Digit Recognition of Handwritten Character using CNN

Dr. S. Siva Nageswararao, professor, CSE
Computer Science & Engineering
Narasaraopeta Engineering College
Narasaraopet, India
profssnr@gmail.com

Chimata Triveni
Computer Science & Engineering
Narasaraopeta Engineering College
Narasaraopet, India
trivenchimata6@gmail.com

Chunchu Meghana
Computer Science & Engineering
Narsaraopeta Engineering College
Narasaraopet, India
chunchumeghana@gmail.com

Adimulam Lakshmi
Computer Science & Engineering
Narsaraopeta Engineering College
Narasaraopet, India
adimulamlakshmi910@gamil.com

Amulya Pallikonda
Computer Science & Engineering
Narsaraopeta Engineering College
Narasaraopet, India
amulyapallikonda47@gmail.com

***Abstract*—Predicting the transcribed digits and numerals is the main goal of handwritten character recognition using CNN. Because there are so many different styles, it might be challenging to identify the digits while using manual data. This model accurately predicts the images, which aids in the solution to the problem. For model training, the MNIST dataset is utilized. To improve the performance of the models, preprocessing methods like normalization are used. With great precision, the Convolutional Neural Network architecture is intended to efficiently extract patterns and spatial hierarchies from the input images. TensorFlow, Pandas, Keras, and Numpy modules are needed to do this. The images are categorized into transcribed digits based on the model's evaluation and training. This paper lays a foundation to predict the numbers.**

Keywords— Deep Learning, Handwritten digit Recognition, MNIST, CNN

I. INTRODUCTION

Neural networks [1] are used in deep learning [2], a branch of machine learning that addresses the drawbacks of conventional machine learning models by using neural networks. Those drawbacks include failing to classify the image and failing to recognize the audio. Deep learning algorithms automatically extract features from raw data through the use of a hierarchical framework. Owing to these characteristics, deep learning finds applicability across multiple fields. Deep learning models can recognize patterns with greater complexity.

Deep learning is based on Neural networks. Input, hidden, and output layers are present in every neural network, as Figure 1 illustrates. The input layer takes the raw data, and it can be images, text, numbers, or any other applicable format. The hidden layer, or therein can be multiple based on the requirement. These multiple layers are stacked together and serve as a foundation for neural networks. In this layer, complex computations are performed on the data that is received from the input layer. This data will pass through interconnected nodes. By processing the data through these multiple hidden layers, the network can extract intricate features and patterns. The layer that is output is the final layer. It generates the final output after receiving the information from the last hidden layer that has been processed.

Deep learning includes supervised learning [3], unsupervised learning [4], and reinforcement learning [5], just like machine learning. Using labeled data, supervised learning develops prediction skills. Under supervised learning, the convolutional neural network [6] is included. For pictures and grid data, the Convolutional Neural Network is employed. There are layers to it. Convolutional, pooling, and fully connected layers are among them. Applying filters is the initial layer. The reduction of spatial dimensions is the second layer. The values are finally flattened.

The model's extreme complexity allows for a wide range of application areas. This paper implements the Convolutional Neural Networks for image recognition. The model is built using the MNIST [7] dataset and evaluated.

II. LITERATURE REVIEW

Recent advancements in image identification and recognition using machine learning techniques, particularly deep learning algorithms, have significantly improved accuracy across various datasets. These algorithms have shown remarkable performance on tasks such as object detection, image classification, and facial recognition, revolutionizing image processing capabilities in fields like computer vision and artificial intelligence.

In a study by Norhidayu binti Abdul Hamid et al. [8] three distinct classifiers—SVM, KNN, and CNN were compared for their performance on the MNIST dataset. The Multilayer Perceptron (MLP) faced challenges on this platform, particularly in the difference of the two numbers 9 and 6 accurately due to being stuck in local optima. However, using the CNN model on the Keras platform significantly enhanced performance, even surpassing other classifiers in accuracy.

