Efficient Use of Machine Learning Models to Evaluate the Parametric Performance of the ML Models for Language Translation from Telugu to Hindi

A. Surya Kausthub, Yerukola Gayatri, Shail Garg, Peeta Basa Pati and Tania Ganguly
Department of Computer Science and Engineering, Amrita School of Computing, Bengaluru, Amrita Vishwa Vidyapeetham, India

Keywords: Natural Language Processing, Deep Learning Models, LSTM, Fairseq, Parametric Tuning.

Abstract: Translation between Telugu and Hindi, two widely spoken languages in India, presents numerous challenges due to significant linguistic, syntactic, and cultural differences. This study focuses on leveraging advanced deep learning models to address these discrepancies and evaluate their performance in translating Telugu to Hindi effectively. The research considers models such as Long Short-Term Memory Networks (LSTM) and Fairseq emphasizing their parametric performance by fine-tuning under various settings. The core objective is to systematically assess these models, uncovering how they respond to parameter optimization and identifying the best methodologies for generating high-quality translations. By analyzing the results, this study aims to pave the way for the development of robust and efficient translation systems tailored to low-resource languages like Telugu. Such systems hold the potential to bridge linguistic gaps and foster more accessible communication across diverse Indian languages, contributing to broader cultural and digital inclusion.From the two models studied Fairseq is a better model with higher accuracy.

1 INTRODUCTION

Language translation can be deemed as one of the primary problems of Natural Language Processing as it mediates between various linguistic areas. Interpreting spoken or written materials from one language to another is indispensable for enabling cross-cultural and business relations and/or social interaction among persons of different cultures.

Interpreting between two languages, which are both pho- netically, syntactically, and semantically distant, such as, for example, Telugu – a Dravidian language spoken in southern India, or Hindi used mainly in the northern and central part of the country. Such differences show that direct translation from one word to the other is not enough, as there seems more to it in translating from these two languages to accord the meaning in the target language. Today, one of the main issues of translating from Telugu to Hindi is the absence of the necessary large parallel corpora that are needed for training deep learning models. Telugu as a low intelligibility language has restricted textual Telugu corpora for computational applications. This lack of data greatly enhances the challenge of the task, as prior translation methods rely on parallel datasets to deliver the best performance.

In recent years deep learning frameworks have become influential in developing mechanisms for machine translation which can further deal with languages no matter the availability of resources of any language. Thus, this research centers on two of the most widely recognized models in the field, namely LSTM, and Fairseq, to evaluate their usefulness in translating low-resource languages such as Telugu to Hindi. These models belong to a variety of architectures of translation, starting from linear models and ending with open-source architectures for the effective processing of natural language.

RNNs, particularly LSTM, are ideal for sequential data in which the order and the dependence of words are important; therefore are good for short–to medium–string translations. Another open-source toolkit is Fairseq which are highly flexible model and supports a range of difficult-to-implement tasks such as sequence-to-sequence modeling; this makes Fairseq a worthy competitor for low-resource language translation.

Kausthub, A. S., Gayatri, Y., Garg, S., Basa Pati, P. and Ganguly, T.
Efficient Use of Machine Learning Models to Evaluate the Parametric Performance of the ML Models for Language Translation from Telugu to Hindi.
DOI: 10.5220/0013642800004664

In *Proceedings of the 3rd International Conference on Futuristic Technology (INCOFT 2025) - Volume 3*, pages 712-720
ISBN: 978-989-758-763-4

However, every one of these models has the potential, although they have their own advantages of some peculiarities. Their applicability in Telugu to Hindi translation is influenced by factors such as the existence of an optimum parameter, linguistic features, and their control and computational limits, etc. Due to the low resource nature of the Telugu language, some difficulties arise because of a scarcity of large parallel corpus which is very important for developing highperformance models.

To determine the efficiency of the methods proposed by LSTM and the Fairseq tool under different circumstances, this investigation seeks to pinpoint potential ways of enhancing translation quality. Thus, the examination not only adds to the improvement in Telugu to Hindi translation but also opens the opportunity for overcoming difficulties that other low-resource languages face. This prolific work seeks to reduce language breakdowns between individuals as well as improve levels of so-called accessibility on the web.

