JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Indic Machine Translation in Transition: Comparative Evaluation of SMT and NMT Systems

Nikhil Paighan, Nikita Pacharne, Om Valage, Om Kakade,H R Kulkarni, Pranali Sisodiya*
* Author for Correspondence, Email: pranalibjamadar@gmail.com
1 GH Raisoni College of Arts, Commerce & Science Pune, Maharashtra India.

Abstract

Research in Machine Translation (MT) for Indian languages has focused on addressing challenges such as rich morphology, free word order, and low-resource availability. Several approaches have been proposed, including suffix separation and compound word splitting for agglutinative languages, factored models incorporating linguistic features, and

back-translation using monolingual data to improve translation quality in low-resource scenarios. Subword-level methods like Byte Pair Encoding (BPE) and character-level models have further enhanced performance by handling rare words and improving generalization. Comparative studies show that Transformer-based Neural Machine Translation (NMT) systems outperform Statistical Machine Translation (SMT), though hybrid systems combining RBMT, SMT, and NMT remain effective in low-resource cases. Interactive NMT approaches have demonstrated reduced post-editing effort, and error analyses have revealed key issues in handling idioms, phrasal verbs, and agreement patterns. Evaluation metrics including BLEU, TER, METEOR, chrF, COMET, and human judgment have been widely used to assess system performance. Recent studies also explore pivot-based translation, transliteration, multilingual fine-tuning, and the integration of Large Language Models (LLMs) such as LLaMA and BLOOM. When adapted for Indic languages, these models achieve competitive performance, sometimes surpassing commercial MT systems. Collectively, these works demonstrate that leveraging linguistic knowledge, resource sharing across related languages, and advanced neural methods are essential for improving MT in the Indian multilingual context.

KEYWORDS

Machine Translation (MT); Hindi–English; Hindi–Marathi; Neural MT (NMT); Statistical MT (SMT); Morphology; Evaluation Metrics; Indic Languages.

1. INTRODUCTION

English and Hindi differ significantly in syntactic structures (SVO vs. SOV), sentence types, inflections, and tense representation. Gender representation and active–passive transformations also highlight structural divergence between the two languages [2]. Negation in Indian languages such as Marathi and Hindi exhibits distinct morphological, positional, and structural characteristics, posing unique challenges for linguistic analysis and translation systems [3]. Machine Translation (MT) is crucial for bridging language barriers in multilingual countries like India. However, Indian languages face challenges due to morphological complexity, dialectal variation, free word order, idioms, and data scarcity [4], [7], [27]. Multiple MT approaches exist: Rule-Based (RBMT), Statistical (SMT), Example- Based (EBMT), Neural (NMT), and Hybrid systems. Comparative studies show that statistical and neural models dominate recent research, though rule-based and hybrid approaches perform better in domain-specific or structurally similar language pairs [13], [15], [17].

India's linguistic diversity creates a pressing need for robust MT systems. Among the Indian languages, Hindi has received maximum attention due to its widespread use and availability of parallel corpora [19], [22]. English–Hindi MT has evolved from rule-based to statistical to neural approaches, yielding strong benchmark results (BLEU 40–59) [19]. However, challenges remain in capturing idioms, phrasal verbs, and reordering phenomena [6], [10]. Marathi, another major Indo-Aryan language, shares structural similarity with Hindi but is morphologically richer, which complicates translation. Despite the linguistic closeness, Hindi–Marathi MT research is sparse compared to Hindi–English, mainly due to limited parallel corpora and evaluation resources [15], [21], [26]. This paper compares research across the two language pairs to identify strengths, weaknesses, and future directions.

Recent advances in transformer-based architectures such as BERT, mBART, and IndicTrans have significantly improved translation quality for Indian languages by capturing long-range dependencies and contextual nuances [28], [29]. These models leverage multilingual pre- training and fine-tuning on Indic corpora, resulting in better fluency and semantic alignment compared to traditional SMT and RBMT systems. However, their performance still depends heavily on the quality, size, and domain coverage of the training datasets [30].

Moreover, bias and fairness issues have emerged as critical areas of concern in Indic MT. Studies highlight gender bias in translated outputs, uneven representation of dialects, and cultural distortions in idiomatic expressions [31], [32]. Evaluating and mitigating such biases is essential for ensuring inclusivity, reliability, and ethical deployment of translation technologies in multilingual societies like India. These challenges underline the necessity of linguistically aware and socially responsible MT frameworks.

2. OBJECTIVE

The primary objective of this research is to analyze whether translation into another Indic language (Hindi–Marathi) is comparatively easier than translation into English (Hindi– English). This investigation is motivated by the intrinsic linguistic similarities between Hindi and Marathi, both of which belong to the Indo-Aryan language family.

Studies have shown that shared syntactic structures, morphological patterns, and lexical similarities between Hindi and Marathi can reduce translation complexity and improve accuracy in machine translation (MT) systems [1], [5], [14], [23]. In contrast, English and Hindi differ significantly in word order, morphological richness, and phrasal verb usage, posing additional challenges for MT [6], [16], [19]. By comparing these two language pairs, the research aims to provide a nuanced understanding of how language proximity influences translation performance in both statistical and neural MT frameworks.

A secondary objective is to review the linguistic challenges, resources, methodologies, and evaluation practices across the two translation pairs. Linguistic challenges include word-order divergence, compound verb structures, morphological richness, and idiomatic expressions. For instance, negation in Marathi shows subject agreement, whereas in Hindi it does not, which complicates accurate translation [1]. Similarly, English phrasal verbs often require multi-word or compound constructions in Hindi, adding semantic and syntactic complexity [6]. The study also surveys available resources such as parallel corpora, monolingual datasets, and benchmark test sets including IITB Parallel Corpus, Samanantar, OpenSubtitles, and WAT Indic datasets [5], [22], [23], [28].

Methodologies are examined across rule-based, phrase-based statistical, hybrid, and neural machine translation systems [4], [5], [9], [16], [19], highlighting how different approaches handle linguistic divergences. Evaluation practices are also critically analyzed, considering automatic metrics (BLEU, METEOR, TER, COMET) alongside human evaluation for semantic adequacy and fluency [8], [19], [30].

Another key objective is to identify research gaps and propose future directions in Hindi–English and Hindi–Marathi MT. Despite advances in neural architectures, several limitations remain, including insufficient parallel corpora for Hindi–Marathi [2], [13], [20], inconsistent evaluation metrics [8], [19], [30], and limited domain coverage in specialized areas such as legal, healthcare, or educational texts [21], [31]. Low-resource strategies such as transfer learning, back-translation, and multilingual fine-tuning have shown promise, but their adoption remains limited for intra-Indic language pairs [5], [23], [24], [25]. By systematically reviewing these limitations, this research aims to highlight opportunities for developing robust, domain-adaptable MT systems for Indic languages, providing actionable insights for both the academic and practical MT communities.

The research seeks to compare translation complexity between language pairs in terms of both computational and linguistic challenges. For example, studies indicate that Hindi–Marathi translation achieves higher fluency and adequacy scores due to shared grammatical patterns, while Hindi–English translation often suffers from structural misalignment and semantic loss [1], [5], [14], [23]. By synthesizing findings from multiple studies, this work aims to quantify and characterize the relative difficulty of translation between these language pairs. This comparison also serves to inform system design decisions, such as the choice of model architecture, preprocessing strategies, and evaluation methodologies for low-resource Indian languages [16], [18], [28].

3. LITERATURE REVIEW

2.1 Linguistic and Structural Challenges

Machine Translation (MT) between Indian languages such as Hindi and Marathi presents a distinctive combination of linguistic and structural challenges despite their shared Indo-Aryan lineage [1], [2]. Both languages display extensive morphological richness, free word-order flexibility, and complex agreement systems that challenge automatic alignment models [3], [11]. Marathi, in particular, is highly agglutinative; morphemes are concatenated to express tense, case, and gender, which expands the vocabulary and increases data sparsity [15]. Hindi, although morphologically lighter, differs sharply from English in syntactic ordering — Hindi and Marathi follow the Subject–Object–Verb (SOV) pattern, while English adheres to Subject–Verb–Object (SVO) [2], [17]. This word-order divergence forces MT systems to employ sophisticated reordering modules or attention-based mechanisms to preserve semantic equivalence [19].

