IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.53

Volume 4, Issue 1, May 2024

Educational Resource Material Translation System

Mr. Satheesh Chandra Reddy[1], Ms. Medhini TR[2], Ms. Shirisha L[3] , Ms. Tanushree P [4], Ms. Supritha KS[5]

Head of Department , Department of Information Science and Engineering[1]

U.G. Students, Department of Information Science and Engineering[2,3,4,5]

S J C Institute of Technology, Chikballapur, India

Abstract*: The proposed "Educational Resource Material Translation System" represents a significant leap in bridging language barriers, particularly for individuals who lack proficiency in English. By leveraging advanced technologies such as Python, Flask, Recurrent Neural Network (RNN) and Large Language Models (LLM), the system offers a robust solution for translating educational content from English to prominent Indian regional languages like Kannada, Telugu and Hindi. Its user- friendly web interface, developed with Flask, ensures a seamless and intuitive translation experience for users of varying technical backgrounds. Beyond mere language translation, this project facilitates cross- cultural communication and enhances accessibility to educational resources, thereby empowering individuals and businesses alike. The utilization of Machine Learning further underscores the system's adaptability and potential for delivering precise and efficient translations. In essence, the "Language Translator" not only serves as a valuable tool for linguistic inclusivity but also fosters knowledge dissemination and cultural exchange on a broader scale.*

Keywords: Python ,Flask (web framework) ,Recurrent Neural Network (RNN) ,Large Language Models (LLM) ,Target Languages

I. INTRODUCTION

The "Educational Resource Material Translation System" breaks language barriers, converting English educational content into Kannada and Hindi, enabling broader access to learning. It promotes inclusivity and enhances understanding for non-English speakers. Leveraging Python, Flask, and RNN algorithms, the Language Translator ensures fluent, contextual translations. It aims to bridge the gap between English and Indian regional languages, fostering effective communication. By capturing linguistic nuances, it improves user experience and promotes cross-cultural understanding. The system's primary goal is to facilitate seamless knowledge exchange across linguistic boundaries. In a diverse, interconnected world, such translation systems play a crucial role. They empower individuals and communities by offering accessible educational resources. Through advanced technology, the Language Translator enhances fluency and maintains context. It fosters global connectivity by facilitating effective communication across languages. Ultimately, it contributes to a more inclusive and accessible educational landscape.

