International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 6 June 2021, pp: 654-660 www.ijaem.net ISSN: 2395-5252

# Handwritten Digit Recognition using Convolutional Neural Networks

B. Sai Kumar[1], L. Vikhyath[2], R. Geetha Krishna Pavansai[3], Mrs. V. Sudha Rani[4], Dr. S.I. Khan[5]
[1,2,3]*Student,Sreenidhi Institute of Science and Technology, Ghatkesar, Telangana*
[4,5]*Associative Professor,Sreenidhi Institute of Science and Technology, Ghatkesar, Telangana.*

---

Submitted: 25-05-2021 Revised: 01-06-2021 Accepted: 05-06-2021

---

**ABSTRACT**:In recent years, with the emergence of Artificial Neural Networks (ANN), deep learning has become increasingly popular in a variety of disciplines, including surveillance, health, medicine, sports, robotics, and drones. Convolutional Neural Network (CNN) is at the forefront of outstanding developments in deep learning, combining Artificial Neural Network (ANN) and previous deep learning methodologies. Handwritten digits recognition in a computer vision system can be a difficult operation, but it's crucial to a variety of new applications.Machine learning and computer vision researchers used to use it substantially towards development and maintaining practical applications including certain automated check number reading. The aim of this study is to examine the performance of CNN with different number of hidden nodes and periods in the classification of handwritten digits. The CMN's Modified National Institute of Standard and Technology (MNIST) dataset is specified by the framework for this performance assessment. The MNIST database shall collect 60,000 digit training pictures and 10,000 digit analysis images. Our main objective is to develop an image recognition human-made neural network.

**KEYWORDS:**Convolutional Neural Networks, Deep Learning, Handwritten Digits, Image Preprocessing, Image Recognition, Machine Learning, MNIST Database.

## I. INTRODUCTION

The number of domains in which deep learning is mostly used for growing all the time. Convolutional Neural Networking (CNN) is a deep learning technique for assessing visual information. The CNN will be applied to detection of objects, face recognition, robots, video analysis, segmentation, pattern identification, tongue processing, fraud detection, categorization of topics, multivariate analysis, speech acknowledgment, and image classification. Deep Convolution Neural Networks have attained human-like levels of performance in many disciplines, including handwritten digit recognition. The biological model for mammalian visual systems is the one that the architecture of the CNN is based on. It was discovered by D. H. Hubel et al. in 1062 that cells inside the cat's visual region are sensitized to a limited portion of the field of vision known as the receptive field. The fundamental computer perception was the pattern recognition model, which was introduced by Fukushima in 1980 and was inspired by the work of D. H. Hubel et al. LeCun et al. defined the framework of CNNs in 1998, with seven layers of convolutional neural networks. It was a technique for classifying handwritten numerals based on image pixel values. The model is trained using a gradient descent and back propagation approach. Characters are used as input in handwriting recognition digits. The system frequently recognizes the model. A network is made of a single input layer, a layer of output and a few invisible layers between the input and output layers (ANN). CNN's architecture is extremely similar to ANN's. Each layer of the ANN has many neurons. The input of a neuron in the next layer is the weighted total of all the neurons in the previous layer, plus a biased value. The layer in CNN has three dimensions. All of the neurons in this picture aren't fully linked. Instead, every neuron in the layer is connected to the receptive field in its immediate vicinity. To coach the network, a price function is generated. It compares the network's output to the specified output.

This portion of the study explains how pre-processing, selection of features and hence classification algorithms contribute to manually recognizing digits. It also provides a comprehensive and exhaustive assessment of recent literature, such as this study. To begin, this section includes an overview of the approaches to template matching Machine Learning (ML) techniques, as well as references to them. The second section contains a study of the factors that influence the popularity error rate. In addition, the third section will go over the used classification approaches in ML and, as a result, the design evaluation. The ultimate section will provide a summary of subsequent stage within the study the network is trained using the probabilistic gradient and backpropagation techniques, and then validated through using forward strategy.

## II. LITERATURE SURVEY

This section of the paper will explain how picture preprocessing, pattern discovery, and thus classification algorithms aid towards handwritten digit recognition. It also provides a comprehensive and exhaustive assessment of current literature, such as this study. To begin, this section includes an overview of the approaches to template matching Machine Learning (ML) techniques, as well as references to them. The second section expresses a study of the factors that influence the popularity error rate.

