jsret.knpub.com

E-ISSN: 2962-6110
P-ISSN: 2964-7843
JSRET
Journal of Scientific Research, Education, and Technology

Comparative Analysis of Adam and RMSprop Optimizers on Bi-LSTM Models for Indonesian–Ngapak Translation

Hery Setiyawan Jordi, Muhammad Zakariyah
Universitas Teknologi Yogyakarta, Yogyakarta, Indonesia

ABSTRACT

This study compares the performance of the Adam and RMSprop optimization algorithms in training a Bidirectional Long Short-Term Memory (Bi-LSTM) model with an Attention mechanism for Indonesian–Ngapak translation. Using a parallel corpus containing 23,592 sentence pairs, both models were tested using identical configurations to ensure an objective comparison. The results of the experiment show that Adam has faster convergence and better translation quality, as indicated by a validation accuracy of 95.5%, a loss of 0.43, and BLEU-1 to BLEU-4 scores of 0.8775–0.7393. In contrast, RMSprop showed slower learning dynamics with a validation accuracy of 93.6% and lower BLEU scores (0.8284–0.6384). These findings confirm that Adam is more adaptive and efficient in optimizing model parameters for low-power language translation tasks. This study makes a unique contribution as the first empirical analysis of optimizer selection in Indonesian–Ngapak translation and supports the preservation of the Ngapak language through artificial intelligence technology.

Keywords: *Neural Machine Translation, Bi-LSTM, Adam, RMSprop, Ngapak Language*

Corresponding author
Name: Hery Setiyawan Jordi
Email:herysetiyawan368@gmail.com

INTRODUCTION

The Ngapak language is a dialect of Javanese that has its own distinctive characteristics in terms of pronunciation, sentence structure, and vocabulary. This dialect is widely spoken by people in the western part of Central Java Province, including the areas of Banyumas, Cilacap, Purbalingga, and surrounding areas. The uniqueness of Ngapak is not only evident in its linguistic characteristics, but also reflects the social identity and local wisdom of the Banyumasan people, who uphold the values of equality, honesty, and openness in communication. Thus, Ngapak serves not only as a means of communication, but also as a cultural symbol that strengthens the identity of its speakers (Nugroho & Kusuma, 2023). However, amid globalization and technological developments, the use of Ngapak language has declined and begun to be abandoned by the younger generation. In the context of language technology development, Ngapak language has not received adequate attention, especially in the development of artificial intelligence-based automatic translation systems. Most automatic translation research in Indonesia still focuses on national languages and popular regional languages such as Sundanese, Madurese, or Minangkabau (Putri et al., 2024). Meanwhile, translation into Ngapak is still done manually. This condition highlights the importance of research that can provide technological solutions to preserve and facilitate the community's understanding of Ngapak in the digital era.

The development of Artificial Intelligence (AI) and Natural Language Processing (NLP) has had a significant impact on the world of language translation. One of the most widely used approaches is Neural Machine Translation (NMT), which is based on Recurrent Neural Networks (RNN). NMT models have the ability to understand context and semantic relationships between words better than previous rule-based or statistical methods(Gunawan et al., 2021). This approach allows machines to learn sentence structures automatically without the need for human intervention in determining language patterns. Previous studies have shown that the application of RNN-based NMT provides more natural and accurate translation results in various languages, including regional languages in Indonesia. This improvement occurs because RNN is able to capture the long-term dependencies between words, which are very important in the context of sentence translation (Fauziyah et al., 2022).

One of the developments of Recurrent Neural Networks (RNN) that is widely used in modern machine translation is Long Short-Term Memory (LSTM) and Bidirectional LSTM (Bi-LSTM). The Bi-LSTM model is capable of processing word sequences from two directions simultaneously, thereby understanding the context of a sentence more comprehensively. This capability makes Bi-LSTM effective in capturing the semantic meaning between words and producing more accurate translations (C. Li & Flanigan, 2022). This advantage makes Bi-LSTM the primary choice for translation systems that require semantic accuracy between sentences. Although the model architecture has developed rapidly, translation performance is also greatly influenced by the model training process, particularly in the selection of optimization algorithms (optimizers). Optimizers play a role in accelerating model convergence, improving the direction of network weight updates, and reducing loss values during training (Abdulkadirov & Lyakhov, 2023). Two widely used optimizers are Adam and RMSprop. The Adam optimizer combines the advantages of RMSprop's adaptive learning rate and momentum from Stochastic Gradient Descent (SGD), enabling it to accelerate the learning process while maintaining model stability. Meanwhile, RMSprop is known to be more stable for long sequential data such as text (Ahda et al., 2024). Although both algorithms are equally popular, empirical comparisons of their performance in the context of regional language translation, such as Ngapak, are still rare. However, the selection of the appropriate optimizer can directly affect convergence speed, training stability, and the resulting BLEU score.

Several previous studies related to regional language translation in Indonesia have shown the successful application of RNN-based NMT models. (Gunawan et al., 2021) developed a translation system from Indonesian to Ketapang Malay using Bahdanau and Luong’s attention mechanisms, achieving higher accuracy than traditional approaches.

(Abidin et al., 2018) demonstrated the effectiveness of attention-based NMT on limited Lampung–Indonesian corpora, while (Indonesia-banyumasan et al., 2023) employed an RNN-GRU model for Indonesian–Banyumasan translation and achieved a BLEU score of 34.1. Collectively, these studies indicate that RNN variants consistently improve translation quality across Indonesian regional languages. However, they also reveal a recurring limitation: the focus remains on architectural improvements, with minimal attention given to how optimization algorithms affect model performance. None of these studies evaluate the influence of Adam or RMSprop on convergence speed, training stability, or BLEU outcomes in low-resource languages such as Ngapak. This lack of empirical comparison forms a clear methodological gap that motivates the present study.

