



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Issue 3, March 2023

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.423

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com

Handwritten Character Recognition

Harsh Shrishrimal, Ganesh Gajengi, Kunal Gole, Pratik Dudhagwali, Prof.Ruchira.P.Pathak

Dept. of Computer Engineering, Smt. Kashibai Navale College of Engineering, Pune, Maharashtra, India

ABSTRACT: Handwritten character recognition is an important task in the field of pattern recognition and machine learning. The goal of this task is to accurately recognize and classify handwritten characters into their respective categories. In recent years, machine learning techniques such as convolutional neural networks (CNNs) have shown significant improvements in handwritten character recognition.

The process of handwritten character recognition involves several steps including preprocessing, feature extraction, and classification. In the preprocessing step, the input image is normalized and binarized to remove noise and enhance the contrast of the image. Feature extraction is performed on the preprocessed image to extract important features such as the shape and size of the characters. These features are then fed into a machine learning algorithm such as a CNN for classification. CNNs have been shown to be highly effective in character recognition due to their ability to learn complex patterns and relationships between features. They are capable of achieving high accuracy rates in recognizing handwritten characters. However, the performance of the CNN is dependent on the quality of the input data and the choice of parameters used in the training process.

Handwritten character recognition has various practical applications such as recognizing postal codes, sorting mail, and recognizing signatures on legal documents. It has also been applied in the field of forensic science for identifying handwriting on evidence.

In conclusion, the use of machine learning techniques such as CNNs has shown significant improvements in handwritten character recognition. The accurate recognition of handwritten characters has various practical applications and continues to be an active area of research in the field of pattern recognition and machine learning.

I. INTRODUCTION

Handwritten character recognition is the process of identifying and classifying handwritten characters or symbols. This task has been the focus of research for several decades, and it has many practical applications, including automatic mail sorting, document digitization, and signature verification. Handwritten character recognition is a challenging problem because of the high variability in handwriting styles and the complex shapes of handwritten characters.

Machine learning has emerged as a powerful tool for solving this problem. In recent years, deep learning techniques, such as convolutional neural networks (CNNs), have achieved state-of-the-art results on benchmark datasets for handwritten character recognition. The success of these methods can be attributed to their ability to automatically learn relevant features from large datasets of labeled handwriting images.

This technology has broad implications in various fields, including finance, government, and healthcare. In finance, it can be used for automated check processing, while in the government, it can be used for processing postal codes, sorting mail, and recognizing signatures on legal documents. Additionally, in healthcare, it can be used to digitize handwritten medical records, enabling doctors and healthcare professionals to access critical patient information easily.

This paper aims to provide an overview of the current state-of-the-art methods for handwritten character recognition using machine learning. We will discuss the key challenges involved in this task, the various techniques used for preprocessing and feature extraction, and the different classification algorithms used for recognition. Finally, we will review some of the practical applications of this technology and the future directions of research in this field.

II. LITERATURE SURVEY

[1] Pal and Singh utilized multilayer perceptron (MLP) for recognizing handwritten English characters and achieved accuracy up to 94% and improved computation time for training the dataset.

[2] Dutt and Dutt demonstrated multilayer CNN using Keras and Theano libraries

which attained 98.7% recognition accuracy on MNIST dataset.

[3] Ghosh and Magharidid comparative study on three neural network approaches demonstrating that DNN was the best algorithm with 98.08% accuracy. However, every neural network has some error rate due to similarity in digit shape (e.g. 3 and 8 and 6 and 9).

[4] Hamid have performed handwritten digit recognition over MNIST dataset using CNN, SVM (Support Vector Machines) and KNN (K-Nearest Neighbour) classifiers. In their work, KNN and SVM predicted the outcomes correctly on datasets but Multilayer perceptron fail to recognize the digit 9 due to non-convex function as it gets stuck in the local minima. It was concluded that the accuracy would improve by using CNN with Keras.

[5] Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review Scientific Figure Research Gate

LeNet-5 Architecture we used in this project was taken from ref. which is going to form core part of our model.

[6] Bag et al propose a method to improve classification performance on Bangla Basic characters using topological features derived from the convex shapes of various strokes.

