

Halegannada to English Translation using Machine Learning

Rajani¹, Bhuvaneshwari V.P², Harshitha H.M³, Prarthana K⁴ and Varshini D⁵

¹⁻⁵Maharaja Institute of Technology Thandavapura, Mysore, India

Email: rajiniravigowda@gmail.com, bharathbhuvana9@gmail.com, hmharshitha909@gmail.com, pprarthanak@gmail.com, varshinidha912@gmail.com

Abstract— Halegannada, the ancient Kannada language, poses key linguistic challenges due to its intricate structure and scarce digital resources. The suggested system facilitates hassle free translation of Halegannada to English through the union of static lexical mapping with intelligent API-based processing. A hybrid pipeline is implemented using a special Halegannada Hosagannada dictionary combined with pre-trained external APIs for improved accuracy. Tokenization, word-level mapping, and sentence reconstruction are employed by the algorithm to preserve grammatical and contextual correctness. Transfer based mapping and semantic reordering address the absence of large parallel corpora. Error handling processes control OCR noise, dictionary misses, and image irregularities. Google Gemini OCR reads text off of historical documents, with the DeepSeek Translation API translating to English. Dictionary lookup, image preprocessing using Pillow, and API interaction are controlled by a modular Python backend through workflows. This complete framework demonstrates an effective, scalable approach for translating low-resource languages and assists in preserving Halegannada literature in a digital format.

Index Terms— Google Gemini OCR, DeepSeek Translation API, Dictionary lookup, Tokenization, word-level mapping, image preprocessing.

I. INTRODUCTION

Keeping historical languages alive matters it's how we hang on to our cultural roots but let's be honest halegannada's old script and words throw a curveball at most modern readers that's why we built the halegannada translation hub. It runs ancient kannada through a hand-crafted dictionary turning it into modern kannada first you can drop in text or even snap a photo and pull text out with google geminis ocr after that the deepseek api takes over and delivers a clean english translation immediately. Centuries of kannada literature aren't locked away anymore they're right there for anyone to explore.

II. LITERATURE SURVEY

A. Computational Approach For Halegannada To Hosagannada Poem Translation

This paper presents an LSTM-based method for translating Halegannada poetry into Modern Kannada, achieving 77.92% accuracy in splitting complex ancient word forms. It combines deep learning with a 20,000-entry custom dictionary and newly built datasets, outperforming GRU and BiLSTM models.

As the first focused approach to Old-to-Modern Kannada poetic translation, it addresses a major gap in prior research. The study also outlines future improvements in semantic handling and data expansion to further enhance translation accuracy and accessibility.

B. Kannada to English Machine Translation using LSTM and GRU

This paper looks at a Kannada-to-English translation system built with LSTM and GRU Seq2Seq models. GRU clearly comes out ahead here, pulling in higher BLEU scores and handling Kannada's tricky grammar better than LSTM. Kannada hasn't gotten as much research love as other Indian languages, so techniques like morphological segmentation and byte-pair encoding turn out to be especially helpful for translating Dravidian languages like this. Honestly, adding attention mechanisms and using richer regional datasets should take translation quality to the next level.

C. Kannada to English Machine Translation using LSTM and GRU

This paper compares LSTM and GRU Seq2Seq models for Kannada-to-English translation and finds GRU performs better, with higher BLEU scores and smoother handling of Kannada's complex grammar. It highlights how techniques like morphological segmentation and byte-pair encoding are especially useful for Dravidian languages, which haven't been studied as deeply. The work also notes challenges with idioms and context, suggesting that attention mechanisms and richer regional datasets could significantly improve future translations.

D. Language Translation Technology Based On Mobile Internet

Mobile internet and the need for quick communication have pushed translation technology to evolve rapidly, moving from manual methods to fast and accurate neural models. The study's big-data system performed impressively, translating over 95% of sentences correctly at about 8 sentences per second. While BLEU scores help measure performance, the research highlights that real quality comes from translations that stay fluent and true to meaning.

