International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

Automatic Handwritten Digit Recognition Model

Kadarla Sahiti[1] , Kura Akhila[2], Maheshwari Ponnan Kumarancheril[3], D. Archana[4]

Department of Computer Science and Engineering, Stanley College of Engineering and Technology for Women, Telangana, India

Abstract.

The Human Visual System is a marvel of the world. People can readily recognise digits. But it is not as simple as it seems. The human brain has a million neurons and billions of connections between them, which makes this exceptionally complex task of image processing easier. People can effortlessly recognize digits. However, it turns into a challenging task for computers to recognize digits. Simple hunches about how to recognize digits become difficult to express algorithmically. Moreover, there is a significant variation in writing from person to person, which makes it immensely complex. The MNIST digit recognition system is the working of a machine to train itself so that it can recognize digits from different sources like emails, bank cheques, papers, images, etc. Convolutional Neural Networks (CNNs) are very effective in perceiving the structure of handwritten characters/words in ways that help in automatic extraction of distinct features and make CNN the most suitable approach for solving handwriting recognition problems. Our aim in the proposed work is to explore the various design options like number of layers, stride size, kernel size for CNN-based handwritten digit recognition. Our objective is to achieve comparable accuracy by using a pure CNN architecture without ensemble architecture, as ensemble architectures introduce increased computational cost and high testing complexity. Thus, a CNN architecture is proposed in order to achieve accuracy, along with reduced operational
complexity and cost.

Keywords: Convolutional Neural Networks, Handwritten Digit Recognition, Pre-processing, Banking Sector

1. Introduction

1.1 About The Article

The aim of a handwriting recognition system is to convert handwritten characters into machine readable formats. Recognition is identifying or distinguishing a thing or an individual from the past experiences or learning. Similarly, Digit Recognition is nothing but recognizing or identifying the digits in any document. Digit recognition framework is simply the working of a machine to prepare itself or interpret the digits. Handwritten Digit Recognition is the capacity of a computer to interpret the manually written digits from various sources like messages, bank cheques, papers, pictures, and so forth and in various situations for web based handwriting recognition on PC tablets, identifying number plates of vehicles, handling bank cheques,

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

digits entered in any forms etc.

1.2 Objectives of the Article

Handwriting varies from person to person and predicting the accurate one each time might be challenging. These challenges can be broadly categorised into Connected Digits, Disjoint Digits and Overlapped Digits. The primary objective of this application is to identify the digits with utmost accuracy i.e., by overcoming the challenges and predicting the correct digit value. To digitize data at workplaces where it is entered manually by the non digitized employees, for instance, at warehouses, by transforming the manually written data into digitized version, misreading of the information and errors can be avoided. To feed this data on to the organization's database as it can be protected from natural calamities and disasters as once mounted onto the cloud, such issues may not rise.

1.3 Scope of the Article

Banks provide deposit and withdraw forms to its customers to get details about the transaction. Different handwritings of people can be misread and this brings a huge loss either to the customer or to the bank. TThe main applications are vehicle license-plate recognition, postal letter-sorting services, Cheque truncation system (CTS) scanning and historical document preservation in archaeology departments, old documents automation in libraries and banks, etc. At warehouses the data entry done by the non digital employees can now be digitalized and mounted onto the database.

2. Literature Survey

2.1 Existing System

These days, an ever-increasing number of individuals use pictures to transmit data. It is additionally mainstream to separate critical data from pictures. Image Recognition is an imperative research area for its generally used applications. In general, the field of pattern recognition, one of the difficult undertakings is the precise computerized recognition of human handwriting. Without a doubt, this is a very difficult issue because there is an extensive diversity in handwriting from an individual to another individual. In spite of the fact that this difference does not make any issues to people, yet, anyway it is increasingly hard to instruct computers to interpret general handwriting. For the image recognition issue, for example, handwritten classification, it is essential to make out how information is depicted onto images [2].

