To optimize, or not to optimize, that is the question: TelU-KU models for WMT21 Large-Scale Multilingual Machine Translation

Sari Dewi Budiwati[1,2]*, **Tirana Noor Fatyanosa**[1]*, **Mahendra Data**[1,3], **Dedy Rahman Wijaya**[2], **Patrick Adolf Telnoni**[2], **Arie Ardiyanti Suryani**[4], **Agus Pratondo**[2], **Masayoshi Aritsugi**[5]

[1]Graduate School of Science and Technology, Kumamoto University, Japan
[2]School of Applied Science, Telkom University, Indonesia
[3]Faculty of Computer Science, Brawijaya University, Indonesia
[4]School of Informatics, Telkom University, Indonesia
[5]Faculty of Advanced Science and Technology, Kumamoto University, Japan
{saridewi, fatyanosa, mahendra.data}@dbms.cs.kumamoto-u.ac.jp
{saridewi, dedyrw, patrickadolf, ardiyanti, pratondo}@telkomuniversity.ac.id
mahendra.data@ub.ac.id, aritsugi@cs.kumamoto-u.ac.jp

Abstract

We describe TelU-KU models of large-scale multilingual machine translation for five Southeast Asian languages: Javanese, Indonesian, Malay, Tagalog, Tamil, and English. We explore a variation of hyperparameters of flores101_mm100_175M model using random search with 10% of datasets to improve BLEU scores of all thirty language pairs. We submitted two models, TelU-KU-175M and TelU-KU-175M_HPO, with average BLEU scores of 12.46 and 13.19, respectively. Our models show improvement in most language pairs after optimizing the hyperparameters. We also identified three language pairs that obtained a BLEU score of more than 15 while using less than 70 sentences of the training dataset: Indonesian-Tagalog, Tagalog-Indonesian, and Malay-Tagalog.

1 Introduction

This paper describes our participation in the WMT21 shared task of large-scale multilingual machine translation. Specifically, we chose small track #2, which involves thirty language pairs, and used the neural machine translation (NMT) method. We call our models TelU-KU (Telkom University - Kumamoto University) as we use our university name in our submissions.

*These authors contributed equally. Everyone contributed to writing this paper.

NMT has been widely used in machine translation research for many languages. Currently NMT has become the state of the art of machine translation with a large number of parallel corpus (Bojar et al., 2017; Nakazawa et al., 2017; Chu and Wang, 2018; Sutskever et al., 2014). Meanwhile, for low resources cases, the NMT tends to give poor translation results (Duh et al., 2013; Sennrich and Zhang, 2019; Zoph et al., 2016; Koehn and Knowles, 2017). In order to get better translation results of NMT for low resources languages, some approaches are applied, such as using a large number of monolingual corpora (Artetxe et al., 2018a,b; Lample et al., 2018b,a), applying transfer learning approach to share lexical and sentence level representation (Gu et al., 2018), using sub-word representation (Durrani et al., 2019), and hyperparameter optimization (HPO) (Sennrich and Zhang, 2019; Rubino et al., 2020).

HPO is an important part of building an NMT system in many real-world applications. In other words, selecting effective hyperparameters is critical to building a strong NMT system. However, in many cases, hyperparameters are often set manually based on intuition and heuristics mechanisms, tedious and error-prone processes that can lead to unreliable experimental results and poor performance of shared tasks or production systems.

Proceedings of the Sixth Conference on Machine Translation (WMT), pages 387–397
November 10–11, 2021.

This is because HPO requires rigorous testing and resources, which makes it a high-cost process. To deal with this problem, table-lookup has been proposed as a benchmark procedure (Zhang and Duh, 2020). Their study provides evaluation protocols and a benchmark dataset for comparing the HPO methods. Moreover, like other NMT models, transformers require setting various hyperparameters, but researchers often use default parameters, even when their data conditions differ substantially from the data conditions previously used to determine those default values.

In low-resource languages cases, the performance of the transformer is highly dependent on the hyperparameter settings (Araabi and Monz, 2020). The experimental results show that the best-suited combination of hyperparameters and regularization methods can produce substantial improvement for low-resource languages data. On the other side, grid search and manual search are the most frequently used strategies for HPO. However, according to the experiment, random search is actually better than grid search in several conditions (Bergstra and Bengio, 2012). Random searches are actually better suited to running on a cluster of computers than grid searches when a group of computers fails. Random search also allows the experimenter to change the "resolution" on the fly. In addition, they have advantages in high-dimensional searching spaces.

In this work, we experimented with two models, that is, TelU-KU-175M and TelU-KU-175M_HPO. Both models are based on a pre-trained model of flores101_mm100_175M. The TelU-KU-175M is our model that manually fine-tuning the hyperparameters, whereas the TelU-KU-175M_HPO is based on hyperparameter optimization. We used a random search method while using 10% of datasets to find the best hyperparameter optimization. In addition, we also included the M2M-100 175M model to compare with our results. This model uses the same pre-trained model as ours but without fine-tuning. Fine-tuning is a common practice in NLP to train a pre-trained model for several epochs on a downstream dataset and has proven to improve performance.

Our experimental results show improvement in most language pairs after optimizing the hyperparameters. The TelU-KU-175M is able to improves the average BLEU scores by 0.35-0.59 over M2M-100 175M. Meanwhile, the TelU-KU-175M_HPO improve the scores by 1.08-1.41 over the baseline. We also identified three language pairs that obtained a BLEU score of more than 15 while using less than 70 sentences of the training dataset: Id-Tl, Tl-Id, and Ms-Tl.

This paper is organized as follows. Section 2 explains the experiment. Section 3 shows the obtained results. Section 4 discusses the effect of HPO and the multilingual model. Section 5 provides the conclusion and future direction of this work.

2 Experiments

In this section, we first describe languages overview of the Southeast Asian language. Then, we discuss data and preprocessing. Finally, we discuss the model and architecture of our model submission.

2.1 Languages overview

We chose small track #2, which involves six languages from Southeast Asia, namely, Javanese (Jv), Indonesian (Id), Malay (Ms), Tagalog (Tl), Tamil (Ta), and English (En).