[7] Das et al. describes handwritten Bangla Character recognition using a soft computing paradigm embedded in a two pass approach. More specifically highly misclassified classes were combined to form a single group in the first pass coarse classification. In the second pass, group specific local features were identified using Genetic Algorithm based region selection strategy to classify the appropriate class from the groups formed in the earlier pass. They used two different sets of features – a) Convex hull based features b) Longest run based features with Support Vector Machines (SVM), a well known classifier for this purpose. They reported a recognition accuracy of 87.26% on a dataset of handwritten Bangla characters consisting of basic characters, compound characters and modifiers.

[8] Sarkhel et al. approached the issue from a perspective of multi-objective based region selection problem where the most informative regions of character samples were used to train SVM classifiers for character recognition. Two algorithms for optimization, specifically a Non-dominated Sorting Harmony Search Algorithm and Non-dominated Sorting Genetic Algorithm II were used to select the most informative regions with the objective of minimal recognition cost and maximal recognition accuracy. A recognition accuracy of 86.65% on a mixed dataset of Bangla numerals, basic and compound characters was reported.

[9] Das et al. described a Quad tree based features used for recognition of 55 frequently occurred compound characters covering 90% of the total of compound characters in Bangla using an Multilayer Perception (MLP) classifier.

[10] Pradeep et al. proposed neural network based classification of handwritten character recognition system. Each individual character is resized to 30 X 20 pixels for processing. They used binary features to train neural network. However such features are not robust. In postprocessing stage, recognized characters are converted to ASCII format. Input layer has 600 neurons equal to number

III. PROBLEM STATEMENT

The problem statement for building a handwritten character recognition system using machine learning is to accurately recognize and classify handwritten characters from a given dataset. This is a challenging problem due to the high variability in handwriting styles, sizes, and orientations.

Additionally, there are several technical challenges that need to be addressed, such as:

Data quality: The quality of the dataset used for training and evaluation can significantly affect the performance of the machine learning model. The dataset must be large enough and diverse enough to ensure that the model can generalize to new, unseen data.

Feature selection: The selection of relevant features to extract from the images is a critical step in the recognition process. Choosing the right features can significantly improve the model's accuracy, while selecting irrelevant or redundant features can negatively impact its performance.

Classification algorithm: The choice of classification algorithm can significantly impact the accuracy and computational efficiency of the system. Different algorithms have their own strengths and weaknesses, and selecting the right one depends on the specific application requirements.

Computational resources: Training and evaluating machine learning models can be computationally intensive, and may require significant resources such as computing power, storage, and memory.



The goal of building a handwritten character recognition system using machine learning is to overcome these technical challenges and develop an accurate and efficient system that can recognize and classify handwritten characters in real-world applications

IV. PROPOSED SYSTEM

The methodology for handwritten character recognition using machine learning typically involves several key steps. These steps are:

Data collection: The first step is to collect a large dataset of handwritten characters, with each character labeled with its corresponding class or category. The dataset should have sufficient variability in handwriting styles, sizes, and orientations to ensure that the machine learning model can generalize to new, unseen data.

Preprocessing: The collected data is preprocessed to prepare it for feature extraction and classification. Preprocessing steps may include noise reduction, image enhancement, and normalization to ensure that the input data is consistent and usable for the machine learning algorithm.

Feature extraction: The next step is to extract relevant features from the preprocessed images. Features may include shape, size, orientation, and other characteristics of the handwritten characters. Feature extraction is a crucial step, as the performance of the machine learning algorithm depends on the quality and relevance of the features selected.

Classification: Once the features are extracted, a machine learning algorithm is trained to classify the characters based on the extracted features. Popular classification algorithms for handwritten character recognition include k-nearest neighbors (KNN), support vector machines (SVM), and convolutional neural networks (CNNs).

Evaluation: The performance of the machine learning model is evaluated using various metrics, such as accuracy, precision, and recall. The model's performance is compared to that of other models or benchmarks to determine its effectiveness in recognizing handwritten characters.

Deployment: Once the model is trained and evaluated, it can be deployed in real-world applications to recognize and classify handwritten characters in new data. The model may require further refinement or fine-tuning based on the specific application requirements.

