





Multiscale Residual Network for Recognizing Handwritten Malayalam Characters



Samatha Pararath Salim¹, Ajay James^{2*}, Philomina Simon³, Bisna Nellichode Divakaran²

¹ School of Psychology and Public Health, La Trobe University, Bundoora 3086, VIC, Australia

² Department of Computer Science and Engineering, Government Engineering College, Thrissur 680009, India

³ Department of Computer Science, University of Kerala, Thiruvananthapuram 695034, India

Corresponding Author Email: ajay@gectcr.ac.in

Copyright: ©2024 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ts.410136>

ABSTRACT

Received: 19 June 2023

Revised: 23 November 2023

Accepted: 23 December 2023

Available online: 29 February 2024

Keywords:

convolutional neural network (CNN), deep learning, handwritten character recognition (HCR), machine learning, multi-scaled features, neural network, residual network, Malayalam

In domains such as banking cheque processing and automated mail sorting, the recognition of handwritten characters is of paramount importance. In Kerala, where Malayalam serves as the primary language for government documentation, the accurate identification of its handwritten characters is crucial. This study introduces a novel approach leveraging a deep residual neural network with multiscale feature extraction for the recognition of Malayalam handwritten characters, encompassing both basic and compound characters as well as signs. Traditional methods of character recognition often rely on handcrafted feature extraction, which, while achieving commendable accuracy, are prone to misclassification due to reliance on low- and mid-level features in the output layer classifier without consideration of parameter modifications. The proposed method addresses these limitations by integrating multiscale features, enhancing the model's ability to discern intricate character details. Evaluated using the P-ARTS Kayyazhuthu Dataset, this approach demonstrated a remarkable accuracy of 99.56%. Additionally, a commendable accuracy of 98% was achieved on other test datasets. The findings underscore the efficacy of deep learning techniques over conventional methods in handwritten character recognition (HCR), particularly in the context of the complex Malayalam script. This study contributes significantly to the field of machine learning and handwriting analysis, offering robust solutions for applications requiring high precision in character recognition.

1. INTRODUCTION

Amidst a period of widespread computerization, there has been significant advancement in the field of HCR. This development can be attributed to the increasing demand for efficient digitization of various linguistic scripts. Malayalam, the official language of the Indian state of Kerala, poses unique challenges compared to other scripts. Malayalam handwritten character recognition (MHCR) refers to the technique or system that can automatically identify and convert handwritten Malayalam characters into digital text. The need for digitizing handwritten Malayalam text documents has arisen due to various reasons, including automated data entry and preservation of archives. In addition, the Government of Kerala has formally designated Malayalam as its official language for governance. This paves the way for the necessity of digitizing Malayalam documents. There is a possibility that, as time passes, these works may deteriorate and be lost. This can be avoided by converting and digitally saving these documents.

Recognizing handwritten characters in Malayalam remains challenging due to the script's cursive nature, variations in individual writing styles, and the presence of ligatures. Despite breakthroughs in optical character recognition (OCR) for printed text, these factors continue to pose difficulties. Furthermore, handwritten character identification must take

into account factors such as the dimensions of the characters, the caliber and type of pen used, the age of the documents, the paper quality, the presence of slanted lines, the variability of the lines, and the proximity of the lines in different writing styles. The recognition of handwritten Malayalam characters is regarded as a difficult undertaking due to the extensive range of character sets, the scarcity of research conducted in Malayalam, the absence of a standardized dataset, and the varied writing styles of Malayalam characters. The significance of this research field lies in the scarcity of research conducted on the scripts of South Indian Dravidian languages.

Machine learning is a prominent and extensively utilized discipline that finds application in OCR. Several techniques, such as statistical and structural feature extraction [1] and extended zone-based algorithms [2], achieved a greater accuracy rate despite the complexity of the dataset. Deep learning is a highly active area of research in computer vision due to its superior capabilities in feature extraction and generalization, which yield impressive outcomes. Convolutional neural networks have demonstrated their efficacy in achieving favorable outcomes for image recognition, particularly in applications linked to OCR. CNN has demonstrated its effectiveness at handling intricate characters such as Chinese, as well as other diverse languages.

This research presents a more efficient method for

identifying handwritten Malayalam characters and conducts a comprehensive assessment of the several OCR algorithms already employed. This study also examines the efficacy of deep learning in comparison to conventional feature extraction methods in MHCR. The experimental analysis demonstrates that the suggested work offers a superior method for the recognition of handwritten characters in the Malayalam language.

2. RELATED WORKS

Tremendous efforts have been made in the field of MHCR research during the last decade. Among these, a few of them achieved around 90% accuracy on the Parts Kayyethuth vowel and consonant dataset [1]. However, the recognition of the complete range of datasets is still challenging. This section gives a study of different character recognition methods and provides methods to improve the classification accuracy of handwritten characters that can be implemented in MHCR.

There are a few established techniques for offline Malayalam handwritten characters, such as hybrid feature extraction, which combines statistical and structural features (SSF). In the research by James et al. [2, 3], 12 distinct features are extracted and classified with 91% accuracy using the decision tree method. Another work by Raveena et al. [4] uses extended zoning for feature extraction, including nine features from the Malayalam character set and classifying them with SVM classifiers. Both methods struggled to classify some of the similarly-shaped characters. Nair et al. [5] suggests a new method for HCR that uses a convolutional neural network on the augmented dataset. The CNN model uses the LeNet-5 architecture and, using backpropagation, achieved an accuracy of 95%. Another attempt by Bhagyasree et al. [6] puts forth a strategy for utilizing DAG-CNN to recognize handwritten cursive English characters. For categorization, this approach leverages features that are retrieved from several network levels.

For handwritten English characters, Parthiban et al. [7] utilizes a recurrent neural network in his work. The network is used to determine the character arrangement. A hybrid handwriting character recognition technique with transfer deep learning was proposed by Can and Yilmaz [8]. Transfer learning minimizes the size of the sample dataset and the processing power. The model is trained on the NIST19 dataset, and the hybrid model produced a 1.1% increase in accuracy. The work by Pragathi et al. [9] recognizes the handwritten Tamil characters using deep learning. The method uses the VGG-16 CNN model and achieved an accuracy of 94.52%. Borad et al. [10] uses augmentation-based CNN for the

recognition of handwritten Gujarati characters. The dataset collected from primary schools was preprocessed and applied to an augmented multi-layer perceptron (MLP) and achieved an accuracy of 98.6%. English HCR using the Hidden Markov Model and deep learning proposed by Alkawaz et al. [11] uses online handwritten input data. For the implementation, they used the Kohonen network together with deep learning, and accuracy improved by 37.49%.

