ISSN: 2093-5374 / E-ISSN: 2093-5382

Neural Machine Translation Model Using GRU with Hybrid Attention Mechanism for English to Kannada Language

Gunti Spandan[1], and Dr. Prasannavenkatesan Theerthagiri[2*]

[1]Department of Computer Science and Engineering, GITAM School of Technology, GITAM University Bengaluru, Bengaluru, India. sgunti@gitam.edu, https://orcid.org/0000-0001-6059-2835

[2*]Department of Computer Science and Engineering, GITAM School of Technology, GITAM University Bengaluru, Bengaluru, India. vprasann@gitam.edu, https://orcid.org/0000-0003-3420-598X

Received: July 29, 2024; Revised: September 08, 2024; Accepted: October 07, 2024; Published: December 30, 2024

Abstract

Neural machine translation is a machine translation system that uses artificial neural networks to identify nonlinear relationships between bilingual sentence pairs. The language English-Kannada pair has met with less attention in the field of Neural Machine Translation (NMT), mostly due to lack of parallel corpora and the complexity of Kannada's linguistic structure. This research proposes a novel NMT model based on Gated Recurrent Units (GRU) and a Hybrid attention mechanism designed specifically for Kannada's morphological complexity and compared with existing models. By adding language-specific preprocessing techniques and employing data augmentation tactics, our model outperforms previous LSTM, GRU and Bi-LSTM-based algorithms in terms of BLEU and METEOR scores. The proposed model with hybrid attention might be a useful substitute in low-resource settings because transformers and other large models require more resources. The comparative study indicates that the proposed methodology improves the translation by 10% increase in Bilingual Evaluation Understudy (BLEU) and accuracy of 80% and achieved a Rouge Score of 0.9.

Keywords: Neural Machine Translation, English-Kannada, Natural Language Processing, Gated Recurrent Networks.

1 Introduction

A combination of computer science, information technology, natural language, and expert systems is known as natural language processing (NLP). it is concerned with building intelligent computer systems that would interpret spoken and written language, translate text, read handwritten and printed texts, analyse text, and interact with databases using natural language (Singh et al., 2022).

It has numerous applications, including text summarization (MT), information extraction, and machine translation. Machine translation is one use case for computational linguistics (NLP). Its purpose is to convert text from one language (source) into another (target).

Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA), volume: 15, number: 4 (December), pp. 259-276. DOI: *10.58346/JOWUA.2024.I4.017*

*Corresponding author: Department of Computer Science and Engineering, GITAM School of Technology, GITAM University Bengaluru, Bengaluru, India.

The technique of translating written words, phrases, and sentences into natural language using computer programs and instructions is known as machine translation. The area of natural language processing that is now the most significant is machine translation. Due to globalization, English language (Al-Barhamtoshy & Metwalli, 2022) resources are available on 52% of websites. Machine translation software has frequently been utilized online. They continue to provide many inaccurate translation results.

Kannada, a Dravidian language with rich historical literature, has few resources in terms of machine learning, making it a tough assignment owing to the semantic and syntactic diversity in its publications (Paul et al., 2023). Kannada, in comparison to other Indian languages, has received less attention in machine translation (Ali & Raj, 2024; Ismail, 2024).

The different types of translation models are Rule-based, Statistical based, hybrid, and neural machine translations for translating from source language to target (Sarode et al., 2023; Balaji et al., 2022). The neural machine translation paradigm offers the advantage of employing a single model for source and destination language decoding and training (Tohma & Kutlu, 2020; Kaggle, dataet).

1.1. Current Research in NMT

The field of Neural Machine Translation (NMT) is undergoing a rapid evolution marked by numerous notable features. First, models with multilingual and zero-shot translation capabilities. The field of NMT is evolving rapidly. It is now more flexible because it can translate across languages without requiring a translation language (Bala Das et al., 2023).

Acquiring more knowledge about the world is essential for providing more accurate translations in fields like law or health. NMT is figuring out ways to improve even for languages with limited resources. Additionally, novel approaches combining machine and human translation to achieve the best of both worlds. In general, NMT is becoming more intelligent, faster, and beneficial to all users (Bala Das et al., 2023).

1.2. Neural Machine Translation

Neural machine translation (Bakarola & Nasriwala, 2021) is a technique that predicts a word sequence's likelihood using artificial neural networks. The newest method for achieving automatic machine translation is called neural machine translation, or NMT. Neural networks mimic the human learning process and relating the source sequence to a suitable target sequence. Recurrent neural networks model the long-term dependencies between the source and target languages.

Sequence to Sequence (Encode Decoder Architecture)

The NMT usually consists of sequence-to-sequence architecture where input sentence language is taken as sequence and will return output as target language. The model is called as encoder-decoder architecture (Nath et al., 2024). The Figure 1 shows the simple encoder and decoder architecture. The encoder’s main task is inputting sequences and convert into the fixed length vector known as context vector. The context vector comprises the meaning of the input sequence and it is taken care by decoder as input to decode the essential output sequence.

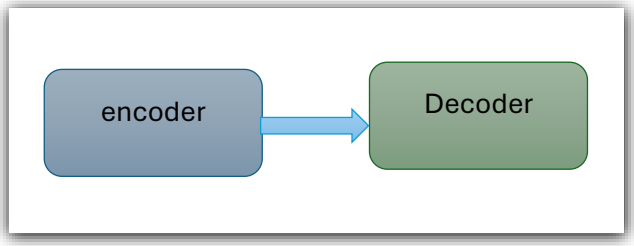

Figure 1: Encoder-Decoder Architecture

Recurrent neural networks are desirable for the neural machine translation and are suitable for sequential and time series data. The RNN (Nath et al., 2024) compromises of memory states which are reasonable for getting sentences which will be helpful during the guessing of sequential data.

Organization of the Paper

The following describes the way the paper is arranged. In Section 2, relevant works in the translation realm are reviewed and discussed. Section 3 deals with the design of the proposed machine translation model with Hybrid attention for taken datasets. The evaluation approach for the suggested neural network translation model and its results comparison with the existing models are described in Section 4. Section 5 discuss about the conclusion of the paper.

2 Literature Survey

In this paper (Unanue et al., 2022), author proposes an approach for regularization during training neural machine translation models on a low resource language pair to increase their accuracy. The experiment was conducted on three datasets low resource and a variable size. The work was conducted on LSTM and transformer models. The limitation was increasing the quality of the translations. Both the models were pretrained on fast text embeddings.

The system is based on Gated Recurrent unit (GRU) and consists of a gating mechanism of recurrent neural network. In this approach they have used 5 different types of RNN Algorithms The paper (Jahan et al., 2021) uses an English to German language pair of 2.2M from manythings.org website. The hybrid model designed provides a good accuracy of 85.69% and a bleu score of 54.40% for unigram.

