Computing, Information Systems, Development Informatics & Allied Research Journal
Vol. 16 No. 2, June 2025 - www.cisdijournal.net

# Domain-Specific Neural Translation: A Transformer-Based Model for Medical Terminology Explanation to Hausa Language

Ladoja, Khadijat T., Hamzat, Zaynab Adeola & Waziri Maryam
[1]Department of Computer Science
University of Ibadan
Ibadan, Nigeria
E-mails: ktladoja35@gmail.com; hazel.adeola@gmail.com, maryamwaxeere@gmail.com

## ABSTRACT

Effective medical communication is crucial for diagnosis and patient safety, particularly in multilingual regions like North-West Nigeria. Hausa-speaking patients often face language barriers with English-speaking healthcare providers, resulting in poor health outcomes. Existing machine translation tools lack accuracy for low-resource languages like Hausa in medical contexts. This study uses a fine-tuned variant of the Helsinki-NLP OPUS-MT (en-ha) model using a curated 2,000 sized EMT dataset focused on medical terminology. The resulting transformer-based Neural Machine Translation model achieved strong performance with BLEU: 34.89, CHRF: 58.84, and ROUGE-L: 61.37. Notably, the fine-tuned model achieved a 68% improvement in BLEU score over the baseline OPUS-MT model (20.74).

Keywords: Neural Machine Translation, Medical Terminology, Transformer Model, Low Resource Language

CISDI Journal Reference Format
Ladoja, K.T., Hamzat, Z.A. & Waziri, M. (2025): Domain-Specific Neural Translation: A Transformer-Based Model for Medical Terminology Explanation to Hausa Language. Computing, Information Systems, Development Informatics & Allied Research Journal. Vol 16 No 2, Pp 45-54. Available online at www.isteams.net/cisdijournal. dx.doi.org/10.22624/AIMS/CISDI/V16N4P4

## 1. INTRODUCTION

Machine Translation (MT), a branch of Natural Language Processing (NLP), enables automated translation of text or speech across different languages, thereby supporting multilingual communication [1]. MT has broad applications in education, commerce, entertainment, and healthcare, fostering global collaboration and inclusivity. Over time, the field has seen significant advancements, moving from early rule-based methods to Neural Machine Translation (NMT) systems [2], which apply deep learning techniques to produce more accurate and context-aware translations [3]. However, challenges remain, particularly for low-resource languages, where limited linguistic datasets hinder effective translation.

In healthcare, precise communication is essential for accurate diagnosis, treatment, and patient safety [4]. Medical terminology comprises specialized terms [5], that require precise translation to avoid misinterpretations that could lead to poor health outcomes [6]. Existing MT systems often struggle with domain-specific translations, particularly in low resource languages with limited digital representation, such as Hausa. This gap affects Hausa-speaking patients in Nigeria, particularly in rural areas, where many rely on Hausa for healthcare communication [7].

Hausa is spoken by over 50 million people in Nigeria and West Africa [8], yet remains underrepresented in digital and technological domains. The absence of tailored translation tools presents significant barriers in healthcare settings, where providers must convey critical medical information accurately. Transformer-based models are particularly well-suited for this context due to their ability to model long-range dependencies, learn contextual representations effectively, and generalize well even with moderately sized datasets when fine-tuned appropriately [9]. This study aims to develop a transformer-based NMT model to enhance English-to-Hausa medical translations. By curating a specialized dataset and fine-tuning an existing model, this research seeks to improve translation accuracy, ensuring contextual and cultural relevance. The study contributes to healthcare accessibility, linguistic inclusivity in NLP, and the advancement of machine translation for low-resource languages. Ultimately, it provides a scalable solution for bridging healthcare communication gaps in Hausa-speaking communities.

The remaining section of this paper is structured as follows: The related works is covered in Section II. The methodology used in this study is covered in Section III and the findings and results are covered in Section IV. This study wraps up in Section V.

