

BVSLP: Enhancing Machine Translation through NMT Fine-Tuning, Back-Translation, and LLM-Assisted Synthetic Data Generation

Palak Arora^{1,2}, Adhikarimayum Meerajita Sharma¹, Mrityunjaya Indoria¹, Keneiwenuo Lhoungu¹,
Lalthafamkimi¹, Bharti Nathani¹, Nisheeth Joshi¹

¹Speech and Language Processing Lab, Centre for Artificial Intelligence, Banasthali Vidyapith, Rajasthan, India

²School of Computing, DIT University, Dehradun, Uttarakhand, India

palak.arora.pa55@gmail.com, meerajitasharma19@gmail.com, mrityunjaya.indoria@gmail.com,
keneiwenuolhoungu@gmail.com, kimkimhnamte24@gmail.com, nbharti@banasthali.in,
nisheeth.joshi@rediffmail.com

Abstract

This paper presents a two-stage neural machine translation framework for English and selected low-resource Northeastern Indian languages. The approach incorporated transfer learning, bidirectional fine-tuning, back-translation, and LLM-assisted synthetic data augmentation. In Stage 1, the NMT model was fine-tuned on cleaned gold-standard parallel data. In Stage 2, additional synthetic sentence pairs were generated using back-translation and Qwen2.5-Instruct, followed by filtering based on language identification, length ratio, duplicate removal, named-entity preservation, semantic similarity, and round-trip consistency. The framework was evaluated on English–Assamese, English–Mizo, English–Manipuri, English–Bodo, English–Khasi, and Nagamese–English translation directions. Results show stronger performance for several Northeastern-language-to-English directions, while the contrastive system notably improved Manipuri (Bengali)–English translation from 26.56 to 30.59 BLEU. The findings indicate that controlled synthetic augmentation can improve low-resource MT, although its effectiveness remains language- and direction-dependent.

1 Introduction

In recent years, we have seen significant improvement in machine translation (MT) technology in particular and natural language processing (NLP) in general. However, most of the Northeastern Indian languages remain underrepresented in the digital space. Their linguistic heterogeneity, multiple scripts, limited

parallel corpora, and inconsistent data quality make the development of reliable MT systems particularly challenging.

Limited availability of quality parallel corpora is one such reason. There are a lot of open source resources which claim to provide parallel corpus for these languages, but they suffer from limited vocabulary coverage, grammatical learning, named-entity handling, and domain-specific translation. Therefore, careful corpus preprocessing, normalization and alignment checking are essential. Multilingual transfer learning can help low-resource languages benefit from shared representations learned across several languages, while back-translation can expand training data by converting monolingual text into synthetic parallel corpora. However, automatically generated data may contain translation errors and must be filtered before use.

Large language models (LLMs) provide an opportunity for low-resource MT, which can generate domain-specific sentences, paraphrases, terminology-rich examples, and corrected synthetic translations. The output of such systems, however, needs strict validation because they may introduce inaccuracies, hallucinations, or culturally inappropriate expressions.

This paper proposes a two-stage NMT framework for English and selected Northeastern Indian languages. In stage 1, an NMT model is trained on a bidirectional gold parallel corpus. The resulting checkpoint is then used for back-translation, while an LLM supports further data augmentation. After semantic, linguistic, script, entity, and quality filtering, the accepted synthetic data is combined with the original corpus for stage 2 training. The framework aims to improve translation quality, vocabulary coverage, domain

diversity, and fluency while maintaining the reliability of human-translated data.

2 Literature Review

Machine translation (MT) research for North-Eastern Indian languages has gradually progressed from linguistically informed statistical systems to neural machine translation (NMT), multilingual pretrained models. The region presents a particularly challenging setting because its languages belong to different linguistic families, use different scripts, and have highly unequal levels of digital representation. Research has primarily concentrated on Assamese, Manipuri, Khasi, Mizo, and Bodo, while Nyishi and Kokborok have gained attention only recently through shared-task initiatives.