Handwritten Recognition from the MNIST dataset is well known among scientists as by utilizing different classifiers for various parameters, the error rate has been decreased, for example, from linear classifier (1-layer NN) with 12% to 0.23% by a board of 35 convolution neural systems. For an undertaking of composing diverse digits (0-9) for various people the general issue confronted would be the digit order issue and the closeness between the digits like 1 and 7, 5 and 6, 3 and 8, 9 and 8 and so forth [3]. Additionally, individuals compose a similar digit from various perspectives, the uniqueness and assortment in the handwriting of various people likewise impact the development and presence of the digits.

2.2 Proposed System

Convolutional Neural Networks (CNN) has become one of the most appealing approaches and has been an ultimate factor in a variety of recent success and challenging machine learning applications such as challenge ImageNet object detection image segmentation and face recognition. Therefore, we choose CNN for our challenging tasks of image classification. We can use it for handwritten digits recognition which is one of high academic and business transactions. There are many applications of handwritten digit recognition in our real life purposes. Precisely, we can use it in banks for reading checks, post offices for sorting letters, and many other related works [4].

Deep Learning has emerged as a central tool for self-perception problems like understanding images, a voice from humans, robots exploring the world. We aim to implement the concept of Convolutional Neural Network for digit recognition. Understanding CNN and applying it to the handwritten digit recognition system is the target of the proposed model. Convolutional Neural Network extracts the features maps from the 2D images. Then it can classify the images using the features maps. The convolutional neural network considers the mapping of image pixels with the neighborhood space rather than having a fully connected layer of neurons. The convolutional neural network is a powerful tool in signal and image processing. Even in the fields of computer vision such as handwriting recognition, natural object classification, and segmentation, CNN has been a much better tool compared to all other previously implemented tools. The broader aim is to develop a machine learning model that could recognize people's handwriting along with exception cases like disjoint digits, inter-connected digits, overlapped digits and poorly written digits.

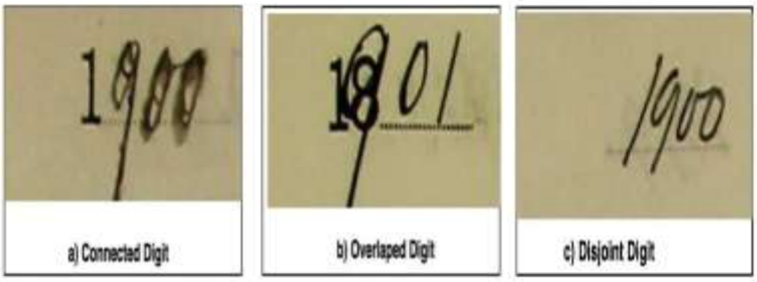


Fig. 1. Exception Cases

3. Proposed Architecture

Fig. 2 illustrates the architecture diagram of the proposed system. The proposed model contains the four stages in order to classify and detect the digits:

