Tulu Text Recognition with Speech Based Output

[1] Rekha A A, [2] Shreya D K, [3] Shradha, [4] Prakash, [5]Avish V Shetty
[1] Assistant Professor, [2] Student, [3] Student, [4] Student, [5] Student
Dept. of ECE,
[1]A J Institute of Engineering & Technology, Mangaluru, India

[1] rekhaanil@ajiet.edu.in , [2]shreyadk2324@gmail.com, [3]shradhakundar123@gmail.com

[4]prakashhiregoudar42@gmail.com , [5]avishshetty21@gmail.com

Abstract— The Tulu language, spoken predominantly in the coastal region of Karnataka, India, possesses a rich cultural heritage and a unique script. Despite its historical significance, the Tulu script is underrepresented in the digital realm. This paper presents a Tulu language character recognition system designed to bridge this digital gap. The system processes images containing handwritten Tulu characters and outputs the corresponding text in English. The workflow begins with pre- processing the images to enhance quality and normalize variations. This is followed by image segmentation using region- based techniques to isolate individual characters. A Convolutional Neural Network (CNN) model, trained on a comprehensive dataset, is then employed for the classification and recognition of these characters. The system meticulously recognizes each character, ensuring best recognition efficiency achieved for Tulu characters from collected dataset. The results verified that the proposed methodology outperforms from the present state of art models. To further enhance the system's utility, the recognized text is mapped to pre-recorded audio clips corresponding to each character. These audio clips are then synthesized sequentially to generate a complete auditory output, providing a spoken version of the recognized character. This feature not only enriches the accessibility of the Tulu language in digital formats but also promotes its preservation and dissemination by catering to auditory learners and visually impaired users.

Index Terms— *Convolutional Neural Network (CNN), Speech synthesis, Audio-text mapping*

I. INTRODUCTION

Tulu language text recognition is an emerging technology aimed at digitizing handwritten and printed Tulu text for preservation and accessibility. Despite being Karnataka's second language and efforts to include it in the 8th Schedule of the Constitution, it remains underrepresented in the digital space. Text recognition technologies help bridge this gap by converting Tulu documents into digital formats for easier storage, retrieval, and sharing.

The Tulu script consists of 50 characters and 12 numerals, historically linked to Kannada. Since Tulu speakers often use Kannada or Roman script for official communication, developing a dedicated recognition system is crucial. The process involves collecting handwritten images, pre- processing them to remove noise, and converting them into binary images. Segmentation techniques then detect and outline character shapes using bounding boxes. A Convolutional Neural Network (CNN) extracts key features through Conv2D layers with ReLU activation, followed by MaxPooling layers to retain essential features. The Flatten and Dense layers further refine character classification, while Dropout layers prevent overfitting. K-Fold cross-validation enhances recognition accuracy by training the model across multiple subsets of data.

Beyond text recognition, an audio-mapping feature further enriches the system by linking recognized text to pre-recorded Tulu audio clips. This allows users to listen to the text, preserving proper pronunciation and improving accessibility for visually impaired individuals and language learners. This integration of text and speech helps maintain Tulu's linguistic heritage while making it more accessible in the digital age. By combining AI-driven text recognition with audio output, this initiative ensures Tulu's continued relevance and promotes its adaptation into modern technology, strengthening its presence in both textual and auditory formats.