II. LITERATURE REVIEW

The literature review for the Educational Resource Material System underscores the significance of technology-driven approaches in addressing language barriers in education. Studies emphasize the importance of language accessibility for effective learning outcomes and highlight the role of language translation systems in promoting inclusivity. Scholarly works explore methodologies, including machine learning algorithms like RNN, to enhance translation accuracy and fluency, while stressing the need for user-friendly interfaces. Additionally, research underscores the potential impact of such systems in bridging the gap between English and regional languages, particularly in diverse societies like India, with case studies demonstrating improved learning outcomes and cultural exchange. Yasir abdelgadir mohamed , (member, ieee), akbar khanan, mohamed bashir1,abdul hakim h. M. Mohamed , (senior member, ieee), mousab a. E. Adiel ,and muawia a. Elsadig [1]It evaluates the current state, challenges, and advancements in AI language translation. Various methods like Statistical Machine Translation (SMT) and NMT techniques are analyzed. The importance of NLP, fuzzy logic, and feature extraction in improving translation quality is highlighted. The document emphasizes the significance of AI techniques like sentiment analysis and word embedding. It mentions the Transformer Architecture and Transfer Learning for better NLP applications. The study suggests further research to enhance translation quality and overcome limitations. Authors advocate for testing across different corpora and model improvements. The paper provides insights into machine translation systems and their capabilities. It concludes by emphasizing the need for continuous advancements in translation technology. Jacques Melitz ,Farid Toubal [2] Variables like common official, spoken, and native languages, distance, adjacency, and historical wars are considered. Language factors have a significant impact on trade, with common official and spoken languages showing positive effects. Common native language results are mixed, while linguistic proximity and common religion also play a role. Data collection involved 42 common second languages spoken by at least 4% of the population in different countries. The methodology includes constructing a common language index based on various linguistic factors. The study addresses issues of double-counting in language data and the normalization of linguistic measures. Results indicate the importance of linguistic influences on trade, with specific coefficients for different types of goods. The analysis also explores the separate importance of English and other major world languages in trade relationships. Overall, the research highlights the complex interplay between language and trade dynamics. Tosho hirasawa, masahiro kaneko , aizhan imankulova,and mamoru Komachi [3] This study focuses on enhancing Multimodal Machine Translation (MMT) models by incorporating pre- trained word embeddings and leveraging vision-language Language Models (LMs). The research proposes a Transformer-based MMT model that demonstrates the effectiveness of visual features in improving translation quality. The study explores debiasing methods to enhance model performance and trains LXMERT-fused models in two steps for improved results. Additionally, the impact of subword- level tokenization on pre-trained embeddings is investigated. Overall, the research aims to advance MMT models by integrating visual features and linguistic knowledge to achieve better translation quality. Prathwini1, anisha p. Rodrigues 2, p. Vijaya 3, and roshan fernandes [4] The PDF discusses Tulu language text recognition and translation. It outlines the methodology for Tulu character recognition, including image collection and preprocessing techniques. The document highlights the use of PCA for feature extraction and dimensionality reduction in image processing. It mentions the challenges in OCR for recognizing Telugu literature characters and the importance of dataset expansion. The research proposes an English-to-Tulu translator using rule-based and neural machine translation methods. It presents the performance analysis of the translation model, achieving 89% accuracy for word-based translation. The study also covers related works on handwritten character recognition and machine translation techniques. The authors include Prathwini, Anisha P. Rodrigues, Vijaya, and Fernandes. Sugyeong eo 1, chanjun park 1,2, hyeonseok moon 1,jaehyung seo 1, and heuiseok lim [5] Cross-Lingual Language Model (XLM):Trains language understanding through self-supervised learning with objectives like CLM, MLM, and TLM.Word-Level Quality Estimation (QE):Utilizes XLM-R, mBART, and XLM-MLM for QE model training.Model Performance Analysis:XLM-R-large shows best performance among mPLMs.Dataset Construction Strategies:M-based, P-based, and H-based strategies for generating word- level Korean-English pseudo-QE datasets. Research Focus:Neural machine translation, quality estimation, multilingual pretrained language models, and natural language processing

III. EXISTING SYSTEM

Addresses challenges like parallel corpora scarcity and system progress needs. Capable of translating Arabic dialects and legal terms. Machine Learning (ML) and Deep Learning (DL):ML and DL contribute to advancements in language translation.ML trains systems to analyse data patterns, improving performance. a gravity equation model to analyse the impact of language on international trade. It considers variables such as common native language, linguistic proximity, common legal system, common religion, and history of wars. The study also discusses the importance of ethnic ties, ease of communication, and trust in influencing trade patterns based on language factors. Utilized LXMERT as feature extractor for MMT. Addressed limitations of MMT models compared to text-only counterparts. Lack of large-scale data and moderate improvements with visual features. Leveraged monolingual corpora for MMT model enhancement. Previous studies focused on textual parallel corpora for MMT training. Proposed debiasing procedures for word embedding and VLM in MMT. The existing system of the paper focuses on Korean-English word-level Quality Estimation (QE) in neural machine translation. It includes the use of XLM and mBART

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, May 2024

models pretrained with Common Crawl corpus for improved performance in low-resource languages. The system also involves data generation strategies such as M-based and H- based methods, along with model analysis considering tokenization methods and