A. Pre-processing

B. Segmentation

C. Feature Extraction

D. Classification and Recognition

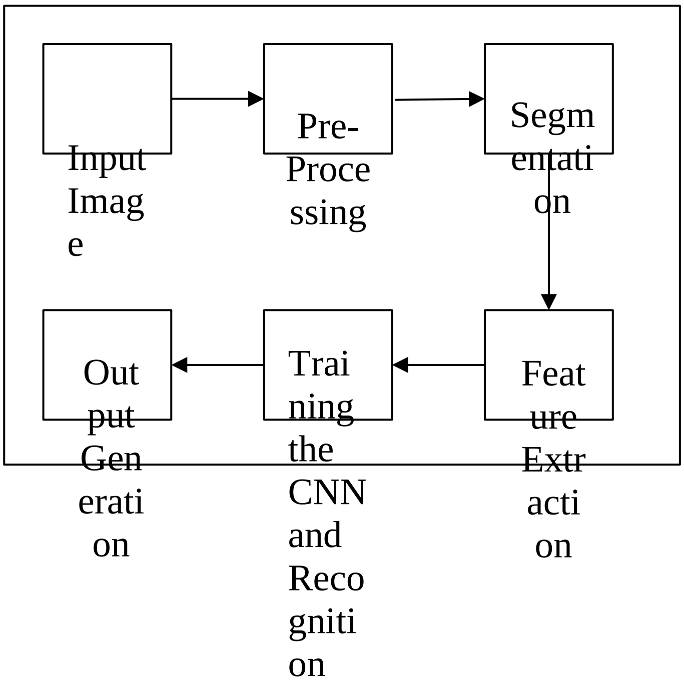


Fig. 2. Architecture of the Proposed System

3.1 Pre-Processing

The role of the pre-processing step is to perform various tasks on the input image. It basically upgrades the image by making it reasonable for segmentation. The fundamental motivation behind pre-processing is to take off a fascinating example from the background. For the most part, noise filtering, smoothing and standardization are to be done in this stage. The pre-processing additionally characterizes a smaller portrayal of the example. Binarization changes over a gray scale image into a binary image. The initial approach to the training set images that are to be processed in order to reduce the data, by thresholding them into a binary image.

MNIST Dataset. The MNIST database (Modified National Institute of Standards and Technology database) is a large database of handwritten digits that is commonly used for training various image processing systems. The database is also widely used for training and testing in the field of machine learning. It was created by "re-mixing" the samples from NIST's original datasets [5]. The MNIST database contains 60,000 training images and 10,000 testing images.

Fig. 3. MNIST Dataset

DIDA Dataset. The DIDA dataset has been collected from the Swedish historical handwritten document images between the year 1800 and 1940 and it is the largest historical handwritten digit dataset to develop handwritten digit recognition methods. To generate DIDA, 250, 000 single digits and 200, 000 multi-digits are cropped from different document images. The DIDA dataset consists of 3 different subdatasets: 1) single digit samples, 2) digit string samples, and 3) digit images with bounding box annotations [6].

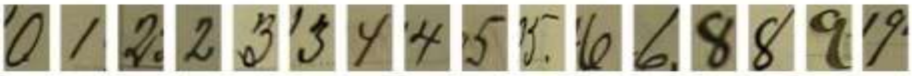


Fig. 4. DIDA Dataset

3.2 Segmentation

Once the pre-processing of the input images is completed, sub-images of individual digits are formed from the sequence of images. Pre-processed digit images are segmented into a sub-image of individual digits, which are assigned a number to each digit. Each individual digit is resized into pixels. In this step an edge detection technique is being used for segmentation of dataset images. Edge detection is a technique of image processing used to identify points in a digital image with discontinuities, simply to say, sharp changes in the image brightness. These points where the image brightness varies sharply are called the edges (or boundaries) of the image [7].

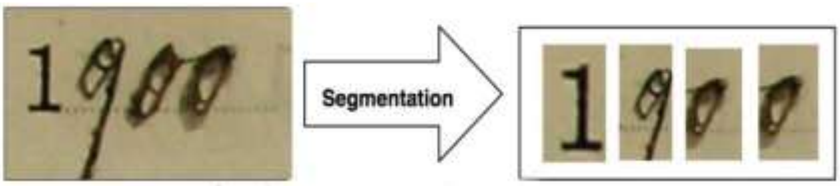


Fig. 5. Segmentation of Inter-Connected Digits

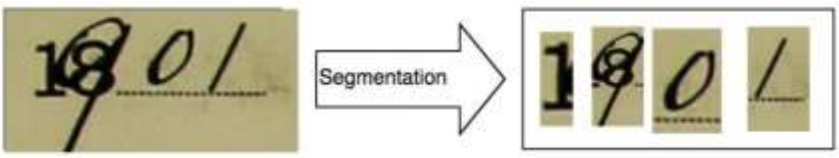


Fig. 6. Segmentation of Overlapped Digits

3.3 Feature Extraction

After the completion of the pre-processing stage and segmentation stage, the pre-processed images are represented in the form of a matrix which contains pixels of the images that are of very large size. In this way it will be valuable to represent the digits in the images which contain the necessary information. This activity is called feature extraction. In the feature extraction stage redundancy from the data is removed [8].

3.4 Classification and Recognition

In the classification and recognition step the extracted feature vectors are taken as an individual input. In order to showcase the working system model, extracted features are combined and defined using CNN.