Early research focused largely on Manipuri. Singh and Bandyopadhyay (2010a, 2010b) developed statistical translation systems incorporating morphology, syntax, semantic information, and dependency relations, demonstrating that linguistic information could partially compensate for limited parallel data. Laskar et al. (2020) introduced EnAsCorp1.0 for English–Assamese, while Laskar et al. (2021) developed EnKhCorp1.0 for English–Khasi. Laitonjam and Singh (2021) explored comparable corpora for Manipuri–English translation. Larger initiatives such as Samanantar (Ramesh et al., 2022) and IndicTrans2 (Gala et al., 2023) considerably expanded multilingual resources for languages such as Assamese, Manipuri, and Bodo.

Huidrom and Lepage (2022) showed that pretrained embeddings improved Manipuri translation, while Moirangthem et al. (2023) demonstrated the usefulness of subword modelling, iterative back-translation, and synthetic-data filtering. Agrawal et al. (2023), Suman et al. (2023), Singh et al. (2023a, 2023b), and Singh et al. (2024) further explored Transformer and multilingual transfer-learning approaches for Assamese and Manipuri, highlighting the effects of domain, script, training size, and translation direction.

Ahmed et al. (2023) demonstrated iterative back-translation for English–Assamese. Nongbri et al. (2023) reported improvements for English–Khasi through monolingual data augmentation. Similar filtering and back-translation strategies were explored by P. M. et al. (2024) and Oinam and Saharia (2025). These studies showed that

synthetic data can improve MT, but its effectiveness depends on the quality of the initial system, semantic filtering, domain compatibility, and the proportion of synthetic and human-translated examples.

The WMT shared tasks have substantially broadened regional MT research. Pal et al. (2023) established common benchmarks for Assamese, Khasi, Manipuri, and Mizo. Subsequent shared tasks expanded experimentation through multilingual fine-tuning, transfer learning, LoRA, semantic filtering, and domain adaptation (Pakray et al., 2024, 2025; Sahoo et al., 2024; Bhaskar & Krishnamurthy, 2024). More recent research has included Bodo, Nyishi, and Kokborok. Boruah et al. (2023) examined subword segmentation for English–Bodo, while Wary et al. (2025) developed an attention-based Bodo system. J et al. (2025) investigated vocabulary enrichment and multilingual transfer for Nyishi, Khasi, and Kokborok, whereas Kumar et al. (2025) explored instruction-tuned LLMs for Bodo and Kokborok. Yadav and Shrivastava (2025) retained phrase-based SMT as a comparative baseline. Parameter-efficient adaptation has also been explored by Khongthaw et al. (2025), while Joshi et al. (2025) incorporated named-entity translation and transliteration into multilingual MT.

3 Methodology

In this study, a neural machine translation framework is used for translation between English and selected Northeastern Indian languages. The framework is intended for low-resource settings where parallel corpora are limited in size, uneven in quality, and often restricted to a small number of domains. To address these limitations, the methodology adopted various approaches, viz NMT fine-tuning through multilingual transfer learning, followed by back-translation and LLM-assisted synthetic data augmentation and filtering.

The framework was applied to language pairs viz English–Assamese, English–Manipuri(Bangla/Meitei Scripts), English–Mizo, English–Bodo, English–Nagamese. The corpus was provided by the shared task organizers (Pakray et al., 2026). It developed framework supports bilingual configuration where, a separate model was trained for each English–regional-language pair. Bidirectional training was adopted so that each aligned pair contributed examples in both

English-to-regional-language and regional-language-to-English directions.

3.1 Corpus Processing

The parallel corpus was cleaned and normalized. Empty lines, duplicate pairs, and excessively long or short sentence pairs were removed. Additional checks were performed for mismatched numbers, dates, unusual length differences, and encoding problems. Sentence-length ratios were examined because extreme source–target differences indicate misalignment. A representative sample from the cleaned corpus was manually inspected to ensure that the filtering procedure was working correctly.

Each sentence pair was stored in JSONL format with fields such as source, target, src_lang, tgt_lang, domain, and provenance. Provenance labels distinguished human-translated data from synthetic examples generated later in the pipeline. Typical labels included gold, backtranslated, llm_generated, llm_refined, and human_corrected. Every gold parallel pair was represented in both translation directions. The dataset was split into training, validation, and test sets before bidirectional expansion to avoid data leakage. The test set remained completely unseen during training, synthetic-data generation, and checkpoint selection.

