JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

A REVIEW ON TRANSLATION TECHNIQUES AND MULTILINGUAL BRIDGING WITH ENGLISH AND KANNADA

[1]Spoorthi, [2]Sudarshan K, [3]Rashmi K P, [4]Greeshma K, [5]Rutika P Gowda

[1]Third year B.E Student, [2]Head of the Department ISE, [2]Third year B.E Student, [3]Third year B.E Student, [4]Third year B.E Student,
[1]Department of Information Science and Engineering
[1]Srinivas Institute of technology, Mangaluru, India.

Abstract: In recent years, the neural machine translation (NMT) has made considerable progress, yet due to lack of parallel corpora and linguistic equipment, has been challenged by a low-resources language pairs such as Kannada-Tulu and English-Tulu. To address this, various strategies -including manual translations, automated text generations, and hybrid approaches -have been employed to create a Kannada -Tulu dataset. Encoder-Decoder architecture such as RNNs, BiRNNs, LSTMs, and transformers is detected, with transformer-based models show promising Bleu and CHRF scores. Linguistic feature-embedded NMT models further improve the quality of translation to the agglutinative Dravidian languages. Additionally, the hybrid system exhibited high accuracy for English-Tulu translation by combining rules-based and nervous approaches. Beyond the textual translation, CNN-based models such as BARAVU-Tulu, effectively identify functional and handwritten Tulu scripts of communication in multilingual areas. Studies also highlight evaluation challenges in morphological -rich languages like Malayalam, indicating the requirement of refined evaluation methods. Collectively, these progresses contribute to bridging digital division for lower-practiced languages**..**

IndexTerms **- Neural Machine Translation, low-resource languages, Kannada–Tulu translation, transformer models, parallel corpus, linguistic features, hybrid translation approach, Tulu script recognition..**

I. INTRODUCTION

Language acts as an important medium for communication, incorporating symbols and sounds that express ideas, feelings and information. In a culturally diverse country like India, where hundreds of languages are in co -existence, effective communication in linguistic boundaries is a requirement and a challenge. Among many language families present, the Dravidian Group - Kannada and Tulu's worshipers - are particularly prominent in southern India. Kannada, spoken by more than 44 million people, is a constitutionally recognized language and has paid moderate attention in the field of computational linguistics. On the other hand, Tulu, although coastal Karnataka and North Kerala are spoken by more than 2.5 million people, are one of the most revived and technically unseen languages in the country.

Machine Translation (MT), a subfield of natural language processing (NLP) has developed more recent and up to powerful nerve machine translation (NMT) model than the initial rule-based machine translation (RBMT) system. While high-resourceful languages benefit from strong parallel corpora and pre-trained models, low-resources languages such as Tulu face many obstacles, including lack of standardized scripts, limited digitized materials, and lack of insufficient linguistic resources. The agglutinate nature of both Kannada and Tulu adds more complexity to the efforts of nature and morphological prosperity translation, demanding a more sophisticated modeling approach.

Tulu, although historically equipped with its script—Sigillary—has seen most of its written communication in the Kannada script due to the absence of Unicode support and limited script literacy. As a result, the recognition and translation of Tulu is forced, which leaves a difference in inclusive digital services such as healthcare, education and governance. In addition, while Kannada has been depicted in previous MT studies, Tulu has not yet found significant technical intervention in this regard. This research addresses pressure to develop an effective MT system for the Kannada-Tulu language pair. The study involves the creation of a parallel corpus through both manual and regulatory approaches and evaluates various NMT architecture including RNN, Birn, and transformers equipped with GRU and LSTM units. Techniques such as Byte Pair Encoding (BPE) are also employed to address the out-of-vocabulary (OOV) issue prevalent in complex languages.

In parallel, the study integrates part-off-speech (POS) tagging to increase additional linguistic characteristics such as model performance. In addition, this task also contributes to script recognition by employing a handwritten Tulu characters to digitize and identify the CNNS, which not only supports translation, but also conservation of cultural identity.