The paper (Al-Barhamtoshy & Metwalli, 2022) consists of a bilingual English to Arabic translation model and aims to expand for multilingual language pairs. The proposed translation systems consist of CNN to generate source languages vectors and RNN as a decoder to generate translated text. The model uses LSTM to find the translation results. The accuracy of 69.61% has obtained on testing dataset and & 70% for training.

The author (Mahmud et al., 2021) has experimented on a gru-based unidirectional recurrent neural networks model translating from English to Bangla in which the research has not conducted yet. Bahdanua attention was used while translation word relevance. The experiment was conducted on a small-sized dataset consists of 1014 sentences. The model achieved Bleu Score 50.07. Future work proposed is working with different types of RNN and could use beam search.

A neural machine translation model from Bengali to English using sequence models has been created in the work (Pathak & Pakray, 2019). The four different Sequence to Sequence models are Long Short-term Memory (LSTM), Gated Recurrent Unit (GRU), Bi-directional (LSTM) and Bi-directional GRU.

The models have achieved Blue score of 47.4,35.8,32.0 and 22.8 and for BLEU-1,2,3 & 4 respectively on LSTM Models.

The proposed neural network model comprises a Gate Recurrent Unit and Recurrent neural networks. The model comprises an RNN Encoder and RNN Decoder and attention mechanism which mainly handles long sentence. BLEU Score was not up to the mark and model performed well on sentences under 8. Result numbers were not presented in paper (Sarode et al., 2023).

The paper (Ismail, 2024) describes the sequence-to-sequence encoder-decoder model comprising GRU & attention mechanism to enhance the performance of NMT. The dataset is collected from an online source partially created and consists of 4500 entities. The accuracy of the model was about 75%. Due to hardware requirement the project was limited.

The author (Saleem & Uprety, 2021) discussed about the translation model of NMT model from English to Hindi. The encoder-decoder model consists of LSTM with Attention mechanism and without Attention was compared. Encoder-Decoder model with Attention achieved 81% accuracy and other without attention model achieved 26%.

The work aims are to create a NMT system using a pre-trained word embedding for Kashmir to English and Kashmir to Hindi translation. The paper implements a transformer based NMT model using a pretrained word and sentences level embeddings and can improve short term translation efficiency. The best bleu score was 6.69. The paper (Sheshadri et al., 2022) suggest that translation efficiency can be improved by more parallel corpus.

The neural machine translation work focus on a hybrid attention model. This hybrid attention model comprises of additive layers and multiplicative attention layers. The limitation of this project (Nath et al., 2024) is out of vocabulary terms need to be addressed and model takes large time for training and can be tested with large datasets.

The study reveals building a neural machine translation from English to Indian Languages using transformer architecture. One of the approaches called back translation methods is used to improve the bleu score for translation of English to Hindi and Hindi to Bengali languages. The paper (Kandimalla et al., 2022) suggests that it can be extended for monolingual datasets of different sizes.

The paper aims to translate from English text into Urdu using a neural machine translation approach. Word2vec word distribution is used in the model. Bleu metric was to evaluate the results which yield a good score. The model (Kumhar et al., 2023) proposed can be applied to Romance languages also. The bleu score drops abruptly after the sequence length of 10.

In this paper (Laskar et al., 2023), the author has considered a neural machine translation approach for a Englis-Assamese language using the transformer based architecture. As the initial process the authors has used RNN-based NMT with attention and transformer model. A bleu scores 15.14 for English to assam and 19.12 for Assam to English was achieved. As a limitation a corpus can be increased and can be apply to the transfer learning-based approach.

The author (Bala Das et al., 2023) proposes a multilingual machine neural translation system to overcome the problems associated to low-resource translation. The model consists of two types i.e. English to Indian languages with a part encoder-decoder consisting of 15 languages translations. Transformer model was used in the model proposed. Some techniques like back-translation improved translation and minor improvements of 0.86 average bleu score were achieved.

The work (Israr et al., 2023) consists of an Urdu to English machine translation using different techniques of encoder decoder like long, short-term memory (LSTM), bi-directional RNN, Gated

Recurrent Units, Convolution neural networks and transformer architecture. The model suffers from under translation and transformer translation results were not good.

The author (Chen, 2023) proposes a new method for word embedding which adds the prior knowledge between the mapping of words during the training process. The experiment was performed on the transformer and Bert model was used to pretrain embeddings and relation embeddings. The experiment was conducted on large language sources like English and German with +0.9 BLEU points and Spanish to Czech low resource language and had an +2.6 Bleu score.

The author (Bakarola & Nasriwala, 2021) describes a NMT model comprising of encoder-decoder including a attention principle with a Gated recurrent unit. The proposed approach performs better than statistical machine system with a bleu score of 53% and compared to 34.56% and accuracy was also good.

Using an attention mechanism the neural machine translation model (Prasanna & Latha, 2023) was used on the Sanskrit to Hindi and Sanskrit to Gujarati. A bleu score of 18.3 was achieved. A bilingual dataset was developed with the Sanskrit-Hindi and Hindi-Sanskrit.

The paper (Kaggle, dataet) aims to create a model on machine translation from English to Tamil Language. The proposed model consists of a Gated Recurrent Unit and Long Short-term memory. The model performed with a bleu score of 0.9 and humans manually evaluated it (Sethi et al., 2023). The model was performed efficiently with less time and space complexity.

3 Proposed Methodology

The Figure 2 presents the proposed methodology for translating English to Kannada using NMT. The Source language is English, passed through the process of preprocessing and tokenization given to a Encoder NMT Model and results in a given context vector. The decoder model will context vector as input and produce the target language

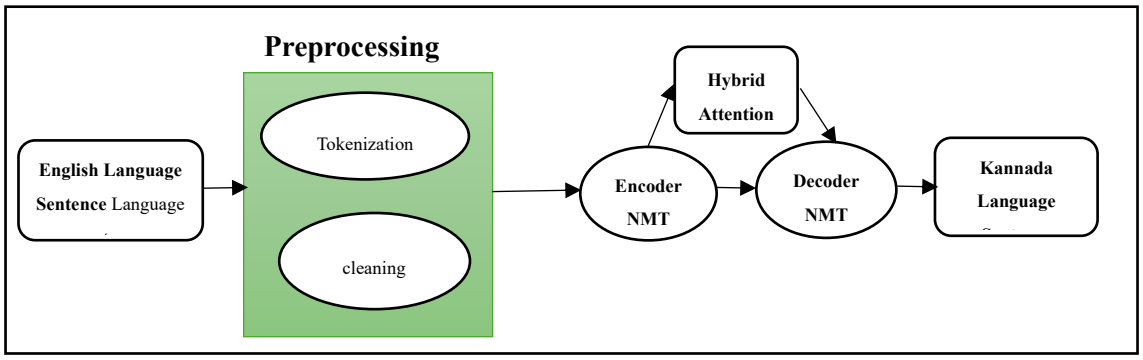

Figure 2: Architecture Flow Model

The proposed methodology will use tokenization and cleaning as a part of preprocessing. The dataset consists of English sentences in which the uppercase letters are converted into lower case, punctuation is eliminated, and digits are removed. Data should be converted into machine understandable model to train the model.

