

Malayalam Handwritten Character Recognition Using AlexNet Based Architecture

Ajay James¹, Manjusha J², Chandran Saravanan³

^{1,2}Department of Computer Science and Engineering, Government Engineering College, India

³Department of Computer Science, NIT Durgapur, India

Article Info

Article history:

Received Jun 1, 2018

Revised Sep 6, 2018

Accepted Dec 4, 2018

Keyword:

AlexNet

Convolutional neural network

Handwritten character

recognition

Malayalam character

recognition

Optical character recognition

ABSTRACT

This research article proposes a new handwritten Malayalam character recognition model based on AlexNet based architecture. The Malayalam language consists of a variety of characters having similar features, thus, differentiating characters is a challenging task. A lot of handcrafted feature extraction methods have been used for the classification of Malayalam characters. Convolutional Neural Networks (CNN) is one of the popular methods used in image and language recognition. AlexNet based CNN is proposed for feature extraction of basic and compound Malayalam characters. Furthermore, Support Vector Machine (SVM) is used for classification of the Malayalam characters. The 44 primary and 36 compound Malayalam characters are recognised with better accuracy and achieved minimal time consumption using this model. A dataset consisting of about 180,000 characters is used for training and testing purposes. This proposed model produces an efficiency of 98% with the dataset. Further, a dataset for Malayalam characters is developed in this research work and shared on Internet.

Copyright © 2018 Institute of Advanced Engineering and Science.
All rights reserved.

Corresponding Author:

Ajay James,

Department of Computer Science and Engineering,

Government Engineering College,

Thrissur, Kerala, 680009, India.

Email: ajayjames80@gmail.com

1. INTRODUCTION

Optical Character Recognition (OCR) is the process of recognising handwritten text and printed text. The recognised texts are converted into an encoded format. Four essential steps in character recognition are: a) data pre-processing, b) segmentation, c) feature extraction, and d) classification [1]. The Malayalam script consists of 15 vowels, 36 consonants, and 5 pure consonants as shown in following Figure 1. There are 12 dependent vowels and 144 compound characters as shown in the Figure 2 and Figure 3 respectively. The characters are compounded both vertically and horizontally in Malayalam script.

Convolutional Neural Network is used for character recognition as well as for other purposes like gender prediction [2], finding fidelity content of document digitalisation [3], image style recognition [4], plant classification [5], rail surface defect detection [6], and several other applications. CNN has its advantage in image classification and feature extraction due to its special characteristics such as local connectivity strategy and weight sharing strategy [7] which drastically reduces usage of many parameters.

Handwritten character recognition is a difficult task, because the writing styles of people differs on a large scale. Apart from the writing styles, other factors like noise and occurrence of skew also increases the difficulty. Unavailability of a standard dataset for Malayalam characters increases the complexity. Several models have been used by researchers to improve the accuracy rate of characters by designing new features, integrating different known features, and using multiple classifiers for classification. Various handcrafted

feature extraction methods have been employed to extract features of Malayalam characters, but, all failed to achieve 100% accuracy. Moreover, these conventional methods are time consuming and most of them failed to achieve higher recognition rate.

A survey of different feature extraction and classification methods is conducted and presented here. G. Pirlo et al. [8] presented a fuzzy-zoning based classification method for handwritten characters. The Recognition and reliability rate obtained for this fuzzy based zoning method is 93% and 95% respectively. Dileep Kumar Patel et al. [9] used Discrete Wavelet Transform (DWT) and Euclidean distance transform to find similar feature patterns. They noticed the minimum distance between the characters during classification. The accuracy obtained is 90%. A neural network based off-line English handwritten character and digit recognition was proposed by Ashok Kumar et al. [10]. ShilpyBansal et al. [11] presented a technique called Neighborhood Foreground Pixels Density for handwritten Gurumukhi character feature extraction. Support Vector Machine (SVM) was used for classification. The accuracy achieved was 91.95%. A novel offline Malayalam handwritten text segmentation method was proposed by Shanjana C et al. [12]. SVM is used for character classification and achieved accuracy of 82%.