By addressing both language modeling and script recognition, this study represents a leading step towards bringing Tulu into digital mainstream. The results are expected to benefit not only native speakers, but also non-Tulu speakers, researchers and policy makers, who promote inclusive access to linguistic resources. Ultimately, this work wishes to bridge the gap between tradition and technology, and two linguistic rich yet between unqualified languages -Kanada and Tulu.

II. Related Works

In [1] paper proposes an efficient system to identify Malayalam text from natural visual images, which take advantage of traditional image processing techniques such as edge detection, binning of Otsu, scalenus improvement and bounding box segmentation. For recognition, the system uses template matching, which works well in a controlled environment with minimal noise and similar lighting, but struggles with real -world challenges such as uneven lighting and complex backgrounds. Being effective in low-resources settings, the dependence of the method on the template limits its adaptability for various fonts and scripts on matching. The author accepts these limitations and suggests future reforms, including deep education, real-time processing, multilingual support and integration of linguistic post-processing for better scalability and accuracy. Overall, the paper provides a strong foundation for Malayalam visual lesson recognition and highlights a clear path for future work to increase the system using modern AI-operated techniques, making it stronger for practical applications.

In [2] author has proposed a hybrid ASR system designed for Malayalam, a morphological language, which is using syllable-based sub word tokens for the production of pronounced L AX Xi'an (PLS) and language models (LMS). They take advantage of the linguistic rules of the Emulf Python Library and incorporate them into the model, comparing syllable token-based systems with traditional word-based models. The study suggests that syllable-based models outperform word-based models, especially in high out-of-vocabulary (OOV) scenarios, by significantly reducing the Word Error Rate (WER).In addition, the syllable-based approach reduces the memory requirements for decoding while maintaining competitive accuracy. However, in a low -year environment, term -based models still perform better. The paper highlights the need for a dictionary expansion to cover rare words and customize trade-bands between memory and WER. Linguistic rules for syllables also require purification to handle different references. This function offers potential instructions for further detection of data-powered, especially in multilingual ASR systems, which can produce more effective models in domains of different languages, submitting diuretics methods.

In [3] study presents significant efforts to create Kannada-Tulu-parallel corpus for both directions and develop baseline neural machine translation models (NMT). Corpus with 117 262 sentence pairs was created by manual translation of original speakers and provides linguistic rules benefits from the automated text generation (ATG) system based on the verb's attitude and syntactic pattern. This study evaluates various NMT architectures, with Chef-score of 0.502 and 0.598, respectively, with blue score of 0.241 (Kannada-Tulu) and 0.341 (Tulu-Karnad), Transformers model with Byte-Precoding (BPE). Manual translation ensured the quality of the data, but was a resource intensive, while ATG provided scalability at the expense of syntactic accuracy due to morphological complexity. The authors have proposed future updates through Rich ATG rules and hybrid NMT models to better capture morphological noise. This research contributes valuable to the less resource -language process and supports the development of strong regional language translation tools.

In [4] study searches for improvement (NMT) by integrating Kannada and Libra-two low-summed, morphologically rich Dravid languages-part-part-bodies (POS) TS Guys into a transformer-based model. The 10,300-wave parallel corpus was pre-processed, with POS TS GS produced using IIIT-Hyderabad shallow parser for Kannada. These TS GS Source was linked to tokens and used to train the model on the OpenNMT framework. POS-UND Model Dale showed significant improvements on the baseline, better Blue (21.99 vs. 20.53), CHRF (35.28 vs. 34.88), and average CHRF (50.78 vs 50.20). The POS-T-GED approach benefited the translation of morphologically complex or influenced sentences, reducing blurring and error. However, due to small corpus size and complex morphology of Tulu, especially the limits for long or rare-words sentences are left. This function indicates the possibility of linguistic feature integration and indicates future directions, including adding morphological and tight information, expanding corporations and expanding the method to other Dravid languages.