Another work by Mandal et al. [12] proposed a method to recognize Indic handwritten characters. The method uses a capsule network, in which kernels can work together in consensus with one another with the help of dynamic routing to employ equivalence among kernels. The method boosts performance over LeNet and AlexNet. A model by Gupta et al. [13] analyzed the effect of optimizers and activation functions on the recognition of the Devanagari dataset. It is observed that Adam optimizers and leakyReLU perform better and produce a testing accuracy of 99.20%.

One of the largest Urdu printed datasets is recognized in the study by Nasir et al. [14], which uses CNN for image feature extraction, passes the extracted features to stacked BLSTM layers to generate a sequence of probabilities at each timestamp, and finally uses CTC loss for training and produces an accuracy of about 99%. Darwish and Elzoghaly [15] developed a better approach for reading Arabic characters printed on paper that makes use of a fuzzy classifier with biological inspiration. The fuzzy K-nearest neighbor classifier (F-KNN) and genetic algorithm (GA) are used in a single framework to increase identification accuracy by 98.69%.

Mohd et al. [16] outline a technique for recognizing Quranic optical text using deep learning models. Comparing the recognition rates of six models was done. The best model, which yields an accuracy of 98%, combined the fundamental CNN with the recurrent neural network.

Song et al. [17] describe a method to increase Chinese character recognition accuracy by improving the dataset quality and quantity using sample set expansion. Together with that, it uses other augmentation transforms on the dataset, which results in an accuracy of 94.47%.

Bangla HCR achieved state-of-the-art accuracy with the architectural enhancements to the deep convolutional neural network suggested by Ashiquzzaman et al. [18] incorporated dropout to reduce overfitting and ELU to overcome the vanishing gradient problem. This results in an overall accuracy of 96.68%.

Another work by Wang and Liu [19] introduce a center loss module to the basic Resnet backbone network for the recognition of Chinese characters. The center loss module enabled matrix learning and achieved an accuracy of 97.03%.

Table 1. Comparison of different enhancement approaches done on neural network-based OCR systems during pre-training, training, and post-training

Model	Language	Method	Accuracy
CNN	Chinese	Enhancement in Pre-training Phase.	94.47%
		Uses sample set expansion method.	
ResNet	Chinese	Transformations on dataset: Gaussian white noise, affine, rotation, scaling.	97.03%
		Enhancement in training phase.	
Deep CNN	Bangla	Uses ResNet with center loss module.	96.68%
		Enhancement in training phase.	
CNN	Chinese	Modified model to include dropout and ELU.	96.03%
		Enhancement in post-training phase.	
		Detect mis-recognized characters through comparing the confidence value with a threshold	

Wang and Liu [19] suggested a method of string-level confidence learning under the minimum classification error criterion. This method was shown to effectively improve recognition performance on Chinese character strings.

The HCR approaches can be broadly classified into traditional feature extraction and modern machine learning methods. Among these, machine learning methods offered better accuracy. Moreover, the performance of machine learning methods can be improved in three ways: Pre-training phase: improving the quality and quantity of the training dataset, training phase: some architectural enhancements to the basic backbone network, and post-training phase: evaluating results, comparing them with ground truth, and then doing some error detection and correction. Table 1 provides a comparison of different such methods.

This work focuses on improving the Malayalam Handwritten Character Recognition system by performing architectural enhancements to the basic backbone residual network.

3. SYSTEM ARCHITECTURE

The architecture of an OCR system can be divided into various sub-systems involving various stages. The stages include data collection, augmentation, pre-processing, CNN modeling, classification, and testing. So, for the proper working of the OCR system for Malayalam handwritten characters, the system is designed, and the overall system architecture is shown in Figure 1.

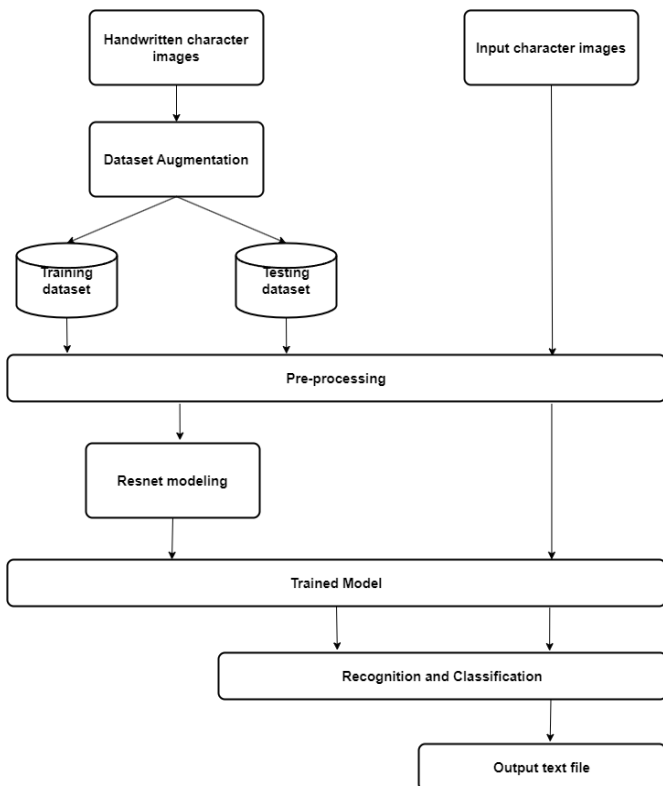


Figure 1. System architecture

The input is scanned using a scanner or taken as a smartphone photograph. The first action that must be taken is pre-processing. The system is loaded with the scanned input image. Next, apply binarization to the image to create a rendition of the character image in black and white. Then the

picture is downsized to 86×86. Since the Malayalam characters have a different style of writing and the use of compound characters varies from character to character, So the input image size is experimentally fixed at 86x86 to capture the Malayalam characters in an efficient way. Also, when the image size becomes small, the convolution layer extracts the features and performs convolution faster compared to images with a larger size. The initialization of kernel weights uses a gaussian distribution. Training images are used to train the network. The CNN is trained using the backpropagation learning rule. The features are extracted using the trained network. The same CNN is used for classification. The layer prior to the softmax layer is a fully connected layer that gives the features of the character at a high level to the softmax classifier. The final one is the softmax layer. This classifies the characters. In this paper, we used a residual network called ResNet [20].

Basic CNN, like LeNet-5, is a simple CNN architecture that has three convolution layers, two max-pooling layers, and two fully connected layers that are fed into the classification layer, which produces the output. This was one of the earliest CNNs that was created, and it was trained to recognize handwritten digits. Residual networks have proven to be highly effective in computer vision tasks. In MHCR, the complexity of character recognition can be high, but utilizing deeper networks can potentially capture more intricate patterns and hierarchies in the data.

