STM JOURNALS

International Journal of
Data Structure Studies

Volume 1, Issue 1, 2023
January—June
DOI (Journal): 10.37591/IJDSS

https://journals.stmjournals.com/ijdss

Review IJDSS

Assessing the Performance of DL Methods in Handwritten Digit Recognition

Radhey Shyam[1,*], Shilpi Khanna[2], Priyanka Verma[2], Sakshi Maurya[2]

Abstract

Handwritten digit recognition is a computer vision task that involves the automatic identification and classification of hand-written digits. The objective is to develop models capable of accurately recognizing and distinguishing digits handwritten by humans. With the development of machine learning and deep learning techniques, this field has advanced remarkably. The convolutional neural network (CNN) is the most often used technique for this purpose. By utilizing CNN, the model can learn to identify unique features of each digit, such as shape, thickness, and curvature patterns, enabling it to effectively classify and differentiate handwritten digits with high precision. Additionally, other machine learning techniques like artificial neural networks (ANN) and recurrent neural networks (RNN) can also be utilized. Handwritten digit recognition has practical uses in postal services, vehicle number plate detection, and extracting numbers from bank cheques. In this particular study, we focus on the detection and classification of handwritten digits by employing various techniques and models. Specifically, we compare the performance of ANN, CNN, and RNN models and found the accuracies to be 97.73, 96.61 and 96.13% respectively based on the evaluation. These evaluations are performed under the most popular dataset i.e., modified national institute of standards technology (MNIST) dataset. The collection consists of about 70,000 photos in grayscale that show handwritten digits from 0 to 9. The study's outcomes have significant implications for practical applications across different domains.

Keywords: Digit recognition, CNN, ANN, RNN, MNIST

INTRODUCTION

Handwritten digit recognition is a computer vision task that involves automatically identifying and classifying handwritten digits from images. The primary objective is to develop algorithms and models that can recognize handwritten digits accurately, much like a human would. This technology has numerous applications, such as digitizing handwritten documents, processing postal addresses, automatic check processing, and digit recognition in forms and surveys. The process of handwritten digit recognition typically involves steps, such as data collection, data pre-processing, feature extraction, model building, training, validation, testing and deployment.

Handwritten digit recognition is a technology that enables computers to interpret and identify numerical digits written by humans. It is a challenging task due to the variability in writing styles and the influence of factors such as writing instruments, paper quality, and scanning conditions. To address these challenges, various machine learning algorithms, including neural networks, decision trees, and support vector machines, have been utilized. These algorithms play a crucial role in recognizing and distinguishing handwritten digits

***Author for Correspondence**
Radhey Shyam
E-mail: shyam0058@gmail.com

[1]Professor & Head, Department of Information Technology, Shri Ramswaroop Memorial College of Engineering and Management, Lucknow, Uttar Pradesh, India
[2]Student, Department of Information Technology, Shri Ramswaroop Memorial College of Engineering and Management, Lucknow, Uttar Pradesh, India

Received Date: July 19, 2023
Accepted Date: July 28, 2023
Published Date: August 14, 2023

Citation: Radhey Shyam, Shilpi Khanna, Priyanka Verma, Sakshi Maurya. Assessing the Performance of DL Methods in Handwritten Digit Recognition. International Journal of Data Structure Studies. 2023; 1(1): 25–32p.

with high accuracy. These algorithms learn from extensive datasets of handwritten digits and utilize that knowledge to accurately recognize and classify new digits [1–3].

This study explores the diverse application of handwritten digit recognition in fields like finance, education, and healthcare. The potential uses include automating form filling, signature verification, and assisting in medical diagnosis. The study focuses on comparing recognition accuracies among the most popular deep learning models, namely artificial neural network (ANN), convolutional neural network (CNN), and recurrent neural network (RNN). Different hyper-parameters (i.e., batch size, epoch, no. of classes, etc.) are chosen to evaluate the models using the modified national institute of standards technology (MNIST) dataset [4]. The dataset sample is illustrated in Figure 1.

Figure 1. Sample of images of MNIST Dataset [5].

Indeed, an Artificial Neural Network (ANN) is a computational model inspired by the architecture of the human brain. It consists of interconnected neurons, and the network structure involves three primary layers: the input layer, the hidden layer(s), and the output layer. The input layer receives data, the hidden layer(s) process the information, and the output layer produces the final results or predictions. Through training and adjusting the connection weights between neurons, an ANN can learn to recognize patterns, make predictions, and solve complex tasks. These layers work together to process and interpret data. The input layer is responsible for receiving data from the external world. This data is then processed through one or more hidden layers before producing the final output through the output layer. The neural network learns and adjusts its internal connections based on a training set, enabling it to respond to input data and provide corresponding outputs [6, 7]. Working of ANN model is illustrated in Figure 2.