## 2. LITERATURE REVIEW

Ngo et al. [10] explored multilingual machine translation (MT) for French-Vietnamese and English-Vietnamese language pairs, addressing rare word challenges. They proposed two strategies: dynamically learning word similarities and updating rare word embeddings during training. Additionally, monolingual data augmented synthetic parallel corpora, mitigating data sparsity. Their approach improved BLEU scores by up to 1.62 and 2.54 points over bilingual baselines. Nyoni and Bassett [11] compared cross-lingual learning techniques for three Bantu languages (Shona, isiXhosa, isiZulu) alongside English. With only 30,000 English-isiZulu sentence pairs, they found language similarity crucial in transfer learning, where using isiXhosa and Shona as parent models led to a BLEU score difference of 5.2. Multilingual learning outperformed other methods, improving BLEU scores by 9.9, 6.1, and 2.0 over the baseline.

Kuwanto et al. [12] investigated neural machine translation (NMT) for Kazakh, Gujarati, and Somali, proposing a training curriculum for low-resource settings. Their approach yielded BLEU score improvements of +12.2 for English-Gujarati and +3.7 for English-Kazakh. On supervised data, their method set a state-of-the-art BLEU score of 29.3 for Somali-English. Gordillo et al. [13] developed a neural MT tool for Spanish-English medical translations using induced word alignment. Training for 15 epochs achieved a BLEU score of 88.55 for English-Spanish, while the Scielo Spanish-English corpus, trained for 25 epochs, reached 53.74, demonstrating improvements over convolutional models and Fairseq baselines.

Susanto et al. [14] established an NMT benchmark for four Indonesian local languages (Javanese, Sundanese, Minangkabau, Balinese). They found that the CodeXL training approach surpassed Scratch and PreXL methods, improving cross-lingual signals. CodeXL-trained systems achieved spm200BLEU scores of 23.80 (translation to Indonesian) and 18.15 (translation from Indonesian), with greater benefits for translations into Indonesian.

Tonja et al. [15] examined Wolaytta-English translation, leveraging both authentic and synthetic datasets. Their self-learning approach enhanced BLEU scores by +2.7 for Wolaytta-English and +2.4 for English-Wolaytta over the best baseline. Fine-tuning yielded further gains of +1.2 and +0.6 BLEU, respectively. Rios et al. [16] analyzed domain-adapted multilingual NMT (MNMT) for medical English-Romanian translation. Their in-domain MNMT model outperformed an out-of-domain model across all automatic metrics, reducing errors. They found COMET (Crosslingual Optimized Metric for Evaluation of Translation) demonstrated the strongest correlation with MQM (Multidimensional Quality Metrics) scores, achieving a correlation coefficient of 0.663.

Rodríguez-Miret et al. [17] applied NMT to clinical case corpora, projecting translated annotations onto Catalan texts for validation by clinical experts. The resulting named entity recognition (NER) system achieved 91.7% precision, 90.9% recall, and a 91.3% F1-score, demonstrating potential for multilingual clinical NLP resource development. While these studies demonstrate promising advancements in NMT for low-resource languages, several methodological limitations remain. Many works focus primarily on general-purpose translation tasks and do not address domain-specific challenges like medical terminology, which requires high precision to avoid clinical risks. Furthermore, studies that do engage with medical translation, such as [13] and [16], are largely concentrated on high-resource languages (e.g., Spanish, Romanian), leaving a significant gap in transformer-based medical NMT for underrepresented African languages.

## 3. METHODOLOGY

The methodology employs the Helsinki-NLP OPUS-MT (en-ha) model [18] as the core translation system for English-to-Hausa medical translations. This transformer-based model is leveraged for its effectiveness in low-resource language translation to address the gaps presented by fine-tuning the transformer-based model specifically for English-to-Hausa medical translation. Unlike prior work, this work leverages a domain-specific dataset (Explanation to Medical Terms) to improve contextual relevance and terminology accuracy in a healthcare setting. This work contributes to both the body of low-resource NMT research and the practical need for accessible medical communication tools in Hausa-speaking communities.

