International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 03 | March - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

A Machine Learning and Deep Learning Approach for Malayalam Handwritten Character Recognition

Surya Ashok[1], Anupama M[2], Rekhalakshmi K R[3]

[1,2,3] *Department of Computer Science*
[1,2,3] *College of Applied Science Chelakkara*

Abstract - *Handwritten character recognition is an important research area in pattern recognition and computer vision with applications in document digitization and automated information processing. Recognition of Malayalam handwritten characters is particularly challenging due to the complex structure of the script, the presence of curves and loops, and variations in individual handwriting styles. This paper presents a system for recognizing isolated Malayalam handwritten characters using two classification approaches: Support Vector Machine (SVM) and Convolutional Neural Network (CNN). In the SVM-based approach, Histogram of Oriented Gradients (HOG) is used to extract discriminative features from character images before classification. In contrast, the CNN model learns feature representations automatically from the image data during the training process. A dataset consisting of handwritten samples of Malayalam characters is used to train and evaluate the models. The study focuses on analyzing the effectiveness of traditional machine learning and deep learning approaches for Malayalam handwritten character recognition.*

Key Words: Malayalam handwritten character recognition, Support Vector Machine, Convolutional Neural Network, Histogram of Oriented Gradients, Optical Character Recognition, Pattern recognition.

1.INTRODUCTION

Handwritten Character Recognition (HCR) is the process of converting handwritten text into a digital form that can be processed by computers. This technology plays a significant role in applications such as document digitization, postal address interpretation, form processing, and digital archiving of historical manuscripts.

Malayalam is one of the widely spoken languages in the southern region of India. The Malayalam script consists of a large number of characters including vowels, consonants, and compound symbols. Many characters contain curved shapes, loops, and visually similar structures, which makes the recognition of handwritten Malayalam characters a challenging problem.

The structural complexity of Malayalam characters and the variations in individual handwriting styles make accurate recognition difficult for traditional recognition systems.

This research focuses on developing a system for recognizing isolated Malayalam handwritten characters using both machine learning and deep learning techniques. In particular, the study evaluates the performance of Support Vector Machine (SVM) and Convolutional Neural Network (CNN) models in recognizing handwritten Malayalam characters.

2. LITERATURE REVIEW

Handwritten character recognition has been an active research area within pattern recognition and computer vision for several decades. Early research focused on rule-based and statistical methods that relied heavily on handcrafted feature extraction techniques.

For Indian language scripts, several feature extraction approaches have been proposed. Techniques such as projection histograms, zoning, and wavelet transforms have been widely used to represent character structures. These methods were commonly combined with machine learning classifiers such as k-Nearest Neighbour (k-NN), Artificial Neural Networks (ANN), and Support Vector Machines (SVM). SVM in particular gained popularity due to its ability to perform well with high-dimensional feature spaces and limited training samples.

In the context of Malayalam handwritten character recognition, previous studies have explored different feature extraction and classification strategies. Wavelet-based feature extraction combined with SVM classifiers has shown promising results in identifying Malayalam characters. Other works have utilized fuzzy zoning techniques and neural network models to improve recognition accuracy.

With the advancement of deep learning techniques, Convolutional Neural Networks (CNN) have become

widely used for image recognition tasks. CNN models automatically learn hierarchical features from images and have significantly improved the performance of handwritten character recognition systems. Several studies have demonstrated that CNN architectures outperform traditional machine learning methods in recognizing complex scripts.

Despite these advancements, Malayalam handwritten character recognition remains a challenging task due to the large character set and structural similarities between characters. Therefore, evaluating both traditional machine learning models and deep learning approaches can provide valuable insights into their effectiveness for this problem.

3. METHODOLOGY

The proposed system for Malayalam handwritten character recognition follows a multi-stage processing pipeline. The main objective of the system is to accurately identify isolated handwritten Malayalam characters from scanned images. The recognition process consists of several stages including image acquisition, preprocessing, segmentation, feature extraction, classification, and output generation.

3.1 System Architecture

The system architecture describes the overall structure and workflow of the Malayalam handwritten character recognition system. The proposed system is designed to process handwritten character images and convert them into digital text through a sequence of processing stages. Each stage performs a specific task that contributes to the recognition process.

The architecture consists of several modules including image acquisition, preprocessing, segmentation, feature extraction, classification, and output generation.