DurjoySenMaitra et al. [13] 5-layer CNN model like LeNet-5 reported an accuracy of 99.10% for English and multi script characters. A CNN based Bangla digit recognition was proposed by M.A.H Akhand et al. [14], 98.80% accuracy was obtained in the proposed model. A study on different applications of deep learning was conducted by S.M. Sofiqul Islam et al. [15]. The review evaluates two models of CNN, AlexNet, and Visual Geometric Group (VGG-S) in nine different benchmark datasets. Gurumukhi character recognition using particle swarm optimization and a neural network has been presented by JaspreetKaur et al. [16]. Their results showed PSNN outperforms ANN in recognizing Gurmukhi characters with 100% accuracy for basic characters. A novel approach to integrate two different classifiers for the recognition of off-line Arabic handwritten characters was made by Mohamed Elleuch et al. [17], 94.9% accuracy was reported in this method. Shenzhen Gu et al. [18] used AlexNet architecture for the detection of a tennis ball in pictures. Ahmed El-Sawy et al. [19] proposed an off-line Arabic handwritten character recognition system using CNN model. A nine-layer CNN was used, and 5.1% misclassification error was reported for the testing data. Darmatasia et al. [20] proposed a CNN based feature extraction model. The overall accuracy, when tested on ten form documents, was obtained as 83.37%. Pranav P et al. [21] worked on Malayalam handwritten character recognition using CNN. The model was compared with LeNet-5 architecture and obtained a higher accuracy result. 95% accuracy was obtained for this method. SVM based character recognition has been proposed by GauriKatiyar et al. [22]. The method emphasizes the advantage of using SVM classifier for character classification. The results show an average accuracy of 95.74% for uppercase characters and 92.19% for lower case characters. AarthnaMaheshwari et al. [23] worked on handwritten English alphabets using PSO algorithm. ANN was used for recognition and obtained an accuracy of 83.8462%.

From the above surveys of object recognition methods, it is observed that the character recognition has scope for further research and development. An emerging trend which has shown advancement in recognition rate and reduction in time consumption is CNN. Here, an attempt is made to improve the recognition rate and reduce time consumption of Malayalam handwritten simple and compound characters using AlexNet based CNN model. The primary objective of this research work is to develop a suitable model that efficiently extract features of Malayalam characters and classifies the character. The proposed new model is having higher recognition rate with higher accuracy and minimal training time. Further, a new dataset for Malayalam characters is developed for testing proposed new model and shared in the Internet.



Figure 1. Vowels, Consonants and Pure consonants

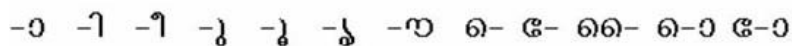


Figure 2. Dependant vowel signs



Figure 3. Compound Characters

2. RESEARCH METHOD

From the research surveys conducted, the drawbacks of handcrafted feature learning from the works of G. Pirlo [8] and Dileep Kumar Patel [9] are identified. In these methods, different features of the characters need to be extracted manually at each iteration, thereby increasing the cost of computation. To avoid the computational cost and to reduce the manual feature computation process, a convolutional feature extraction technique using neural network model to extract and train the network using different features of the same data is introduced. This improves the efficiency of character recognition and reduces the cost of computation which occurred while using manual feature extraction method.

One of the features of CNN is that its recognition rate improves as the training set increases. This requires collecting a large dataset from people of different ages and different sections. Since, Malayalam does not have a large dataset, collecting handwritten character data is the first requirement for this research work. The collected data is then pre-processed in a small amount and augmented to enlarge the size of the dataset. This data is further divided into training and test set in the ratio 80:20, respectively. The primary aim of this research work is to automate the feature extraction by using CNN architecture. The CNN model proposed here consists of a 24-layer architecture, like AlexNet, which is used for extracting features of characters and SVM is used for classifying the output characters.