II. LITERATURE SURVAY

This provides a brief overview of what research has been done before by reviewing the numerous papers and setting the stage for the current study. C K Savitha, P J Antony[1] titled " PNN and Deep Learning Based Character Recognition System for Tulu Manuscripts" This research introduces a novel technique for recognizing handwritten Tulu characters, numerals, and palm leaf manuscript texts. It employs noise removal, adaptive thresholding, and segmentation, comparing Probabilistic Neural Network (PNN) and Deep CNN. The Deep CNN achieves superior accuracy and efficiency, pioneering Tulu numeral recognition for digital preservation. Midhun Varghese, Suraksha, Abhinav Vinod, Suhail Abdul Nazir, Dr.Anoop B K[2]titled " A Comprehensive Review on BARAVU-Tulu Lipi Identification" The Tulu Lipi identification project employs CNNs in PyTorch for automated Tulu character recognition. Data preprocessing, augmentation, and loss monitoring enhance model performance. Evaluation confirms robustness, while visualizations validate accuracy. Prathwini, Anisha P Rodrigues, P.Vijaya, Roshan Fernandes[3]titled " Tulu Language Text Recognition and Translation" A handwritten character recognition system was developed using CNN, achieving 92% accuracy for Tulu characters and numerals. An algorithm using a rule-based method was incorporated into the research work for English to-Tulu translation, achieving 89% accuracy for simple words and sentences. Neural machine technology was applied to increase efficiency and achieved a blue score of 0.83. Prof. Saurabh Saoji, Anshul Arora, Rajat Singh, Ankit Mangal, Ashiq Eqbal[4]titled "E Extraction And Detection Of Text From Images"This paper explores text recognition from images, emphasizing key preprocessing steps like noise removal, normalization, and binarization for enhanced OCR accuracy. It details segmentation, feature extraction, and classification using ANN and pattern matching. Post-processing ensures proper text formatting. Covering texture- and region-based detection methods, it presents a complete framework for efficient image-based text recognition. Banoth R , Dhir R[5]titled " Identification of Telugu script in a bilingual document image", This study explores Telugu-English script identification in bilingual documents, addressing the need for accurate OCR in India's multilingual landscape. Sachin Bhat, G. Seshikala[6]titled "Character recognition of Tulu script using convolutional neural network", this study focuses on Tulu handwriting recognition using Deep Convolutional Neural Networks (DCNNs), achieving 92.41% accuracy on a dataset of 90,000 characters. Preprocessing techniques like noise reduction and segmentation enhance recognition. Additionally, integrating speech synthesis converts recognized text into audio, aiding language learners and visually impaired users. This research advances Tulu language preservation, promoting digital accessibility and reducing manual efforts in digitizing handwritten texts. Mahadeva Prasad Y. N, Chethan H. K[7] titled " Recognition of Printed Kannada Text in Scene Images using Machine Learning Techniques", This study enhances Kannada scene text recognition using a four-stage process: preprocessing, detection, feature extraction, and recognition. A YOLO V7 model detects Kannada text, while a Dense Stacked LSTM extracts features. A Convolutional Residual Network-assisted autoencoder ensures accurate recognition. This approach surpasses existing methods, improving text detection in complex images, making it valuable for real-world applications in electronic and communication media. Ayush Purohit, Shardul Singh Chauhan[8] titled " A Literature Survey on Handwritten Character Recognition", Handwriting recognition is a key focus in pattern recognition and machine learning, enhancing Optical Character Recognition (OCR) and Handwritten Character Recognition (HCR). This paper reviews methodologies, improvements, and challenges in HCR, emphasizing its role in creating a paperless environment. As technology evolves, handwriting recognition will revolutionize document management and digital transformation, making physical archives more accessible and efficient. Sanket Gandhare, P. Jyothi, P. Bhattacharyya[9] titled " Literature Survey : Spoken Language Translation" Spoken language translation systems combine Automatic Speech Recognition (ASR), Machine Translation (MT), and Machine Learning (ML) to convert speech from one language into text in another. This paper reviews the workflow, focusing on ASR techniques, neural machine translation, and integration. It explores text processing, hypothesis selection, feature extraction, and advanced decoding strategies like lattice and confusion network decoding. These methods enhance translation accuracy by resolving ASR output ambiguities, addressing challenges in system integration, and optimizing performance for seamless spoken language translation. Parikshith Honnegowda, S M Naga Rajath, D Shwetha, C M Sindhu[10]titled " Handwritten Character Recognition of Kannada Language Using Convolutional Neural Networks and Transfer Learning", this study proposes a novel approach for recognizing Kannada handwritten characters using Convolutional Neural Networks (CNNs) and transfer learning. After pre-processing steps like noise removal and image resizing, the model efficiently classifies Kannada vowels, consonants, and numerals, even with structural similarities between characters. Transfer learning enhances adaptability, enabling accurate recognition of new handwriting styles. The proposed method outperforms traditional techniques, achieving high accuracy and showcasing its potential for practical applications in handwritten Kannada character recognition.

