Review on Hand Written Character Recognition and Speech Recognition Approaches

Asra Masrat
Dept. of Information Technology
K.J Somaiya College of Engineering
Mumbai, India
asra.masrat@somaiya.edu

Deepa Kumari
Dept. of Information Technology
K.J Somaiya College of Engineering
Mumbai, India
deepa.kumari@somaiya.edu

Hardika Gawde
Dept. of Information Technology
K.J Somaiya College of Engineering
Mumbai, India
hardika.gawde@somaiya.edu

***Abstract*—Language is a very essential part of human connection and communication. It enables the flow of thoughts and sharing of ideas and feelings with others. Writing and speaking are the most basic ways to learn a new language. Many platforms utilize different methodologies to facilitate the learning of languages via flashcards, Speech processing, handwritten character recognition, etc. In this paper, we have presented a thorough review of different established or proposed methods for handwritten character recognition, speech processing, Testing mechanisms which could be useful for an AI-enabled language learning platform. The results of the systematic literature review has concluded prominent trends of Deep Learning and Machine Learning approaches which could aid the process of learning a language digitally.**

***Index Terms*—Convolution Neural Network, Hidden-Markov model, Handwritten character recognition, OCR, ResNet, ANN, Warping,Feature Extraction**

I. Introduction

Language gives us the power to share our ideas and feelings with the world. It essentially enables communication and helps forge relations. One participates more effectively and responsibly in a multicultural world if one knows other languages. Recently, there has been a steady rise in the demand for self-learning platforms. Many people find themselves comfortable learning at their own pace, location, and time. In the traditional teaching methods, the instructor teaches the students how to draw a letter and guides them if they make any errors. The same level of sophistication can also be provided on an online platform with various added benefits. However, there is no single platform that boasts the ability to teach multiple Indian regional languages. The number of multilingual people in India is declining due to lack of resources for learning new regional languages. People who travel to rural places in India face many challenges in communicating with the local people. To solve all of these above issues, we are developing an AI-enabled platform to ease the process of learning new languages for all users and bridge the gap between traditional and self-learning methods using an AI-enabled language learning platform.

In order to build this system a thorough literature review of different hand written character recognition systems and speech recognition mechanisms for the purpose of pronunciation checking needs to be conducted. In this paper we have presented a detailed literature survey of existing handwritten character recognition systems, speech recognition mechanisms and other related work which is applicable to this concept.

The paper is assembled as follows. The section II encompasses the literature review and Section III presents the results and conclusions of the review.

II. Literature Review

A. Handwritten Devanagari Character Recognition using Convolutional Neural Network

In this paper, they have described how Artificial Neural Network has taken inspiration from the biological neural network and how Deep Learning like Convolution neural network has solved the problems of variation, deformation, etc under image classification. They also described how OCR systems are available largely for English and not for regional languageslike Devanagiri due to its complex features like curves, seg- ments and edges. Different Devanagiri Character recognition systems right from basics which classified characters using loops and lines to more sophisticated feature extraction and classification technique to CNN-based system which extracts and classifies features automatically were researched by them.Apart from this they also explained various applications of CNN algorithms in face recognition, brain tumor detections, etc.

A detailed explanation of the CNN algorithm and its different layers like Convolution layer, rectification linearunit, pooling layer and fully connected layer was also given by them. Then they compared and explained how different pre-trained architecture like Alexnet, ZF Net, VGG Net, Googlenet and ResNet saves our time to develop CNN algorithm from scratch. Then they also developed a DevanagiriCharacter recognition system where they used Alexnet for their database. They used ISI Kolkata Handwritten Devanagari character and numeral image database, they pre-processed them, resized images to 227x227x3, divided them into training and testing data and applied Alexnet architecture consisting of 8 layers finally, they send it to the classifier to predict the output. They obtained 91.23% accuracy in character

recognition and 100% accuracy in numeral recognition [1]. Lastly, they explained how accuracy can be improved by using preprocessing algorithms on the images before applying a deep learning network.