3.1 Residual network (ResNet)

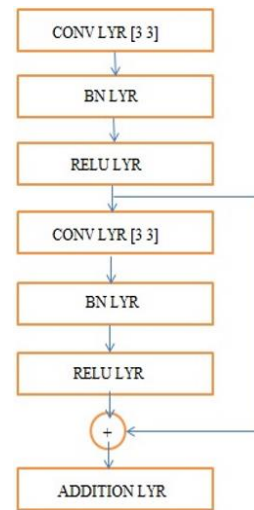


Figure 2. Residual block

ResNet-50 was one of the most popular deep CNN models. Residual network, abbreviated as ResNet, was built of residual blocks that implement skip connections. Residual learning is enabled by a skip connection. That is, instead of feeding input $x[i]$ to layer i as in a plain network, it is also fed to layer $(i + 1)$. So the local characteristics of the input data are directly fed to higher-level layers. This fixes the vanishing gradient issue, which arises when a deep multilayer feed-forward network is unable to transmit useful gradient data from the model's output end to the layers close to its input end. In Figure 2, it is depicted. Deep network training is essential in MHCR, as character recognition frequently requires knowledge of both local and global properties. Moreover, residual connections enable the learning of residual features, which can be important for capturing fine details and nuances in the input

data. This can be beneficial in character recognition, where characters in Malayalam scripts have subtle variations, and the network needs to discern these details for accurate recognition.

3.2 Multi-scaled ResNet

A multi-scaled residual network is a new approach to character recognition in a deep learning framework. Contemporary approaches extract features from a single output layer for classification and prediction. Since feature representation plays a vital role in MHCR, which includes complex, curved, yet not cursive scripts, methods to incorporate intricate features need to be done in addition to skip connection by ResNet. Classification accuracy can be increased by utilizing multi-scaled features that are collected from numerous low-, mid-, and high-level layers. The network adds multi-scaled features to the addition block's classification by skipping connections. Coarse and fine-grained classification tasks can efficiently share the multi-scaled data. There are five possible network designs, depending on how the addition blocks are entered in different combinations. The addition block of multi-scaled ResNet-1 has a combination of resblk-1 and resblk-6; multi-scaled ResNet-2 has resblk-2 and resblk-6; and so on up to multi-scaled ResNet-5 has resblk-5 and resblk-6.

In the multi-scaled ResNet-1 architecture, the output of the first and final residual blocks is connected to the addition block by a skip connection and given to the classifier. In the multi-scaled ResNet-2 architecture, the output of the second

and final residual blocks is connected to the addition block by a skip connection and given to the classifier. In the multi-scaled ResNet-3 architecture, the output of the third and final residual blocks is connected to the addition block by a skip connection and given to the classifier. In the multi-scaled ResNet-4 architecture, the output of the fourth and final residual blocks is connected to the addition block by a skip connection and given to the classifier. In the multi-scaled ResNet-5 architecture, the output of the fifth and final residual blocks is connected to the addition block by a skip connection and given to the classifier. These architectures are plotted in Matlab and shown in Figures 3-7.