Therefore, this study focuses on a comparative analysis of the performance of Adam and RMSprop optimizers on a Bidirectional Long Short-Term Memory (Bi-LSTM) model in Indonesian–Ngapak translation tasks. The evaluation is conducted using three main metrics—accuracy, loss function, and BLEU (Bilingual Evaluation Understudy)—to determine the effectiveness and efficiency of the model training process. This experimental design aims to provide empirical insights into how optimizer selection affects translation quality, training stability, and the model's ability to reach optimal convergence. Beyond its technical contribution to the development of Bi-LSTM-based Neural Machine Translation (NMT) systems, this study also carries socio-cultural significance by supporting the preservation of the Ngapak language through the application of artificial intelligence technology. By integrating linguistic approaches with modern computational methods, this research is expected to strengthen the digital representation of regional languages and expand the use of NLP within local cultural contexts in Indonesia.

METHOD

1. Dataset

This study uses a parallel corpus of Indonesian–Ngapak as the main data in training artificial neural network models. This corpus consists of parallel sentence pairs between the source language (Indonesian) and the target language (Ngapak) collected from various informal text sources such as online conversations and social media posts, then validated with native speakers to ensure equivalence of meaning and semantic context. The dataset is stored in CSV format with two main columns, namely Indonesian and Ngapak, containing a total of 23,592 sentence pairs. The data was divided using the train_test_split function from the scikit-learn library with a ratio of 80% for training, 10% for validation, and 10% for testing. After compilation, the entire corpus underwent a pre-processing stage to ensure linguistic consistency before being used in model training.

2. Data Preprocessing

Pre-processing is an essential initial stage in the development of a Neural Machine Translation (NMT) model, as it serves to prepare the Indonesian–Ngapak parallel corpus so that it has a consistent linguistic structure, is free of irrelevant symbols or characters, and is uniform in input format so that the model can learn optimally. In the context of languages with limited data sources, as discussed by (Marco & Fraser, 2022), the application of appropriate preprocessing strategies can improve training efficiency and the accuracy of model prediction results. To provide a clearer picture of the steps involved, the text cleaning and normalization process is summarized systematically in Table 1 below.

Table 1. Data Preprocessing

Process	Description	Input	Output
Cleaning	Remove non-alphabetic characters such as numbers, punctuation marks, and symbols to ensure that the text is clean and uniform.	WAH!! 43 Enak Banget nasi Goreng-nya , TOP deh...	WAH Enak Banget nasi Gorengnya TOP deh
Normalization	Change all letters to lowercase to maintain consistency in spelling and text format.	WAH Enak Banget nasi Gorengnya TOP deh	wah enak banget nasi gorengnya top deh
Tokenization	Breaking sentences into word units so that the model can recognize linguistic structures and semantic patterns.	wah enak banget nasi gorengnya top deh	['wah', 'enak', 'banget', 'nasi', 'gorengnya', 'top', 'deh']
Special Token Insertion	Add special tokens to mark the beginning and end of the target sentence.	['wah', 'enak', 'banget', 'nasi', 'gorengnya', 'top', 'deh']	['', 'wah', 'enak', 'banget', 'nasi', 'gorengnya', 'top', 'deh', '']
Vocabulary Construction	Convert tokens into numerical representations based on a vocabulary index built from the entire dataset.	['', 'wah', 'enak', 'banget', ..., '']	[2, 50, 60, 70, ... , 3]
Padding (until 17)	Align the length of the sequences so that all data has a uniform input dimension for the Bi-LSTM network.	[2, 50, 60, 70, 30, 150, 100, 90, 3]	[2, 50, 60, 70, 30, 150, 100, 90, 3, 0, 0, 0, 0, 0, 0, 0, 0]

3. NMT Model

Neural Machine Translation (NMT) is the latest approach in machine translation based on artificial neural networks and trained end-to-end to translate sentences from the source language to the target language. Unlike statistical or rule-based methods, NMT uses an encoder–decoder architecture, where the encoder component converts the sequence of words in the source language into a vector representation, and the decoder then generates words in the target language based on that representation (Puspitaningrum, 2021). In addition, the application of attention mechanisms has become a common practice to improve translation quality by providing dynamic focus on important parts of the source sentence.

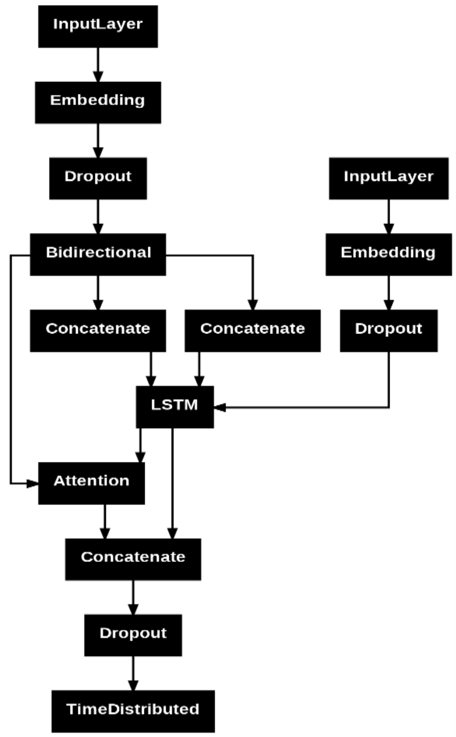

Figure 1. Bi-LSTM Model Architecture + NMT Attention

The architecture used in this study is Bidirectional Long Short-Term Memory (Bi-LSTM) in a sequence-to-sequence (seq2seq) encoder–decoder framework equipped with an Attention mechanism. In the encoder stage, each word in the source sentence is projected into a vector space through an Embedding Layer that converts words into numerically meaningful semantic representations. These representations are then processed bidirectionally by Bidirectional LSTM, where the outputs from the forward and backward directions are concatenated to form a context representation that can capture word relationships from both directions simultaneously. The final result of this process is a hidden state and cell state that are combined as the initial context for the subsequent decoding process. This bidirectional approach has proven effective in improving semantic context understanding and sentence structure, as explained by (T. Xie et al., 2023), who found that Bi-LSTM is capable of strengthening long-term context representation in neural-based automatic translation systems.