Archit-ecture	Comparison Factors					
	Year	Layer	Filter Size	Images used	Days and GPU Usage	Error Rate %
Alex net	2012	8	11x11	15 million over 22000 categories	5-6 days /GTX 580 GPU	15.4
ZF Net	2013	8	7x7	1.3 million	12 days /GTX 580 GPU	11.2
VGG Net	2014	19/16	3x3	1.3 million	2 Week/ Nvidia Titan black	7.3
Google Net	2015	22	3x3/5 x5	1.2 million	Within week	6.7
Res Net	2015	152	3x3	1.28 million	8 GPU	3.6

Table. 1. Case Studies of CNN Architectures

B. Recognition of Handwritten Devanagari Character using Convolutional Neural Network

This paper has explored the use of Convolution Neural Network on Devanagari characters. They used multiple ways in CNN architecture in this study and compared their accuracyto identify the optimum match depending on each approach's accuracy. They have discussed different approaches used by various other people using CNN as well as some other methods like ANN, SVM, etc. They have used two different datasets which are Devanagari Handwritten character Dataset Devanagari (Nepali) Handwritten Character Dataset, both of which included a wide range of data and were large enough to ensure correctness.[2] These datasets included consonants, numerical as well as vowels. Their main goal for this system was on Marathi language characters. This study provides a detailed explanation of how the convolution neural network was developed and how it was used to classify the images, as well as a detailed comparison of all four CNN architectures and their respective accuracy.

The CNN algorithm is presented in full, including all of the various layers and their roles. They also investigated alternative CNN architectures with various layer compositions and calculated the accuracy of each architecture. The accuracyin training and testing data with constants and numbers is explained in detail in Architecture 1 and 2. Furthermore, architectures 3 and 4 describe performance using training and testing accuracy, as well as the epochs employed. They then compared the performance outcomes of each architecture in detail. In conclusion, the results show that architecture 2 has a greater testing accuracy of 98% for consonants and 99% for numerals, whereas architecture 4 has a testing dataset accuracy of 99%. The overall accuracy of Devanagari consonants is 98%, vowels are 97.56%, and Devanagari numbers are 99% [2].

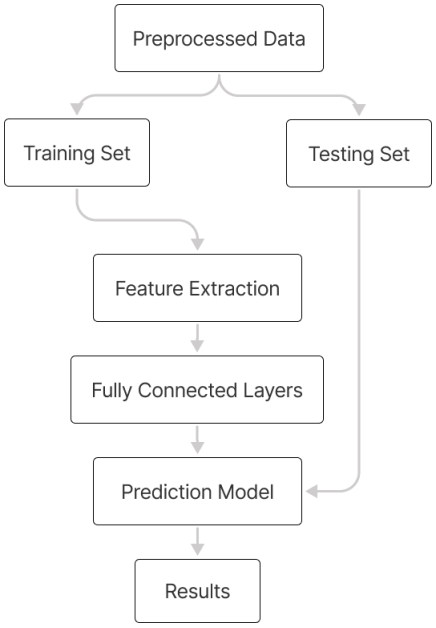

Fig. 1. Flowchart of CNN

C. SVM Classifier for Recognition of Handwritten Devanagari Numeral

Jangid M and colleagues provided an SVM-based framework in this research, which contains standardised pictures of samples with 32, 40, and 48 pixel dimensions. The characteristics for each region were calculated by splitting the standardised photos into 4*4, 5*5, and 6*6 zones, correspondingly. For their experiment, they utilized two types of features: zonal density (ZD) and background directional distribution (BDD). There are nine features in each zone: one density feature and eight background directional distribution features. The zonal density feature is determined by dividing the total number of pixels in each zone by the number of foreground pixels in each zone, which is 64 [14]. The remaining 8 qualities are based on values from the background directional distributionin eight directions. The values corresponding to surrounding background pixels indicated in the appropriate mask for each direction are added to each foreground pixel to produce these directional values. The directional distribution characteristics for all pixels in each zone are totaled for each direction. As a result, the ultimate number of features used for recognition is 144, 225, and 324 for samples of various sizes in ascending order [14]. A support vector machine (SVM) classifier withan RBF kernel is used for classification.

Total data size is 22546, with 18783 for training and 3763 for testing. For changing sizes of samples in ascending order, the best 5-fold cross validation accuracies of training data are

98.76 percent, 98.91 percent, and 98.94 percent, respectively [14]. Inference drawn is Cross validation can be used to ensure that the models are accurate. If strong image processing techniques are not used, the accuracy of the SVM system may be low. The image will be turned into a numpy array in SVM, and many of the details in the image will be lost.. As a result, using SVM over the CNN is inconvenient.