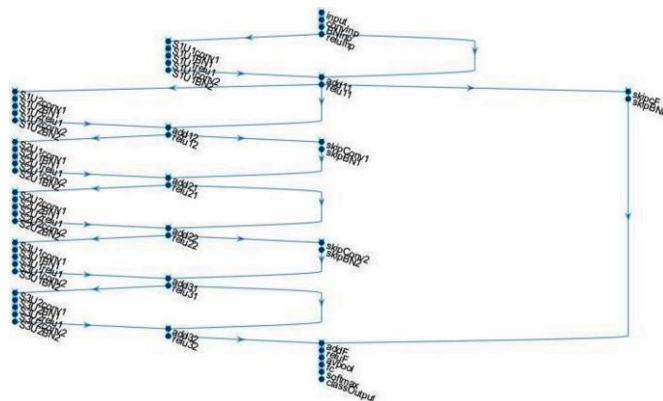


Figure 3. Visualization of multi-scaled ResNet-1 architecture

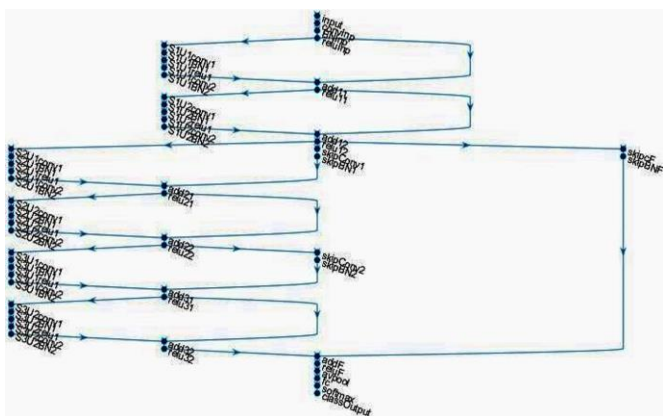


Figure 4. Visualization of multi-scaled ResNet-2 architecture plotted in MATLAB

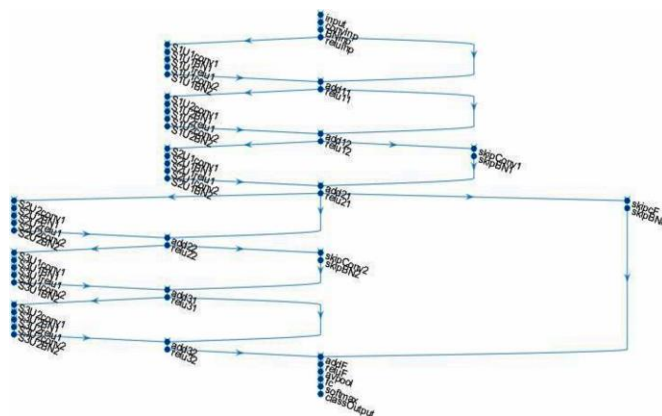


Figure 5. Multi-scaled ResNet-3 architecture visualized in MATLAB

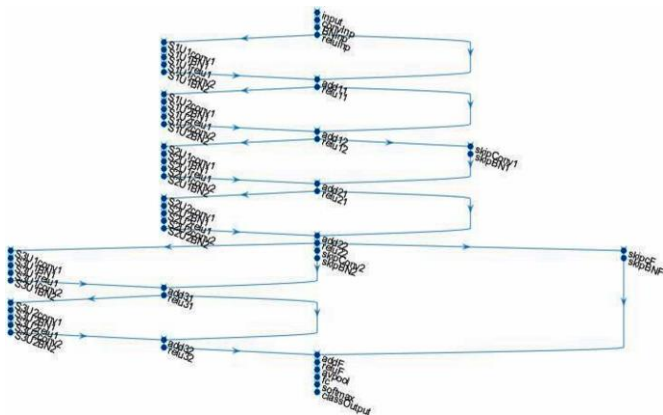


Figure 6. Multi-scaled ResNet-4 architecture visualized in MATLAB

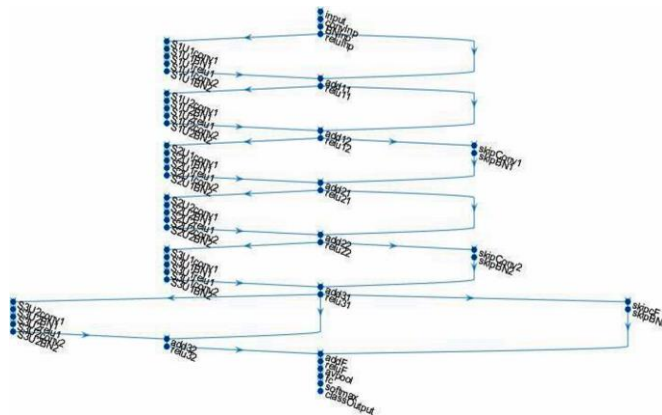


Figure 7. Multi-scaled ResNet-5 architecture visualization

4. RESULTS AND DISCUSSION

The entire dataset is first divided into a training dataset and a testing dataset. The training set is picked at random from the dataset, while the testing set is chosen at random from the remaining 20%. This training set is assessed using Multi-Scaled ResNet after training. Images used for testing first hide their class names. The class of the input image is then estimated using the CNN model. The accuracy of the CNN is consequently determined by dividing the number of correct predictions by the total number of test images. This network also takes into account and tests fresh raw inputs from diverse people. For this, the input images are first binarized and then scaled to an 86x86 size. By feeding this image to the network, the class label of the input image is then predicted.

4.1 Dataset

There was no complete dataset available. There is a lack of a benchmark dataset for Malayalam character recognition. Hence, it was necessary to create a new dataset for consonant and vowel signs. 44 basic Malayalam characters and 36 compound characters are directly available from the P-ARTS Kayyazhuthu [15] online datastore. For each character, a total of 2000 samples are created. The dataset collected from different individuals is then augmented to increase the number to 2000. The collected dataset is broken down into two in an 80:20 proportion. 1600 samples are labeled for training, and the remaining 400 are for testing. The size of all the samples is normalized to 86×86 . Characters are labeled as CHAR1 to CHAR44 for the basic Malayalam characters. And CHAR45 to CHAR81 for compound characters. The consonant and vowel signs are labeled SYM01 to SYM15. These datasets are shown in Figures 8-10. A total of 19,0000 images are used in the work.

CHAR1	അ	CHAR12	ഇ	CHAR23	എ	CHAR34	ഉ
CHAR2	ആ	CHAR13	ഈ	CHAR24	ഈ	CHAR35	ഊ
CHAR3	ഈ	CHAR14	ഊ	CHAR25	ഊ	CHAR36	ഋ
CHAR4	ഋ	CHAR15	ൠ	CHAR26	ൠ	CHAR37	ൡ
CHAR5	ഌ	CHAR16	഍	CHAR27	഍	CHAR38	എ
CHAR6	ഏ	CHAR17	ഐ	CHAR28	ഐ	CHAR39	഑
CHAR7	ഐ	CHAR18	ഓ	CHAR29	ഓ	CHAR40	ഔ
CHAR8	ഓ	CHAR19	ഔ	CHAR30	ഔ	CHAR41	ഓ
CHAR9	ഔ	CHAR20	ഓ	CHAR31	ഔ	CHAR42	ഐ
CHAR10	ഐ	CHAR21	ഔ	CHAR32	ഐ	CHAR43	ഏ
CHAR11	ഏ	CHAR22	ഐ	CHAR33	ഏ	CHAR44	ഏ

Figure 8. Basic Malayalam characters and their labels

CHAR45	ഏ	CHAR54	ഏ	CHAR63	ഏ	CHAR72	ഏ
CHAR46	ഏ	CHAR55	ഏ	CHAR64	ഏ	CHAR73	ഏ
CHAR47	ഏ	CHAR56	ഏ	CHAR65	ഏ	CHAR74	ഏ
CHAR48	ഏ	CHAR57	ഏ	CHAR66	ഏ	CHAR75	ഏ
CHAR49	ഏ	CHAR58	ഏ	CHAR67	ഏ	CHAR76	ഏ
CHAR50	ഏ	CHAR59	ഏ	CHAR68	ഏ	CHAR77	ഏ
CHAR51	ഏ	CHAR60	ഏ	CHAR69	ഏ	CHAR78	ഏ
CHAR52	ഏ	CHAR61	ഏ	CHAR70	ഏ	CHAR79	ഏ
CHAR53	ഏ	CHAR62	ഏ	CHAR71	ഏ	CHAR80	ഏ

Figure 9. Compound Malayalam characters and their labels

SYM01	ഓ	SYM06	ഔ	SYM11	ഐ
SYM02	ഐ	SYM07	ഏ	SYM12	ഏ
SYM03	ഏ	SYM08	ഏ	SYM13	ഏ
SYM04	ഏ	SYM09	ഏ	SYM14	ഏ
SYM05	ഏ	SYM10	ഏ	SYM15	ഏ

Figure 10. Malayalam consonant and vowel signs and their labels

4.2 Experimental setup

Compute Unified Device Architecture (CUDA) is a parallel computing platform and programming model established by NVIDIA. A Dell Precision Tower 5810 workstation with 64 GB of RAM and a 4 GB DDR5 NVIDIA Quadro M2000 CUDA graphics card is used for this task. The entire work is implemented in MATLAB 2016b. Accuracy is the criterion for evaluation in this case. Accuracy corresponds to testing and real-world images. Hence, the ratio of Malayalam characters with predicted labels that match real labels to the total number of testing images represents the testing accuracy.

$$Accuracy = \frac{\text{Number of correctly predicted images}}{\text{Total number of input images}} \times 100$$

4.3 Quantitative analysis

The five different architectures of the multi-scaled residual network are trained using three classes of characters. The resultant accuracy is tabulated below. Table 2 shows the result of training the multi-scaled Resnet-1 network with different initial learning rates for basic Malayalam characters of 44 classes, consonant characters of 36 classes, and consonant and vowel signs of 15.

From Table 2, it can be inferred that for epochs 2, 3, and 4, accuracy is higher for learning rate 0.01. And slightly decreases as the learning rate increases. Table 2 also shows the result of training the multi-scaled Resnet-2 network with different initial learning rates for different classes of Malayalam characters.

For epochs 2, 3, and 4, accuracy is higher for learning rate 0.01. And slightly decreases as the learning rate increases. Table 3 demonstrates the result of training the multi-scaled Resnet-3 network with different initial learning rates for different classes of Malayalam characters.

For epochs 2, 3, and 4, accuracy is higher for learning rate 0.01. And slightly decreases as the learning rate increases. In the multi-scaled ResNet-4 architecture, the output of the fourth and final residual blocks is connected to the addition block by a skip connection and given to the classifier. Table 4 shows the result of training the same network with different initial learning rates for different classes of Malayalam characters.