3.2 Model Training

3.2.1 Model Training

The NMT component followed a Transformer-based encoder-decoder architecture. NLLB-200 multilingual model was preferred over training from random initialization because low-resource corpora were generally too small to support high-capacity models effectively. The encoder produced contextual representations of the source sequence, while the decoder predicted the target sentence token by token. Source and target language identifiers were used to control the translation direction in multilingual training.

Subword tokenization was used to reduce out-of-vocabulary problems and handle rare, morphologically complex, borrowed, or inflected words. Maximum source and target lengths were selected after examining corpus statistics, with values around 96–128 tokens being suitable for sentence-level training in many cases.

Stage 1 training used only the cleaned gold-standard corpus. During each training step, source

sentences were encoded and target tokens were predicted by the decoder. Mixed precision and gradient checkpointing was enabled to reduce memory consumption. Several checkpoints were saved during Stage 1. The best checkpoint was selected using validation loss and translation metrics, viz BLEU and chrF++. This checkpoint formed the basis for Stage 2.

3.2.2 Synthetic Data Generation

To expand the training corpus, monolingual data was collected for English and the selected Northeastern languages. Back-translation was then used to convert monolingual sentences into synthetic bilingual pairs. A Stage 1 model translated the monolingual sentence into the opposite language, and the generated sentence was paired with the original text.

To expand the training corpus, monolingual data is collected for English and the selected Northeastern languages. Back-translation is then used to convert monolingual sentences into synthetic bilingual pairs. A Stage 1 model translates the monolingual sentence into the opposite language, and the generated sentence is paired with the original text.

Qwen2-5-Instruct with 3 billion parameter was used for augmentation. This generated domain-specific sentences, paraphrases, and corrections to low-quality synthetic translations. Prompts were controlled to avoid unnecessary explanations. Named entities and numbers were preserved for the languages for which we had a POS tagger. We generally looked at name, organization and location NEs for preservation.

As the synthetic data would have had hallucinations, mistranslations, unnatural phrasing, or entity errors. All generated pairs underwent filtering. The filtering stage included language identification, length-ratio checks, duplicate removal, repetition detection, named-entity consistency, number preservation, semantic similarity, round-trip consistency, and quality estimation. Multilingual sentence embeddings were used for calculating cosine similarity between source and target representations. Low-scoring pairs were either rejected or manually reviewed.

Filtered synthetic data was then combined with the original gold corpus. Gold data remained the most reliable source of supervision and was given greater weight so that synthetic examples did not

dominated training. In multilingual settings, balanced or temperature-based sampling was used to prevent languages with larger corpora from overwhelming extremely low-resource languages. Stage 2 training began from the best Stage 1 checkpoint. Since the model had already learned the fundamental language mappings. The augmented corpus exposed the model to broader vocabulary, new domains, paraphrases, and naturally occurring monolingual patterns. Validation was performed using the same development set used in Stage 1, allowing a direct comparison between the gold-only and augmented systems. The final system was selected using validation performance and was evaluated separately for every language and translation direction.

4 Evaluation

Two types of systems were submitted to the shared task: primary and contrastive systems. The primary systems were trained exclusively on the datasets provided by the shared-task organizers, whereas the contrastive systems incorporated additional data collected from publicly available online repositories, including Kaggle and Hugging Face. The performance of these systems across the different translation directions is presented below.

For English-Assamese, the systems produced one of its strongest results, achieving 16.44 BLEU, 38.10 METEOR, 48.64 chrF++, 85.44 BERTScore, and 83.23 COMET, with TER of 66.71. By BLEU, this placed the primary system fourth among the listed primary systems. More importantly, its COMET score of 83.23 was extremely competitive, exceeded only slightly by the best-performing system at 83.31 and higher than the BLEU-leading by another best-performing system’s primary system at 83.17. This suggests that although BVSLP does not always reproduced the reference translation closely at the surface n-gram level, the generated translations preserved semantic content well. The contrastive Assamese system performed worse, confirming that the primary configuration was the more effective variant for this direction.