The tokenization process (Mahmud et al., 2021) will always occur before the model is designed. A token of <start> and <end> token is padded to both the sides of the sentence to get the front and back of the sequence while training the model. NLTK Tokenizer is being used and padding is applied to ensure that all the sequences in the dataset will have equal length.

3.1. Pre-trained Word Embeddings

The fundamental purpose of word embedding is to keeps the words in vectors. Word embedding have a vector space representation of words with comparable meanings Different types of word embedding techniques exits for word vectors (Sheshadri et al., 2022). Several word embeddings have word vector representation including fast text, word2vec, and Glove. In this study, Glove word Embedding representation was used. Glove embedding consists of pre-trained word embeddings and is used to initialise the embedding layer of a neural network by generating an embedding matrix.

3.2. NMT Model Architecture

The proposed model has been built upon the sequence-to-sequence learning framework with hybrid attention. The model consists of a hybrid attention network, a decoder network, and an encoder network.

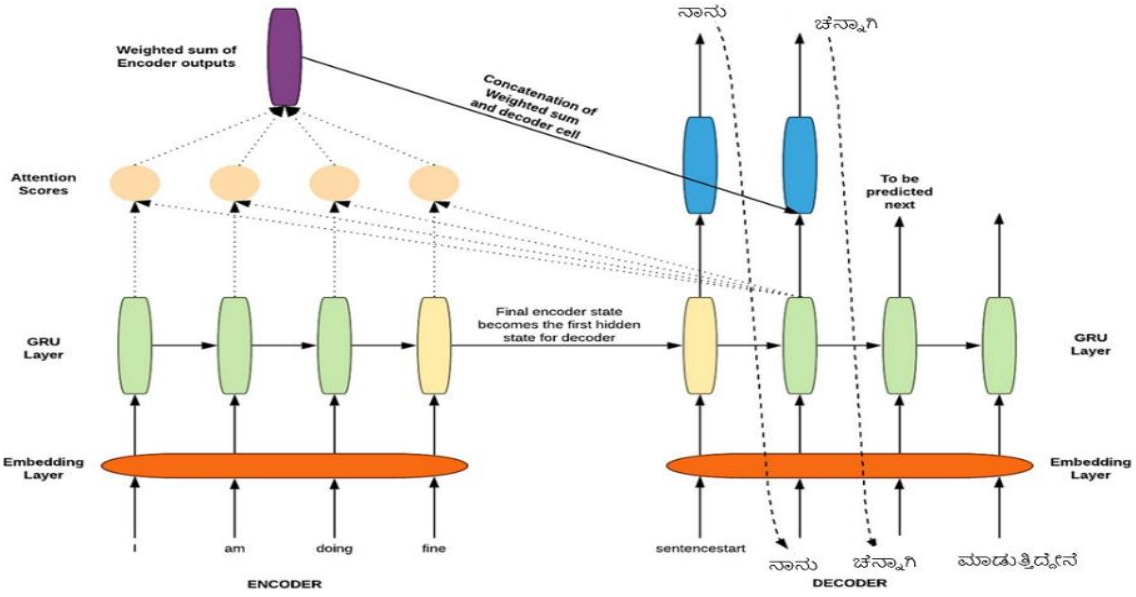

Figure 3: Proposed GRU NMT Architecture

The encoder part encrypts a source sentence using RNN and GRU units. The decoder module consists of RNN With GRU units to produce the target sentence (Ismail, 2024). The proposed GRU encoder and decoder architecture and hybrid attention mechanism is illustrated in the Figure 3.

Encoder and Decoder

The sequence model comprising encoder and decoder is commonly used for performing the neural machine translation. The key benefit of the approach is that we may train directly from source to destination by providing the source sentences and target sentences it can handle the various length of given the input and output sequences respectively. The architecture (Ismail, 2024) will have two LSTMs: one at the encoder and another at the decoder. This model uses GRU (Gated Recurrent Units) to replace the two LTSMs since GRU Computes less processing power while providing the same results as LTSMs.

Steps in Encoding

a) The tokenized sentences consist of small units called words embedded with the input sentences and are defined with different space dimensions that are defined in glove embeddings (Ismail, 2024).

b) The embeddings will subsequently be submitted to the GRU.The encoder GRU, characterized by the final hidden state, will be used as the decoder system's beginning hidden state.

Steps in Decoding

a) An embedded layer is in the decoder for the sequence in the destination language. In the embedded space every word is defined which contains meaning the proximity.

b) On the decoder side, an attention layer (Ismail, 2024) is used to weight the total of the current decoder hidden state and encoder output to obtain the encoder outputs.

c) The results obtained from the last two steps are put together and this result is called the final tensor which will be to send to decoder comprising of GRU Layer.

d) The product of the GRU Layer passes to the density layer, which calculates the likelihood of all words that will emerge.

Gated Recurrent Unit

The Standard RNN have a memory state which will perform well for the sequential data by storing the previous dependencies which help predict sentences. Problem with Standard RNN''s will suffer from vanishing and exploding gradient issues and model would not be able to cover since the gradients stop changing will have an increasing epoch. The different variants of standard RNN's are LSTM and GRU (Nath et al., 2024). The abstract view of GRU is shown in Figure 4.

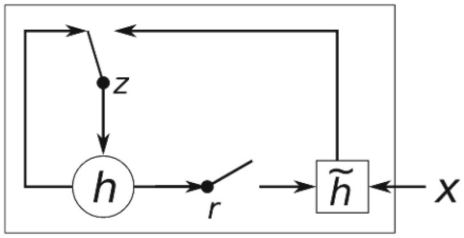

Figure 4: Abstract View of GRU

GRU's is type of RNN which will reduce the architecture of LSTM and consume less computational time and training time. GRU consists of two gates update gates and reset gates and single final hidden state (Nath et al., 2024).

Update Gate: This gate determines how much previously hiding state information will be passed on to the present hidden state. This gate is like output gate of LTSM repeating unit. The update gate determines whether to add the new hidden state to the hidden state. it is mathematically expressed as Equation 1.

$$\text{Update Gate, } zi = \sigma([Wz\, x] + [Uzht - 1]i) \quad (1)$$

where x and $h_t - 1$ are the input and prior hidden state, σ is the activation function, and [i] denotes the i^{th} element of a vector. Trainable weight matrices are W_z and Uz.

Reset Gate: This gate becomes helpful for filtering information that won't be relevant in the future. The reset gate determines whether to eliminate the previously hidden state data.

Reset, update, and hidden state gates can be computed while considering the ith time step, and the hidden states can be mathematically expressed as Equation 2.

$$\text{Reset gate, } ri = \sigma([Wr\ x] + [Ur\ ht - 1]i) \quad (2)$$

Where x and $h_t - 1$ are the input and prior hidden state, σ is the activation function, and [i] denotes the i^{th} element of a vector. Trainable weight matrices are W_r and Ur.