2.1. Collecting Character data set

In this proposed method, 36 compound characters and 44 basic characters of Malayalam language are considered. Since Malayalam characters do not have a standardised dataset, the 44 primary characters are collected from a new Malayalam character dataset called P-ARTS KAYYEZHUTU [24] dataset. The 36 compound characters are collected manually from people of different age groups. Grids of size 14 x 8 are printed in A4 papers, and character set is written in these columns of the grid. Following Figure 4 shows the architecture of the proposed system.

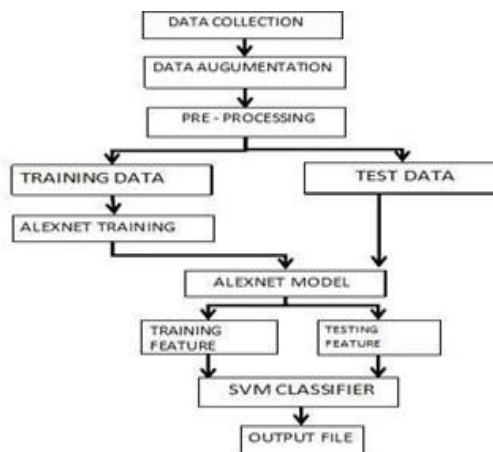


Figure 4. Architecture of the proposed system

2.2. Augmenting and Pre-processing data

The data collected are cropped and characters are separated into 36 different labels. Each character image is of size 86 x 86 pixels. This dataset is augmented to develop a larger dataset. The augmentation is performed by applying affine transformation. The affine transformation includes translation, rotation, and scaling. The affine transformations preserve edge points, shapes, curves etc. So the basic shapes of characters remain the same. Rotation is used for augmenting the dataset, 8 distinct degrees of rotations are used (-2, 2, 4, -4, 8, -8). Following Figure 5(a) shows the character “nna” with eight different degrees of rotation.

After augmentation, these dataset images are preprocessed. Negative of the character image is created to check whether any pixel discontinuities are present in the created dataset. Sample negative characters are shown in the following Figure 5(b). The model also predicted the noise present in the image. The final dataset of about 180,000-character images is obtained for the final training and testing.

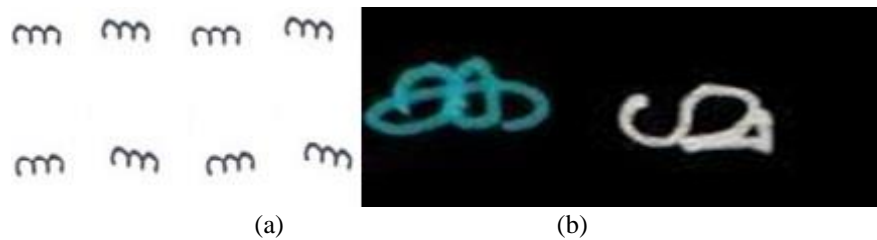


Figure 5. (a) Eight different rotated angles of character, NNA
(b) Sample negative of characters, NKA and GMA

2.3. AlexNet-24 Modeling and Architecture

The next step is building a CNN architecture that effectively trains and tests the data. The different frameworks of CNN are studied to select the best model that helps in solving the recognition problems. The number of layers, type of layers, number of neurons per layer, size of the kernel, stride in each operation, and the type of classifier are selected. Different settings of the network are tested on the various selected network configuration to find the best accuracy. The proposed model used consists of 24 layers with five convolutions, 7 activation layers, five pooling layers, two drop outs, and 5 fully connected layers. The architecture is like that of AlexNet convolutional neural network. The difference lies in the number of ReLU layers and normalization layers used. Also, instead of conventional softmax layer SVM is used for classification. The first layer other than the data input layer is called the convolutional layer. The operation performed by this layer is called convolution operation.

Step 1: The input data is represented in the form of pixels of size 227x227x3.

Step 2: A filter of size $k \times k$ is selected from the image and is convolved over the entire image pixels.