### Data Collection and Description

To construct a medically relevant corpus, we began by extracting 2,000 core medical terms from MedlinePlus using keyword matching and concept extraction techniques. These terms formed the basis of the Explain Medical Terms (EMT) dataset, which consists of English medical terms paired with their corresponding lay explanations to support the understanding of clinical language. Unlike traditional corpora, the EMT dataset does not contain bilingual translations. This dataset focuses on improving comprehension of clinical terminology, particularly in the context of pediatric patient–clinician communication. In parallel, we used an English-Hausa parallel corpus sourced from opensubtitles and Hugging Face (https://huggingface.co/datasets/Kumshe/English-Chinese-Hausa_Dataset), to fine-tune the translation model for the English-to-Hausa task.

### Preprocessing and Feature Extraction

To ensure linguistic consistency, preprocessing steps include:

- Text Cleaning: Removal of noise (e.g., extra spaces, HTML tags, special characters).
- Normalization: Standardization of text case and preservation of Hausa diacritics.

- Tokenization: Use of subword-based methods like Byte Pair Encoding (BPE) to handle Hausa's morphological richness.
- Noise Removal & Parallel Alignment: Ensuring well-matched English-Hausa sentence pairs.
- Feature Extraction: TF-IDF vectorization highlights domain-specific terms, while pretrained FastText embeddings capture semantic relationships. Positional encoding preserves sentence structure, and custom embeddings focus on medical terminologies for improved translation.

## Model Development

The translation system was developed using a fine-tuned version of OPUS-MT (en-ha) model trained on the EMT dataset. Fine-tuning was conducted using the dataset available at Hugging Face. This bilingual corpus was used to adapt the OPUS-MT (en-ha) model for English-to-Hausa medical translation, ensuring the model could generalize well to health-related language in low-resource settings. The model uses an encoder-decoder architecture, where the encoder processes input English text into contextualized vector representations, and the decoder generates Hausa translations from those embeddings. The model was paired with a corresponding tokenizer to maintain consistency in preprocessing and vocabulary alignment. GPU acceleration was utilized via PyTorch to enhance computational efficiency during inference. During inference, batches of English text were tokenized and translated using the.generate() method with consistent generation parameters: a maximum sequence length of 128, beam search with four beams, and early stopping enabled. Translations were then decoded and cleaned before evaluation.

- **Dataset Split:** A total of 100,000 from the English-Hausa dataset was partitioned into training, validation, and test subsets. First, 90% was allocated for training and 10% for validation. Then, the training portion was further split, with 90% retained for training and the remaining 10% set aside as the test set.
- **Hyperparameter Tuning:** Hyperparameter tuning was conducted to optimize model performance, focusing on several key parameters that significantly influence the effectiveness of the model. Key parameters are presented in table I

TABLE I: MODEL PARAMETERS

| Parameter | Value |
|---|---|
| Learning rate | 0.0008 |
| Max Epoch | 10 |
| Batch size | 16 |
| Optimizer | Adam optimizer |
| Weight decay | 0.01 |
| Training size | 81000 |
| Validation size | 10000 |
| Testing size | 9000 |

## Evaluation Metrics

Model performance was assessed using:

- **BLEU Score:** Measures translation accuracy.
- **ROUGE Score:** Evaluates contextual accuracy.
- **chrF Score:** Measure character-level translation accuracy.

## 4. RESULTS AND DISCUSSION

The results of this study are presented through various figures and tables that illustrate the performance of the developed model.

### Validation Loss

The validation loss over the epochs is illustrated in Fig. 1. The observed trend indicates a consistent decrease from epoch 1 to approximately epoch 8, suggesting effective learning and generalization to unseen data. The loss curve converges towards a lower value and plateaus around 1.8, implying that additional training may not significantly enhance performance. Validation Scores The validation scores, including BLEU, chrF, and ROUGE scores, as shown in Fig. 1, illustrate the model's performance. These metrics collectively provide insights into the effectiveness of the model in capturing linguistic nuances, maintaining fluency, and ensuring accurate medical term translation.