Next, in the decoder stage, the target sentence input is processed through the Decoder Embedding Layer and LSTM Decoder to generate an autoregressive output sequence. The Attention Layer mechanism then calculates the context vector as a weighted average of the encoder output based on its relevance to the current decoder state. This value is combined with the decoder output to help the model focus its attention on the most relevant parts of the input. This process is followed by a Concatenate layer and Dropout 0.3 to prevent overfitting, as well as a TimeDistributed Dense + Softmax Layer that generates a probability distribution over the target vocabulary. The complete model architecture is shown in Figure 4, which shows the data flow from the input layer (encoder_input) to the model output (translated sentence). All hyperparameter configurations such as embedding dimensions, number of LSTM units, dropout rate, batch size, and number of epochs were made identical in both experiments, ensuring reliability by keeping all settings consistent so that any performance differences arise solely from the optimizer used, namely Adam and RMSprop, to assess their effect on convergence speed, training loss stability, and translation quality as described by (Wu & Xing, 2024).

4. Adam Optimization

The Adam algorithm (Adaptive Moment Estimation) is one of the optimization methods widely applied in neural network training processes due to its ability to adaptively adjust the learning rate. According to (Loshchilov & Hutter, 2019), Adam was developed as an improvement on the Adaptive Gradient Algorithm (AdaGrad) and Root Mean Square Propagation (RMSProp) methods, which had weaknesses in convergence stability. The combination of these two approaches allows Adam to work effectively on data with uneven gradient values, both large and very small.

In general, this algorithm stores two important statistical values from the gradient, namely the average gradient called the first moment (m_t) and the average square gradient called the second moment (v_t). Both values are continuously updated at each iteration by considering the latest gradient (g_t) using the following equation:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \qquad (1)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \qquad (2)$$

After both moments are obtained, corrections are made for biases that typically arise at the beginning of training. The model parameters are then updated with the formula:

$$\theta_t = \theta_{t-1} - \eta \frac{\widehat{m_t}}{\sqrt{\widehat{v_t}} + \epsilon} \qquad (3)$$

Where η denotes the learning rate, β_1 dan β_2 are exponential decay factors for each moment, and ϵ epsilon is a small constant value to maintain numerical stability during the parameter update process.

According to (Keskar & Socher, 2017), Adam has advantages in terms of computational efficiency and does not require much adjustment to hyperparameters. In addition, Adam is also able to accelerate the convergence process without sacrificing training stability, especially in complex and sequential models.

5. RMSProp Optimization

The RMSProp (Root Mean Square Propagation) algorithm is an adaptive optimization method widely used in neural network training due to its ability to dynamically adjust the learning rate of each parameter. RMSProp works by updating weights based on the mean square gradient, thereby stabilizing parameter updates, especially for sequential data such as text. According to (Elshamy et al., 2023), this approach can accelerate convergence and reduce gradient fluctuations in deep learning models.

In general, parameter updates in RMSProp follow the following equation:

$$E[g^2]_t = \gamma E[g^2]_{t-1} + (1-\gamma) g_t^2 \qquad (4)$$

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{E[g^2]_t + \varepsilon}} g_t \qquad (5)$$

Here, η denotes the learning rate, γ is the exponential decay factor for the moving average of squared gradients, and ε is a small constant added to ensure numerical stability during parameter updates.

RMSProp excels at controlling the size of update steps for each parameter, allowing the model to learn more efficiently without experiencing large oscillations. According to (Malladi et al., 2022), this adaptive adjustment is very effective in accelerating the training of Recurrent Neural Network (RNN) or Long Short-Term Memory (LSTM) models that are sensitive to long gradients. However, the selection of hyperparameter values such as η dan γ must still be considered because inappropriate values can hinder the convergence process or cause divergence in training.

6. Model Training Procedure

The training procedure in this study uses a deep neural network-based Neural Machine Translation (NMT) approach that works end-to-end to translate Indonesian text into Ngapak. This model applies an encoder–decoder architecture, where the encoder converts the source sentence into a contextual vector representation and the decoder generates the target sentence based on that representation. This approach is capable of learning semantic and syntactic patterns between words without requiring explicit linguistic rules (Israr et al., 2023).

The model was trained using Bidirectional Long Short-Term Memory (Bi-LSTM) combined with Attention Mechanism to improve its ability to capture sentence context and long-term dependencies. Training was carried out using two optimization algorithms, namely Adam and RMSprop, to compare their effectiveness in the convergence process and training stability. According to (Wu & Xing, 2024), the combination of Bi-LSTM and Attention is effective in improving translation accuracy and efficiency because it allows the model to focus on important words during the learning process. In order for the training process to run optimally and the evaluation results to be analyzed objectively, the parameter settings were made according to the configuration summarized in Table 2.

Table 2. Configuration Parameters

No	Parameter Name	Value / Configuration
1	Embedding Dimension	128
2	LSTM Units (Dasar)	256
3	Batch Size	64
4	Epoch (Maximum)	200
5	Dropout Rate	0.3
6	Activation Function	Softmax
7	Optimizer	Adam, RMSprop
8	Loss Function	Categorical Crossentropy
9	Validation Split	0.1 (10%)
10	Callbacks	EarlyStopping & ModelCheckpoint
11	Hardware	GPU (Colab / CUDA enabled)

7. Evaluation Metrics

Accuracy evaluation is used to assess the accuracy of the model in predicting target tokens during the training and validation processes. This metric shows the extent to which the model is able to recognize patterns of relationships between words in sequential data. According to (Rotskoff & Vanden-Eijnden, 2022), consistent accuracy between training and validation data indicates that the optimization process is stable and that the model has good generalization capabilities. In this study, accuracy was calculated using the built-in TensorFlow–Keras metric at each epoch to monitor model performance and assess the effectiveness of the Adam and RMSProp optimizers.The loss function is used to measure the difference between the model's prediction and the actual target.

This study applies Categorical Crossentropy, which is commonly used in models with softmax outputs for sequence-based translation tasks. According to (Elharrouss et al., 2025), this function is able to accelerate network convergence because it updates the gradient proportionally to the error rate. The loss value is observed at each epoch to ensure learning stability and to compare the performance of the two optimizers in minimizing prediction errors.