For Assamese-English, the primary system obtained 20.67 BLEU, 54.38 METEOR, 51.58 chrF++, 94.19 BERTScore, and 84.74 COMET. Although its BLEU value was below the leading systems, the high BERTScore and COMET indicated that the output retained considerable

semantic correspondence with the references. The contrastive version fell to 19.74 BLEU and 82.26 COMET, accompanied by a rise in TER from 72.70 to 80.81. Thus, the primary configuration was better.

The bidirectional English-Mizo results reveal a marked directional difference. English-Mizo achieved 16.29 BLEU and 63.26 COMET, whereas Mizo-English achieved a considerably stronger 24.85 BLEU, 59.44 METEOR, 55.50 chrF++, 94.74 BERTScore, and 78.91 COMET. The contrastive English-Mizo model slightly improved some measures such as chrF++, BERTScore, and COMET, but BLEU declined and TER increased sharply. The Mizo-English contrastive model performed substantially worse, falling from 24.85 to 9.90 BLEU.

One of the most interesting findings appears for English-Manipuri in Bengali script. The primary system achieved 9.18 BLEU, whereas the contrastive system fell to 7.21 BLEU. However, almost every other metric improved in the contrastive system: METEOR rose from 16.71 to 19.17, TER improved from 89.27 to 84.36, chrF++ increased from 35.99 to 39.60, BERTScore rose from 82.97 to 84.54, and COMET increased from 66.76 to 67.85. This was an important result because it demonstrated that BLEU alone would provide an incomplete interpretation. The contrastive output appeared to have weaker exact n-gram correspondence but better character-level, semantic, and edit-based correspondence with the reference.

The reverse Manipuri (Bengali)-English direction was one of the clearest successes. The primary system achieved 26.56 BLEU, while the contrastive system improves to 30.59 BLEU. METEOR increased from 62.47 to 68.09, TER decreased from 63.97 to 59.01, and chrF++ rose from 57.79 to 63.12. BERTScore and COMET also increased slightly to 94.95 and 81.14, respectively. The consistency of improvement across almost all metrics indicated that the modifications introduced in the contrastive system genuinely enhanced translation quality rather than merely optimizing for one evaluation measure.

The Meitei Mayek results showed another strong directional asymmetry. English-Manipuri (Meitei Mayek) obtained only 3.82 BLEU, although its BERTScore of 96.23 and COMET of 87.54 suggested that some semantic information was preserved despite very low lexical overlap. In

contrast, Manipuri (Meitei Mayek)-English performed much better with 24.61 BLEU, 57.99 METEOR, 54.60 chrF++, 94.48 BERTScore, and 76.37 COMET. The contrastive Meitei-English system declined substantially to 15.30 BLEU, indicating that the primary configuration was considerably more reliable for this direction. For Khasi, submitted contrastive systems. English-Khasi achieved 15.69 BLEU, 36.73 METEOR and 59.16 COMET, which represents moderate translation capability. However, Khasi-English was substantially weaker at 8.68 BLEU and 55.42 COMET. The difference suggests that the model configuration used for Khasi does not generalize equally well in both directions and requires further language-specific adaptation.

In Category 2, the English-Bodo contrastive result was comparatively weak, with 5.20 BLEU, 17.52 chrF++, and 39.58 COMET. However, Bodo-English improved markedly to 20.56 BLEU, 55.64 METEOR, 50.81 chrF++, 93.36 BERTScore, and 76.79 COMET. This again supported the broader observation that BVSLP performed better when English was the target language. For Nagamese-English, the primary system obtained 8.18 BLEU, 30.68 METEOR, 32.13 chrF++, 88.95 BERTScore, and 58.93 COMET. This indicates usable semantic transfer but considerable room for improvement in lexical accuracy and fluency relative to stronger systems in this direction.

5 Conclusion

This paper described a two-stage NMT framework for English and selected Northeastern Indian languages using transfer learning, back-translation, and LLM-assisted synthetic data augmentation. In Stage 1, the NMT model was fine-tuned on cleaned gold-standard parallel data, while Stage 2 incorporated filtered synthetic data generated through back-translation and Qwen2.5-Instruct. Quality control included language identification, duplicate removal, length checks, named-entity preservation, semantic similarity, and round-trip consistency.