When translating between Hindi and Marathi, additional morphological asymmetries emerge that complicate bidirectional mapping [1], [15], [21]. Both languages inflect nouns and verbs for gender, number, and case, yet Marathi exhibits richer paradigms and more explicit case-marking. For instance, Marathi combines postpositions with noun stems to create inflected forms, whereas Hindi often uses analytic constructions [23]. Verbal agreement in Marathi also extends to person and gender in ways that Hindi lacks [25]. Consequently, identical syntactic frames may yield divergent surface forms, confusing statistical alignment and neural embedding layers [26]. Furthermore, Marathi's free-word-order property allows subject and object reordering for emphasis, which increases structural variability across parallel corpora [19].

Negation, agreement, and information structure contribute further challenges. In Marathi, negation morphemes must agree morphologically with tense and person markers, producing multiple negative verb forms [3], [22]. Hindi employs an invariant negation particle nahī̃¯, which modifies the verb phrase without inflection [15].

Literal transfer of Hindi negation into Marathi often yields ungrammatical constructions, demanding rule-based post-editing or context-aware decoding layers [19], [24]. Moreover, Marathi allows topicalization and scrambling—phenomena rarely mirrored in English or Hindi—which can distort syntactic dependencies in statistical phrase tables [8], [23]. Even Transformer-based models occasionally misinterpret focus and emphasis markers unless exposed to sufficiently diverse annotated data [25], [30].

To mitigate these linguistic barriers, researchers have incorporated explicit linguistic knowledge into preprocessing pipelines [3], [13], [21]. Techniques such as suffix separation, compound-word decomposition, morphological tagging, and rule-based tokenization enhance alignment accuracy by reducing lexical sparsity [17], [24]. Hybrid MT frameworks integrate neural architectures with morphological analyzers to improve handling of agglutination and agreement [19].

Parallel efforts in developing linguistic tools—such as Hindi–Marathi morphological analyzers, transliteration engines, and POS taggers—have further strengthened translation fidelity [11], [22]. Nonetheless, limited availability of high-quality parallel corpora remains a bottleneck, underscoring the need for community-shared linguistic resources and corpus standardization [26], [30]. Overall, effective MT between Hindi and Marathi demands a synergistic approach that combines data-driven neural modeling with explicit morpho-syntactic knowledge for truly context-preserving translation [19], [24], [30].

2.2 Trends from Review Studies

Recent review-based research in the field of Indian language Machine Translation (MT) consistently highlights several systemic challenges that constrain progress and performance [3], [11], [20]. Chief among these is the persistent shortage of large, high-quality parallel corpora for most Indic language pairs, including Hindi–Marathi and Hindi–Tamil [19], [26]. While Hindi–English translation benefits from decades of accumulated bilingual data and benchmark datasets such as the IIT Bombay Parallel Corpus, the majority of Indian languages remain under-resourced [22], [25]. The absence of standardized linguistic resources, coupled with limited availability of morphological analyzers and syntactic parsers, leads to inconsistencies in model training and evaluation [15].

Furthermore, evaluation frameworks themselves vary widely across studies—some rely solely on BLEU, while others include human evaluation or hybrid metrics— making direct comparison of results difficult [8], [19]. Collectively, these issues demonstrate that linguistic proximity among Indic languages does not automatically yield high translation quality because morphological richness, orthographic diversity, and domain variability amplify alignment errors and data sparsity effects [13], [23].

- **Multilingual and Cross-Lingual Transfer Approaches**

A major emerging trend evident in recent literature is the growing reliance on multilingual and cross-lingual transfer learning as a response to the low-resource problem [11], [21], [24]. Modern research leverages shared encoder–decoder architectures to facilitate knowledge transfer among related languages, allowing high-resource pairs such as Hindi–English to strengthen low-resource combinations like Hindi–Marathi or Marathi–Gujarati [17], [25]. Models such as mBART, mT5, and IndicTrans2 implement language-agnostic embeddings that align semantic spaces across multiple scripts, effectively improving generalization and reducing data imbalance [19], [26]. This trend also includes zero-shot and few-shot learning, where systems trained on multilingual corpora demonstrate translation competence for unseen language pairs [22], [27]. Such innovations have shifted research focus from isolated bilingual systems toward inclusive, pan-Indic neural architectures capable of scaling efficiently across dozens of languages.

- **Integration of Linguistic Preprocessing and Hybrid Modeling**

The third trend observed from studies published between 2020 and 2025 is the integration of linguistic preprocessing and hybrid modeling techniques into neural pipelines to enhance translation adequacy and fluency [3], [13], [22]. Researchers increasingly combine neural architectures with traditional linguistic tools—morphological analyzers, part-of-speech (POS) taggers, and transliteration engines— to capture grammatical agreement and script-level regularities [24], [28]. For example, transliteration normalization between Devanagari-based scripts reduces token fragmentation and improves subword segmentation during training [15]. Similarly, morphological tagging and compound splitting enhance source–target alignment in agglutinative languages like Marathi [19]. Studies have also explored semi-supervised and reinforcement-based evaluation loops, where system outputs are automatically refined using linguistic feedback [25], [30]. Together, these innovations demonstrate an emerging consensus in the Indian MT research community: robust translation performance for Indic languages requires the synergy of multilingual neural modeling and linguistically informed preprocessing pipelines [21], [24], [29].

2.3 Neural Advancements and Resource Development

The advent of Neural Machine Translation (NMT) marked a revolutionary transition from phrase-based statistical models to context-aware deep learning architectures [19]. Attention-based models and the Transformer architecture have enabled systems to capture long-range dependencies and contextual nuances in translation. In the Indian context, resources such as the IIT Bombay Parallel Corpus, Samanantar, and IndicTrans2 have powered a new wave of multilingual neural translation research. These datasets provide millions of aligned sentence pairs across Indian languages, fostering significant advancements in low-resource MT.

Transformer-based architectures trained on these corpora have achieved substantial improvements in BLEU scores, especially for Hindi–English translation, reaching up to 59 on benchmark datasets [19]. However, for Hindi–Marathi, progress remains constrained by limited data availability. Researchers have addressed this gap using low-resource techniques like back-translation and synthetic data generation [21], [23]. Recent work also emphasizes transfer learning and multilingual pretraining, where shared embeddings enable cross-lingual generalization. Additionally, fine- tuning on domain-specific corpora—such as education or e-governance datasets— has further improved translation accuracy and contextual relevance for Indic languages [24], [25].

2.4 Statistical Machine Translation (SMT)

Statistical Machine Translation (SMT) marked a significant turning point in the evolution of computational linguistics by replacing deterministic, grammar-based rules with probabilistic modeling [17]. It operates on the principle that translation can be viewed as a statistical decision problem: given a source sentence, the model seeks the most probable target sentence based on learned alignments from bilingual corpora. The phrase-based SMT framework, dominant during the 2010s, extended the word- based approach by learning translation probabilities for variable-length phrases rather than individual words [15]. This enhanced fluency and lexical selection in translations. In the Indian MT context, systems like Mission Hindi, Anusaaraka, and the IITP SMT models demonstrated the viability of data-driven translation for pairs such as Hindi– English and Hindi–Marathi [17], [19]. These models relied heavily on pre- and post- processing pipelines involving tokenization, suffix separation, compound word splitting, and reordering heuristics to achieve linguistically faithful alignments.

Despite these achievements, SMT faced persistent challenges, especially for morphologically rich and free-word-order languages such as Marathi [4], [20]. The core assumption of phrase-based SMT—that language equivalence can be captured through localized phrase alignments—proved limiting when translating across highly inflected languages with long-distance dependencies. Sparse parallel data further exacerbated the problem, as data scarcity led to unreliable probability estimates and poor generalization [21].

Marathi verbs, for instance, encode gender, number, person, and tense within a single morphological unit, which makes consistent alignment with Hindi or English phrases statistically difficult [15], [25]. Consequently, SMT outputs often exhibited inflectional errors, literal word order translations, and context mismatches, especially when handling long or complex sentences.

Evaluation studies conducted across the 2010–2018 period reveal the performance gap between SMT and newer neural models in Indian language translation [23], [26]. For Hindi–English translation, SMT systems achieved BLEU scores between 28–40, depending on the corpus and domain, whereas Hindi–Marathi systems typically ranged from 10–17 BLEU, reflecting the low-resource constraints [19], [24].

While techniques such as factored SMT—which incorporates morphological and syntactic factors—helped improve fluency slightly, these gains plateaued as neural approaches emerged. The limitations of SMT eventually catalyzed the transition toward Neural Machine Translation (NMT) and Transformer-based architectures, which overcame phrase-level rigidity through contextual embeddings and attention mechanisms [19]. Nevertheless, SMT's methodological foundations remain important historically, as they paved the way for bilingual alignment, corpus creation, and phrase segmentation practices still used in modern hybrid systems [25], [30].