2 LITERATURE SURVEY

Xu et al. (Xu, Xie, et al. , 2023) Did a Critical Review and Assessment While preserving performance, parameter-efficient fine-tuning (PEFT) techniques use less memory and fewer fine-tuning parameters while maintaining the same level of performance as conventional fine-tuning methods. Hence, they are applied to cross-lingual transfer, backdoor attack, and multi-task learning. In order to attain better outcomes they either introduce new parameters into the methodology or include different aspects of PEFT. Subsequent research will focus on PEFT strategies applied to multi-modal learning and computer vision and enhance PEFT’s performance and interpretability.

Mohamed et al. (Mohamed, Khanan, et al. , 2024) Worked on machine translation developments, under the TLDR umbrella, which focused on the improvements of machine translation, including neural machine translation (NMT). It offers details on the facets where utilization of deep learning and artificial neural networks can be utilized to have a higher quality, efficiency, and accuracy in translation. The study also calls for more studies for better translation quality and values of translating culture with diverse methods; moreover, evaluating the effectiveness of an MT system by both algorithm and human slicer.

Cayamcela et al. (Cayamcela and Lim, 2019) It focuses on two areas related to NMT and current discussion surrounding diversity and representation issues that include topics such as cultural sensitivity in an attempt to understand how computational intelligence is influencing the field of language translation. Here there is Semantic fuzziness and language variability handling, as well as feature extraction, intelligent recognition, and maximum entropy. Artificial intelligence is revolutionalizing the translation processes.

Zhang et al. (Zhang, 2021) Illustrates how back-translation and cross lingual embeddings together with creativity improve results of translation. It underscores the importance of fixing problems with neural machine translation (NMT) training methods for far better and accurate translations with the illustrated improvements over conventional unsupervised models. Another strength in the strategy of the study is that, since assessments focus on developing ways of measuring translation quality through automated and human approach, the study also highlights the use of machine translation as an efficient solution and a relief to the burden oftentimes placed on translators.

Mantoro et al. (Mantoro, Asian et al. , 2016) Improved a statistical machine translator by applying sequence IRSTLM translation parameters and pruning. It discusses the challenges of translating and presents a process it says one can use to get translations that are accurate without necessarily having to master the language being used. The importance of interface, customization, and pruning is stressed in the context of machine translation and factors concerning IRSTLM language modeling are compared. The proposed approach eclipses conventional strategies that require language proficiency, and I found effectiveness in the proposed strategy generating promising profiles.

Sun et al. (Sun, Hou, et al. , 2023) developed a novel way of enhancing translation for the languages which are not so popular and which contain minimal data. To enhance preciseness, especially when scant bilingual data are available, it refers to CeMAT, an extremely powerful pre-trained model. This brings out one of the major issues of how to prevent the model from making similar errors is mentioned. They address this through proposing an approach that localizes the development of the model from the mistakes that it makes. Further they provide an intelligent training plan that changes with the data and the model confidence especially useful for low resource languages. The experiments they demonstrate indicate that these approaches translate significantly better, arguing for the value of pretraining in combination with this ground-breaking learning technique for low-resource languages.

Thillainathan et al. (Thillainathan, Ranathunga, et al. , 2021) examines enhancing Accurate Translation of Low Resource Languages employing mBART or other related NMT pre-trained models. The study introduces translation from and to Sinhala and Tamil and shows by fine-tuning mBART with little parallel data (e.g., 66,000 sentences), we can achieve substantial BLEU gains over a comparable transformer-based NMT model. According to the findings, the quality of translation is significant on the amount of monolingual corpus for the target language and the linguistic density of the language in question. This research proves that the power of multilingual models can be effective in the extreme low-resource setting further implying that the research direction can proceed toward joint multilingual finetuning or using even more advanced models such as mT5.

Tran et al. (Tran, 2024) explore ways by which we can obtain good translations between low resource language pairs including Lao -Vietnamese. Based on a dataset of the VLSP 2023 MT challenge, the study investigates hyperparameters tuning, back translation, and fine-tuning of multilingual pre-trained models that include mT5 and mBART. From the experiments, it can be seen that hyperparameters tuning yields 22 more BLEU points than experiment without tuning, back translation increases scores to 27.79 and fine tuning mT5 got the highest score of 28.05. The results show that integrating optimization with the application of pre-trained models significantly improve the translations and future work on low-resource languages.