The results showed that system performance varied considerably across languages and translation directions. English-Assamese produced one of the strongest results, with 16.44 BLEU and 83.23 COMET. In general, translation from Northeastern languages into English performed better than translation in the reverse direction, as

observed for Mizo, Manipuri, and Bodo. The contrastive system produced its clearest improvement for Manipuri (Bengali)-English, where BLEU increased from 26.56 to 30.59. However, contrastive configurations did not consistently improve all language pairs, indicating that synthetic augmentation must be tailored to individual languages and directions. Overall, the findings demonstrate the potential of controlled multilingual NMT and synthetic-data augmentation for low-resource Northeastern Indian language translation while highlighting the need for stronger language-specific preprocessing and filtering.

References

- Agrawal, G., Das, R., Biswas, A., & Thounaojam, D. M. (2023). Neural machine translation for English-Manipuri and English-Assamese. In *Proceedings of the Eighth Conference on Machine Translation* (pp. 931–934). Association for Computational Linguistics.
- Ahmed, M. A., Kashyap, K., Talukdar, K., & Boruah, P. A. (2023). Iterative back translation revisited: An experimental investigation for low-resource English-Assamese neural machine translation. In *Proceedings of the 20th International Conference on Natural Language Processing* (pp. 172–179). NLP Association of India.
- Bhaskar, Y., & Krishnamurthy, P. (2024). Yes-MT’s submission to the low-resource Indic language translation shared task in WMT 2024. In *Proceedings of the Ninth Conference on Machine Translation* (pp. 788–792). Association for Computational Linguistics.
- Boruah, P. A., Talukdar, K., Ahmed, M. A., & Kashyap, K. (2023). Neural machine translation for a low-resource language pair: English-Bodo. In *Proceedings of the 20th International Conference on Natural Language Processing* (pp. 295–300). NLP Association of India.
- Doshi, M., Gaikwad, P., Deoghare, S., & Bhattacharyya, P. (2023). Machine translation advancements for low-resource Indian languages in WMT23: CFILT-IITB’s effort for bridging the gap. In *Proceedings of the Eighth Conference on Machine Translation*. Association for Computational Linguistics.
- Gala, J., Chitale, P. A., Raghavan, A. K., Gumma, V., Doddapaneni, S., Kumar, A., et al. (2023). IndicTrans2: Towards high-quality and accessible machine translation models for all 22 scheduled Indian languages. *Transactions on Machine Learning Research*.