In addition, the third section will go over the used classification approaches in ML and, as a result, the design evaluation. The final section will include a description of the study's subsequent stages, as well as what will be done next and what will be provided for the experiment's preparation.

Since handwritten Marathi number are not covered by standard large datasets, Mane & Kulkarni (2018) performed numerous adjustments to emphasize the dimensions of the data set. Scaling, by way of example; Horizontal and Vertical Skewing: each image was positioned and relocated in a randomly generated location, using a 0.5-element vertical and horizontal tilting. The dataset has grown manyfold as a result of these transformations.

Slant correction, thinning, and smoothing are three distinct preprocessing approaches proposed by Hanmandlu and Murthy in their study. The pitch that is defined as the trend to write relative to the vertical line is one of the principal variances in the written text for handwritten characters. Slant correction also has to take place before other preprocessing tasks, as correction 22, generally, generates a harsh character contour, whereas smoothing changes the picture structure.For generating alternative deformed character representations at the desired resolution, bilinear interpolation is useful (29x29). Moreover, defining a continuous thickness "centre line" for each letter for identification purposes is one approach of attaining non-uniform thickness invariance. As a result, the approach is known as Skeletonization. Murthy, Hanmandlu also proposed a substitute approach for smoothing and removing the virtual slant of distorted values. Bui, Sadri, Suen found characteristic locations on the string image that were consistent with the established technique and generated possible segmentation hypotheses, as well as determining which group had the highest segmentation recognition reliability. Platt, Simard and Steinkraus, another group of researchers, suggested that if data is insufficient and the distribution being examined contains transform-invariant properties, using transformations can provide additional data and even increase performance. Elastic deformation, which in handwriting recognition may be carried out by computing a substitute target location in relation with to the initial post, is one of the ways used to increase the training session. Suen, Bloch and Lauer then employed affinity transformations and hence past knowledge of static qualities to quadruple the MINST dataset size.The flexible deformation are performed to every specimen of the training samples to increase nine new samples in the raised approach.

Furthermore, Keysers et al. proposed that using an area deformation methodology, more sophisticated designs (e.g., 2-d warp) aren't necessarily better models than basic picture distortion models.

Digital identification is one of the most important pattern identification applications. Many researchers studied various datasets and recognize them. The US letter zip code, for example, has been gathered in the CEDAR digital data set and used by researchers as a typical database (Sarkhel et al., 2016). As Marathi is currently not provided with a criterium database, a data set of 2000 images with 0-9 marathi numerals from age categories was collected (Mane& Kulkarni, 2018).

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 6 June 2021, pp: 654-660 www.ijaem.net ISSN: 2395-5252

| Work reference | Techniques | Database | RR |
|---|---|---|---|
| Phangtriastu et al., 2017 | SVM, ANN | Chars74K | 94.43% |
| Mane & Kulkarni, 2018 | CCNN | Self created | 94.93% |
| Hochuli et al., 2018 | CNN | NIST SD19 | 97% |
| Mahto et al., 2015 | SVM, KNN | Self created | 98.06% |
| Roy et al., 2017 | DCNN | CEDAR | 90.33% |
| Cecotti, 2016 | K-NN | MNIST | 98.54% |
| Karimi et al., 2015 | Bagging | TMU | 98.06% |

**Table.1.** Comparative description of various databases

Furthermore, the SD19 database of the National Institute of Standards and Technology was infectious around scientists as well (Hochuli et al., 2018). The CMATERdb 3.1.3.3 may be a database containing 171 distinct grayscale classes.However, images within the dataset are not central or even, causing difficult problems in identifying patterns (Roy et al., 2017). Table 1 compares the rates for handwritten characters (RR) or figures from the various databases of classification techniques recognition rates (RR).

The effectiveness of a classification model may depend on the classifier's standard features (Kherallah, Elleuch&Maalej 2016). Regrettably, many classifiers such as SVM and RF cannot efficiently process raw images or data, as the extraction of adequate structural properties from complex forms can be a significant challenge (Bag &Pramanik, 2018). The main problem with OCR in handwritten digital recognition is therefore the way to apply the combination of sophisticated extraction and classification features.

An experiment in 2017 by Tanoto, Phangtriastu, Harefa compared the main commonly used SVM and ANN graders and achieved 94,43 percent of the best possible accuracy using the SVM gradation with the combination of projection histogram and HOG functional extracting algorithms.