Convolutional Neural Networks(CNNs). Convolutional Neural Networks are a special kind of multi-layer neural networks designed to recognize visual patterns directly from pixel images with minimal preprocessing. The typical convolutional neural network architecture with three convolutional layers is well adapted for the classification of handwritten images as shown in Figure 3.1.1 and Figure 3.1.2. It consists of the input layer, multiple hidden layers (repetitions of convolutional, normalization, pooling) and a fully connected and an output layer. Neurons in one layer connect with some of the neurons present in the next layer, making the scaling easier for the higher resolution images. The operation of pooling or sub-sampling can be used to reduce the dimensions of the input. In a CNN model, the input image is considered as a collection of small sub-regions called the “receptive fields” [9]. A mathematical operation of the convolution is applied on the input layer, which emulates the response to the next layer. The response is basically a visual stimulus. The detailed description is given below:

Input Layer. The input data is loaded and stored in the input layer. This layer describes the height, width and number of channels (RGB information) of the input image.

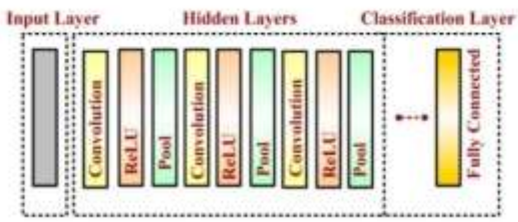


Fig. 7. Typical Convolutional Neural Network Architecture

Hidden Layers. The hidden layers are the backbone of CNN architecture. They perform a feature extraction process where a series of convolution, pooling and activation functions are used. The distinguishable features of handwritten digits are detected at this stage.

Convolutional Layer. The convolutional layer is the first layer which can extract features from the images. Because pixels are only related to the adjacent and close pixels, convolution allows us to preserve the relationship between different parts of an image. Convolution is filtering the image with a smaller pixel filter to decrease the size of the image without losing the relationship between pixels. When we apply convolution to the 5x5 image by using a 3x3 filter with 1x1 stride (1-pixel shift at each step), we will end up having a 3x3 output (64% decrease in complexity) [10].

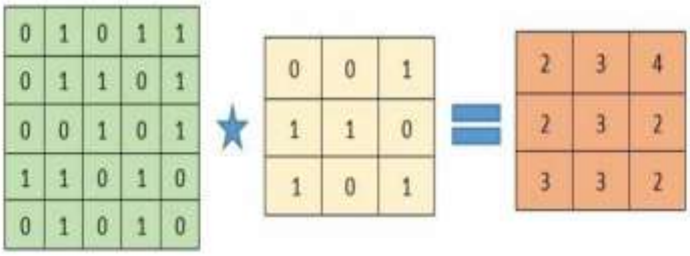


Fig. 8. Convolution Operation

Pooling Layer. When constructing CNN, it is common to insert pooling layers after each convolution layer to reduce the spatial size of the features maps. Pooling layers also help with the overfitting problem. We select a pooling size to reduce the amount of the parameters by selecting the maximum, average, or sum values inside these pixels. Max Pooling, is demonstrated as follows:

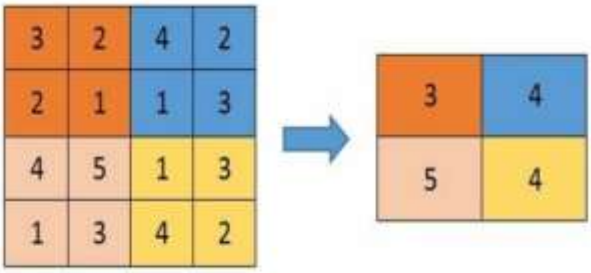


Fig. 9. Max Pooling Operation

Activation Layer. Just like regular neural network architecture, CNN architecture also contains the activation function to introduce the non-linearity in the system. It has been observed that the sigmoid activation function might weaken the CNN model because of the loss of information present in the input data. The activation function used in the present work is the non-linear rectified linear unit (ReLu) function, which has output 0 for input less than 0 and raw output otherwise. Some advantages of the ReLu activation function are its similarity with the human nerve system, simplicity in use and ability to perform faster training for larger networks [11].

