AKSHARAM

An Educational Platform for Malayalam Language

Guide Name
Ms. Philo Sumi
Assistant Professor
Department of AIDS

Team Members:
Marc George, Merlin Sarah Jiju, Timon.K.John
Batch No: 2
Team No: 21

Contents

Objective

- The objective of this project is to develop an AI-powered Malayalam learning platform that enhances language acquisition through interactive experiences.
- The system integrates a LeNet-based handwritten character recognition model to assist users in learning how to write Malayalam characters with real-time feedback.
- It also provides pronunciation assistance, contextual learning, and conversation-based examples.
- The platform employs OCR and a translation model to extract and translate Malayalam text from images.
- Additionally, it features a text-to-speech module for accessibility.

Title	Summary	Technology Used	Advantages	Disadvantages
"Handwritten Malayalam Character Recognition System using Artificial Neural Networks" (**Authors:** Vaisakh V K ,Lyla B Das)(**Year:** 2020)	The paper presents a system for recognizing handwritten Malayalam characters using Convolutional Neural Networks (CNN). The process includes image acquisition, preprocessing using OpenCV, segmentation of characters using contour detection, and feature extraction via CNN. Real-time testing is performed by processing input images and classifying characters into text.	Preprocessing ,Segmentation ,Feature Extraction ,Classification.	Utilizes deep learning for improved performance , Real-time character recognition capability.	Struggles with variations in individual handwriting styles ,Computationally expensive, requiring powerful hardware, Requires a large labeled dataset for training ,

Table: Comparison of Existing Systems

Title	Summary	Technology Used	Advantages	Disadvantages
"Segmentation of Malayalam Handwritten Characters into Pattern Primitives and Recognition using SVM" (**Authors:** Baiju.K.B, Sabna.T.S, Lajish.V.L)(**Year:** 2020)	Proposes a segmentation-based approach for Malayalam handwritten character recognition. Uses RDP and EDFC for segmentation, extracts key features, and classifies using SVM (RBF).	Hi-Tech e-Writemate, Min-Max Normalization, RDP, EDFC, SVM (RBF), Python (Jupyter) .	High accuracy, Efficient feature extraction, Real-time applicability.	Limited to 8 vowels, Requires manual reference set, Struggles with visually similar characters.

Table: Comparison of Existing Systems

Title	Summary	Technology Used	Advantages	Disadvantages
"On developing handwritten character image database for Malayalam language script" (**Authors:** Manjusha K,Anand Kumar Madasamy,Soman Kp)(**Year:** 2019)	The paper focuses on developing a handwritten character image database for Malayalam script, essential for research in handwritten text recognition. It presents Amrita MalCharDb, an open-source database containing 85 character classes collected from 77 native Malayalam writers. The database is segmented using active contour models and tested with various feature extraction and classification techniques to evaluate recognition accuracy.	Image Processing ,Feature Extraction ,Classification,Dataset.	Provides a standardized dataset ,High recognition accuracy ,Open-source and extensible for further research.	Computationally expensive feature extraction techniques ,Requires further expansion to include all valid character shapes.

Table: Comparison of Existing Systems

Title	Summary	Technology Used	Advantages	Disadvantages
"A Novel Method for Malayalam Handwritten Character Recognition" (**Authors:** Anish S, Preeja V)(**Year:** 2015)	The project focuses on developing an offline handwritten character recognition (HCR) system for the Malayalam language. It employs a texture extraction model using a co-occurrence matrix and Euclidean distance for character identification.	Image Processing: Binarization, Segmentation, Feature Extraction,Classification	High recognition accuracy ,Effective for complex Malayalam characters ,Robust texture-based feature extraction	Struggles with highly similar characters,Limited to offline recognition ,Requires extensive training data for improved accuracy

Table: Comparison of Existing Systems

Title	Summary	Technology Used	Advantages	Disadvantages
"An Efficient Character Recognition System for Handwritten Malayalam Characters Based on Intensity Variations" (**Authors:** Abdul Rahiman M ,Rajasree M S)(**Year:** 2011)	The document presents a method for recognizing handwritten Malayalam characters using intensity variations. The approach addresses the challenges posed by the complexity and connectivity of Malayalam script characters. By classifying characters into three categories—Ra type, Pa type, and Special symbols—the method uses an algorithm based on HLH intensity patterns to identify these characters with high accuracy.	Optical Character Recognition (OCR) ,Feature Extraction Techniques ,Image Processing.	High Accuracy ,Versatility ,Editable Output.	Limited to Noiseless Environment ,Complexity in Processing Connected Characters ,Limited Generalization.