BLEU (Bilingual Evaluation Understudy) is a quantitative metric that measures the similarity between machine translation and human translation results by calculating the similarity of word sequences or short phrases (n-grams) between the two (Elharrouss et al., 2025). BLEU scores range from 0 to 1, with higher values indicating more accurate translations. BLEU also uses a penalty component (brevity penalty, BP) to avoid overly short results. The BLEU formula is expressed as follows:

$$\mathrm{BLEU} = \mathrm{BP} \cdot \exp\left(\sum_{n=0}^{N} w_n \log P_n\right) \quad (6)$$

Description:

- BP is a penalty factor that takes into account sentence length,
- w_n is the weight for each n-gram precision,
- P_n shows n-gram precision at level n.

The logarithm function is applied to balance the contribution of each precision, while the exponential function modifies the final result to keep it within the range of 0–1. Penalties for sentences that are too short are applied using the following formula:

$$\mathrm{BP} = \exp\left(1 - \frac{r}{c}\right), \quad c \le r \quad (7)$$

$$\mathrm{BP} = 1, \quad c > r \quad (8)$$

With r as the reference length and c as the translation length. If the translation is shorter than the reference ($c < r$) then the BP value will be less than 1, causing the final score to decrease. Conversely, if the translation result is longer ($c > r$), then $BP = 1$, indicating no penalty because the sentence length is considered reasonable as long as the n-gram alignment remains high. However, this study has methodological limitations, particularly the limited size of the parallel corpus and the informal nature of some data sources, which may influence the generalization capability of the model.

RESEARCH RESULT

1. Translation Results

As a qualitative evaluation, several random test sentences were tested to assess the model's ability to translate Indonesian into Ngapak. This stage complements the quantitative results by comparing the model's output with human reference translations. The translation results from models with Adam and RMSprop optimizers are shown in Table 3.

Table 3. Translation Results of the Model (Adam) And (RMSprop) on Test Set

Indonesia	Ngapak (Original)	Ngapak (Adam Prediction)	Ngapak (RMSprop Prediction)
apakah saya meninggalkan sesuatu	apa enyong ninggalna sesuatu	apa enyong ninggalna sesuatu	apa enyong ninggalna sesuatu
mari kita coba lihat	ayuh dewek njajal ndeleng	ayuh dewek njajal ndeleng	ayuh dewek njajal ndeleng

hujan belum turun juga	udan urung medun mbireng	udan urung medun mbireng	udan urung medun mbireng
dia telah menjadi kepala suku selama telung puluh lima tahun	kae uis dadi kepala sukune selama telung puluh lima taun	kae uis dadi kepala sukune selama telung puluh lima taun	kae uis dadi ndas pamanke selama telung puluh lima taun
sudah lama sejak saya mengunjungi nenek saya	uwis sue kawit enyong ngunjungi nini enyong	uwis sue kawit enyong ngunjungi nini enyong	uwis sue kawit enyong ngunjungi rokok enyong
kesepakatanya batal di menit terakhir	kesepakatane batal neng menit terakir	nusawungu seni nang menit terakir	garingna celanane nang awakmu terakhir
dia disarankan olehnya untuk tidak pergi	kae disaranaken nang kae nggo ora lunga	kae disaranaken nang kae nggo ora lunga	kae disaranaken nang kae nggo ora lunga
telurnya lumayan besar	ndoge lumayan gede	ndoge lumayan gede	bima lumayan gede
lihatlah peta dunia	delenglah peta dunia	delenglah peta dunia	delenglah netes dunia
tom berpikir kalau kau berbohong	tom mikir mek kowe nglombo	tom mikir mek kowe nglombo	tom mikir mek kowe nglombo

The results in Table 3 show that the Adam and RMSprop models have different characteristics in producing translations from Indonesian to Ngapak. Both are capable of forming sentences that are generally consistent with the structure and meaning of the source, especially in simple conversation contexts. The Adam model tends to produce translations that are consistent in phrase structure and sentence patterns, while RMSprop shows variations in form and word order that give different nuances of expression.

However, in some cases, RMSprop outputs display minor semantic errors, such as the addition of irrelevant words or changes in meaning in certain parts. This shows that differences in optimization algorithms not only affect translation style, but also the level of accuracy of the results produced. Overall, both models are still able to capture the basic lexical and semantic relationships between the source and target languages with their respective characteristics.

2. Training Performance

The training process of the Bidirectional Long Short-Term Memory (Bi-LSTM) model based on the Attention mechanism was carried out to compare the effects of two optimization algorithms, namely Adam and RMSprop, on the speed and stability of convergence. Based on the experimental results, the model with Adam achieved a training accuracy of 99.7% and a validation accuracy of 95.5% at the 53rd epoch with a validation loss of 0.43, while RMSprop only achieved similar conditions at the 90th epoch with a training accuracy of 99.2%, a validation accuracy of 93.6%, and a validation loss of 0.49. The training process was automatically stopped by the EarlyStopping mechanism after the validation loss did not improve for 10 consecutive epochs, so Adam stopped faster because it had reached optimal convergence, while RMSprop required longer iterations to stabilize learning.

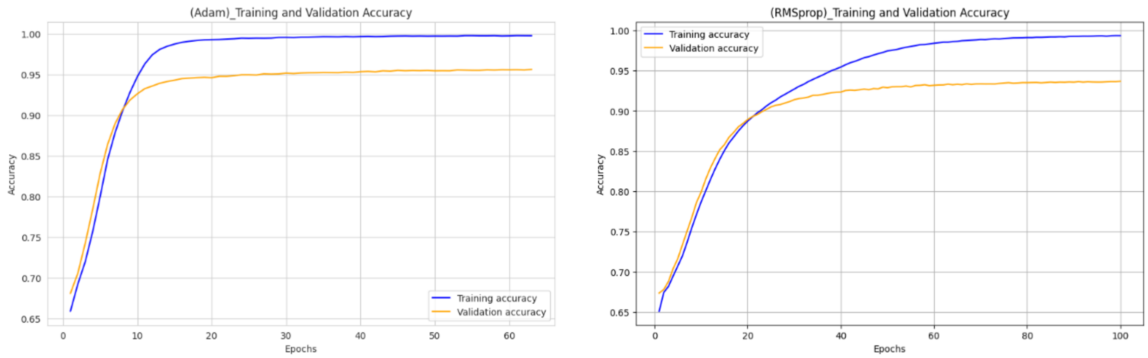

Figure 2. Accuracy Comparison of Adam and RMSProp

The visualization results in Figure 2 show that the accuracy curve of the model trained with Adam increases sharply from the start of training and reaches the convergence point relatively quickly. The gap between the training and validation accuracy lines appears narrow, indicating that the model's generalization ability to new data is performing well. Meanwhile, the RMSprop accuracy curve shows a slower rate of increase and requires more iterations to reach stability. Nevertheless, RMSprop still shows a consistent learning pattern and does not experience significant fluctuations during the training process.