The Convolutional Neural Network (CNN) is a specialized artificial intelligence technique used primarily for image classification. It is specifically designed to handle and analyze image data, offering superior performance in this domain. A CNN (Convolutional Neural Network) is composed of three essential layers: the convolutional layer, the fully connected layer, and the pooling layer. These layers work together to identify images more accurately by processing and extracting key information from them [8–13]. Figure 3 shows the CNN model's intricate operation.

The Recurrent Neural Network (RNN) is a type of neural network specifically designed to handle sequential data. In contrast to conventional feedforward neural networks, RNNs (Recurrent Neural Networks) employ a feedback loop mechanism that allows information from previous inputs to impact the current output. This feedback loop allows the network to have a memory of past information and make decisions based on the context of the entire sequence [14]. Rather than following a linear flow, RNNs utilize recurrence relations and learn through a process known as back-propagation through time [15–19]. Illustration of RNN model is shown in Figure 4.

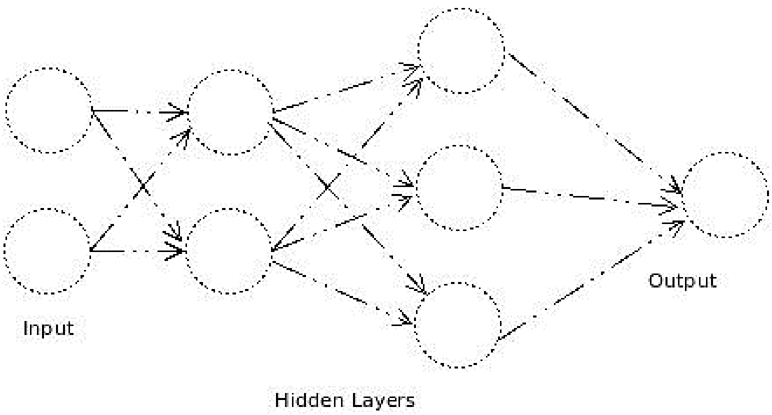

Figure 2. Architecture of a typical artificial neural network.

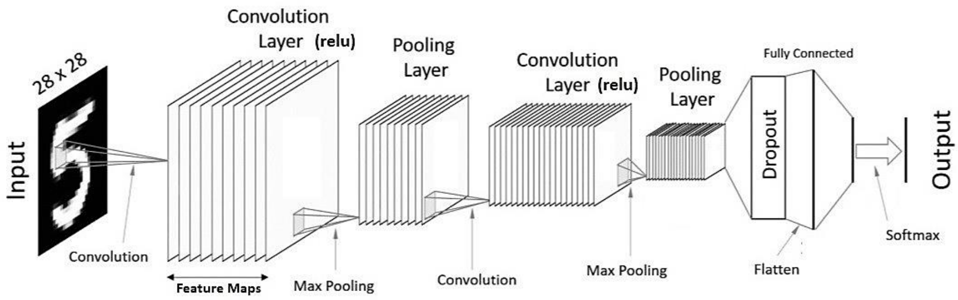

Figure 3. Architecture of a typical convolutional neural network [20].

