DEEP LEARNING BASED HANDWRITTEN DIGIT RECOGNITION SYSTEM

Abstract

Handwritten recognition pertains to the capacity of a data processing system to discern patterns of handwritten information, such as digits and characters, from diverse sources like emails, papers, images, and reports. Within domains like computer vision and applications of machine learning (ML), the challenge of recognizing handwritten digits has emerged as an extensively studied problem. Numerous ML techniques have been devised to address this issue of identifying handwritten digits. With the advancement of various machine learning and deep learning approaches, handwritten digit recognition has evolved into an exceptionally captivating field for researchers. In this study, the outcomes of some of the most commonly employed deep learning algorithms, namely the Deep Neural Network (DNN), Convolutional Neural Network (CNN), and Recurrent Neural Network (RNN) are evaluated. Notably, CNN has demonstrated superior accuracy compared to conventional image recognition techniques.

Keywords: Handwriting Recognition, Deep Learning, CNN, RNN, DNN.

Author

Varun P. Sarvade
Assistant Professor
Department of CSE
B. G. M. Institute of Technology
Mudhol,Karnataka, India.
varunpsarvade@gmail.com

I. INTRODUCTION

Handwriting analysis has been present since the 1980s, and the rapid advancements in Machine Learning are revolutionizing this field. Each day, novel computations are being introduced, leading to remarkable improvements in system capabilities. These machines possess the ability to distinguish between various handwriting styles that often elude human perception. Machine Learning, a facet of computerized reasoning, empowers systems to autonomously enhance and refine their performance based on past experiences, thus paving the way for future development [1].

This study employs deep learning computations to validate handwritten digits, encompassing three core architectures: the Deep Neural Network, Convolutional Neural Network, and Recurrent Neural Network. TensorFlow, a versatile library for dataflow programming, spans a range of tasks. As an open-source software and a potent mathematical toolkit, it finds extensive application in Machine Learning, including neural connections modeling. Developed by Google Brain's team for internal use, TensorFlow represents the second generation of their systems, with Version 1.0.0 released on February 11, 2017 [2].

II. BACKGROUND

In this segment, an outline of the different Machine Learning techniques used in our paper is given:

1. **Neural Network (NN):** A Neural Network is a fundamental concept in machine learning as shown in figure 1, inspired by the structure and functioning of the human brain's interconnected neurons. It's a computational model designed to recognize patterns, relationships, and features in data, particularly when dealing with complex and high-dimensional information. Neural Networks consist of layers of interconnected nodes, called neurons. Each neuron processes input data, applies a set of weights and biases to it, and then passes the result through an activation function. These neurons are organized into layers: an input layer, one or more hidden layers, and an output layer. The connections between neurons are characterized by weights, which are adjusted during training to optimize the network's performance [3].

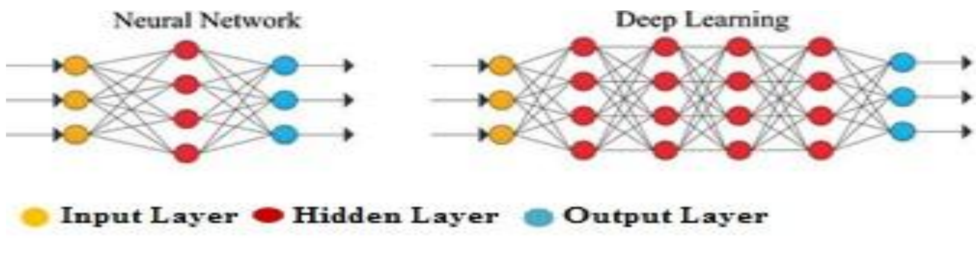

Figure 1: Neural Network Structure.

2. **The Deep Neural Network (DNN):** The Deep Neural Network (DNN) stands as a cornerstone within the realm of both Deep Learning and the broader Machine Learning domain, built upon the principle of acquiring informative representations. Comprising an input layer, an output layer, and typically one or more hidden layers [4], DNNs represent a dynamic field that has propelled machine learning towards attaining unprecedented

levels of precision. The synergy between extensive datasets and robust computational capabilities assumes a pivotal role in the triumphant formulation of DNN models [5].

