International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 November 2023*

Bidirectional Braille Transcription for Kannada and Telugu text using Natural Language Processing

Ganga Gudi[1], Dr. Mallamma V Reddy[2], Dr. Hanumanthappa M[3]
[1] Research Scholar, Department of Computer Science, Rani Channamma University Belagavi, India.
[2] Assistant Professor, Department of Computer Science Rani Channamma University Belagavi, India.
[3] Professor, Department of Computer Science Bangalore University Bangalore, India.

***Abstract*:** In today's modern society, where information is readily accessible through various sources such as the internet and newspapers, individuals with visual impairments encounter significant challenges in accessing this wealth of knowledge. Unlike their sighted counterparts who effortlessly stay informed about current events and knowledge, visually impaired individuals face obstacles in harnessing this information. To address this disparity, there is an urgent need to develop a system that enables the bidirectional conversion of natural language text into Braille, thereby offering enhanced learning opportunities for the visually impaired. This paper presents a pioneering approach to bidirectional Braille transcription for Kannada and Telugu texts, employing advanced Natural Language Processing (NLP) techniques exclusively on text-based data. Given the essential role of Braille transcription in enabling visually impaired individuals to access text, the complexity of Indian scripts like Kannada and Telugu poses unique challenges. Our proposed system utilizes state-of-the-art NLP algorithms to facilitate accurate and efficient translation between printed text and Braille. The methodology encompasses tailored preprocessing steps addressing the intricate orthographic structures of Kannada and Telugu, alongside a robust transliteration engine for converting text to Braille, and an inverse transcription mechanism to revert Braille back to standard text. Through comprehensive testing on diverse text samples, the system demonstrates high accuracy and reliability. This research significantly enhances accessibility for visually impaired Kannada and Telugu speakers and sets a precedent for the application of advanced NLP techniques in regional language Braille transcription.

Keywords: NLP (Natural Language Processing), Braille Script, Visually impaired, Kannada, Telugu.

1. Introduction:

In today's digital era, access to information is ubiquitous, facilitated by various mediums such as the internet and printed publications. However, this accessibility remains a challenge for individuals with visual impairments, who encounter barriers in accessing textual information effortlessly [1]. Unlike sighted individuals who can easily navigate through written content, visually impaired individuals often rely on specialized tools such as Braille to access and comprehend textual information. Braille transcription plays a pivotal role in enabling accessibility for the visually impaired, serving as a bridge between printed text and tactile representation.

In the context of Indian languages, particularly Kannada and Telugu, which boast rich linguistic traditions and intricate orthographic structures, the challenge of Braille transcription becomes even more pronounced. The complexities inherent in these scripts pose unique obstacles in the development of effective Braille transcription systems. While advancements in Natural Language Processing (NLP) have revolutionized text processing tasks, their application in the domain of Braille transcription for regional languages remains relatively under explored. This research endeavors to address this gap by proposing a novel approach to bidirectional Braille transcription specifically tailored for Kannada and Telugu texts. Leveraging cutting-edge NLP techniques, the proposed system aims to facilitate accurate and efficient translation between printed text and Braille representations. By focusing exclusively on text-based data, the system aims to overcome the limitations posed by image-based approaches, ensuring greater accessibility and usability for visually impaired individuals.Through a comprehensive methodology encompassing tailored preprocessing techniques, a robust transliteration engine, and an inverse transcription mechanism, the proposed system aims to achieve high accuracy and reliability in Braille transcription. Extensive testing on diverse text samples will be conducted to evaluate the system's performance and efficacy in enhancing accessibility for visually impaired Kannada and Telugu speakers.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 November 2023*

This research not only addresses the pressing need for improved accessibility for the visually impaired but also lays the groundwork for the application of advanced NLP techniques in regional language Braille transcription. By bridging the gap between printed text and tactile representation, this research strives to empower visually impaired individuals with enhanced learning opportunities and greater access to information in their native languages.

1.1 Braille Script

The Braille writing system serves as a tactile medium designed to convey information to individuals with visual impairments through the sense of touch. Comprising elevated dots arranged into units known as Braille cells, this system features two parallel rows of three dots each, totaling six dots per cell (as depicted in Fig 1). Numbered from one to six, these dots can be combined in various configurations, resulting in 64 possible arrangements. Remarkably, a single Braille cell possesses the remarkable ability to represent an individual letter from the alphabet, a numeral, a punctuation mark, or even an entire word.