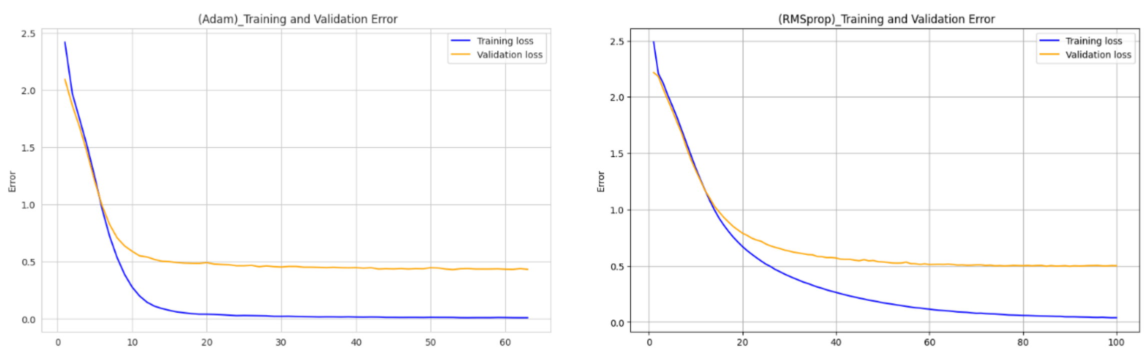

Figure 3. Loss Comparison of Adam and RMSProp

The curve in Figure 3 shows that the training loss in Adam decreases sharply at the beginning of the epoch and quickly reaches stability at a relatively small value. The validation loss value also appears to decrease proportionally with the training loss, indicating that the learning process is running efficiently and in a controlled manner. In contrast, RMSprop shows a more gradual decline and only reaches stability near the 80th epoch. Although the rate of decline is slower, RMSprop provides fairly good stability in weight updates on long sequential data, so it remains effective in maintaining training consistency.

2. BLEU Score

he quality of the translation results was evaluated using the Bilingual Evaluation Understudy (BLEU) metric, taking into account four levels of n-grams, namely BLEU-1 to BLEU-4. The BLEU score was used to measure the suitability of the model's translation results against human references, where a higher score indicates a greater degree of lexical and sentence structure similarity. Based on the test results, the model with the Adam optimizer obtained higher BLEU scores at each n-gram level compared to RMSprop. The BLEU-1 to BLEU-4 scores for Adam were 0.8775, 0.8317, 0.7887, and 0.7393, respectively, while RMSprop obtained scores of 0.8284, 0.7636, 0.7034, and 0.6384. This difference shows that models trained with Adam produce translations that are closer to human reference sentences and have a more natural structure.

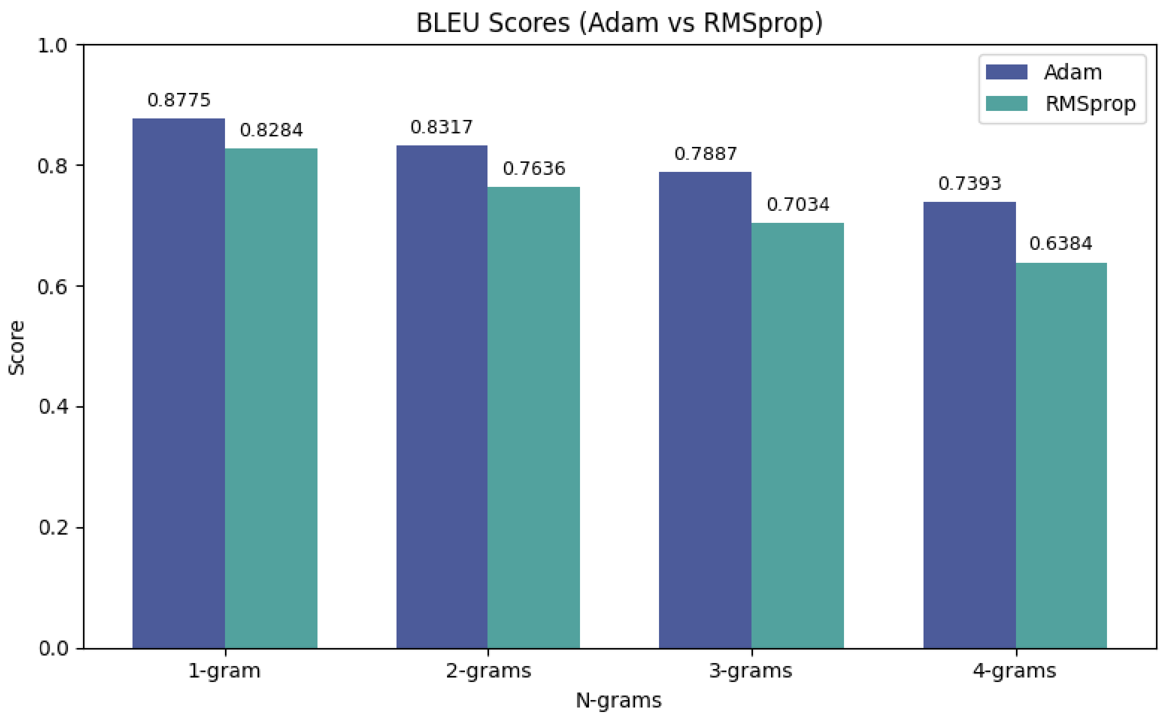

Figure 4. BLEU Score Comparison of Adam and RMSProp

Based on Figure 4, the difference in BLEU scores becomes more pronounced at higher n-grams. Adam is able to maintain BLEU score stability up to 4-grams with relatively small decreases, indicating that the model can capture word relationships in sentence context more completely. In contrast, RMSprop shows a sharper decline in scores at higher n-grams, indicating limitations in recognizing long-term dependencies. This condition is consistent with the previous training results, where Adam showed faster and more stable convergence, resulting in network weights that are more representative of language patterns.

