EasyChair Preprint

№ 9618

Handwritten Digit Recognition Using CNN

V Viswanatha, A C Ramachandra, Satya Dev Nalluri,
Sai Manoj Thota and Aishwarya Thota

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

January 24, 2023

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 11, Issue 1, January 2023 ||

| DOI: 10.15680/IJIRCCE.2023.11010012 |

Handwritten Digit Recognition Using CNN

Viswanatha V, Ramachandra A C, Satya Dev Nalluri, Sai Manoj Thota, Aishwarya Thota

Asst. Professor, Department of Electronics and Communication Engineering, Nitte Meenakshi Institute of Technology, Bangalore, India

Professor, Department of Electronics and Communication Engineering, Nitte Meenakshi Institute of Technology, Bangalore, India

UG Student, Department of Electronics and Communication Engineering, Nitte Meenakshi Institute of Technology, Bangalore, India

UG Student, Department of Electronics and Communication Engineering, Nitte Meenakshi Institute of Technology, Bangalore, India

UG Student, Department of Electronics and Communication Engineering, Nitte Meenakshi Institute of Technology, Bangalore, India

ABSTRACT: In this paper, we are going to see how we can train a neural network model to recognize a handwritten digit which is given as an input to the model. The algorithm used to realize it is Convolution Neural Network (CNN).It is a network architecture for deep learning. It learns from the data which isthrough the images. It finds patterns in images and recognizes objects and categories. The CNN has several layers which takes the input, analyzes the input, and producesoutput. It's very much used in Deep Learning and very efficient in the modern world filled with AI. It's a part of ANN that has been the superior algorithm incomputer vision tasks. It has achieved top-level performances in various fields like medical research, AI, etc. CNN, which is used for processing data, is a type ofdeep learning model that has grid patternsi.e.-images. It is a construct that has three kinds of layers namely convolution, pooling, and fully connected layersrespectively.The first few layers do the feature extraction, whereas the next layer maps the extracted features into final output. The convolution layer plays an importantrole in CNN. It is composed of a stack of mathematical operations such as convolution. The pixel values are stored in a two-dimensional array in the digital images and small grid of parameters called kernel is applied at each image position. This makes the CNN highly efficient for image processing. The layers perform convolutionand subsampling one after another. Output of one layer is input of the next layer. The output of the final layer is our predicted value. Extracted features can progressivelybecome more complex, as one layer feeds its output to the next layer.

KEYWORDS: Character recognition, ReLU function, Feature extraction, Handwriting Recognition, Deep learning

I. INTRODUCTION

Artificial intelligence (AI) in simple words is basically making a computer do the work that traditionally requires the human brain. AI has the ability to take in large amounts of data unlike the human and uses that data to recognize patterns, make decisions, and give judgment. In this AI we have a subset which is called Machine learning. ML is used to make computers to learn to behave ashumans. This is done by two ways, supervised learning in which the computer is given a set of input data and the required outputfor it. Now it uses ML to learn the algorithm to understand how that particular input gives this particular output. Now the unsupervised learning is when input data is provided but with no output, so the ML has to learn to analyze and clutter the datasets into categories.

ML (Machine Learning) is the branch of computer science which helps the machines to learn without being programmed. A program learns from data given. It performs tasks and measures accuracy and tells if it's performance at doing tasks improved with the experience or not. We use algorithms and other techniques in Machine Learning instead of doing the programming part. Machines learn from past experiences and examples. A model can be built based on their past experiences, so that it can be used to predict the new values. If the question or the problem is too large and difficult to solve, then it can be used. It can also help finding the answers to questions based on the analysis of data. It requires very less time to find important things when a large amount of data is given. It can also solve very complex problems as machines can learn faster than humans and they may even exceed humans in some fields. As a result, its demand is rising continuously. Machine Learning is catching up with cloud computing and big data as it can solve many difficult problems with ease. It's used in many applications.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 11, Issue 1, January 2023 ||

| DOI: 10.15680/IJIRCCE.2023.11010012 |

1. In Medicare, it helps in discovering new medicines and allows doctors to make many good diagnoses which leads to prediction of new and unknown diseases in advance.
2. In the internet community, it targets users and gives them better suggestions according to their gender, location, age et cetera, hence analyzing their purchasing behavior.
3. It can be used to detect scams and online frauds.
4. It's very useful in speech and face recognition, natural language processing, automotive, automated trading, etc.

In the past few years, information and data has grown and is growing in abundance. So we must find something that can lead to accurate decisions. Machine learning has helped us a lot in this field. It's a field of ANN which makes machines learn from their experiences and examples just like humans do. Model is built from the data which is given to the algorithm. It can predict new values based on this model. It helps us to find and explore new and unknown things. Machine learning can be used in various fields like finance, health, media, travel, image processing and computer vision, automated trading, aerospace, natural language processing, manufacturing, automotive and many more. This report provides us with a clear view on the basics of machine learning, the algorithms used and its applications in various fields.
The algorithm used in this paper is CNN. It's a network architecture for deep learning. It learns from the data which is through the input images. It finds patterns in images and recognizes objects and categories, this isfeature extraction. Now max pooling and convolution takes place. The CNN has several layers which take the input, analyzes it andproduces output as shown in Fig.