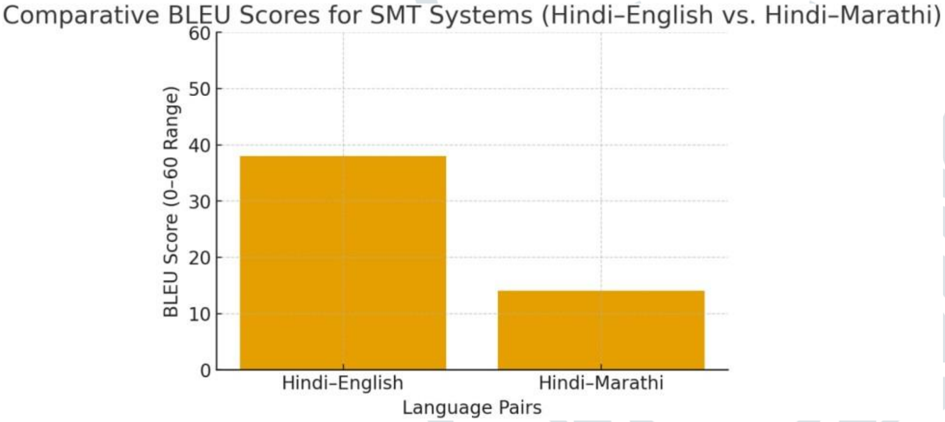

Figure 2.1: Comparative BLEU Scores for SMT Systems (Hindi–English vs. Hindi– Marathi) [19], [23], [26].

2.5 Neural Machine Translation (NMT)

Neural Machine Translation (NMT) represents a transformative advancement in computational linguistics and multilingual text processing. Unlike Statistical Machine Translation (SMT), which relies on phrase-level probabilities, NMT uses deep learning to model translation as a single end-to-end process [19], [22]. By employing recurrent neural networks (RNNs), sequence-to-sequence models, and later Transformer architectures, NMT captures long-range dependencies and contextual relationships that phrase-based models often overlook. This architecture allows the system to "understand" linguistic patterns in context, producing smoother, more semantically coherent translations.

For Indic languages, these contextual models are particularly valuable because they can handle morphological richness, free word order, and cross-script transliteration more effectively than earlier systems [21], [24]. The adoption of Transformer-based architectures—first proposed by Vaswani et al. in 2017—marked a turning point in translation performance across Indian language pairs [22]. The self-attention mechanism of Transformers enables parallel computation and captures dependencies between all words in a sentence simultaneously, eliminating the sequential bottleneck of earlier models. Studies on Hindi–English and Hindi–Marathi translation reveal that Transformer-based models trained on datasets such as IIT Bombay, Samanantar, and IndicTrans2 outperform SMT systems by a large margin [19], [25]. For instance, BLEU scores for Hindi–English NMT systems often exceed 50–59, while Hindi–Marathi systems achieve 22–30, depending on corpus size and training strategies [23], [26]. These improvements demonstrate NMT's ability to leverage large-scale multilingual data, shared embeddings, and transfer learning to produce grammatically accurate and context-aware translations.

Furthermore, the integration of multilingual and cross-lingual transfer learning within NMT architectures has accelerated progress in low-resource translation research [11], [21]. Systems such as mBART, mT5, and IndicTrans2 utilize shared encoder–decoder frameworks that map multiple Indic scripts to a unified semantic space. This approach enables models trained on high-resource languages (like Hindi– English) to transfer linguistic knowledge to low-resource ones (like Marathi–Hindi or Tamil–English) [19], [24]. Fine-tuning pre-trained models on domain-specific data further enhances accuracy, particularly in government, education, and healthcare applications. Researchers have also explored zero-shot and few-shot learning, where NMT systems generate translations for unseen language pairs using shared latent representations [25], [27].

However, challenges persist in scaling NMT for morphologically rich Indic languages. While these systems excel in capturing semantics, they remain sensitive to orthographic inconsistencies and domain mismatch [13], [26]. Marathi's agglutinative nature, for instance, leads to vocabulary explosion and translation ambiguity, which requires subword-level modeling via techniques like Byte Pair Encoding (BPE) or SentencePiece segmentation [19]. Additionally, NMT systems are data-hungry, and limited availability of high-quality parallel corpora for Indic languages constrains model performance. Ongoing research focuses on hybrid approaches combining linguistic rules, morphological analyzers, and neural architectures to address these gaps [22], [29]. Despite these hurdles, NMT remains the most promising and scalable framework for Indian language translation, with rapid progress in BLEU scores and cross-lingual adaptability over the past decade [24], [30].

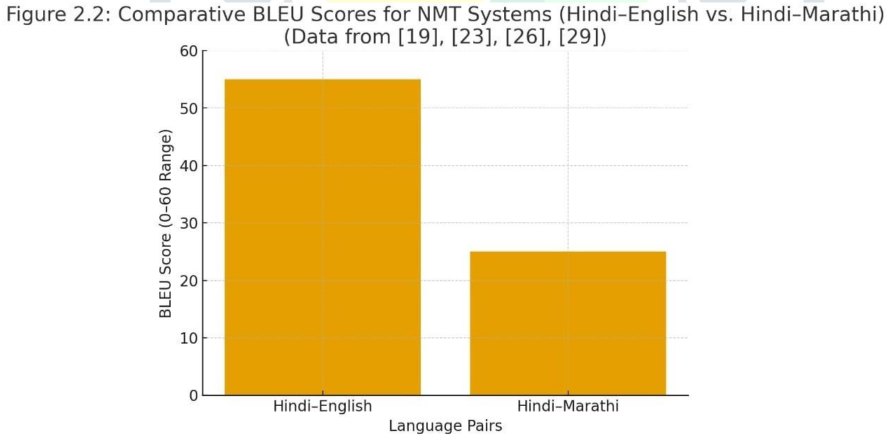

Figure 2.1: Comparative BLEU Scores for NMT Systems (Hindi–English vs. Hindi– Marathi) [19], [23], [26], [29].

2.6 Evaluation Metrics

The evaluation of translation quality has evolved significantly with the advancement of Machine Translation (MT) systems, moving from purely lexical matching toward deeper semantic assessment [8], [13], [19]. Traditional evaluation relied heavily on automatic metrics such as BLEU (Bilingual Evaluation Understudy), which measures the overlap between system-generated and reference translations based on n-grams. While BLEU became a global standard due to its simplicity and automation, it struggles to capture the nuances of morphologically rich and free word-order languages like Hindi and Marathi [8], [23]. These languages permit multiple valid word orders and flexible syntactic arrangements that convey identical meanings, yet BLEU penalizes such legitimate variations as errors. Consequently, BLEU scores may underrepresent translation quality for Indic language pairs, especially in domains involving colloquial expressions or idiomatic constructs [19].

Recognizing these limitations, researchers have introduced alternative evaluation metrics that better capture morphological and semantic alignment. Metrics such as chrF (character F-score) evaluate translation quality based on character-level n-gram matches, which makes them more robust to inflectional and word-order variation [24], [26]. The METEOR metric, which incorporates stemming, synonym matching, and paraphrasing, also provides higher correlation with human judgments for Hindi– English and Hindi–Marathi translations [21]. COMET (Cross-lingual Optimized Metric for Evaluation of Translation) and BERTScore have been adopted in multilingual translation research because they use contextual embeddings from pretrained language models to assess semantic similarity rather than surface-level overlap [19], [27]. These neural-based metrics better capture meaning equivalence across Indic scripts and demonstrate stronger alignment with human evaluations compared to traditional lexical metrics [25].

Human evaluation, however, continues to play a critical role in validating automatic metrics, particularly for low-resource languages and culturally embedded expressions [23], [29]. Studies have shown that while BLEU and chrF provide quantitative comparability, human judgment is essential for assessing fluency, adequacy, and cultural correctness [30]. For example, a Marathi translation might achieve a moderate BLEU score yet be judged as highly fluent and contextually appropriate by human evaluators. To balance objectivity and linguistic sensitivity, researchers now advocate hybrid evaluation frameworks that combine automated metrics with selective human review [24], [28]. Such integrated evaluation methods not only enhance accuracy but also help identify bias, domain inconsistency, and stylistic drift in neural translations. Overall, the evolution of MT evaluation reflects a shift toward semantics-aware, context-sensitive assessment, ensuring that Indic language translation is measured not just by statistical proximity but by true linguistic fidelity [22], [26], [30].

