

# AKSHARAM

An Educational Platform for Malayalam Language

## Guide Name

Ms. Philo Sumi  
Assistant Professor  
Department of AIDS

## Team 21:

Marc George - Roll No : 43  
Merlin Sarah Jiju - Roll No : 45  
Timon.K.John - Roll No : 67

# Contents

- ① Problem Statement
- ② Objective
- ③ Existing System and Literature Survey
- ④ Inference from Literature Survey
- ⑤ Proposed System
- ⑥ Methodology
  - Data Flow Diagram
  - Architecture Diagram
  - Module Details
  - Software Requirements
- ⑦ Pending Tasks
- ⑧ Conclusion
- ⑨ References

# Problem Statement

## Problem

Traditional textbook-based learning methods for concepts, subjects, and languages have become increasingly unappealing to modern learners. With the average adult attention span now at just 8.25 seconds, there is a growing demand for faster, more engaging, and cost-effective approaches to language acquisition. In our state, Malayalam remains predominantly taught in conventional classroom settings, limiting accessibility and failing to align with evolving learning preferences.

## Solution

**Aksharam, AI Powered Malayalam Learning Platform**

# 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.

# Existing System and Literature Survey

Title	Summary	Technology Used	Advantages	Disadvantages
Vaisakh V K ,Lyla B Das, "Handwritten Malayalam Character Recognition System using Artificial Neural Networks", (2020)	The paper presents a CNN-based system for recognizing handwritten Malayalam characters, involving image acquisition, preprocessing with OpenCV, character segmentation using contour detection, and feature extraction.	Preprocessing ,Segmentation ,Feature Extraction ,Classification.	Utilizes deep learning for improved performance , Real-time character recognition capability.	Computationally expensive, requires powerful hardware, Requires a large labeled dataset
Baiju.K.B,Sabna.T.S, Lajish.V.L, "Segmentation of Malayalam Handwritten Characters into Pattern Primitives and Recognition using SVM" (2020)	Proposes a segmentation-based approach for Malayalam handwritten character recognition using RDP and EDFC, with feature extraction and SVM (RBF) classification.	Hi-Tech e-Writemate, Min-Max Normalization, RDP, SVM	High accuracy, Efficient feature extraction, Real-time applicability.	Limited to 8 vowels, Requires manual reference set, Struggles with visually similar characters.

Table: Literature Survey

# Existing System and Literature Survey

Title	Summary	Technology Used	Advantages	Disadvantages
<b>Manjusha K, Anand Kumar Madasamy, Soman Kp, "On developing handwritten character image database for Malayalam language script" (2019)</b>	The paper introduces Amrita MalCharDb, an open-source handwritten character database for Malayalam script. It is tested with various feature extraction and classification techniques	Image Processing ,Feature Extraction ,Classification.	standardized dataset ,High recognition accuracy ,Open-source	Computationally expensive ,Requires further expansion to include all valid character shapes.
<b>Anish S, Preeja V, "A Novel Method for Malayalam Handwritten Character Recognition" (2015)</b>	The project develops an offline Malayalam HCR system using texture extraction with a co-occurrence matrix and Euclidean distance for character identification.	Image Processing, Segmentation, Feature Extraction, Classification	High accuracy ,Effective for complex Malayalam characters ,Robust feature extraction	Limited to offline recognition ,Requires extensive training data for improved accuracy

Table: Literature Survey

# Existing System and Literature Survey

Title	Summary	Technology Used	Advantages	Disadvantages
<b>M. Abdul Rahiman and Rajasree M. S., “An Efficient Character Recognition System for Handwritten Malayalam Characters Based on Intensity Variations” (2011)</b>	OCR for handwritten Malayalam using “HLH” intensity-based segmentation to handle single and connected characters.	- HLH (High–Low–High) segmentation, Categorizes vertical/horizontal/special forms	High accuracy, Handles older/complex Malayalam forms, Simple segmentation	Needs clean input, Language-specific approach, Slightly lower accuracy for horizontal recurrences

Table: Literature Survey

# Inference from Literature Survey

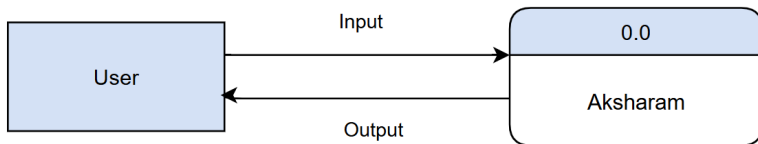
- Convolutional Neural Networks (CNNs) are widely used for handwritten Malayalam character recognition due to their strong feature extraction capabilities.
- Various segmentation techniques like contour detection, RDP improve character isolation for better recognition accuracy.
- Different methods such as texture extraction, co-occurrence matrices, and deep learning-based feature extraction enhance classification performance.
- Combining multiple techniques, such as SVM with RBF improves system robustness.
- Large, well-structured datasets like Amrita MalCharDb are crucial for advancing handwritten character recognition research.