Step 3: The first layer calculates the match of feature to a patch of the image, which multiplies each pixel in a feature by the corresponding pixel value in the image.

Step 4: Add up the answers and divide by total pixel values.

Step 5: In the first layer, if both pixels are black, $(-1) \times (-1) = 1$ or if both pixels are white then $1 \times 1 = 1$, every matching pixel result in a 1.

The third dimension 3 represents the RGB colour channels of the image. In our experiment input size is 227x227x3, kernel $k = 11$, with a stride of 4 and zero-padding. The first convolution operation Conv1 yields a feature map with 96 features. The output size of a convolution layer is calculated as follows:

$$C = \frac{(W-K+2P)}{S} \quad (1)$$

where, C is the output layer size, W is the input height / length, K is the kernel size, and P represent padding and S represent stride value.

A general convolution operation is represented as:

$$h(t)*g(t) = (h*g)(t) \quad (2)$$

Let h and g are two functions, whose convolution is written as $h * g$. This is the integral of the product of two functions after one is reversed and shifted. The convolution is a type of integral transform as shown below:

$$\begin{aligned} h * g(t) &= \int_{-\infty}^{\infty} h(\tau)g(t - \tau)d\tau \\ &= \int_{-\infty}^{\infty} h(t - \tau)g(\tau)d\tau \end{aligned} \quad (3)$$

here, t need not represent time domain and this operation is defined as a weighted average of the function $h(\tau)$ at moment t where $g(-\tau)$ is weighting shifted by amount t . Following Figure 6 shows image of a compound character before and after convolution operation.

The next necessary layer is called activation function layer. This layer mainly helps to introduce some non-linearity to the network. Also, it helps to convert the outputs of neurons in previous layer as the input to the neurons in the next layer that aids in preventing linear mapping. Max-pooling finds the maximum value in a proposed region. A max-pool layer MP1, with a kernel size of 3×3 with stride 2 is used here. This reduces the input size to $112 \times 112 \times 48$. Following Figure 6 shows the sample image before and after the convolution operation. The second convolution layer Conv2, with kernel $5 \times 5 \times 48$, stride 1 and padding 2 produces a feature map of size 256.



Figure 6. Image before and after convolution operation

ReLU is applied after each convolution operation, and the output fed as input to max-pool layer 2. Max-pool MP2 with kernel 3×3 and stride two is supplied as input to next convolution layer Conv3. Following Figure 7(a) shows the sample feature map or activation map obtained for the characters after convolution operation. Following Figure 7(b) shows the different features that are obtained after the convolution operation of Malayalam character “AH”.



Figure 7. (a) Sample feature map obtained for character
(b) Different features of character “AH” extracted by the model

2.4. Classification and Model training

Support Vector Machine (SVM) is used for classification instead of Soft-max classifier of CNN for two main reasons. First, SVM reduces the over fitting problem that occurs in CNN. Drop out layer [25] is used for handling the misclassifications. By appropriately tuning the margin control parameter λ over fitting problem is reduced. The λ is selected using cross-validation methods. Secondly, Softmax classifier is made for 1000 class problem, whereas to deal with lesser number of classes, in this research work 80 classes, thus, SVM is more suitable. The goal of SVM is to find the optimum separating hyper plane that maximizes the margin of training data. A multi-class SVM is used here for training the data. It works as follows:

For a training set $(x_1, y_1), (x_2, y_2) \dots (x_n, y_n)$ with labels y_i in $[1\dots g]$, it finds the solution of the following optimization problem during training.

$$\min \frac{1}{2} \sum_1^g W1 * W2 + \frac{a}{n} \sum_1^n \xi_i \quad (4)$$

$$\text{s.t. for all } y \text{ in } [1\dots g] [x_i \cdot w^{yi}] \geq [x_i \cdot w_y] + 100 * \Delta(y_1, y) - \xi_i$$

$$\text{s.t. for all } y \text{ in } [1\dots g] [x_n \cdot w^{yi}] \geq [x_i \cdot w_y] + 100 * \Delta(y_n, y) - \xi_i$$

where, a is the regularization parameter which adjusts the margin size and training error. $\Delta(y_n, y)$ is the loss function. If $y_n=y$ then it returns 0 else 1.