In [5] study presents an integrated system for the Tulu Language Accreditation and Translation, which aims to remove the digital gap for this under-resourced Dravidian language. It contains three key components: Neural Machine Translation (NMT) using CNN-based handwritten character evalidation, rule-based English-to-fool translation, and a coder-diver-diver model with LSTM. Dataset with 30,500 handwritten compatible characters was done to train the CNN model, with 92% accuracy. The translation component works 1,481 manually overall words and 1000 sentence pairs, which reach 89% for rule -based translation words and 81% for sentences. The NMT system performed strongly with a blue score of 0.83 (Tulu to Tulu) and 0.65 (Tulu to English). When CNN character shows strong performance in characteristics, translation systems with syntactically complex sentences are struggling due to data restrictions. This study suggests expanding data sets, integrating Unicode support and developing real-time applications in future work. This research plays an important role in digitalization and maintenance of Tulu language.

In [6] examine explores screwing to Malayalam the usage of Neural Machine Translation (NMT) using English paraphrase datasets including GyAFC. Four fashions - Multi -Oog, naked, opus and a synonymous replacement approach - were tested by using translating English paraphrases in Malayalam thru Google Translate and refining them using NMT fashions. A guide curated dataset with 800 pairs of Malayalam sentences supported the automatic and human assessment. While OPUS finished the highest rating in Bleu (0.34), Meteor (zero.63) and Kosinu equality (0.83), man evaluated the most effective techniques as synonymous substitute and found out a disconnection between metric score and skilled quality. Interestingly, the Malayalam -precise multi-Indian model is higher adapted to human judgment, even though it is decrease in computerized assessments, highlighting the restrictions on cutting-edge calculations for agglutinative languages. The observe emphasizes the need for Malayalam -specific assessment measurements, large native commercial enterprise models and hybrid -language to enhance paraphrases generation. Overall, it offers precious records on NMT model to regional language remedy and improvement of semantic relevance in low resource environments.

In [7] BARAVU-Tulu Lipi identity mission offers a commendable initiative to keep the Tulu script through an automated device that uses Neural Convolution Networks (CNNS). By healing facts units with published and manuscript Tulu indicators and the use of information growth strategies, the mission correctly educated CNN fashions, including VGG and redefinar, using Pytorch. The gadget demonstrated excessive accuracy and robustness in the identification of Tulu scripts in special writings and manuscript patterns, with validation and take a look at that showed their generalization capabilities. The foremost challenges blanketed encompass marked facts shortage, handwriting variability and want to capture linguistic tones. Pre -treatment care and fashions optimization performed a crucial role in overcoming those barriers. The imaginative and prescient of the assignment of expanding for actual -time packages for training, cultural conservation and tourism, along with capacity integration in cellular platforms and OCR systems, emphasizes its broader meaning. This work contributes significantly to the virtual revitalization of Script Tulu and improves the supply of sub -supplied languages.

In [8] study presents a concentrated discovery of the transfer of teaching techniques for visual lesson recognition in six Indian languages, and deals with script complexity and lack of data. By using deep learning models such as CRNN and Star-Net, supplemented with a corrective Blastoma module, research emphasizes adaptability and accuracy through training on synthetic datasets. Conclusions suggest that the cross transfer between Indian script achieved much better due to linguistic differences when transferred learning from English to Indian script, especially the degree of recognition (WRR) for Hindi, Telugu, Bungalow and Malayalam improved. Major challenges include dealing with morphological prosperity of Indian languages, improving synthetic data set realism and increasing the strength of models in writings and writing styles. The study emphasizes the boundaries of the English -centered model and requires more language -specific approaches. This provides valuable insight into developing multilingual OCR systems and highlights the need for datasets of the real world, rich in future OCR progress, rich in supporting lower resource languages.