Attention Mechanism

Whenever longer sequences in NMT Models consists of an ordinary encoder-decoder architecture cannot generate a correct translation (Nath et al., 2024). The decoder uses the source sequence's "meaning" is utilised by the decoder to generate the target sequence. The context vector actively represents the encoder's hidden states with attention process. it allows the decoder to access separate encoder hiding states at various times.

Proposed Hybrid Attention Mechanism

This paper's primary feature is the integration of a suggested hybrid attention mechanism (Singh et al., 2022) in a neural translation model. For a given corpus, hybrid attention can determine which of the additive and multiplicative attention mechanisms is best. The loss calculated for each prediction made by both attention systems is the fundamental concept of hybrid attention.

A brief overview of hybrid attention architecture is given in Figure 5, where the hybrid attention mechanism is divided into one section. A single layer of unidirectional GRU with 1024 hidden units makes up the encoder and decoder. The word embedding vectors of the source and target tokens are obtained using source and target embedding layers. A completely connected layer with 1024 units serves as the decoder's final output layer.

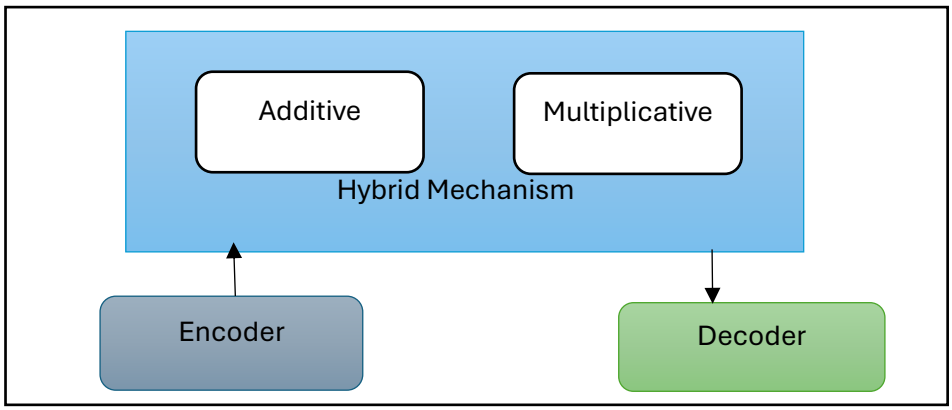

Figure 5: Hybrid Attention Mechanism

3.3. Dataset

The dataset that has been used is a bilingual sentence pairs which consists of three separate datasets. Among this ted (Shah & Bakrola, 2019) and Samantar (Tiwari et al., 2020) are online dataset collected, and one dataset named FUS (OPUS, 2020) which is taken from Kaggle repository. The Table 1 shows the different datasets with the number of sentences.

Table 1: Datasets

Dataset Name	Sentences
Ted	2.3 k
FUS	5 K
Samantar	4M

The three datasets utilized included 85% data for the training set and 15% for the testing set. The number of samples used to train, test and validate while preserving the 10% data is spitted for validation reasons. A sample of sentences from all three datasets are represented as the Table 2. Table 3. the Table 4.

Table 2: FUS Dataset

English Sentences	Kannada Sentences
I will be back soon.	ನಾನು ಶೀಘ್ರದಲ್ಲಿ ಮರಳುತ್ತೇನೆ.
I'm at a loss for words.	ನಾನು ಪದಗಳಿಗೆ ನಷ್ಟದಲ್ಲಿದ್ದೇನೆ.
This is never going to end.	ಇದು ಎಂದಿಗೂ ಕೊನೆಗೊಳ್ಳುವುದಿಲ್ಲ.

Table 3: Ted Dataset

English Sentences	Kannada Sentences
It's very complicated to explain.	ಇದನ್ನ ವಿವರಿಸೋದು ಸ್ವಲ್ಪ ಕಷ್ಟ.
But anyway, the public still doesn't buy it.	ಜನರಂತೂ ಅದನ್ನು ನಂಬಲಿಲ್ಲ .
It is still boiling.	ಈಗಲೂ ಅದರ ಬಿಸಿಯಾಡುತ್ತಿದೆ.

Table 4: Samantar Dataset

English Sentences	Kannada Sentences
I read a lot into this as well.	ಇದರ ಬಗ್ಗೆ ನಾನೂ ಸಾಕಷ್ಟು ಓದಿದ್ದೇನೆ.
Granted, the standard of living varies from country to country.	ಜೀವನಮಟ್ಟವು ದೇಶದಿಂದ ದೇಶಕ್ಕೆ ಬದಲಾಗುತ್ತದೆ ನಿಜ.
Thats a fine plan.	ಇದೊಂದು ಉತ್ತಮ ಯೋಜನೆ.

3.4. Evaluation Metrics

ROUGE: A collection of criteria called the Recall-Oriented Understudy for Gusting Evaluation, or ROUGE Score, is used to assess how well summarization and document translation models perform (Bakarola & Nasriwala, 2021).

A number around zero indicates low similarity between the candidate and references, whereas a score near one indicates significant similarity. The score goes from 0 to 1. ROUGE scores are divided into three categories: ROUGE-1, ROUGE-2, and ROUGE-L. The formula for calculating ROUGE-N is showing in Equation-3.

$$ROUGE - N = \frac{\sum s \in \{reference\ Summaries\} gram_n \in s \sum Count_{match}(gram_n)}{\sum s \in \{reference\ Summaries\} gram_n \in s \sum Count(gram_n)} \quad (3)$$

BLEU: When assessing the accuracy of human-generated reference translations, the bleu score is an essential indicator. To evaluate the adequacy and fluency of machine translation output, the Bilingual Evaluation Understudy (BLEU) measure assigns a score of 0 to 1 (Bakarola & Nasriwala, 2021). The more closely the test sentences align with their human reference translations, the closer they score to 1. The Formula for calculating blue is shown in equation 4.

$$BLUE = Min(1, \exp(1 - \frac{refereence-length}{output-length})) \quad (4)$$

Accuracy: Performance is one of the metrics we look for in the machine learning space to gauge how well our model is working. The ability of a neural network to anticipate outcomes is measured by its accuracy. It is determined by dividing the total number of predictions by the proportion of accurate predictions (Bakarola & Nasriwala, 2021).

4 Implementation and Evaluation

The proposed model has been implemented using the following requirements. it consists of a CPU with i7 Processor with 16 GB Ram and internal hard disk of 512 GB and graphic card of 8 GB Nvidia GE RTX Graphics Memory

We have used three datasets namely ted, Samantar and collection of sentences from a website. Ted consists of 2254 parallel sentences and Samantar consists of 40 million out of only 10k sentences has been taken due to memory constraint and third datasets from Kaggle repository consists of 5k English and Kannada Sentences respectively.

4.1. Training Parameters

The below Table 5 shows the different parameters used to implement the NMT with GRU and hybrid attention mechanism with different datasets.