In general, a two-class classifier is built over a feature vector $\phi(x, y)$ which consists of input features and class of the datum. At test time classifier chooses the class.

$$y = \operatorname{argmax} Y^1 W^1 \quad (5)$$

Finally, the classifier classifies these data using 80 classes. As mentioned, FC7 layer consist of 4096 connections of neurons and about 122,880 different features of each character are used as training features. That is about 4096 x 122,880 features are fed as input to the SVM classifier. Gradient features of each character are considered to calculate the feature vector values. SVM uses feature vector values for classification.

3. RESULTS AND ANALYSIS

The compound character dataset consists of 36-character classes. This dataset is tested against two other architecture other than AlexNet-24 architecture, namely LeNet-5 with seven-layer architecture and architecture with 152 layers called ResNet.

Following Table 1 shows the comparison between three CNN models with two different datasets, raw data, and pre-processed data. The LeNet-5 model produces an accuracy of 85.30% to the raw dataset and 87.54% to the pre-processed dataset. The LeNet-5 architecture consists of seven layers, and it takes a long time for executing the data of about 180,000-character images. The ResNet network consists of 152-layer architecture and there occur the problem of vanishing gradient and degradation. It produces better accuracy rates than LeNet-5 but lesser than the proposed AlexNet-24.

The first parameter is initial learning rate. The learning rate indicates the time required for the neural networks to learn different features. If it is low the time in learning increases, whereas if it is high the learning time is reduced but the prediction accuracy decreases. The learning rate is set as 0.001. Other parameters are mini-batch size and epoch. The above Table 2 shows the different mini-batch size used with corresponding accuracy obtained. Mini-batch size is the amount of data that needs to be used for a single iteration. The value after comparison is considered as 128, which produces the highest accuracy in short time. Epoch determines the number of forward and backward pass required for iterations. Here, different epoch values such as 5, 10, 20, and 30 are tried and this value is set to be 20.

The two different datasets collected are tested separately using the three models, i.e LeNet-5, ResNet, and AlexNet-24. The input images of 36 compound characters are tested and number of classifications is recorded. The compound characters are classified with the basic character data. The following Table 3 shows that the accuracy obtained for both the basic characters collected from P-ARTS Kayyethuthu dataset and tested with those of compound characters, AlexNet-24 shows a higher average accuracy level.

The Mean Square Error (MSE) rate is 1.58%. Following Table 4 shows the characters, their corresponding misclassified classes, and their percentage of misclassification. The confusion matrix is used for evaluation that produces an accurate level of correct recognition and misclassified character percentage. The average accuracy level obtained is above 98.4%. The model is tested several times with different iterations, and a different number of test images are used. A consistent accuracy level of 98.42% is obtained in most of the cases. The experiment is carried out in MATLAB 2017a.

Table 1. Comparison of raw data set with Pre-processed data set

CNN	Raw data accuracy	Pre-processed data accuracy
LeNet-5	85.30%	87.54%
ResNet	96.33%	97.13%
AlexNet-24	96.42%	98.42%

Table 2. Accuracy Vs training time with Different batch size

Mini batch size	Accuracy	Training time (s)
16	98.21	1208
32	98.29	1190
64	98.34	1233
128	98.40	0956
256	98.39	1075

Table 3. Comparison of three methods using both two datasets

CNN Models	Accuracy of data set from P-ARTS	Accuracy of compound characters
	Kayyethuthu	collected
LeNet-5	85.25%	87.54%
AlexNet-24	98.74%	98.42%
ResNet	94.65%	97.13%