Indonesian and Malay are considered closely related languages due to being mutually intelligible in morphology, and both languages belong to the Malayo-Polynesian language family (Susanto et al., 2012). The base of formal Indonesian is from Malayo-Riau (Abas, 1987). The main difference is the influence of the vocabulary. Indonesian is largely influenced by Dutch, whereas English influences Malay. The Tagalog language has the same language family as Indonesian and Malay. However, it has different morphology characteristics, and the vocabulary is influenced by several countries, such as Spain, America, and Malay. Javanese is one of the Indonesian ethnic languages used by more than 42% of Indonesia's population, mostly from the central and eastern parts of Java (Novitasari et al., 2020). Javanese is also used in Suriname and New Caledonia. Currently, the Javanese is influenced by Indonesian. This is because Indonesian is used in

Language	Family	Alphabet
Indonesian	Malayo-Polynesian	Latin
Malay	Malayo-Polynesian	Latin
Tagalog	Malayo-Polynesian	Latin
Tamil	Dravidian	Tamil
English	Germanic	Latin
Javanese	Malayo-Polynesian	Latin

Table 1: Language family and writing system.

formal documents and as a daily conversation. Last, Tamil belongs to Dravidian, a unique family where the language is mostly spoken in a southern state (Tamil Nadu) of India (Kumar and Singh, 2019). Table 1 shows the general characteristic of the selected languages.

2.2 Data and preprocessing

We used a dataset provided by the WMT21 organizers. Thus, our system was considered a constrained system. We used three types of datasets, that is, training, evaluation, and hidden test datasets. The training dataset is a parallel corpus from Opus monolingual and Wikipedia, as shown in Table 2. The evaluation dataset is a parallel corpus from Flores101 (Goyal et al., 2021). The evaluation dataset consists of two evaluations, that is, *dev* and *devtest*, as much as, 997 and 1,012 sentences, respectively. Last, the hidden test dataset is an unknown parallel corpus provided by the organizer through the Dynabench leaderboard.[1]

We tokenized all the training and evaluation datasets by SentencePiece tokenizer (Sennrich et al., 2016). This tokenizer is an unsupervised text tokenizer and detokenizer, where the vocabulary size is predetermined prior to the neural model training. We preprocessed dataset according to the guideline,[2] that is, encode and binarize.

2.3 Models & Architectures

We use an NMT system with big Transformer architecture (Ng et al., 2019; Vaswani et al., 2017), i.e., *transformer_wmt_en_de_big*, as implemented in the Fairseq toolkit (Ott et al., 2019). We run experiments on Standard NC24 of Microsoft Azure virtual machine consisting of 4 NVidia Tesla K80 with 12 GB GPU memory.[3] We experimented with the following two models:

Language pairs	Sentences
En - Id	1,019,169
En - Jv	13,049
En - Ms	120,016
En - Ta	95,162
En - Tl	75,447
Id - Jv	42
Id - Ms	1,167
Id - Ta	24,648
Id - Tl	56
Jv - Ms	18
Jv - Ta	1,296
Jv - Tl	2,251
Ms - Ta	3,920
Ms - Tl	5
Ta - Tl	1,478

Table 2: Training datasets of each language pair.

- **TelU-KU-175M** is a pre-trained flores101_mm100_175M with fine-tuning. We manually tune the hyperparameters, as shown in Table 3, column 4.
- **TelU-KU-175M_HPO** is a pre-trained flores101_mm100_175M with HPO. The hyperparameters and their ranges are shown in Table 3. Some of these hyperparameters are based on (Ravikumar, 2020). Figure 1 shows the logical flow of our approach. We run 30 iterations of random searches for two epochs. Due to costly training, we only run the optimization using only 10% *training*, *dev*, and *devtest*. From those 30 models, we select the best model based on the results from the *devtest*. Then, we use the hyperparameter from the best models to fine-tune the flores101_mm100_175M model. The hyperparameter optimization results are shown in Table 3, column 5.

3 Results

We evaluate the generated texts of our models using the sentence-piece BLEU (spBLEU). The spBLEU uses a SentencePiece tokenizer with 256k tokens, and then the BLEU score is computed on the SentencePiece tokenized text. The results are shown in Table 4.

[1] https://www.dynabench.org/flores
[2] https://github.com/facebookresearch/flores
[3] The source code of our experiments is available at https://github.com/fatyanosa/WMT21

Hyper-parameter	Definition	Range	Hyperparameter value	
			TelU-KU-175M	TelU-KU-175M_HPO
BS	Batch size	Min: 8, Max: 128	128	15
LR	Learning rate	Min: 3e-05, Max: 3e-04	3e-05	0.000181463
BT1	Beta1	Min: 0.7, Max: 0.9999	0.9	0.745923
BT2	Beta2	Min: 0.7, Max: 0.9999	0.98	0.948909
EPS	Epsilon	Min: 9.98e-09, Max: 9.99e-06	1e-06	9.62e-06
WD	Weight Decay	Min: 0.0, Max: 0.018	0.0	0.00946414
AD	Attention Dropout	Min: 0.0, Max: 0.5	0.1	0.182843
DR	Dropout	Min: 0.0, Max: 0.5	0.3	0.0162452
SE	Seed	Min: 0, Max: 300	222	72

Table 3: Hyperparameter range for each hyperparameter and the values for each model.

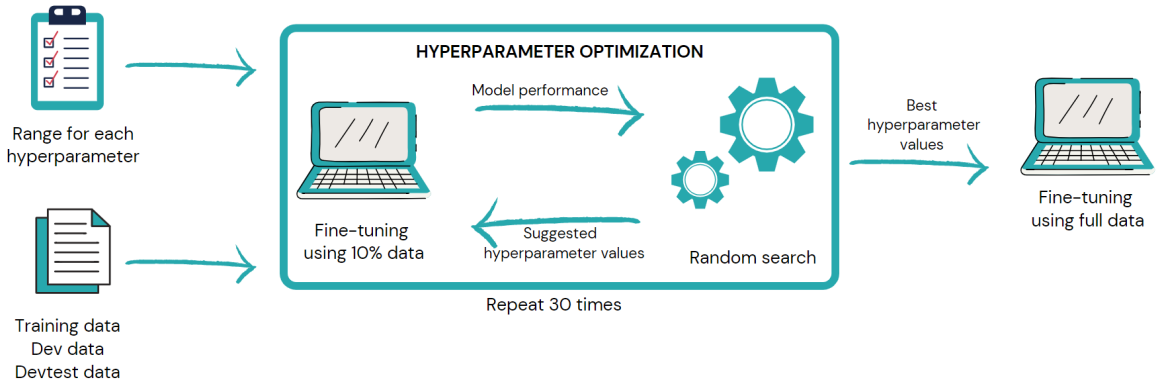

Figure 1: Logical flow of hyperparameter optimization.

We also compare our results with the pretrained model without fine-tuning (M2M-100 175M) as the baseline. Our models: the TelU-KU-175M and TelU-KU-175M_HPO, improved the BLEU scores on 16-17 language pairs over the M2M-100 175M model. In terms of the improvement in each language pair, TelU-KU-175M and TelU-KU-175M_HPO improved BLEU scores by 0.04-5.39 and 0.01-12.22, respectively. The BLEU scores also decreased on several language pairs between 0.1 to 3.25 and 0.02 to 6.5 for TelU-KU-175M and TelU-KU-175M_HPO, respectively.