Hallac et al. (Hallac, Ay, et al. , 2018) further investigates pretraining and finetuning of deep learning models for the classification of tweet data using a large corpus of news articles labeled for the same topic and a small set of tweets. The authors employ models such as CNN, Bi-LSTM-CONV, and MLP first on news data and then fine-tuning them on tweets to categorise content into culture, economy, politics, sports, and technology. Altogether, the experimental evaluation indicate that the fine-tuned model that performs the best is the Bi-LSTM-CONV model with high extra accuracy beyond the models trained solely with tweets. The study implies that the classification of texts could be improved during pre-training on similar large datasets and activation of step-by-step fine-tuning in data-deficient environments.

Saji et al. (Saji, Chandran, Alharbi, 2022) discusses an architecture of English-to-Malayalam machine translation exploiting transformers while emphasizing translation quality enhancement to low-resource languages such as Malayalam. It compares multiple architectures of NMT: Seq2Seq models with Bahdanau, multi-head and scaled dot product attention mechanisms, and MarianMT. Adjustment of the MarianMT model considerably improves performance, and the solutions obtained have the highest BLEU and E-values with subjective estimations. The work also shows that attention mechanisms help in the enhancement of translation quality and indicates how these models can be used in low-resource languages.

Premjith et al. (Premjith, Kumar, et al. , 2019) The study introduces a Neural Machine Translation (NMT) system that uses parallel corpora to translate English into four Indian languages: Tamil, Punjabi, Hindi, and Malayalam. It draws attention to issues like the dearth of high-quality datasets and the morphological diversity of Indian languages, and it suggests solutions including transliteration modules to handle terms that are not in the vocabulary and attention mechanisms for processing lengthy phrases.

Nair et al. (Nair, Krishnan, et al. , 2016) In order to handle grammatical subtleties like declensions and sentence reordering, the study suggests a hybrid strategy for an English-to-Hindi machine translation system that combines rule-based and statistical techniques. Its potential for more extensive multilingual applications is shown by its better accuracy as compared to current systems.

Unnikrishnan et al. In order to overcome linguistic disparities, the study presents a Statistical Machine Translation (SMT) system for English to South Dravidian languages (Malayalam and Kannada). It incorporates morphological information, syntax reordering, and optimized bilingual corpus construction. It offers a framework that may be modified to accommodate other Dravidian languages and exhibits increased translation accuracy and a smaller corpus size.

KM et al (KM, Namitha, et al. , 2015) In this paper, two different corpora—a general text corpus and a Bible text corpus—are used to compare English-to-Kannada statistical machine translation (SMT). The difficulties presented by Kannada's morphological diversity are emphasized, and methods for boosting translation quality are covered, with a focus on how corpus size and token frequency might raise the baseline SMT systems' BLEU score.

The next section explains the methodology of our proposed fine-tuned models.

3 METHODOLOGY

3.1 Dataset collection

The dataset consists of 21,404 bilingual words and phrases, Hindi and Telugu. All these books have been collected from different linguistic resources particularly books which are used in teaching Telugu to Hindi speaking persons. These educational materials provide carefully structured and context based examples and thus ensure high accuracy in translation from one language to the other. Furthermore, parallel corpora, current affairs articles, and other free bilingual datasets used in the current study's dataset. To collect this data, it requires manual extraction, expert translation and most important, automated alignment to ensure quality and consistency is achieved.

3.2 Dataset Preprocessing :

The dataset underwent thorough preprocessing to guarantee cleanliness and consistency for subsequent analysis or modeling. First of all, rows with missing values at the Hindi or Telugu columns as well as rows with an empty string were removed. All entries were converted to lower case for uniformity to remove variability and any single or double quotes were erased. The punctuation marks and numbers were excluded to remain focused only with the textual data, the strings of text; also leading and trailing spaces and multiple successive spaces within the strings were removed. The mean of both Hindi and Telugu sentences are segregated as strings and then, certain operations were performed to remove all the usual white spaces. Moreover, the additions of start and end tokens to Telugu translations ensured the dataset's relevance to sequence-to-sequence task as in machine translation. These preprocessing procedures provided a normalized, clean and immediately usable data which can be fed to a linguistic programme or NLP application.