## III. METHODOLOGY

**DATA COLLECTION:** On the Y. Lecun website you are often searched for a commonly used data set for handwritten digit recognition, called MNIST. This is a collection of 70, thousand digits by the different 750 staff and high school students of the Census office. This dataset can be a well-known criterion with a training set of almost 60,000 pictures and a test set of 10,000. The numbers have been sized, centred and sequentially saved in graylevel bitmaps with 28 to 28 pixels. Each image contains one digit. The results are given to the label. The data were applied in the experiments below in this ready-to-use dataset. Figure 1 shows a number of samples in the training set.

**Figure 1:** The sample numbers in the training set

The MNIST database consists of the unique NIST 3 and 1 database consisting of hand-written graphics. In instance, there are 30 000 SD3 patterns and 30,000 SD-1 patterns on the MNIST training system, which includes examples from around 250 authors. The test set has 5,000 SD-3 patterns and 5,000 SD-1 patterns. The SD-3 for the training set and the SD-1 for the test set were previously identified by NIST. SD-3 is more accessible and more susceptible to identification than SD-1.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 6 June 2021, pp: 654-660 www.ijaem.net ISSN: 2395-5252

**CNN:** A CNN may be a multiple layer neuronal feed-forward system with a profoundly supervised learning architecture, a combination of two components: an automatic extractor and an exercisable classifier. The classifier and the weight of the method for background spread are used in the extractor feature. Convolution Neural Network (CNN) can also retrieve properties for topological photos. The initial picture in the first layer is summarised and the final layer layout is classified. Mission pattern recognition was implemented with the simplest property. For example, Hochuli et al. (2018) suggested that the CNN classification be more efficient than all the algorithms in the segmentation provided in the literature in handling the complexities of touch numbers to achieve a detection performance of 97%.

The CNN is used to identify highly complex data and to vary the demand for convolution and sub-sampling layers. It is a different structure. The main 34 layer generally consists of the alternating of the convolution layer and hence the layer or co-sample sub-sample.Theconvolutionary layer is used for removal from the local receptive domain of essential visual properties. During an aeroplane it is organised, known as an easy neuron unit, also known as feature mapping. Each group has 25 inputs, the local area receptive, connected to the 5/5 area of the input frame. In addition, a two-ratio is available for downsampling through convolution filtering.

For various problems, such as visual perception and recognition of handwriting nature, many CNN constructions are proposed. In addition, CNN blends three primary areas of hierarchy, including the receptive area, weight sharing and spatial subsequent sampling, to ensure a certain level of invariant in the scope, shift and distortion. Each connection for the neural quality network has trainable weights, but all feature elements share the same weight. The very fact that the first feature sensors useful for some of the picture can be helpful throughout the image also demonstrates this characteristic. Weight sharing methods also leave the number of trainable parameter with a discount.

Initially, all the modules that we need to train our model are being imported. Some datasets are available in the Keras library and one of them is MNIST. We can therefore easily import and start to work with the dataset. The method mnist.load data() returns training data, labels and test data and labels.
The data of the image cannot be transmitted directly to the model, so that the info is ready for our neural network to be processed. The training data dimension (60000, 28, 28). The CNN model would need a different dimension to shape the matrix (60000, 28, 28, 1).

A CNN model usually has overlapping and pooling layers. It works better for grid-structured data. This is often the reason why CNN works well in the case of image classification issues. The drop-out layer deactivates specific neurons and lowers the fit of the model throughout exercise. Then with the Adam optimizer, we will compile the model.

Keras' model.fit() function will begin the model training. Training data, validation data, times and batch size are required.

The 10,000 pictures in the dataset that can be used to assess how well our model works. The test data were therefore not incorporated into the information training, they are new data for our model. The MNIST data set is well balanced so that we have approximately 99% precision.
To check the performance of our model, we printed the confusion matrix and classification report.
The model is evaluated on test data and the matplotlib.pyplot foresees 25 test images.
The GUI consists of creating the latest entry which we create a window for drawing numbers on a canvas and recognise the digit with a button. The Tkinter library is provided in the standard Python library. In order to predict the digit with a trained model, we constructed an input picture predict digit.
Then we start creating the class App that is responsible for our app's GUI. We create a canvas where we draw the predict_digit() function by capturing the mouse event and by pressing a button.
This Graphical User Interface block is interface of our model where we offer canvas to input the hand written image by writing thereon and therefore the response of the model is shown below it.