III. METHODOLOGY

The methodology for this work consists of several key steps that ensure efficient recognition of handwritten and printed Tulu characters with speech-based output. The process involves structured stages, including image acquisition, preprocessing, character segmentation, classification using deep learning techniques, and speech- based output generation. Each step is critical in achieving high accuracy and reliability in character recognition.

A. Image Acquisition:

Image acquisition is the foundational step in building the Tulu text recognition system. This stage involves collecting a diverse set of handwritten and printed Tulu characters to train the deep learning model effectively. The dataset consists of 50 Tulu character classes and 12 numerical digits, resulting in a total of 10,000 images. These images were sourced from different individuals to capture a wide range of handwriting styles and variations. To maintain consistency, all images were resized to 100x100 pixels, ensuring that the deep learning model receives uniform input. The images were acquired using mobile phone cameras and scanners, allowing for high-quality digital representations. Standardizing the image dimensions reduces discrepancies caused by varying resolutions and aspect ratios, making the dataset suitable for robust training. Figure 1 shows the pre- processed images of 50 characters, displaying how the dataset is structured with clearly defined classes.

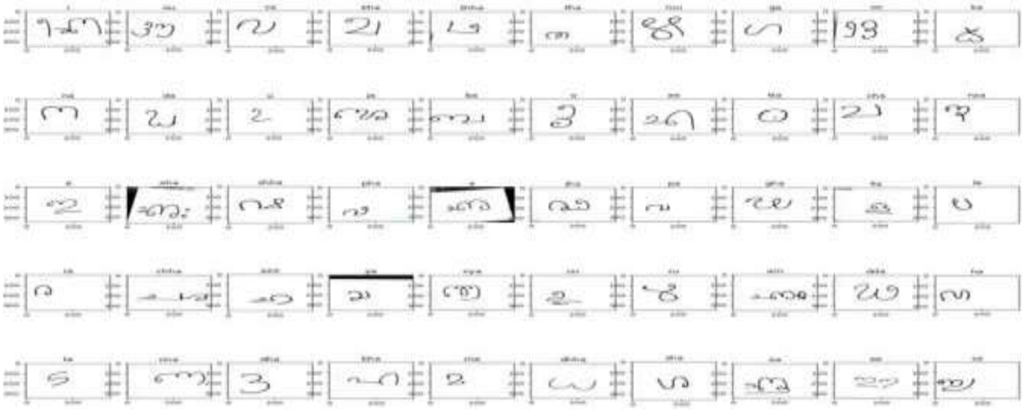

Fig 1. Pre-processed image of 50 characters of dataset with 50 classes

B. Image Preprocessing:

Image preprocessing is a crucial step that enhances the quality of input images and prepares them for feature extraction. Since the dataset contains handwritten characters with potential noise, distortions, and varying lighting conditions, preprocessing helps improve clarity and ensures accurate recognition.

1.Load images from the dataset using OpenCV and resize them to a uniform size of (100,100).
2Apply affine transformations, including scaling, rotation, and shearing, to introduce geometric variations.
3Apply Gaussian blur and additive Gaussian noise to simulatereal-world distortions.
4Adjust brightness using pixel multiplication to enhance contrast variations.
5.Apply horizontal flipping to increase dataset diversity.
6.Convert image labels into numerical values and apply one- hot encoding for classification.
7.Split the dataset into training and testing sets using the `train_test_split` method to ensure balanced data distribution.

Figure 2 shows a pre-processed Tulu character, highlighting the enhanced clarity and uniformity of the images after applying this technique.