3. **Convolutional Neural Network (CNN):** A Convolutional Neural Network (CNN) is a specialized type of neural network designed primarily for processing and analyzing visual data, such as images and videos. It's particularly effective at capturing spatial patterns and features within the data. CNNs use a specialized layer called a "convolutional layer" that applies filters or kernels to the input data, scanning it to detect different features like edges, textures, and shapes. These learned features are then hierarchically combined in subsequent layers to understand more complex structures in the data. CNNs have proven highly successful in various tasks such as image classification, object detection, and image segmentation due to their ability to automatically learn and extract relevant features from raw visual data. The CNN depicted in figure 2 consists of an input layer, an output layer, and several hidden layers [6]. The input layer of the network is composed of input neurons that encode the extracted values from the input pixels of our handwritten digits. For our training data, sourced from the MNIST dataset, a multitude of 28 by 28-pixel images is present. Consequently, the input layer comprises 784 information neurons. Subsequently, the hidden layer receives the output from the information layer and employs pattern recognition capabilities to discern distinctive patterns within the input images [7].

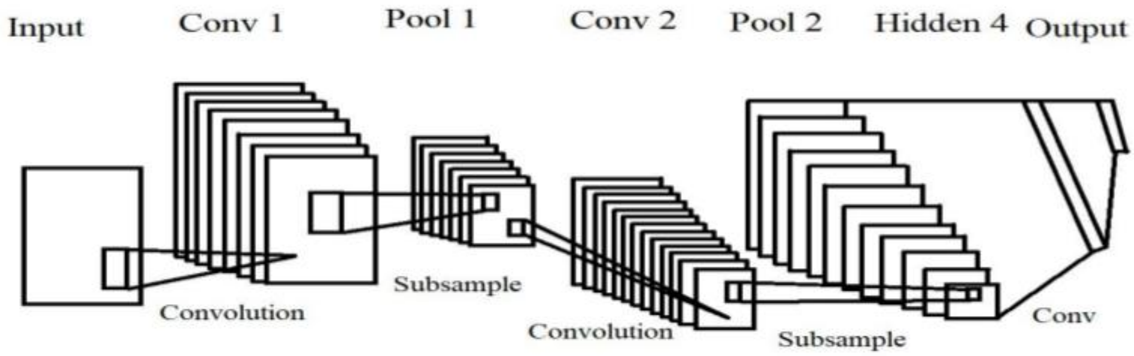

Figure 2: Simplified CNN structure.

4. **Recurrent Neural Network (RNN):** A Recurrent Neural Network (RNN) as shown in figure 3, is a type of artificial neural network designed for handling sequential data and tasks that involve a temporal dimension. Unlike traditional feedforward neural networks, which process inputs in isolation, RNNs maintain an internal memory or state that captures information from previous time steps. This memory enables RNNs to capture patterns and relationships within sequences, making them well-suited for tasks like natural language processing, speech recognition, and time series analysis. RNNs work by repeatedly applying the same set of weights and biases to each input in a sequence, while also incorporating information from previous inputs through their internal state. This architecture allows RNNs to model dependencies and context within sequential data. However, standard RNNs can suffer from the "vanishing gradient" problem, where the influence of early inputs diminishes quickly as the sequence progresses. This helps to unveil dynamic temporal deeds for a time sequence [8]. Self-Organizing Map (SOM) is likewise utilized, intended for self-coordinating picture highlights from crude pictures [9].

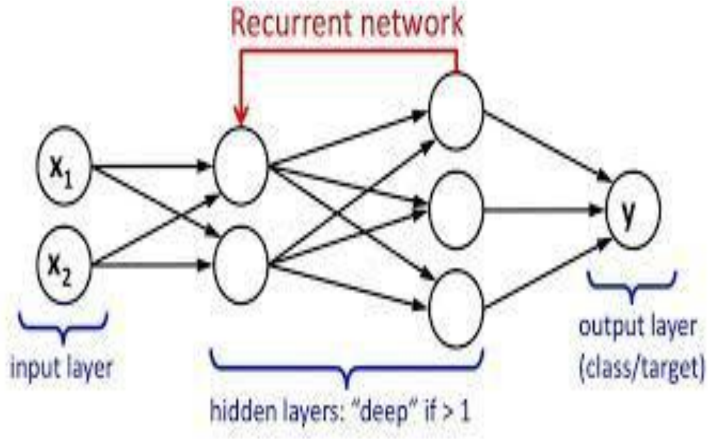

Figure 3: A simple structure of RNN

III. RELATED WORK

Yan Lecun et al.[10] put out the paper with the title "Deep Learning" for computational models, on the way to learn representations of data deep learning which consist of multiple processing layers and also with multiple levels of abstraction. In discourse recognition, visual article identification, object recognition and numerous different areas, for example, drug revelation and genomics Deep Learning methods have improved the state-of-the-art. Deep learning finds complex construction on heft of datasets utilizing the back spread calculation to show how a machine can change interior boundaries and is additionally used to figure the information portrayal in each layer. Deep convolutional nets zeroed in primarily on handling of pictures, video, discourse, and sound, though repetitive nets uncovered the light on successive information like text and discourse.