Fully Connected Layer. A fully connected network is in any architecture where each parameter is linked to

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

one another to determine the relation and effect of each parameter on the labels. Since convolution and pooling layers reduce time-space complexity, we can construct a fully connected network in the end to classify the images. The neurons in the fully connected layers are connected to all the neurons of the previous layer. This layer calculates predicted classes by identifying the input image, which is done by combining all the features learned by previous layers. The number of output classes depends on the number of classes present in the target dataset. In the present work, the classification layer uses the 'softmax' activation function for classifying the generated features of the input image received from the previous layer into various classes based on the training data.

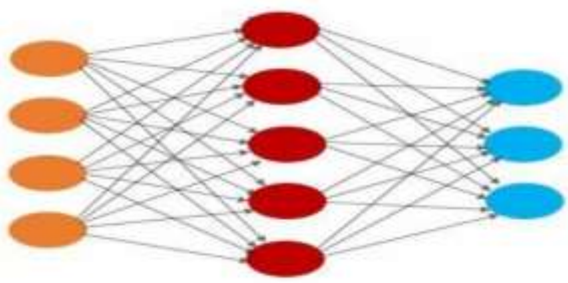
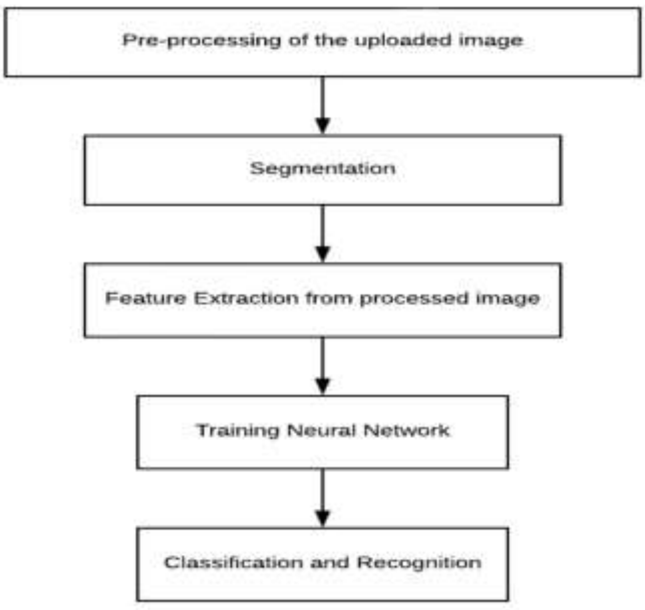


Fig. 10. Fully Connected Layers

4. Implementation

The implementation of handwritten digit recognition by Convolutional Neural Network is done using Keras. The initial step to be carried out is to place the dataset, which can be effectively done through the Keras programming interface. On running the system code the output is generated that shows which is the number input by the user.



International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

Fig. 11. Flow of Implementation

The CNN model works in the following sequence. User writes a particular number or digits which he wants to recognize. The drawn number is captured and the captured image will be processed by the system. On running the system code the output is generated that shows which is the number or digits input by the user.

The role of the pre-processing step is to perform various tasks on the input image.For the most part, noise filtering, smoothing and standardization are to be done in this stage. Once the pre-processing of the input images is completed, sub-images of individual digits are formed from the sequence of images.Pre-processed digit images are segmented into a sub-image of individual digits, which are assigned a number to each digit.After the completion of pre-processing stage and segmentation stage, the pre-processed images are represented in the form of a matrix which contains pixels of the images that are of very large size. In the classification and recognition step the extracted feature vectors are taken as an individual input to the CNN model.

4.1 Algorithm

- Convolutional Neural Network is a deep learning algorithm that is widely used for image recognition and classification. It is a class of deep neural networks that require minimum pre-processing. The implementation of handwritten digit recognition by Convolutional Neural Network is done using Keras.
- From Keras, we have used a Sequential class which allowed us to create model layer-by-layer. The dimension of the input image is set to 28(Height), 28(Width), 1(Number of channels).
- Next, we created the model whose first layer is a Conv layer. This layer uses a matrix to convolve around the input data across its height and width and extract features from it. This matrix is called a Filter or Kernel. The values in the filter matrix are weights. We have used 32 filters each of the dimensions (3,3) with a stride of 1.