Similarly, Gawlikowski et al. [9] explored the performance of CNN, DBF, and DNN classifiers on the MNIST dataset. Their comparative analysis revealed that the DNN model achieved the highest accuracy of 98.08%, outperforming the others. The different execution times and error rates of each classifier, however, emphasize how important it is to choose the best model depending on the demands of the particular application and performance indicators.

The method of Principal Component Analysis (PCA) was employed by Yehya Abouelnaga et al. [10] to decrease overfitting in a combined method that Involved KNN and CNN. When compared to using each classifier separately, this special combination generated a huge 0.7% gain in accuracy, which is a considerable improvement.

Researchers are actively improving CNN accuracy by minimizing errors, with recent studies showing that deeper networks perform better due to reduced overfitting to study by Samay et al. [11] Compared to NORB and CIFAR10 on MNIST, their designs exhibit fewer errors in handwriting recognition. One study used a 3-NN arrangement on MNIST and was able to achieve a low error rate of 1.19%. Innovations like the coherent recurrent convolutional network (CRCN) are being explored to extract words from images. Methods like Ncfm (No combination of feature maps) are enhancing CNN performance on MNIST datasets, reaching 99.81% accuracy. CNN is also being applied to improve dark image enhancement and traffic sign verification models in Germany, achieving an accuracy rate of 98%.

Numerous deep learning as well as machine learning strategies have been investigated by researchers, including SVM, RFC, KNN, MLP, and CNN, for handwritten digit recognition. This study compares Ritik Dixit et al. [12] CNN's effectiveness with SVM and KNN, proposing CNN as a superior deep learning method using Keras for MNIST digit recognition. The suggested CNN, utilizing the RMSprop optimizer, achieves high training (99.06%) and testing (98.80%) accuracy through convolutional layer augmentation, pooling, dropout, and model hyper parameter tuning.

III. PROPOSED SYSTEM

Our Model is Proposed based on certain criteria as follows.

- **Data Analysis**
- **Data visualization**
- **Preprocessing techniques**
- **Model creation and evaluation**
- **Accuracy**

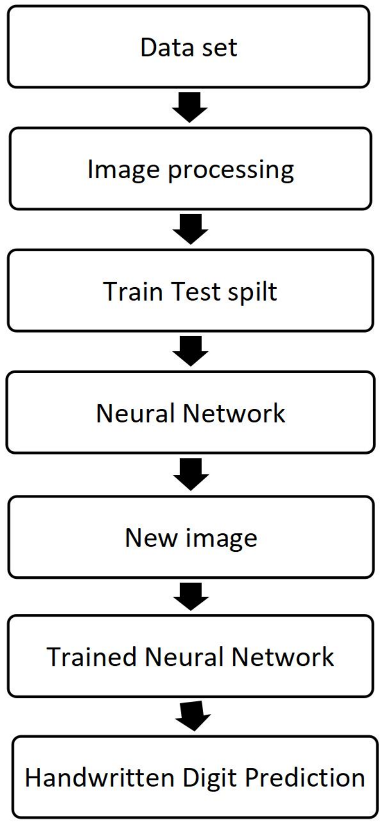

Fig. 1. Workflow for prediction of images

A. Data Analysis

we use the MNIST [7] dataset. The MNIST dataset is imported from the Modified National Institute of Standards and Technology database. To use this first we need to use keras [13]. dataset import mnist. Then the dataset is loaded using the load data function. So that without manual intervention the training data and testing data are separated.

There are 70,000 photos in the MNIST collection, each measuring 28 by 28 pixels. With 60,000 photos in the training set and 10,000 images in the testing set, this dataset is split into training and testing sets. These photos are used as input for the digit recognition process, and the digits 0 through 9 are represented by 10 output class labels.

MNIST [7] is a popular benchmark for deep learning tasks that assesses how well algorithms perform in the area of image categorization, especially when it comes to handwritten digit recognition. It is perfect for testing with different deep learning models, such as Convolutional Neural Networks (CNN's) [6], which are excellent at picture recognition tasks, because of their standardized format and manageable size.

