
CNN BASED HANDWRITTEN TEXT RECOGNIZER

Jaya Patil*¹, Sourabh Dalal*², Dheeraj Joshi*³,

Ritik Mahajan*⁴, Anup Patil*⁵

*^{1,2,3,4,5}, Department of Computer Engineering, Shram Sadhna Bombay Trust's College of Engineering and Technology, Bambhori, Jalgaon, Maharashtra, India.

ABSTRACT

The project's goal is to create OCR software that can recognise Marathi text. OCR, or optical text recognition, converts typewritten or handwritten images into text that may be edited by a computer by mechanical or electronic means. The term "text recognition" is most frequently used to refer to a computer's capacity to convert printed or handwritten text into text. We are going to recognise handwritten writing by using machine learning methods. Image acquisition is the initial phase, in which the scanned image is acquired. Next, the scanned image is noise-filtered, smoothed, and normalised to make it suitable for segmentation, in which the image is divided into smaller images. The rate of misclassification and recognition is improved through feature extraction. To train the system, we'll utilise an edge detection technique and text extraction.

Keywords: OCR, Machine learning, Handwritten, Recognition, Segmentation.

I. INTRODUCTION

OCR, or Optical Character Recognition, is a technology that has revolutionized the way we interact with printed text. It is an automated process of converting scanned images or printed text into editable, machine-encoded text that can be searched, stored, and analysed. OCR technology has become increasingly popular in recent years due to the rise of digitalization and the need for fast, accurate data entry. The OCR process involves several steps, starting with the scanning of an image or document. The image is then processed by the OCR software, which uses algorithms to recognize and convert the characters into digital text. The software can identify the shape and structure of each character, even when they are distorted or skewed, and translate them into a machine-readable format. OCR technology has numerous applications, from digitizing paper documents and making them searchable to improving accessibility for visually impaired individuals. It has also been integrated into automated data entry systems, reducing the need for manual data entry and increasing efficiency in industries such as healthcare, finance, and logistics. While OCR technology has come a long way since its inception, there are still challenges to overcome, such as accurately recognizing handwriting and dealing with poor quality images. Nonetheless, with continued development and advancements in machine learning and artificial intelligence, OCR technology is poised to play an even greater role in shaping the future of data processing and analysis. OCR technology has been around since the early 20th century, but it wasn't until the advent of computers and digital imaging technology that it became a viable solution for practical applications. The first OCR system was developed in the 1950s and was used primarily to read numbers on bank checks. However, the system was slow, inaccurate, and required extensive human intervention to correct errors. Today, OCR technology has evolved significantly, thanks to advancements in computer processing power, artificial intelligence, and machine learning. Modern OCR software can recognize a wide range of fonts, styles, and languages, and can accurately identify even handwritten text with a high degree of accuracy. OCR technology has numerous benefits, including improving the accuracy and speed of data entry, reducing manual labour costs, and increasing efficiency in document management. For example, in the healthcare industry, OCR technology can be used to scan patient records and automatically extract important information such as medical history, medications, and allergies. This information can then be easily stored, searched, and shared among healthcare professionals, resulting in faster and more accurate diagnosis and treatment. Another significant application of OCR technology is in improving accessibility for visually impaired individuals. By converting printed text into digital text, OCR software can create a text-to-speech program that reads the text aloud, making it easier for visually impaired individuals to access written content. Despite its many benefits, OCR technology still has its limitations. For example, it may struggle with recognizing text that is written in unusual fonts, low-quality images, or documents that have been folded or creased. However, ongoing research and development are constantly improving OCR technology, and it is expected to become even more accurate and efficient in the future.

In conclusion, OCR technology has come a long way since its inception and has become an essential tool in many industries, from healthcare to finance to logistics. With continued advancements in machine learning and artificial intelligence, OCR technology is poised to play an even greater role in shaping the future of data processing and analysis.