For epochs 2, 3, and 4, accuracy is higher for learning rate 0.01. And slightly decreases as the learning rate increases. The multi-scaled ResNet-5 architecture output of the fifth and final residual blocks is connected to the addition block by a skip connection and given to the classifier. Table 5 shows the result of training the same network with different initial learning rates for different classes of Malayalam characters.

Table 2. Comparison of accuracy for different learning rate and epoch for basic, compound Malayalam characters and consonant and vowel signs using Multi-scaled ResNet-1

Epoch	Basic Learning Rate	Basic Malayalam Characters		Compound Characters		Consonant and Vowel Signs	
		Accuracy	Training Time	Accuracy	Training Time	Accuracy	Training Time
2	0.004	92.02%	75 min 23 sec	92.02%	75 min 23 sec	92.30%	30 min 40 sec
	0.006	93.12%	70 min 14 sec	92.62%	70 min 14 sec	92.62%	25 min 11 sec
	0.008	93.59%	69 min 32 sec	93.59%	67 min 32 sec	93.00%	24 min 50 sec
	0.01	95.02%	61 min 23 sec	95.02%	65 min 23 sec	94.30%	20 min 40 sec
	0.02	94.42%	60 min 41 sec	93.22%	63 min 14 sec	92.02%	19 min 41 sec
	0.03	92.38%	60 min 38 sec	92.59%	62 min 32 sec	92.00%	19 min 30 sec
3	0.004	93.32%	72 min 03 sec	93.02%	75 min 23 sec	92.30%	30 min 40 sec
	0.006	93.62%	70 min 11 sec	93.92%	70 min 14 sec	93.24%	29 min 41 sec
	0.008	93.86%	68 min 50 sec	94.59%	67 min 32 sec	93.90%	25 min 30 sec
	0.01	96.42%	61 min 23 sec	96.82%	60 min 43 sec	94.00%	20 min 40 sec
	0.02	94.42%	60 min 41 sec	95.42%	59 min 41 sec	92.02%	19 min 41 sec
	0.03	93.38%	60 min 38 sec	94.38%	59 min 38 sec	92.00%	19 min 30 sec
4	0.004	93.23%	79 min 47 sec	94.16%	76 min 26 sec	92.67%	35 min 12 sec
	0.006	93.98%	78 min 51 sec	94.99%	72 min 14 sec	92.35%	30 min 01 sec
	0.008	95.02%	76 min 40 sec	95.19%	69 min 32 sec	93.11%	29 min 47 sec
	0.01	96.17%	75 min 52 sec	96.69%	78 min 52 sec	94.17%	25 min 12 sec
	0.02	95.17%	72 min 31 sec	96.57%	69 min 31 sec	93.57%	24 min 31 sec
	0.03	94.09%	72 min 16 sec	96.09%	66 min 16 sec	92.09%	23 min 57 sec

Table 3. Comparison of accuracy for different learning rate and epoch for basic, compound Malayalam characters and consonant and vowel signs Malayalam characters using Multi-scaled ResNet-2

Epoch	Basic Learning Rate	Basic Malayalam Characters		Compound Characters		Consonant and Vowel Signs	
		Accuracy	Training Time	Accuracy	Training Time	Accuracy	Training Time
2	0.004	92.52%	73 min 23 sec	92.26%	70 min 52 sec	92.10%	33 min 22 sec
	0.006	93.22%	70 min 53 sec	92.49%	68 min 31 sec	92.42%	24 min 51 sec
	0.008	93.79%	68 min 32 sec	93.00%	65 min 16 sec	93.20%	24 min 23 sec
	0.01	95.12%	61 min 23 sec	93.26%	60 min 52 sec	94.73%	20 min 32 sec
	0.02	94.63%	59 min 41 sec	94.01%	59 min 31 sec	92.05%	19 min 31 sec
	0.03	92.38%	58 min 08 sec	94.00%	59 min 16 sec	92.10%	18 min 10 sec
3	0.004	93.42%	71 min 59 sec	94.07%	65 min 12 sec	92.30%	30 min 40 sec
	0.006	93.72%	70 min 11 sec	94.58%	61 min 01 sec	93.24%	29 min 41 sec
	0.008	93.97%	68 min 50 sec	95.31%	56 min 16 sec	94.90%	25 min 30 sec
	0.01	96.77%	61 min 03 sec	97.47%	60 min 52 sec	95.84%	20 min 40 sec
	0.02	94.11%	60 min 41 sec	96.38%	59 min 31 sec	93.02%	17 min 01 sec
	0.03	93.18%	60 min 38 sec	96.11%	59 min 16 sec	93.00%	17 min 30 sec
4	0.004	93.33%	78 min 47 sec	92.98%	71 min 32 sec	93.82%	35 min 12 sec
	0.006	93.99%	77 min 00 sec	93.03%	69 min 10 sec	93.99%	30 min 01 sec
	0.008	95.42%	76 min 42 sec	94.11%	65 min 16 sec	94.34%	29 min 47 sec
	0.01	96.67%	75 min 12 sec	96.12%	70 min 54 sec	94.45%	21 min 32 sec
	0.02	95.77%	72 min 45 sec	95.52%	69 min 41 sec	94.37%	22 min 17 sec
	0.03	94.00%	72 min 34 sec	95.08%	66 min 38 sec	93.85%	24 min 10 sec

Table 4. Comparison of accuracy for different learning rate and epoch for basic, compound Malayalam characters and consonant and vowel signs Malayalam characters using Multi-scaled ResNet-3

Epoch	Basic Learning Rate	Basic Malayalam Characters		Compound Characters		Consonant and Vowel Signs	
		Accuracy	Training Time	Accuracy	Training Time	Accuracy	Training Time
2	0.004	93.62%	73 min 23 sec	94.42%	61 min 12 sec	93.33%	32 min 22 sec
	0.006	94.12%	71 min 00 sec	95.13%	59 min 59 sec	93.82%	27 min 54 sec
	0.008	94.29%	69 min 02 sec	96.11%	56 min 46 sec	94.09%	26 min 03 sec
	0.01	95.62%	60 min 23 sec	96.96%	52 min 56 sec	95.73%	25 min 21 sec
	0.02	94.03%	54 min 41 sec	96.73%	50 min 28 sec	93.86%	22 min 12 sec
	0.03	93.38%	53 min 08 sec	95.30%	48 min 17 sec	92.34%	19 min 28 sec
3	0.004	94.52%	65 min 59 sec	95.07%	55 min 12 sec	94.21%	26 min 40 sec
	0.006	95.61%	60 min 00 sec	95.18%	51 min 01 sec	95.04%	24 min 41 sec
	0.008	96.97%	54 min 50 sec	97.34%	49 min 16 sec	96.12%	20 min 23 sec
	0.01	99.56%	45 min 10 sec	99.47%	42 min 52 sec	98.40%	18 min 28 sec
	0.02	98.34%	43 min 12 sec	98.49%	46 min 31 sec	97.35%	17 min 12 sec
	0.03	96.68%	42 min 14 sec	97.81%	49 min 16 sec	95.39%	16 min 43 sec
4	0.004	94.13%	79 min 39 sec	95.28%	59 min 32 sec	94.36%	32 min 23 sec
	0.006	95.99%	78 min 30 sec	95.50%	58 min 10 sec	94.76%	30 min 56 sec
	0.008	96.82%	77 min 52 sec	96.71%	56 min 16 sec	95.18%	29 min 19 sec