4 Discussion

This section discusses the effect of HPO and NMT of the multilingual model against our models' evaluation results.

4.1 Effect of HPO

The main objective of HPO in this task is to explore a high-dimensional search space in NMT. As we mentioned in Section 2.3, we run HPO using random search for 30 iterations using 10% of the dataset. The best hyperparameter values were determined based on *devtest*. The best configurations based on the average BLEU scores for each iteration were saved for running the pretrained model, M2M-100 175M, using full datasets (TelU-KU-175M_HPO). The detail of the best configurations is shown in Table 3. The evaluation results were conducted by translating the *dev*, *devtest*, and *test* datasets. We show our detailed evaluation results in Table 4, while Table 5 is our average evaluation among other participants.

We found that fine-tuning of 10% dataset does not lead to the best results for all language pairs translation compared to the manual hyperparameter tuning (TelU-KU-175M) and fine-tuning using 100% dataset (TelU-KU-175M_HPO). The best HPO using 10% of datasets resulted in an average BLEU of 12.75 for dev, 12.33 for devtest, and 12.39 for test datasets. Nevertheless, these values are higher compared to the baseline (M2M-100 175M) in Table 5. This means that fine-tuning with only 10% dataset using a basic method, random search (Bergstra and Bengio, 2012), indeed increases the BLEU scores.

Lang pairs	Dev			Devtest			Test		
	M2M-100 175M	TelU-KU-175M	TelU-KU-175M_HPO	M2M-100 175M	TelU-KU-175M	TelU-KU-175M_HPO	M2M-100 175M	TelU-KU-175M	TelU-KU-175M_HPO
En-Id	28.13	33.36	39.58	28.25	33.34	40.47	28.82	32.97	40.33
Id-En	28.42	30.6	35.54	26.92	29.55	35.33	27.48	29.62	35.1
En-Jv	10.13	8.79	7.32	9.79	8.67	7.02	9.5	8.45	7.2
Jv-En	14.98	14.55	11.01	14.62	14.12	11.43	15.12	13.83	10.59
En-Ms	25.86	24.05	27.35	25.13	22.98	27.39	26.32	23.98	27.63
Ms-En	27.45	28.48	31.16	25.89	27.1	29.68	27.06	27.83	30.8
En-Ta	3.83	2.44	2.27	3.43	2.4	2.41	3.82	2.34	2.07
Ta-En	4.76	5.9	4.97	4.29	5.25	4.41	4.71	5.59	4.42
En-Tl	11.16	16.11	21.16	10.46	15.85	21.38	10.67	16.16	21.18
Tl-En	20.44	22.68	24.88	17.94	21.08	23.59	19.26	22.06	24.4
Id-Jv	12.42	10.72	7.24	12.24	10.92	7.79	11.65	10.1	6.7
Jv-Id	17.89	16.9	16.13	18.22	17	16.09	17.9	16.04	15.44
Id-Ms	26.33	23.53	25.21	25.85	22.92	22.32	26.61	23.36	24.08
Ms-Id	28.99	29.03	29.76	28.64	28.54	29.96	28.71	28.74	29.1
Id-Ta	1.02	2.23	2.35	0.89	2.19	2.23	1.02	2.01	2.31
Ta-Id	4.05	4.34	4.12	3.75	4.36	3.9	3.82	4.24	3.89
Id-Tl	8.11	12.98	18.43	7.41	12.4	17.95	7.75	12.78	17.94
Tl-Id	15.73	17.51	20.66	14.85	16.68	20.43	15.91	17.35	19.98
Jv-Ms	15.19	12.99	8.77	14.21	12.82	7.71	14.61	12.84	8.46
Ms-Jv	11.1	9.64	8.43	10.01	9.2	7.81	9.94	8.68	7.12
Jv-Ta	2.48	1.5	1.08	2.32	1.15	1.04	2.49	1.27	1.02
Ta-Jv	0.88	1.39	0.89	0.7	1.31	1.02	0.76	1.23	0.81
Jv-Tl	8.37	8.37	9.1	7.78	8.24	8.81	7.86	8.41	8.65
Tl-Jv	8.12	7.56	5.59	7.58	7.3	4.82	7.91	7.15	5.25
Ms-Ta	2.71	2.81	3.77	2.29	2.53	3.86	2.64	2.53	3.64
Ta-Ms	3.78	3.9	2.1	3.46	3.71	2.29	3.64	3.9	2.18
Ms-Tl	9.57	12.48	16.42	8.88	11.9	16.3	8.91	12.03	16.09
Tl-Ms	14.59	14.24	14.57	12.53	12.24	11.64	13.5	13.12	12.92
Ta-Tl	2.63	3.03	4.62	2.67	3.31	4.35	2.62	3.04	4.17
Tl-Ta	2.64	2.27	2.57	2.4	2.12	2.24	2.42	2.12	2.31

Table 4: Summary of results for all language pairs based on BLEU scores. Blue font means that there is an improvement over the M2M-100 175M model, while red font means a decline over the M2M-100 175M model.

The BLEU scores improved even more after using the full dataset with the same hyperparameter values (TelU-KU-175M_HPO). This means that the number of datasets influences the performance. We left for future work discussing the effect of the number of datasets in HPO for NMT.

We also study the hyperparameter importance of all optimized hyperparameters using Hyanova[4], a python implementation of a functional analysis of variance (fANOVA) algorithm (Hutter et al., 2014). The algorithm partitions the observed variation of a response value into components against its inputs (Klein and Hutter, 2019). In this study, the response value is the BLEU score, while hyperparameters are the inputs. The higher the fANOVA values, the more important the hyperparameter. Table 6 shows that LR is the most important hyperparameter, while BS is the least important. This means that the LR influence the achieved BLEU scores. From our observation, within our selected range in Table 3, the higher the learning rate, the higher the average BLEU score. Therefore, it is important to tune the LR within a higher range.

Furthermore, we investigate the statistical significance across language pairs from Table 4 using Wilcoxon signed-rank test with $\alpha = 0.05$. We show the p-values of all models in Table 7. Unfortunately, all the results demonstrate statistically non-significant as all of the p-values were more than 0.05. Although the average BLEU can be increased by optimizing the hyperparameter values, this finding shows that HPO might not contribute much to the performance.