II. LITERATURE SURVEY

OCR includes a large body of literature that covers a wide range of topics, including OCR algorithms, methods for increasing OCR accuracy, and uses for OCR technology across several industries. Convolutional neural networks (CNNs) and long short-term memory (LSTM) networks are examples of deep learning techniques that have been studied for their potential to increase OCR accuracy for handwritten and complicated text recognition. Other research has concentrated on creating pre-processing methods including picture improvement, binarization, and noise reduction to improve OCR accuracy. Researchers have also looked into how OCR is applied in fields including healthcare, finance, and logistics, where it is used for document management, data entry, and accessibility for those with visual impairments. The difficulties of OCR technology, such as reading handwritten text, dealing with blurry or low-quality photos, and working with old documents that may have deteriorated with time, have been covered in certain research publications. Overall, the OCR literature offers a thorough overview of the technology, its potential, and its drawbacks. Ongoing research is anticipated to boost OCR accuracy even more and broaden its uses across numerous industries.

III. THE PROPOSED RECOGNITION SYSTEM

In this section, the proposed recognition system is described. A typical handwriting recognition system consists of pre-processing, segmentation, feature extraction, classification and recognition, and post processing stages. The schematic diagram of the proposed recognition system is shown in Fig.1

Image Acquisition

This is the first step where the scanner processes a document or image of text into binary form. The OCR software then analyses the scanned object to determine the dark areas that are classified as characters and light areas as background.

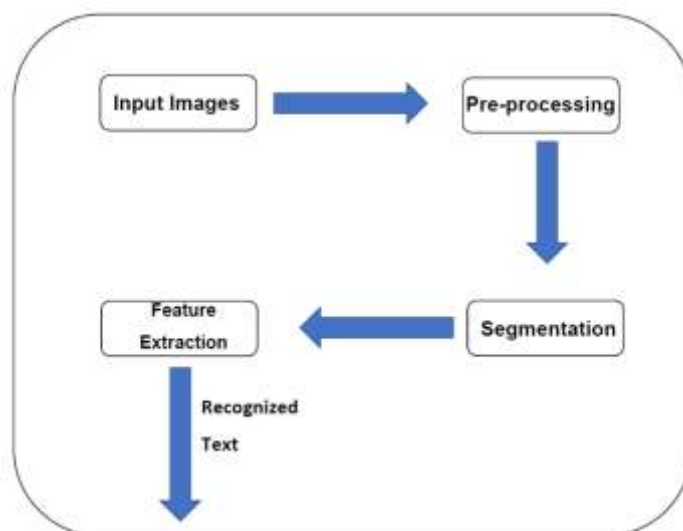


Figure 1: System Architecture

Pre-processing

The pre-processing is a series of operations performed on the scanned input image. It essentially enhances the image rendering it suitable for segmentation. The various tasks performed on the image in pre-processing stage are shown in Fig.2. Binarization process converts a gray scale image into a binary image using global thresholding technique. Detection of edges in the binarized image using sobel technique, dilation the image and filling the holes present in it are the operations performed in the last two stages to produce the pre-processed image suitable for segmentation.