S.No.	Sample Size	No. of Features	Cross validation Accuracy	Classifier parameters
1.	32*32	144	98.94%	C=12, γ=0.35
2.	40*40	225	98.99%	C=12, γ=0.13
3.	48*48	324	99.08%	C=12,γ=0.12

Table. 2. 5-fold Cross validation results

D. Deep Learning Based Large Scale Handwritten Devanagari Character Recognition

The Devanagari Handwritten Character Dataset (DHCD) is a new public Devanagari character image dataset. In addition to the dataset, this research proposed a deep learning architec- ture based on CNN to identify these characters. The DHCD was created by the images of handwritings of multiple people, with physical edits made to the writings. Each picture in the DHCD collection is unique and 32 × 32 pixels in size, with the actual character centred at 28 x 28 pixels. Via this paper, Acharya S. and colleagues suggested a CNN system comprised of two convolutional layers and two subsampling layers. This system suggests using DCNN for feature extraction and classification. It makes use of successive convolutional layers. Consecutive convolutional layers help extract the high-level features prior to pooling [6].

E. Neural Network Based Offline Tamil Handwritten Character Recognition System

To recognise handwritten Tamil letters, this study proposes the use of a multilayer perceptron having one hidden layer. The handwritten character's Fourier Descriptors are a feature retrieved from it. A study was undertaken to ascertain how many hidden layer nodes are required to achieve high back propagation network performance in the recognition of handwritten Tamil characters. A variety of handwriting styles submitted by male and female participants of varying ages were used to train the system. Fourier Descriptors combined with a back propagation network generate a 97 percent recog- nition accuracy for handwritten Tamil letters, according to testdata [11].

F. Bengali Handwritten Character Transformation: Basic to Compound and Compound to Basic Using Convolutional Neural Network

In this paper, they have recommended a system that helps to identify the basic Bengali characters and then turn these characters into compound characters. The approach for recognizing the Bengali handwritten characters is complicated by the letters' peculiar shape and stroke. In Bengali scripts,a compound character is made up of two or more basic characters which make it challenging for kids to understand. There are 50 primary alphabets in the Bengali language. Thereare 11 vowels and 39 consonants among them.

They employed image processing techniques in this work, and all of the photos in the dataset were preprocessed first. The pattern is then identified using a CNN model for both basic and complex characters. If basic characters are supplied into this suggested system, it automatically gives an accurate output, which is a composite of these simple characters [9]. When it receives data of the compound type, it outputs all of the simple characters that the compound character has created. The proposed methodology consists of four steps: dataset acquisition and description, image preprocessing, an explanation of the convolutional neural network training procedure, and finally, the MapReduce strategy for locating the suitable key-value combination. They have optimized the CNN model using the Adam optimizer, which is a powerful optimization tool. The precision rate of 89.20 percent and loss rate of 32.22 percent is obtained [9].

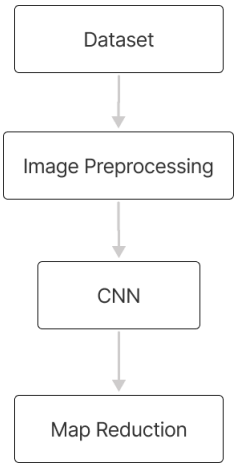

Fig. 2. Methodology for Conversion

G. Handwritten Devanagari Compound Character Recognition Using Legendre Moment: An Artificial Neural Network Approach

The authors of this paper propose a character recognition system from which Legendre moment-based features are extracted and the system is trained and tested using a basic FFNN. Moment functions have been effectively used to a widerange of pattern recognition problems; as a result,they have a proclivity for capturing global traits, making them ideal as feature descriptors[12]. The process picture is normalised to 30 x 30 pixel size and separated into zones, from which structural and statistical features are retrieved.

The suggested system has been trained and evaluated on 27000 handwritten samples obtained from various persons.