3.3 Design:

Fig.1 provides a clear view on how to fine-tune already developed machine translation models FairSeq and LSTM for synthesizing Hindi language translation to Telugu. The overall goal is then achieved by collecting the datasets and preprocessing them, as well as, training the models. For now, hyperparameters affecting FairSeq include dropout rate, learning rate, and the number of embedding layers, whereby results are checked using a validation loss plot. In LSTM, fine-tuning simply implies changing the dataset size and number of epochs by comparing and contrasting validation loss and accuracy graphs. The best-performing configurations from the two models are then determined from these evaluations to arrive at the best model to be implemented.

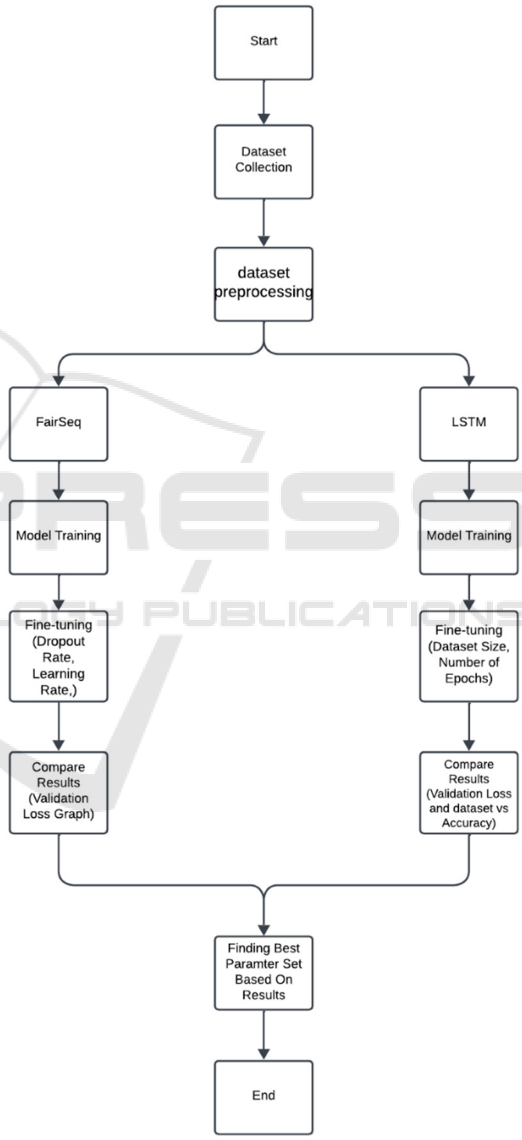

Figure 1: Flow chart of the proposed Framework

3.4 Models

Encoder Decoder LSTM: Long Short-Term Memory (LSTM) networks are a specific implementation of Recurrent Neural Network (RNN) that is meant to address the problems of the standard RNN, first of which is the problems with handling long-term dependencies. Memory cell is used in LSTMs and it is accompanied by three gates including input, the forget and the output gates. Based on these gates, what information should be stored, which information discarded or which information should be utilized in order to influence the output; this makes LSTMs identify patterns across sequences. This capability is necessary to employ in turn-based operations such as language modeling, in which interpretation of a particular word depends on the definition of the complete sentence or paragraph. Because of their ability to learn long- term dependencies this make LSTMs to be widely used in applications area including speech recognition, text generation, and machine translation.

Translating from Telugu to Hindi entails understanding and mapping the semantic structures of two languages that often differ in grammatical norms and word arrangement.Unlike Hindi which belongs to the Indo Aryan branch of the Indo European family of language the Telugu language which belongs to the Dravidian linguistic family poses quite dif- ferent syntactic and morphological translation issues. LSTMs especially can help extract the contextual meaning of input sequences (Telugu sentences) and through the encoder-decoder technique in which the encoder part converts the entered sequences into a fixed-size vector called the context vector. This vector is used as something like understanding of the original sentence with respect to basic semantics that is then translated into a grammatically correct Hindi equivalent. Still, since one translates one language to many languages or vice versa the kind of mapping that is enjoyed by conventional machine learning models is not good here this is why LSTMs work really well with these types of mapping. For example, one Telugu word could be several Hindi words that are quite beyond the ability of LSTMs to handle. The efficiency with which they learn to preserve the words' dependencies across long sequences helps keep the translations truly capturing language and context.