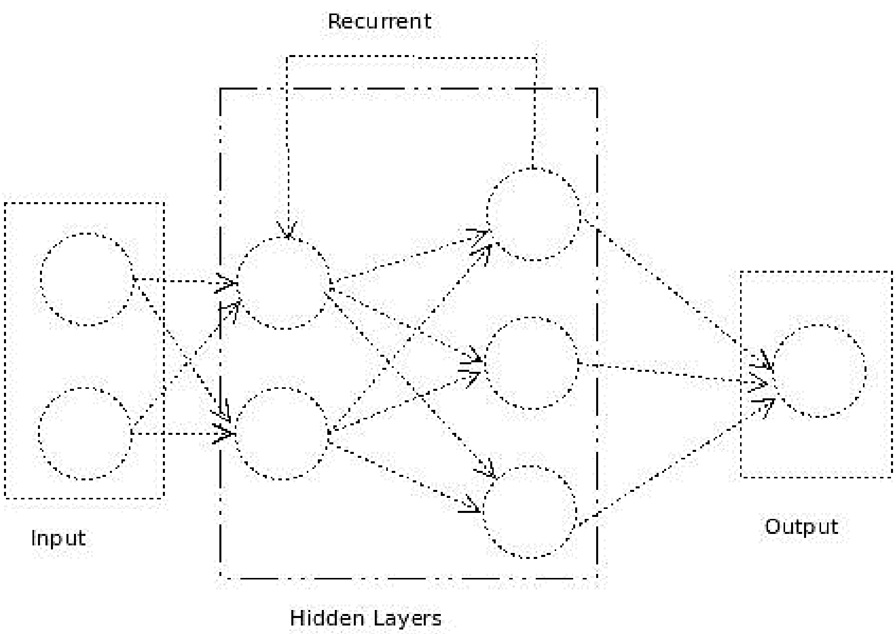

Figure 4. Architecture of a typical Recurrent neural network.

The structure of the paper is as follows: The related works are summarized in the next Section. The Section after that describes the methodologies employed in framing the models. The evaluation of the findings is discussed afterwards. The Section after that discusses and presents observations based on the findings. Ultimately, the last Section provides the concluding remarks of the paper.

RELATED WORK

The dataset mentioned in the article by Brownlee is extensively employed and thoroughly comprehended in the field [21]. It is considered to be mostly "solved", as state-of-the-art models, specifically deep learning convolutional neural networks, consistently achieve classification accuracy rates exceeding 99%. The error rate on the test dataset not used during training typically ranges between 0.4 and 0.2%.

Ahlawat *et al.* accomplished an outstanding recognition rate of 99.89% on the MNIST database by employing the Adam optimizer [22]. Their achievement exceeds all previously reported results in this field. Furthermore, their research effectively demonstrates the positive impact of increasing the number of convolutional layers in the CNN architecture on the performance of handwritten digit recognition.

In their study, Pham *et al.* applied the dropout regularization technique to improve the performance of recurrent neural networks (RNNs) in recognizing unconstrained handwriting [23]. Their research showcased significant enhancements in RNN performance, leading to a substantial decrease in both the character error rate (CER) and word error rate (WER) during the recognition process. During the recent years, the CNN model has gained extensive popularity in the domain of handwritten digit recognition, particularly with the MNIST benchmark database. Various researchers have documented remarkable accuracy rates, achieving as high as 98 or 99% in recognizing handwritten digits [24].

Gupta *et al.* introduced a unique multi-objective optimization framework aimed at identifying the most informative local regions within a character image [25]. They evaluated this framework on various datasets, including isolated handwritten English numerals i.e., MNIST images and three other prominent Indic scripts: handwritten Bengali numerals and handwritten Devanagari characters. The researchers integrated features extracted from a convolutional neural network into their model, resulting in an outstanding recognition accuracy of 95.96%.

Nguyen *et al.* employed a multi-scale CNN to extract spatial classification features from handwritten mathematical expressions (HME) [26]. CNN was used to capture both local features and spatial information from HME images, which were then utilized for clustering the HME images. The study observed excellent performance on the CROHME dataset. Furthermore, the authors reached the conclusion that additional enhancements in classification can be attained by training the CNN using a combination of global max pooling and global attentive pooling techniques. The effectiveness of CNNs is largely influenced by the selection of hyper-parameters [27], which are typically determined through a trial-and-error approach. Vital hyper-parameters, such as the activation function, number of epochs, kernel size, learning rate, hidden units, hidden layers, and others, significantly influence how the algorithm learns from the given data [28]. It is essential to distinguish hyper-parameters from model parameters, as they need to be defined before the training process begins.

Some popular CNN models, namely ResNet-52 [29], GoogleNet [30], VGG-16 [30], and AlexNet [31], possess varying numbers of hyper-parameters: 150, 78, 57, and 27, respectively. The choice of hyper-parameters significantly influences the performance of CNNs, and selecting inappropriate values can result in high computational costs and poor overall CNN performance.

The expertise of researchers is crucial in determining the optimal configuration of hyper-parameters, necessitating a well-thought-out and strategic approach. This raises several important questions concerning CNN design for handwriting recognition tasks. Firstly, CNNs excel in extracting distinct features from handwritten characters, making them particularly well-suited for this purpose. Secondly, different hyper-parameters have distinct effects on CNN performance, and their proper selection is critical for achieving optimal results. Lastly, design parameters, including hyper-parameters, play a significant role in enhancing CNN performance, and their careful tuning can lead to notable improvements in handwriting recognition tasks.