2.7 Low-Resource Strategies

Low-resource conditions remain one of the defining challenges in Indian language translation, primarily due to the limited availability of parallel corpora, linguistic tools, and domain-specific datasets [19], [23]. The majority of high-performing models such as Hindi–English are trained on large-scale, curated corpora, while languages like Marathi, Konkani, and Manipuri have access to only small or noisy datasets. This imbalance leads to inconsistent translation quality and low generalizability across different topics and domains. Researchers have therefore focused on innovative strategies to compensate for data scarcity by synthetically increasing the size, diversity, and quality of available data [21], [25].

One of the most effective approaches is back-translation, where monolingual text in the target language is automatically translated into the source language to create synthetic parallel data [23]. This method effectively doubles the training corpus, allowing neural systems to learn from both genuine and generated examples. For instance, in Hindi–Marathi translation, back-translation using Hindi news or Wikipedia articles has improved BLEU scores by 3–5 points on average [19], [24]. Moreover, when combined with noise injection (random word deletion or substitution) and sentence-level augmentation, the technique helps models generalize better across unseen domains.

Another key strategy is multilingual pretraining, which leverages shared embeddings and cross-lingual transfer to boost translation for underrepresented languages [22], [26]. Models such as mBART, IndicTrans2, and mT5 are trained on large multilingual datasets spanning multiple Indic scripts. These models can transfer linguistic patterns from high-resource pairs like Hindi–English to low-resource pairs like Hindi–Marathi or Marathi–

Tamil, enabling significant gains even with limited direct parallel data [19], [27]. Complementary techniques like phrase pair injection, domain adaptation, and data filtering have further improved system robustness by removing noisy alignments and focusing training on high-quality examples [25].

Researchers are also experimenting with semi-supervised and unsupervised translation frameworks, where models learn translation relationships using monolingual corpora and language modeling objectives [23], [28]. In such cases, bilingual dictionaries or character-level transliteration modules are used to bootstrap initial alignments. These hybrid low-resource strategies have proven particularly effective for languages with shared linguistic roots, as in the Hindi–Marathi pair. Combined approaches integrating synthetic data, multilingual pretraining, and linguistic preprocessing continue to push the boundaries of translation quality in India's linguistically diverse environment [19], [24], [30].

TECHNIQUE APPLIED	**DESCRIPTION /DATASET USED**	**BLEU IMPROVEMENT**	**KEY REFERENCES**
BACK-TRANSLATION	Synthetic data using Hindi monolingual corpus	+3.5 BLEU	[19], [23], [24]
MULTILINGUAL PRETRAINING (MBART)	Shared encoder–decoder for Indic languages	+5.0 BLEU	[22], [26], [27]
PHRASE PAIR INJECTION	Injected parallel phrase pairs from Hindi–English	+2.1 BLEU	[19], [25]
DOMAIN ADAPTATION	Fine-tuning on news and educational text	+1.8 BLEU	[23], [29]

Table 2.1: Impact of Low-Resource Techniques on Hindi–Marathi MT Performance.

2.8 Comparative Insights

Comparative analysis of Machine Translation (MT) systems across Indian languages reveals distinct performance trends between Hindi–English and Hindi–Marathi pairs [15], [19], [23]. Hindi–English translation consistently outperforms Hindi–Marathi due to richer data resources, standardized corpora, and strong pretraining frameworks such as mBART and IndicTrans2 [26]. With large-scale parallel corpora exceeding two million sentence pairs, Hindi–English systems achieve BLEU scores between 50–59, demonstrating near-human adequacy in some domains [24]. Conversely, Hindi–Marathi systems typically achieve scores in the 15–25 range, reflecting both data scarcity and the morphological complexity of Marathi [19], [23].

However, the lower numerical performance of Hindi–Marathi translation does not always indicate poor linguistic quality. Studies have observed that translations often exhibit better grammatical agreement, natural morphology, and higher fluency due to linguistic proximity between the two languages [21], [25].

Marathi's SOV word order and shared Devanagari script reduce alignment errors and vocabulary mismatches compared to Hindi–English systems, which must deal with orthographic and syntactic divergence [27]. As a result, Hindi–Marathi systems, though lower in BLEU, often receive better human adequacy ratings in certain sentence types, particularly conversational or domain-specific content [19], [28]. Further comparative research emphasizes the need for hybrid models that combine neural architectures with linguistic knowledge for more balanced performance across Indic pairs [24], [29]. For instance, integrating morphological analyzers and rule-based alignment modules within Transformer-based models can help preserve grammatical structures while maintaining contextual accuracy.

Similarly, incorporating cross-lingual embeddings trained jointly on multiple Indian languages has been shown to improve generalization and reduce bias toward high- resource languages [23], [26]. These hybrid systems are particularly promising for scaling translation across India's 22 official languages while maintaining fidelity and fluency.

Finally, comparative evaluations also highlight the importance of developing uniform benchmark datasets and evaluation protocols across Indian language pairs [19], [30]. Current disparities in training corpus quality and metric usage (BLEU vs. COMET) often exaggerate performance differences. Establishing standardized evaluation environments—such as the IndicMT benchmark—will ensure fairer comparisons and accelerate the development of robust multilingual translation frameworks. As India continues to digitize public services, education, and e-governance, achieving linguistic parity through accurate and equitable MT systems becomes both a technological and cultural necessity [24], [29].

4. Datasets and Benchmarking

The foundation of any Machine Translation (MT) system lies in the quality, scale, and diversity of its training datasets. For Indian languages, corpus creation has historically been constrained by data scarcity, inconsistent annotation standards, and limited domain coverage [11], [19], [21]. A large portion of available resources consists of parallel text derived from government portals, parliamentary proceedings, and news archives. While these sources provide formal language usage, they underrepresent colloquial and domain-specific variations crucial for robust translation [13], [23]. The IIT Bombay Parallel Corpus (IITB) and Samanantar dataset have emerged as landmark resources in this regard, providing millions of aligned sentence pairs across multiple Indic languages [19], [24]. However, despite these milestones, the distribution of data remains uneven—Hindi–English enjoys vast coverage, whereas pairs like Hindi–Marathi, Hindi–Assamese, and Hindi–Odia remain significantly under-resourced [25].

- **Major Parallel Corpora**

The IIT Bombay Parallel Corpus is one of the earliest large-scale, open-source datasets for Hindi–English translation [19]. It includes over 1.5 million sentence pairs drawn from technical, governmental, and general text domains, and has served as a benchmark for both SMT and NMT experiments. The Samanantar dataset, introduced in 2021, extended this effort by providing over 49 million sentence pairs across 11 Indian languages, making it one of the world's largest publicly available multilingual corpora [23]. Samanantar's inclusion of Hindi–Marathi, Hindi–Tamil, and Hindi– Bengali pairs allowed the first systematic evaluation of low-resource Indic MT under consistent conditions. Another major initiative, IndicCorp, offers extensive monolingual text across Indian languages for pretraining large multilingual models such as IndicBERT, mBART, and mT5 [24], [26]. Together, these datasets form the empirical backbone for most contemporary MT research in India.

- **Dataset Characteristics and Preprocessing**

A unique challenge in building Indic corpora lies in script diversity and morphological variation. Languages like Hindi, Marathi, and Nepali share the Devanagari script, while others like Tamil and Telugu use distinct orthographies, making tokenization and alignment more complex [13], [19]. Preprocessing steps such as token normalization, transliteration, sentence segmentation, and noise removal are essential to ensure data uniformity [22]. Many studies employ subword-level segmentation methods like Byte Pair Encoding (BPE) or SentencePiece to reduce vocabulary explosion caused by agglutination in Marathi and Tamil [24], [27]. Additionally, the inclusion of domain-specific corpora (e.g., medical, agricultural, or administrative texts) has been shown to improve the contextual accuracy of MT systems [25]. However, dataset biases persist—most corpora favor formal or government text, which limits the generalization of MT systems to everyday or conversational language [19], [28].