Table 5: Parameters

Framework	Tensor flow
Neural Network Type	Gated Recurrent Unit (GRU)
The encoder-decoder's layer count	1
The encoder-decoder's hidden units per layer	1024
Batch size	16
Word embedding dimension	300
Epochs	25
Optimizer used	Adam
Loss function	Sparse categorical cross-entropy

4.2. Accuracy and Loss

The Table 6 indicates the accuracy and loss obtained when the model deployed with different datasets and graph of three datasets plotted by accuracy and loss. The three datasets are Samantar, Ted and FUS.

Table 6: Accuracy and Loss

Performance	**Samantar**	**Ted**	**FUS**
Accuracy	0.80	0.75	0.50
Loss	0.08	0.38	0.08

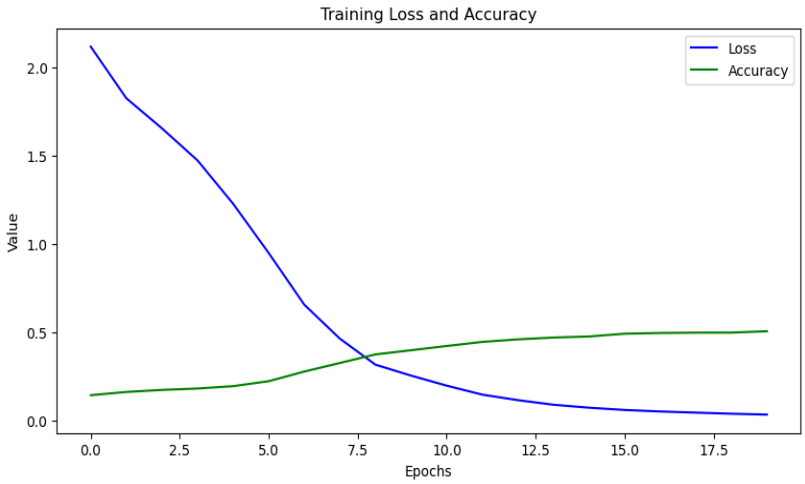

Figure 6: Graph of Accuracy v/s loss FUS

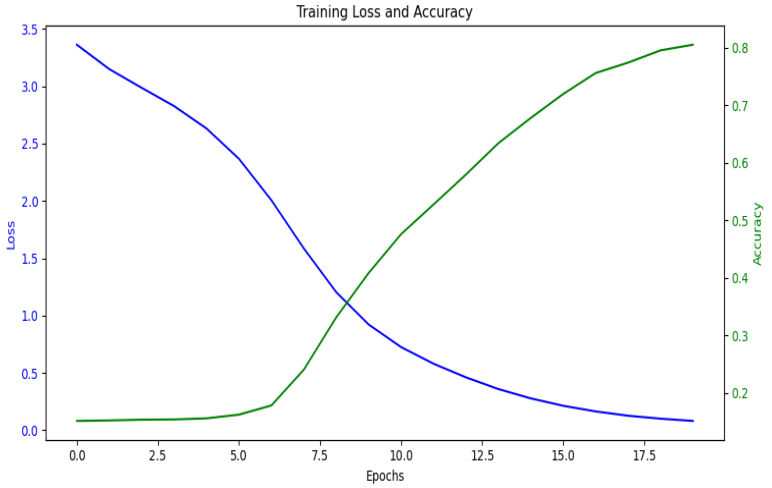

Figure 7: Graph Accuracy v/s loss Samantar

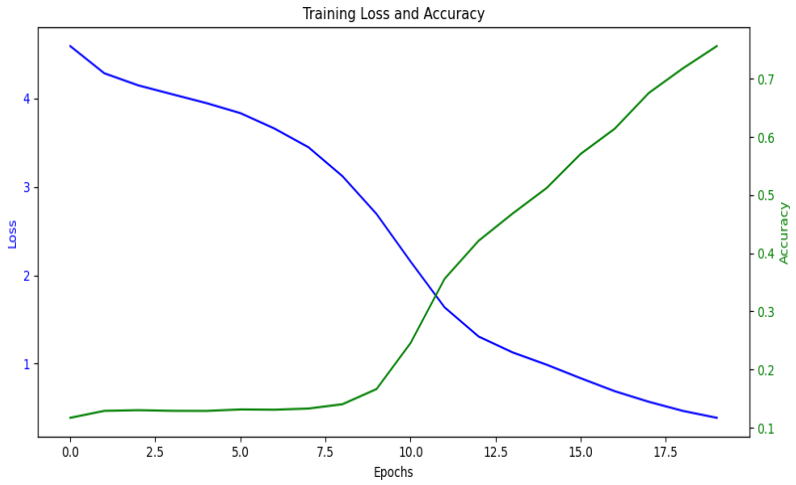

Figure 8: Graph Accuracy v/s Loss ted Dataset

The Figure 6 predicts that loss and accuracy will be consistent for 7 epochs, increasing accuracy 50% for 20 epochs and loss is 0.0330 for 20 epochs. The Figure 7 depicts that the loss and accuracy will be consistent for starting 8 epochs and accuracy is increasing after 20 epochs to around 80% and loss is decreased with 0.0861 for 20 epochs. The Figure 8 indicates that accuracy and loss is consistent for 20 epochs and accuracy is increasing with 72% after 25 epochs and loss is 0.0387 for 20 epochs.

4.3. ROUGE Score

One of the metrics which was used to evaluate machine translation was ROGUE. The below Table 7 shows the ROUGE-1, ROUGE-2 and ROUGE-N Score for the following three datasets.

Table 7: Rogue Score

Dataset	Size	ROGUE-1	ROGUE-2	ROGUE-L
FUS	5k	0.96	0.97	0.96
Samantar	10k	0.90	0.96	0.90
Ted	2.3k	0.56	0.55	0.64

The proposed model obtained on dataset of frequent Used Sentences (FUS) have good rogue score of ROUGE-1 of 0.96and ROUGE -2 of 0.97 and ROUGE-l 0.96. The Proposed model obtained on dataset of Samantar have achieved a good score like ROUGE-1 with 0.90, ROUGE-2 with 0.96 and ROUGE-L with 0.90.

4.4. BLEU Score and Graphs

The test set chosen from the various datasets was used to evaluate the trained model. The Table 8 displays the top BLEU scores on the test dataset. The Figure 9 to 12 shows the BLEU Scores of all datasets with bar graphs indicating that Samantar has a high BLEU Score. The results of BLEU Score was compared with existing baseline model (Singh et al., 2022).