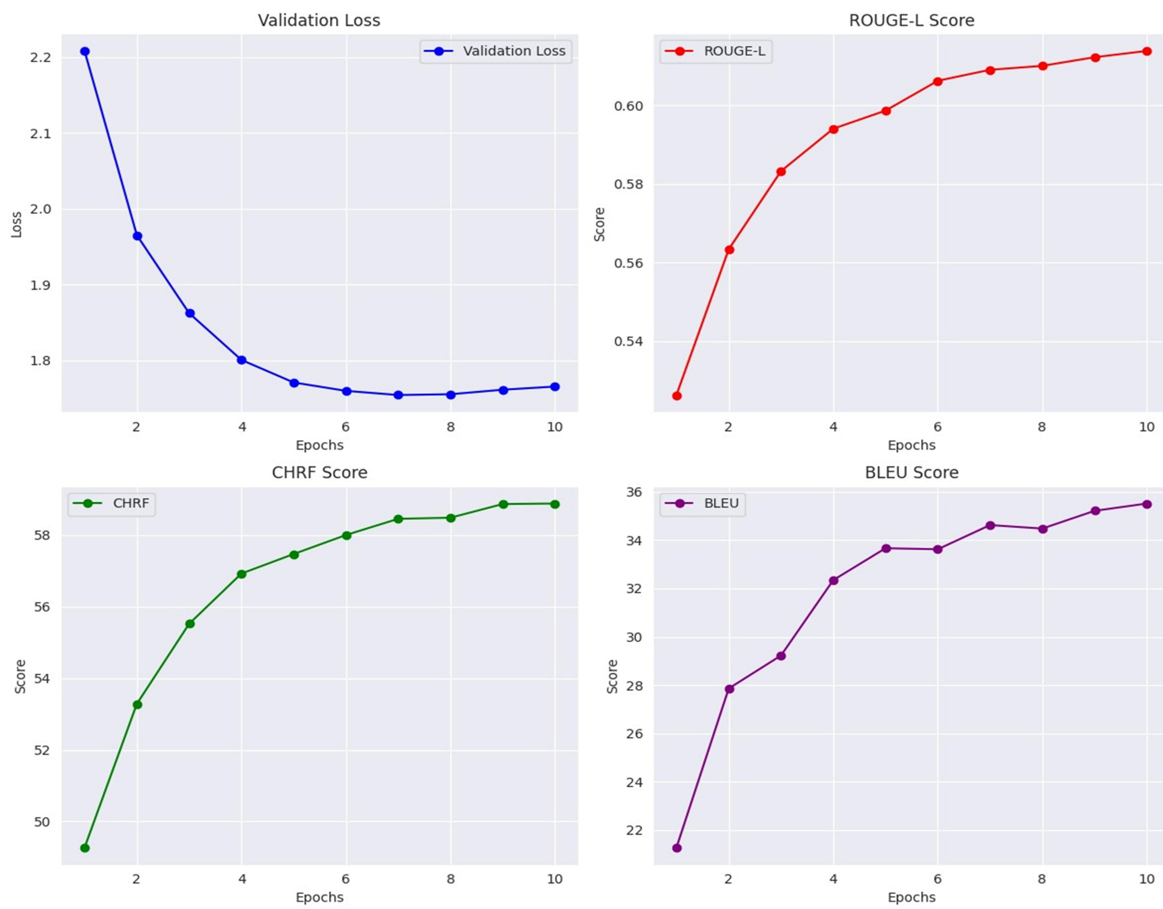


Fig. 1. Model Performance MetricsThe BLEU score demonstrates an upward trajectory throughout training, peaking at the final evaluation. This improvement reflects the model's increasing ability to generate translations closely aligned with human references.

