ISSN No:-2456-2165

Handwritten Hindi Character Recognition using Multiple Classifiers in Machine Learning

Md Ziaul Haque
Department of CS & IT
Maulana Azad National Urdu University
Hyderabad, India

Mohd Omar
Department of CS & IT
Maulana Azad National Urdu University
Hyderabad, India

Abstract:- Hindi is a national language of India spoken in many states in our countries, like Bihar, Uttar Pradesh, Madhya Pradesh, Jharkhand, and Delhi. The Hindi language is 3rd most popular language globally, which is the script of Devanagari. It consists of 36 primary alphabets and ten digits. We present sophisticated handwritten Hindi character recognition (2HCR) using machine learning techniques to implement Hindi characters and digits. A dataset consists of Ninety-Two Thousand images of 46 different types of characters and digits in the Hindi language segmented from handwritten documents. Nowadays, it has become easy totrain data because of the availability of various algorithms and methodology. We have used many classification algorithms for implementing and improving accuracy. Classification algorithms are Linear-Regression (LR), Logistic-Regression (LGR), Support-Vector-Machine (SVM), Random-Forest (RF), and Naïve-Bayes (NB) to classify the model and improve the accuracy. Handwritten Character Recognition, the area for research is still an active platform because of individuals' different human writing styles, shapes, and sizes. Also, it is used in many applications such asreading license plate numbers, document reading, cheque numbers, postcodes on envelopes, verification of signatures, etc. This system, that we have developed, designed, and implement, has been done using python programming. After completing, we analyzed the performance and accuracy of thesystem.

Keywords:- *Machine Learning, Python, 2HCR, OCR, Hindi Character, Devanagari, LR, LGR, SVM, RF, NB.*

I. INTRODUCTION

Handwritten Character Recognition (HCR), also familiar as Handwritten Text Recognition is the capability of the machine to collect and process the recognition of different handwritten input image data from a source such as paper images, screen images, scanning document devices, etc.[1][2]. In addition, cursive writing is an upward slant, so handwriting will be more difficult to detect[3]. We face many difficulties because of individual people's different styles, shapes, and designs of writing[4]. Also, Handwritten recognition is a highly active research domain where machine learning is utilized[5][6]. Nowadays, HCR is a current technology that will be helpful in the 21st century because HCR technology is responsible for the automatic conversion of handwritten text into computerized text[7][8]. The HCR system is commonly used in various applications such as document reading, mail sorting and verifying, bank processing, and postal card address recognition[9][10]. The application of handwritten recognition is a type of optical recognition. In such a situation, when a person behaves toward another language, they can take a picture of a handwritten image ordocument and forward it to the HCR algorithm[7]. These applications help people to develop reading and writing skills[11]. India is a multi-dialectal country comprised of eighteen official languages. Hindi is the national language of India, and it is used in many Indian states[12]. It is the third most popular language globally and is written in the Devanagari script[13][14]. Our main challenging task isthe devising of the dataset. So, we have introduced a new publicly accessible MNIST dataset of 2HCR[15]. The Hindi Handwritten Character Dataset (2HCD), of Ninety- Two thousand images of 36 Hindi characters and ten digits[16][17]. Before implementing, we have worked on pre-processing because of removing the noisy data, Segmentation is converting input images into individual characters, feature extraction is bringing out of the features using feature extraction techniques, and classification isthe phase to classify the image data using LR, LGR, SVM, RF, NB algorithms, training the image data, and finally testing it to achieve the target[18]. We conducted those experiments on the recognition of Hindi Characters and used a classification algorithm to improve the accuracy of the method. We will use machine learning techniques to implement handwritten recognition in Hindi characters or digits.

Hindi	हि◌न्द◌ी	Meaning	Pronunciation
०	शून्य	0	Shunya
१	एक	1	Ek
२	दो	2	Do
३	तीन	3	Teen
४	चार	4	Char
५	पाँच	5	Paanch
६	छे	6	Che
७	सात	7	Saat
८	आठ	8	Aath
९	नौ	9	Nau

Fig 1 Table of Hindi Numeric values

II. RELATED WORKS

Many organizations have developed and designed methods for handwritten characters and digits as a deep experiment is going on in this particular field. We have studied HCR methods to implement the characters anddigits:

A large amount of research is accomplished on handwritten characters and digit recognition. The research in this field started in 1970. Devanagri text consists of basic, combination, hybrid, and numeral characters[19]. A Handwritten character recognition system has been designed and implemented using a fuzzy logic algorithm based on Hindi Character recognition. They have usedhamming neural network techniques in their approach[20][21]. A creative method for recognizing handwritten Hindi characters approach using Neural Networks algorithm has been developed and designed by using Self Organizing Map(SOM) techniques.[22]. Their system executes the close accurate results but occasionally gets errors if the handwritten character is not fragmented or divided correctly[23]. A few research reports are

III. PROPOSED WORK

In this field, we have discussed how our system has been implemented and how the model of our system works. The modelof our system performs many functions. Out of Ninety-Thousand image data 80% data is to be trained and the rest 20% is to be tested by us.

The given diagram in figure.2 has been designed to achieve accuracy. The dataset preparation phase can be divided into somephases like pre-processing, segmentation, feature extraction, and classification with training and testing data. First, we have to collect the image data from different types of handwritten MNIST datasets in Hindi characters. This dataset consists ofNinety-Two Thousand images of 46 different types of characters and digits. After collecting the image data is accessed from the memory location. After the image is accessed, it has to be sent to available on recognizing handwritten Hindi characters[24].And these Hindi character recognition of handwritten Indian scripts is more complicated than in other languages[25]. 2HCR has the same importance as character recognition for they can be found on cheques, envelopes, Optical Mark Recognition (OMR) sheets,etc[26]. The researcher has presented in this paper that the size of a handwritten text is eccentric and unique for every person and proposed system for recognizing and identifying a different people from their handwriting[27]. In the Devanagari text, there is a lot of old literature available which contains Devanagari characters and digits[28]. The structure of characters in the Hindilanguage varies from character to character [29]. SVM- RBF kernel had the highest accuracy achieved with it was 91.63%, MPL algorithm determined to be the lowest with an accuracy of 86.72%, and 90% accuracy with the KNN classifier[30][26]. Some algorithms have been developed to recognize handwritten characters for languages like English, Hindi, Gujarati, Tamil, French, etc. Many research developers in the area of computer science and machine learning have considerably scrutinized handwriting recognition research work[31].

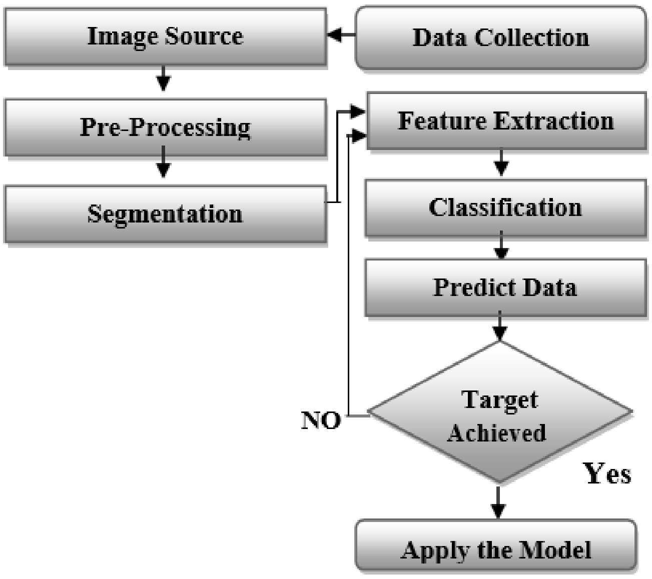

Fig:2 Flow chart of proposed 2HCR Systems

A. *Pre-processing:*

Pre-processing is an initial stage that focuses on improving the input data by reducing unwanted noises and redundancy and the image quality so that we can analyze it better[4][14].

Pre-processing for further process. Pre-processing is the pioneer to start the main work of processing the images. It is done by reducing unwanted noises and redundancy of the image quality. Pre-processing is followed up by another important step called segmentation. Segmentation is the process of converting input images into individual characters. After segmentation, the process to be followed up is feature extraction. Feature extraction is the process by which characteristics of images are extracted by feature extraction methods. Feature Extraction is lastly followed up by the process called classification. We have used the algorithms of machine learning to classify the data for better accuracy. Logistic Regression, Naïve Bayes, SVM, Random Forest. After classifying, predict whether the target's dataaccuracy has been achieved. If the target has been achieved then apply the model to check the accuracy of the results. If not thengo back again to feature extraction and repeat the same process. By doing this the data accuracy will be achieved.