By systematically adjusting the number of epochs and the size of the dataset used for the LSTM model, we see a notable effect on the model's accuracy in translating between Hindi and Telugu. A structured approach was adopted. The dataset, consisting of 21,404 bilingual sentence pairs, was divided into six subsets, representing 10%, 20%, 40%, 60%, 80%, and 100% of the data. Each subset was used to train the model independently, ensuring consistent settings for parameters such as batch size, optimizer, and learning rate. This step-by-step increase in dataset size allowed us to systematically examine how the amount of training data influences the model's ability to learn and generalize.The LSTM model, built with an encoder-decoder architecture, was trained to convert Telugu sentences into semantic context vectors and decode them into their corresponding Hindi translations. To prepare the data, start and end tokens were added to Telugu sentences, and padding was applied to standardize sequence lengths. The training was carried out using a TPU((Tensor Processing Unit), which provided the computational efficiency needed to handle the varying dataset sizes effectively.

To explore the impact of training duration, the number of epochs was varied for each dataset size. Increasing epochs gave the model more opportunities to refine its understanding of linguistic patterns, capturing the complex relationships between Hindi and Telugu. At the same time, training and validation loss were monitored to observe trends in convergence and generalization. By using a consistent validation set across all experiments, we ensured a fair comparison of the model's performance across different configurations.This provides a systematic way to evaluate the role of dataset size and training duration in improving translation accuracy. By fine-tuning these parameters, we aimed to identify the best practices for building effective machine translation models, particularly for low-resource language pairs like Hindi and Telugu.

FAIRSEQ: FairSeq is a sequence-to-sequence transformer-based model designed by AI researchers from Facebook and is used for applications like machine translation, text summarization, and language modeling. Among them, recurrent neural networks (RNN), LSTM, and transformers are supported. This program is very flexible and fast and has some cool features such as distributed training, mixed precision optimization, and a pre-trained model – all of which make it perfect for fine-tuning on large datasets. In this way, it offers an opportunity to tune the hyperparameters that are necessary to obtain the models providing a high degree of translation.

Table 1

Hyperparameter	Values Tested	Description
Dropout Rate	0.1, 0.3	Helps prevent overfitting by setting random input units to zero during training.
Learning Rate	0.001, 0.005	Controls the size of steps taken during gradient descent optimization.
Embedding Layers	2, 4	Adjusts the number of layers and neurons to capture semantic relationships.

With the relatively complex syntax and semantics of Hindi and Telugu languages, the structural and functional architec- ture of the FairSeq model's architecture makes it specifically capable of translating between these two languages. Hindi and Telugu are two different types of language groups, Indo- Aryan, and Dravidian respectively; and its translation from one particular language to the other is more complex. Thanks to the possibilities of changing model parameters, FairSeq is ready to address such linguistic diversity by using appropriate values of dropout rates, learning rates embedding layers, etc. Moreover, its capability for fine-tuning pre-trained models helps to con- verge and achieve better results on the adopted dataset while offering an optimal solution to this type of translation. The study makes optimization of the selected model FairSeq for Hindi to Telugu translation, with variations of hyperparameters such as dropout rate, learning rate, and embedding layers as shown in Table I, and assesses its performance through the validation loss graph. The objective here is to find out which of these hyperparameters provides the lowest validation loss and the highest quality of the translation done.

The model is trained for 10 epochs on each of the 8 combinations of those hyperparameters. As for each training run the validation loss is measured after each epoch is com- pleted. It also enables us to determine how the model performs when applied to new data and consider how each setting of hyperparameters performs.

A graph of the validation loss is constructed for each of the eight scenarios to analyze the model's performance. The configuration that achieves the least validation loss is chosen as the best solution. This approach aids in determining the optimal hyperparameter settings for Hindi to Telugu translation using FairSeq; this is while optimizing the strengths of the language pair translated.

To enhance the translation quality and to reduce overfitting, here the study tried different strategies including dropout rates, learning rates, and different embedding layers, and selected the model configuration where the model gave the minimum validation loss. This fine-tuning is critical to fine-tuning the model for the specific task of translating Hindi to Telugu for which it has not been specifically designed.

The next section includes results which were obtained after training our fine tuned models

4 RESULTS

4.1 Exploration and Cleaning of the Dataset

For the experiment, the data comprised a Hindi-to-Telugu translation dataset containing 21,403 sentence pairs. During the initial data inspection, a single missing value was identified in the Hindi column, which was promptly removed to ensure clean data integrity. The preprocessing pipeline was meticulously designed to prepare the data for effective model training.