One of the possible causes is the utilization of random search, which is categorized as an uninformed search. This category does not learn from previous results, and therefore, each solution is independent of the other. Moreover, uninformed search is proven to be inferior to the informed search, i.e.,

[4] https://pypi.org/project/hyanova/

	Model	Average Bleu Score		
		Dev	Devtest	Test
Other participants	DeltaLM+Zcode (Microsoft-Small)	34.09	33.94	33.89
	615m (Baohao Liao)	33.74	33.51	33.34
	TenTrans (Wanying Xie)	29.25	28.94	28.89
	adaavg (Danni Liu)	28.70	29.09	28.64
	huawei-tsc1 (huaweitsc)	28.64	28.34	28.40
	srph-large (jcblaisecruz02)	22.92	23.14	22.97
	finetune-saptarashmi (saptab)	16.05	15.45	15.72
	615m-new (zizhenlian)	15.50	14.89	15.10
Ours	TelU-KU-175M_HPO	13.57	13.19	13.19
	TelU-KU-175M	12.81	12.37	12.46
Baseline	M2M-100 175M	12.39	11.78	12.11

Table 5: Average BLEU scores from all submitted systems.

Hyper-parameter	Importance value
BS	0.94
LR	1.02
BT1	1.00
BT2	1.00
EPS	1.00
WD	1.00
AD	1.00
DR	1.00
SE	0.99

Table 6: Hyperparameter importance.

bayesian optimization or evolutionary algorithm (Fatyanosa and Aritsugi, 2020, 2021).

This work only calculates the statistical significance across language pairs and leaves the calculation per language pair for future studies.

4.2 Effect of multilingual model

The TelU-KU-175M and TelU-KU-175M_HPO models produced 16-17 language pairs that have higher BLEU scores compared to M2M-100 175M, as shown in Table 4 with blue colors. Among them, we identified seven language pairs that obtained a BLEU score below 10: Id-Ta, Ta-Id, Ta-Jv, Ms-Ta, Ta-Ms, Ta-Tl, and Jv-Tl. Most of these language pairs were related to the Tamil language. We found that most of the Tamil translation results had an English sentence as unknown words (see Tables 8 and 9 in appendix). The translation results leading to not the same as a reference file. As a result, these language pairs had a lower BLEU score.

Surprisingly, we identified three language pairs that obtained a BLEU score of more than 15: Id-Tl, Tl-Id, and Ms-Tl, while using less than 70 sentences of the training dataset. This could be because of the attention mechanism in NMT of multilingual models. The attention mechanism, which was initially called a soft-alignment model in (Bahdanau et al., 2015), aligns a source phrase to a target word. Training this attention-based model is done by maximizing the conditional log-likelihood. After training, the model can do translation from any of the source languages to any of the target languages included in the parallel training corpora (Firat et al., 2016).

The Id-Tl, for example, obtained a BLEU score of 17.94 using only 56 sentences of training datasets. The NMT system that trained with fewer training datasets, e.g., below 1M, usually obtained lower BLEU scores. However, the Id-Tl results indicated that this language pair obtained an advantage from the attention mechanism of the multilingual model using 30 language pairs. In this study, these 30 language pairs are considered as low-resource languages and mostly have the same language family. Table 2 shows that our models used a small number of training datasets, e.g., below 1M, in all language pairs. Whereas, Table 1 shows that most languages have the same language family: Malayo-Polynesian. Therefore, we argue that low-resource language with the same language family should be considered in the NMT of the multilingual model. For example, if we want to improve the Tamil language performance, we should considered to add other languages with the same (or closely) language family as Tamil, e.g., Kannada, Bengali, Hindi.

Model	Dev			Devtest			Test		
	M2M-100 175M	TelU-KU-175M	TelU-KU-175M_HPO	M2M-100 175M	TelU-KU-175M	TelU-KU-175M_HPO	M2M-100 175M	TelU-KU-175M	TelU-KU-175M_HPO
M2M-100 175M	x	0.567	0.399	x	0.360	0.289	x	0.734	0.572
TelU-KU-175M		x	0.229		x	0.329		x	0.360
TelU-KU-175M_HPO			x			x			x

Table 7: Results of Wilcoxon signed-rank test.

5 Conclusion

We described our team submission for WMT21. Our results show improvement in most language pairs after optimizing the hyperparameters.

Furthermore, we also found three language pairs that obtained a BLEU score of more than 15 while using less than 70 sentences of the training dataset. In this study, we used 30 language pairs that are considered as low-resource language and mostly have the same language family. This result indicated that low-resource language with the same language family should be considered in the NMT of the multilingual model.

As future work, we plan to use a more sophisticated optimization algorithm, specifically informed searches such as bayesian optimization or evolutionary algorithm. Additionally, we want to try other percentages of the optimized dataset to see the effect of the number of training data on the performance. We also plan to use a specific tokenizer for Tamil, e.g., Indic NLP library (Kunchukuttan, 2020), iNLTK (Arora, 2020). The Tamil language needs a particular pre-processing due to its writing system that differs from other languages. Last, we plan to clean the dataset in pre-processing steps, considering that the dataset used in this work is noisy. We expect this will maximize the attention mechanism in the NMT of a multilingual model. Therefore, our model could produce better translation results.

Acknowledgements

This project was funded by Compute Grants: Large-Scale Multilingual Machine Translation of Conference on Machine Translation (WMT) and Microsoft Azure. This project was also funded by the PDT research scheme, Telkom University.

References

Husen Abas. 1987. Indonesian as a unifying language of wider communication : a historical and sociolinguistic perspective.

Ali Araabi and Christof Monz. 2020. Optimizing transformer for low-resource neural machine translation. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 3429–3435, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Gaurav Arora. 2020. iNLTK: Natural language toolkit for indic languages. In *Proceedings of Second Workshop for NLP Open Source Software (NLP-OSS)*, pages 66–71, Online. Association for Computational Linguistics.

Mikel Artetxe, Gorka Labaka, and Eneko Agirre. 2018a. Unsupervised statistical machine translation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3632–3642, Brussels, Belgium. Association for Computational Linguistics.

Mikel Artetxe, Gorka Labaka, Eneko Agirre, and Kyunghyun Cho. 2018b. Unsupervised neural machine translation. In *International Conference on Learning Representations*.

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural machine translation by jointly learning to align and translate. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.

James Bergstra and Yoshua Bengio. 2012. Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13.

Ondřej Bojar, Rajen Chatterjee, Christian Federmann, Yvette Graham, Barry Haddow, Shujian Huang, Matthias Huck, Philipp Koehn, Qun Liu, Varvara Logacheva, Christof Monz, Matteo Negri, Matt Post, Raphael Rubino, Lucia Specia, and Marco Turchi. 2017. Findings of the 2017 conference on machine translation (WMT17). In *Proceedings of the Second Conference on Machine Translation*, pages 169–214, Copenhagen, Denmark. Association for Computational Linguistics.