E. Unsupervised Neural Machine Translation For English To Kannada Using Pre- Trained Language Model

The paper introduces an unsupervised English-Kannada translation system that works with just monolingual data, powered by a pre-trained cross-lingual language model (XLM). By aligning both languages in a shared space and refining translations through iterative back-translation, the system achieves BLEU scores of 0.61 (EN→KA) and 0.32 (KA→EN). Experiments with IndicCorp and Samanantar show that transfer learning and adding more languages could further boost performance for low-resource translations.

F. Data Pre-Processing Framework For Kannada Vachana Sahitya

This paper introduces a preprocessing framework that cleans and standardizes the poetic, irregular text of Kannada Vachana Sahitya, achieving about 35% normalization. It shows that while many NLP systems rely on simple tokenization, morphology-aware methods like segmentation, semantic parsing, and subword models work far better for Kannada and similar Dravidian languages. By creating a bilingual corpus and a tailored pipeline, the study lays a strong foundation for future translation tools and helps preserve Karnataka's literary heritage.

III. EXISTING SYSTEM

Translating Halegannada currently depends almost entirely on scholars, making the process slow and labor-intensive. Classical works like Vachana Sahitya use old vocabulary, grammar, and context-sensitive expressions that modern NLP tools struggle to interpret. There are very few digital resources, parallel corpora, or annotated datasets available for Halegannada, which makes automation difficult. Most existing tools are built for modern Kannada and cannot handle archaic spellings, poetic structures, or historical word forms. These limitations make large-scale translation and digitization challenging, slowing down efforts to preserve and share classical Kannada literature with today's readers.

IV. PROPOSED SYSTEM

The Halegannada Translation Hub is designed to convert classical Kannada into English using both text and image inputs. It relies on a carefully curated Dictionary.pkl to map Halegannada words to Modern Kannada, which helps improve translation accuracy. Text inputs are preprocessed through tokenization, compound splitting, and encoding standardization, while images are enhanced, segmented, and processed through OCR to extract the text. A Seq2Seq machine learning framework—using LSTM, GRU, or Transformer layers—captures the context and generates English translations with attention mechanisms. The system combines local resources

with external APIs like Google Gemini OCR for reliable text recognition. Altogether, this pipeline provides an accurate, scalable, and user-friendly way to make classical Kannada literature accessible.

V. METHODOLOGY

A. Overall System Architecture

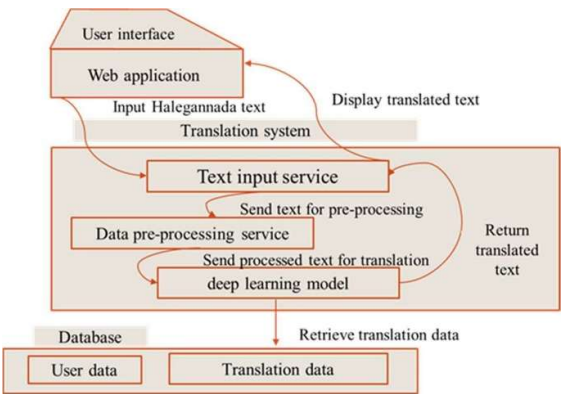


Fig 1. Proposed System Architecture

Users interact with the system through a web application, submitting Halegannada text or images for translation. The web app sends the input to the backend and displays the translated text in a user-friendly format. For images, Gemini OCR extracts the text before processing. The translation system includes modular services: a Text Input Service formats the input, and a Data Pre-processing Service cleans, tokenizes, normalizes, and encodes it for translation. A Deep Learning Model (Seq2Seq, Transformer, or LSTM/GRU) generates Modern Kannada or English translations. Translated text is returned to the UI, while a database stores user data, input history, and translation references. The model can leverage this data to improve accuracy and efficiency.

VI. RESULTS

This translation hub makes it easy to connect ancient language with the modern world. It works through a simple, intelligent process each time you upload something. When a historical script image is added, the system steps in like a digital scribe, carefully reading the old text using powerful optical recognition. After the Halegannada text is pulled from the image, the words are gently updated into their modern Kannada forms using a carefully built internal glossary. If a word isn't found in this glossary, it's kept as it is so nothing is lost. The modernized text is then passed on to a highly capable external language model, which creates the final English translation. What you get is a smooth, accurate, and easy-to-read translation delivered straight to the interface. The best part is that the whole process works without any local machine-learning training or complicated dataset preparation. Instead, it uses strong, existing tools for text recognition and language understanding. Altogether, this design offers a fast and reliable way to explore and appreciate Old Kannada literature.