# Proposed System

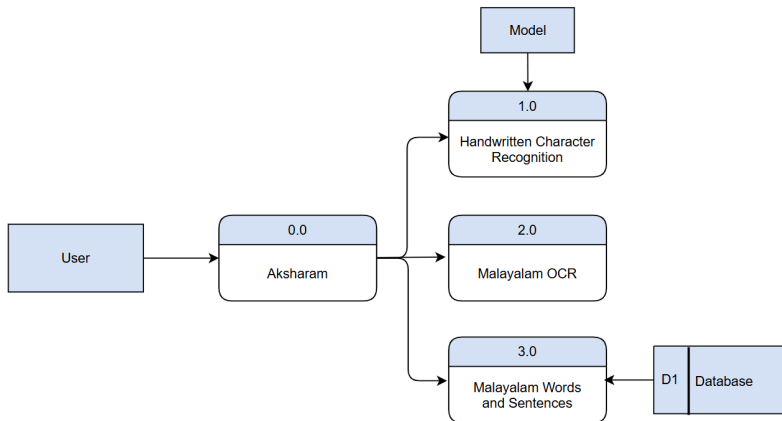
- The proposed system aims to assist users in learning Malayalam characters, words, and sentences through AI-powered handwritten character recognition and language translation.
- Utilizes computer vision and optical character recognition technologies to enhance language comprehension by accurately extracting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

## A. Data Flow Diagrams



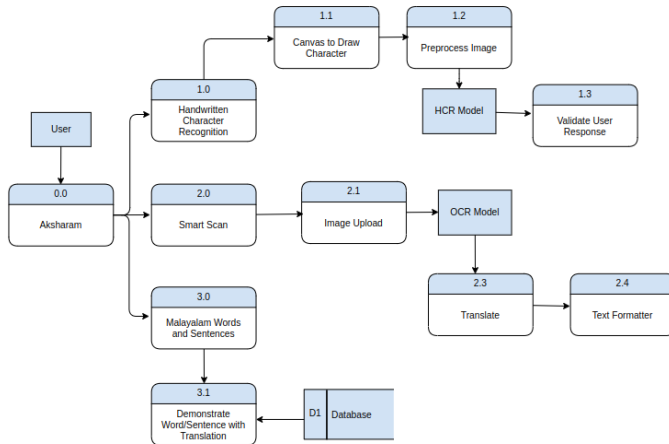
# Methodology

## A. Data Flow Diagrams



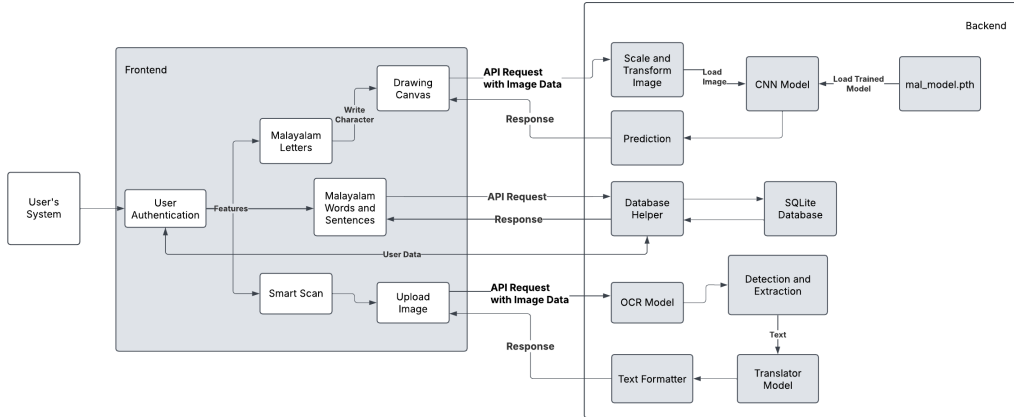
# Methodology

## A. Data Flow Diagrams



# Methodology

## B. Architecture Diagram



## C. Module Details

### • 1. Letter Classification Module:

- The user is shown how to write the character and then their written input is captured in the canvas provided. The image is pre-processed and sent to a classification model, which determines the correct alphabet and decides whether the user can proceed to the next letter.

### • Components of Letter Classification Module:

- Demo: Demonstrates how to write the letter.
- Canvas: The user then writes the letter on the given space.
- Classification Model: The written letter is then inputted to the model where it is classified.
- Result: Determines whether the user can move on to the next character or not.

## C. Module Details

### ● 2. Smart Scan Module:

- The Image-Based Learning Module helps users learn Malayalam by extracting and translating text from images. Using an OCR model, the system detects Malayalam words in images, while a translation model converts them into the user's preferred language. This feature enhances learning by providing real-world examples, making language acquisition more interactive and engaging.
- **Components of Smart Scan Module:**
  - Image Input: Users upload or capture an image containing Malayalam text.
  - OCR Model: Extracts Malayalam words from the image.
  - Translation Model: Translates the extracted text into the desired language.
  - Result: Shows the original and translated text for learning.

## C. Module Details

- **3. Words and Sentences Learning Module:**
- Helps users build their Malayalam vocabulary and conversational skills using a structured database. It teaches Malayalam words by displaying their meanings, pronunciation. Additionally, it introduces basic Malayalam sentences providing example sentences along with its English translations to enhance understanding and fluency.
- **Components of Words and Sentences Learning Module:**
- Demonstrates the Word/Sentence
- Provides English translation for understanding



## D. Software Requirements

- Pytorch - For the handwritten Character recognition model and translator model.
- SuryaOCR - For Optical Character Recognition of Malayalam Words.
- OpenCV - For Preprocessing images.
- Django - For the backend of the platform.
- ReactJS - For an interactive and user friendly Interface
- SQLite - For storing User data, malayalam words and sentences

## Pending Tasks

- **Smart Scan:** Integrating the OCR Translator to the frontend application
- **Frontend:** Develop and implement the user interface, ensuring a responsive and intuitive design for seamless user interaction.
- **Letter Classificaton Model Optimization:** Trying to achieve better accuracy for the test results and how to optimize the model.

# 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

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

**Thank You!**