Table 8: BLEU Scores

Dataset	BLEU Score
Frequent Sentences	72
Samantar	64
Ted	29

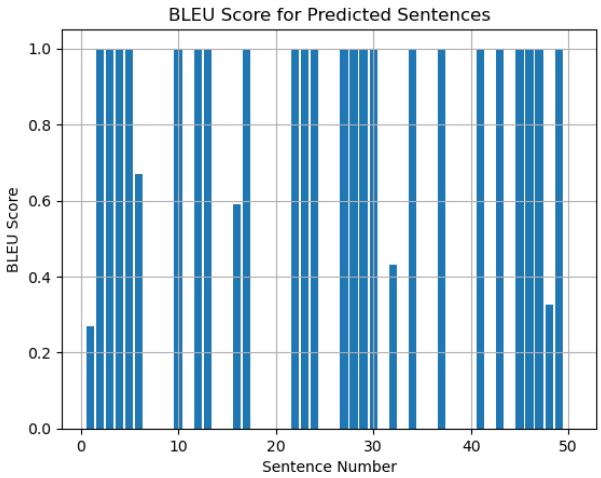

Figure 9: BLEU for FUS

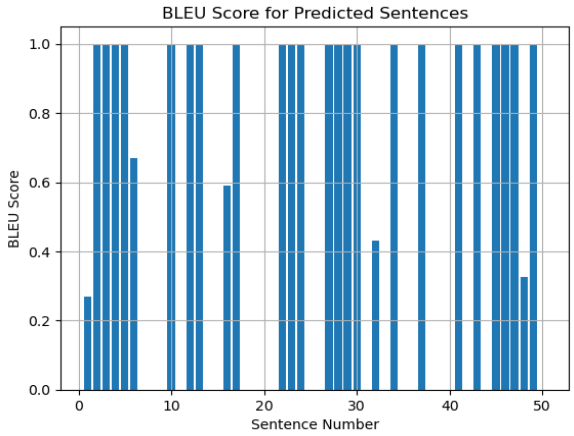

Figure 10: BLEU for Samantar

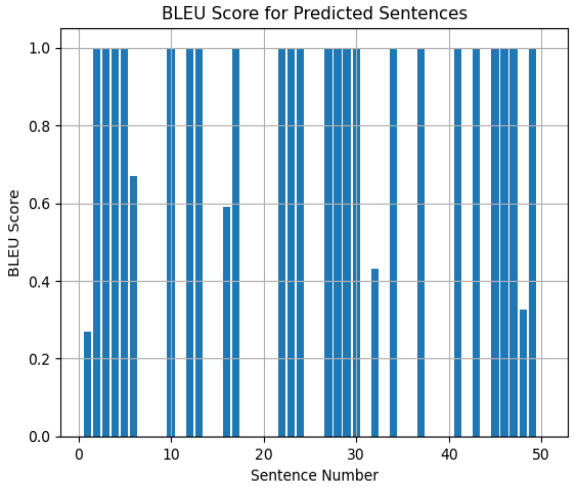

Figure 11: BLEU for Ted

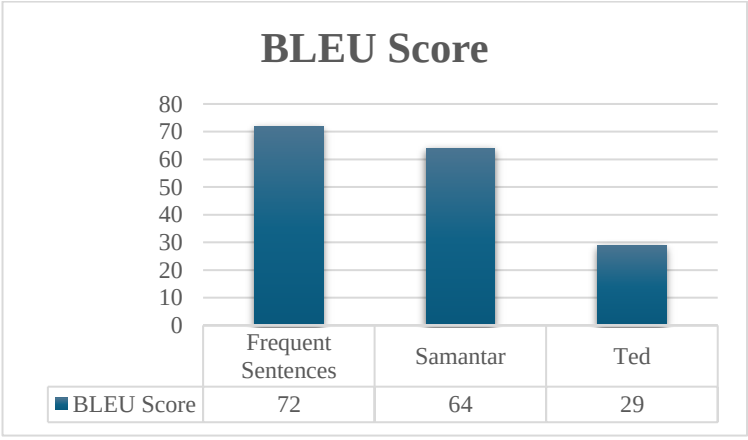

Figure 12: BLEU Scores for All Datasets

Table 8 shows the BLEU achieved for the FUS, Samantar and Ted dataset. Fus dataset has achieved a BLEU Score 72 of for the test dataset and Samantar has got a BLEU 64 Score of for the test dataset and ted has achieved a less BLEU Score of 29 due to the Sentence length in ted being more than 20.

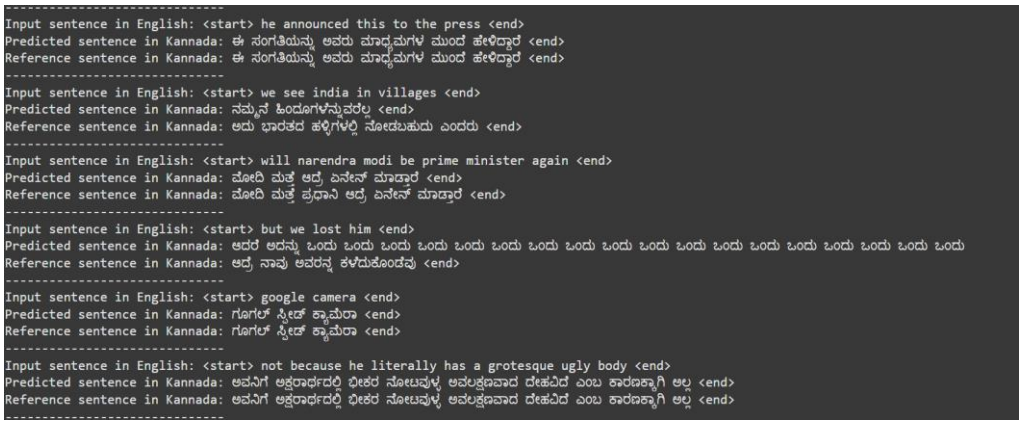

Figure 13: Snapshots of Translation of Frequent Used Sentences

The Figure 13 shows some sample translation of sentences of FUS dataset, including input sentences in English and predicted and references sentences in Kannada. Figure 13 out of 4 sentences 3 are predicted correctly and fourth sentences small mismatch was found.

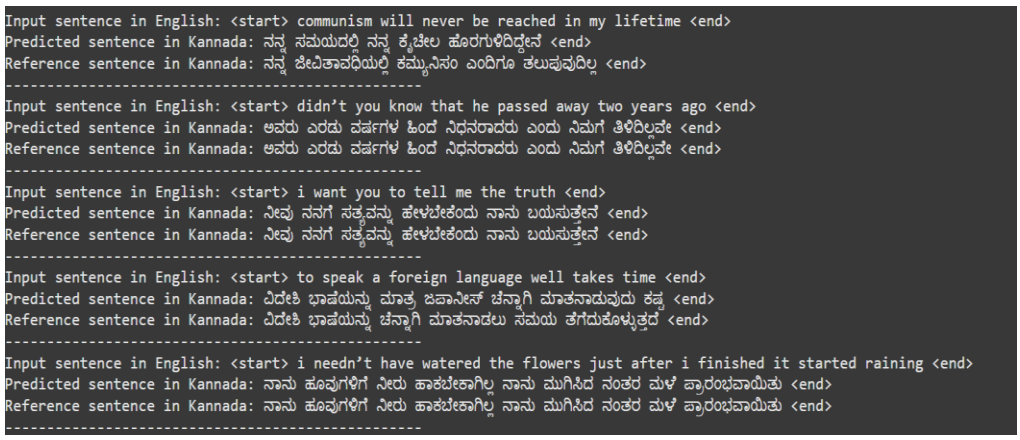

Figure 14: Snapshots of Translation for Samantar Dataset

The Figure 14 shows some sample translation of sentences for Samantar dataset. including input sentences in English and predicted and references sentences in Kannada. In the Figure 14 out of 4 sentences 3 are predicted correctly and fourth sentences slight mismatch was found.