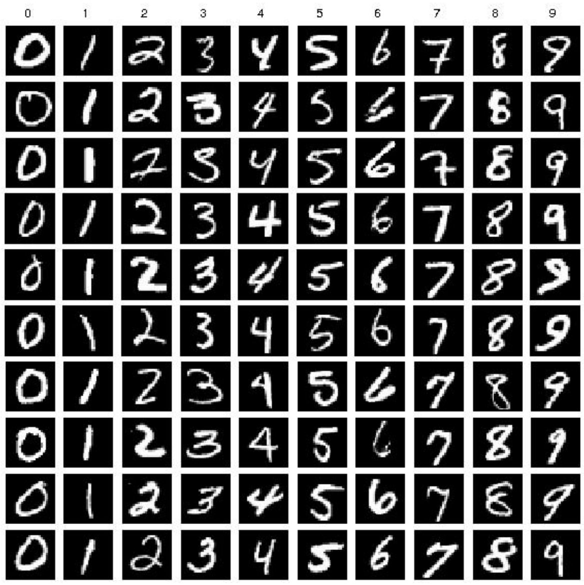

Fig. 2. A sample of MNIST dataset

B. Data visualization

Data visualization is a process conducted to understand the data easily and detect the outliers. Here, the dataset is the MNIST [7] dataset that contains the images from 0 to 9. We should ensure that whether one class is of more samples or the other is too less. So, we visualize the data in a bar graph to know about the number of images that each class has. After plotting, we came to know that all are in the same range approximately and image 1 has the highest samples, and image 5 has the least samples.

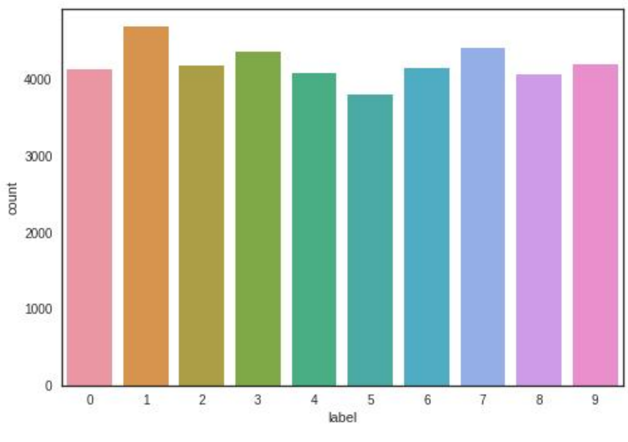

Fig. 3. Visualizing the number of samples

C. Preprocessing techniques

Firstly, the dataset should be collected. The mnist dataset is gathered. Secondly, the preprocessing should be done. The data is not acceptable to feed into the algorithm without preprocessing. For this the techniques used are

- Normalization
- Grayscale Conversion
- One hot encoding

- Normalization:

The data values in the training and testing sets range from 0 to 255. The Min-Max normalization is performed by dividing the pixel values by 255 if the photo is loaded with the pixel values represented by a range between 0 and 255. A popular method for scaling features to a range between 0 and 1 is min-max normalization. We make sure that every pixel value is inside this normalized range by dividing the pixel values by 255. Additionally, by ensuring that the model is less sensitive to the scale of the input characteristics, min-max normalization produces predictions that are more dependable and strong.

- Grayscale Conversion:

Another preprocessing technique we use is the conversion of RGB values to grayscale. By default, the dataset contains the images in RGB format. To save the computational power we convert them to grayscale i.e., the images will be in black and white colors only.

- One hot encoding:

One hot encoding [14] refers to transforming the categorical input to numerical data. We use one-hot encoding to ensure correct predictions and not to prone them to Ordinality. Here the ordinality means that the model doesn't choose 0 or 9 to be the predictions in most cases because 0<1<2<3<4<5<6<7<8<9. This refers to ordinality. So to avoid these the data must undergo encoding.

D. Model creation and evaluation

The construction of the model comes next after preprocessing. The appropriate algorithm is chosen at this stage. A convolutional neural network [6] is chosen for this purpose. The main reason behind selecting the Convolutional Neural Network is its ability to deeply analyze every part of the image, taking into consideration the intricate hierarchical relationships. This sequential model goes through several layers to comprehend the complexity of the data.