The chrF score shows a steady increase, starting below 50 and reaching around 58 at the final evaluation. The smooth curve indicates consistent improvement across epochs, highlighting the model's enhanced translation fluency and accuracy as training progresses. The ROUGE-L score reflects a positive trend over the training period, beginning at approximately 0.54 at epoch 2 and climbing to around 0.62 by epoch 10. This progression signifies improved alignment with reference translations, affirming the model's effectiveness in generating coherent and contextually appropriate outputs.

### Test Evaluation

The test evaluation results provide insights into model performance on an independent test set. Table II summarizes key training and evaluation metrics over ten epochs. The gradual decline in loss, alongside increasing BLEU, chrF, and ROUGE-L scores, indicates progressive model improvement. Notably, BLEU improves from 21.27 in epoch 1 to 35.52 in epoch 10, reinforcing the model's growing translation accuracy. Test performance results in Table III confirm the model's generalization ability, with a test loss of 1.7595, aligning closely with final evaluation loss. The BLEU score of 34.8862, chrF score of 58.8359, and ROUGE-L score of 0.6137 demonstrate strong translation quality. These metrics, including a test runtime of 798.17 seconds indicate computational efficiency.

TABLE II: TRAINING AND EVALUATION METRICS

| Epoch | Eval Loss | Eval BLEU | Eval CHRF | Eval ROUGE-L |
|---|---|---|---|---|
| 1.0 | 2.2080 | 21.2688 | 49.2651 | 0.5263 |
| 2.0 | 1.9650 | 27.8577 | 53.2834 | 0.5634 |
| 3.0 | 1.8621 | 29.2240 | 55.5203 | 0.5833 |
| 4.0 | 1.8004 | 32.3456 | 56.9254 | 0.5941 |
| 5.0 | 1.7706 | 33.6664 | 57.4720 | 0.5987 |
| 6.0 | 1.7596 | 33.6236 | 58.0030 | 0.6063 |
| 7.0 | 1.7541 | 34.6257 | 58.4563 | 0.6091 |
| 8.0 | 1.7551 | 34.4803 | 58.4872 | 0.6101 |
| 9.0 | 1.7611 | 35.2150 | 58.8692 | 0.6123 |
| 10.0 | 1.7653 | 35.5172 | 58.8837 | 0.6139 |

TABLE III: TEST RESULTS

| Test Loss | Test BLEU | Test CHRF | Test ROUGE-L | Test Runtime |
|---|---|---|---|---|
| 1.7595 | 34.8862 | 58.8359 | 0.6137 | 798.173 |

### Sample Translations

Table IV below presents example outputs generated by the translation models and validated by a human evaluator, for selected English medical sentences.

TABLE IV: SAMPLE TRANSLATIONS

| Original Text | Translated Text | Reference Translations |
|---|---|---|
| He was told the unusual chest pain might be related to a condition called Cardiomyopathy. | An gaya masa zafin kirji na ba a sani ba na iya danganta shi da yanayin da ake kira Carylosy. | Cardiomyopathy: Lokacin da zuciya ta yi rauni ko lalacewa, ba zai iya fitar da jini yadda yakamata ba.Wannan na iya haifar da alamu kamar gajiya, ɗan gajeren numfashi, da walƙiya a cikin kafafun da ba a warware su ba, zai iya haifar da rashin nasara. |
| Membranous Nephropathy was mentioned in the report, and the doctor said more tests were needed. | An ambaci Membranous nephropathy a cikin rohoto, Kuma likita ya ce ana bukatar karin gwaje-gwaje. | Membranous Nephropathy: Wata cuta ce ta koda da ke haifar da canje-canje da kumburi a cikin tsarin koda da ke taimakawa tace datti da ruwa. Kumburin na iya haifar da matsala wajen aikin koda. |

### Comparison with Other Models

The fine-tuned model was evaluated against three established translation models to benchmark its performance:

- **OPUS-Base**: The original Helsinki-NLP model developed for general-purpose English–Hausa translation.
- **T5-Base**: A multilingual sequence-to-sequence model from Google's Text-to-Text Transfer Transformer (T5) family.
- **M2M100 (418M)**: A many-to-many multilingual translation model developed by Facebook AI, capable of direct translation between 100 languages.