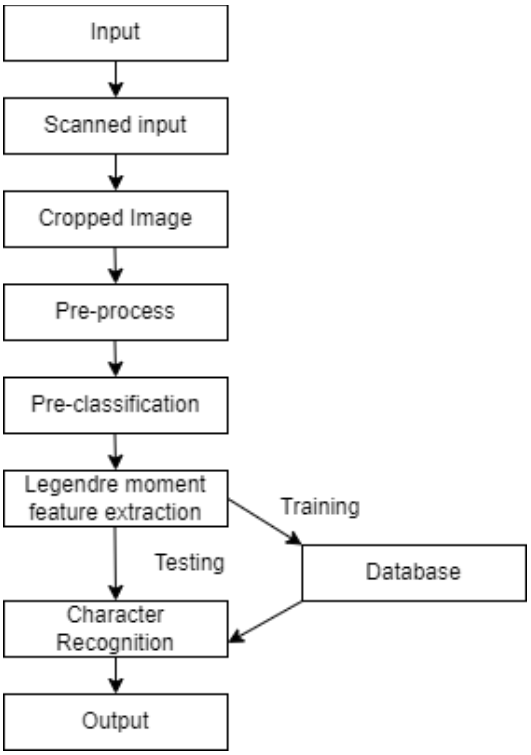

Fig. 3. Recognition System using Legendre Moment Block diagram

They claim to have gotten a better outcome using the Legendre moment-based feature. To relegate the data, they used an ANN. The proposed system has an accuracy of up to 98.25 percent for simple characters and 98.36 percent for all compound characters.

H. Malayalam Handwritten Character Recognition Using Convolutional Neural Network

This study proposes a CNN-based approach for extracting characteristics. Handwritten character recognition is difficult because characters can have a variety of looks depending on the writer, writing style, and background noise. This method is different from the usual method, which requires the use of handcrafted features to discover textual qualities. With a new dataset of six Malayalam characters, they put the network to the test. This method is distinct from the old method, which requires the use of handcrafted features to discover textual qualities.[8]

The input is scanned or pictured with a smartphone first. The Gaussian distribution is used to initialize the kernel weights. Pre-processing, Dataset Creation, Dataset Augmentation, CNN Modelling, Classification, and Testing are the stages of the suggested method. In this case, CNN has proven to be the state-of-the-art technology for some languages, allowing for a higher precision rate for Malayalam characters as well.

I. Support Vector Machine for Handwritten Character Recognition

A technique for recognising unconstrained handwritten Malayalam letters is proposed in this work. This study makes use of a collection of 10,000 character samples from 44 basic Malayalam characters. To train and test the SVM classifier, the accuracy was measured using a discriminatory feature set comprising 64 local and 4 global features and it was 92.24 percent [7]. The purpose of this research is to use a support vector machine classifier to recognise handwritten basic Malayalam letters.

The data was collected from many people in Kerala, encompassing diverse age groups and educational levels, without imposing any limits in order to reflect a wide range of writing styles. In the pre-processing stage, to reduce undesired noise, a median filter with a 3x3 mask is used [7]. Binarization is accomplished using the traditional Otsu technique. Otsu's technique employs the global thresholding method, which has proven to be the most effective on average. Following this, skew detection is done and skews in the image are corrected, line segmentation is done and then each picture is converted to 32x32 using a bicubic interpolation algorithm, with the output pixel value being a weighted average of pixels in the adjacent 4x4 neighbourhood, followed by feature extraction.

A grid of size 4 X 4 is used to extract discriminative information from the dataset characters. A vector of 68 characteristics is used to represent each character. For classification, SVM along with RBF and polynomial kernel has been utilised. The RBF kernel with parameter 0.02 achieves the maximum classification accuracy of 92.24 percent[7]. The research also indicates that improved classification accuracy may be obtained by using more robust features to decrease misclassified characters or by employing a multi stage classifier.

J. Survey paper on Different Speech Recognition Algorithm: Challenges and Techniques

This study has looked at the many and most common voice recognition issues and approaches. They have gone to great lengths regarding the significance of communication and how speaking is the most common way of communication. The categorization of algorithms based on different methodologies is presented in this study. There is also a detailed description of the issues that arise in voice recognition. They have also examined other utterance techniques such as isolated words, linking words, and different utterance styles such as continuous speech and spontaneous speech, as well as different sorts of speaker models. They have described the method of voice recognition in this work. This study provides in-depth information on feature extraction in speech as well as the categorization of different voice recognition algorithms.

In this article, various techniques for feature extraction were discussed and their meaning explained. A detailed explanation about the taxonomy of Speech Recognition is described. They have discussed mainly the 3 approaches used for speech recognition like Acoustic Phonetic, Pattern recognition, artificial intelligence. They have also described the knowledge based

approach used for continuous speech recognition. Different models like HMM, Dynamic Time Warping (DTW), Neural networks are discussed in detail with their working architecture. They have also argued on the neural network that it is capable of solving more complex problematic detection tasks, but these are not as scalable as the HMM.[3]

K. A Study on HMM based Speech Recognition System

This paper provides a high-level explanation of the Hidden Markov Model, a statistical method to voice recognition. The study explains the many factors used to classify voice recognition systems: speaker, speech (isolated word recognition, linked word recognition, continual speech recognition, spontaneous speech recognition), and vocabulary size.