0.01	97.67%	65 min 12 sec	97.45%	53 min 54 sec	97.84%	25 min 12 sec
0.02	96.17%	62 min 45 sec	96.12%	52 min 41 sec	96.17%	22 min 17 sec
0.03	94.20%	60 min 34 sec	95.68%	50 min 38 sec	95.89%	20 min 10 sec

Table 5. Comparison of accuracy for different learning rate and epoch for basic, compound Malayalam characters and consonant and vowel signs Malayalam characters using Multi-scaled ResNet-4

Epoch	Basic Learning Rate	Basic Malayalam Characters		Compound Characters		Consonant and Vowel Signs	
		Accuracy	Training Time	Accuracy	Training Time	Accuracy	Training Time
2	0.004	92.13%	72 min 13 sec	94.10%	60 min 12 sec	92.03%	31 min 08 sec
	0.006	93.02%	72 min 00 sec	94.23%	58 min 11 sec	92.79%	26 min 23 sec
	0.008	94.10%	68 min 34 sec	95.08%	57 min 32 sec	93.98%	24 min 23 sec
	0.01	94.54%	64 min 43 sec	95.98%	51 min 45 sec	94.89%	23 min 09 sec
	0.02	94.03%	62 min 57 sec	94.98%	49 min 34 sec	92.56%	21 min 00 sec
	0.03	93.69%	59 min 12 sec	92.12%	48 min 12 sec	92.00%	17 min 34 sec
3	0.004	93.12%	63 min 13 sec	94.76%	54 min 50 sec	93.23%	27 min 23 sec
	0.006	94.45%	62 min 34 sec	95.34%	53 min 12 sec	94.45%	25 min 34 sec
	0.008	96.98%	60 min 34 sec	96.12%	49 min 06 sec	95.34%	22 min 56 sec
	0.01	97.45%	57 min 54 sec	97.99%	45 min 54 sec	97.76%	20 min 34 sec
	0.02	95.12%	55 min 06 sec	97.19%	45 min 34 sec	97.09%	18 min 54 sec
	0.03	93.98%	50 min 34 sec	96.09%	48 min 23 sec	94.23%	17 min 23 sec
4	0.004	93.56%	74 min 23 sec	94.00%	58 min 55 sec	93.81%	30 min 23 sec
	0.006	94.34%	72 min 14 sec	94.34%	57 min 35 sec	94.67%	28 min 56 sec
	0.008	95.06%	69 min 08 sec	95.34%	56 min 00 sec	95.23%	28 min 00 sec
	0.01	96.12%	64 min 45 sec	96.98%	52 min 09 sec	96.98%	25 min 51 sec
	0.02	95.97%	62 min 23 sec	95.23%	52 min 45 sec	95.37%	23 min 34 sec
	0.03	95.00%	57 min 12 sec	94.12%	48 min 10 sec	93.19%	22 min 09 sec

Table 6. Comparison of accuracy for different learning rate and epoch for basic, compound Malayalam characters and consonant and vowel signs Malayalam characters using Multi-scaled ResNet-5

Epoch	Basic Learning Rate	Basic Malayalam Characters		Compound Characters		Consonant and Vowel Signs	
		Accuracy	Training Time	Accuracy	Training Time	Accuracy	Training Time
2	0.004	92.09%	71 min 00 sec	94.09%	59 min 12 sec	92.00%	30 min 08 sec
	0.006	93.00%	71 min 23 sec	94.12%	57 min 34 sec	92.56%	25 min 09 sec
	0.008	94.03%	67 min 23 sec	95.01%	56 min 56 sec	93.71%	22 min 03 sec
	0.01	94.34%	65 min 45 sec	95.83%	50 min 51 sec	94.62%	20 min 34 sec
	0.02	94.00%	63 min 35 sec	94.34%	48 min 42 sec	92.39%	18 min 05 sec
	0.03	93.47%	57 min 45 sec	92.02%	47 min 30 sec	91.78%	16 min 39 sec
3	0.004	93.06%	64 min 54 sec	94.46%	50 min 45 sec	93.12%	25 min 13 sec
	0.006	94.23%	63 min 13 sec	95.04%	51 min 23 sec	94.36%	24 min 00 sec
	0.008	95.76%	59 min 12 sec	96.07%	48 min 11 sec	95.27%	21 min 23 sec
	0.01	96.35%	56 min 45 sec	97.53%	44 min 11 sec	97.67%	19 min 54 sec
	0.02	95.07%	53 min 00 sec	97.11%	46 min 51 sec	97.00%	17 min 57 sec
	0.03	93.45%	49 min 48 sec	96.00%	47 min 35 sec	94.15%	16 min 00 sec
4	0.004	93.45%	73 min 13 sec	93.56%	58 min 25 sec	93.12%	31 min 03 sec
	0.006	94.12%	71 min 00 sec	94.07%	56 min 55 sec	94.11%	29 min 51 sec
	0.008	95.00%	68 min 34 sec	95.27%	55 min 34 sec	95.13%	29 min 40 sec
	0.01	96.04%	63 min 52 sec	96.48%	54 min 43 sec	96.45%	26 min 24 sec
	0.02	95.69%	61 min 12 sec	95.12%	51 min 51 sec	95.17%	22 min 04 sec
	0.03	94.70%	56 min 41 sec	94.00%	47 min 30 sec	93.31%	20 min 48 sec

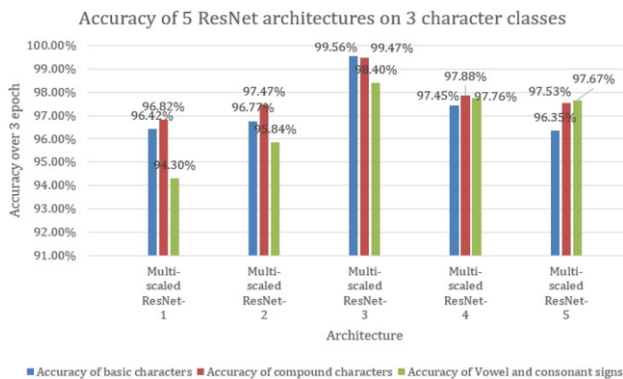


Figure 11. Comparison of different multi-scaled ResNet architectures over 3 epochs

From Table 6, it can be understood that for epochs 2, 3, and 4, accuracy is higher for learning rate 0.01. And slightly decreases as the learning rate increases. The five architectures for the multi-scaled ResNet are created and trained using three categories of dataset, including 44 classes of basic characters, 36 classes of compound characters, and 15 classes of vowel and consonant signs. The architectures differ only by the combination of layers used for the classification block. A comparison of five multi-scaled ResNet architectures is elucidated in Figure 11. It clearly depicts the superiority of multi-scaled ResNet-3 over other multi-scaled residual networks.