Chenhui Chu and Rui Wang. 2018. A survey of domain adaptation for neural machine translation. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1304–1319, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Kevin Duh, Graham Neubig, Katsuhito Sudoh, and Hajime Tsukada. 2013. Adaptation data selection using neural language models: Experiments in machine translation. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 678–683, Sofia, Bulgaria. Association for Computational Linguistics.

Nadir Durrani, Fahim Dalvi, Hassan Sajjad, Yonatan Belinkov, and Preslav Nakov. 2019. One size does not fit all: Comparing NMT representations of different granularities. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 1504–1516, Minneapolis, Minnesota. Association for Computational Linguistics.

Tirana Noor Fatyanosa and Masayoshi Aritsugi. 2020. Effects of the Number of Hyperparameters on the Performance of GA-CNN. In *2020 IEEE/ACM International Conference on Big Data Computing, Applications and Technologies (BDCAT)*, pages 144–153. IEEE.

Tirana Noor Fatyanosa and Masayoshi Aritsugi. 2021. An Automatic Convolutional Neural Network Optimization Using a Diversity-Guided Genetic Algorithm. *IEEE Access*, 9:91410 – 91426.

Orhan Firat, Kyunghyun Cho, and Yoshua Bengio. 2016. Multi-way, multilingual neural machine translation with a shared attention mechanism. In *NAACL HLT 2016, The 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego California, USA, June 12-17, 2016*, pages 866–875. The Association for Computational Linguistics.

Naman Goyal, Cynthia Gao, Vishrav Chaudhary, Peng-Jen Chen, Guillaume Wenzek, Da Ju, Sanjana Krishnan, Marc'Aurelio Ranzato, Francisco Guzmán, and Angela Fan. 2021. The FLORES-101 evaluation benchmark for low-resource and multilingual machine translation. *CoRR*, abs/2106.03193.

Jiatao Gu, Hany Hassan, Jacob Devlin, and Victor O.K. Li. 2018. Universal neural machine translation for extremely low resource languages. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 344–354, New Orleans, Louisiana. Association for Computational Linguistics.

Frank Hutter, Holger Hoos, and Kevin Leyton-Brown. 2014. An efficient approach for assessing hyperparameter importance. In *Proceedings of the 31st International Conference on International Conference on Machine Learning - Volume 32*, ICML'14, page I–754–I–762. JMLR.org.

Aaron Klein and Frank Hutter. 2019. Tabular benchmarks for joint architecture and hyperparameter optimization.

Philipp Koehn and Rebecca Knowles. 2017. Six challenges for neural machine translation. In *Proceedings of the First Workshop on Neural Machine Translation*, pages 28–39, Vancouver. Association for Computational Linguistics.

Amit Kumar and Anil Kumar Singh. 2019. NLPRL at WAT2019: transformer-based tamil - english indic task neural machine translation system. In *Proceedings of the 6th Workshop on Asian Translation, WAT@EMNLP-IJCNLP 2019, Hong Kong, China, November 4, 2019*, pages 171–174. Association for Computational Linguistics.

Anoop Kunchukuttan. 2020. The IndicNLP Library. `https://github.com/anoopkunchukuttan/indic_nlp_library/blob/master/docs/indicnlp.pdf`.

Guillaume Lample, Alexis Conneau, Ludovic Denoyer, and Marc'Aurelio Ranzato. 2018a. Unsupervised machine translation using monolingual corpora only. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*. OpenReview.net.

Guillaume Lample, Myle Ott, Alexis Conneau, Ludovic Denoyer, and Marc'Aurelio Ranzato. 2018b. Phrase-based & neural unsupervised machine translation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 5039–5049, Brussels, Belgium. Association for Computational Linguistics.

Toshiaki Nakazawa, Shohei Higashiyama, Chenchen Ding, Hideya Mino, Isao Goto, Hideto Kazawa, Yusuke Oda, Graham Neubig, and Sadao Kurohashi. 2017. Overview of the 4th workshop on Asian translation. In *Proceedings of the 4th Workshop on Asian Translation (WAT2017)*, pages 1–54, Taipei, Taiwan. Asian Federation of Natural Language Processing.

Nathan Ng, Kyra Yee, Alexei Baevski, Myle Ott, Michael Auli, and Sergey Edunov. 2019. Facebook fair's WMT19 news translation task submission. In *Proceedings of the Fourth Conference on Machine Translation, WMT 2019, Florence, Italy, August 1-2, 2019 - Volume 2: Shared Task Papers, Day 1*, pages 314–319. Association for Computational Linguistics.

Sashi Novitasari, Andros Tjandra, Sakriani Sakti, and Satoshi Nakamura. 2020. Cross-lingual machine speech chain for javanese, sundanese, balinese, and bataks speech recognition and synthesis. In *Proceedings of the 1st Joint Workshop on Spoken Language Technologies for Under-resourced languages and Collaboration and Computing for Under-Resourced Languages, SLTU/CCURL@LREC 2020, Marseille, France, May 2020*, pages 131–138. European Language Resources association.

Myle Ott, Sergey Edunov, Alexei Baevski, Angela Fan, Sam Gross, Nathan Ng, David Grangier, and Michael Auli. 2019. fairseq: A fast, extensible toolkit for sequence modeling. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Demonstrations*, pages 48–53. Association for Computational Linguistics.

Meghana Ravikumar. 2020. Efficient BERT: Finding Your Optimal Model with Multimetric Bayesian Optimization.

Raphael Rubino, Benjamin Marie, Raj Dabre, Atsushi Fujita, Masao Utiyama, and Eiichiro Sumita. 2020. Extremely low-resource neural machine translation for asian languages. *Mach. Transl.*, 34(4):347–382.

Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016. Neural machine translation of rare words with subword units. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1715–1725, Berlin, Germany. Association for Computational Linguistics.

Rico Sennrich and Biao Zhang. 2019. Revisiting low-resource neural machine translation: A case study. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 211–221, Florence, Italy. Association for Computational Linguistics.

Raymond Hendy Susanto, Septina Dian Larasati, and Francis M. Tyers. 2012. Rule-based Machine Translation between Indonesian and Malaysian. In *Proceedings of the 3rd Workshop on South and Southeast Asian Natural Language Processing, WSSANLP@COLING 2012, Mumbai, India, December 8, 2012*, pages 191–200.

Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. 2014. Sequence to sequence learning with neural networks. *CoRR*, abs/1409.3215.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is All You Need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 6000–6010, Red Hook, NY, USA. Curran Associates Inc.

Xuan Zhang and Kevin Duh. 2020. Reproducible and Efficient Benchmarks for Hyperparameter Optimization of Neural Machine Translation Systems. *Transactions of the Association for Computational Linguistics*, 8:393–408.

Barret Zoph, Deniz Yuret, Jonathan May, and Kevin Knight. 2016. Transfer learning for low-resource neural machine translation. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1568–1575, Austin, Texas. Association for Computational Linguistics.

A Appendix

Lang pairs	Model	Text
English	Source	We now have 4-month-old mice that are non-diabetic that used to be diabetic, he added.
En-Id	Reference	“Saat ini ada mencit umur 4 bulan nondiabetes yang dulunya diabetes,” tambahnya.
	M2M-100_175M	Kami sekarang memiliki kuku 4 bulan yang tidak diabetik yang digunakan untuk menjadi diabetik, katanya.
	TelU-KU-175M	Kami sekarang memiliki mice berusia 4 bulan yang bukan diabetik yang digunakan untuk menjadi diabetes, ia menambahkan.
	TelU-KU-175M_HPO	Kami sekarang memiliki tikus berusia 4 bulan yang non-diabetik yang digunakan untuk diabetik, dia menambahkan.
En-Jv	Reference	Saiki kita nduweni tikus umur-4-sasi sing ora-nduweni-diabetes sing sadurunge nduweni diabetes, ujare.
	M2M-100_175M	Kita ora ana 4 bulan madu yang non-diabetik sing digunakake kanggo diabetes, katanya.
	TelU-KU-175M	Dhèwèké ya iku mice 4 bulan kang non-diabetik kang digunakaké diabetes,” dhèwèké tambah.
	TelU-KU-175M_HPO	Nanging, dhèwèké bisa nganti 4 bulan kang ora diwènèhi yèn diwènèhi diapit.
En-Ms	Reference	“Kami kini mempunyai seekor anak tikus yang berusia 4 bulan yang sudah tidak menghidap diabetes,” beliau menambah.
	M2M-100_175M	Kami kini mempunyai kuku 4 bulan yang tidak diabetik yang digunakan untuk menjadi diabetik, katanya.
	TelU-KU-175M	Kami sekarang mempunyai mice berusia 4 bulan yang bukan-diabetik yang digunakan untuk menjadi diabetes, ia tambah.
	TelU-KU-175M_HPO	Kami sekarang mempunyai tikus berusia 4 bulan yang bukan-diabetik yang digunakan sebagai diabetik, ia menambah.
En-Ta	Reference	"எங்களிடம் இப்போது 4-மாத-வயதுடைய எலி ஒன்று உள்ளது, முன்னர் அதற்கு நீரிழிவு இருந்தது தற்போது இல்லை" என்று அவர் மேலும் கூறினார்.
	M2M-100_175M	நாங்கள் இப்போது 4 மாதங்களுக்குள் சுழுதைகள் உள்ளன, இது இரகசியமாக பயன்படுத்தப்படாத இரகசியமாக உள்ளது, அவர் கூறினார்.
	TelU-KU-175M	We now have 4-month-old mice that are non-diabetic that are used to be diabetes, he added.
	TelU-KU-175M_HPO	We now have 4-month-old mice that are non-diabetic that used to be diabetic, he added.
En-Tl	Reference	Mayroon na tayong 4 na buwang gulang na daga na hindi diabetic na dating diabetic, dagdag niya.
	M2M-100_175M	Kapag medyo nakarating ka na sa age na alam mo na ang dami mo nangpinagdaanan... a good soldier must knowwhen to surrender.
	TelU-KU-175M	Tapos ay may 4-month-old mice na hindi-diabetic na ginagamit upang maging diabetes,” siya ay hinaharap.
	TelU-KU-175M_HPO	We now have four-month-old mice that are non-diabetic that used to be diabetic, he added.
Indonesia	Source	“Saat ini ada mencit umur 4 bulan nondiabetes yang dulunya diabetes,” tambahnya.
Id-En	Reference	We now have 4-month-old mice that are non-diabetic that used to be diabetic, he added.
	M2M-100_175M	“Now there is a age of 4 months of non-diabetes that has long been diabetes,” she said.
	TelU-KU-175M	This day there is a 4 months old nondiabetes who had previously been diabetes, his debut.
	TelU-KU-175M_HPO	In this period there is a four-month-old nondiabetes that was formerly diabetes, his tambahnya.
Id-Jv	Reference	Saiki kita nduweni tikus umur-4-sasi sing ora-nduweni-diabetes sing sadurunge nduweni diabetes, ujare.