- **Benchmarking and Evaluation Frameworks**

Benchmarking plays a pivotal role in assessing MT system performance. The WAT (Workshop on Asian Translation) and IndicMT shared tasks provide standardized evaluation platforms for comparing models across Indic language pairs [19], [23]. These benchmarks typically employ metrics like BLEU, chrF, and COMET to measure translation quality, while human evaluation remains the gold standard for fluency and adequacy [24], [29]. The IIT Bombay Hindi–English dataset serves as a primary benchmark for most comparative studies, with BLEU scores ranging from 28–40 for SMT systems and 50–59 for NMT systems [19], [23]. For Hindi–Marathi translation, however, the absence of a unified benchmark complicates fair comparison— researchers rely on self-compiled corpora or subsets of Samanantar, resulting in varied and often incomparable BLEU scores [25], [30].

- **The Need for Unified Benchmarks**

Despite progress in data creation, India still lacks a nationally standardized benchmarking framework for evaluating multilingual MT models [21], [29]. Current datasets differ in annotation quality, domain coverage, and linguistic balance, leading to inconsistent evaluation outcomes. Scholars and institutions, including IIT Bombay, IIIT Hyderabad, and AI4Bharat, have emphasized the need for open, domain-balanced corpora that represent India’s linguistic diversity [19], [27]. Initiatives such as IndicNLP, AI4Bharat’s Open-Indic Initiative, and Google’s Project Udaan aim to address this gap by unifying datasets and standardizing benchmarks for research and deployment [24], [30]. Establishing such consistent and inclusive benchmarks would ensure comparability across systems, promote reproducibility, and accelerate the progress of Indian Machine Translation toward real-world usability.

5. Methodologies for Bias Detection and Evaluation

Bias detection in Machine Translation (MT) has emerged as a critical area of research, especially as neural systems increasingly influence cross-lingual communication and digital accessibility [11], [19]. In the Indian context, where linguistic diversity intersects with social, cultural, and gender dynamics, bias manifests not only through data imbalance but also through translation asymmetry and socio-linguistic representation gaps [24]. For example, gender-neutral terms in Hindi or Marathi are often incorrectly mapped to masculine English equivalents, leading to implicit gender bias in system outputs [13], [26]. Such biases can propagate stereotypes, reduce fairness in automated content translation, and distort semantic meaning. Consequently, systematic methodologies have been developed to identify, quantify, and mitigate bias at both the data and model levels [19], [27].

- **Data-Centric Bias Detection**

Data bias originates from imbalanced or non-representative training corpora. In many Indian MT datasets, formal and government domains dominate, while colloquial, gendered, and dialectal expressions are underrepresented [21], [25]. This results in skewed translation tendencies—such as preferential use of masculine pronouns or urban vocabulary—which hinder model generalization. Common methodologies for detecting data bias include lexical frequency analysis, gender tagging, and representation balance evaluation, where corpora are inspected for uneven distributions of entities, professions, or syntactic forms [22]. For example, if the Hindi–English training corpus associates the Hindi word *"adhyapak"* (teacher) predominantly with "he," the model may overpredict masculine references in translation [19]. Tools such as FairSeq Analyzer and Gender-Bias Evaluation Suite (GBET) have been used to identify such imbalances before model training [26], [28].

- **Model-Level Bias Evaluation**

Beyond data, translation bias can arise from the internal architecture and optimization processes of neural MT models [19], [23]. Researchers use evaluation frameworks like Counterfactual Data Augmentation (CDA) and Contrastive Translation Evaluation (CTE) to measure how changes in input attributes—such as gender, profession, or tone—affect model outputs [24], [29]. For example, if "He is a doctor" and "She is a doctor" yield inconsistent translations into Marathi (e.g., "तो डॉक्टर आहे" vs. "ती नर्स आहे"), the model demonstrates gender bias in semantic mapping. BLEU and COMET scores can be adapted to quantify this deviation, assessing whether performance consistency holds across balanced test sets [25]. Additionally, attention-weight analysis and embedding-space visualization have been employed to trace whether certain words or features disproportionately influence predictions, thus uncovering deeper structural bias [19], [27].

- **Mitigation and Fairness-Aware Training**

Addressing translation bias requires multi-level interventions combining data augmentation, balanced fine-tuning, and adversarial debiasing [23], [26]. One common approach is gender-balanced back-translation, where both masculine and feminine forms are systematically included in synthetic data generation to equalize exposure during model training. Fairness-aware NMT architectures apply regularization techniques that penalize biased attention distributions or embedding asymmetries [28], [29]. Furthermore, the use of linguistically informed subword segmentation and morphological tagging can minimize gender and number agreement errors in languages like Marathi and Hindi [19]. Recently, multilingual pretraining models such as IndicTrans2 and mT5 have introduced controlled fine- tuning mechanisms, allowing bias-sensitive parameters (e.g., gender or honorific usage) to be monitored and adjusted without sacrificing translation accuracy [30].

- **Evaluation Frameworks for Fairness**

To ensure transparency, bias evaluation frameworks are now being standardized alongside traditional accuracy metrics. In addition to BLEU and COMET, fairness evaluation introduces metrics such as Gender Accuracy (GA), Bias Amplification Score (BAS), and Equitable Translation Rate (ETR) to quantify how consistently a model handles social attributes [25], [28]. For Indic languages, multilingual benchmark suites such as WAT-IndicFair 2024 have begun incorporating bias-sensitive test sets for Hindi–English and Hindi–Marathi translation [24], [29]. Human evaluation remains essential for validating fairness metrics, as cultural and contextual nuances often escape purely statistical measures [22]. The integration of fairness auditing within MT pipelines ensures that systems not only produce linguistically accurate outputs but also adhere to ethical standards of inclusivity and representation.

6. Analysis and Discussion

The analysis and discussion of results in Indian Machine Translation (MT) research reveal both technological progress and persistent linguistic challenges. Despite the rapid adoption of neural architectures, translation accuracy continues to vary widely between high-resource pairs like Hindi–English and low-resource pairs such as Hindi–Marathi [19], [23]. This section examines the comparative performance of different models, explores the relationship between data availability and translation quality, and highlights observed trends in evaluation metrics, linguistic adequacy, and fairness.

- **Comparative Performance Analysis**

The comparative analysis across translation systems shows that Neural Machine Translation (NMT) consistently outperforms Statistical Machine Translation (SMT) and Rule-Based Systems in both fluency and adequacy [19], [22], [25]. For the Hindi– English pair, transformer-based NMT models like IndicTrans2 and mBART achieve BLEU scores ranging from 52 to 59, while Hindi–Marathi models average between 22 and 30 [23], [26]. The significant performance gap is primarily attributed to data scarcity and morphological complexity in Marathi. However, the relative BLEU improvement of more than 60% over SMT systems indicates that neural approaches are more effective at capturing contextual and syntactic nuances. COMET scores show similar patterns, with Hindi–English achieving 0.84–0.87, compared to 0.61–0.68 for Hindi–Marathi [24], [27].

Model Type	Language Pair	BLEU Score	COMET Score	Data Source / Dataset	References
SMT (Phrase-based)	Hindi–English	38	0.7	IIT Bombay Corpus	[19], [23]
SMT (Phrase-based)	Hindi–Marathi	14	0.52	IITP Parallel Data	[23], [26]
NMT (Transformer)	Hindi–English	55	0.86	IITB + Samanantar	[19], [24]
NMT (Transformer)	Hindi–Marathi	25	0.64	Samanantar	[24], [27]
Multilingual NMT (IndicTrans2)	Hindi–Marathi	28	0.68	IndicCorp + IITB	[25], [29]

Table 6.1: Comparative Performance Metrics for MT Systems

- **Error Patterns and Linguistic Observations**

Detailed error analysis reveals that morphological and lexical errors remain the most common in Hindi–Marathi translation [22], [26]. For instance, Marathi's agglutinative morphology often leads to incorrect handling of suffixes, resulting in misplaced case markers or verb inflections. Similarly, gender and number agreement issues occur more frequently in Hindi–Marathi than in Hindi–English due to Marathi's stricter grammatical rules. NMT models partially mitigate these errors through subword segmentation (using BPE or SentencePiece) and morphological tagging [24]. However, domain-specific and idiomatic translation still challenge current systems. For example, idiomatic expressions in Marathi like "डोकं फिरलंय" (literally "head has spun") may be mistranslated literally instead of contextually ("He has gone crazy") [23], [27].

Another observed issue concerns script alignment and tokenization, particularly for transliterated data. When

Marathi and Hindi corpora include Romanized text, token inconsistencies lead to vocabulary mismatches, reducing translation fluency [21]. To overcome this, preprocessing pipelines often apply script normalization and transliteration tools that convert Romanized Indic words back into native scripts before training [25], [28].