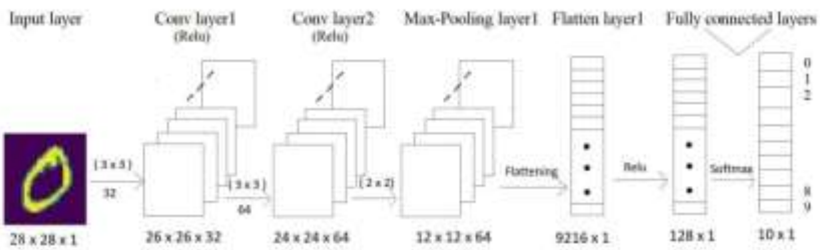


Fig. 12. Detailed Architecture of Convolutional Neural Network

- Stride determines the number of pixel shifts. Convolution of filters over the input data gives us activation maps whose dimension is given by the formula: $((N + 2P - F)/S) + 1$ where N= dimension of input image, P= padding, F= filter dimension and S=stride. In this layer, Depth (number of channels) of the output image is equal to the number of filters used.

- To increase the non-linearity, we have used an activation function that is Relu. Next, another convolutional layer is used in which we have applied 64 filters of the same dimensions (3,3) with a stride of 1 and the Relu function.
- We have employed MAX-pooling which keeps only the maximum value from a pool. The depth of the network remains unchanged in this layer. We have kept the pool-size (2,2) with a stride of 2, so every 4 pixels will become a single pixel. To avoid overfitting in the model, a Dropout layer is used which drops some neurons which are chosen randomly so that the model can be simplified.
- We have set the probability of a node getting dropped out to 0.25 or 25%. Following it, Flatten Layer is used which involves flattening i.e. generating a column matrix (vector) from the 2-dimensional matrix. This column vector will be fed into the fully connected layer. This layer consists of 128 neurons with a dropout probability of 0.5 or 50%.
- After applying the Relu activation function, the output is fed into the last layer of the model that is the output layer. This layer has 10 neurons that represent classes (numbers from 0 to 9) and the SoftMax function is employed to perform the classification. This function returns probability distribution over all the 10 classes. The class with the maximum probability is the output.

4.2 Code Implementation

Tensorflow. TensorFlow is an amazing information stream in machine learning library made by the Brain Team of Google and made open source in 2015.It is intended to ease the use and broadly relevant to both numeric and neural system issues just as different spaces. Fundamentally, TensorFlow is a low level tool for doing entangled math and it targets specialists who recognize what they're doing to construct exploratory learning structures, to play around with them and to transform them into running programs.

Python 3.7. Python is broadly utilized universally and is a high-level programming language. It was primarily introduced for prominence on code, and its language structure enables software engineers to express ideas in fewer lines of code. Python is a programming language that gives you a chance to work rapidly and coordinate frameworks more effectively.

Anaconda3 5.3.1. Anaconda is a free and open-source appropriation of the Python and R programming for logical figuring like information science, AI applications, large-scale information preparing, prescient investigation, and so forth. Anaconda accompanies in excess of 1,400 packages just as the Conda package and virtual environment director, called Anaconda Navigator, so it takes out the need to figure out how to introduce every library freely. to Anaconda appropriation that enables clients to dispatch applications and oversee conda packages, conditions and channels without utilizing command line directions.

5. Result

After implementing the algorithm CNN, we have determined the accuracy and execution time with the help of experimental graphs for perspicuous understanding. We have taken into account the Training and Testing Accuracy of the model. After executing the model, we found that on the testing dataset CNN accomplishes utmost accuracy. Generally, the running time of an algorithm depends on the number of operations it has performed. So, we have trained our deep learning model up to 30 epochs.