In conclusion, the methodology for handwritten character recognition using machine learning involves data collection, preprocessing, feature extraction, classification, evaluation, and deployment. Each step is critical to the overall performance of the machine learning model and its effectiveness in recognizing handwritten characters in real-world applications.

DIAGRAM OF PROPOSED SYSTEM

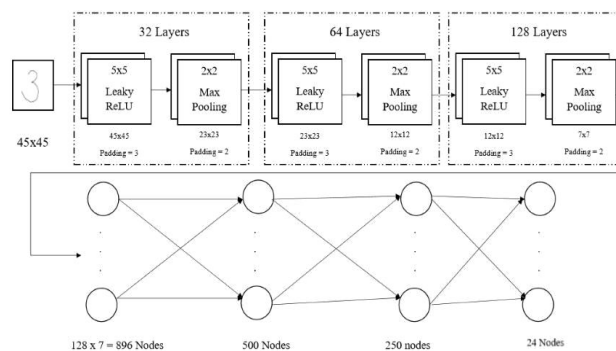


Fig. 1. Proposed System

V. ALOGORITHM USED

Several machine learning algorithms can be used for building a handwritten character recognition system. The choice of algorithm depends on the specific requirements of the application, as well as the size and complexity of the dataset. Some popular algorithms for building a handwritten character recognition system include:

k-NearestNeighbors (k-NN): k-NN is a simple and effective algorithm that is often used for classification tasks. In handwritten character recognition, k-NN can be used to classify a new image by comparing it to the k most similar images in the training dataset.

Support Vector Machines (SVM): SVM is a powerful algorithm for binary classification tasks. In handwritten character recognition, SVM can be used to classify each character as belonging to one of two classes: either the character is present in the image or it is not.

Random Forests: Random forests are an ensemble learning method that can be used for classification tasks. In handwritten character recognition, random forests can be used to classify each character by constructing multiple decision trees and combining their predictions.

Convolutional Neural Networks (CNNs): CNNs are a type of neural network that are highly effective for image recognition tasks. In handwritten character recognition, CNNs can be used to automatically learn relevant features from raw image data, which can significantly improve the accuracy of the system.

Overall, the choice of algorithm depends on the specific requirements of the application, as well as the size and complexity of the dataset. It is important to evaluate the performance of each algorithm and select the one that provides the best balance between accuracy and computational efficiency.

VI. CONCLUSION

Handwritten character recognition using machine learning is a challenging but important problem with numerous applications, including optical character recognition (OCR), signature verification, and document analysis.

The methodology for building a handwritten character recognition system typically involves several key steps, including data collection, preprocessing, feature extraction, classification, evaluation, and deployment. Each step is critical to the overall performance of the system and must be carefully optimized to achieve the best results.

There are several machine learning algorithms that can be used for building a handwritten character recognition system, including k-Nearest Neighbors (k-NN), Support Vector Machines (SVM), Random Forests, and Convolutional Neural Networks (CNNs). The choice of algorithm depends on the specific requirements of the application and the size and complexity of the dataset.

Despite the challenges, advances in machine learning and image processing technologies have led to significant improvements in the accuracy and efficiency of handwritten character recognition systems. With further development and refinement, these systems are expected to become even more accurate and useful in a wide range of applications.

REFERENCES

- [1] PalA,SinghD(2010)Handwrittenenglishcharacterrecognitionusingneuralnetwork.IntJComputSciCommum1(2):141-144
- [2] Dutt A, Dutt A (2017) Handwritten digit recognition using deep learning. Int J AdvResComputEngTechnol6(7):990-997
- [3] Ghosh MMA, Maghari AY (2017) A comparative study on handwriting digit recognitionusing neural networks. In: International conference on promising electronic technologies, pp77-81
- [4] Hamid,N.B.A.,Sjarif,N.N.B.A.:HandwrittenrecognitionusingSVM,KNNandneuralnetwork
- [5]. U. Pal, B. B. Chaudhuri, "Indian Script Character recognition: A survey", Pattern Recognition, vol. 37, pp. 1887-1899, 2004.
- [6]. Dholakia, K., A Survey on Handwritten Character Recognition

Techniques for various Indian Languages, International Journal of Computer Applications, 115(1), pp 17–21, 2015.