The working of speech recognizer involves inputting the speech into the system, which is then converted to an analog signal which contains the information about the speaker. Preprocessing is performed on the signal for noise reduction and feature extraction. Features that can be extracted are pitch, duration, band energy, spectrum, etc. The method given for feature extraction is MFCC. This step is called acoustic analysis, and it converts the signal into a series of feature vectors. The steps include applying a hamming window, performing Discrete Fourier Transform (DFT), mel frequency warping, taking log of the coefficients obtained, performing Inverse DFT. After successful feature extraction an acoustic model and a language model are used. The statistical representations of sound created in the acoustic model, HMM, is used to represent the distinct sounds that form a word. A phonetic dictionary is also used for the matching process. The n-gram language model is described[4]. It captures the features of the language to predict the next word of the speech sequence. The last step of the recognizer model is searching, which is done to generate the target sequence.

The paper depicts the architecture of the HMM based recognizer and the probabilistic parameters of the model. The HMM model is computationally feasible and easy to use. It's performance is measured using the speed and World-Error-Rate. This paper sets an overall idea on the basis of which an HMM model can be implemented and it's performance can be gauged.

L. An artificial neural network approach to automatic speech processing

This research paper deeply delves into ANN. The authors claim that during their experiments, ANNs trained across long-span temporal patterns performed best on the Wall Street Journal large vocabulary continuous speech recognition. The authors used the word lattice rescoring process to integrate the phoneme posterior probabilities provided by the ANN- based phoneme classifier's output into an existing LVCSR system [10]. These approaches are summarised based on the researchers' laboratory experience and cover a wide range of topics including speech attribute modelling and detection, error analysis, phonetic classification and recognition, bottom-up LVCSR, word lattice (re-)scoring, density prediction for the Hidden Markov Model, and speech tokenization in automatic language recognition. The neural architecture used in all of the experiments in the study paper is a feed-forward multi-layer perceptron with a single or several hidden layers.The paper very well describes the experimental setup and findings of all the above mentioned experiments along with proper pictorial representations and formulae.

M. English Pronunciation Recognition and Detection Based on HMM-DNN

The authors of this study developed a hybrid voice recognition model based on HMM-DNN. This research proposes a new method for detecting pronunciation errors based on HMM-DNN, as well as an efficient classifier based on multi-task learning of neural networks [13]. The binary classifier of each phoneme is included in the same neural network, and the multi-task learning issue is solved using high-dimensional phonetics as the basic input feature.

They employed various frame lengths to verify frame length, accuracy, network training was carried out during training with a single frame input, and 11-frame expansion and 21-frame expansion were evaluated in this paper. The results of error detection for the three phonemes with the biggest sample size and the two phonemes with the underlying network of DNN and HMM-DNN are compared in this study.

III. Results And Conclusion

In this paper, we survey the literature on handwritten character and speech recognition. First, a thorough study on Bangla Handwritten Character Recognition using Autoencoder and DCNN was done. The accuracy of the models for different datasets were below 96%. Upon studying Devanagari Character Recognition using the CNN model, it was understood that an accuracy of 91.23% was achieved. Another paper with CNN model for Devanagari character recognition giving an accuracy of 98% for consonants and 97.56% for vowels was studied. After which a conclusion was reached that Auto-Encoder with DCNN although being efficient, has a lesser accuracy as compared to the other models under study. CNN can be used for various classification problems including Handwritten Character Recognition and various trained architectures can also be used to save time while building CNN algorithms. A proper understanding of how convolution neural networks are implemented and how it is used to classify images which gives high accuracy helped in real-time work on the project.

Different challenges and techniques used for speech recognition were explored such as HMM and ANN. Different approaches and their models have also been reviewed in this paper along with the process of speech recognition. The HMM, a statistical method, is the most commonly used model for speech recognition. The entire process along with core steps involved in implementation of the HMM model has been outlined in this paper.