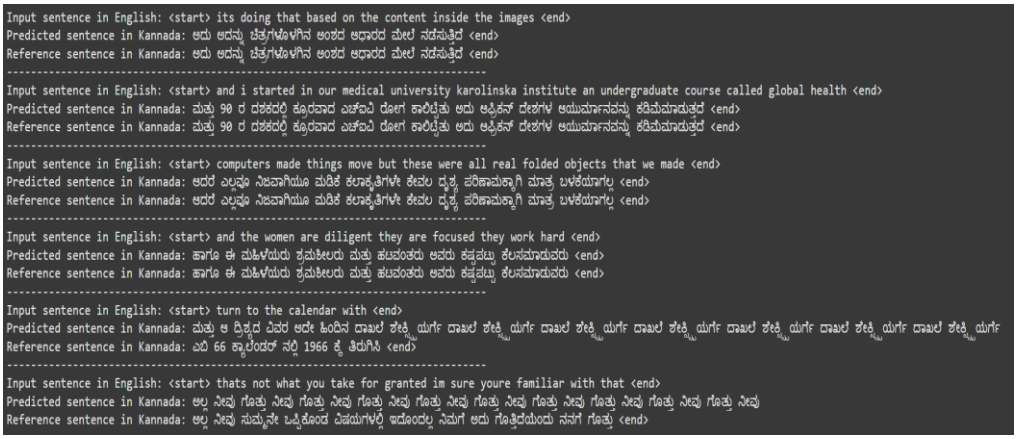

Figure 15: Snapshots of Translation for Ted Dataset

The model didn't achieve a good translation for ted dataset since the length of the sentences was very long and attention model was not predicting the good translation. In the Figure 15 out of 4 sentences 1 are predicted correctly and three sentences mismatch was found.

4.5. Comparison of Different Models

The proposed model was compared with state-of-the-art like LSTM, BI-LTSM with attention. ROUGE-1 is calculated for unigram, ROUGE-2 bi-gram and ROUGE-L for n grams it will be range 0 -1 and BLEU Score will range from 0 to 100.

Table 9: Frequently Used Sentences

Model	Size	BLEU	ROUGE-1	ROUGE-2	ROUGE-L
LSTM Attention	5K	30	0.28	0.13	0.28
GRU with Attention	5K	62	0.86	0.82	0.83
Bi-LTSM with Attention	5K	60	0.78	0.70	0.78
GRUwith hybrid Attention	**5k**	**72**	**0.96**	**0.97**	**0.96**

The Table 9 shows the NMT Model names with BLEU, ROUGE Scores for the FUS dataset with proposed model. Our Proposed model outperforms a good BLEU Score with 72 and other models with Attention have a good BLUE Score.

Table 10: Samantar Dataset

Model	Size	BLEU	ROUGE-1	ROUGE-2	ROUGE-L
LSTM without Attention	10k	12	0.10	0.01	0.10
GRU with Attention	10k	60	0.76	0.78	0.82
Bi-LTSM with Attention	10k	57	0.77	0.70	0.77
GRU with hybrid attention	**10k**	**64**	**0.90**	**0.96**	**0.90**

The Table 10 shows the NMT Model names with BLEUROUGE Scores for the Samantar dataset with proposed model. Our Proposed model outperforms a good BLEU Score with 64 and other models with Attention have a good BLUE Score.

Table 11: Ted Dataset

Model	Size	BLEU	ROUGE-1	ROUGE-2	ROUGE-L
LSTM with Attention	2.3K	13	0.05	0.006	0.005
GRU with Attention	2.3K	16	0.68	0.65	0.69
Bi-LTSM with Attention	2.3K	13	0.21	0.10	0.21
GRU with hybrid Attention	**2.3k**	**29**	**0.56**	**0.55**	**0.64**

The Table 11 shows the NMT Model names with BLEU, ROUGE Scores for the ted dataset with proposed model. Our Proposed model outperforms BLEU Score with 23 and other models with Attention have a least BLUE Score. Models without attention have significantly less score since without attention it will not perform good for lengthy sentences without Attention has significantly less score of 4.

5 Conclusion & Future Work

In this paper we propose a Neural Machine Translation Model with Pretrained Embeddings using GRU and hybrid Attention Mechanism from English to Kannada. Pre-trained embeddings reduce the effects of unusual or unseen words in the test data, improving accuracy. We have used three datasets like FUS,

Samantar and ted which is of different sizes. The Model consists of a hybrid attention network, a decoder network, and an encoder network. his proposed neural machine translation model has increase of +10% in BLEU Score and Rouge Score of 0.90. We have compared with LSTM, BILSTM with Attention and the proposed model has outperformed. Since the size of the dataset was small BLEU Score obtained was high on the model. To obtain more accurate results, study can improve by employing large corpora and the latest neural-based models, such as transformers, transfer learning, back translation.

References

[1] Al-Barhamtoshy, H. M., & Metwalli, A. S. Q. (2022, October). Neural Networks for Bilingual Machine Translation Model. In *2022 20th International Conference on Language Engineering (ESOLEC)* (Vol. 20, pp. 84-91). IEEE. https://doi.org/10.1109/ESOLEC54569.2022.10009266

[2] Ali, R., & Raj, M. (2024). Evaluating the Efficacy of Translanguaging Approach for Language Learning through K-Means Clustering Analysis. *Indian Journal of Information Sources and Services, 14*(3), 232–240. https://doi.org/10.51983/ijiss-2024.14.3.30

[3] Bakarola, V., & Nasriwala, J. (2021, September). Attention based neural machine translation with sequence to sequence learning on low resourced indic languages. In *2021 2nd International Conference on Advances in Computing, Communication, Embedded and Secure Systems (ACCESS)* (pp. 178-182). IEEE. https://doi.org/10.1109/ACCESS51619.2021.9563317

[4] Bala Das, S., Biradar, A., Kumar Mishra, T., & Kr. Patra, B. (2023). Improving multilingual neural machine translation system for indic languages. *ACM Transactions on Asian and Low-Resource Language Information Processing*, *22*(6), 1-24. https://doi.org/10.1145/3587932

[5] Balaji, R., Logesh, V., Thinakaran, P., & Menaka, S. R. (2022). E-Learning Platform. *International Academic Journal of Innovative Research*, *9*(2), 11–17. https://doi.org/10.9756/IAJIR/V9I2/IAJIR0911