Table 4. Misclassification table

CHARACTER CLASS	CHARACTER	ACCURACY	MIS-CLASSIFIED CLASSES	MISS-CLASSIFIED RATE	MIS-CLASSIFIED CHARACTERS
CHAR 1		98.13%	2,22,24	0.27%	
CHAR 2		95.39%	1,13,18,22	3.5%	
CHAR 3		99.7%	4	0.16%	
CHAR 4		97.65%	36,44,33	0.81%	
CHAR 5		99.78%	36	0.1%	
CHAR 6		97.56%	24,7,10	0.27%	
CHAR 7		96.2%	6,10,12	0.54%	

4. CONCLUSION

Recognising Malayalam handwritten characters is one of the challenging research areas. In general, the handwritten characters vary in writing style, from person to person, thus, an automated feature extraction process makes the Malayalam character recognition easier. Compound characters of Malayalam are most prone to misclassification with their basic characters. A dataset consisting of 90,000 basic characters are collected from P-ARTS Kayyethuthu dataset, and 100,000 compound characters are newly developed in this research work manually using different augmentation methods. In this research experiment an overall training accuracy of 99.9% and a testing accuracy of 98.42% are achieved. Several combinations are tried out with various functions and methods in this research work. These methods are selected by accuracy and processing time for executing the data. A comparison of the proposed new model AlexNet-24 architecture is made with two other existing recent models LeNet-5 and ResNet. Also, a complete Malayalam character dataset is newly developed by these basic and compound characters made available through P-ARTS Kayyethuthu drive. The experimental results show that the proposed model AlexNet-24 outperforms other recent existing models LeNet-5 and ResNet.

REFERENCES

- [1] Alex Krizhevsky, Ilya Sutskever, "Geoffrey E. Hinton, ImageNet Classification with Deep Convolutional Neural Networks," *Proceedings of the 25th International Conference on Neural Information Processing Systems*, Lake Tahoe, NV, pp. 1097-1105, Dec 2012.
- [2] Angel Morera, Ángel Sánchez, José Francisco Vélez, and Ana Belén Moreno., "Gender and Handedness Prediction from Offline Handwriting Using Convolutional Neural Networks," *Hindawi Complexity*, vol. 2, 2018.
- [3] LI Pengchao, PENG Liangrui, WEN Juan., "Rejecting Character Recognition Errors Using CNN Based Confidence Estimation," *Chinese Journal of Electronics*, vol. 25(3), 2016.