Algorithm/ Approach Used	Language Implemented	Dataset Size (Samples)	Architecture followed	Number of layers in the Network	Accuracy Achieved
Convolutional Neural Network	*Devanagari (Consonants and Vowels)*	*22,556*	*Alexnet*	*8*	*91.23%*
	Devanagari (Numerals)		*Alexnet*	*8*	*100%*
Convolution Neural Network	*Devanagari (Consonants & Numerals)*	*92,000*	*-*	*6*	*98.76%*
	Devanagari (Vowels)	*2,700*	*-*	*12*	*98%*
Fourier Descriptors with Back propagation network	*Tamil*	*9880*	*Feed forward Multilayer Perceptron (MLP) network with One hidden layer*	*-*	*97%*
Support Vector Machine	*Malayalam*	*10,000*	*SVMlight*	*-*	*92.24%*
Artificial Neural Network Approach (Legendre Moment)	*Devanagari*	*27,000*	*Feed Forward Neural Network*	*2 layers with 108 neurons each*	*98.25%*

Tabel. 3. Comparison of surveyed hand written character recognition systems

REFERENCES

[1] A. Mohite and S. Shelke, "Handwritten Devanagari Character Recognition using Convolutional Neural Network," 2018 4th International Conference for Convergence in Technology (I2CT), 2018, pp. 1-4, doi: 10.1109/I2CT42659.2018.9057991.

[2] I. Dokare, S. Gadge, K. Kharde, S. Bhere and R. Jadhav, "Recogni-tion of Handwritten Devanagari Character using Convolutional Neural Network," 2021 3rd International Conference on Signal Processing and Communication (ICPSC), 2021, pp. 353-359, doi: 10.1109/ICSPC51351.2021.9451716.

[3] Ayushi Y Vadwala, Krina A Suthar, Yesha A Karmakar and Nirali Pandya., "Survey paper on Different Speech Recognition Algorithm: Challenges and Techniques", International Journal of Computer Applications 175(1):31-36, October 2017.

[4] S. Boruah and S. Basishtha, "A study on HMM based speech recog- nition system," 2013 IEEE International Conference on Computational Intelligence and Computing Research, 2013, pp. 1-5, doi: 10.1109/ICCIC.2013.6724147.

[5] Y. Gurav, P. Bhagat, R. Jadhav, S. Sinha, "Devanagari Handwritten Character Recognition using Convolutional Neural Networks," International Conference on Electrical, Communication, and Computer Engineering (ICECCE), 2020.

[6] S. Acharya, A. K. Pant and P. K. Gyawali, "Deep learning based large scale handwritten Devanagari character recognition," 9th International Conference on Software, Knowledge, Information Management and Applications,(SKIMA), 2015.

[7] Jomy John, "Support Vector Machine for Handwritten Character Recognition," KKTM Cognizance - A Multidisciplinary Journal, March 2016, ISSN:2456-4168, arXiv:2109.03081

[8] P. P. Nair, A. James and C. Saravanan, "Malayalam handwritten character recognition using convolutional neural network," 2017 International Conference on Inventive Communication and Computational Technologies (ICICCT), 2017, pp. 278-281, doi: 10.1109/ICICCT.2017.7975203.

[9] S. Chakraborty and S. Paul, "Bengali Handwritten Character Transformation: Basic to Compound and Compound to Basic Using Convolutional Neural Network," 2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST), 2021, pp. 142-146, doi: 10.1109/ICREST51555.2021.9331247.

[10] Sabato Marco Siniscalchi, Torbjørn Svendsen, Chin-Hui Lee, "An artificial neural network approach to automatic speech processing," 2014, Neurocomputing, Volume 140, 2014, Pages 326-338, ISSN 0925-2312.

[11] J. Sutha and N. Ramaraj, "Neural Network Based Offline Tamil Handwritten Character Recognition System," International Conference on Computational Intelligence and Multimedia Applications (ICCIMA 2007), 2007, pp. 446-450, doi: 10.1109/ICCIMA.2007.86.

[12] K. V. Kale, S. V. Chavan, M. M. Kazi and Y. S. Rode, "Handwritten Devanagari Compound Character Recognition Using Legendre Moment: An Artificial Neural Network Approach," 2013 International Symposium on Computational and Business Intelligence, 2013, pp. 274-278, doi: 10.1109/ISCBI.2013.62.

[13] W. Ying, "English Pronunciation Recognition and Detection Based on HMM-DNN," 2019 11th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA), 2019, pp. 648-652, doi: 10.1109/ICMTMA.2019.00148.

[14] M. Jangid and S. Srivastava, "Handwritten devanagari character recognition using layer-wise training of deep convolutional neural networks and adaptive gradient methods," Journal of Imaging, vol. 4, no. 2, p. 41, 2018