	M2M-100_175M	“Itu ora tumitake umur 4 bulan nondiabetes sing iku diabetes,” tambahnya.
	TelU-KU-175M	]]]]]]]]]]]]]]]]
	TelU-KU-175M_HPO	“Selang iki duwé umur 4 bulan nondiabetes kang formeré diabetes,” tambahnya.
Id-Ms	Reference	“Kami kini mempunyai seekor anak tikus yang berusia 4 bulan yang sudah tidak menghidap diabetes,” beliau menambah.
	M2M-100_175M	“Pada masa ini, ia mempunyai umur 4 bulan nondiabetes yang sejak itu diabetes,” tambahnya.
	TelU-KU-175M	Saat ini sudah berusia 4 bulan nondiabetes yang dulunya diabetes, tambahnya.
	TelU-KU-175M_HPO	“Saat ini terdapat anak berusia 4 bulan nondiabetes yang dulunya diabetes,” tambahnya.
Id-Ta	Reference	"எங்களிடம் இப்போது 4-மாத-வயதுடைய எலி ஒன்று உள்ளது, முன்னர் அதற்கு நீரிழிவு இருந்தது தற்போது இல்லை" என்று அவர் மேலும் கூறினார்.
	M2M-100_175M	இந்த நிலையில், 4 மாதங்களுக்கு முன்னர் சிகிச்சை அளிக்கப்பட்டு சிகிச்சை அளிக்கப்பட்டு சிகிச்சை பெற்றுள்ளது.
	TelU-KU-175M	This time there is a four-month-old nondiabetes that had previously been diabetes, his debut.
	TelU-KU-175M_HPO	This time there is a four-month-old nondiabetes that was formerly diabetes, his tambahnya.
Id-Tl	Reference	Mayroon na tayong 4 na buwang gulang na daga na hindi diabetic na dating diabetic, dagdag niya.
	M2M-100_175M	Kapag medyo nakarating ka na sa age na alam mo na ang dami mo nangpinagdaanan... a good soldier must knowwhen to surrender.
	TelU-KU-175M	Ang ito ay may edad na 4 buwan ng nondiabetes na dating diabetes,” tumutukoy niya.
	TelU-KU-175M_HPO	Sa panahong ito ay mayroong apat na buwan ng nondiabetes na dating diabetes, kanyang tambahnya.
Javanese	Source	Saiki kita nduweni tikus umur-4-sasi sing ora-nduweni-diabetes sing sadurunge nduweni diabetes, ujare.
Jv-En	Reference	We now have 4-month-old mice that are non-diabetic that used to be diabetic, he added.
	M2M-100_175M	Now we have four-year-old age, which are not diagnosed with diabetes, diabetes.
	TelU-KU-175M	We have four-year-old children who have diabetes, ujare.
	TelU-KU-175M_HPO	Today we have a four-year-old mice that is not-to-be-diabetes who now has diabetes, then.
Jv-Id	Reference	“Saat ini ada mencit umur 4 bulan nondiabetes yang dulunya diabetes,” tambahnya.
	M2M-100_175M	Tetapi kita memiliki usia 4 tahun yang tidak memiliki diabetes yang sering memiliki diabetes, seperti itu.
	TelU-KU-175M	Saiki kita memiliki tikus berusia 4 tahun yang sekarang-nduweni-diabetes yang memiliki diabetes, ujare.
	TelU-KU-175M_HPO	Tetapi kita memiliki tikus berusia 4 tahun yang tidak-wangian-diabetes yang sekarang memiliki diabetes, ujare.
Jv-Ms	Reference	“Kami kini mempunyai seekor anak tikus yang berusia 4 bulan yang sudah tidak menghidap diabetes,” beliau menambah.
	M2M-100_175M	Sekarang kita mempunyai usia 4 tahun yang tidak mempunyai diabetes yang mempunyai diabetes, boleh.
	TelU-KU-175M	Saiki kita mempunyai tikus berusia 4-sasi yang sekarang-nduweni-diabetes yang mempunyai diabetes, ujare.
	TelU-KU-175M_HPO	Terdapat kita mempunyai tikus berusia-4-sasi yang kini mempunyai diabetes, ujare. ”
Jv-Ta	Reference	"எங்களிடம் இப்போது 4-மாத-வயதுடைய எலி ஒன்று உள்ளது, முன்னர் அதற்கு நீரிழிவு இருந்தது தற்போது இல்லை" என்று அவர் மேலும் கூறினார்.
	M2M-100_175M	நாங்கள் இப்போது 4 வயதில் 4 வயதைக் கொண்டே இருக்கிறோம், இது நோயாளிகளின் நோயாளிகள், நோயாளிகள்.
	TelU-KU-175M	We have four-year-old tikus who are now-anduweni-diabetes with diabetes, ujare.
	TelU-KU-175M_HPO	நாங்கள் 4 வயதிற்குட்பட்ட குழந்தைகளைக் கொண்டிருக்கின்றோம்" (ஆங்கிலம்: "ஆங்கிலம்: <unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk>)".
Jv-Tl	Reference	Mayroon na tayong 4 na buwang gulang na daga na hindi diabetic na dating diabetic, dagdag niya.
	M2M-100_175M	Sa ngayon, kami ay nagkaroon ng isang edad ng 4 na edad na hindi nag-aalok sa diabetes na may diabetes, na nag-aalok.
	TelU-KU-175M	Sa ngayon ay may 4-sasyong babae na hindi-duweni-diabetes na nagkaroon ng diabetes, ujare.
	TelU-KU-175M_HPO	Sa ngayon ay mayroong 4 taong gulang na mice na hindi-milyang-diabetes na nakarating na may diabetes, pangalang.”