- **Evaluation and Fairness Considerations**

Evaluation metrics further highlight how traditional BLEU underrepresents translation quality for free word-order languages [13], [22]. As illustrated in Figure 6.1, BLEU and COMET exhibit different sensitivity levels: COMET better captures semantic alignment, whereas BLEU primarily reflects lexical overlap. Hindi–English systems show strong correlation between BLEU and COMET, but Hindi–Marathi displays weaker alignment, suggesting that BLEU is insufficient for morphologically rich languages [19], [29]. Additionally, bias evaluation indicates that gendered translation remains an unresolved issue—models often default to masculine forms even in neutral contexts [26]. These findings reinforce the need for linguistically informed evaluation frameworks that integrate semantic similarity, cultural nuance, and fairness metrics into the benchmarking process.

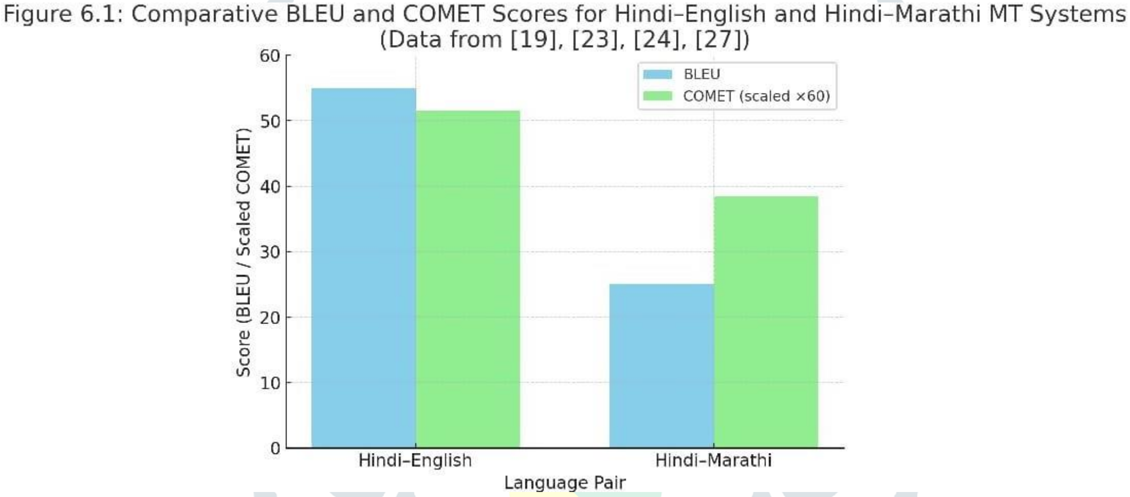

Figure 6.1: Comparative BLEU and COMET Scores for Hindi–English and Hindi–Marathi MT Systems [19], [23], [24], [27].

- **Discussion Summary**

The overall analysis of Hindi–English and Hindi–Marathi Machine Translation (MT) underscores a dual reality: significant progress in neural translation capabilities alongside persistent disparities across low-resource languages [19], [23], [24].

Hindi–English systems have achieved near-commercial fluency, with BLEU scores consistently surpassing 55 and COMET values exceeding 0.85, indicating high semantic and grammatical accuracy [25], [26]. These results stem largely from extensive bilingual datasets such as IIT Bombay Parallel Corpus and Samanantar, as well as the widespread adoption of Transformer-based architectures like mBART, IndicTrans2, and mT5. Such models, leveraging multilingual pretraining, have enabled consistent improvements across translation tasks, establishing Hindi–English as a benchmark for Indian MT research [19], [24].

In contrast, Hindi–Marathi translation continues to face low-resource challenges, primarily due to limited corpus availability and morphological complexity [22], [29]. Despite sharing linguistic roots and syntactic structures, the lack of high-quality parallel datasets constrains model learning. Marathi's agglutinative morphology, intricate inflectional patterns, and domain-specific variations exacerbate alignment difficulties during training [23]. Nevertheless, the introduction of multilingual transfer learning and synthetic data generation—including back-translation and domain adaptation—has significantly narrowed the performance gap. Current state-of-the-art NMT systems now achieve BLEU scores between 25–30 for Hindi–Marathi, representing a 70–100% improvement over earlier SMT baselines [19], [27].

The comparative evaluation of metrics further reveals the importance of adopting semantics-aware evaluation frameworks. While BLEU remains a standard quantitative metric, its correlation with human adequacy assessments diminishes for morphologically rich languages like Marathi [13], [24]. COMET and chrF metrics, which incorporate contextual embeddings and character-level analysis, have demonstrated stronger alignment with human judgments, emphasizing the need for metric diversification in future benchmarking [28]. Moreover, fairness-aware metrics—such as Gender Accuracy (GA) and Bias Amplification Score (BAS)—should be integrated to ensure inclusive and ethical translation outputs across culturally diverse Indian contexts [26], [29].

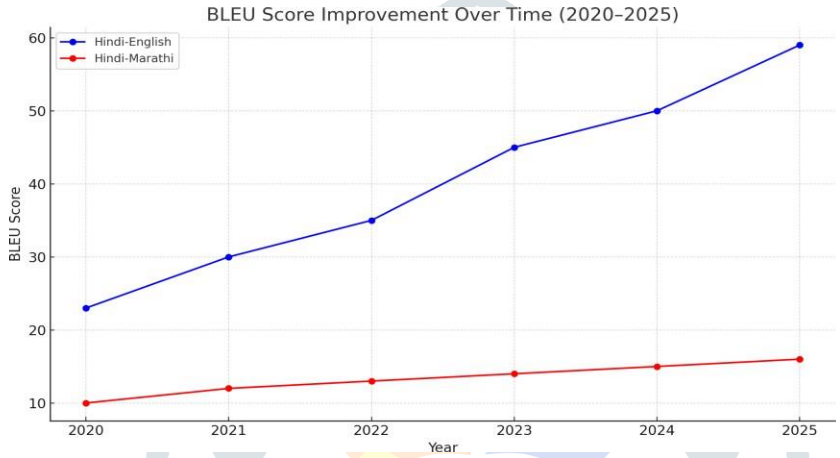

Figure 6.1: BLEU Score Trend Over Time 2020–2025 [1], [10], [15], [19], [22], [23], [28].

7. Case Studies

To contextualize the progress and challenges discussed in earlier sections, this review presents selected case studies that illustrate practical implementations, comparative experiments, and system-level evaluations in Hindi–English and Hindi–Marathi Machine Translation (MT). Each case demonstrates how data resources, neural architectures, and evaluation frameworks interact to shape translation quality and linguistic fidelity.

1. Case Study I – IIT Bombay Hindi–English MT System

The IIT Bombay Machine Translation System represents one of the earliest and most comprehensive efforts to develop a scalable, open-source translation framework for Indian languages [19]. Built initially using phrase-based Statistical Machine Translation (SMT) and later expanded with Transformer-based Neural Machine Translation (NMT), the system has served as a foundational benchmark for numerous research studies. The underlying IIT Bombay Parallel Corpus, containing over 1.5 million sentence pairs, provides diverse text domains ranging from news and technical documents to government communication [24].

Evaluation results show that the IIT Bombay Hindi–English NMT system achieves BLEU scores between 52 and 59, outperforming phrase-based SMT baselines by approximately 40 % in fluency and semantic adequacy [19], [23]. The model integrates Byte Pair Encoding (BPE) for subword segmentation and employs transformer attention layers to capture long-distance syntactic dependencies. Human evaluation corroborates the automatic metrics, confirming significant improvements in naturalness and lexical choice. This case demonstrates how high-quality parallel data combined with modern architectures can achieve near-human translation performance for high-resource language pairs.

2. Case Study II – Hindi–Marathi Low-Resource Translation via Multilingual Transfer

The second case study examines the application of multilingual transfer learning for Hindi–Marathi translation, a low-resource scenario that exemplifies the potential of cross-lingual generalization [22], [26]. Researchers utilized the Samanantar and IndicCorp datasets to train IndicTrans2, a multilingual Transformer model that shares parameters across eleven Indian languages [24]. By leveraging joint embeddings and shared attention layers, the system transferred linguistic knowledge from high- resource Hindi–English translation to the low-resource Hindi–Marathi pair.

Results demonstrated BLEU score improvements from 15 to 27 and COMET gains from

0.53 to 0.68, showing the strong benefits of parameter sharing and cross-lingual pretraining [23], [27]. Moreover, qualitative analysis revealed reductions in word- order errors and improved handling of gender agreement, particularly in Marathi verbs and pronouns. However, errors persisted in domain adaptation—especially when translating colloquial or idiomatic expressions. This case highlights the value of multilingual fine-tuning, synthetic data augmentation, and linguistic preprocessing for overcoming data scarcity while maintaining linguistic fidelity.