Dan Saadati et al. [11] published a paper with the title "Handwriting text recognition using deep learning", Authors used mainly two approaches to establish the classification of an individual manually written word so that transcribed text can be meant a computerized structure by ordering words straightforwardly and character division. They have utilized Convolutional Neural Network (CNN) with different structures to prepare a model that can precisely characterize words. For the last option, they utilize Long Short-Term Memory organizations (LSTM) with convolution to develop bouncing boxes for each person. They then pass the segmented characters to a CNN for order, and afterward remake each word as per the consequences of characterization and division.

San Ni Yun et al. [12] have explained that the significance and intricacy of manually written character input in human-system collaboration framework, a non-contact character input strategy has been proposed. An individual holding a light pen writes in the air in this strategy, the video of visual light spot development in air is recorded by a camera and handled using computer system vision innovation, then the manually written character picture is remade, at last, the recreated character is perceived. The strategy is using the light spot variety elements and trademark boundaries shape and pixel region to portion it in the video which is recorded by a common camera. Subsequent to distinguishing the light spot, they incorporated a person which adjusts to the state recommended norms of the info character and remembered it on line. By utilizing the transcribed person input strategy embraced in this paper, it is exceptionally useful for future clients to dispose of the customary man machine intelligent method of console, mouse and so on.

Louis Vuurpijl et al. [13] have described that the new strategy is utilized for ordering character shapes (allograph) in huge informational collections of penmanship into a various leveled structure. Such a procedure might be utilized as the reason for an efficient naming plan of character shapes. Each bunch addresses an allograft model. The advantage of the technique is most noteworthy in the lower parts of the group dendrogram, where countless individuals might have a place with single hubs. In the event that the objective is to recognize an essential construction looking like characters, the decision of the list of capabilities is as yet significant. Future work is aimed at a solidification of results utilizing others including portrayals, for example, picture based highlights.

IV. METHODOLOGY

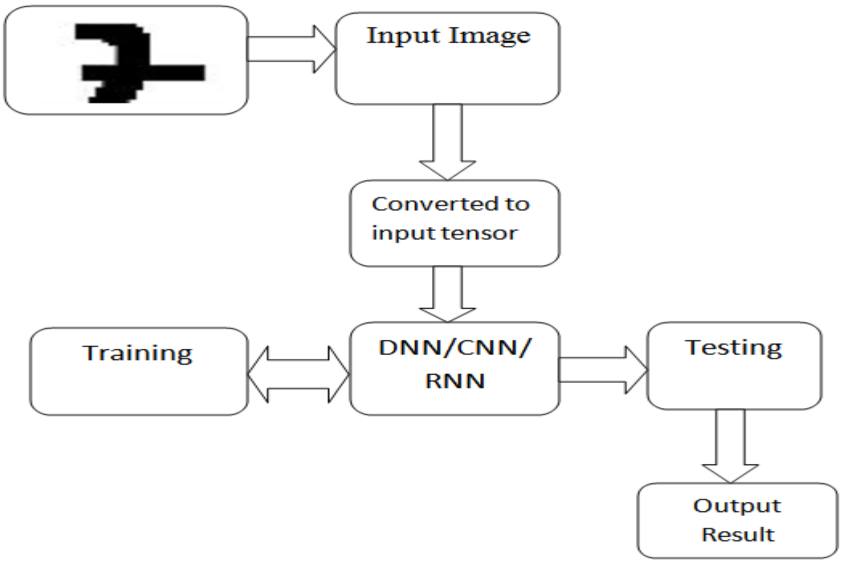

Figure 4: Block diagram of digit recognition system using Neural Network, DNN, CNN & RNN

The proposed system shown in figure 4 is sorted into five stages: preprocessing, Segmentation, Feature Extraction, Training, classification and recognition are explained as below.

1. **Preprocessing:** At this point, the pictures in the data set are perfect and silent, thus there is no prerequisite of sound decrease procedure. However, in a genuine framework, the need is to dispense with clamor from the pictures. Since there could be optical clamors present close by of the archives. Totally in the transcribed records, the flighty shapes may not be constantly special. Thus the preprocessing step is compulsory. At first apply Erosion with 3 X 3 organizing components, so it will kill the slightest bit of blunders and produce a smooth edge. Then the characters are extended with 2 X 2 components.