English	Tulu
How are you?	ಎಂಚ ಉಳ್ಳರ್?
Is everyone good at home?	ಇಲ್ಲಡ್ ಮಾತೆಲ್ಲ ಉಷಾರ್ ಉಳ್ಳರ
What are you having today?	ಇನಿ ದಾದಾ ತಿನಿ ನಿ?
Where did you go?	ಒಡ ಪೋಯಿನಿ?
Puru went to Bangalore.	ಪುರು ಬಂಗಳೂರುಗ್ ಪೋಯೆ.
Does it take a while?	ಮಸ್ತ್‌ಪರ್ತಲ ದೆತೊನುನ?
Give me a masala dosa.	ಎಂಕ್ ಮಸಾಲ್ ದೋಸೆ ಕೊಲ್ಲಲ.
I know little Tulu.	ಎಂಕ್ ಚುರು ರ್ತಳು ಬಪುಲಂಡು.
Pratheeksha likes it.	ಪ್ರತೀಕ್ಷಗ್ ಇಷ್ಟ ಆಪುಂಡು.
We will come back.	ಎಂಕುಲು ಪಿರ ಬರ್ಪಲ.
Is there a telephone nearby?	ದಯುವಿಟ್ಟುಟ ಸಮೋಪ್ದಲ್ಲು ಟೆಲ್ಲ್ಫೋನ್ ಇದೆಯಾ?
May I use your telephone?	ಇನಲ ಮೊಬೈಲ್ ಉಪ್ಯೋಗ ಮನ್ನಿ ಲ್ಲಯ?
My name is Prathwini.	ಎನಿ ಪುದರ್ ಪ್ರಥ್ವಿ ನಿ.
Please give the bill.	ದಯಾಮಲ್ ರಶೋØ ಕೊಲ್ಲಲ.
Please wait for a moment.	ದಯಾಮಲ್ ಚುರು ಕಾಪುಲ್ಲ.
Thank you very much!	ಮಸ್ತ್‌ಸೊಲು!
Ganesh went to London.	ಗಣೇಶ ಲ್ಂಡನ್ ಪೋಯೆ.

Table No1: Training sentences of neural machine translation

noising schemes. Additionally, the system explores QE applications for document and word-level translation quality assessment. There were no existing translators available for English to Tulu or Tulu to English at the time of the literature survey. The research aimed to fill this gap by developing a system for translating English to Tulu and vice versa

IV. PROPOSED SYSTEM

Objectives of the Proposed Work

The objective of the proposed system is to address the lack of available translators for the Tulu language by creating a system that can accurately translate English to Tulu and vice versa. The goal is to fill the gap in Tulu language translation tools by developing a system that can precisely translate between the two languages. The research aims to create a system that can recognize Tulu characters through various steps like image pre-processing, segmentation, feature extraction, classification using machine learning algorithms, and performance comparison. Additionally, the system endeavors to utilize both rule-based and neural machine translation techniques to achieve accurate English-to-Tulu language translation. The proposed system also focuses on improving translation accuracy for complex sentences and phrases by combining rule-based and neural machine translation methods.

Precision:

$$\text{Precision} = TP / (TP + FP)$$

Recall :

$$\text{Recall} = TP / (TP + FN)$$

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, May 2024

F1 Score:

F1 Score = (2- Precision -Recall) / (Precision +Recall)

Precision. The precision or positive predictive value (PPV) is described as the ratio of correct prediction to the total correct values including the true and false predictions and is depicted mathematically as follows:
Precision= TP / TP+ FP.
Recall. The recall or sensitivity or true positive rate (TPR) is described as the ratio of correct predicted values to the sum of correct positive predictions and the incorrect negative predicted values and is depicted mathematically as follows:
Recall= TP/ TP + FN
F1-Score
The F-measure (Fβ) is described as the weighted average of the values obtained from the calculation of precision and recall parameters. Whenever the distribution of class is not even, then the value of F1 − Score is highly important than the accuracy value. The F1 − Score is depicted mathematically

V. METHODOLOGY

Tulu Character Recognition: Follows steps like Image Pre-processing, Image Segmentation, Feature Extraction, Classification using machine learning algorithms, and Performance Comparison. Data Preparation and Collection: Involves collecting various handwritten characters of about 62 classes, including 50 characters and 12 numerical, totalling 31,000 images. Image Preprocessing: Applies augmentation techniques like rotation, zoom, and brightness adjustments to enhance image quality. Image Segmentation: Utilizes contour techniques to identify distinct shapes or patterns within images. Classification Using Machine Learning Algorithms: Trains on a dataset with Convolutional Neural Network for multi-class categorization with 62 classes. Translation Technques: Utilizes rule-based and neural machine translation methods for English-to-Tulu language translation. Performance Evaluation: Achieves high accuracy in Tulu character recognition and translation tasks through a combination of techniques.

