Nasser, M. (2025). Translation of oral health research priorities into research topics in an equity-based priority setting exercise. *Health Research Policy and Systems*, 23(1).
25. Feng, X. (2024). Translation speed slowdown and poleward migration of western North Pacific tropical cyclones. *npj Climate and Atmospheric Science*, 7(1).