Among the models tested, the fine-tuned model demonstrated significantly superior performance. The comparison results, measured using the BLEU score metric, are presented in fig. 2.

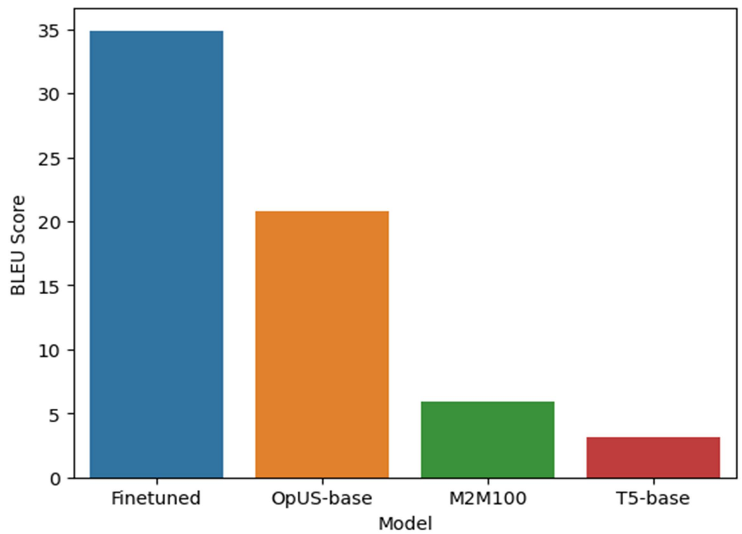


Fig. 2. BLEU Scores by model

## 5. CONCLUSION

This study developed and evaluated a transformer-based machine translation model for English–Hausa medical terminology to improve healthcare communication for Hausa-speaking communities. Built on the Helsinki-NLP OPUS-MT (en-ha) model, the system was fine-tuned using a curated EMT corpus and public bilingual datasets. Key challenges addressed included limited data availability, linguistic complexity, and cultural context. Through rigorous preprocessing and domain-specific adaptation, the model achieved strong results: BLEU score of 35.52, CHRF of 58.88, and ROUGE-L of 0.6139, outperforming baseline models. Human evaluation further confirmed its accuracy and contextual relevance. This research marks a significant contribution to low-resource machine translation and healthcare NLP, offering a foundation for inclusive, AI-driven health communication tools. Future work should focus on expanding the dataset, integrating speech components, and incorporating expert validation.

## REFERENCES

[1] M. S. H. Ameur, F. Meziane, & A. Guessoum (2020). Arabic Machine Translation: A survey of the latest trends and challenges. *Computer Science Review*, *38*, 100305. https://doi.org/10.1016/j.cosrev.2020.100305

[2] Z. Tan, S. Wang, Z. Yang, G. Chen, X. Huang, M. Sun, & Y. Liu (2020). Neural machine translation: A review of methods, resources, and tools. *AI Open*, *1*, 5–21. https://doi.org/10.1016/j.aiopen.2020.11.001

[3] M. Ashraf, (2024). Innovations and Challenges in Neural Machine Translation: A Review. *International Journal of Science and Research (IJSR)*, *13*(10), 656–662. https://doi.org/10.21275/SR241008113336