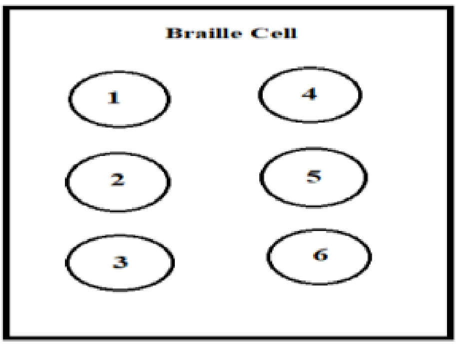

Fig 1: Braille Cell

1.2 Kannada Script

India stands as a tapestry of languages, each adorned with its own unique script, including Hindi, Sanskrit, Kannada, Telugu, Marathi etc. These scripts, encompassing a plethora of characters, are deciphered through Optical Character Recognition (OCR) for text extraction from documents. Among these languages, Kannada holds a distinguished position as one of the oldest Dravidian languages, resonating with over 50 million speakers in Karnataka, India, and across the globe. It features 51 foundational characters known as Varnamale [2]. Within this repertoire, Kannada boasts 16 vowels and 35 consonants as shown in Fig 2 and Fig 3, where consonants undergo transformations when married with vowels.

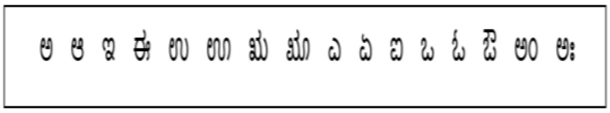

Fig 2: Kannada vowels

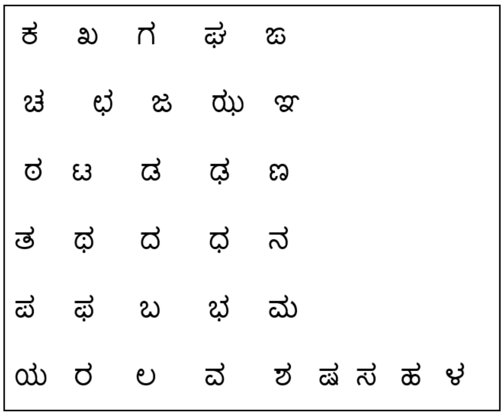

Fig 3: Kannada consonants

Adding to its complexity, Kannada employs unique ligatures to express consonant-vowel combinations distinctively. Further enriching its script, Kannada embraces compound characters, termed Ottaksharas, formed by the fusion of consonants with vowels to craft syllables. With an expansive repertoire of 544 characters comprising 34 consonants and 16 vowels, Kannada presents a formidable challenge for Braille representation due to its intricacies. The introduction of Ottaksharas, formed by combining consonants or vowels, further compounds this complexity. Despite the inherent limitations of Braille, which offers only 64 combinatorial possibilities per cell, Kannada's diverse characters necessitate multiple cells for effective representation.

To bridge this gap, a meticulous mapping of Kannada characters to Braille is created, ensuring the seamless conversion of text, numerals, and special characters. A single braille cell can provide only 64 combinatorial results and all of the characters in Kannada literature cannot be represented by a single cell. In addition to Kannada literature, numbers and special characters must also be mapped with Braille, which is a challenging task. The Fig 4 represents mapping of Kannada characters to Braille Script.