In[9] take a look at presents a sizeable development in scene textual content recognition for Indic scripts by using adapting a hybrid CNN-RNN architecture for Devanagari, Telugu, and Malayalam. By combining convolutional layers for feature extraction with bidirectional LSTMs for collection modeling, and schooling give up-to-cease the usage of CTC loss, the version successfully bypasses the want for phrase segmentation. The advent of two massive-scale datasets—a synthetic dataset with 4 million word pictures in keeping with script and the actual-global IIIT-ILST dataset—enables thorough benchmarking. The hybrid version outperforms conventional RNN-OCR strategies, with first rate gains in Word and Character Recognition Rates, accomplishing seventy three . Four% WRR and 92.Eight% CRR for Malayalam. The model demonstrates robustness towards noise, font, and heritage variability in real scenes. However, the take a look at also recognizes limitations in artificial records and emphasizes the need for greater great real-scene datasets. This work lays a robust basis for reinforcing scene text recognition in beneath-resourced Indic scripts and paves the manner for destiny multilingual OCR systems.

In[10] study introduces a robust and systematic method for Kannada text recognition from stage images, structured to pre -treatment, detection, characteristic extraction and stages of recognition. By using bilateral and Viennese filters for noise discount, the study uses a conventional block attention-based-based YOLO V7 model for text detection and ensures that accurate delimitation box identification. For functional extraction addresses a tightly stacked LSTM network combined with the Densenet-121 structure disappearing gradient problems, while recognizing is achieved through a conventional remaining network-assisted autoocoder (Cresnet-AE), and provides tremendously correct Kannada text strings. The machine achieves an outstanding 97.eight% recognition price, and surpasses conventional strategies such as CNN and horror in precision, remembering and accuracy, and effectively manages different text patterns and noise. However, the challenges continue to improve the generalizability of version for actual global situations and overcome the lack of significant actual international datasets. Future work should be awareness of increasing the data sets, improving multilingual recognition talents and optimizing processing speed for wider usefulness, especially for low resources.

In[11] project introduces a rule-based machine translation (RBMT) system for English-to-Tulu translation, which addresses the language barrier in coastal Karnataka. The system employs linguistic rules, grammar and morphological, syntax and a bilingual dictionary for semantic analysis. It passes the English input using a morphological analyzer, coincides with Tulu structures, and a transfer generates accurate translation relevant through the lexicon and morphological generator. With the user and administrator interfaces, system lessons and image translations, as well as dictionary management allows. Initial results indicate strong performances for simple sentences and direct word translations, making it a promising tool for basic communication in Tulu. However, challenges remain in complex syntax structures, handicaps of idioms and ensuring accuracy in the rules application, as the rigid nature of RBMT can lead to errors. Future work includes expanding lexicon, involving statistical or nerve models for better relevant understanding and making the system accessible on mobile platforms to increase rural access.

In[12] observation introduces a neural machine translation (NMT) unit to translate Kannada into English using a sequence-to-seq (SEQ2SEQ) version based on long short-term memory (LSTM) network. The system uses a coder decoder structure, with pre-treatment steps such as tokenization, vectorization and serial adjustment performed into a customized parallel corpus of 41,000 Kannada-English sentence pairs. When we achieve a Bleu score of 31. Venty eight% on school data and 17.38% on take on data, the version of traditional statistical machine translation (SMT) exceeds Moses. With an accuracy of 86.32% and a validation deficiency of 0.849 after forty eras, the model demonstrates strong translation for simple sentences. However, challenges such as the limited availability of Kannada-English datasets, sentence form complexity and semantic protection are referred to, in connection with the absence of interest mechanisms, which affect handling long sequences.