Firstly, a Convolutional 2D layer is used, which includes filters, kernel size, activation function, and input shape. Secondly, the pooling layer is employed to reduce the spatial dimensions. These layers may be used multiple times based on the specific requirements.

After the completion of these layers, flattening is performed. Flattening is used to convert the multidimensional data into a one-dimensional array. Finally, the dense layer, also known as the fully connected layer, is used, which includes the number of units and activation function.

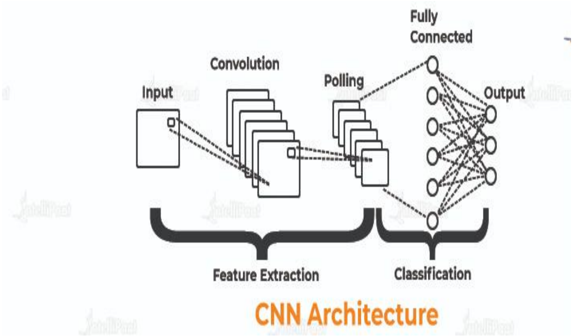

Fig. 4. CNN Architecture

Hyper parameter tuning [15],including batch size, activation function, optimizes, number of epochs, number of dense layers, and number of neurons, is carried out later in the training phase. These parameters are used to evaluate the model's performance. The training data is then used to evaluate the model once it has been trained.

$$R(x) = \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

$$s\left(x_i\right) = \frac{e^{x_i}}{\sum_{j=1}^{n} e^{x_j}} \quad (2)$$

E. Accuracy

Accuracy is a fundamental metric used to evaluate the performance of the model. Without accuracy, it is difficult to assess the effectiveness of the model's training. It measures the proportion of correctly classified instances out of the total instances evaluated, offering a straightforward way to gauge performance.

MNIST Dateset	Model	Accuracy
Existing System [16]	Simple CNN	93%
	Deep CNN	91%
Proposed System	CNN	98%

Table 1: Accuracy of the Existing and Proposed System

The table describe that the existing system is built on using simple cnn and deep cnn and their accuracies are 93% and 91% respectively. The proposed system is built on CNN where the accuracy is 98%.

In this paper, accuracy is determined using the evaluation method, and it can also be assessed through the confusion matrix [17]. These methods provide valuable insights into the model's predictive capabilities and overall performance.

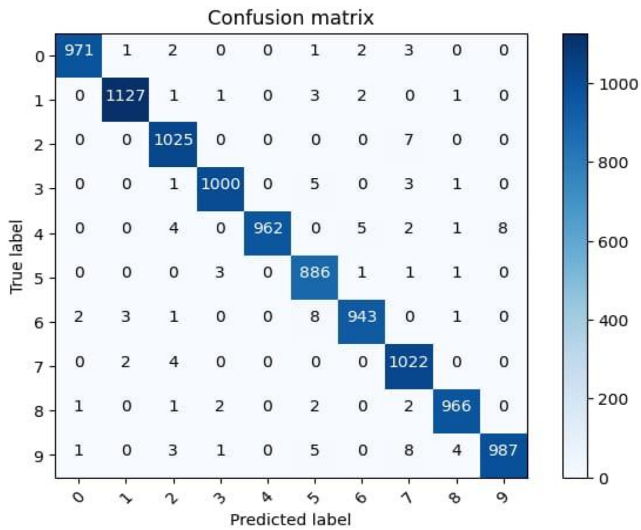

Fig. 5. Confusion matrix

IV. CONCLUSION AND FUTURE SCOPE

In summary, this study investigated the use of convolutional neural networks for number and digit prediction. The outcomes show that the predictions were correct. The accuracy on the mnist dataset is 98.9%.

Those who need to know about the unclear digits in photos might utilize this model. This guarantees that picture recognition is one of the highlights of deep learning. To guarantee even greater accuracy, this work can be extended in the future to include work on hybrid algorithms.

V. REFERENCES

[1] Rene Y. Choi, Aaron S. Coyner, Jayashree Kalpathy-Cramer, Michael F. Chiang, and J. Peter Campbell "Introduction to Machine Learning, Neural Networks, and Deep Learning, " Translational Vision Science & Technology, 2020 Feb.