2. **Segmentation:** After the preprocessing step, sub-pictures of a singular digit are ready by a deteriorating series of pictures of a digit. The preprocessed input picture is fragmented

into isolated digits by relegating a "Number" to every digit independently through a naming strategy. Surmised measurements with respect to various digits in the picture are given by marking. Thereafter every individual digit is uniformly resized into 100 X 70 pixels for additional procedures (for example grouping and identification stages).

3. **Feature Extraction:** After the division stage, The result of this stage, for example the sectioned picture will be sent as a contribution to the following module for highlight extraction. The measurable highlights, for example, histogram, mean and standard deviation variables will be pull out from the pictures.

4. **Training:** After the Feature Extraction stage, the proposed calculations (CNN, DBN, DNN) are prepared discreetly with the functioning out pictures.

5. **Classification & Recognition:** After the preparation period, Classification and Recognition strategies are utilized as a dynamic proportion of an identification framework and it likewise utilizes the above removed highlights. Two secret layers are available in the feed forward back proliferation brain network having a design of 54-100-100-38 for play out the characterization. The secret layers use log sigmoid enactment capability, and the result layer is a cutthroat layer, as one of the digits is to be recognized. The element vector is signified as X where X = (f1, f2,… ,fd) where f indicates highlights and d is the quantity of zones into which every digit is separated. The quantity of information not entirely settled by the length of the element vector d. The absolute quantities of digits' n decide the quantity of neurons in the result layer. The quantity of neurons in the secret layers is obtained by experimentation. It is to perceive manually written digits involving the three calculations where every calculation perceives the picture in its own particular manner. After the preparation cycle, the Digits are contrasted by a specialist to evaluate the exactness of the tip. Additionally, the accuracy, the cost of execution and execution time are analyzed [14].

V. EXPERIMENTAL SETUP

1. **Dataset:** The MNIST dataset comprises 60,000 training images and 10,000 testing images. The training set and testing set are divided into two significant portions, with one part sourced from MNIST's standard dataset and the other from its challenging dataset [15]. Kaggle serves as a platform for competitive predictive modeling and analytical competitions, where data professionals and analysts engage in creating optimal models to predict and characterize datasets provided by various entities. Within the Kaggle dataset, there are 21,000 images designated for training and 9,000 images for testing purposes [16].

2. **Simulation Tool:** The system employs Ubuntu 16.04 (64-bit) as its operating system, accompanied by Anaconda and Google Colab, which are essential for importing diverse deep learning techniques and conducting simulations.

3. **Performance Metrics:** Accuracy is considered for measuring the performance of different deep learning techniques.

VI. RESULT ANALYSIS

In the initial stages of instruction, errors are notably elevated, but as the instructional process advances, the loss experiences a substantial reduction. This trend is illustrated in figure 5, where the Neural Network achieves an accuracy level of 83.03%. The MNIST dataset consists of 60,000 training data points and 10,000 testing data points, originally employed to assess the precision of the DNN (Deep Neural Network).

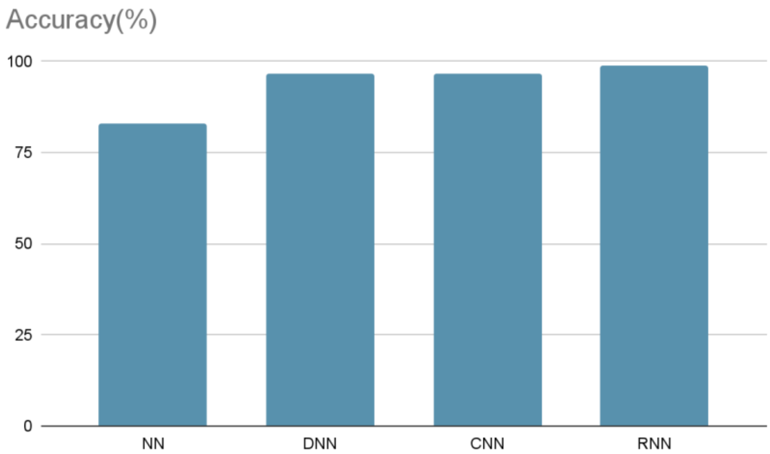

Figure 5: Accuracy comparison of NN,DNN,CNN and RNN.

Initially, there is an 83.03% accuracy achieved using conventional methods, such as studying neural networks, which is significantly lower compared to contemporary approaches. Introducing the Deep Neural Network raises the accuracy to 96.46%, surpassing the performance of brain networks and concurrently reducing the error rate. Notably, the Convolutional Neural Network achieves an even higher accuracy of 96.78% than the Deep Neural Network and further minimizes the error rate. Progressing to the Recurrent Neural Network, an impressive accuracy of 98.78% is attained, outperforming the Convolutional Neural Network while continuing to decrease the loss rate.