Fig 2. Pre-processed image of a Tulu character 'ae'

C. Character Segmentation:

Character segmentation is the process of isolating individual Tulu characters from an image. This step is crucial because handwritten text may have overlapping characters or varying spacing, making direct recognition challenging. The segmentation process begins with edge detection using the Canny algorithm. This method highlights the boundaries of the characters by detecting sharp changes in pixel intensity. Once the edges are detected, morphological operations such as dilation and opening are applied to refine the detected edges and remove small noise components.

Following edge enhancement, bounding box extraction is performed to enclose each character within a distinct rectangular region. This ensures that only the relevant character portions are processed, reducing computational complexity and improving accuracy. Figure 3 displays the segmented images of 50 characters from the dataset, demonstrating the effectiveness of the segmentation process.

Fig 3. Segmented Images of 50 Characters of dataset with 50 Classes

D. Classification Using Convolutional Neural Networks (CNNs):

Once the characters are segmented, they are passed through a Convolutional Neural Network (CNN) for classification. CNNs are highly effective in image recognition tasks due to their ability to automatically learn spatial hierarchies of features, making them ideal for handwritten character recognition.

1.Construct a Convolutional Neural Network (CNN) model using TensorFlow/Keras for character recognition.
2.Add multiple convolutional layers (Conv2D) with activation functions to extract spatial features from input images.
3.Apply max pooling layers (MaxPooling2D) to reduce the dimensionality and retain essential features.
4.Use dropout layers to prevent overfitting and improve generalization.
5.Flatten the extracted features into a one-dimensional array for classification.
6.Add fully connected (dense) layers with activation functions to learn complex patterns.
7.Use the softmax activation function in the output layer for multi-class classification.
8.Compile the model with an appropriate optimizer, loss function, and evaluation metrics.
9.Train the model using the prepared dataset and validate performance using test data
10.Evaluate the trained model on unseen data to measure accuracy and generalization

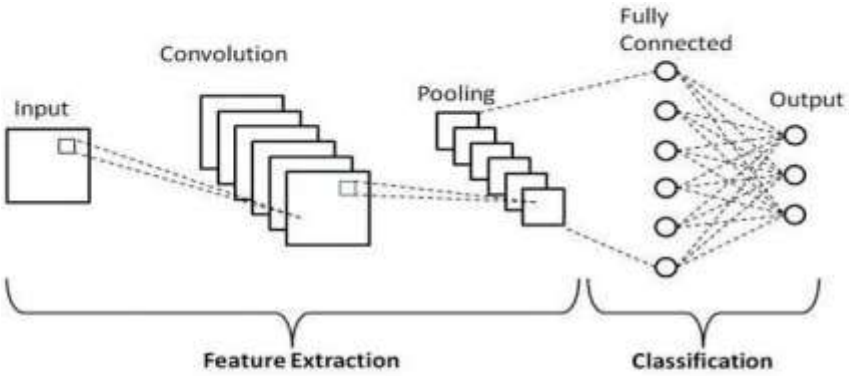

Fig 4. Feature Extraction and classification of images using CNN

To optimize the training process, the Adam optimizer with a learning rate of 0.001 was used, along with categorical cross-entropy loss. The model was trained using K-Fold cross- validation (K=3) to ensure that it generalizes well across different data subsets. Early stopping was also implemented to prevent overfitting, halting training if the validation loss did not improve for three consecutive epochs. The final model was evaluated on a test dataset, achieving an accuracy of 90.72%, demonstrating its robustness in handling different handwriting styles.

E. Speech-Based Output Generation

The final step in the methodology involves converting the recognized text into speech to enhance accessibility, particularly for visually impaired users and language learners. Each recognized character is mapped to a pre- recorded audio clip stored in a speech database. When a character is identified, the system retrieves the corresponding audio file and plays it, allowing users to hear the spoken version of the text.

To ensure high-quality speech output, a text-to-speech (TTS) module was integrated into the system. The module takes the recognized text as input and generates natural- sounding speech output, maintaining proper pronunciation and linguistic accuracy. This feature enhances user interaction by providing both visual and auditory feedback, making the system more inclusive.