- [4] Tiancheng Sun, Yulong Wang, Jian Yang, and Xiaolin Hu., "Convolution Neural Networks With Two Pathways for Image Style Recognition," *IEEE Transactions on Image Processing*, vol. 26(9), Sep 2017.
- [5] HulyaYalcin, SalarRazavi., "Plant Classification using Convolutional Neural Networks," *Fifth International Conference on Agro-Geoinformatics (Agro-Geoinformatics)*, IEEE, 2016.
- [6] Lidan SHANG, " Qiushi YANG, Jianing WANG, Shubin LI, WeiminLEI, "Detection of Rail Surface Defects Based on CNNImage Recognition and Classification," *International Conference on Advanced Communications Technology (ICACT)(0975-8887)*, Feb 2018.
- [7] Li Chen, Song Wang, Wei Fan, Jun Sun, Satoshi Naoi., "Beyond Human Recognition: A CNN-Based Framework for Handwritten Character Recognition," *3rdIAPR Asian Conference on Pattern Recognition*, 2015.
- [8] G.Pirlo and D. Impedovo, "Fuzzy-Zoning-Based Classification for Handwritten Characters," *IEEE Transactions on Fuzzy Systems*, vol. 19(4), Aug 2011.
- [9] Dileep Kumar Patel, Tanmoy Som, Sushil Kumar, Yadav, Manoj Kumar Singh., "Handwritten Character Recognition using Multi-Resolution Technique and Euclidean Distance Metric," *Journal of Signal and Information Processing*, vol. 3, pp. 208-214, 2012.
- [10] Ashok Kumar, Pradeep Kumar Bhatia., "Offline Handwritten Character Recognition Using Improved Back-Propagation Algorithm," *Proceedings of 2ndInternational Conference on Emerging Trends in Engineering and Management, ICETEM*, 2013.
- [11] Shilpy Bansal, Mamta Garg, Munish Kumar., "A Technique for Offline Handwritten Character Recognition," *IJCAT International Journal of Computing and Technology*, vol. 1(2), pp. 2415-1521, Mar 2014.
- [12] Shanjana C., Ajay James., "Offline Recognition of Malayalam Handwritten Text," *8th International Conference Interdisciplinarity in Engineering INTER-ENG 2014, Tirgu-Mures Romania*, pp. 22-28, 2014.
- [13] Durjoy Sen Maitra, Ujjwal Bhattacharya, and Swapan K. Parui., "CNN Based Common Approach to Handwritten Character Recognition of Multiple Scripts," *13thInternational Conference on Document Analysis and Recognition (ICDAR)*, 2015.
- [14] M.A.H Akhand, Mahtab Ahamed, M.M. HafizurRahman., "Multiple Convolutional Neural Network Training for Bangla Handwritten Numeral Recognition," *International Conference on Computer & Communication Engineering*, 2016.
- [15] S.M. Sofiqul Islam, Shanto Rahman, Emon Kumar Dey., "Application of Deep Learning to Computer Vision: A Comprehensive Study," *5thInternational Conference on Informatics, Electronics and Vision (ICIEV)*, 2016.
- [16] Jaspreet Kaur, B. S. Dhaliwal, S. S. Gill., "Offline Handwritten Gurmukhi Character Recognition using Particle Swarm Optimized Neural Network," *IJCA Proceedings on International Conference on Advances in Emerging Technology (ICAET) 2016 (12)*, pp. 35-41, Sep 2016.
- [17] Mohamed Elleuch, Rania Maalej, and Monji Kherallah., "A New Design Based-SVM of the CNN Classifier Architecture with Dropout for Offline Arabic Handwritten Recognition," *The International Conference on Computational Science (ICCS)*, vol. 80, 2016.
- [18] Shenzhen Gu, Lu Ding, Yue Yang, and Xinyi Chen., "A New Deep Learning Method Based on AlexNet Model and SSD Model for Tennis Ball Recognition," *IEEE 10thInternational Workshop on Computational Intelligence and Applications, Hiroshima, Japan*, pp. 11-12, Nov 2017.
- [19] Ahmed El-Sawy, Mohamed Loey, Hazem EL-Bakry., "Arabic Handwritten Characters Recognition using Convolutional Neural Network," *WSEASTransactions on Computer Research*, E-ISSN. 2415-1521, vol. 5, 2017.
- [20] Darmatasia, Mohamad Ivan Fanany., "Handwriting Recognition on Form Document Using Convolutional Neural Network and Support Vector Machines (CNN-SVM)," *Fifth International Conference on Information and Communication Technology (IcoICT)*, 2017.
- [21] Pranav P. Nair, Ajay James, C.Saravanan., "Malayalam Handwritten Character Recognition Using Convolutional Neural Network," *International Conference on Inventive Communication and Computational Technologies (ICICCT)*, 2017.
- [22] Gauri Katiyar, Ankita Katiyar, Shabana Mehfooz., "Off-Line Handwritten Character Recognition System Using Support Vector Machine," *American Journal of Neural Networks and Applications*, vol. 3(2), pp. 22-28, 2017.
- [23] Aarthna Maheshwari, Khushboo Shah., "Offline Handwritten Character Recognition using MLPNN and PSO Algorithm," *International Journal of Innovative Science and Research Technology*, vol. 3(2), Feb 2018.
- [24] <https://drive.google.com/drive/folders/0B1eLyjUeuERZWUVIOU9OZm40RHc>
- [25] N. Srivastava, G. Hinton, Alex Krizhevsky, IlyaSutskever, RuslanSalakhutdinov., "Dropout: A Simple way to prevent Neural Network from Overfitting," *Journal of Machine Learning Research*, vol. 15, pp. 1929-1958, 2014.