- Huidrom, R., & Lepage, Y. (2022). Introducing EM-FT for Manipuri–English neural machine translation. In Proceedings of the WILDRE-6 Workshop (pp. 1–6). European Language Resources Association.
- J, S., Kankanwadi, D., Mishra, A., & Paul, B. (2025). ANVITA: A multi-pronged approach for enhancing machine translation of extremely low-resource Indian languages. In Proceedings of the Tenth Conference on Machine Translation (pp. 1240–1247). Association for Computational Linguistics.
- Joshi, N., Arora, P., Krishnia, A., Lonchenpa, R., & Vizo, M. (2025). BVSLP: Machine translation using linguistic embellishments for IndicMT shared task 2025. In Proceedings of the Tenth Conference on Machine Translation (pp. 1265–1270). Association for Computational Linguistics.
- Khongthaw, O., Salvin, G. L., Budde, S., Chigwededza, A. N., Malkar, D., & Hingmire, S. (2025). DoDS-IITPKD: Submissions to the WMT25 low-resource Indic language translation task. In Proceedings of the Tenth Conference on Machine Translation (pp. 1248–1252). Association for Computational Linguistics.
- Kumar, D., Singh, K. B., & Ekbal, A. (2025). Tackling low-resource NMT with instruction-tuned LLaMA: A study on Kokborok and Bodo. In Proceedings of the Tenth Conference on Machine Translation (pp. 1215–1221). Association for Computational Linguistics.
- Laitonjam, L., & Singh, S. R. (2021). Manipuri–English machine translation using comparable corpus. In Proceedings of the 4th Workshop on Technologies for MT of Low Resource Languages (pp. 78–88). Association for Machine Translation in the Americas.
- Laskar, S. R., Khilji, A. F. U. R., Kaushik, D., Pakray, P., & Bandyopadhyay, S. (2021). EnKhCorp1.0: An English–Khasi corpus. In Proceedings of the 4th Workshop on Technologies for MT of Low Resource Languages (pp. 89–95). Association for Machine Translation in the Americas.
- Laskar, S. R., Khilji, A. F. U. R., Pakray, P., & Bandyopadhyay, S. (2020). EnAsCorp1.0: English–Assamese corpus. In Proceedings of the 3rd Workshop on Technologies for MT of Low Resource Languages (pp. 62–68). Association for Computational Linguistics.
- Moirangthem, G., Nongbri, L., Salam, S. S., & Nongmeikapam, K. (2023). Impacts of approaches for agglutinative low-resource-language neural machine translation: A case study on the Manipuri–English pair. In Proceedings of the 20th International Conference on Natural Language Processing (pp. 191–201). NLP Association of India.
- Nongbri, L., Moirangthem, G., Salam, S., & Nongmeikapam, K. (2023). Bidirectional neural machine translation using monolingual data for the Khasi–English pair. In Proceedings of the 20th International Conference on Natural Language Processing (pp. 318–325). NLP Association of India.
- Oinam, D. S., & Saharia, N. (2025). DELAB-IIITM WMT25: Enhancing low-resource machine translation for Manipuri and Assamese. In Proceedings of the Tenth Conference on Machine Translation (pp. 1222–1226). Association for Computational Linguistics.
- P. M., A., Shetye, K., & Krishnamurthy, P. (2024). MTNLP-IIITH: Machine translation for low-resource Indic languages. In Proceedings of the Ninth Conference on Machine Translation. Association for Computational Linguistics.
- Pakray, P., Krishna, R., Pal, S., Vetagiri, A., Dash, S., Maji, A. K., et al. (2025). Findings of WMT 2025 shared task on low-resource Indic languages translation. In Proceedings of the Tenth Conference on Machine Translation (pp. 532–553). Association for Computational Linguistics.
- Pakray, P., Pal, S., Vetagiri, A., Krishna, R., Maji, A. K., Dash, S., et al. (2024). Findings of WMT 2024 shared task on low-resource Indic languages translation. In Proceedings of the Ninth Conference on Machine Translation (pp. 654–668). Association for Computational Linguistics.
- Pal, S., Pakray, P., Laskar, S. R., Laitonjam, L., Khenglawt, V., Warjri, S., Dadure, P. K., & Dash, S. K. (2023). Findings of the WMT 2023 shared task on low-resource Indic language translation. In Proceedings of the Eighth Conference on Machine Translation (pp. 682–694). Association for Computational Linguistics.
- Partha Pakray, Santanu Pal, Advaita Vetagiri, Kshetrimayum Boynao Singh, Sandeep Kumar Dash, Arnab Kumar Maji, Saralin A. Lyngdoh, Lenin Laitonjam, Anupam Jamatia, Ajit Das, Sunita Warjri, Uzzal Sharma, Koj Sambyo and Riyanka Manna, “Findings of WMT 2026 shared task on Low-resource Indic Languages Translation”, In Proceedings of the Eleventh Workshop on Machine Translation (WMT) under EMNLP 2026, 28-29 October, 2026, Budapest, Hungary.
- Ramesh, G., Doddapaneni, S., Bheemaraj, A., Jobanputra, M., Raghavan, A. K., Sharma, A., et al. (2022). Samanantar: The largest publicly available parallel corpora collection for 11 Indic languages. Transactions of the Association for Computational Linguistics, 10, 145–162.
- Sahoo, P., Brahma, M., & Desarkar, M. S. (2024). NLIP_Lab-IITH low-resource MT system for

WMT24 Indic MT shared task. In Proceedings of the Ninth Conference on Machine Translation. Association for Computational Linguistics.

Singh, K. B., Singh, N. A., Meetei, L. S., Bandyopadhyay, S., & Singh, T. D. (2023b). NITS-CNLP low-resource neural machine translation systems for the English–Manipuri language pair. In Proceedings of the Eighth Conference on Machine Translation (pp. 967–971). Association for Computational Linguistics.