The BLEU evaluation results prove that Adam provides better translation performance than RMSprop. The efficiency of Adam's parameter updates contributes to improved semantic alignment between sentences, while RMSprop tends to be slower in adjusting gradients, which affects the final translation quality. These findings confirm that the choice of optimization algorithm has a direct impact on the output quality of Bi-LSTM-based automatic translation systems.

DISCUSSION

The results show that the Adam optimization algorithm provides superior performance compared to RMSprop in the training process of the Bi-LSTM model with the Attention mechanism for Indonesian–Ngapak translation. Models trained using Adam show an increase in validation accuracy and a more stable decrease in loss values compared to RMSprop in the same training phase. In addition, the translation quality produced by the model with Adam also shows better lexical and semantic alignment with the reference translation, as reflected in the BLEU metric-based evaluation. These findings indicate that Adam has a more effective ability to adapt to gradient dynamics during the learning process, enabling the model to achieve optimal performance with greater time efficiency and faster convergence.

Adam's superiority is mainly due to its adaptive mechanism in adjusting the learning rate for each parameter, which makes the gradient update process more stable without large fluctuations and helps the model converge faster while reducing the risk of overfitting. These results are in line with the study (Ahda et al., 2024) which compares Adam and RMSprop in Minangkabau–Indonesian translation, where Adam shows faster convergence and more consistent translation results on limited source data. Meanwhile, RMSprop showed a slower but stable learning rate, as explained by (Z. Xie et al., 2022) that the adaptive characteristics of RMSprop are effective in maintaining gradient stability in long sequential data. In the context of Ngapak language translation, both training stability and efficiency are important due to the diversity of structures and complex vocabulary. Therefore, Adam is more suitable for accelerating learning in NMT systems with limited data, while RMSprop remains useful in maintaining training stability in sequential models such as Bi-LSTM.

The results of this study are in line with a number of previous studies that show the effectiveness of adaptive optimization algorithms such as Adam in accelerating the convergence process and maintaining training stability in sequential neural networks. Adam's adaptive mechanism allows automatic adjustment of the learning rate for each parameter, so that weight updates are more efficient and gradient fluctuations can be minimized. Similar findings are described in (Zou et al., 2023), which states that regularization in Adam can improve generalization capabilities and accelerate neural network learning stability. On the other hand, research shows that RMSprop remains superior in maintaining weight update consistency through(Elshamy et al., 2023) an exponential mean square mechanism that reduces gradient instability in long sequential data. In the context of Indonesian–Ngapak automatic translation based on the Bi-LSTM model with the Attention mechanism, both algorithms play an important role in balancing training efficiency and stability. This approach has been proven to help the model understand the semantic relationships between words more contextually and produce more natural sentence structures in regional languages with limited data sources.

This research is related to the development of an Indonesian–Banyumasan translation system based on RNN-GRU architecture, which is a dialect of Ngapak. In a previous study by Model Recurrent Neural Network Gated Recurrent Unit for Building an Indonesian-Banyumasan Translation Engine (Indonesia-banyumasan et al., 2023), a parallel corpus of 1,302 sentence pairs was collected and produced a BLEU score of 34.1. In contrast, this study uses a much larger corpus of 23,592 sentence pairs—and applies a Bi-LSTM model with an Attention mechanism and two optimization algorithms, namely Adam and RMSprop. The results show a significant improvement in all BLEU metrics: BLEU-1 to BLEU-4 reached 0.8775, 0.8317, 0.7887, and 0.7393, respectively, more than double that of the previous study. These observations confirm that the selection of model architecture and optimization algorithms has a direct impact on translation accuracy, training stability, and output quality, while also contributing to the preservation of the Ngapak language through the application of artificial intelligence technology that is adaptive to local culture.

Overall, the findings of this study support the conclusions of (Ahda et al., 2024), who reported that Adam consistently achieves faster convergence and higher BLEU scores in low-resource translation tasks. At the same time, our results extend the findings of (Z. Xie et al., 2022) by demonstrating that RMSprop's stability on long sequential data does not necessarily translate into higher translation quality for Ngapak. Furthermore, compared to (Indonesia-banyumasan et al., 2023), this study achieves substantially higher BLEU values, indicating that the combination of Bi-LSTM and Adam provides a more effective approach for local language translation.

The consistency between the increase in BLEU score and validation accuracy shows that the model is not only capable of replicating lexical structure, but also maintaining semantic meaning at the sentence level. This indicates that the Bi-LSTM model learning process with the Attention mechanism successfully captures contextual relationships between words, in line with the findings of (CAI & ZHU, 2025) and (Hosseinzadeh & Sadeghzadeh, 2025), which confirm that the attention mechanism allows the model to dynamically focus on the most relevant parts of the source sentence for the translation result.

This study has several limitations that need to be considered when interpreting the results. In terms of data, the number of parallel corpora used, amounting to 23,592 sentence pairs, is still relatively limited for a Neural Machine Translation (NMT)-based translation system, which generally requires a large volume of data for the model to adequately capture linguistic variation. This contrasts with the reality that NMT for low-resource languages often struggles to perform well due to the scarcity of parallel data and the heterogeneity of linguistic expressions, as described in a survey on low-resource NMT (Wang et al., 2021).

In addition, some of the research data was sourced from informal texts such as online conversations, so there may be inconsistencies in sentence structure, spelling variations, and typos that could affect the model's learning process. In low-resource environments, these conditions add to the burden on the model to generalize well due to linguistic noise and lack of domain control, as found in experimental studies for other low-resource languages (Tonja et al., 2023).