## IV. ANALYSIS

In order to gauge our system, we used classification report, confusion matrix and accuracy and loss finder. Further the system is evaluated across prediction and true values and therefore the performance is analyzed. Ultimately, handwritten numbers never seen by the systems are used to test the exact models and show the expected figures.

**CLASSIFICATION REPORT:** A classification report is used to live the classification algorithm standard of the predictions. What are the percentages and how many are false. In particular, the metropic of the classification report will not be forecast in True negatives, False negatives, False positives and True positives.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 6 June 2021, pp: 654-660 www.ijaem.net ISSN: 2395-5252

Classification report

========================

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 1.00 | 1.00 | 1.00 | 980 |
| 1 | 0.99 | 1.00 | 1.00 | 1135 |
| 2 | 1.00 | 1.00 | 1.00 | 1032 |
| 3 | 1.00 | 1.00 | 1.00 | 1010 |
| 4 | 0.99 | 1.00 | 0.99 | 982 |
| 5 | 0.99 | 1.00 | 0.99 | 892 |
| 6 | 1.00 | 0.99 | 1.00 | 958 |
| 7 | 0.99 | 1.00 | 0.99 | 1028 |
| 8 | 1.00 | 0.99 | 0.99 | 974 |
| 9 | 1.00 | 0.98 | 0.99 | 1009 |
| micro avg | 0.99 | 0.99 | 0.99 | 10000 |
| macro avg | 0.99 | 0.99 | 0.99 | 10000 |
| weighted avg | 0.99 | 0.99 | 0.99 | 10000 |

Our model predicted digits with a precision of just about 100% (or 99%) and also re- call of 100%, thus leading to f1-score of just about 99-100% and support around 1000 for every digit.

**CONFUSION MATRIX:** Tons of false positive warnings are generated (These alerts are then manually investigated by investigation team). We had to use the machine to automatically close the false alerts. The assessment criteria for the machine learning model were a metric negative predicted value, which is a percentage of cases identified correctly, based on total negative forecasts of the model.

Confusion matrix

```
======================
[[ 977    0    0    0    0    0    1    1    1    0]
 [   0 1133    0    1    0    0    0    1    0    0]
 [   1    0 1028    0    0    0    0    3    0    0]
 [   0    0    0 1005    0    2    0    2    1    0]
 [   0    0    0    0  980    0    0    0    0    2]
 [   0    0    0    1    0  889    1    1    0    0]
 [   1    2    0    0    0    2  953    0    0    0]
 [   0    2    1    0    0    0    0 1025    0    0]
 [   0    1    1    0    2    0    1    1  967    1]
 [   1    1    0    0    9    2    0    3    2  991]]
```

1. TN/True Negative: if negative, negative is predicted by CNN 1.
2. TP / True Positive: if a case was positive and CNN predicts positive.
3. FN / False Negative: if a positive case, but CNN is negative predicted.
4. FP/False Positive: if a negative case but the CNN was positive predicted.

The confusion matrix shows almost perfect predictions and there are only minute differences across digits with similar geometry.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 6 June 2021, pp: 654-660 www.ijaem.net ISSN: 2395-5252

>>>Test loss: 0.016411524798042955
Test accuracy: 0.9947999715805054
The accuracy of our model is 99.48% and the loss is 1.6%. This shows that the model is almost perfect.

## V. CONCLUSION

In this study, the handwritten numbers were recognized by using machine learning tools to instruct the classification. In addition, the use of Computer Vision techniques has been studied to investigate the effect of image preprocessing, feature extraction, and general accuracy classifier selection. The data set for the experiment was originally an MNIST dataset comprising 60,000 exercises and 10,000 photos of 28 x 28 (0- 255) gray - scale and bitmap formats. The machine learning and character recognition methods are an excellent database while minimizing preprocessing and formatting efforts.