Singh, K. B., Singh, N. A., Meetei, L. S., Singh, N. J., Singh, T. D., & Bandyopadhyay, S. (2023a). A comparative study of Transformer and transfer-learning MT models for English–Manipuri. In Proceedings of the 20th International Conference on Natural Language Processing (pp. 791–796). NLP Association of India.

Singh, N. J., Singh, K. B., Singh, N. A., Phijam, S., & Singh, T. D. (2024). WMT24 system description for the MultiIndic22MT shared task on Manipuri language. In Proceedings of the Ninth Conference on Machine Translation (pp. 797–803). Association for Computational Linguistics.

Singh, T. D., & Bandyopadhyay, S. (2010a). Statistical machine translation of English–Manipuri using morphosyntactic and semantic information. In Proceedings of the 9th Conference of the Association for Machine Translation in the Americas: Student Research Workshop.

Singh, T. D., & Bandyopadhyay, S. (2010b). Manipuri–English bidirectional statistical machine translation systems using morphology and dependency relations. In Proceedings of the 4th Workshop on Syntax and Structure in Statistical Translation (pp. 83–91).

Suman, D., Mandal, A., Pal, S., & Naskar, S. (2023). IACS-LRILT: Machine translation for low-resource Indic languages. In Proceedings of the Eighth Conference on Machine Translation (pp. 972–977). Association for Computational Linguistics.

Wary, S., Borgoyary, B., Ahmed, A., Sah, M., & Senapati, A. (2025). An attention-based neural translation system for English to Bodo. In Proceedings of the Tenth Conference on Machine Translation. Association for Computational Linguistics.

Yadav, S., & Shrivastava, M. (2025). A preliminary exploration of phrase-based SMT and multi-BPE segmentations through concatenated tokenised corpora for low-resource Indian languages. In Proceedings of the Tenth Conference on Machine Translation (pp. 1253–1258). Association for Computational Linguistics.

Table 1: Evaluation Results of Indic MT Shared Task 2026 for BVSLP

Language Pair	System	BLEU ↑	METEOR ↑	TER ↓	chrF++ ↑	BERTScore ↑	COMET ↑
English - Assamese	Primary	16.44	38.1	66.71	48.64	85.44	83.23
English - Assamese	Contrastive	12.86	33.7	76.53	46.73	85.07	80.58
Assamese – English	Primary	20.67	54.38	72.7	51.58	94.19	84.74
Assamese – English	Contrastive	19.74	53.09	80.81	50.71	93.25	82.26
English – Mizo	Primary	16.29	42.6	74.2	43.98	79.92	63.26
English – Mizo	Contrastive	15.08	42.72	87.24	45.35	80.06	64.12
Mizo – English	Primary	24.85	59.44	65	55.5	94.74	78.91

Language Pair	System	BLEU ↑	METEOR ↑	TER ↓	chrF++ ↑	BERTScore ↑	COMET ↑
Mizo - English	Contrastive	9.9	39.55	103.92	40.05	89.78	60.62
English - Khasi	Contrastive	15.69	36.73	76.82	37.16	76.12	59.16
Khasi - English	Contrastive	8.68	30.84	130.64	32.05	89.26	55.42
English - Manipuri (Bengali)	Primary	9.18	16.71	89.27	35.99	82.97	66.76
English - Manipuri (Bengali)	Contrastive	7.21	19.17	84.36	39.6	84.54	67.85
Manipuri (Bengali) - English	Primary	26.56	62.47	63.97	57.79	94.69	80.74
Manipuri (Bengali) - English	Contrastive	30.59	68.09	59.01	63.12	94.95	81.14
English - Manipuri (Meitei Mayek)	Primary	3.82	12.37	102.3	27.68	96.23	87.54
Manipuri (Meitei Mayek) - English	Primary	24.61	57.99	65.78	54.6	94.48	76.37
Manipuri (Meitei Mayek) - English	Contrastive	15.3	48.61	88.37	46.69	91.4	66.2
English - Bodo	Contrastive	5.2	12.73	95.77	17.52	66.91	39.58
Bodo - English	Contrastive	20.56	55.64	67.96	50.81	93.36	76.79
Nagamese - English	Primary	8.18	30.68	89.11	32.13	88.95	58.93

652

653

654