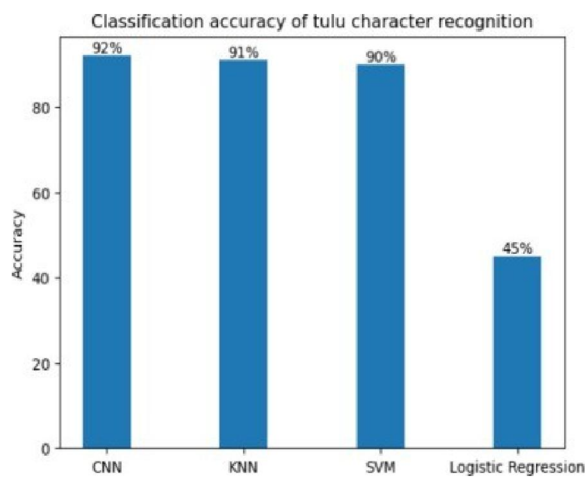

Fig No 1:Tulu character recognition accuracy classification

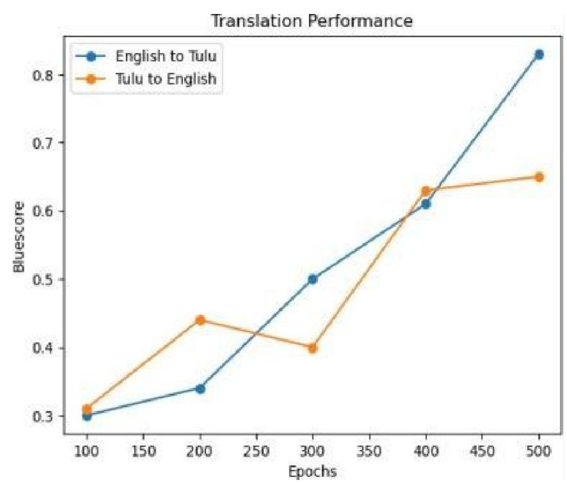

Fig No 2: Analysis of blue score.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, May 2024

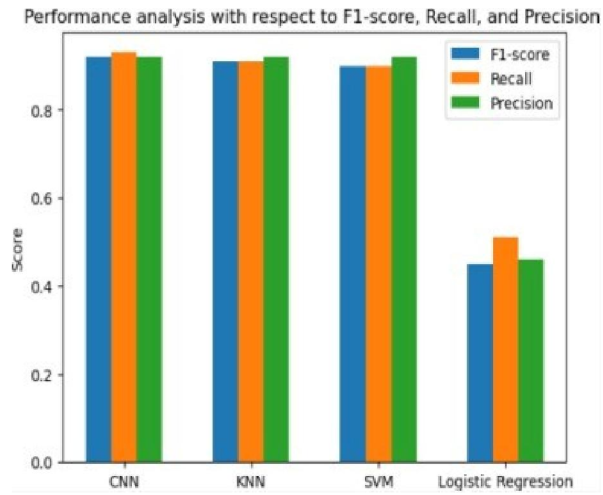

Fig No:3 Performance analysis of Tulu character.