B. *Segmentation:*

The noise-free image is passed to the segmentation after pre-processing and cleaning the document into itscomponents (i.e. paragraph, sentences, words, and letters). Segmentation is converting input images into individual characters[2][8]. Segmentation of an image is in practice for classifying an image pixel.

C. *Feature Extraction:*

After segmentation, the process to be followed upis feature extraction. Feature extraction is the process by which characteristics of images are extracted by feature extraction methods. Feature Extraction is the important phase in

ISSN No:-2456-2165

character recognition. Recognition accuracy mainly depends on extracted features[2][14]. To extract the features of individual characters techniques likeZoning, Histogram, Principle Components Analysis (PCA), and Gradient-based features can be applied[32].

D. Classification:

The classification algorithm is the decision-making stage of recognition in which objects are categorized into classes. The extracted features are used for recognizing characters. The relevant characteristics are classified using a different neural network, Multilayer perceptron, fuzzy logic, Logistic Regression, Naïve Bayes, KNN, SVM, and CNN[2][8][14].

E. Algorithms:

- *Support Vector Machine (SVM):* Support Vector Machine or SVM is one of the most effective and efficient supervised learning methods, which is used for classification. We have applied the SVM algorithm to predict the data. After applying, this model it has achieved the target of 98.88% accuracy.
- *Random Forest:* Random Forest is an effective and efficient machine learning algorithm that belongs to the supervised learning methods used for classification problems. We have applied the Random Forest algorithm to predict the data. After applying, this model it has achieved the target of 97.22% accuracy.
- *Logistic Regression:* Logistic Regression is the most favored supervised learning method. This method is usedto predict the categorical dependent variable using a given set of independent variables. We have applied a logistic regression algorithm to predict the data. After applying,this model it has achieved the target of 95.83% accuracy.
- *Linear Regression:* Linear Regression is the most popular and most effective supervised learning method in predictive analyses is linear regression. We have applied a linear regression algorithm to train the data and predict the value. After applying, this model it has achieved the target of 52.43% accuracy.
- *Naïve Bayes:* The naïve Bayes algorithm is the most effective and efficient supervised learning method. This method is used to solve the classification and prediction problems, which are based on the Bayes theorem. We have applied the Naïve Bayes algorithm to predict the data.After applying, this model it has achieved the target of 52.6% accuracy.

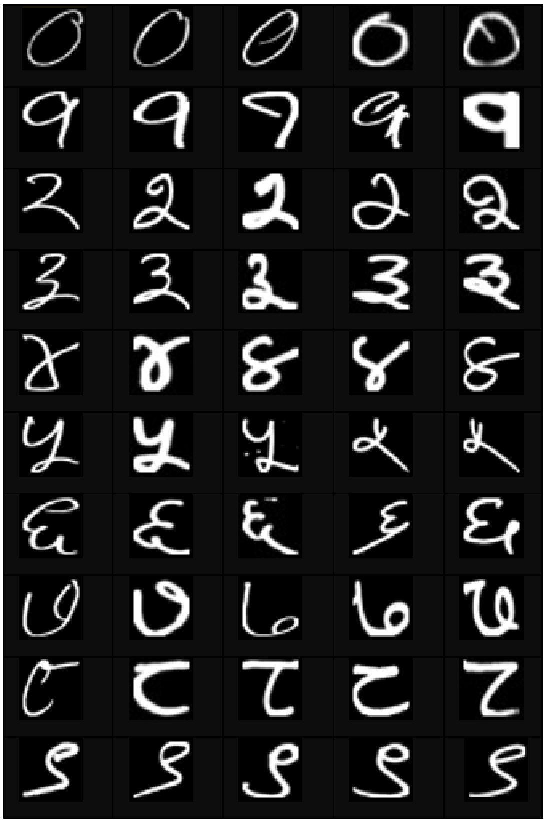

Fig 3. Sample of dataset

IV. RESULTS AND ANALYSIS

The dataset contained 92000 images of 46 different types of Hindi characters and digits. The dataset was randomly shuffled before implementation. The comparison of resultsaccuracy is presented in form of Table I.