This study also has methodological and evaluative aspects that need to be considered. From a methodological perspective, the focus is still limited to Bi LSTM architecture with conventional Attention mechanisms and a comparison of two main optimization algorithms, namely Adam and RMSprop. However, recent studies show that architectures such as Transformer and hybrid mechanisms can significantly improve contextual representation and performance in resource-limited conditions. For example, (Araabi & Monz, 2020) argue that optimizing Transformers for low-resource conditions can add up to 7.3 BLEU points compared to the default settings.

This study utilizes the BLEU score metric to measure lexical similarity between translation results and human references. Although this metric provides a useful quantitative measure, research by (Lee et al., 2023) shows that BLEU tends to be less effective in capturing overall semantic similarity and sentence fluency. Therefore, further research development is recommended to include additional evaluations such as manual evaluation by native speakers, as well as expansion of the data corpus and application of modern models so that the results are more comprehensive and implementable in the context of local language translation such as Ngapak.

From a technical perspective, the results of this study show that the application of adaptive optimization algorithms such as the Adam optimizer on the Bi LSTM model with the Attention mechanism can strengthen training efficiency and translation quality in limited data conditions, as discussed in a review of adaptive optimization algorithms and the latest neural network architecture (R. Li & Liu, 2025). These findings provide an important basis for the development of regional language-based automatic translation systems. Therefore, it is recommended that future research integrate more advanced architectures such as Transformers and expand parallel corpora so that models can handle more complex and varied sentence contexts.

This research makes an important socio-cultural contribution in the context of preserving the Ngapak language as part of local linguistic wealth. By implementing artificial intelligence technology in automatic translation systems, not only is the accuracy of interlingual communication improved, but digital tools are also provided that enable regional languages to remain active in the modern era. As indicated by (Sihite & Sibarani, 2024), digital tools such as mobile applications, online dictionaries, and social platforms have successfully strengthened documentation, learning, and community involvement in the preservation of regional languages in Indonesia. Furthermore, study (Review, 2025) emphasizes that the integration of technology and community education is key to language revitalization strategies. Thus, this research is not only a technical step in the development of translation systems, but also part of a broader technological ecosystem that supports the preservation and revitalization of local languages in Indonesia for future generations.

This finding also opens up opportunities for applying a similar approach to other regional languages in Indonesia that have morphological characteristics similar to Ngapak, such as Banyumasan, Madurese, or Minangkabau. The Bi-LSTM and Attention-based Neural Machine Translation approach can be adapted to overcome the limitations of low-resource languages, as recommended by (Buchanan, 2023).

CONCLUSION

This study compares the performance of two optimization algorithms, Adam and RMSprop, on a Bidirectional Long Short-Term Memory (Bi-LSTM) model with an Attention mechanism in the Indonesian–Ngapak translation task. The test results show that the model with Adam converges faster with a validation accuracy of 95.5%, validation loss of 0.43, and BLEU-1 to BLEU-4 scores of 0.8775, 0.8317, 0.7887, and 0.7393, respectively. Meanwhile, the model with RMSprop showed slower convergence with a validation accuracy of 93.6%, validation loss of 0.49, and BLEU-1 to BLEU-4 scores of 0.8284, 0.7636, 0.7034, and 0.6384, respectively. This comparison proves that Adam produces better and more stable translation quality compared to RMSprop.

From these results, it can be concluded that Adam is more efficient and adaptive in updating model parameters, thereby accelerating convergence and producing more accurate translations in regional language-based translation systems with limited data. However, RMSprop remains superior in gradient stability during long-term training. Technically, this research contributes to the development of Neural Machine Translation (NMT) for low-resource languages, while socio-culturally, it supports the preservation of the Ngapak language through the integration of artificial intelligence technology. Further research is recommended to develop Transformer-based models and expand parallel corpora so that translation systems can represent local linguistic and cultural contexts more comprehensively.

REFERENCES

Abdulkadirov, R., & Lyakhov, P. (2023). *Survey of Optimization Algorithms in Modern Neural Networks*. 1–37.

Abidin, Z., Sucipto, A., & Budiman, A. (2018). *PENERJEMAHAN KALIMAT BAHASA LAMPUNG-INDONESIA DENGAN PENDEKATAN NEURAL MACHINE TRANSLATION BERBASIS ATTENTION TRANSLATION OF SENTENCE LAMPUNG-INDONESIAN LANGUAGES WITH NEURAL MACHINE TRANSLATION ATTENTION BASED*. *06*(02), 191–206.

Ahda, F. A., Wibawa, A. P., Prasetya, D. D., & Sulistyo, D. A. (2024). Comparison of Adam Optimization and RMSprop in Minangkabau- Indonesian Bidirectional Translation with Neural Machine Translation. *International Journal on Informatics Visualization*, *8*(1), 231–238. https://doi.org/10.62527/joiv.8.1.1818

Araabi, A., & Monz, C. (2020). Optimizing Transformer for Low-Resource Neural Machine Translation. *COLING 2020 - 28th International Conference on Computational Linguistics,* *Proceedings of the Conference*, *2019*, 3429–3435. https://doi.org/10.18653/v1/2020.coling-main.304

Buchanan, L. (2023). A new condition on the Jones polynomial of a fibered positive link. *Journal of Knot Theory and Its Ramifications*, *32*(14), 1–12. https://doi.org/10.1142/S0218216523500797

CAI, G., & ZHU, J. (2025). Interpreting Attention Mechanisms of NMT with Linguistic Features. *IEICE Transactions on Information and Systems*. https://doi.org/10.1587/transinf.2024edp7292

Elharrouss, O., Mahmood, Y., Bechqito, Y., Serhani, M. A., Badidi, E., Riffi, J., & Tairi, H. (2025). *Task-based Loss Functions in Computer Vision: A Comprehensive Review*. http://arxiv.org/abs/2504.04242

Elshamy, R., Abu-Elnasr, O., Elhoseny, M., & Elmougy, S. (2023). Improving the efficiency of RMSProp optimizer by utilizing Nestrove in deep learning. *Scientific Reports*, *13*(1), 1–16. https://doi.org/10.1038/s41598-023-35663-x

Fauziyah, Y., Ilyas, R., & Kasyidi, F. (2022). *MESIN PENTERJEMAH BAHASA INDONESIA-BAHASA SUNDA MENGGUNAKAN RECURRENT NEURAL NETWORKS*. *16*, 313–322.