Addressing these questions is essential to provide guidance for future research in the handwriting recognition field, paving the way for the development of more effective and accurate CNN models.

METHODOLOGY

1. The initial step involved acquiring an appropriate dataset for handwritten digit recognition from various sources. The MNIST Dataset is commonly used due to its performance and accuracy in recognizing handwritten digits. This dataset, introduced in 1998, offers an abundant amount of training and testing data for models. It consists of images represented as clusters of 28×28 values and has been widely employed to evaluate the effectiveness of classification algorithms in the domain of handwritten digit recognition.
2. Data pre-processing is a crucial step in which the focus is on enhancing the input data by reducing or eliminating unnecessary noise, impurities, and redundancies.
3. Data visualization involves presenting data or information in visual formats such as graphs or charts. The purpose of visualization is to make the data easily understandable, allowing patterns and trends within large datasets to be identified more effectively.
4. The subsequent step involves dividing the complete dataset into distinct training and testing datasets.
5. In the dataset, the training data is utilized for constructing the model, while the testing or validation data is employed to assess the performance and validate the model. Consequently, the training data is utilized to train the model, while the testing data is employed to evaluate its performance.
6. The trained models are then utilized to make predictions on unseen data, known as the test dataset. This process entails using a subset of images from the test dataset to assess and measure the performance of the models.

IMPLEMENTATION

We have used python’s library containing Keras and TensorFlow for implementation in this study. Keras is an open-source neural network library that includes the MNIST dataset.

(a) Initially, we imported all the essential libraries necessary for training the model.
(b) Retrieve the MNIST dataset and load it using the built-in library provided by Keras.
(c) Preparation of the data for training, including various pre-processing steps.
- During this step, we import the data and split it into training and testing variables.
- Restructuring the data’s shape and performing normalization.
- Generating a validation set of images.

(d) Developing multiple deep learning models and conducting a comparative analysis of their recognition accuracies.
- Developing an Artificial Neural Network (ANN) Model.
- Building a CNN Model that includes a Convolutional Layer and a Max Pooling Layer.
- Constructing a Recurrent Neural Network (RNN) Model.

(e) Configuring the Model with an Optimizer for Enhanced Performance.
(f) Training the Model with Multiple Epochs and Batch Size while Conducting Cross-Validation.
- Visualizing the Loss and Accuracy across Epochs.

(g) Assessing the Model Accuracy on the Test Dataset.

RESULT ANALYSIS

In this research, we assessed different deep learning models using the widely known MNIST dataset, which consists of approximately 70,000 gray-scale handwritten digit images ranging from 0 to 9. The dataset was already divided into training and testing sets, with 60,000 images used for training and 10,000 for testing. Each image had a resolution of 28×28 pixels. The MNIST dataset has 10 different classes.