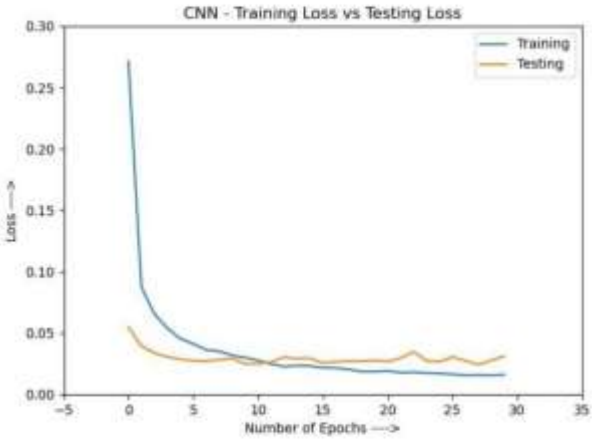


Fig. 13. Loss rate v/s Number of epochs

The above figure or graph illustrates the transition of training loss with increasing number of epochs in the Convolutional Neural Network(CNN) Model. The training loss initially is high and decreases with increasing number of epochs whereas, the Testing loss comparatively is lesser than training loss initially and decreases furthermore with increasing number of epochs.

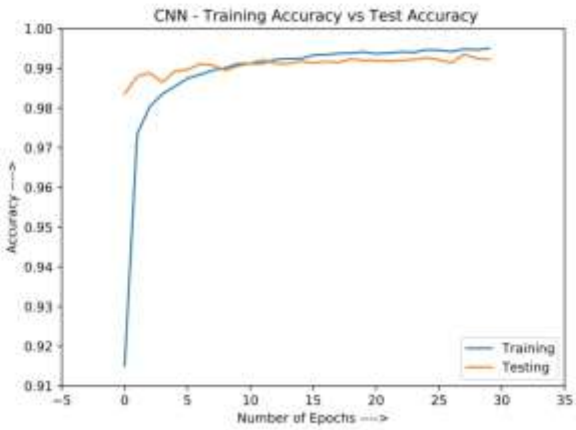


Fig. 14. Accuracy v/s Number of epochs

The above figure or graph illustrates the transition of training accuracy with increasing number of epochs in the Convolutional Neural Network(CNN) Model. The training accuracy initially is low and increases with

International Journal for Innovative
Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

increasing number of epochs whereas, the Testing accuracy comparatively is higher than training accuracy initially and increases furthermore or stays saturated with increasing number of epochs.

Testing accuracy is 99.15% which implies that the model is trained well for prediction. Training set size affects the accuracy and accuracy increases as the number of data increases. The more data in the training set, the smaller the impact of training error and test error, and ultimately the accuracy can be improved. Although there are some digits which are not good handwriting, our model will be able to classify them correctly.

Among 10,000 test cases, our model misclassifies a total 85 digits after eight epochs which correspond to 99.15% recognition rate. The results are pretty good for such a simple model with CPU training and less training time. The model works on different exception cases such as, poorly handwritten digits, Dis-joint digits, Connected digits, Overlapped digits.

Fig. 14a. Poorly Written Digit

Fig. 14b. Disjoint Digit

Fig. 15a. Interconnected Digit

Fig. 15b. Overlapped Digit

6. Conclusion

In this research, we have implemented the CNN model for handwritten digit recognition using MNIST datasets, based on deep and machine learning algorithms. We compared them based on their characteristics to appraise the most accurate model among them. We have found that CNN gave the most accurate results for handwritten digit recognition. So, this makes us conclude that CNN is best suitable for any type of prediction problem including image data as an input. Next, by comparing execution time of the algorithms we have concluded that increasing the number of epochs without changing the configuration of the algorithm is useless because of the limitation of a certain model and we have noticed that after a certain number of epochs the model starts overfitting the dataset and give us the biased prediction. Using Keras as backend and Tensorflow as the software, a CNN model is able to give accuracy of about 98.72%. The final outcome of the Handwritten Digit Recognition resulted in successful prediction of the digits by overcoming the challenges like Overlappped and Disjoint Digits.

7. Future Scope

The future development of the applications based on algorithms of deep and machine learning is practically boundless. In the future, we can work on a denser or hybrid algorithm than the current set of algorithms with more manifold data to achieve the solutions to many problems. In future, the application of these algorithms lies from the public to high-level authorities, as from the differentiation of the algorithms above and with future development we can attain high-level functioning applications which can be used in the classified or government agencies as well as for the common people, we can use these algorithms in hospitals application for detailed medical diagnosis, treatment and monitoring the patients, we can use it in surveillances system to keep tracks of the suspicious activity under the system, in fingerprint and retinal scanners, database filtering applications. The advancement in this field can help us create an environment of safety, awareness and comfort by using these algorithms in day to day application and high-level application. Applications-based on artificial intelligence and deep learning are the future of the technological world because of their absolute accuracy and advantages over many major problems.