[6] Chen, Q. (2023). A smaller and better word embedding for neural machine translation. *IEEE Access*, *11*, 40770-40778. https://doi.org/10.1109/ACCESS.2023.3270171

[7] https://ai4bharat.iitm.ac.in

[8] https://opus.nlpl.eu/TED2020/en&kn/v1/TED2020

[9] https://www.kaggle.com/datasets/spandangunti/eng-kan

[10] Ismail, W. S. (2024). Human-Centric AI: Enhancing User Experience through Natural Language Interfaces. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications, 15*(1), 172-183. https://doi.org/10.58346/JOWUA.2024.I1.012

[11] 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 & Continua*, *75*(2). http://dx.doi.org/10.32604/cmc.2023.035814

[12] Jahan, Z., Lateef, K. F., & Paul, J. (2021, August). Bangla-German Language Translation Using GRU Neural Networks. In *2021 International Conference on Emerging Techniques in Computational Intelligence (ICETCI)* (pp. 147-152). IEEE. https://doi.org/10.1109/ICETCI51973.2021.9574076

[13] Kandimalla, A., Lohar, P., Maji, S. K., & Way, A. (2022). Improving English-to-Indian language neural machine translation systems. *Information*, *13*(5), 245. https://doi.org/10.3390/info13050245

[14] Kumhar, S. H., Ansarullah, S. I., Gardezi, A. A., Ahmad, S., Sayed, A. E., & Shafiq, M. (2023). Translation of English Language into Urdu Language Using LSTM Model. *Computers, Materials, And Continuum (In English)*, *74*(2), 3899-3912. http://dx.doi.org/10.32604/cmc.2023.032290

[15] Laskar, S. R., Paul, B., Dadure, P., Manna, R., Pakray, P., & Bandyopadhyay, S. (2023). English–Assamese neural machine translation using prior alignment and pre-trained language model. *Computer Speech & Language*, *82*, 101524. https://doi.org/10.1016/j.csl.2023.101524

[16] Mahmud, A., Al Barat, M. M., & Kamruzzaman, S. (2021, December). GRU-based encoder-decoder attention model for English to Bangla translation on novel dataset. In *2021 5th International Conference on Electrical Information and Communication Technology (EICT)* (pp. 1-6). IEEE. https://doi.org/10.1109/EICT54103.2021.9733595

[17] Nath, B., Sarkar, S., Das, S., & Mukhopadhyay, S. (2024). Neural machine translation for Indian language pair using hybrid attention mechanism. *Innovations in Systems and Software Engineering*, *20*(2), 175-183. https://doi.org/10.1007/s11334-021-00429-z

[18] Pathak, A., & Pakray, P. (2019). Neural machine translation for indian languages. *Journal of Intelligent Systems*, *28*(3), 465-477. https://doi.org/10.1515/jisys-2018-0065

[19] Paul, N., Faruki, I., Pranto, M. I., Shawon, M. T. R., & Mandal, N. C. (2023, February). Bengali-English Neural Machine Translation Using Deep Learning Techniques. In *2023 International Conference on Electrical, Computer and Communication Engineering (ECCE)* (pp. 1-6). IEEE. https://doi.org/10.1109/ECCE57851.2023.10101491

[20] Prasanna, A. S. D., & Latha, C. B. C. (2023). Bi-Lingual Machine Translation Approach using Long Short–Term Memory Model for Asian Languages. *Indian Journal of Science and Technology*, *16*(18), 1357-1364. https://doi.org/10.17485/IJST/v16i18.176

[21] Saleem, M. W., & Uprety, S. (2021). Neural Machine Translation with Attention. https://doi.org/10.13140/RG.2.2.29381.37607/1.

[22] Sarode, S., Thatte, R., Toshniwal, K., Warade, J., Bidwe, R. V., & Zope, B. (2023, March). A System for Language Translation using Sequence-to-sequence Learning based Encoder. In *2023 International Conference on Emerging Smart Computing and Informatics (ESCI)* (pp. 1-5). IEEE. https://doi.org/10.1109/ESCI56872.2023.10099876

[23] Sethi, N., Dev, A., & Bansal, P. (2023). A Novel Neural Machine Translation Approach for low-resource Sanskrit-Hindi Language pair. *ACM Transactions on Asian and Low-Resource Language Information Processing*. https://doi.org/10.1145/3591207

[24] Shah, P., & Bakrola, V. (2019, February). Neural machine translation system of indic languages-an attention based approach. In *2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP)* (pp. 1-5). IEEE. https://doi.org/10.1109/ICACCP.2019.8882969

[25] Sheshadri, S. K., Gupta, D., & Costa-Jussá, M. R. (2022, December). Neural machine translation for Kashmiri to English and Hindi using pre-trained embeddings. In *2022 OITS International Conference on Information Technology (OCIT)* (pp. 238-243). IEEE. https://doi.org/10.1109/OCIT56763.2022.00053

[26] Singh, J., Sharma, S., & Briskilal, J. (2022, May). Natural language processing based machine translation for hindi-english using gru and attention. In *2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC)* (pp. 965-969). IEEE. https://doi.org/10.1109/ICAAIC53929.2022.9793214

[27] Tiwari, G., Sharma, A., Sahotra, A., & Kapoor, R. (2020, July). English-Hindi neural machine translation-LSTM seq2seq and ConvS2S. In *2020 International Conference on Communication and Signal Processing (ICCSP)* (pp. 871-875). IEEE. https://doi.org/10.1109/ICCSP48568.2020.9182117

[28] Tohma, K., & Kutlu, Y. (2020). Challenges Encountered in Turkish Natural Language Processing Studies. *Natural and Engineering Sciences, 5*(3), 204-211. https://doi.org/10.28978/nesciences.833188

[29] Unanue, I. J., Borzeshi, E. Z., & Piccardi, M. (2022). Regressing word and sentence embeddings for low-resource neural machine translation. *IEEE Transactions on Artificial Intelligence*, *4*(3), 450-463. https://doi.org/10.1109/TAI.2022.3187680

Authors Biography

Gunti Spandan, is working as an Assistant Professor in the CSE department at GITAM University Bengaluru campus. He has nine years of experience. He completed her post-graduation at Jain University Bangalore and is pursuing a PhD at Gitam University Bengaluru campus. His research is Natural Language Processing with Machine Translation using Neural Networks.

Dr. Prasannavenkatesan Theerthagiri, was awarded PhD (Full-Time) from Anna University Chennai. He received Bilateral Mobility grant award from Slovenia. Published his research works in 17+ SCI(E) indexed journals, 41+ SCOPUS indexed journals, 10+ national/international conferences, 10+ book chapters in Elsevier, Springer, Wiley, T&F, Emerald, etc. publishers. He is a Reviewer for 10+ SCI Indexed journals. His research interests are Smart Agriculture with AIML & IoT, Healthcare analysis using ML.