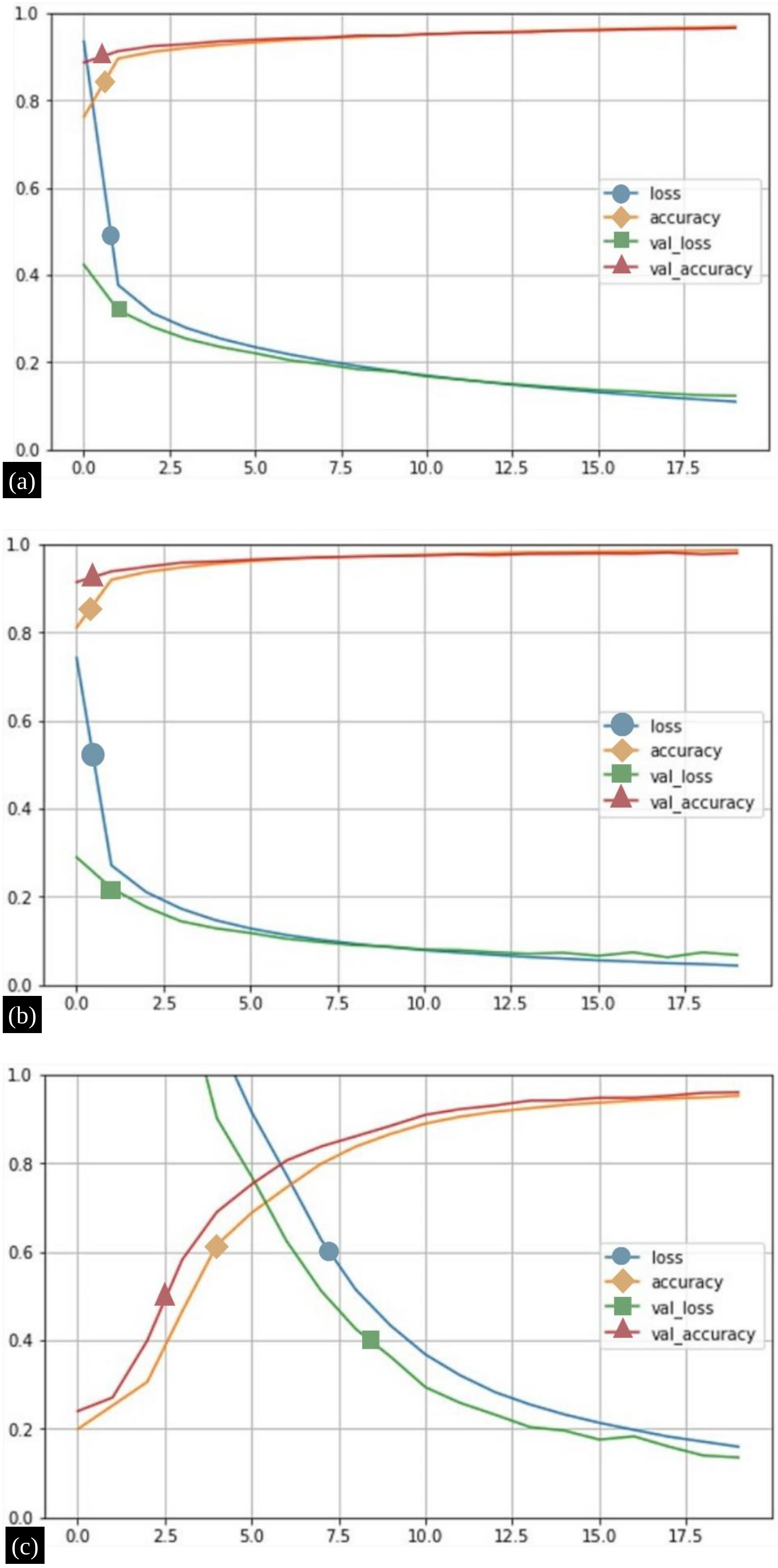

Figure 5. Performance of various deep learning models under MNIST dataset: (a) ANN, (b) CNN, and (c) RNN.

Upon conducting training and testing on the various deep learning models, namely ANN, CNN, and RNN, we obtained recognition accuracies of 97.73, 96.61 and 96.13%, respectively under different hyper parameters (i.e., batch size, epoch, no. of classes, etc.). The findings clearly indicated that the CNN model performed better than the other models in accurately recognizing handwritten digits. The summarized results are presented in Table 1. The results are also verified with graphical representations (Figure 5).

Table 1. Performance Analysis of Popular Deep Learning Models Under MNIST Dataset.

S.N. (#)	Deep Learning Models	Image Size (Rows × Columns)	No. of Classes (#)	Accuracy (%)
1.	ANN	28×28	10	96.61
2.	**CNN**	28×28	10	**97.73**
3.	RNN	28×28	10	96.13

CONCLUSION

This study aimed to investigate the recognition of handwritten digits using various deep learning models and the popular MNIST dataset. The study involved a comparative analysis of these models, considering different hyper-parameters such as batch size, epochs, and the number of classes, with the ultimate goal of achieving higher accuracy. The results indicated that the CNN model achieved the highest recognition accuracy of 97.73%, followed by the ANN model with 96.61%, and the RNN model with 96.13%. The results led to the conclusion that the CNN model exhibited superior performance in terms of recognition accuracy compared to the ANN and RNN models. The graphical representation of the data also offered further evidence to corroborate this discovery.