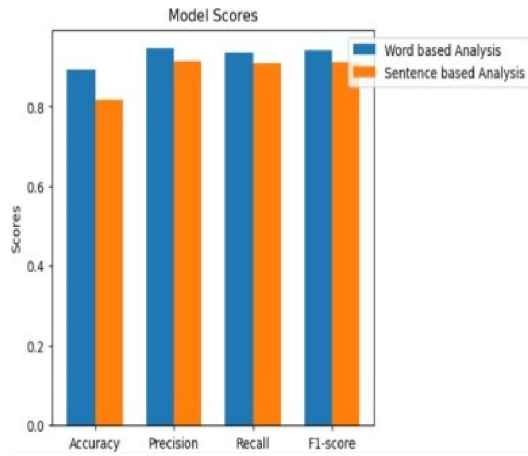

Fig No:4 Performance analysis of rule based method

Outcomes of the Proposed System

The outcomes of the proposed system could be manifold, including: Increased Availability of Translators: By developing a system capable of accurately translating between English and Tulu, the project could contribute significantly to addressing the lack of available translators for the Tulu language. This could have positive implications for communication, education, business, and various other domains where language translation is essential. Improved Translation Accuracy: The system's integration of both rule-based and neural machine translation techniques aims to enhance translation accuracy. Rule-based methods can handle specific linguistic rules and patterns, while neural machine translation can capture the nuances of language through extensive training on large datasets. By combining these approaches, the system can provide more precise translations, especially for complex sentences and phrases. Recognition of Tulu Characters: Through the implementation of image pre-processing, segmentation, feature extraction, and classification using machine learning algorithms, the system aims to achieve accurate recognition of Tulu characters. This would facilitate better input processing for translation tasks, particularly in scenarios where text is not readily available in digital formats. Performance Comparison: The research includes performance comparison among different stages of the system, such as image pre- processing, segmentation, feature extraction, and classification. This evaluation could provide insights into the effectiveness of various techniques and algorithms employed in the system, helping to refine and optimize its performance further. Contribution to Research and Development: The development of this system involves research into the integration of multiple approaches and techniques, including machine learning, image processing, and natural language processing. The outcomes of this research could contribute to advancements in these fields, potentially leading to improvements in other language translation systems or related technologies. Overall, the proposed system aims to address the specific challenges associated with translating between English and Tulu, with the potential to make significant strides in improving translation accuracy and accessibility for the Tulu language

VI. CONCLUSION

In conclusion, the Educational Resource Material Translation System is instrumental in overcoming linguistic barriers by translating educational content from English to regional languages like Kannada, Hindi, and Telugu. This initiative champions inclusivity, cultural appreciation, and equal access to educational materials, fostering a diverse learning environment. By leveraging this system, individuals from different linguistic backgrounds can access knowledge effectively. It contributes to a more enriched educational landscape by ensuring that learning resources are accessible to all. Ultimately, it plays a crucial role in promoting cross-cultural understanding and democratizing education.

REFERENCES

[1] Yasir abdelgadir mohamed , (member, ieee), akbar khanan, mohamed bashir1,abdul hakim h. M. Mohamed , (senior member, ieee), mousab a. E. Adiel ,and muawia a. Elsadig " The Impact of Artificial Intelligence on Language Translation" Received 24 January 2024, accepted 4 February 2024, date of publication 16 February 2024, date of current version 23 February 2024.

[2] Jacques Melitz , Farid Toubal "Native language, spoken language, translation and trade" Received 23 January 2013Received in revised form 1 April 2014 Accepted 1 April 2014

[3] Tosho hirasawa , masahiro kaneko , aizhan imankulova, and mamoru Komachi "Pre-Trained Word Embedding and Language Model Improve Multimodal Machine Translation" Received 26 April 2022, accepted 4 June 2022, date of publication 22 June 2022, date of current version 30 June 2022.

[4] Prathwini1, anisha p. Rodrigues 2, p. Vijaya 3, and roshan fernandes "Tulu Language Text Recognition and Translation" Received January 30, 2022, accepted April 18, 2022, date of publication April 21,
2022, date of current version May 2, 2022.

[5] Sugyeong eo 1, chanjun park 1,2, hyeonseok moon 1,jaehyung seo 1, and heuiseok lim "Word- Level Quality Estimation for Korean-English Neural Machine Translation" Received January 30, 2022, accepted April 18, 2022, date of publication April 21, 2022, date of current version May 2, 2022