As these networks differ only by which level residual block is given to the addition layer to sum up the classification, the accuracy depends on the residual block level. Figure 12 shows

how the residual block level affects the training accuracy of multi-scaled ResNet on basic characters, compound characters, and vowel and consonant signs. Here in the figure, the block level describes the level of features (lower, middle, or final level features) given to the addition layer. It is clear that the residual block-3 that contains middle-level features contributes more to classification accuracy.

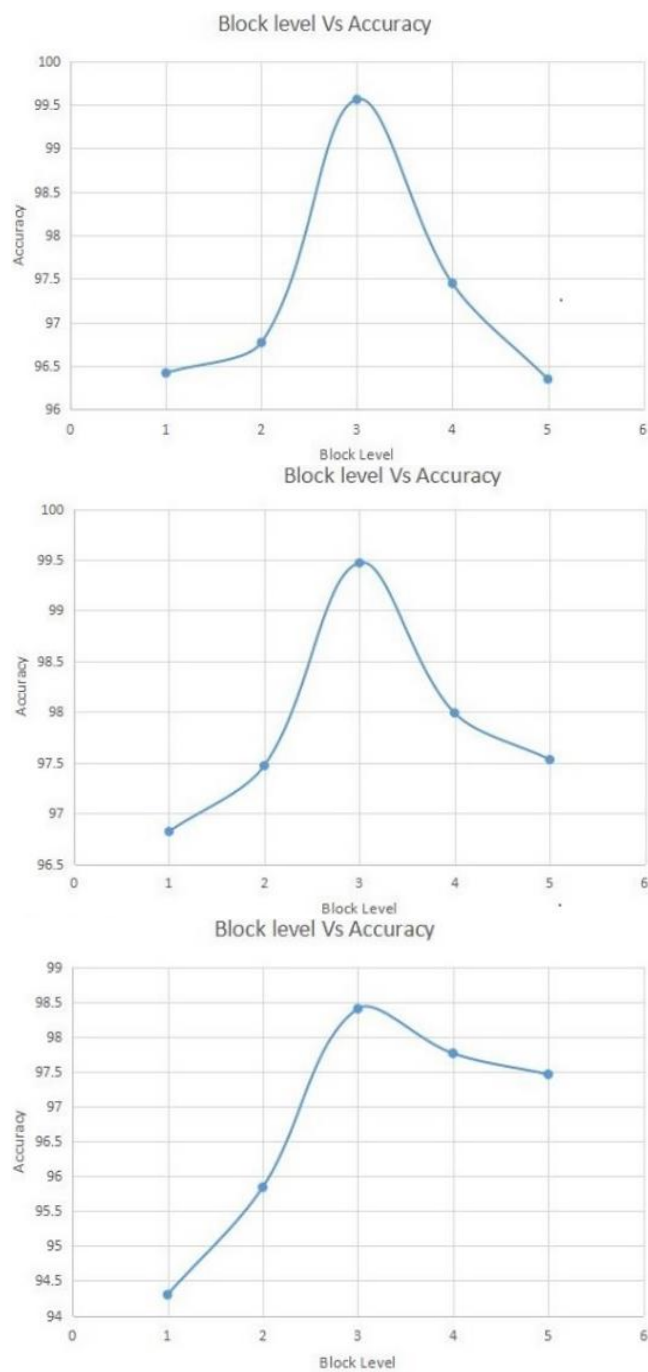


Figure 12. Block level vs accuracy of different multi-scaled ResNet architectures for basic Malayalam characters, compound characters and consonant and vowel signs

It was required to look for the ideal parameters in order to acquire the best network configuration for the multi-scaled ResNet. The network was somewhat modified, and several parameter values and pooling methods were tested. The two most widely used sampling techniques are maximum pooling and average pooling. On the network, both sampling techniques are used. The results are displayed in Table 7.

Table 7. Comparison of pooling strategies

Pooling Method	Testing Accuracy	Training Time in Seconds
Max pooling	99.16%	3320
Average pooling	99.56%	3281

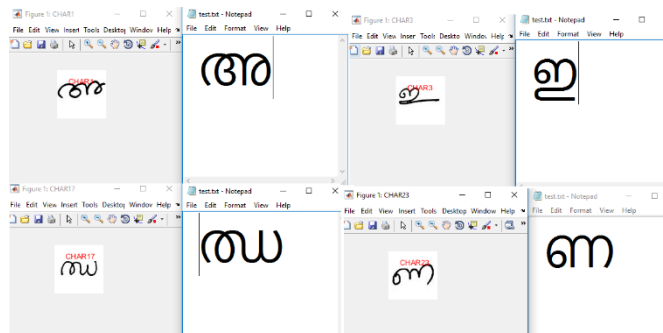


Figure 13. Test data samples and their classification using multi-scaled ResNet-3

The results show that the average pooling method had an accuracy of 99.56% in 3281s of training time, compared to 99.16% accuracy in 3320s of time when using max pooling. The multi-scaled ResNet-3 architecture was tested using a real-time dataset. The data collected from different individuals is tested using the network. Results show an accuracy of 98% was achieved using the network. Some of the raw data samples and their classification are depicted in Figure 13.

The comparison of different Malayalam handwritten techniques on the P-ARTS Kayyazhuthu dataset is presented in Table 8. It is clear that the classification accuracy of the multi-scaled ResNet surpasses all the existing methods.

Table 8. Comparison of different MHCR techniques on P-Arts Kayyazhuth dataset

Sl. No	Technique Used	Accuracy	Traditional/Deep-Learning Approach
1	Chain code histogram	72.1%	Traditional
2	Wavelet Transform	90.25%	Traditional
3	Intensity Variation	92%	Traditional
4	Density based features	81.82%	Traditional
5	Lenet-5	95%	Deep Learning
6	AlexNet	97%	Deep Learning
7	Multi-scaled ResNet	99.56%	Deep Learning

5. CONCLUSION

OCR is utilized in various everyday tasks, such as identifying number plates and automating office processes. Despite the existence of multiple Malayalam HCR systems, the majority of them have failed to reduce the misclassification of handwritten characters that have similar shapes. The methodology is based on deep learning, wherein the machine is trained using extensive datasets of handwritten characters. The system performs automatic feature extraction, eliminating the requirement for user-defined features and relieving the programmer from the task of creating hand-crafted features. A novel dataset has been generated and is currently accessible to the public for further investigation. The network is a deep

residual network that has been improved by using multi-scaled features. The residual network utilizes skip connections to address the issue of gradient vanishing. The suggested approach is highly valuable since it incorporates features from both low- and middle-level layers, as well as the final layer output or processed characteristics, in the classification of data. It significantly enhances the recognition rate. The multi-scaled ResNet was assessed using the same method, resulting in a found accuracy of 99.56%. Different arrangements of the deep network are being evaluated. The network hyperparameters are evaluated, and the ultimate model is chosen based on the optimal combination for achieving high accuracy and efficient training. The system is provided with external test data, and it was discovered that out of 35 input photos, an accuracy of 98% was achieved. The model's high level of precision makes it possible to automate the digitization of governance documents written entirely in Malayalam scripts. Furthermore, as a potential improvement, this project might be expanded to include challenges involving recognition at the level of individual strings and words. Word-level recognition is often regarded as a highly viable approach for digitizing extensive handwritten manuscripts. The design of this model can be modified to enable its usage in recognizing other forms of handwritten script.