- [7]. Nafiz Arica, Fatos T. Yarman-Vural, “An Overview of Character Recognition Focused On Off-line Handwriting”, C99-06-C-203,IEEE, 2000.
- [8]. R. J. Ramteke, S. C. Mehrotra, “Recognition of Handwritten Devnagari Numerals”, International Journal of Computer Processing of Oriental Languages, 2008.
- [9]. Harikesh Singh, R. K. Sharma, “Moment in Online Handwritten Character Recognition”, National Conference on Challenges & Opportunities in Information Technology (COIT-2007) Gobindgarh. March 23, 2007.
- [10]. Aggarwal, A., Rani, R. and Dhir, R. (2012), “Handwritten Devanagari character recognition using Gradient features”, International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 2, 5, pp. 85-90.
- [11]. Shah, M. and Jethava, G. B. (2013), “A literature review on handwritten character recognition”, International Journal of Indian Streams Research Journal, Vol. 3, 2, pp. 1-19.
- [12]. Das, N., Das, B., Sarkar, R., Basu, S., Kundu, M. and Nasipuri, M. (2010), “Handwritten Bangla Basic and Compound Character recognition using MLP and SVM classifier”, Journal of Computing, Vol. 2, 2, pp. 109- 115.
- [13]. Vaidya, S.A. and Bombade, B.R. (2013), “A Novel Approach of Handwritten Character Recognition using Positional Feature Extraction”, International Journal of Computer Science and Mobile Computing, Vol. 2, 6, pp. 179-186.
- [14]. Bag, S., Bhowmick, P., Harit, G., 2011. Recognition of Bengali handwritten characters using skeletal convexity and dynamic programming, in: Emerging Applications of Information Technology (EAIT), 2011 Second International Conference on, pp. 265–268.
- [15]. Das, N., Sarkar, R., Basu, S., Saha, P.K., Kundu, M., Nasipuri, M., 2015. Handwritten bangla character recognition using a soft computing paradigm embedded in two pass approach. Pattern Recognition 48, 2054 – 2071.
- [16]. Sarkhel, R., Das, N., Saha, A.K., Nasipuri, M., 2016. A multiobjective approach towards cost effective isolated handwritten bangla character and digit recognition. Pattern Recognition 58, 172–189.
- [17]. Das, N., Basu, S., Sarkar, R., Kundu, M., Nasipuri, M., Basu, D., 2009. Handwritten bangla compound character recognition: Potential challenges and probable solution, in: IICAI, pp. 1901–1913.
- [18]. J. Pradeepa, E. Srinivasana, S. Himavathib, "Neural Network Based Recognition System Integrating Feature Extraction and Classification for English Handwritten", International journal of Engineering, Vol. 25, No. 2, pp. 99-106, May 2012.
- [19]. Ashutosh Aggarwal, Rajneesh Rani, Renu Dhir, "Handwritten Devanagari Character Recognition Using Gradient Features", International Journal of Advanced Research in Computer Science and Software Engineering (ISSN: 2277-128X), Vol. 2, Issue 5, pp. 85- 90, May 2012.
- [20]. Imaran Khan Pathan, Abdulbari Ahmed Bari Ahmed Ali, Ramteke R.J., "Recognition of offline handwritten isolated Urdu character ", International Journal on Advances in Computational Research, Vol. 4, Issue 1, pp. 117-121, 2012
- [21]. Pritpal Singh, Sumit Budhi raja, “Handwritten Gurumukhi Character Recognition using Wavelet Transforms”, International Journal of Electronics, Communication & Instrumentation Engineering Research and Development, Vol. 2, Issue 3, pp. 27-37, Sept 2012.
- [22]. Pal U, Roy RK, Roy K, Kimura F (2012) Multi-lingual city name recognition for Indian postal automation. In: 2012 International conference on frontiers in handwriting
- [23]. Pal S, Alireza A, Pal U, Blumenstein M (2012) Multi-script offline signature identification. In: 12th International conference



INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.118



ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



www.ijirset.com

Scan to save the contact details