The integration of speech output not only supports individuals with visual impairments but also serves as an educational tool for those learning the Tulu language. By enabling users to listen to the pronunciation of characters, the system bridges the gap between textual recognition and spoken communication, reinforcing language preservation efforts.

IV. RESULT AND DISCUSSIONS

The Tulu character recognition model was successfully implemented and trained using a dataset consisting of 4,979 images spanning 50 distinct character classes. The development and implementation of the model involved several critical steps, including dataset preparation, image preprocessing, image segmentation, data augmentation, model training, and evaluation. This section presents the experimental results and performance analysis of the system in recognizing handwritten Tulu characters.

A. Dataset and Preprocessing

The dataset used in this study comprises grayscale images of handwritten Tulu characters, with each character class representing a different Tulu symbol. To improve the model's robustness, several preprocessing techniques were applied. First, the images were resized to a consistent dimension of 100x100 pixels to standardize the input data. Grayscale conversion was performed to simplify the images by eliminating color information, which reduces computational complexity. To enhance the model's ability to recognize diverse handwriting styles, data augmentation techniques were employed. This included random rotations, zooming, brightness adjustments, and horizontal flips, which introduced additional variations in character appearances. Additionally, pixel normalization was applied to scale all pixel values to a range of [0,1], improving computational efficiency and training stability. The dataset was then split into training (80%) and testing (20%) sets, ensuring that the model was trained on a substantial portion while maintaining a separate set for evaluation.

Figure 5 shows an example of a raw 'ae' character from the dataset, while Figure 6 illustrates the same character after preprocessing and segmentation, ensuring that only relevant features are retained for recognition.

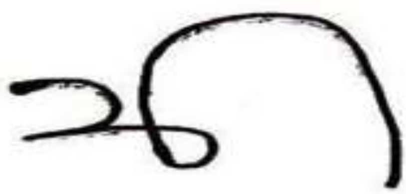

Fig 5. Character input from Dataset

Fig 6. The Preprocessed and Segmented image

B. Model Architecture and Training

The Convolutional Neural Network (CNN) model was employed to recognize Tulu characters due to its strong capability in processing image-based data. The architecture consisted of three convolutional blocks with ReLU activation, followed by max-pooling layers to reduce the spatial dimensions and capture essential features. The model also included a fully connected dense layer with 128 neurons and a dropout layer for regularization, helping to prevent overfitting. The final output layer used softmax activation, allowing the model to classify images into one of the 50-character classes.

To improve the model's generalization and reduce overfitting, data augmentation techniques, such as rotation, scaling, and flipping, were applied. This artificially increased the dataset size, enabling the model to learn from a more diverse set of images. The model was trained using K-fold cross-validation (K=3) to evaluate its performance more reliably, ensuring that the model was not biased toward a particular subset of the data. Early stopping was also implemented to prevent overfitting by halting training when the validation loss did not improve for three consecutive epochs. During training, the model was optimized using the Adam optimizer and categorical cross-entropy loss, with the goal of maximizing accuracy. The average cross-validation accuracy achieved across the folds was used to assess the overall performance, and the final model was evaluated on the test set. This approach ensured robust performance on unseen data, resulting in high accuracy for character recognition tasks. Figure 7 illustrates the classification and recognition process of the 'ae' character, demonstrating how the model extracts and interprets the character's features for accurate identification.

```
1/1 [==============================] - 0s 25ms/step
Prediction array: [[2.8676262e-05 2.2611633e-18 9.9997067e-01 1.5385133e-10 4.2372445e-16
  1.1155224e-15 1.6761115e-12 4.8108916e-07 7.6139986e-15 1.9241429e-16
  2.6601366e-19 9.7324590e-18 1.7001839e-22 2.5558016e-25 4.1486311e-11
  6.2906440e-14 3.7877202e-17 1.9463088e-19 3.5689717e-18 7.9183354e-22
  2.4628528e-12 1.5449134e-12 2.4873722e-20 6.5496204e-15 3.6108605e-19
  4.0466943e-16 1.3340464e-15 2.2709525e-11 2.6129391e-14 5.5046191e-13
  1.0841712e-07 6.2995284e-14 2.1506532e-14 2.3255948e-12 1.6494807e-12
  2.0542762e-20 8.9523039e-15 1.4859042e-16 6.1681834e-14 1.6957850e-08
  1.3020413e-21 7.8149601e-21 5.7242305e-19 1.8976743e-13 1.9789103e-29
  5.4976428e-09 5.3444078e-20 7.9335902e-18 2.1169083e-20 3.9821484e-12]]
Predicted class: 2
Predicted character: ae
```