Algorithms	**Comparison**[18][26]	**Accuracy**
SVM	96.25%	**98.88%**
RandomForest	98.44%	97.22%
Logistic Regression	86.23%	95.83%
Linear Regression	-	52.43%
Naïve Bayes	89.47%	52.68%

Table 1:- Accuracy achieved by different algorithms

It can be winded up that the performance of the algorithm using 2HCR proposed methods that SVM gives the highestaccuracy with **98.88%.** Generally, we can say that SVM resulted in good performance accuracy on a recognition problem.

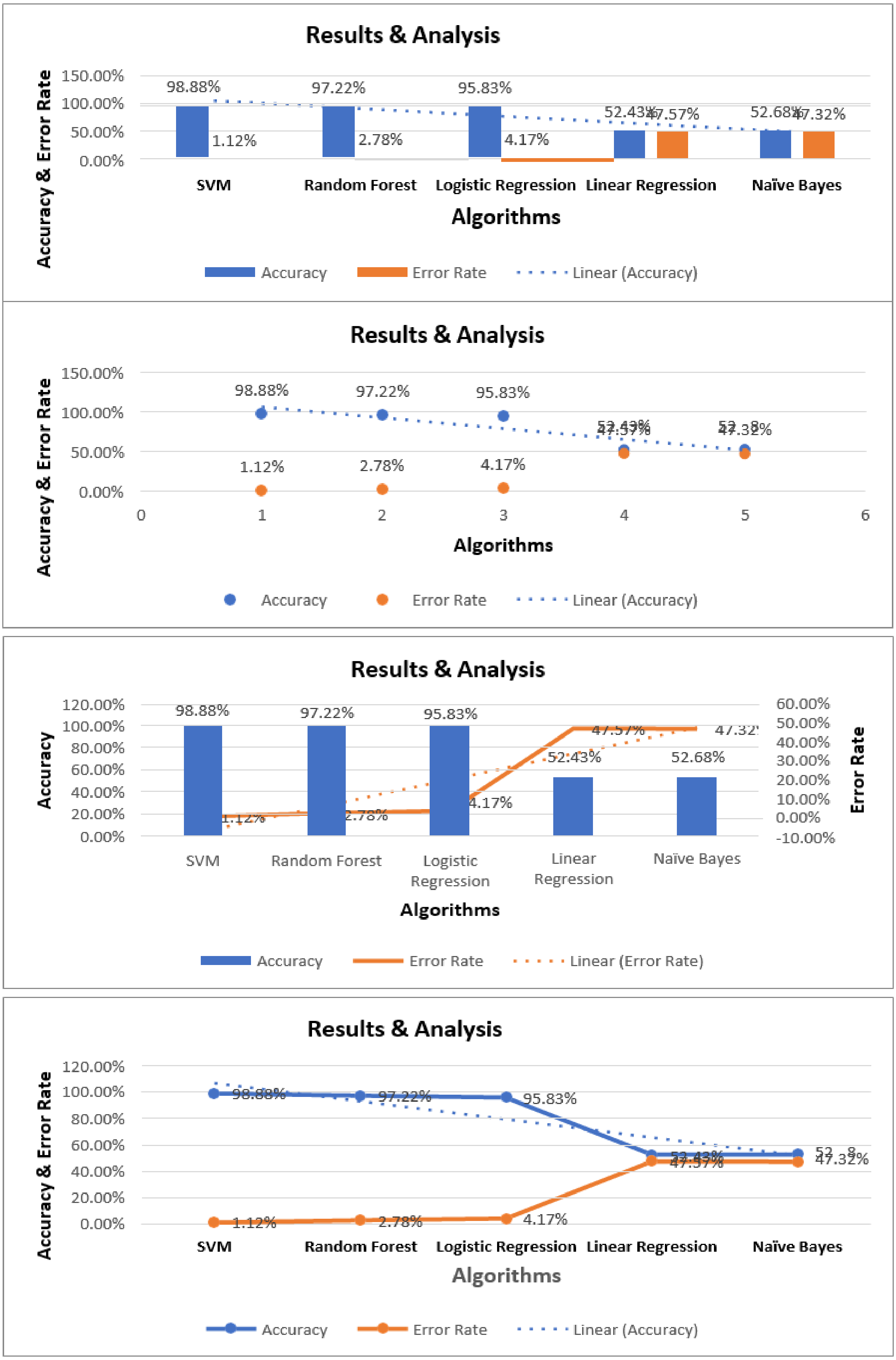

Fig 4 Results and Analysis

V. CONCLUSION

Handwritten Hindi character recognition is a challenging task for the researchers, which is not simply solvable. There are many developments possible in character recognition machines in near future. Nowadays, the machine can only recognize characters and numbers. We can build up the recognition of special characters in the future. The accurate recognition is directly turned on the nature of the written by different users. We presented an MNIST dataset of Handwritten Hindi Characters which is publicly accessible. It consists of 92 thousand images of 36 primary alphabets and ten digits of Devanagari Script. ThisLearning paper presents handwritten Hindi character recognition based on a machine-learning algorithm to improve
accuracy.