In[13] research presents an innovative system for conversion of speech to Malayalam text, followed by English translation and offers a promising solution for multilingual communication. The system works in three steps: speech is transcribed to the text using the speech recognition library, followed by early treatment with the indecently library for tokenization and standardization, and eventually the Malayalam tab is translated into English using a custom machine translation model with Ctranslant2 and the sentence. The evaluation of an oral metal data set showed high precision in voting recognition and fluid translation in English. The system shows the effective use of modern NLP tools and personalized models to face the challenges of translation of complex and rich morphological languages as Malayla. Although the study recognizes limitations as difficulties with dialectal variations and complex phrases structures, the results are promising. Future jobs can focus on refining the model to better handle dialectal differences and further improve the quality of the most robust applications in the real world.

In[14] research introduces Bhasha Verse, a robust multilingual translation ecosystem designed to aid 36 Indian languages, consisting of Kannada and Tulu. The system integrates human-established, artificial, and pivot-based translation methods to create massive parallel corpora, and employs a multi-mission model that handles translation, grammar correction, submit-editing, and mistakes identification thru encoder-decoder architectures. Evaluated with over 10 billion sentence pairs and area-precise corpora, Bhasha Verse excels in discourse-degree translation and addresses key linguistic demanding situations, together with script variant, morphological complexity, and code-blending. The evaluation framework consists of each reference-primarily based and reference-free metrics, ensuring excessive translation nice. However, challenges stay in resource shortage for positive dialects and enhancing fashions for low-useful resource languages. Despite those hurdles, Bhasha Verse indicates robust promise in enhancing multilingual communication across India. Future work must consciousness on increasing help for underrepresented dialects and improving discourse coherence to in addition elevate translation performance in diverse actual-global packages.

In[15] article examines the use of machine learning to identify handwritten Tulu signs, which use shallow and deep learning techniques. For shallow learning, writers use artificial nervous networks (Ann), Support Vector Machines (SVMS) and Adbosts, which extract properties such as zone -wide density and shields against handwritten signs. Deep Learning Approach uses Deep Con Vision Neural Network (Deep CNN), which improves shallow models. The results show that Deep CNN obtains 98.49% accuracy in identifying Tulu signs from modern documents. However, the accuracy falls to 80.49% when the old Tulu Palm Leaf is used on manuscripts, which highlight the challenge of identifying the characters in humiliated or old texts. The conclusion of the paper is that although deep CNN is very effective for modern Tulu character recognition, further research is needed to improve the performance of old manuscripts. This task contributes significantly to the development of Tulu language therapy, potential, potential, for future improvement in historical document recognition.

III. CHALLENGES

The primary challenge in the progression of the Neural Machine translation (NMT) lies in its limited efficiency with under-re-language couples, roughly due to the lack of parallel corpora, in contrast to the high French available. This limit has imposed researchers against developing NMT techniques that are specially adapted to the lower purpose language pairs, focusing on the production and collection of parallel companies.

When it comes to Indian languages, this challenge is especially clear due to linguistic diversity and the presence of many languages with low resources. Indian languages, including Kannada and Tulu, are characterized by their morphological prosperity and agglutinative nature, which further complicates the development of effective MT systems. Parallel Corpora deficiency for these languages, combined with the complexity of their linguistic structures, is a significant obstacle to high -quality machine translations.

Current MT systems normally warfare to seize the complicated language homes of those languages, resulting in a negative translation point. This is in particular clear in the translation between Kannada and Tulu, in which Tulu's specific grammatical aspects and oral nature gives unique demanding situations.

In addition, these languages require a pleasing technique to assess the MT system that is going past the standard automatic matrix as blue. These matrices are typically not able to reflect the first-rate of translations nicely and highlight the want to encompass human selections and linguistic insights about the assessment system.

Combating these challenges requires a concrete effort in lots of regions: the development of strategies to create a parallel company for the low aid language, the development of NMT fashions that could efficiently deal with morphological and visual ceiling headaches of the ceiling of these languages and examine the translation high-quality accurately.