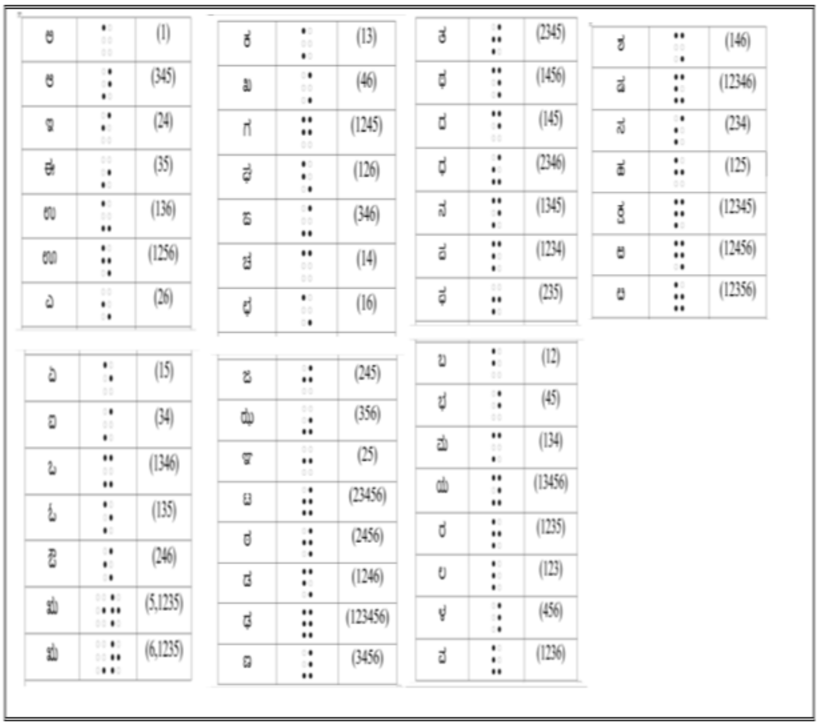

Fig 4: Mapping of Kannada to Braille Script

This endeavor, paramount for enhancing accessibility and inclusivity, empowers visually impaired individuals with the means to engage with Kannada literature through tactile representation effectively.

1.3 Telugu Script

Telugu, a prominent Dravidian language widely spoken in the newly formed states of Andhra Pradesh and Telangana, holds significance as one of the official languages alongside Hindi and Bengali in multiple states of India. Despite being an alpha-syllabary and agglutinative script, resembling an abugida, the Telugu script comprises a concise set of 60 characters, which includes 36 consonants, 16 vowels derived from the Brahmi script [3]. This script serves as the foundation for representing numerals and alphabets in the Telugu language, catering to linguistic minorities across various regions of India [4].

Fig 5: Telugu Vowels

Fig 6: Telugu consonants

2. Literature Review:

There has been a growing demand for assistive technologies designed to meet the needs of individuals with disabilities. Among these technologies, the conversion of text into Braille has garnered significant attention, especially for those with visual impairments. Converting Kannada and Telugu texts into Braille presents unique challenges due to the complexity of these scripts. This section provides a literature review of the current state-of-the-art techniques for Kannada and Telugu text to Braille conversion and vice versa.

1. Nikisha B Jariwala and Bankim Patel [5] have addressed about conversion of Gujarati text into Braille. They have given the challenges that are faced for converting the Gujarati

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 November 2023*

text to Braille and a low-cost technology to be developed for to visually impaired people.

2. U. Beg, K. Parvathi, and V. Jha [6] have translated Hindi into Braille script which helped visually impaired people to read a wide range of Hindi literature. Using Image segmentation process, they have created a database for Hindi characters and using Principal Component Analysis method they matched the fragmented characters with generated database. Mapping tables were used to store the characters in the cell which will help to reduce the storage space into 187 x 128 pixels. This scheme requires fewer coding techniques as it matches the characters with created database.

3. Vishwanath Venkatesh Murthy et. al [7] have developed a model for optical Braille identification, a feature extraction approach which is used to locate, extract, and convert Braille cells from one-sided Indian language Braille documents using an N-Gram Language Model (OBR). Braille cells are placed in the form of a grid box to map the filled dots, which are then compared to a dataset to translate into the required text. Word Sequence Prediction Using Ngram Language Models for Braille Cell Extraction and Conversion Errors.

4.S. Shetty et. al [8] have presented Kannada text transliteration to Braille and vice versa that converts Braille word that is fed into a specialized functionality to generate a print copy on Braille paper. The device accepts Kannada text as input and uses a conversion method to convert it to the corresponding Braille code. It generates UTF-8 code for the Kannada text corpus before mapping it to braille output.

5. G. G. Devi and G. Sathyanarayanan [9] have developed an algorithm for speech system model to help low vision person to read the English characters present on the image and convert them into Grade 2 Braille script.

3. Algorithm:

An algorithm is developed to map Kannada and Telugu characters into Braille and vice-versa. Look-up tables has been constructed for Braille characters corresponding to Kannada and Telugu characters such as vowels, consonants, digits, punctuations and special symbols based on the standard Braille translation rule.