Table 8: Translation results.

Lang pairs	Model	Text
Malay	Source	“Kami kini mempunyai seekor anak tikus yang berusia 4 bulan yang sudah tidak menghidap diabetes,” beliau menambah.
Ms-En	Reference	We now have 4-month-old mice that are non-diabetic that used to be diabetic, he added.
	M2M-100_175M	“We now have a four-month old baby that has no diabetes,” he added.
	TelU-KU-175M	We now have a four-month boy that has no diabetes, he added.
	TelU-KU-175M_HPO	We now have a four-month-old mice child who has no diabetes, he added.
Ms-Id	Reference	“Saat ini ada mencit umur 4 bulan nondiabetes yang dulunya diabetes,” tambahnya.
	M2M-100_175M	“Kami sekarang memiliki seekor tikus yang berusia 4 bulan yang belum menderita diabetes,” katanya.
	TelU-KU-175M	Kami sekarang memiliki seekor tikus berusia 4 bulan yang sudah tidak menghidap diabetes, dia menambahkan.
	TelU-KU-175M_HPO	“Kami kini memiliki seorang anak tikus yang berusia 4 bulan yang sudah tidak menghidap diabetes,” beliau menambahkan.
Ms-Jv	Reference	Saiki kita nduweni tikus umur-4-sasi sing ora-nduweni-diabetes sing sadurunge nduweni diabetes, ujare.
	M2M-100_175M	“Kita ora nduweni anak tikus 4 bulan lan ora ana diabetes,” tambah.
	TelU-KU-175M	We now have a four-month boy tikus who has no menghidap diabetes, she added.
	TelU-KU-175M_HPO	“Kami kini mempunyai seekor anak tikus yang umur 4 bulan yang sudah tidak menghidap diabetes,” beliau menambahkan.
Ms-Ta	Reference	"எங்களிடம் இப்போது 4-மாத-வயதுடைய எலி ஒன்று உள்ளது, முன்னர் அதற்கு நீரிழிவு இருந்தது தற்போது இல்லை" என்று அவர் மேலும் கூறினார்.
	M2M-100_175M	“ஒரு வயதிலிருந்து 4 மாதங்களுக்கு முன்னர் ஒரு குழந்தையைப் பெற்றோம்” என்று அவர் கூறியுள்ளார்.
	TelU-KU-175M	We now have a four-month-old girl who has no diabetes, she added.
	TelU-KU-175M_HPO	We now have a four-month-old tikus child who already does not live on diabetes, he added.
Ms-Tl	Reference	Mayroon na tayong 4 na buwang gulang na daga na hindi diabetic na dating diabetic, dagdag niya.
	M2M-100_175M	Kapag medyo nakarating ka na sa age na alam mo na ang dami mo nangpinagdaanan... a good soldier must know when to surrender.
	TelU-KU-175M	Tapos ay may isang 4 buwan na anak na hindi nakapagpatuloy ng diabetes,” siya ay nagsasabi.
	TelU-KU-175M_HPO	“Kami ngayon ay may isang anak na tikus na 4 na buwan na hindi namuhay ng diabetes,” he added.
Tamil	Source	"எங்களிடம் இப்போது 4-மாத-வயதுடைய எலி ஒன்று உள்ளது, முன்னர் அதற்கு நீரிழிவு இருந்தது தற்போது இல்லை" என்று அவர் மேலும் கூறினார்.
Ta-En	Reference	We now have 4-month-old mice that are non-diabetic that used to be diabetic, he added.
	M2M-100_175M	He said, We have now one of the four-year-old rolls, and it was not until it was long,” he further said.
	TelU-KU-175M	He said to me, I have a four-thirds light now, but it was not until it was lost.”
	TelU-KU-175M_HPO	He said, ”I am now one of the four-member elephants, and that it was not until it had been passed.”
Ta-Id	Reference	“Saat ini ada mencit umur 4 bulan nondiabetes yang dulunya diabetes,” tambahnya.
	M2M-100_175M	“Kami sudah memiliki satu dari empat orang, dan tidak ada lagi yang terjadi sebelumnya,” katanya.
	TelU-KU-175M	Ia berkata kepada kami sekarang, ”Saya sudah ada empat jam, dan sebelumnya tidak ada waktu yang lama.”
	TelU-KU-175M_HPO	Ia berkata, ”Tentara ini sekarang memiliki empat kursi, dan sebelumnya sudah tidak ada lagi,” katanya.
Ta-Jv	Reference	Saiki kita nduweni tikus umur-4-sasi sing ora-nduweni-diabetes sing sadurunge nduweni diabetes, ujare.
	M2M-100_175M	Itu uga ana ing kita minangka salah siji saka 4 taun, nanging ora ana ing nganti-nganti iku ing nganti, katanya.
	TelU-KU-175M	Dhèwèké diprodhuksi déning dhèwèké nganti 4 kaliyan, lan ora diprodhuksi”.
	TelU-KU-175M_HPO	Piyambakipun nggantosaken ”Sang” ingkang dipunsebat ”Sang” inggih punika ”Sang” ingkang dipunsebat ”Sang” lan ”Sang” punika ”Sang ”Sang” ingkang dipunsebat ”Sang”.
Ta-Ms	Reference	“Kami kini mempunyai seekor anak tikus yang berusia 4 bulan yang sudah tidak menghidap diabetes,” beliau menambah.
	M2M-100_175M	Beliau berkata, “Kami mempunyai satu daripada empat orang, dan ia tidak pernah berlaku sebelum ini,” katanya.
	TelU-KU-175M	Saya sekarang mempunyai satu lilin empat, ia juga berkata kepada saya tidak akan lama lagi, katanya.
	TelU-KU-175M_HPO	Lagi ini, ia berkata Terdapat unsur api empat, tetapi ia tidak lagi menyahpepijat.”
Ta-Tl	Reference	Mayroon na tayong 4 na buwang gulang na daga na hindi diabetic na dating diabetic, dagdag niya.
	M2M-100_175M	Kapag medyo nakarating ka na sa age na alam mo na ang dami mo nangpinagdaanan... a good soldier must know when to surrender.
	TelU-KU-175M	Siya ay isang seryeng tao, at hindi siya nag-iisip sa kanya.
	TelU-KU-175M_HPO	Iminungkahi na ”ang isang four-mga relyo sa kami ngayon ay isang relasyon, na hindi ito pinapatakbo nang hindi ito naganap.”
Tagalog	Source	Mayroon na tayong 4 na buwang gulang na daga na hindi diabetic na dating diabetic, dagdag niya.
Tl-En	Reference	We now have 4-month-old mice that are non-diabetic that used to be diabetic, he added.
	M2M-100_175M	We have 4 small areas that do not have diabetic dating diabetic, he said.
	TelU-KU-175M	May we have 4 new gullies that are not diabetic dating diabetesic, he says.
	TelU-KU-175M_HPO	There are four-year-old, non- diabetic, former diabetic, he added.
Tl-Id	Reference	“Saat ini ada mencit umur 4 bulan nondiabetes yang dulunya diabetes,” tambahnya.
	M2M-100_175M	Kami memiliki 4 negara yang tidak diabetik dan tidak diabetik, kata dia.
	TelU-KU-175M	Kami memiliki 4 warna putih yang tidak diabetes yang dating diabetes, dia tahu.
	TelU-KU-175M_HPO	Tetapi ada empat orang tua tua yang tidak diabetik yang sebelumnya diabetic, katanya.
Tl-Jv	Reference	Saiki kita nduweni tikus umur-4-sasi sing ora-nduweni-diabetes sing sadurunge nduweni diabetes, ujare.
	M2M-100_175M	Kita ana 4 negara sing padha padha padha padha diabetes sing ora diabetik, ujar dia.
	TelU-KU-175M	Dhèwèké dhèwèké dhèwèké nggèké 4 gulang kang tidak diabetes kang dating diabetes,” dhèwèké.
	TelU-KU-175M_HPO	Tetapi ana 4 gulungan kang umuré lan ora diabetik kang former diabetic, dhèwèké maragani.
Tl-Ms	Reference	“Kami kini mempunyai seekor anak tikus yang berusia 4 bulan yang sudah tidak menghidap diabetes,” beliau menambah.
	M2M-100_175M	Kami mempunyai 4 buah negara yang tidak diabetik dan tidak diabetik, katanya.
	TelU-KU-175M	Kami mempunyai 4 warna putih yang tidak diabetes yang dating diabetes, diadagangkan.
	TelU-KU-175M_HPO	Terdapat ana 4 orang tua yang gaya yang bukan diabetik yang bekas diabetic, kata-kata dia.
Tl-Ta	Reference	"எங்களிடம் இப்போது 4-மாத-வயதுடைய எலி ஒன்று உள்ளது, முன்னர் அதற்கு நீரிழிவு இருந்தது தற்போது இல்லை" என்று அவர் மேலும் கூறினார்.
	M2M-100_175M	நாங்கள் 4 மாதங்கள் கழித்து நோய் நோய் நோய் நோயாளிகள் இல்லை என்று அவர் கூறியுள்ளார்.
	TelU-KU-175M	We have 4 new colored days that are not diabetesic dating diabetesic, hedaged.
	TelU-KU-175M_HPO	நாம் 4வது வயதிலேயே, முன்னேற்றமான, முன்னேற்றமான, முன்னேற்றமான, முன்னேற்றத்தின் முன்னேற்றத்தில்" என்று அவர் கூறுகிறார்.

Table 9: Translation results.