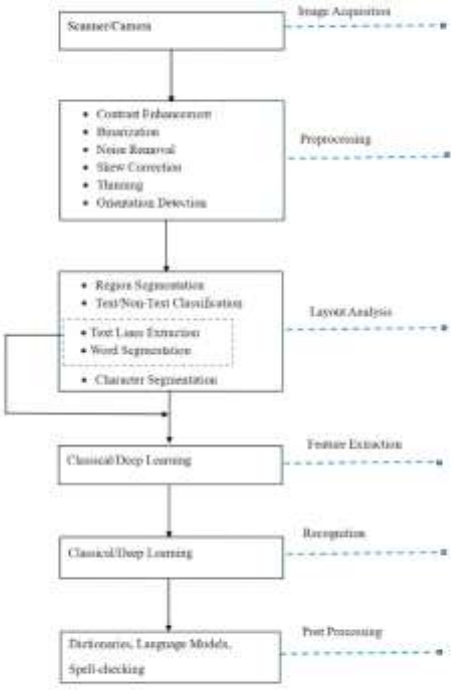

Fig 1. Architecture Diagram

3.2 System Overview

The overall system architecture is designed to process handwritten character images and classify them into corresponding Malayalam character classes. The workflow of the system can be summarized as follows:

1. Image acquisition
2. Image preprocessing
3. Character segmentation
4. Feature extraction
5. Character classification
6. Output generation

3.3 Image Acquisition

Image acquisition is the first stage of the recognition system. In this step, handwritten Malayalam characters are collected and converted into digital images using a scanner or camera device. The captured images are stored in standard formats such as PNG or JPEG.

The dataset used in this study consists of handwritten samples collected from multiple writers in order to capture variations in handwriting styles.

3.4 Image Preprocessing

Preprocessing is an essential step that prepares the input images for feature extraction and classification. Raw handwritten images may contain noise, uneven illumination, or background variations that can affect recognition accuracy.

The preprocessing stage includes the following operations:

- Conversion of RGB images into grayscale format.
- Noise removal using filtering techniques.
- Image binarization to convert grayscale images into binary form.
- Normalization to standardize image size and orientation.

These operations improve image quality and ensure consistency across the dataset.

3.5 Character Segmentation

Segmentation is the process of isolating the region of interest containing the handwritten character from the background. In this stage, the system identifies the boundaries of the character and extracts the relevant portion of the image.

The segmentation process reduces unnecessary image data and focuses only on the character region, which improves the efficiency of feature extraction and classification. Accurate segmentation is essential because improper separation of characters may lead to incorrect feature extraction and reduced recognition accuracy.

3.6 Feature Extraction

Feature extraction aims to represent the character image using numerical descriptors that capture its structural characteristics. In the SVM-based approach, Histogram of Oriented Gradients (HOG) is used as the feature extraction method.

HOG analyses the distribution of gradient orientations in localized portions of the image. This technique effectively captures edge directions and shape information, which are important for distinguishing between different characters. The resulting feature vectors are then used as input for the classification stage.

3.7 Classification

In the classification stage, the extracted features are used to determine the class label of the handwritten character. Two classification approaches are considered in this study:

Support Vector Machine (SVM)

Support Vector Machine (SVM) is a supervised learning algorithm widely applied in pattern recognition and classification problems. The main objective of SVM is to identify an optimal separating boundary, known as a hyperplane, that distinguishes data points belonging to different classes.

In the proposed recognition system, SVM is employed to classify handwritten Malayalam characters based on features extracted from the character images. To represent the structural information of characters, Histogram of Oriented Gradients (HOG) is used as the feature extraction method. The Histogram of Oriented Gradients (HOG) technique captures the distribution of gradient directions in an image. By analysing edge orientations and local intensity changes, HOG effectively represents the shape and structural patterns present in handwritten characters. After extracting HOG features, the resulting feature vectors are used to train the SVM classifier. During the testing phase, the trained model receives feature vectors from new handwritten images and determines the most probable character class based on the learned decision boundary.

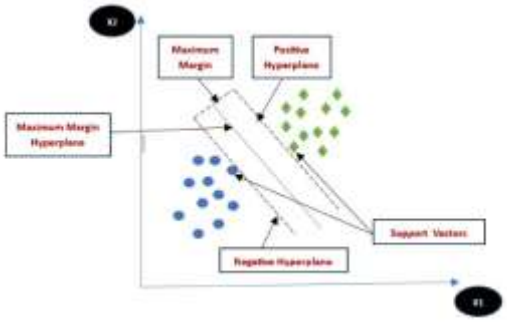

Fig 2. SVM Diagram

Convolutional Neural Network (CNN)

Convolutional Neural Networks (CNN) are deep learning models specifically developed for processing image data. Unlike traditional machine learning techniques, CNNs automatically learn relevant features from images during the training process without requiring manual feature extraction. The CNN model implemented in this study

consists of several layers that work together to extract features and perform classification. These layers include:

- An input layer that receives the character image.
- Convolutional layers that detect spatial features such as edges and curves.
- Activation layers that introduce non-linear transformations.
- Pooling layers that reduce the dimensionality of feature maps.
- Fully connected layers that perform classification

The final classification is performed using a Softmax layer, which predicts the most likely class among the 48 Malayalam characters. CNN models are particularly effective for handwritten character recognition because they can automatically identify patterns such as stroke direction, curvature, and character structure directly from the input images.