[2] Amitha Mathew, P. Amudha, and S. Sivakumari, "Deep Learning Techniques: An Overview, " In Advanced Machine Learning Technologies and Applications, pp. 599-608, January 2021.

[3] Nafea, Ahmed Adil, Saeed Amer Alameri, Russel R Majeed, Meaad Ali Khalaf, and Mohammed M AL-Ani, "A Short Review on Supervised Machine Learning and Deep Learning Techniques in Computer Vision, " Babylonian Journal of Machine Learning 2024, pp. 48-55, 2024.

[4] Wani, M Arif, Farooq Ahmad Bhat, Saduf Afzal, Asif Iqbal Khan, M Arif Wani, Farooq Ahmad Bhat, Saduf Afzal, and Asif Iqbal Khan, "Unsupervised deep learning architectures, " Advances in Deep Learning, pp. 77-94, 2020.

[5] Yutaka Matsuo, Yann LeCun, Maneesh Sahani, Doina Precup, David Silver, Masashi Sugiyama, Eiji Uchibe, and Jun Morimoto, "Deep learning, reinforcement learning, and world models," Neural Networks,2022, 152: 267-275.

[6] Laith Alzubaidi, Jinglan Zhang, Amjad J Humaidi, Ayad Al-Dujaili, Ye Duan, Omran Al-Shamma, J. Santamaría, Mohammed A Fadhel, Muthana Al-Amidie, and Laith Farhan. "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," J Big Data. 2021, 8(1): 53. 2021 Mar 31. doi: 10.1186/s40537-021-004448

[7] https://www.tensorflow.org/datasets/catalog/mnist

[8] Abdullah, Dakhaz Mustafa, Abdulazeez, and Adnan Mohsin, "Machine learning applications based on SVM classification: a review," Qubahan Academic Journal 1 (2), 81-90 (2021).

[9] Gawlikowski, Jakob, Cedrique Rovile Njieutcheu Tassi, Mohsin Ali, Jongseok Lee, Matthias Humt, Jianxiang Feng, Anna Kruspe, Rudolph Triebel, Peter Jung, Ribana Roscher, Muhammad Shahzad, Wen Yang, Richard Bamler, and Xiao Xiang Zhu, "A survey of uncertainty in deep neural networks," Artificial Intelligence Review, vol. 56, no. Suppl 1, pp. 1513-1589, 2023.

[10] Yehya Abouelnaga, Ola S. Ali, Hager Rady, and Mohamed Moustafa, " CIFAR-10: KNN-based ensemble of classifiers", 66 IEEE, March 2017.

[11] Samay, Pashine, Ritik Dixit, and Rishika Kushwaha, "Handwritten digit recognition using machine and deep learning algorithms," arXiv preprint arXiv:2106.12614, 2021

[12] Ritik Dixit, Samay, Pashine, and Rishika Kushwaha, "Handwritten digit recognition using machine and deep learning algorithms," arXiv preprint arXiv:2106.12614, 2021.

[13] https://keras.io/

[14] Hancock, John T., and Taghi M. Khoshgoftaar. "Survey on categorical data for neural networks," Journal of Big Data 7 (1), 1-41, 2020.

[15] Antunes, André, Bruno Ferreira, Nuno Marques, and Nelson Carriço. "Hyperparameter optimization of a convolutional neural network model for pipe burst location in water distribution networks," Journal of Imaging 9 (3), 68, 2023.

[16] SakhtimohanM,ElizabethRani.G,NavaneethaKrishnan. M.Abhinaya.M,Karthigadevi.K and P.V.Siddhartha, "Digit Recognition of MNIST Handwritten Using Convolutional Neural Networks (CNN)," 2023 International Conference on Intelligent Systems for Communication, IoT and Security (ICISCoIS), Coimbatore, India, 2023, pp. 328-332, doi: 10.1109/ICISCoIS56541.2023.10100602.

[17] Liang, and Jingsai, "Confusion matrix: Machine learning, " POGIL Activity Clearinghouse 3 (4), 2022.