Fig 7. Classification and Recognition of character 'ae' for the segmented image

C. *Performance Analysis*

The CNN model achieved 90.66% cross-validation accuracy and 90.72% test accuracy in recognizing handwritten Tulu characters, demonstrating strong performance. This success is due to preprocessing (grayscale conversion, noise removal, normalization), segmentation (isolating characters), and data augmentation (rotation, scaling, flipping) to enhance generalization. The architecture, with three convolutional blocks and dropout layers, effectively captured features while preventing overfitting. Early stopping ensured optimal training by halting when validation loss stabilized. The model consistently exceeded 90% validation accuracy, highlighting its robustness in handling character recognition challenges.

Figure 8 presents the speech output for a recognized character, while Figures 9 and 101 depict the model's accuracy and loss curves, showing how the model converged over multiple training epochs.

Fig 8. Speech Output

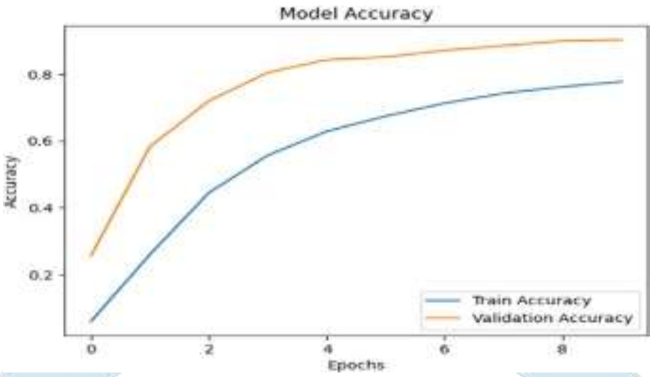

Figure 9. Model accuracy

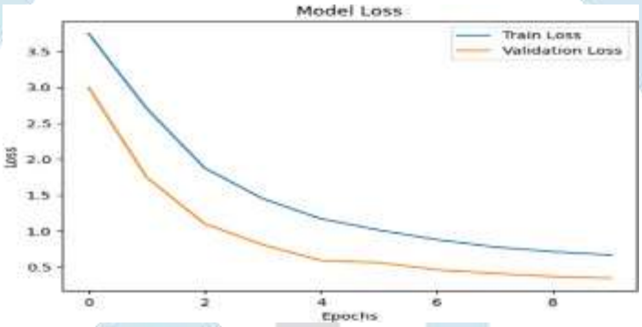

Figure 10. Model loss

D. *Speech Output and User Interaction*

The CNN model achieved 90.72% test accuracy in recognizing handwritten Tulu characters. Success stems from preprocessing, segmentation, and data augmentation to enhance generalization. Its three convolutional blocks with dropout layers captured features while preventing overfitting. Early stopping optimized training, ensuring robust performance.

Figure 11 presents an example of the recognized character "ka", while Figure 12 demonstrates the corresponding speech output. This integration of visual and auditory outputs makes the system more interactive and user-friendly, further contributing to the digital preservation of the Tulu language.