VII. CONCLUSION

The developed a framework for identifying handwritten digits that leverages the power of Deep Learning through TensorFlow. Initially, a neural model was used, which yielded an accuracy of only 83%, which fell short of expectations. However, through subsequent efforts, it effectively boosted the accuracy to an impressive 98.72% by implementing advanced Deep Learning techniques such as DNN, CNN, and RNN. The training and testing phases involved the utilization of publicly available datasets from MNIST and Kaggle.

VIII. FUTURE WORK

It is conceivable that futuristic trends in AI will usher in transformative developments. One such potential paradigm shift lies in the integration of Quantum AI systems, which hold the promise of revolutionizing the speed and accuracy of recognizing handwritten digits.

Quantum AI systems leverage the principles of quantum mechanics to process information in a fundamentally different manner compared to classical computing. By harnessing the power of superposition and entanglement, these systems exhibit an unprecedented capacity to handle complex computational tasks. When applied to the domain of handwritten digit recognition, Quantum AI systems may offer a revolutionary leap forward in terms of processing capabilities.

The utilization of Quantum AI for handwritten digit recognition may lead to a dramatic acceleration in the speed at which algorithms can accurately interpret and classify handwritten characters. This accelerated processing could have far-reaching implications for applications ranging from automated data entry to image recognition in various domains. Moreover, the potential for greater precision and efficiency in recognizing handwritten digits could significantly enhance the overall performance and usability of AI-driven systems.

While this application of Quantum AI in handwritten digit recognition is currently speculative, it underscores the transformative potential that lies ahead in the intersection of AI and quantum computing. As research and development in this field continue to progress, it is conceivable that Quantum AI systems may play a pivotal role in shaping the future landscape of digit recognition technologies.

REFERENCES

[1] Online: "Introduction of handwriting recognition", https://en.wikipedia.org/wiki/Handwriting recognition.
[2] Online: “TensorFlow”, www.tensorflow.org.
[3] Online: “What is a Neural Network? ”https://www.investopedia.com/terms/n/neuralnetwork.asp
[4] “Deep Neural Network”, https://en.wikipedia.org/wiki/ Deep Neural Network.
[5] Moazam Soomro, and Rana Hammad Raza, "Performance Evaluation of Advanced Deep Learning Architecture of Offline Handwritten Character Recognition", 2017 International Conference on Frontiers of Information Technology.
[6] “Convolutional Neural Network”, https://en.wikipedia.org/wiki/Convolutional_neural_network.
[7] T Siva Ajay “Handwritten Digit Recognition Using Convolutional Neural Networks”. International research journal of engineering and technology, July 2017.
[8] “Recurrent Neural Network”,https://en.wikipedia.org/wiki/Recurrent_neural_network.
[9] S. Nishide, H. G. Okuno, T. Ogata and J. Tani, "Handwriting prediction based character recognition using recurrent neural network," 2011 IEEE International Conference on Systems, Man, and Cybernetics, 2011, pp. 2549-2554, doi: 10.1109/ICSMC.2011.6084060.
[10] LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. Nature, 436–444 (2015). https://doi.org/10.1038/nature14539
[11] Balci, Batuhan, Dan Saadati, and Dan Shiferaw. "Handwritten text recognition using deep learning." CS231n: Convolutional Neural Networks for Visual Recognition, Stanford University, Course Project Report, Spring (2017): 752-759.
[12] Vuurpijl, Louis, and Lambert Schomaker. "Finding structure in diversity: A hierarchical clustering method for the categorization of allografts in handwriting." Proceedings of the Fourth International Conference on Document Analysis and Recognition. Vol. 1. IEEE, 1997.
[13] San Ni Yun , Lan Xiang Zhong "Experimental Research on Handwritten Character Written in the Air Recognition Based on Computer Vision", 4th International Congress on Image and Signal Processing, pp.530-532, IEEE 2011.
[14] Ghosh, Mahmoud M. Abu, and Ashraf Y. Maghari. "A comparative study on handwriting digit recognition using neural networks." 2017 international conference on promising electronic technologies (ICPET). IEEE, 2017.
[15] Online: MNIST Dataset for digit recognition, “http://yann.lecun.com/exdb/mnist/”.
[16] Online: Anthony Goldbloom “Kaggle dataset for digit recognition”, “www.kaggle.com” .