Vocabulary Creation:

- **Hindi Vocabulary:** A total of 16,068 unique Hindi words were identified, representing the language's full spectrum in terms of richness.
- **Telugu Vocabulary:** To capture the detailed syntactical structures of Telugu, a larger vocabulary comprising 32,316 words was established.

After preprocessing, the dataset was divided into three distinct sets to facilitate balanced training and evaluation:

- **Training Set:** 12,841 samples
- **Validation Set:** 4,281 samples
- **Testing Set:** 4,281 samples

This partitioning ensured that the models were trained on a substantial number of samples while retaining adequate data for reliable validation and testing.

4.2 Encoder-Decoder LSTM Model Performance

In the Encoder-Decoder LSTM model, three sub-models were employed and analyzed: the Autoencoder model, Decoder model, and Encoder model. The Encoder-Decoder LSTM is designed for sequence-to-sequence translation of Hindi to Telugu sentences. The architecture utilized two distinct embedding layers with 256 embedding dimensions for the encoder and decoder components, followed by single-layer LSTM networks. This setup was intended to incorporate the temporal dependencies present in language translation tasks.

Training Dynamics:

The LSTM model was trained up to 40 epochs with a batch size of 64. Throughout the training phase, both training and validation losses exhibited a consistent downward trend, indicating effective learning. However, despite the decreasing loss values, the accuracy metrics showed only marginal improvements, reflecting challenges in capturing the complexities of the translation task.

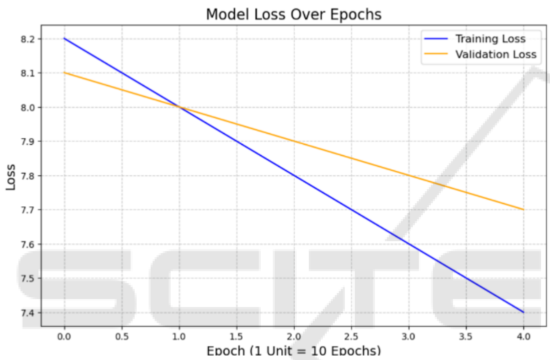

Figure 2: Training and Validation Loss Over Epochs for Encoder-Decoder LSTM

Fig.2 presents the training and validation loss curves over the epochs for the LSTM model. Both losses decreased steadily, showcasing the model's ability to minimize errors during training. By the end of training, the model achieved a training loss of approximately 7.26 and a validation loss of 7.63. The training accuracy reached 82.37%, while the validation accuracy was 81.36%.

Impact of Dataset Size on LSTM Performance: To evaluate the influence of dataset size on model performance, the LSTM was trained on subsets comprising 10%, 20%, 40%, 60%, 80%, and 100% of the total dataset. The accuracies for these subsets are summarized in Table 2.

Table 2: Dataset Size Vs. Accuracy For Encoder-Decoder LSTM

Dataset Size (%)	Accuracy (%)
10	35.06
20	50.06
40	65.41
60	72.13
80	78.26
100	85.41

Fig.3 illustrates the relationship between dataset size and accuracy for the LSTM model.

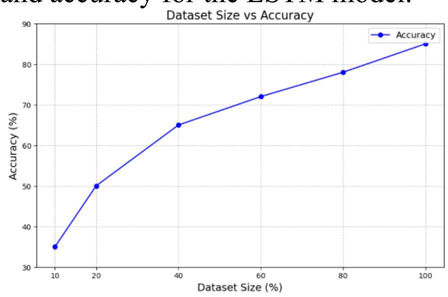

Figure 3: Dataset Size vs. Accuracy for Encoder-Decoder LSTM

The results show a slight improvement in accuracy as the dataset size increases from 10% to 40%. Beyond the 40% threshold, the gains in accuracy become marginal, suggesting that additional data provides limited benefits for the LSTM model's performance. This plateau indicates that the model's architectural constraints hinder its ability to leverage larger datasets effectively for capturing complex translation nuances.

4.3 Fairseq Transformer Model Performance

The Fairseq Transformer model was employed to leverage the advanced capabilities of Transformer architectures in handling complex translation tasks. The model underwent meticulous hyperparameter tuning, focusing on dropout rate, learning rate, and number of encoder layers to optimize its performance as detailed in Table I.