FUTURE WORK

A lot of research work on recognition systems is still needed for utilizing new features to improve the current performance. In the future, we will develop a deep learning model which is used for recognizing Hindi words and sentences.

REFERENCES

[1]. G. A. Fink, "Handwriting Recognition," Markov Model. Pattern Recognit., pp. 237–248, 2014, doi: 10.1007/978-1-4471-6308-4_14.

[2]. S. N. R. S and S. Afseena, "Handwritten Character Recognition – A Review," vol. 5, no. 3, pp. 1–6, 2015.

[3]. S. S. Rosyda and T. W. Purboyo, "A Review of Various Handwriting Recognition Methods," Int. J. Appl. Eng. Res., vol. 13, no. 2, pp. 1155– 1164, 2018, [Online]. Available: http://www.ripublication.com

[4]. R. Dixit, R. Kushwah, and S. Pashine, "Handwritten Digit Recognition using Machine and Deep Learning Algorithms," Int. J. Comput. Appl., vol. 176, no. 42, pp. 27–33, 2020, doi: 10.5120/ijca2020920550.

[5]. D. Eshwar Reddy, K. V. Pranathi Naidu, M. Kartheek Srinivas, A. Raheem, and S. Sureddy, "Handwritten character recognition using SVM," Int. J. Adv. Sci. Technol., vol. 29, no. 5, pp. 4001–4007, 2020, doi: 10.55014/pij.v3i2.98.

[6]. D. D. Frp, "+Dqgzulwwhq +Lqgl 1Xphulf &amp;Kdudfwhu 5Hfrjqlwlrq Dqg Frpsdulvrq Ri Dojrulwkpv," pp. 13–16, 2017.

[7]. S. Preetha, I. M. Afrid, K. H. P, and S. K. Nishchay, "Machine Learning for Handwriting Recognition," vol. 4523, pp. 93–101.

[8]. A. Indian, G. K. Vishvidyalaya, K. Bhatia, and G. K. Vishvidyalaya, "A Survey of Offline Handwritten Hindi Character Recognition," 2017.

[9]. V. L. Sahu and B. Kubde, "Offline Handwritten Character Recognition Techniques using Neural Network : A Review," vol. 2, no. 1, pp. 87–94, 2013.

[10]. P. Bojja, N. Sai, S. Teja, G. K. Pandala, and S. D. L. R. Sharma, "Handwritten Text Recognition using Machine Learning Techniques in Application of NLP," Int. J. Innov. Technol. Explor. Eng., vol. 9, no. 2, pp. 1394–1397, 2019, doi: 10.35940/ijitee.a4748.129219.

[11]. J. Memon, M. Sami, R. A. Khan, and M. Uddin, "Handwritten Optical Character Recognition (OCR): A Comprehensive Systematic Literature Review (SLR)," IEEE Access, vol. 8, no. August, pp. 142642–142668, 2020, doi: 10.1109/ACCESS.2020.3012542.

[12]. D. Singh and M. M. M. E. College, "Neural Network Based Handwritten Hindi Character Recognition System," pp. 9–12, 2009.

[13]. V. J. Dongre and V. H. Mankar, "Devnagari Document Segmentation Using Histogram Approach," Int. J. Comput. Sci. Eng. Inf. Technoogy, vol. 1, no. 3, pp. 46–53, 2011, doi: 10.5121/ijcseit.2011.1305.

[14]. A. Gaur, "Handwritten Hindi Character Recognition using K- Means Clustering and SVM," pp. 65–70, 2015.

[15]. ["search Hindi / Devanagari MNIST Data About Dataset," p. 1700.