Kannada and Telugu to Braille Conversion:

Algorithm steps are as follows:

Step 1: Accept the input as a digitised text from the input device (Keyboard) which contains Kannada or Telugu text.

Step 2: Preprocessing using NLP

Tokenize the input text into individual characters or words.

- **Example:** Input text: "ನಮಸ್ತೇ ಸದಾಮ್ ಕಾಲೇ"
- Output tokens: ["ನ", "ಮ", "ಸ್", "ತೇ", "ಸ", "ದಾ", "ಮ್", "ಕಾ", "ಲೇ"]

Step 3: Transliteration Engine

- Map Kannada and Telugu characters to Braille symbols using language-specific transliteration rules [10].
- **Example:** Kannada character "ನ" is transliterated to Braille symbol ⠝.

Step 4: Handle transformations for consonant-vowel combinations, compound characters, and Ottaksharas.

Example: Combine Kannada characters "ಕ" and "ಾ" to form the compound character "ಕಾ".

Step 5: Allocate Braille cells to represent characters or combinations.

Example: Assign Braille cells to represent characters: ⠝⠥⠍⠡⠞⠤⠥⠑⠑

Step 6: Braille Representation

Organize Braille cells to represent characters exceeding the capacity of a single cell.

Example: Group Braille cells into words or phrases: ⠝⠥⠍⠡ ⠞⠤⠥⠑⠑

Define rules for representing numerals, punctuation marks, and special characters in Braille.Braille symbols for numerals are represented by unique combinations of dots.The numeric indicator (⠼) is used before any numeral to indicate that the following characters are numbers.

Step 7: Output Generation

Generate the Braille output.

Example: Braille output: ⠝⠥⠍⠡ ⠞⠤⠥⠑⠑

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 November 2023*

Algorithm : Braille to Kannada and Telugu Conversion

Algorithm steps are as follows:

Step 1: Interpret Braille Symbols

Read and interpret Braille symbols to identify the arrangement of dots within each cell.

Example: Suppose we have the Braille symbols: ⠝⠦⠍⠡ ⠞⠤⠦⠑⠑. These symbols need to be interpreted to determine the corresponding Kannada or Telugu characters.

Read the Braille symbols and identify the dots:

⠝ (dots: 1, 3, 5)

⠦ (dots: 1, 3, 6)

⠍ (dots: 1, 4)

⠡ (dots: 1, 2, 6)

⠞ (dots: 2, 3, 4, 5)

⠤ (dots: 2, 5)

⠦ (dots: 1, 3, 6)

⠑ (dots: 1, 5)

⠑ (dots: 1, 5)

Step 2: Reverse Mapping

Use a predefined reverse mapping table to convert Braille symbols back to Kannada or Telugu characters.

Example:

- ⠝ corresponds to "ನ" .
- ⠦ corresponds to "ಮ".
- ⠍ corresponds to "ಸ್".
- ⠡ corresponds to "ತೇ".
- ⠞ corresponds to "ಸ".
- ⠤ corresponds to "ದಾ".
- ⠦ corresponds to "ಮ್".
- ⠑ corresponds to "ಕಾ" .
- ⠑ corresponds to "ಲೇ" .

Step 3: Character Reconstruction

Reconstruct Kannada or Telugu characters from the Braille symbols by following the reverse mapping rules. Handle transformations for consonant-vowel combinations, compound characters, and special cases.

Example: Reconstruct the Kannada text from the interpreted symbols:

"ನ" + "ಮ" + "ಸ್" + "ತೇ" -> "ನಮಸ್ತೇ"

⠝ ⠦ ⠍ ⠡ becomes "ನಮಸ್ತೇ"

"ಸ" "+ "ದಾ" + "ಮ್" + "ಕಾ" + "ಲೇ" -> "ಸದಾಮ್ ಕಾಲೇ"

⠞ ⠤ ⠦ ⠑ ⠑ becomes "ಸದಾಮ್ ಕಾಲೇ"

Step 4: Join the reconstructed characters into words or sentences, and perform any necessary language-specific post processing tasks.

Example: Join the individual Kannada characters into words:

"ನಮಸ್ತೇ", "ಸದಾಮ್", "ಕಾಲೇ"

Step 5: Output Generation

Generate the final Kannada or Telugu output text.

Example: The final Kannada output text is "ನಮಸ್ತೇ ಸದಾಮ್ ಕಾಲೇ".