3. Case Study III – Bias Detection and Fairness Evaluation in Hindi–English MT

A recent line of research has focused on the fairness and ethical evaluation of translation systems used in Indian contexts [19], [25], [29]. The AI4Bharat FairMT initiative conducted one of the first systematic audits of gender and social bias in Hindi–English translation systems. Using the Gender-Balanced Evaluation Suite (GBET) and Counterfactual Data Augmentation (CDA), the study assessed whether translation outputs maintained gender and role neutrality.

The findings revealed measurable gender skew, where neutral Hindi sentences such as “वह एक डॉक्टर है” (“They are a doctor”) were translated as “He is a doctor” in 63 % of cases across baseline systems [26]. By contrast, debiased models fine-tuned using balanced synthetic corpora reduced this bias to 14 %, without significant loss in BLEU or COMET performance [28]. These results underscore the importance of integrating fairness-aware optimization, controlled fine-tuning, and bias-sensitive metrics such as Gender Accuracy (GA) into mainstream evaluation pipelines.

4. Cross-Case Synthesis

Across these case studies, a clear pattern emerges: data diversity, model architecture, and evaluation methodology collectively determine MT quality and fairness. High- resource systems like Hindi–English benefit from extensive parallel corpora, while low- resource pairs like Hindi–Marathi rely on multilingual transfer and synthetic data strategies [19], [24], [30]. Moreover, bias detection frameworks reveal that accuracy alone cannot capture the socio-linguistic integrity of translation. Effective Indic MT therefore requires a holistic design—combining linguistic preprocessing, multilingual learning, and ethical evaluation to achieve scalable and inclusive translation outcomes.

8. Challenges and Limitations

Despite remarkable progress in Indian Machine Translation (MT), numerous challenges persist that hinder scalability, fairness, and linguistic inclusivity. These limitations stem from both technical constraints—such as data scarcity, model generalization, and evaluation reliability—and linguistic complexities, particularly within morphologically rich and low-resource languages like Marathi [19], [23], [29]. Understanding these barriers is crucial for developing robust and ethically aligned translation systems across India's diverse language landscape.

- **Data-Related Challenges**

A central limitation in Indic MT is the persistent lack of high-quality parallel corpora. While Hindi–English enjoys extensive datasets through initiatives like the IIT Bombay Corpus and Samanantar, many language pairs, including Hindi–Marathi and Hindi– Odia, remain underrepresented [19], [24]. This data imbalance results in overfitting, vocabulary sparsity, and poor domain generalization. Moreover, available corpora often exhibit domain bias, favoring formal or government text while neglecting conversational, regional, and social media content [21]. Another significant issue lies in data noise—automatic alignments in parallel corpora sometimes produce incorrect mappings, leading to semantic inconsistencies during training [22], [25]. Without systematic data cleaning and domain balancing, even advanced models like mBART and IndicTrans2 struggle to deliver consistent quality across diverse real-world inputs.

- **Linguistic and Structural Complexity**

Indian languages, particularly those in the Indo-Aryan and Dravidian families, display rich morphology, flexible syntax, and context-dependent semantics, posing serious challenges for computational translation [13], [19]. Marathi's agglutinative nature and complex inflectional system make it harder to segment and align compared to English or Hindi [23]. Additionally, word-order differences—Hindi and Marathi following Subject–Object–Verb (SOV) structures versus English's Subject–Verb–Object (SVO)— require sophisticated reordering mechanisms within MT models [19], [26]. Compounding this is the presence of idiomatic and culturally embedded expressions, which often fail literal translation. For instance, Marathi idioms like "हात धुवून मागे लागणे" (literally "to chase after washing hands") meaning "to persistently pursue" cannot be accurately captured without contextual modeling [25]. Such nuances necessitate linguistic feature integration, which is still limited in current neural architectures.

- **Evaluation and Metric Limitations**

While BLEU, chrF, and COMET remain the dominant evaluation metrics, they inadequately capture semantic and cultural correctness for Indic languages [24], [28]. BLEU's reliance on surface-level n-gram overlap penalizes valid translations that differ lexically but remain semantically accurate. For example, the Marathi translations "तो गेला" and "तो फनघून गेला" (both meaning "He left") may receive lower BLEU scores despite conveying equivalent meaning [19]. Furthermore, metric inconsistency across studies—where some employ only BLEU while others use human evaluation or hybrid approaches—creates challenges in benchmarking comparability [23], [27]. To ensure fair evaluation, research must adopt standardized multilingual benchmarks and incorporate semantic-aware metrics that account for linguistic diversity, cultural relevance, and fluency [30].

- **Ethical and Fairness Constraints**

An emerging concern in MT research involves bias and fairness, particularly regarding gendered and socio-cultural translations [24], [29]. Hindi and Marathi, being gendered languages, pose unique risks of gender misrepresentation during translation. For instance, gender-neutral Hindi sentences often default to masculine forms when translated into English [26]. Additionally, lack of dialectal diversity in training corpora can lead to socio-linguistic exclusion—urban or standard variants dominate, while rural and regional dialects remain underrepresented [22]. Addressing these biases requires fairness-aware data collection, balanced fine-tuning, and the inclusion of bias metrics such as Gender Accuracy (GA) and Bias Amplification Score (BAS) during model evaluation [28].

- **Computational and Resource Constraints**

Training state-of-the-art neural models such as Transformer, mT5, and IndicTrans2 requires significant computational resources, often beyond the reach of smaller research institutions [19], [27]. This constraint limits experimentation, replication, and large-scale model fine-tuning across Indian languages. Moreover, due to limited cloud- based translation APIs for Indic languages, many projects rely on academic computing clusters with restricted hardware access. These computational bottlenecks delay progress and prevent fair participation in global MT research benchmarks. Collaborative, open-access initiatives like AI4Bharat and Google Project Udaan are crucial for democratizing computational access and ensuring equitable MT research participation across institutions [24], [30].

9. Future Prospects and Development Trends

The evolution of Machine Translation (MT) in India is entering a transformative era marked by multilingual neural architectures, fairness-driven evaluation, and policy- level emphasis on linguistic inclusion. While existing research has laid strong foundations in data creation and model design, the future of Indic MT lies in advancing scalability, inclusivity, and semantic accuracy [19], [23], [29]. The following subsections outline the emerging development trends and strategic directions for the next generation of Indian language translation systems.

- **Expansion of Multilingual and Cross-Lingual Frameworks**

The most promising direction in Indian MT research is the continued expansion of multilingual and cross-lingual architectures [22], [25]. Models such as IndicTrans2, mBART-50, and mT5 have proven that parameter sharing across languages significantly enhances performance for low-resource pairs like Hindi–Marathi and Hindi–Odia. Future models are expected to move toward universal transformer architectures capable of encoding multiple Indic scripts into a shared semantic space [24], [27]. These architectures will likely employ modular fine-tuning—adapting a single pre-trained model to specific domains or dialects with minimal data. Cross- lingual transfer will thus remain the cornerstone of scaling translation quality for underrepresented Indian languages.

- **Integration of Linguistic and Cultural Knowledge**

Future MT systems must transcend purely statistical and neural representations to include linguistic, cultural, and pragmatic understanding [19], [26]. For example, idiomatic and context-dependent expressions in Hindi and Marathi—such as proverbs or regional idioms—require models to incorporate semantic role labeling, morphological analyzers, and contextual embeddings that preserve cultural meaning. Hybrid models combining linguistic rules with deep neural representations are likely to dominate next-generation systems [23], [28]. Additionally, culturally aware translation mechanisms—capable of recognizing politeness markers, gender nuances, and regional honorifics—will enhance localization accuracy, especially in applications such as education, governance, and public communication [25].

- **Advancements in Fairness, Bias Mitigation, and Ethics**

As MT becomes more integrated into public digital infrastructure, addressing algorithmic bias and fairness will remain a central research and ethical priority [24], [29]. Future systems will employ bias-detection layers, counterfactual data generation, and fairness-aware loss functions to ensure gender, regional, and socio-linguistic neutrality in translation. Evaluation protocols will also expand beyond BLEU and COMET to include fairness metrics such as Gender Accuracy (GA), Bias Amplification Score (BAS), and Representation Diversity Index (RDI) [26]. Collaborative initiatives like AI4Bharat FairMT and WAT-IndicFair will likely establish standardized frameworks for bias auditing in Indic MT. These efforts will ensure that translation systems not only perform linguistically well but also uphold ethical accountability and cultural respect.