Gunawan, W., Sujaini, H., & Tursina, T. (2021). Analisis Perbandingan Nilai Akurasi Mekanisme Attention Bahdanau dan Luong pada Neural Machine Translation Bahasa Indonesia ke Bahasa Melayu Ketapang dengan Arsitektur Recurrent Neural Network. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, *7*(3), 488. https://doi.org/10.26418/jp.v7i3.50287

Hosseinzadeh, R., & Sadeghzadeh, M. (2025). Attention Mechanisms in Transformers: A General Survey. *Technology Journal of Artificial Intelligence and Data Mining*, *13*(3), 359–368. https://doi.org/10.22044/jadm.2025.15584.2679

Indonesia-banyumasan, B., Wijanarko, A., Nisa, A., Haura, A., Puspitaningrum, I., & Surya, D. I. (2023). *Model Recurrent Neural Network-Gated Recurrent Unit untuk Membangun Mesin Penerjemah*. 218–226. https://doi.org/10.30864/eksplora.v13i2.977

Israr, H., Khan, S. A., Tahir, M. A., Shahzad, M. K., Ahmad, M., & Zain, J. M. (2023). Neural Machine Translation Models with Attention-Based Dropout Layer. *Computers, Materials and Continua*, *75*(2), 2981–3009. https://doi.org/10.32604/cmc.2023.035814

Keskar, N. S., & Socher, R. (2017). *Improving Generalization Performance by Switching from Adam to SGD*. *1*. http://arxiv.org/abs/1712.07628

Lee, S., Lee, J., Moon, H., Park, C., Seo, J., Eo, S., Koo, S., & Lim, H. (2023). A Survey on Evaluation Metrics for Machine Translation. *Mathematics*, *11*(4), 1–22. https://doi.org/10.3390/math11041006

Li, C., & Flanigan, J. (2022). *Improving Neural Machine Translation with the Abstract Meaning Representation by Combining Graph and Sequence Transformers*. *Dlg4nlp*, 12–21.

Li, R., & Liu, R. (2025). The Central Role of Adaptive Optimization Algorithms in Deep Learning: A Cross-Domain Survey from CNNs to Transformers. *Applied and Computational Engineering*, *158*(1), 1–10. https://doi.org/10.54254/2755-2721/2025.tj23299

Loshchilov, I., & Hutter, F. (2019). Decoupled weight decay regularization. *7th International Conference on Learning Representations, ICLR 2019*.

Malladi, S., Lyu, K., Panigrahi, A., & Arora, S. (2022). On the SDEs and Scaling Rules for Adaptive Gradient Algorithms. *Advances in Neural Information Processing Systems*, *35*.

Marco, M. W. Di, & Fraser, A. (2022). Findings of the WMT 2022 Shared Tasks in Unsupervised MT and Very Low Resource Supervised MT. *Conference on Machine Translation - Proceedings*, 801–805.

Nugroho, C., & Kusuma, I. P. (2023). Identitas Budaya Banyumasan dalam Dialek Ngapak. *Jurnal Ilmu Komunikasi*, *21*(2), 333. https://doi.org/10.31315/jik.v21i2.4556

Puspitaningrum, D. (2021). A Study of English-Indonesian Neural Machine Translation with Attention (Seq2Seq, ConvSeq2Seq, RNN, and MHA): A Comparative Study of NMT on English-Indonesian. *ACM International Conference Proceeding Series*, 271–280. https://doi.org/10.1145/3479645.3479703

Putri, F. I., Wibawa, A. P., & Collante, L. H. (2024). Refining the Performance of Indonesian-Javanese Bilingual Neural Machine Translation Using Adam Optimizer. *ILKOM Jurnal Ilmiah*, *16*(3), 271–282. https://doi.org/10.33096/ilkom.v16i3.2467.271-282

Review, A. N. (2025). *Lingua : Journal of Linguistics and Language*. *2*, 77–88.

Rotskoff, G., & Vanden-Eijnden, E. (2022). Trainability and Accuracy of Artificial Neural Networks: An Interacting Particle System Approach. *Communications on Pure and Applied Mathematics*, *75*(9), 1889–1935. https://doi.org/10.1002/cpa.22074

Sihite, M. R., & Sibarani, B. (2024). Technology and Language Revitalization in Indonesia: A Literature Review of Digital Tools for Preserving Endangered Languages. *International Journal of Educational Research Excellence (IJERE)*, *3*(2), 610–620. https://doi.org/10.55299/ijere.v3i2.988

Tonja, A. L., Kolesnikova, O., Gelbukh, A., & Sidorov, G. (2023). Low-Resource Neural Machine Translation Improvement Using Source-Side Monolingual Data. *Applied Sciences (Switzerland)*, *13*(2). https://doi.org/10.3390/app13021201

Wang, R., Tan, X., Luo, R., Qin, T., & Liu, T. Y. (2021). A Survey on Low-Resource Neural Machine Translation. *IJCAI International Joint Conference on Artificial Intelligence*, 4636–4643. https://doi.org/10.24963/ijcai.2021/629

Wu, Y., & Xing, Y. (2024). *Efficient Machine Translation with a BiLSTM-Attention Approach*. http://arxiv.org/abs/2410.22335

Xie, T., Ding, W., Zhang, J., Wan, X., & Wang, J. (2023). Bi-LS-AttM: A Bidirectional LSTM and Attention Mechanism Model for Improving Image Captioning. *Applied Sciences (Switzerland)*, *13*(13). https://doi.org/10.3390/app13137916

Xie, Z., Wang, X., Zhang, H., Sato, I., & Sugiyama, M. (2022). Adaptive Inertia: Disentangling the Effects of Adaptive Learning Rate and Momentum. *Proceedings of Machine Learning Research*, *162*, 24430–24459.

Zou, D., Cao, Y., Li, Y., & Gu, Q. (2023). Understanding the Generalization of Adam in Learning Neural Networks With Proper Regularization. *11th International Conference on Learning Representations, ICLR 2023*.