Training Dynamics: Compared with the LSTM, the Transformer model indicated better performance in learning. Across 10 epochs, all model configurations exhibited significant reductions in both training and validation losses. For instance, a configuration with a dropout rate of 0.1, learning rate of 0.0005, and encoder layers of 2 achieved a training loss of 6.126 and a validation loss of 7.267 by the end of the training phase.

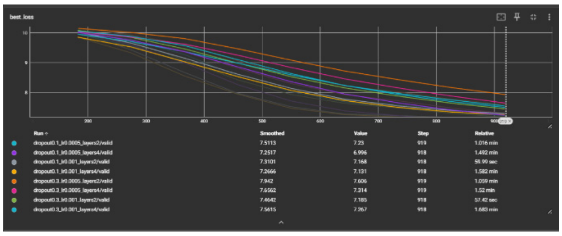

Figure 4: Training and Validation Loss Over Epochs for Fairseq Transformer

Fig. 4 showcases the training and validation loss curves for various hyperparameter configurations of the Transformer model. The rapid decrease in loss values across epochs indicates effective optimization and learning, surpassing the performance observed in the LSTM model. The Transformer's architecture, which incorporates multi-head attention and positional encoding, facilitates superior feature extraction and contextual understanding, contributing to its enhanced performance.

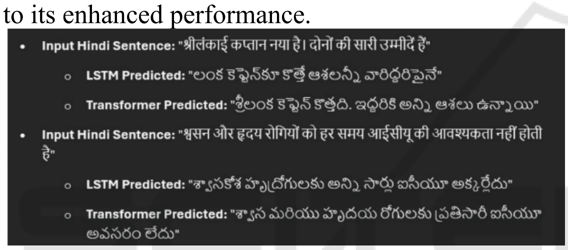

Figure 5: Example Translations Comparison

Fig.5 underscores the qualitative differences between the LSTM and Transformer models. The Transformer's translations are notably more fluent and semantically accurate, effectively capturing the essence and nuances of the source sentences. In contrast, the LSTM's outputs lack coherence and contextual relevance, highlighting the Transformer's superior translation capabilities. The Transformer's proficiency in maintaining grammatical correctness and contextual integrity demonstrates its advanced understanding of linguistic structures, making it a more reliable model for translation tasks.

4.4 Model Comparison

Comparing Encoder-Decoder LSTM and Fairseq Transformer, the latter performs better in every aspect. Concerning the performance measures, it was apparent that training and validation losses of Transformer were getting lower (6.126 to 6.486) as opposed to a higher training loss (7.26) and validation loss (7.63) of the LSTM, because the Transformer is capable to optimize and generate better representations by learning syntactic and semantic structures. The Transformer was trained with exceptional speed, and it reached the point of convergence in 10 epochs, while LSTM took about 40 epochs making little changes in accuracy. There is improved efficiency due to the Transformer model with the multi-head attention and deeper layers in the work, it makes learning faster. Moreover, quality criteria focused on the Transformer's efficiency in producing crime-legal and natural translations of text, thus the quality of translation was far from that of the Transformer and LSTM, the latter often resulted in less coherent and contextually inconsistent translations. These advantages make the Transformer a better and reliable model as compared to others for the translation tasks.

The Transformer's advanced architectural features enable it to learn more effectively from the dataset, resulting in lower training and validation losses within fewer epochs. Additionally, the qualitative evaluation of translation outputs demonstrated the Transformer's superior ability to generate coherent and contextually accurate translations, whereas the LSTM model's outputs were less reliable and fluent. These findings collectively highlight the Transformer's advantage in machine translation tasks, particularly in handling complex language pairs like Hindi and Telugu.

The next section includes conclusion and future scope of our research.

5 CONCLUSION

The comparison research demonstrates that the Fairseq Transformer model is the preferable choice for Hindi-to-Telugu translation jobs, as it achieves much lower training and validation losses, faster convergence, and semantically richer translations. The Transformer's sophisticated architecture, which relies on multi-head attention mechanisms and positional encoding, allows it to handle complicated linguistic patterns more successfully than the Encoder-Decoder LSTM. However, this work demonstrates that fine-tuning is critical to improving model accuracy for both approaches. Fine-tuning the dataset size and number of epochs dramatically enhanced the LSTM's performance, resulting in better generalization across different training sizes. In contrast, fine-tuning hyperparameters like as dropout rates, embedding sizes, and learning rates improved the Fairseq Transformer's optimization and translation quality. These findings emphasize the importance of hyperparameter optimization in realizing the full potential of machine translation models, paving the way for more resilient and effective systems designed for low-resource language pairs.