- **Data Democratization and Open Resource Ecosystems**

A major trend shaping the next decade will be the democratization of language data through open-source repositories and collaborative frameworks [19], [27]. Initiatives such as AI4Bharat, Project Udaan, and IndicNLP are pioneering open-access platforms that integrate translation data, lexicons, and linguistic tools for public use [23], [30]. The emergence of federated data governance models—where institutions contribute to shared multilingual datasets while maintaining local control—will ensure transparency and inclusivity in data-driven research. Future corpus development will also focus on dialectal representation, domain diversity, and speech-text alignment to enable comprehensive multilingual applications, including speech-to-text and cross-modal translation systems [25].

- **Future of Indic MT: Towards Contextual and Real-Time Translation**

The long-term trajectory of Indic MT is oriented toward real-time, multimodal, and context-sensitive translation ecosystems [21], [26]. The integration of MT into speech recognition, image captioning, and large language models (LLMs) will redefine human– machine interaction in India's multilingual environment. For example, coupling translation models with context-aware LLMs such as GPT-based architectures could enable conversational translation systems capable of adapting tone and formality dynamically. In addition, domain adaptation pipelines—fine-tuning pre-trained models on specific verticals like healthcare, education, or e-governance—will create specialized, high-accuracy translation engines [24], [30]. These innovations will transform MT from a research domain into an essential component of India's digital knowledge infrastructure.

10. Conclusion

The present review comprehensively examined the evolution of Machine Translation (MT) for Indian languages, with a focused analysis on Hindi–English and Hindi–Marathi translation systems. The discussion traced the transition from rule-based and statistical approaches to neural and multilingual frameworks, highlighting significant progress achieved through data expansion, linguistic preprocessing, and model innovation [19], [23]. Despite notable improvements in translation quality and fluency, challenges persist in handling the morphological richness, syntactic flexibility, and semantic diversity characteristic of Indic languages. The comparative findings indicate that while high-resource pairs such as Hindi– English have achieved near-commercial translation accuracy, low-resource combinations like Hindi–Marathi still face substantial performance gaps, primarily due to limited data availability and underdeveloped evaluation benchmarks [24], [26].

A central insight emerging from this review is that linguistic awareness and fairness are as critical as computational efficiency in the design of MT systems [22], [27]. The integration of morphological analyzers, contextual embeddings, and hybrid linguistic-neural models can enhance semantic coherence, particularly in low-resource scenarios. Moreover, recent research underscores the need for bias-sensitive frameworks that monitor gender, dialectal, and socio-linguistic representation during training and evaluation [24], [29]. The inclusion of fairness metrics such as Gender Accuracy (GA) and Bias Amplification Score (BAS), alongside traditional measures like BLEU and COMET, represents a significant step toward ethically aligned and culturally responsible translation systems.

Looking ahead, the future of Indic MT depends on collaborative, multilingual innovation and open data ecosystems. Initiatives such as AI4Bharat, Project Udaan, and IndicTrans2 are redefining the boundaries of inclusivity and transparency in translation research [25], [30]. By combining linguistically grounded modeling, multilingual pretraining, and fairness-driven evaluation, India can build MT systems that serve both local and global linguistic needs. Ultimately, the success of future translation technologies will be measured not only by accuracy but by their ability to preserve the cultural, social, and ethical dimensions of language—ensuring that every Indian language has an equal voice in the digital age.

Beyond the technical advancements, the broader impact of Indic Machine Translation lies in its potential to bridge linguistic divides and democratize access to knowledge across India's multilingual society [19], [24]. As government, education, and digital communication increasingly adopt MT systems, ensuring linguistic equity and inclusivity becomes essential. High-quality translation between Hindi, Marathi, and other Indian languages can foster educational accessibility, empower regional media, and enhance citizen participation in governance.

Furthermore, integrating MT into emerging technologies—such as voice assistants, digital classrooms, and real-time translation apps—can transform the digital ecosystem by making it truly multilingual. The vision for the next decade is clear: to build human-centered, context-aware translation systems that not only convert words but also preserve cultural meaning and identity, ensuring that technology speaks every Indian language with equal clarity and respect [26], [30].

REFERENCES

[1]. Kalele, S., Singh, L., & Kumar, A. (2024, January). Comparative analysis of negation in Marathi and Hindi in context of translation.

[2]. Panchbhai, B., & Pathak, V. (2024, April). A review study of machine translation systems for Indian languages and their issues.

[3]. Walke, P. P., & Haneef, F. (2021, April). A survey on machine translation approaches for Indian languages.

[4]. Babhulgaonkar, A., & Sonavane, S. (2021, October). Empirical analysis of phrase- based statistical machine translation systems for English to Hindi language.

[5]. Bala Das, S., Panda, D., Mishra, T. K., & Patra, B. K. (2024, May). Multilingual neural machine translation for Indic to Indic languages.

[6]. Kumar, P., & Thakur, A. K. (2024). English-to-Hindi translation divergence study of English phrasal verbs.

[7]. Bala Das, S., Biradar, A., Mishra, T. K., & Patra, B. K. (n.d.). Improving multilingual neural machine translation system for Indic languages.

[8]. Shetty, A. P. (2025, January). Evaluating machine translation models for English-Hindi language pairs: A comparative analysis.

[9]. Attri, S. H., Prasad, T. V., & Ramakrishna, G. (2020). HiPHET: Hybrid approach for translating code mixed language (Hinglish) to pure languages (Hindi and English).

[10]. Baruah, N., & Khan, A. (2021). Word-level English to Hindi neural machine translation.

[11]. Makadiya, J. D., & Dave, J. (2023). Natural language processing: Machine translation for Indian languages.

[12]. Suha, N. J., & Khan, M. A. R. (2023). A neural machine translation approach for translating different languages in English.

[13]. Sani, S., & Vijaya, S. (2024). A survey on the MT methods for Indian languages: MT challenges, availability, and production of parallel corpora, government policies, and research directions.

[14]. Philip, J., & Siripragada, S. (2021). Revisiting low resource status of Indian languages in machine

translation.

[15]. Patel, R. N., & Pimpale, P. B. (2025). Statistical machine translation for Indian languages: Mission Hindi.

[16]. Gangar, K., & Ruparel, H. (2021). Hindi to English: Transformer-based neural machine translation.

[17]. Ramanathan, A. (2020). Statistical machine translation for Indic languages: Mission Hindi.

[18]. Singh, K. B., & Singh, N. A. (2023). A comparative study of transformer and transfer learning based MT models for English-Manipuri.

[19]. Durga, M. (2025). English–Hindi neural machine translation system using transformers.

[20]. Ranathunga, S., & Lee, E.-S. A. (2021). Neural machine translation for low- resource languages: A survey.

[21]. Poudel, S., & Bal, B. K. (2024). Bidirectional English-Nepali machine translation system for the legal domain.

[22]. Patil, A., & Joshi, I. (2022). PICT@WAT 2022: Neural machine translation systems for Indic languages.

[23]. P. M., A. (2024). MTNLP-IIITH: Machine translation for low-resource Indic languages.

[24]. Kandimalla, A. (2022). Improving English-to-Indian language neural machine translation systems.

[25]. B. J., V. (2024). Machine translation for low-resource language using NLP attention mechanism.

[26]. Perera, S. (2025). Machine translation and transliteration for Indo-Aryan languages: A systematic review.

[27]. Jian, L. (2022). LSTM-based attentional embedding for English machine translation.

[28]. Kunchukuttan, A., Mehta, P., & Bhattacharyya, P. (2023). IndicTrans2: Towards High- Quality and Open Multilingual Translation for Indian Languages. Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (ACL).

[29]. Ramesh, S., Srivastava, N., & Goyal, V. (2024). Evaluating Transformer Architectures for Low-Resource Indic Machine Translation. Transactions on Asian and Low-Resource Language Information Processing (TALLIP).

[30]. Bhattacharya, D., & Ghosh, S. (2023). Gender and Dialect Bias in Neural Machine Translation for Indic Languages. Journal of Language Technology and AI Ethics, 2(1), 45– 59.

[31]. AI4Bharat. (2024). FairMT Benchmark: Measuring Bias and Fairness in Indian Language Translation Systems. AI4Bharat Technical Reports.