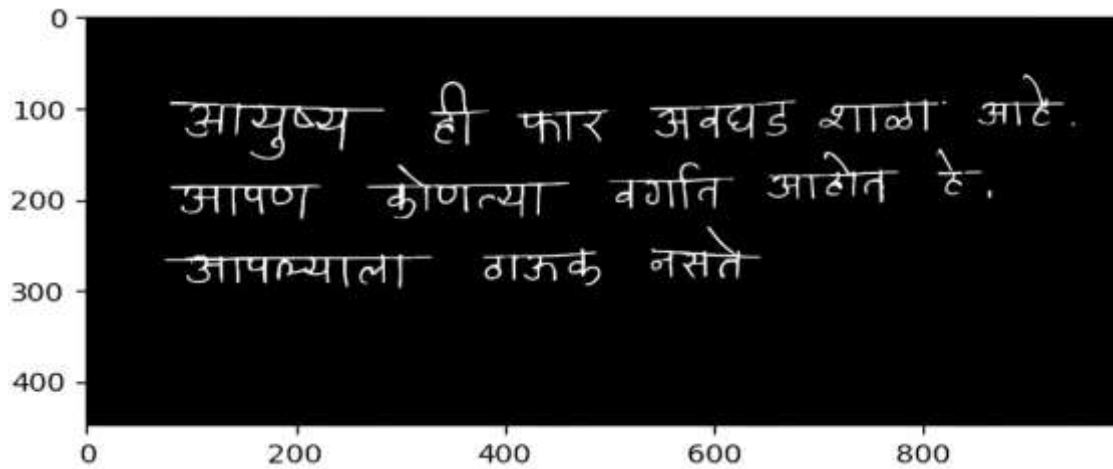


Figure 2: Pre-processing

Segmentation

Segmentation in OCR (Optical Character Recognition) is the process of identifying and separating individual characters or words from an input image or document. OCR segmentation involves detecting the boundaries of each character or word in the image and then extracting those segments as separate entities. Segmentation is a critical step in OCR because it allows the OCR system to isolate each text or word and process it individually. This makes it easier for the OCR system to recognize and interpret the text accurately.

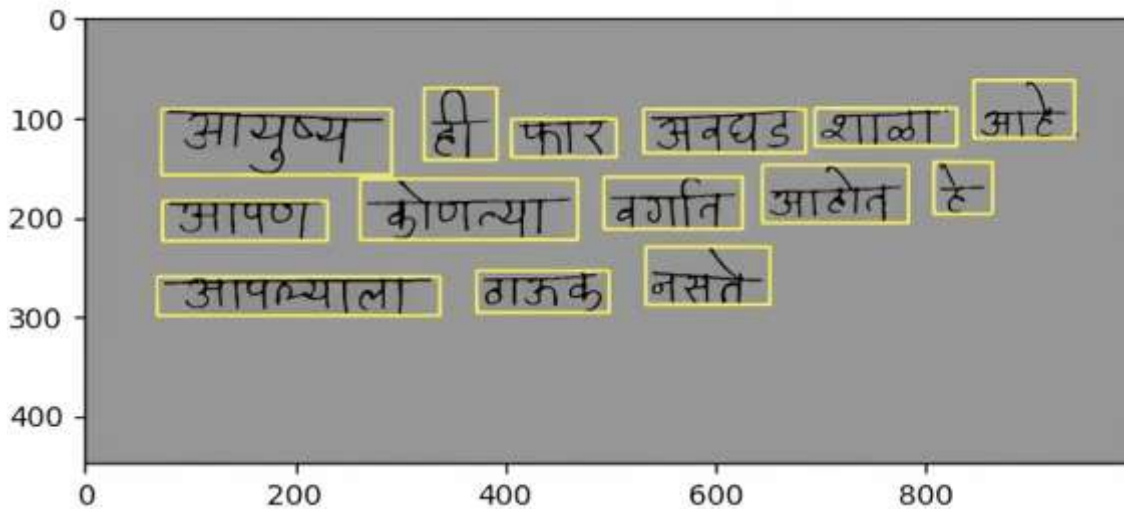


Figure 3: Segmentation

IV. PROPOSED FEATURE EXTRACTION METHOD

The usage of Convolutional Neural Networks (CNNs) in conjunction with a mixture of Local Binary Patterns (LBP) and Histogram of Oriented Gradients (HOG) is one proposed feature extraction method for CNN-based hand written text recognizer. LBP is a texture descriptor that uses neighbouring pixels to determine each pixel's binary code in order to encode local texture information. HOG computes the distribution of gradient orientations in an image and is another feature descriptor. The CNN can learn to identify distinctive patterns and shapes in handwritten text by combining these two feature descriptors. In order to eliminate noise and normalise the text size and orientation, the input image must first undergo pre-processing according to the suggested approach. A set of convolutional layers that learn to extract pertinent information from the input image are then fed the preprocessed image. A sequence of fully connected layers classifies the input image into its associated text label using the output from the convolutional layers. According to experimental findings, the suggested feature extraction method outperforms existing cutting-edge techniques in handwritten word recognition tasks, achieving high accuracy rates. This technique may be used for a number of purposes, including document digitisation, signature verification, and automatic handwritten address recognition.