Table: Comparison of Existing Systems

Proposed System

- The proposed system aims to assist users in learning Malayalam characters, words, and sentences through AI-powered handwritten character recognition and language translation.
- Utilizes computer vision and optical character recognition technologies to enhance language comprehension by accurately extracting and analyzing Malayalam text and translating it so users can understand.
- Integrates gamification elements such as levels, achievements, and quizzes to make language learning engaging and interactive.
- Focused on usability, accuracy, and accessibility to help learners master Malayalam in an interactive and enjoyable manner.

Proposed System

- **Hand Written Character Recognition**
  - AI model analyzes handwritten Malayalam characters and provides accuracy feedback.
  - Assists users in improving handwriting through visual guidance and corrections.
- **Malayalam Word Learning**
  - Displays Malayalam words and meanings for enhanced vocabulary building.
  - Provides pronunciation assistance for better spoken language proficiency.
- **Sentence Translation and Contextual Learning**
  - Provides pronunciation assistance for better spoken language proficiency.
  - Offers context-based examples to help users understand sentence structures.
- **Image-based Text Recognition**
  - AI-powered OCR extracts Malayalam text from images.
  - Translates text into English to assist in reading comprehension.
- **Scalability and Adaptability**
  - Suitable for students, language learners, and non-native speakers.
  - Scalable to incorporate advanced features such as voice recognition and AI-driven personalized learning paths for a more immersive and adaptive learning experience.

Advantages of the Proposed System

- **User-Friendly and Accesible**
  - With an intuitive interface, the system is easy to use for learners of all ages and technical expertise.
- **Comprehensive Learning Approach**
  - Combining text, images, and handwriting recognition, along with AI-powered pronunciation and contextual learning, it ensures effective comprehension of Malayalam.
- **Engaging and Interactive Learning**
  - The system incorporates gamification with levels, achievements, and quizzes to keep learning fun, rewarding, and motivating.
- **Scalability and Future Growth**
  - Built to scale, the app can expand to include voice recognition and personalized learning paths, with potential support for more languages.
- **Free To Use**
  - The system is completely free, making it accessible to everyone who wants to learn Malayalam without any cost.

Software Requirements

- Pytorch - For the handwritten character recognition model.
- TensorFlow - For sequence-to-sequence translation model.
- SuryaOCR - For Optical Character Recognition of Malayalam Words.
- OpenCV - For preprocessing images.
- Django - For backend development.
- ReactJS - For frontend application.
- SQL - For database management

Conclusion

- In conclusion, the proposed AI-powered Malayalam learning platform utilizes advanced handwritten character recognition and OCR-based text extraction to enhance the language learning experience.
- By leveraging AI-driven translation and pronunciation assistance, the system provides users with a structured and interactive way to learn Malayalam characters, words, and sentences.
- Additionally, the platform facilitates contextual learning by offering real-world usage examples, helping users grasp the language more effectively .
- This project not only promotes efficient self-paced learning but also encourages linguistic and cultural awareness, making Malayalam more accessible to a broader audience.
- With its scalable and intelligent approach, the platform serves as an innovative solution to modern language learning challenges, bridging the gap between traditional and AI-powered education.

References

- “Handwritten Malayalam Character Recognition System using Artificial Neural Networks”, Vaisakh V.K, Lyla B. Das, 2020.
- “Segmentation of Malayalam Handwritten Characters into Pattern Primitives and Recognition using SVM”, Baiju K. B, Sabna T. S, Lajish V. L, 2020.
- ”On developing handwritten character image database for Malayalam language script”, P Shourie, V Anand, D Upadhyay, 2019.
- ”A Novel Method for Malayalam Handwritten Character Recognition”, Anish S, Preeja V, 2015.
- ”An Efficient Character Recognition System for Handwritten Malayalam Characters Based on Intensity Variations”, Abdul Rahiman M, Rajasree M S, 2011.

Thank You!