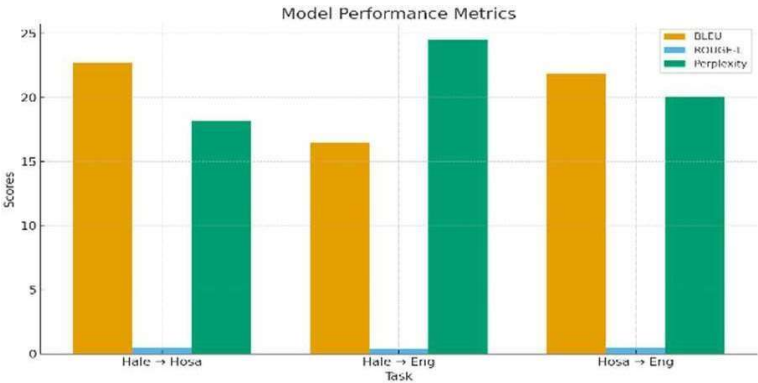


Fig 2: Results

VII. CONCLUSION

The Halegannada to English translation project brings ancient Kannada to life using modern technology. By combining a carefully curated dictionary with smart AI APIs, it ensures the original meaning is preserved. Its user-friendly web interface makes it easy to process both text and images, with OCR tools reading old scripts and updating them to modern Kannada. The DeepSeek API then creates clear, meaningful English translations, opening classical literature to students, researchers, and history enthusiasts. By blending culture with AI, the system delivers fast, accurate results, and future features like voice translation and offline use promise to make it even more accessible, connecting the wisdom of the past with today's digital world.

ACKNOWLEDGEMENT

We would like to express our sincere gratitude to Dr. P V Narayan, a poet and scholar, for his invaluable contributions to the study of Halegannada, particularly in the field of Pampa literature. His extensive work and insights have been instrumental in shaping the direction of this research.

REFERENCES

- [1] S. K. Sheshadri, B. S. Bharath, A. H. N. S. C. Sarvani, P. R. V. B. Reddy, and D. Gupta, "Unsupervised Neural Machine Translation for English to Kannada Using Pre Trained Language Model," in Proc. 13th Int. Conf. on Computing Communication (ICCCNT), 2022, pp. 16, doi:10.1109/ICCCNT54827.2022.99 84521.
- [2] Sheshadri, S. K., Gupta, D., & Costa-Jussà, M. R. (2023). "A voyage on neural machine translation for Indic languages". *Procedia Computer Science*, 218, 2694– 2712.
- [3] B. Zahoor, L. Santhosh, A. K. N., V. Potdar, R. B., and R. D. K., "Rule-Based Lexical Modelling for Translating Proverbs in Kashmiri and Kannada to English: Corpus Creation and Analysis," *Journal of Information Systems Engineering and Management*, vol. 10, no. 38s, 2025.
- [4] Ravish, G. K., Sowjanya, M. N., Bhaskar, K. S., Rohan, G. K., & Hittalamani, V. (2024). Novel translation system. Intelligent Kannada inscription character recognition using OCR and machine learning. *International Research Journal*
- [5] Upadhye, G. D., Phatangare, S. A., Barhate, M., Anekar, D. R., Jagtap, S., Dhainje, A., & Ghuge, (2025). Multilingual language translator using ML. *Advances in Nonlinear Variational Inequalities*, 28(2s), 37–46.
- [6] P. K. N., 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. doi: 10.18280/isi.260113.
- [7] Shivakumar, B., Asha, K. R., & Kavyashree, N. (2024). A survey on intelligent Kannada inscription character recognition using OCR and machine learning. *International Research Journal of Innovations in Engineering and Technology (IRJIET)*, 8(7), 154–158.
- [8] Vijay Kumar, M. S., Bhavana, M., Neha, L., Prerana, B., & Thanushree, G. P. (2024). Halegannada to Hosa Kannada Translator. *International Research Journal of Modernization in Engineering Technology and Science*, 6(4). <https://doi.org/10.56726/IRJMET54873>.