IV. SIGNIFICANCE AND IMPLICATION

Reviewed paper for under-resourced language such as Kannada and Tulu has a significant value in promoting neural machine translation (NMT), which is linguistic rich but digitally signed. This emphasizes the important need to develop MT systems to fit the linguistic diversity of multilingual countries such as India, especially in areas where local languages dominate daily communication. By focusing on Kannada and Tulu - Language - Language - Language - Language - Paper with limited parallel Corpora addresses a pressure difference in today's NLP research. Integration of linguistic characteristics such as POS tag and morphology in the NMT model improves translation translation accuracy and relevant understanding, making them more effective for real -world -applications in education, health care and governance. In addition, the development of the handwritten character data sets and lesson recognition systems for Tulu not only supports technical inclusion, but also contributes to the protection of cultural and linguistic heritage. The combination of rule -based and statistical techniques provides practical solutions when hybrid function and data are rare.

Overall, the task places emphasis on lingually informed models, sewn data sets and the evaluation beyond the Bleu score beyond the calculations to ensure cementic accuracy. Its widespread implications lies in promoting inclusion of indigenous languages through inclusiveness, digital access and AI, which establishes a foundation for future research in low-resources language processing .The development of NMT model corresponds to these under-relieved languages, as detected in papers, is important to increase the quality and accuracy of translation. In addition, the emphasis on making parallel corpora and employing advanced techniques such as transformer architecture and hybrid machine translation reflects a commitment to overcome the boundaries generated by lack of data. Research also contributes to the broad area of natural language processing (NLP) by addressing the nuances of low-resources languages and requires more refined assessment matrix beyond standard automatic measures such as blue. The development of resources and equipment for Tulu , including lesson recognition and translation systems, further underlines the importance of this work in preserving and promoting linguistic diversity.

V. FUTURE SCOPE

Future efforts will focus on many major directions to further improve the quality of translation and strengthen the system. The purpose of increasing the ATG component with more comprehensive linguistic rules is to solve the current boundaries, such as the generation of small sentences. Inclusion of more linguistic characteristics - stress, gender, number, and intensity changes - models promise the way to enrich input and improve reference handling. Additionally, the construction of a POS tagger for Tulu will enable both translations symmetrical POS-Embedded NMT for directions. For expansion of dataset, especially for complex and long sentences, model is a priority for improving normalization. For script recognition, future plans include expanding the CNN model into a real-time application with Tulu Unicode support and evaluating it on a comprehensive handwritten dataset.. In addition, standardized evaluation for Dravidian languages will enable more accurate benchmarking in the construction of metrics. These initiatives collectively aim to create more reliable, scalable and culturally important language technologies.

VI. CONCLUSION

This body of work represents an important step in the domain of low-resources language processing, especially the Dravidian languages focus on Kannada and Tulu. Construction of Kannada-Tulu parallel corpus is manually and capable of development and evaluation of various nerve machine translations (NMT) models with the help of rules-based ATGs. Among them, the Transformer+LSTM+BPE model demonstrated the advantages of a hybrid architecture, consistently outperforming others in both Kan–Tul and Tul–Kan translation directions . Additionally, introducing linguistic characteristics such as Part-of-Speech (POS) tags showed an average improvement in translation quality, validating the hypothesis that morphologically rich languages benefit from such integration.. Beyond translation, the Tulu script recognition through CNN, and the development of English -Tulu translation systems has designed the basis for practical applications in language conservation and education. Deep learning use for handwritten character recognition gained strong accuracy, which shows the feasibility of computational methods in processing low-term scripts such as the Tulu script. Together, these initiatives not only increase language technology equipment for low-resources languages, but also make meaningful contributions to their cultural and linguistic conservation.

REFERENCES

[1] R. Emmanuel and J. George, “Automatic Detection and Recognition of Malayalam Text from Natural Scene Images,” *IOSR Journal of VLSI and Signal Processing*, vol. 3, no. 2, pp. 55–61, 2013.