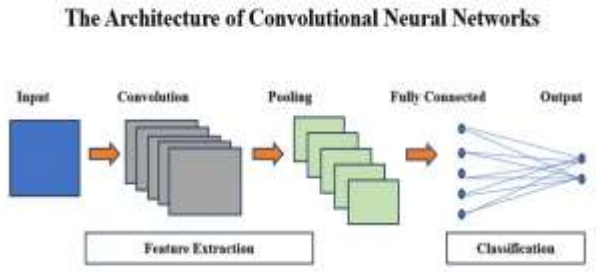

Fig 3. CNN Diagram

3.8 Output Generation

The final stage of the system generates the recognition result. The predicted class label is mapped to the corresponding Malayalam character using Unicode representation. The recognized character can then be displayed or stored as editable text. This enables the handwritten input to be converted into a digital format that can be easily processed, edited, or archived.

4 Dataset and Experimental Results

The experimental evaluation was conducted using a dataset containing 5099 handwritten Malayalam character samples, representing 48 different characters. From this dataset, 987 samples were reserved for testing, while the remaining samples were used for training the models.

4.1 Performance Evaluation

Accuracy = Correct Predictions/Total Predictions

Error Rate = 1- Accuracy

Performance Comparison

Model	Accuracy	Error Rate
SVM	96.35%	3.65%
CNN	98.12%	1.88%

Fig 4. Result Comparison Table

4.2 Accuracy Comparison Graph

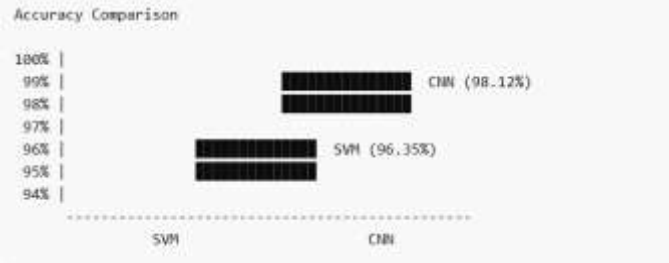

Fig 5. Accuracy Comparison Graph

The results indicate that the CNN model achieved higher recognition accuracy because it is capable of automatically learning complex visual features from handwritten character images.

4.3 Confusion Analysis

Some Malayalam characters have similar shapes which leads to misclassification.

Character 1	Character 2	Reason
ച	ച	Similar curve structure
ജ	ഇ	Similar stroke pattern

Fig 6. Character Analysis

5. CONCLUSIONS

This research presented a handwritten Malayalam character recognition system using both Support Vector Machine and Convolutional Neural Network models. The proposed system processes handwritten character images through multiple stages including preprocessing, segmentation, feature extraction, and classification.
Experimental results showed that the SVM-based model achieved an accuracy of 96.35%, while the CNN-based

model obtained an accuracy of 98.12%. These findings demonstrate that deep learning techniques can provide improved performance for recognizing complex handwritten scripts such as Malayalam.

Future research may focus on extending the system to recognize compound characters, handwritten words, and complete document images. Increasing the size and diversity of the dataset may also further improve recognition accuracy.

ACKNOWLEDGEMENT

The authors would like to express their sincere appreciation to all individuals who provided guidance, encouragement, and valuable suggestions during the course of this research work. Their support and constructive feedback greatly contributed to the completion of this study.

REFERENCES

[1] B. Jose and K. Pushpalatha, "Intelligent handwritten character recognition for Malayalam scripts using deep learning approach," *IOP Conference Series: Materials Science and Engineering*, vol. 1085, no. 1, 2021.

[2] B. Jose, "Optimized Malayalam handwritten character recognition using hybrid deep learning model," *ACM Transactions on Asian and Low-Resource Language Information Processing*, 2025.

[3] A. James and M. Mathew, "Malayalam handwritten character recognition using convolutional neural network architecture," *International Journal of Advanced Research in Science, Communication and Technology*, 2024.

[4] R. C. Karpagalakshmi, "Deep learning-based recognition of handwritten characters using convolutional neural networks," *Procedia Computer Science*, 2025.

[5] M. A. Liman, "Handwritten character recognition using deep learning models," *Journal of Imaging and Vision*, vol. 8, no. 2, pp. 120–130, 2024.

[6] X. F. Wang et al., "A survey of text detection and recognition algorithms based on deep learning," *Neurocomputing*, 2023.

[7] W. AlKendi et al., "Advancements and challenges in handwritten text recognition systems," *Journal of Imaging*, 2024.

[8] S. K. Singh and R. K. Sharma, *"Handwritten Character Recognition Using Deep Convolutional Neural Networks,"* International Journal of Computer Vision and Image Processing, vol. 13, no. 2, pp. 45–56, 2023.

[9] P. Kumar and A. Gupta, *"Deep Learning-Based Optical Character Recognition for Handwritten Text Recognition,"* IEEE Access, vol. 11, pp. 78234–78245, 2024.