6 FUTURE WORK

Future research should focus on optimizing model hyperparameters such as attention heads and layers, or on hybrid architectures that combine LSTMs and Transformers to better capture linguistic nuances in low-resource languages such as Hindi and Telugu. Extended training with more epochs and enhanced evaluation measures, such as ROUGE or METEOR have the potential to increase translation quality and assessment. These solutions are intended to improve performance and handle issues in low-resource machine translation.

REFERENCES

Xu, L., Xie, H., Qin, S.Z.J., Tao, X. and Wang, F.L., 2023. Parameter-efficient fine-tuning methods for pretrained language models: A critical review and assessment. arXiv preprint arXiv:2312.12148.

Mohamed, Y.A., Khanan, A., Bashir, M., Mohamed, A.H.H., Adiel, M.A. and Elsadig, M.A., 2024. The impact of artificial intelligence on language translation: a review. IEEE Access, 12, pp.25553-25579.

Cayamcela, M.E.M. and Lim, W., 2019, February. Fine-tuning a pretrained convolutional neural network model to translate American sign language in real-time. In 2019 International Conference on Computing, Networking and Communications (ICNC) (pp. 100-104). IEEE.

Zhang, K., 2021, May. Application of Pretrained Models for Machine Translation. In 2021 International Conference on Communications, Information System and Computer Engineering (CISCE) (pp. 849-853). IEEE.

Mantoro, T., Asian, J. and Ayu, M.A., 2016, October. Improving the performance of translation process in a statistical machine translator using sequence IRSTLM translation parameters and pruning. In 2016 International Conference on Informatics and Computing (ICIC) (pp. 314-318). IEEE.

Sun, S., Hou, H.X., Yang, Z.H. and Wang, Y.S., 2023, June. Multilingual Pre-training Model-Assisted Contrastive Learning Neural Machine Translation. In 2023 International Joint Conference on Neural Networks (IJCNN) (pp. 01-07). IEEE.

Thillainathan, S., Ranathunga, S. and Jayasena, S., 2021, July. Fine-tuning self-supervised multilingual sequence-to-sequence models for extremely low-resource NMT. In 2021 Moratuwa Engineering Research Conference (MERCon) (pp. 432-437). IEEE.

Tran, Q.D., 2024, August. Exploring Low-Resource Machine Translation: Case Study of Lao-Vietnamese Translation. In 2024 International Conference on Multimedia Analysis and Pattern Recognition (MAPR) (pp. 1-6). IEEE.

Hallac, I.R., Ay, B. and Aydin, G., 2018, September. Experiments on fine tuning deep learning models with news data for tweet classification. In 2018 International Conference on Artificial Intelligence and Data Processing (IDAP) (pp. 1-5). IEEE.

Saji, J., Chandran, M., Pillai, M., Suresh, N. and Rajan, R., 2022, November. English-to-Malayalam Machine Translation Framework using Transformers. In 2022 IEEE 19th India Council International Conference (INDICON) (pp. 1-5). IEEE.

Premjith, B., Kumar, M.A. and Soman, K.P., 2019. Neural machine translation system for English to Indian language translation using MTIL parallel corpus. Journal of Intelligent Systems, 28(3), pp.387-398.

Nair, J., Krishnan, K.A. and Deetha, R., 2016, September. An efficient English to Hindi machine translation system using hybrid mechanism. In 2016 international conference on advances in computing, communications and informatics (ICACCI) (pp. 2109-2113). IEEE. [13] Unnikrishnan, P., Antony, P.J. and Soman, K.P., 2010. A novel approach for English to South Dravidian language statistical machine translation system. International Journal on Computer Science and Engineering, 2(08), pp.2749-2759.

KM, S.K., Namitha, B.N. and Nithya, R., 2015. A comparative study of English to Kannada baseline machine translation system with general and bible text corpus. International Journal of Applied Engineering Research, 10(12), pp.30195-30201.