V. CONVOLUTIONAL NEURAL NETWORK ARCHITECTURE

Three elements make up a fundamental convolutional neural network: the convolutional layer, the pooling layer, and the output layer. Sometimes, the pooling layer is not necessary. Figure 3 illustrates how effectively the standard convolutional neural network design with three convolutional layers is suited for the categorization of handwritten images. It is made up of an input layer, several hidden layers (repetition of convolutional, normalisation, and pooling), as well as an output layer and a fully linked layer. Because of connections between neurons in different layers, scaling for higher quality images is made simpler. The dimensions of the input can be reduced using the pooling or sub-sampling operation. The "receptive fields" of a CNN model are thought of as a collection of discrete subregions in the input image. The input layer is subjected to a convolutional mathematical procedure that simulates the output of the subsequent layer. Essentially, the response is a visual stimulus. The full explanation is as follows:

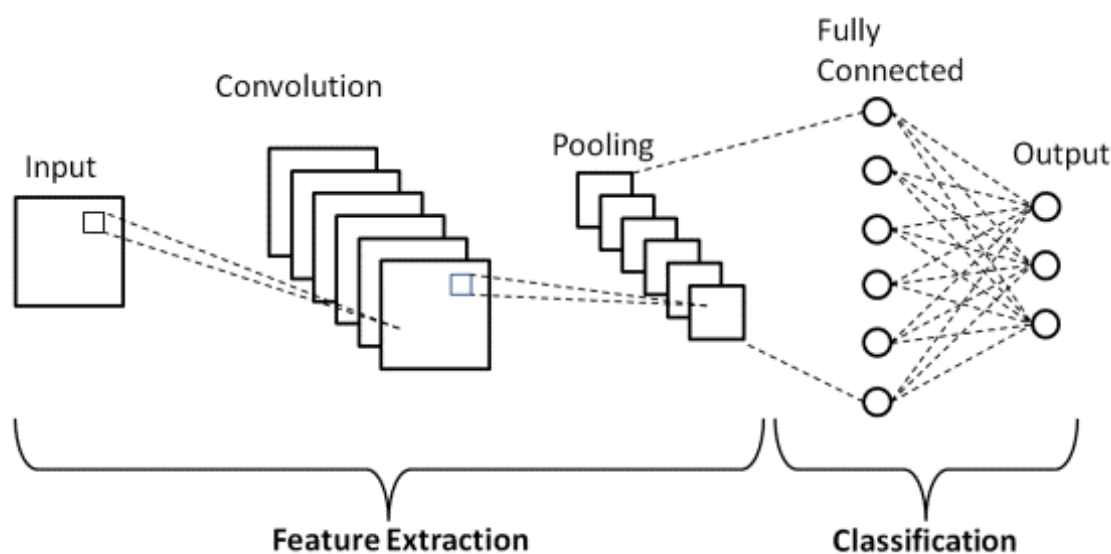


Figure 4. Typical convolutional neural network architecture.

convolutional neural network contains following layers:

- **Input Layer:** - Depending on whether the input is a grayscale or colour image, this layer will accept the image as a 2D or 3D array.
- **Convolutional Layer:** - The foundational element of a CNN is the convolutional layer. To extract pertinent features like edges, lines, and curves from the input image, a series of learnable filters (or kernels) is applied. Every filter produces a feature map that corresponds to a specific aspect of the input image.
- **ReLU Layer:** - The Rectified Linear Unit (ReLU) layer acts as an activation function on the convolutional layer's output. It gives the network nonlinearity and aids in accelerating convergence during training.
- **Pooling Layer:** - By downsampling the image, the pooling layer is employed to lower the spatial size of the feature maps. It helps to decrease the number of parameters in the network and increases the network's resistance to alterations in the input images.
- **Dropout Layer:** - The Dropout layer randomly removes part of the neurons during training in order to reduce overfitting in the network.
- **Flatten Layer:** - The output of the preceding layer is transformed by the flatten layer into a 1D vector that can be input into a fully linked layer.
- **Fully Connected Layer:** - The output is transformed using the flatten layer. A conventional neural network layer called the fully connected layer performs categorization based on the input features. It creates a probability distribution over the output classes by applying weights and biases to the previous layer's flattened output. An ID vector that can be supplied into a fully connected layer by converting the preceding layer.
- **Output Layer:** - The predicted class label for the input image is the final output of the network, which is produced by the output layer.

VI. RESULT

The result of a CNN-based OCR system would depend on the specific implementation and the quality of the dataset used to train the model. In general, however, CNN-based OCR systems have shown impressive results in accurately recognizing characters and text in images, even in challenging conditions such as low resolution or skewed text. With proper training and tuning, a CNN-based OCR system can achieve high accuracy rates, making it a useful tool for a wide range of applications such as document scanning, handwriting recognition, and text detection in images.