[16]. S. Acharya, A. K. Pant, and P. K. Gyawali, "Deep learning based large scale handwritten Devanagari character recognition," Ski. 2015 - 9th Int. Conf. Software, Knowledge, Inf. Manag. Appl., 2016, doi: 10.1109/SKIMA.2015.7400041.

[17]. S. K. Singh and A. Khamparia, "Deep Learning Architecture for Large Scale Hand Written Devanagari Character Recognition," vol. 5, no. 10, pp. 222–228, 2018.

[18]. P. Chaudhary, "Handwritten Hindi Character Recognition using Machine ThisLearning and Deep Learning," pp. 48–53.

[19]. S. D. Pande et al., "Digitization of handwritten Devanagari text usingCNN Informatics, vol. 2, no. 3, p. 100016, 2022, doi: 10.1016/j.neuri.2021.100016.

[20]. W. Lu, Z. Li, and B. Shi, "v p j".

[21]. H. Zhan, S. Lyu, U. Pal, and Y. Lu, "CNN-based Hindi numeral string recognition for Indian postal automation," 2019 Int. Conf. Doc. Anal. Recognit. Work. ICDARW 2019, vol. 5, pp. 77–82, 2019, doi: 10.1109/ICDARW.2019.40085.

[22]. P. Banumathi and G. M. Nasira, "Handwritten Tamil character recognition using artificial neural networks," Proc. 2011 Int. Conf. Process Autom. Control Comput. PACC 2011, 2011, doi: 10.1109/PACC.2011.5978989.

[23]. B. V. S. Murthy, "Handwriting recognition using supervised neural networks," Proc. Int. Jt. Conf. Neural Networks, vol. 4, pp. 2899–2902, 1999, doi: 10.1109/ijcnn.1999.833545.

[24]. N. K. Garg, L. Kaur, and M. Jndal, "Recognition of Offline Handwritten Hindi text using middle zone of the words," 2015 IEEE/ACIS 14th Int. Conf. Comput. Inf. Sci. ICIS 2015 - Proc., pp. 325–328, 2015, doi: 10.1109/ICIS.2015.7166614.

[25]. J. M. R. D and A. V. Reddy, "Recognition of Handwritten Characters using Deep Convolutional Neural Network," no. 6, pp. 314–317, 2019, doi: 10.35940/ijitee.F1064.0486S419.

[26]. M. Yadav and R. Purwar, "Hindi handwritten character recognition using multiple classifiers," Proc. 7th Int. Conf. Conflu. 2017 Cloud Comput. Data Sci. Eng., pp. 149–154, 2017, doi: 10.1109/CONFLUENCE.2017.7943140.

[27]. J. Pradeep, E. Srinivasan, and S. Himavathi, “Neural network based handwritten character recognition system without feature extraction,” 2011 Int. Conf. Comput. Commun. Electr. Technol. ICCCET 2011, pp. 40–44, 2011, doi: 10.1109/ICCCET.2011.5762513.

[28]. Y. Gurav, P. Bhagat, R. Jadhav, and S. Sinha, “Devanagari Handwritten Character Recognition using Convolutional Neural Networks,” 2nd Int. Conf. Electr. Commun. Comput. Eng. ICECCE 2020, no. June, pp. 1–6, 2020, doi: 10.1109/ICECCE49384.2020.9179193.

[29]. N. Singh, “An Efficient Approach for Handwritten Devanagari Character Recognition based on Artificial Neural Network,” 2018 5th Int. Conf. Signal Process. Integr. Networks, SPIN 2018, pp. 894–897, 2018, doi: 10.1109/SPIN.2018.8474282.

[30]. A. Sahu and S. N. Mishra, “Odia handwritten character recognition with noise using machine learning,” Proc. - 2020 IEEE Int. Symp. Sustain. Energy, Signal Process. Cyber Secur. iSSSC 2020, pp. 20–23, 2020, doi: 10.1109/iSSSC50941.2020.9358804.

[31]. I. Khandokar, M. Hasan, F. Ernawan, S. Islam, and M. N. Kabir, “Handwritten character recognition using convolutional neural network,” J. Phys. Conf. Ser., vol. 1918, no. 4, 2021, doi: 10.1088/1742-6596/1918/4/042152.

[32]. M. Agarwal, V. Tomar, and P. Gupta, “Handwritten Character Recognition using Neural Network and Tensor Flow,” no. April, pp. 1445–1448, 2019, doi: 10.35940/ijitee.F1294.0486S419.