REFERENCES

- [1] P-ARTS Kayyvezhuthu Dataset 2017. <https://drive.google.com/drive/folders/Dataset>.
- [2] James, A., Sujala, K., Saravanan, C. (2018). A novel hybrid approach for feature extraction in Malayalam handwritten character recognition. *Journal of Theoretical & Applied Information Technology*, 96(13): 4191-4202.
- [3] James, A., Raveena, P.V., Saravanan, C. (2018). Handwritten Malayalam character recognition using regional zoning and structural features. *International Journal of Engineering & Technology*, 7(4): 4629-4636. <https://doi.org/10.14419/ijet.v7i4.12551>
- [4] Raveena, P.V., James, A., Saravanan, C. (2017). Extended zone based handwritten Malayalam character recognition using structural features. In 2017 Second International Conference on Electrical, Computer and Communication Technologies, Coimbatore, India, pp. 1-5. <https://doi.org/10.1109/ICECCT.2017.8117898>
- [5] Nair, P.P., James, A., Saravanan, C. (2017). Malayalam handwritten character recognition using convolutional neural network. In 2017 International Conference on Inventive Communication and Computational Technologies, Coimbatore, India, pp. 278-281. <https://doi.org/10.1109/ICICCT.2017.7975203>
- [6] Bhagyasree, P. V., James, A., Saravanan, C. (2019). A proposed framework for recognition of handwritten cursive English characters using DAG-CNN. In 2019 1st International Conference on Innovations in Information and Communication Technology, Chennai, India, pp. 1-4. <https://doi.org/10.1109/ICIICT1.2019.8741412>
- [7] Parthiban, R., Ezhilarasi, R., Saravanan, D. (2020). Optical character recognition for English handwritten text using recurrent neural network. In 2020 International Conference on System, Computation, Automation and Networking (ICSCAN), Pondicherry, India, pp. 1-5. <https://doi.org/10.1109/ICSCAN49426.2020.9262379>
- [8] Can, F., Yilmaz, A. (2019). Hybrid handwriting character recognition with transfer deep learning. In 2019 27th Signal Processing and Communications Applications Conference (SIU), Sivas, Turkey, pp. 1-4. <https://doi.org/10.1109/SIU.2019.8806364>,
- [9] Pragathi, M.A., Priyadarshini, K., Saveetha, S., Banu, A. S., Aarif, K.M. (2019). Handwritten tamil character recognition using deep learning. In 2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN), Vellore, India, pp. 1-5. <https://doi.org/10.1109/ViTECoN.2019.8899614>
- [10] Borad, P., Dethaliya, P., Mehta, A. (2020). Augmentation based convolutional neural network for recognition of handwritten Gujarati Characters. In 2020 IEEE International Conference for Innovation in Technology, Bangaluru, India, pp. 1-4. <https://doi.org/10.1109/INOCON50539.2020.9298192>
- [11] Alkawaz, M.H., Seong, C.C., Razalli, H. (2020). Handwriting detection and recognition improvements based on hidden Markov model and deep learning. In 2020 16th IEEE International Colloquium on Signal Processing & Its Applications, Langkawi, Malasia, pp. 106-110. <https://doi.org/10.1109/CSPA48992.2020.9068682>
- [12] Mandal, B., Dubey, S., Ghosh, S., Sarkhel, R., Das, N. (2018). Handwritten Indic character recognition using capsule networks. In 2018 IEEE Applied Signal Processing Conference, Kolkata, India, pp. 304-308. <https://doi.org/10.1109/ASPCON.2018.8748550>
- [13] Gupta, A., Jain, A., Saifi, J., Singh, A.K. (2021). Devanagari character recognition using deep convolution neural networks. In 2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), Greater Noida, India, pp. 927-930. <https://doi.org/10.1109/ICAC3N53548.2021.9725416>
- [14] Nasir, T., Malik, M.K., Shahzad, K. (2021). MMU-OCR-21: Towards end-to-end Urdu text recognition using deep learning. *IEEE Access*, 9: 124945-124962. <https://doi.org/10.1109/ACCESS.2021.3110787>
- [15] Darwish, S.M., Elzoghaly, K.O. (2020). An enhanced offline printed Arabic OCR model based on bio-inspired fuzzy classifier. *IEEE Access*, 8: 117770-117781. <https://doi.org/10.1109/ACCESS.2020.3004286>
- [16] Mohd, M., Qamar, F., Al-Sheikh, I., Salah, R. (2021). Quranic optical text recognition using deep learning models. *IEEE Access*, 9: 38318-38330. <https://doi.org/10.1109/ACCESS.2021.3064019>
- [17] Song, X., Gao, X., Ding, Y., Wang, Z. (2016). A handwritten Chinese characters recognition method based on sample set expansion and CNN. In 2016 3rd International Conference on Systems and Informatics, Shanghai, China, pp. 843-849. <https://doi.org/10.1109/ICSAI.2016.7811068>
- [18] Ashiquzzaman, A., Tushar, A.K., Dutta, S., Mohsin, F. (2017). An efficient method for improving classification accuracy of handwritten Bangla compound characters using DCNN with dropout and ELU. In 2017 Third International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), Kolkata, India, pp. 147-152. <https://doi.org/10.1109/ICRCICN.2017.8234497>

- [19] Wang, D.H., Liu, C.L. (2012). String-level learning of confidence transformation for Chinese handwritten text recognition. In Proceedings of the 21st International Conference on Pattern Recognition, Tsukuba, Japan, pp. 3208-3211.
- [20] Mascarenhas, S., Agarwal, M. (2021). A comparison between VGG16, VGG19 and ResNet50 architecture frameworks for image classification. In 2021 International Conference on Disruptive Technologies for Multi-Disciplinary Research and Applications, Bengaluru, India, pp. 96-99. <https://doi.org/10.1109/CENTCON52345.2021.9687944>