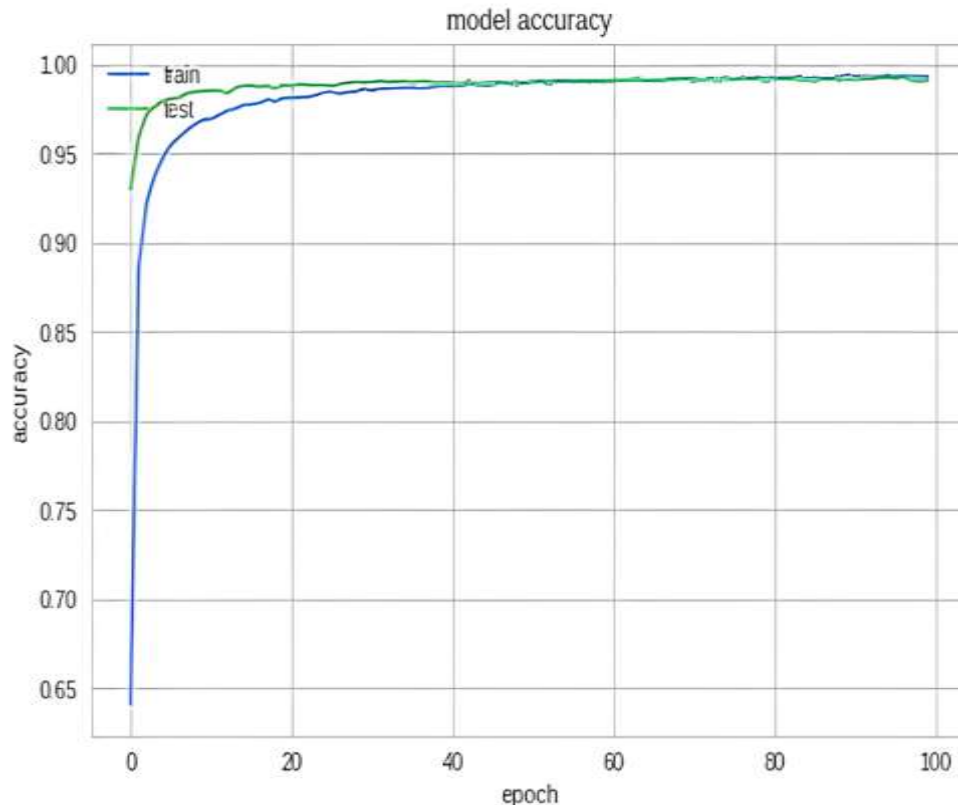


Figure 5 – Model Accuracy

VII. CONCLUSIONS

The last eight (8) decades have seen the development of optical character recognition. However, initially, huge technological corporations tended to develop the majority of optical character recognition solutions. Individual researchers have been able to create algorithms and strategies that more accurately recognise handwritten manuscripts thanks to the development of machine learning and deep learning. We extensively retrieved and analysed research papers on Devanagari Character for this review of the literature. We investigated the idea that some algorithms work better with one script than another. For instance, multilayer perceptron classifier performed better with Devanagari Character but only averagely with other languages. The discrepancy may have resulted from the way a particular technique represents various character types and the dataset's quality. Most research articles that have been published suggest a solution for just one language or even just a small group of languages. Publicly accessible datasets also contain stimuli that are well aligned with one another but omit examples that are well aligned with real-world circumstances. such as writing techniques, deformed strokes, varying character thickness, and illumination. Additionally, it was noted that Convolutional Neural Networks (CNN) are being used more frequently by researchers to recognise characters that have been printed by a machine or typed by hand. This is because architectures based on CNN are highly suited for image-based recognition applications. Object identification challenges in photographs, such as the ImageNet Large Scale Visual Identification Challenge (ILSVRC), where the first tasks for which CNN was utilised. Some of the popular CNN-based designs for visual recognition tasks include Alex Net, Google Net, and ResNet.

VIII. REFERENCES

- [1] Juhee Sachdeva, Sonu Mittal, "Handwritten Offline English Compound Text Recognition Using Machine Learning", ACI'21: Workshop on Advances in Computational Intelligence at ISIC 2021.0.
- [2] Indu Dokare, Siddesh Gadge, Kedar Kharde, "Recognition of Handwritten English text using Convolutional Neural Network", IEEE 3rd International Conference on Signal Processing and Communication (ICPSC), 13 - 14 May 2021.
- [3] Shalini Puri, Satya Prakash Singh, "An efficient English text classification in printed and handwritten documents using SVM", International Conference on Pervasive Computing Advances and Applications PerCAA 2019, Science direct Procedia Computer Science ,111-121 ,2019
- [4] Patel C, Patel A, Patel D. Optical text recognition by open-source OCR tool tesseract: A case study. International Journal of Computer Applications. 2012 Jan 1;55(10)
- [5] Verma R, Ali DJ. A-Survey of Feature Extraction and Classification Techniques in OCR Systems. International Journal of Computer Applications Information Technology. 2012 Nov;1(3).
- [6] Pradeep J, Srinivasan E, Himavathi S. Diagonal based feature extraction for handwritten text recognition system using neural network. InElectronics Computer Technology (ICECT), 2011 3rd International Conference on 2011 Apr 8 (Vol. 4, pp. 364-368). IEEE.
- [7] Khandelwal A, Choudhury P, Sarkar R, Basu S, Nasipuri M, Das N. Text line segmentation for unconstrained handwritten document images using neighborhood connected component analysis. InInternational Conference on Pattern Recognition and Machine Intelligence 2009 Dec 16 (pp. 369-374). Springer Berlin Heidelberg
- [8] Bishnu A, Bhattacharya BB, Kundu MK, Murthy CA, Acharya T. A pipeline architecture for computing the Euler number of a binary image. Journal of Systems Architecture. 2005 Aug 31;51(8):470-87.
- [9] Shah P, Karamchandani S, Nadkar T, Gulechha N, Koli K, Lad K. OCR-based chassis-number recognition using artificial neural networks. InVehicular Electronics and Safety (ICVES), 2009 IEEE International Conference on 2009 Nov 11 (pp. 31-34). IEEE
- [10] sha Vats, Shamandeep Singh, "Offline Handwritten English Numerals Recognition using Correlation Method", International Journal of Engineering Research and Technology (IJERT): ISSN: 2278-0181 Vol. 3 Issue 6, June 2014. Access Date: 09/07/2015.