1. It's one of the most used algorithms in deep learning. It's very much used in Deep Learning and very efficient in the modern world filled with AI. CNN doesn't require hand crafted feature extraction.CNN architectures don't need segmentation of organs or tumors by the experts.

CNN has the scope to take a lot more data because it has highly complex learnable parameters. Also, it's very expensive. It has three types of layers which perform different operations.
a. **Convolution Layer**- It creates a feature map using all the already used examples and predicts the probabilities for each feature byscanning the image. It flattens all the pixels and analyzes a few at a time.
b. **Pooling layer-** It basically scales down the information which is generated as the output of the convolution layer. Also known as down sampling. Both the above layers repeat several times for each feature extracted from the image.
c. **Fully Connected Layers**- It consists of 3 layers namely fully connected input layers, hidden layers and output layers.
Fully connected input layers flatten the pixels of the image for easier analysis. Hidden layers apply weights to the inputs generated and applies the activation function. Output layer generates the final probabilities, predicts and determines the class and category of the image.

1). In this implementation, we used the Convolution Neural Network algorithm to recognize the digits written with our own style and font.

2). The types of layers used were an input layer, one hidden layer and one output layer in the CNN algorithm.

3). We used the Keras library and the Tensorflow library to load the datasets and train the model.

4). The activation function used for the hidden layer is ReLU. Since this activation function removes all the negative values and lets the positive values pass, this activation function is very simple, super-efficient and easy to understand.

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 11, Issue 1, January 2023 ||

| DOI: 10.15680/IJIRCCE.2023.11010012 |

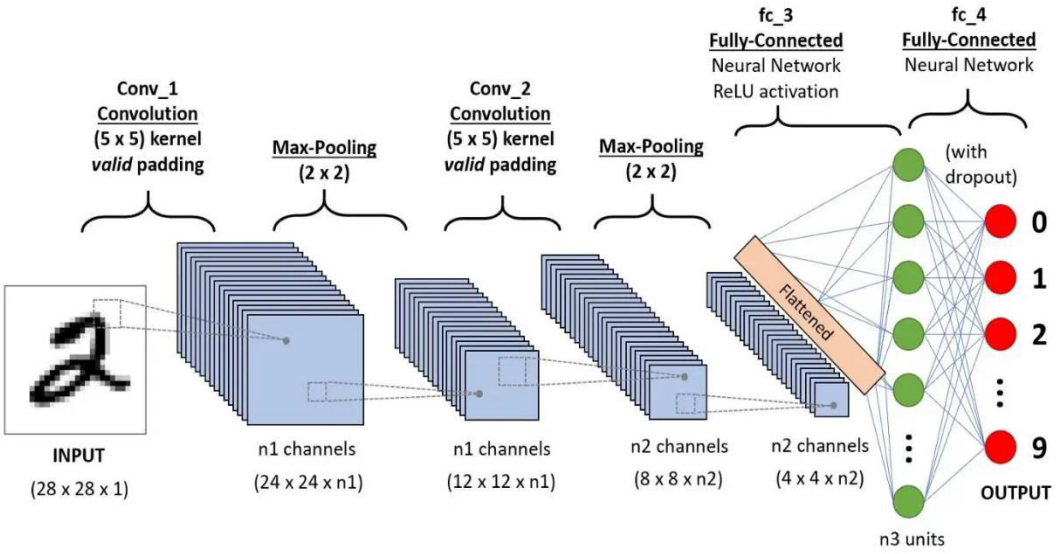


Fig.1 Convolution Neural Network

II. METHODOLOGY

1. Importing the libraries: Libraries are useful tools that can make a web developer's job more efficient. It's a set of prewritten code, that we can call while programming our own code. Basically, it's the work that's already done by someone else that youcanmake use of, without having to do it yourself. You can also use it in your own code. Different libraries have differentrestrictionson fair use, but this is a code that was designed to be used by others, instead of just standing alone.

The libraries used in this code are -

a. *Tensorflow* - Tensorflow'sframework is open source, it is used in machine learning and other computations on various data. It has the symbol of TF. It allows the developers to create the learning algorithms on their set of data and models, and also to experiment with diverse algorithms. It allows the developers to create data flow structures, which shows how the data will move through a series of graphs or nodes.

b. *OpenCV* – OpenCV (Open-Source Computer Vision Library) is an open-source computer vision and machine learning software libraries. It's a library in python which is designed to solve computer vision problems. OpenCV (also known as CV2) was built to provide a common infrastructure for computer vision applications and to speed up the use of machine perception in all kinds of products.

c. *NumPy* – NumPy stands for Numerical Python. We use the NumPy library to work with arrays. It can also be used to work in the domain of linear algebra, matrices and Fourier transform. NumPy was created by Travis Oliphant in 2005. It's an open source project and can be used freely.