8. References

1. [1] Ahlawat, S.; Rishi, R. A genetic algorithm based feature selection for handwritten digit recognition. Recent Pat. Comput. Sci. 2019, 12, 304–316.
2. [2] Alvear-Sandoval, R.; Figueiras-Vidal, A. On building ensembles of stacked denoising auto-encoding classifiers and their further improvement. Inf. Fusion 2018, 39, 41–52.
3. [3] Cai, Z.W.; Li-Hong, H. Finite-time synchronization by switching state-feedback control for discontinuous Cohen–Grossberg neural networks with mixed delays. Int. J. Mach. Learn. Cybern. 2018, 9, 1683–1695.
4. [4] Choudhary, A.; Ahlawat, S.; Rishi, R. A binarization feature extraction approach to OCR: MLP

vs. RBF. In Proceedings of the International Conference on Distributed Computing and Technology ICDCIT, Bhubaneswar, India, 6–9 February 2014; Springer: Cham, Switzerland, 2014; pp. 341–346.

5. [5] Choudhary, A.; Ahlawat, S.; Rishi, R. A neural approach to cursive handwritten character recognition using features extracted from binarization technique. Complex Syst. Model. Control Intell. Soft Comput. 2015, 319, 745–771.
6. [6] Choudhary, A.; Rishi, R.; Ahlawat, S. Handwritten numeral recognition using modified BP ANN structure. In Proceedings of the Communication in Computer and Information Sciences (CCIS-133), Advanced Computing, CCSIT 2011, Royal Orchid Central, Bangalore, India, 2–4 January 2011; Springer: Berlin/Heidelberg, Germany, 2011; pp. 56–65.
7. [7] Choudhary, A.; Rishi, R. Improving the character recognition efficiency of feed forward bp neural network. Int. J. Comput. Sci. Inf. Technol. 2011, 3, 85–96.
8. [8] Chuangxia, H.; Liu, B. New studies on dynamic analysis of inertial neural networks involving non-reduced order methods. Neurocomputing 2019, 325, 283–287.
9. [9] Ciresan, D.C.; Meier, U.; Masci, J.; Gambardella, L.M.; Schmidhuber, J. High-performance neural networks for visual object classification. arXiv 2011, arXiv:1102.0183v1.
10. [10] Ciresan, D.C.; Meier, U.; Schmidhuber, J. Multi-column deep neural networks for image classification. arXiv 2012, arXiv:1202.2745.
11. [11] Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05), San Diego, CA,USA, 20–26 June 2005; Volume 1, pp. 886–893.
12. MVSPN Dr Y V S Sai Pragathi, Analysis and implementation of realtime stock prediction using reinforcement frameworks, Journal of Physics: Conference Series 2089 (1).2021
13. DYVSS Pragathi, MVSP Narasimham, Realtime stock prediction using Transformations and Modeling, Journal of Contemporary Issues in Business and Government 27 (3), 2129-2135, 2021
14. YVSSP M V S Phani Narasimham,Realtime Cost performance Improved Reservoir Simulator Service using ANN and Cloud Containers, International Journal of Innovative Technology and Exploring Engineering 9 …, 2020
15. P Yellanki, MVSP Narasimham,Secure Routing Protocol for VANETS using ECC, 2020 International Conference on Computer Science, Engineering and …, 2020
16. YVSSP Dumala Anveshini, Real-Time Emulation to Investigate the performance of LANMARK Routing Protocol in MANET, International Journal of Advanced Science and Technology 29 (3), 9395-9404, 2020
17. A Parveen, YVS Sai Pragathi,A study of routing protocols for energy conservation in manets, Advances in Decision Sciences, Image Processing, Security and Computer …, 2020
18. YVSSP M V S Phani Narasimham,Development of Realistic models of oil well by modeling porosity using modified ANFIS technique, International Journal on Computer Science and Engineering 11 (07), 34-39, 2019