REFERENCES

1. Goodfellow I, Bengio Y, Courville A. Deep learning. Massachusetts: MIT press; 2016 Nov 10.
2. Krizhevsky A, Hinton G. Learning multiple layers of features from tiny images. Technical report, University of Toronto; 2009.
3. Shyam R, Singh R. A taxonomy of machine learning techniques. *J Adv Robot*. 2021; 8(3): 18–25.
4. LeCun Y, Cortes C, Burges CJ. MNIST handwritten digit database. 2010. URL http://yann. lecun. com/exdb/mnist. 2010; 7(23): 6.
5. Aggarwal CC. Neural networks and deep learning. Cham: Springer; 2018 Sep; 10(978): 3.
6. Wan L, Zeiler M, Zhang S, Cun YL, Fergus R. Regularization of neural networks using dropconnect. In *Proceedings of Proceedings of the International Conference on Machine Learning*. 2013; 1058–1066.
7. Shyam R, Chakraborty R. Machine learning and its dominant paradigms. *J Adv Robot*. 2021; 8(2): 1–10.
8. Zhang X, Zhou Y, Lin M, Sun J. Shufflenet: An extremely efficient convolutional neural network for mobile devices. In *Proceedings of the IEEE Conference on Pattern Recognition*. 2015; 684–691.
9. Simran, Tandon S, Khanna S, Shyam R. Detection of traffic sign using CNN. *Recent Trends Parallel Comput*. 2022; 9(1): 14–23.
10. Pandey A, Shyam R. Analysis of road lane detection using computer vision. *International Journal of Software Computing and Testing* (IJSCT). 2023; 9(1): 7–14.
11. Simard PY, Steinkraus D, Platt JC. Best practices for convolutional neural networks applied to visual document analysis. In *Proceedings of International Conference on Document Analysis and Recognition*. 2003; 958–962.
12. Srivastava V, Shyam R. Enhanced object detection with deep convolutional neural networks. *Int J All Res Educ Sci Methods*. 2021; 9(1): 4093–4102.
13. Verma S, Jaiswal V, Shyam R. Intensifying security with smart video surveillance. *Recent Trends Program Lang*. 2022; 9(1): 24–30.
14. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput*. 1997; 9(8): 1735–1780.

15. Ciresan D, Meier U, Schmidhuber J. Multi-column deep neural networks for image classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2012; 3642–3649.
16. Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proc IEEE*. 1998; 86(11): 2278–2324.
17. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556 [cs.CV]. 2015.
18. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016 Jun; 770–778.
19. Graves A, Mohamed AR, Hinton G. Speech recognition with deep recurrent neural networks. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. 2013; 6645–6649.
20. Shyam R. Convolutional neural network and its architectures. *J Comput Technol Appl*. 2021; 12(2): 6–14.
21. Brownlee J. (2019 May 8). How to Develop a CNN for MNIST Handwritten Digit Classification. [Online]. Machine Learning Mastery. Available from: https://machinelearningmastery.com/how-to-develop-a-convolutional-neural-network-from-scratch-for-mnist-handwritten-digit-classification/
22. Ahlawat S, Choudhary A, Nayyar A, Singh S, Yoon B. Improved handwritten digit recognition using convolutional neural networks (CNN). *Sensors*. 2020; 20(12): 3344.
23. Pham V, Bluche T, Kermorvant C, Louradour J. Dropout improves recurrent neural networks for handwriting recognition. *2014 14th International Conference on Frontiers in Handwriting Recognition*, Hersonissos, Greece. 2014; 285–290.
24. Jarrett K, Kavukcuoglu K, Ranzato MA, Yann LeCun. What is the best multi-stage architecture for object recognition? In *2009 IEEE 12th International Conference on Computer Vision*. 2009; 2146–2153.
25. Gupta A, Sarkhel R, Das N, Kundu M. Multiobjective optimization for recognition of isolated handwritten indic scripts. *Pattern Recognit Lett*. 2019; 128: 318–325.
26. Nguyen CT, Khuong VTM, Nguyen HT, Nakagawa M. CNN based spatial classification features for clustering offline handwritten mathematical expressions. *Pattern Recognit Lett*. 2020; 131: 113–120.
27. Hua C, Jie B. A new hyperparameters optimization method for convolutional neural networks. *Pattern Recognit Lett*. 2019; 125: 828–834.
28. William TW, Baris B, Efstratios PN. Hy-pop: Hyperparameter optimization of machine learning models through parametric programming. *Comput Chem Eng*. 2020; 139: 106902.
29. Jian S, Kaiming H, Shaoqing R, Xiangyu Z. Deep residual learning for image recognition. InIEEE Conference on Computer Vision & Pattern Recognition 2016 (pp. 770-778).
30. Christian S, Wei L, Yangqing J, Pierre S, Scott R, Dragomir A, Dumitru E, Vincent V, Andrew R. Going deeper with convolutions. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA. 2015; 1–9.
31. Krizhevsky A, Sutskever I, Hinton G. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems* 25, NIPS. 2012.