[4] I. J. Udo, & M. E. Ekpenyong, (2020). Improving Emergency Healthcare Response using Real-Time Collaborative Technology. *Proceedings of the 4th International Conference on Medical and Health Informatics*, 165–173. https://doi.org/10.1145/3418094.3418118
[5] B. Džuganová, (2019). Medical language – a unique linguistic phenomenon. *JAHR*, *10*(1), 129–145. https://doi.org/10.21860/j.10.1.7
[6] A. Tiwary, A. Rimal, B. Paudyal, K. R. Sigdel, & B. Basnyat. (2019). Poor communication by health care professionals may lead to life-threatening complications: examples from two case reports. *Wellcome Open Research*, a*4*, 7. https://doi.org/10.12688/wellcomeopenres.15042.1
[7] A. Bashir, & M. A Umar, (2024). Hausa in Hospitals: Exploring Healthcare Communications in Yariman Bakura Specialist Hospital Gusau, Zamfara State, Nigeria. *Middle East Research Journal of Linguistics and Literature*, *4*(02), 35–42. https://doi.org/10.36348/merjll.2024.v04i02.003
[8] P. J. Jaggar (2006). Hausa. In *Encyclopedia of Language & Linguistics* (pp. 222–225). https://doi.org/10.1016/B0-08-044854-2/02071-X
[9] R. Gruetzemacher, & D. Paradice. (2022). Deep transfer learning & beyond: Transformer language models in information systems research. *ACM Computing Surveys (CSUR)*, *54*(10s), 1-35.
[10] T. V. Ngo, P. T. Nguyen, T. L. H, K. Q. Dinh, & L. M. Nguyen. (2020). Improving Multilingual Neural Machine Translation For Low-Resource Languages: French, English - Vietnamese. *Proceedings of the 3rd Workshop on Technologies for MT of Low Resource Languages*, 55–61. https://doi.org/10.18653/v1/2020.loresmt-1.8
[11] E. Nyoni, & B. A. Bassett, (2021). *Low-Resource Neural Machine Translation for Southern African Languages*. http://arxiv.org/abs/2104.00366
[12] G. Kuwanto, A. F. Akyürek, I. C. Tourni, S. Li, A. G. Jones, & D. Wijaya. (2021). *Low-Resource Machine Translation Training Curriculum Fit for Low-Resource Languages*. http://arxiv.org/abs/2103.13272
[13] A. Gordillo, A. Uyaguari, & L. Garces (2022). Neural Machine Translation tool from Spanish to English in the Medical Domain. *2022 Third International Conference on Information Systems and Software Technologies (ICI2ST)*, 79–84. https://doi.org/10.1109/ICI2ST57350.2022.00019
[14] L. Susanto, R. Diandaru, A. Krisnadhi, A. Purwarianti, & D. Wijaya (2023). *Replicable Benchmarking of Neural Machine Translation (NMT) on Low-Resource Local Languages in Indonesia*. http://arxiv.org/abs/2311.00998
[15] A. L. Tonja, O. Kolesnikova, A. Gelbukh, & G. Sidorov (2023). Low-Resource Neural Machine Translation Improvement Using Source-Side Monolingual Data. *Applied Sciences*, *13*(2), 1201. https://doi.org/10.3390/app13021201
[16] M. Rios, A. Secar˘, R. M. Chereji, & C. Dragos (2023). *Quality Analysis of Multilingual Neural Machine Translation Systems and Reference Test Translations for the English-Romanian language pair in the Medical Domain*. https://dl.fbaipublicfiles.com/fairseq/

[17] J. Rodriguez-Miret, E Farre-Maduell, S. Lima-Lopez, L.Vigil, V.Briva-Iglesias, & M.Krallinger (2024). Exploring the Potential of Neural Machine Translation for Cross-Langauge Clinical Natural Language Processing (NLP) Resource Generation through Annotation Projection. *Information*, 15(10), 585.
https://doi.org/10.3390/info15100585
[18] J. Tiedemann, M. Aulamo, D.Bakshandaeva, M. Boggia, S. A. Gronroos, T. Nieminen, A. Raganato, Y. Scherrer, R. Vazquez, & S. Virpioja (2022). *Democratizing Neural Machine Translation with OPUS-MT*
https://arxiv.org/abs/2212.01936