[2] K. Manohar, A. R. Jayan, and R. Rajan, “Syllable Subword Tokens for Open Vocabulary Speech Recognition in Malayalam,” in *Proceedings of the Third International Workshop on NLP Solutions for Under Resourced Languages (NSURL 2022)*, Trento, Italy, Dec. 2022, pp. 1–7.

[3] A. Hegde and H. L. Shashirekha, “KT2: Kannada-Tulu Parallel Corpus Construction for Neural Machine Translation,” in *Proceedings of the 20th International Conference on Natural Language Processing (ICON)*, Goa University, Goa, India, pp. 743–753, Dec. 2023.

[4] M. Supriya, U. D. Acharya, and A. Nayak, “Enhancing Neural Machine Translation Quality for Kannada–Tulu Language Pairs Through Transformer Architecture: A Linguistic Feature Integration,” *Designs*, vol. 8, no. 5, p. 100, 2024.

[5] A. P. Rodrigues, P. Vijaya, and R. Fernandes, “Tulu Language Text Recognition and Translation,” *IEEE Access*, vol. 12, pp. 12734–12744, 2024.

[6] C. Varghese, S. Koshelev, and I. P. Yamshchikov, “Neural Machine Translation for Malayalam Paraphrase Generation,” in *Proceedings of the Fourth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*, St. Julian's, Malta, pp. 10–15, Mar. 2024.

[7] M. V., S. Suraksha, A. V., S. A. N., and A. B. K., “A Comprehensive Review on BARAVU-Tulu Lipi Identification,” *Global Journal of Research in Engineering & Computer Sciences*, vol. 4, no. 2, pp. 105–109, Apr. 2024.

[8] S. Gunna, R. Saluja, and C. V. Jawahar, “Transfer Learning for Scene Text Recognition in Indian Languages,” in *Proceedings of the Document Analysis and Recognition Workshops (ICDARW)*, Lausanne, Switzerland, Sep. 2021, pp. 182–197.

[9] M. Mathew, M. Jain, and C. V. Jawahar, “Benchmarking Scene Text Recognition in Devanagari, Telugu and Malayalam,” in *Proc. Document Analysis and Recognition Workshops (ICDARW)*, 2021, pp. 1–6.

[10] M. Prasad Y.N. and C. H.K., “Recognition of Printed Kannada Text in Scene Images using Machine Learning Techniques,” *International Journal of Intelligent Systems and Applications in Engineering (IJISAE)*, vol. 11, no. 1, pp. 4233–4238, 2023.

[11] A. Shenoy, M. P. Rao, V. K. Shenoy, V. P. Kudva, and V. Nayak, “English to Tulu Translator,” *International Research Journal of Engineering and Technology (IRJET)*, vol. 7, no. 4, pp. 6227–6231, Apr. 2020.

[12] P. K. Nagaraj, K. S. Ravikumar, M. S. Kasyap, M. H. S. Murthy, and J. Paul, “Kannada to English Machine Translation Using Deep Neural Network,” *Ingénierie des Systèmes d’Information*, vol. 26, no. 1, pp. 123–127, Feb. 2021.

[13] A. K. Sunny, D. Thomas, J. S. Simon, M. Z. Rafeeque, and T. M. Yaseen, “A Novel Approach to Malayalam Speech-to-Text and Text-to-English Translation,” *International Journal of Scientific Research in Engineering and Management (IJSREM)*, vol. 6, no. 5, pp. 1–5, May 2022.

[14] V. Mujadia and D. M. Sharma, “BhashaVerse: Translation Ecosystem for Indian Subcontinent Languages,” *arXiv preprint arXiv:2412.04351*, Dec. 2024.

[15] C. K. Savitha and P. J. Antony, “Machine Learning Approaches for Recognition of Offline Tulu Handwritten Scripts,” *Journal of Physics: Conference Series*, vol. 1142, no. 1, p. 012005, 2019.