In this study, we assessed the neural network versions in order to prevent extensive pre-processing, costly extraction of features and the high-class ensemble method of a system of standard recognition to improve the performance of hand-written digital recognition. This paper shows the role of several hyper parameters through a comprehensive examination using an MNIST dataset. We have also confirmed that high-quality hyperparameter tuning is essential for improving the performance of the CNN architecture. With the Adam Optimizer for the MNIST database, we achieved an accuracy of 99.48 percent, better than all previous findings.The findings clearly indicate the effect on the performance of handwritten digital detection of the growing number of convolutionary layers in CNN architecture. The study is unusual in that it examines all of the CNN architectural parameters that best recognise the MNIST dataset accuracy. According to peer researchers, a pure CNN model could not match that accuracy. In order to enhance the accuracy and precision of detection in the balance of increased processing cost and high testing complexity, some researchers employed the CNN network ensemble for an analogous dataset.

## VI. FUTURE SCOPE

Various architectures for CNN are often researched in the future, such as CNN hybrid, CNNRNN, CNN-HMM and domain-specific recognition systems. CNN-Larning parameters optimise the quantity of layers, the pace of learning and kernel sizes of convolutionary filters often are investigated with evolutionary algorithms.

The current study is also expanded naturally, so that outcomes may be further enhanced and strengthened. For this purpose, the MNIST benchmark database offers a fantastic data source for machine learning and pattern identification while minimizing preprocessing and formatting effort. All manuscript digit sets, however, are not all normalized or sequentially centered as 28x28 pixel gray-scale pictures in the real scale circumstances. Similar tests with different databases on the array dimension of characteristics and different language scripts such as Chinese, Arabic, French, and others may therefore be essential.

A stimulant subject for future study topics is the complicated recognition difficulty associated with handwriting. For example, when some anonymous handwritten parts of the crime are located on the web and automatically it can be identified that the author may also be a "left-handed guy," that may minimize the number of suspected individuals to be probed. These challenges in classification are often quite difficult, given that the handwriting properties of each class concerned are fairly difficult to discern (Morera et al., 2018). A clear illustration of this may be seen in the gender categorization. Although female writing is more round and consistent than male writing, there are some examples of male writing having a "feminine" look. In the realm of handwritten digital recognition for future work this might be a further exact topic.

## REFERENCES

[1]. Phon-Amnuaisuk, S. (2013). Applying Non-negative Matrix Factorization to Classify Superimposed Handwritten Digits. Procedia Computer Science,24, 261-267. doi:10.1016/j.procs.2013.10.049

[2]. Pramanik, R., & Bag, S. (2018). Shape decomposition-based handwritten compound character recognition for Bangla OCR. Journal of Visual Communication and Image Representation,50, 123-134. doi:10.1016/j.jvcir.2017.11.016

[3]. Rashad, M., Amin, K., Hadhoud, M., &Elkilani, W. (2012). Arabic character recognition using statistical and geometric moment features. 2012 Japan-Egypt Conference on Electronics, Communications and Computers. doi:10.1109/jec-ecc.2012.6186959

[4]. Roy, S., Das, N., Kundu, M., &Nasipuri, M. (2017). Handwritten isolated Bangla compound character recognition: A new benchmark using a novel deep learning approach. Pattern Recognition Letters,90, 15-21. doi:10.1016/j.patrec.2017.03.004

[5]. Rusu, C. (2012). Fast design of efficient dictionaries for sparse representations. 2012 IEEE International Workshop on Machine

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 6 June 2021, pp: 654-660 www.ijaem.net ISSN: 2395-5252

Learning for Signal Processing,1134-1144. doi:10.1109/mlsp.2012.6349795

[6]. Sadri, J., Suen, C. Y., & Bui, T. D. (2007). A genetic framework using contextual knowledge for segmentation and recognition of handwritten numeral strings. Pattern Recognition,40(3), 898-919. doi:10.1016/j.patcog.2006.08.002

[7]. Sarkhel, R., Das, N., Das, A., Kundu, M., &Nasipuri, M. (2017). A multi-scale deep quad tree based feature extraction method for the recognition of isolated handwritten characters of popular indic scripts. Pattern Recognition,71, 78-93. doi:10.1016/j.patcog.2017.05.022

[8]. Sarkhel, R., Das, N., Saha, A. K., &Nasipuri, M. (2016). A multi-objective approach towards cost effective isolated handwritten Bangla character and digit recognition. Pattern Recognition,58, 172-189. doi:10.1016/j.patcog.2016.04.010

[9]. Sethi, I. K., & Chatterjee, B. (1976). Machine Recognition of Hand-printed Devnagri Numerals. IETE Journal of Research,22(8), 532-535. doi:10.1080/03772063.1976.11451104