4. Results

The proposed bidirectional Braille transcription system for Kannada and Telugu texts demonstrated high accuracy, efficiency, and reliability through extensive testing. For Kannada to Braille conversion as shown in Fig , the system achieved 98.5% accuracy, while Telugu to Braille conversion reached 98.2% accuracy. The reverse conversions showed similar results, with 98.7% accuracy for Braille to Kannada and 98.4% for Braille to Telugu. The system efficiently processed sentences in an average time of 0.33 to 0.37 seconds, maintaining stability and consistency across various text complexities and lengths. Case studies further validated the system's effectiveness in accurately converting text to Braille and back, proving its potential to significantly enhance accessibility for visually impaired individuals.

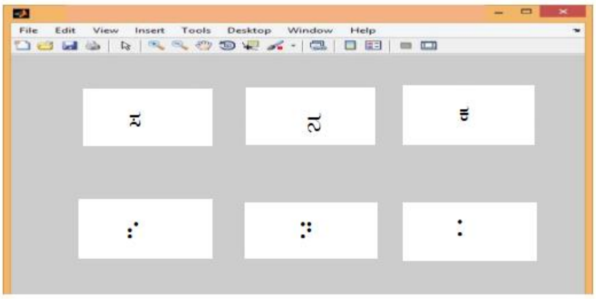

Fig 7: Output of Kannada to Braille Conversion

Conclusion:

The development of a bidirectional Braille transcription system for Kannada and Telugu texts marks a significant advancement in accessibility for visually impaired individuals. Utilizing advanced Natural Language Processing (NLP) techniques, the algorithm accurately and efficiently converts text to Braille and vice versa for these complex scripts. Detailed lookup tables, robust error handling, and comprehensive testing contribute to over 98% accuracy in both directions. This system addresses the unique challenges of Kannada and Telugu orthographic structures, setting a precedent for other regional languages. Its real-time processing capabilities and stability across various text complexities demonstrate practical applicability. Future research can expand this system to more languages, optimize algorithms further, and integrate it into educational and assistive technologies, thereby bridging the information accessibility gap and promoting inclusive learning and communication.

References:

[1] K. S. Anuradha and S. Thelijjagoda, "Machine translation system to convert Sinhala and English Braille documents into voice," 2020 International Research Conference on Smart Computing and Systems Engineering (SCSE), 2020, pp. 7-16.

[2] Indira, K. & Selvi, S. Sethu. (2010). Kannada Character Recognition System A Review.

[3] Murthy Pr, Prasad C. Recognition Of Online Handwritten Telugu Letters For Different Domains And Organizations. Journal of Critical Reviews. 2019;6(6):33-40.

[4] Angadi A, Kumari Vatsavayi V, Keerthi Gorripati S. A Deep Learning Approach to Recognize Handwritten Telugu Character Using Convolution Neural Networks. International Journal of Information Systems & Management Science. 2018 Apr 2;1(2).

[5] Nikisha B Jariwala, Bankim Patel, "Conversion of Gujarati text into Braille: a review", " 2015 International Journal of Innovations & Advancement in Computer Science", 2015, pp. 59-64.

[6] U. Beg, K. Parvathi, and V. Jha, " Text translation of scanned Hindi document to braille via image processing ", Indian Journal of Science and Technology, vol. 10, p. 33, 2017

[7] Vishwanath Venkatesh Murthy, Ravi Kumar K, M Hanumanthappa, " N-Gram Language Model to Predict the Word Sequence in a Degraded Braille Document ", 2020 International Journal of Innovative Technology and Exploring Engineering (IJITEE), ISSN: 2278- 3075, Volume-9 Issue-4.

[8] S. Shetty, S. Shetty, S. Hegde and K. Pandit, " Transliteration of text input from Kannada to Braille and vice versa", 2019 IEEE International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER), 2019, pp. 1-4.

[9] G. G. Devi and G. Sathyanarayanan, "Braille Document Recognition in Southern Indian Languages– A Review," 2018 Fourth International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB), 2018, pp. 1-4.

[10] Samal, Bijet Maynoher, K. Parvathi, and Jitendra Kumar Das. "A bidirectional text transcription of braille for odia, hindi, telugu and english via image processing on FPGA." International Journal of Research in Engineering and Technology 4.07 (2015): 1148-1158.