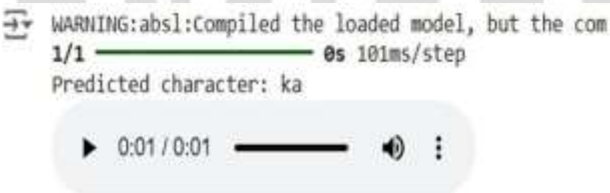

Fig 11. Character 'ka'

Fig 12. Recognized speech output

V. FUTURE WORK

Future work will focus on enhancing the system's capabilities by shifting from character-level to word-level recognition, improving usability, and expanding scalability. Implementing sequence-based models like RNNs, LSTMs, or Transformer architectures will enable the system to understand word structures and contextual relationships, enabling it to process entire sentences for applications like document digitization and real-time translation. Additionally, an interactive platform will be developed for real-time handwritten Tulu input, offering text and speech outputs with features like live handwriting recognition and feedback for model improvement. Multimodal input options, including speech-to-text conversion, will increase accessibility, and advanced neural TTS models will enhance speech synthesis quality.

VI. Conclusion

The Tulu character recognition project successfully implemented a deep Convolutional Neural Network (CNN) model, achieving high accuracy on a dataset of 4,979 images across 50 classes. This success highlights the potential of advanced deep learning techniques in preserving underrepresented languages like Tulu. The project digitizes handwritten and printed Tulu documents, facilitating efficient storage, retrieval, and sharing of linguistic content while contributing to the preservation of Tulu's cultural heritage.

The system encompasses image acquisition, preprocessing, segmentation, feature extraction, and classification, demonstrating its effectiveness in recognizing complex handwritten characters. Additionally, the recognized Tulu text is converted into speech, providing an auditory output that enhances accessibility for visually impaired users and supports language learning. This feature enables users to engage with the language through correct pronunciations, promoting inclusivity. The project represents a significant step towards linguistic preservation, ensuring the Tulu language thrives in the digital era.

References

[1] Savitha, C., Antony, P.J. (2019). PNN and Deep Learning Based Character Recognition System for Tulu Manuscripts 1855.

[2] Publication, Gjr. (2024). A Comprehensive Review on BARAVU-Tulu Lipi Identification. 4. 105-109. 10.5281/zenodo.10969139.

[3] Prathwini, A. P. Rodrigues, P. Vijaya and R. Fernandes, "Tulu Language Text Recognition and Translation," in IEEE Access, vol. 12, pp. 12734- 12744, 2024, doi: 10.1109/ACCESS.2024.3355470.

[4] Arora, Anshul & Vidyapeeth, Bharati & Singh, Rajat & Eqbal, Ashiq & Mangal, Ankit & Saoji, Saurabh. (2021). EXTRACTION AND DETECTION OF TEXT FROM IMAGES. International Journal of Research in Engineering and Technology. 8. 2395 0056.

[5] Banoth, R. Dhir, R. (2016). Identification of Telugu script in a bilingual document image. International Journal of Scientific Progress and Research, 20(2), 107-113. 1

[6] Bhat, Sachin, and G. Seshikala. "Character recognition of Tulu script using convolutional neural network." In International Conference on Artificial Intelligence and Data Engineering, pp. 121-131. Singapore: Springer Nature Singapore, 2019.

[7] Prasad Y. N., M. ., & H. K., C. . (2023). Recognition of Printed Kannada Text in Scene Images using Machine Learning Techniques. International Journal of Intelligent Systems and Applications in Engineering,12(8s),600–614.Retrieved from https://ijisae.org/index.php/IJISAE/article/view/4233

[8] Purohit, Ayush & Chauhan, Shardul. (2016). A Literature Survey on Handwritten Character Recognition. International Journal of Computer Science and Information Technology. 7. 1-5.

[9] Gandhare, S., Jyothi, P., & Bhattacharyya, P. (2018). Literature Survey : Spoken Language Translation.

[10] Honnegowda, Parikshith & Rajath, S & Shwetha, D & Sindhu, C & Ravi, P. (2021). Handwritten Character Recognition of Kannada Language Using Convolutional Neural Networks and Transfer Learning. IOP Conference Series: Materials Science and Engineering. 1110. 012003. 10.1088/1757-899X/1110/1/012003