2. Loading datasets: For the training and testing data, we will be using a Dataset which is present in the Tensorflow API.

This specific dataset is the Mnist data that contains around fifty thousand images for the training data and another ten thousand images for testing data confined to a dimension of 28X28.

3. Creating a model: We are creating a model using a sequential neural network to recognize handwritten digits. We are using Three dense layers namely input hidden and output layers. Here we used 128, 128, 10 parameters for the previously mentioned layers respectively.

4.Training the Model: We are training the model using 3 dense layers, one using input layer, one using hidden layer and the last one using the output layer. The 3 dense layers take 128, 128, 10 parameters each. We flatten the pixels of the greyscale image in the input layer. Then we apply the activation functions to the weights in the hidden layer. The output layer gives the result as a

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 11, Issue 1, January 2023 ||

| DOI: 10.15680/IJIRCCE.2023.11010012 |

prediction.

5.Testing the Model: Once the model is trained, we can use the loss function and accuracy function to test the model. We should have accuracy as high as possible and loss as less as possible to get the desired output (with accuracy being close to 1 and loss being close to 0). We used "ADAM" as our optimizer, "cross-entropy" as our loss and "accuracy" as our metrics.

6.Getting the output: We take a set of inputs and upload them into our model. Then we use a while loop to analyze each input individually and give the predictions for each image. This process has been shown in a flowchart below in Fig. 2

FLOWCHART

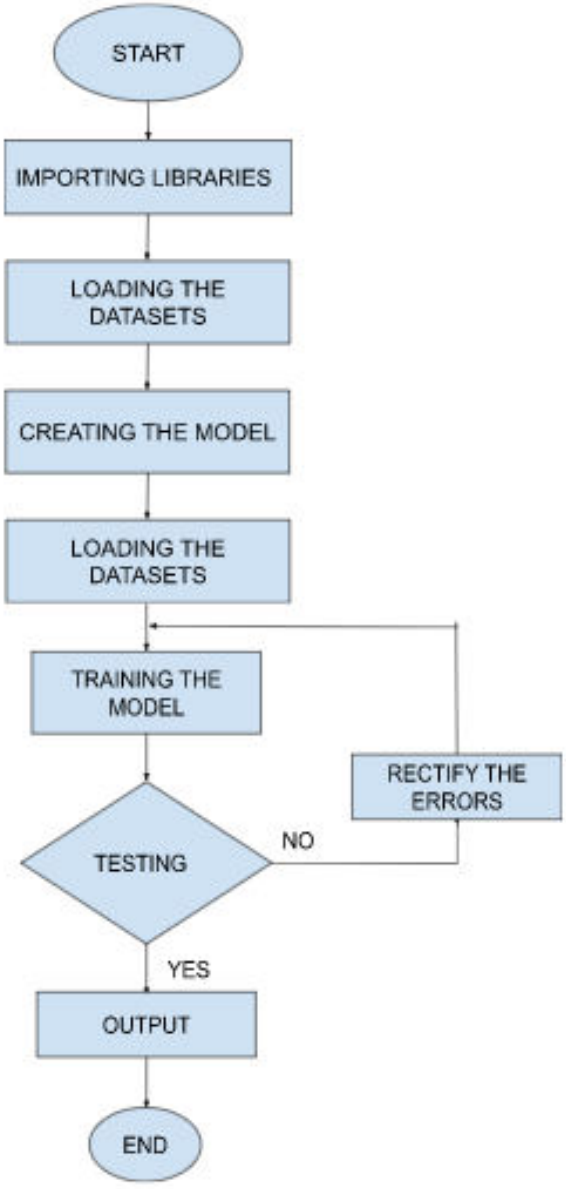


Fig.2 Methodology used in Implementation of Handwritten Recognition

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 11, Issue 1, January 2023 ||

| DOI: 10.15680/IJIRCCE.2023.11010012 |

III. TOOLS AND ALGORITHMS USED

1. The neural network used isSequential Model - Sequence models are basically the models in machine learning that analyze input sequences and produce output sequences of data.

Sequential data includes audio clips, text streams, time-series data, video clips, etc. Recurrent Neural Networks (RNNs) is a majorlyused algorithm used in sequence models. It's basically a linear path which connects all the input, hidden and the output layers withoutskipping any layers in between, as shown in Fig. 3. A Sequential model is ideal for a plain stack of layers where each layer has exactly one input tensor, one or more hidden tensors and one output tensor.

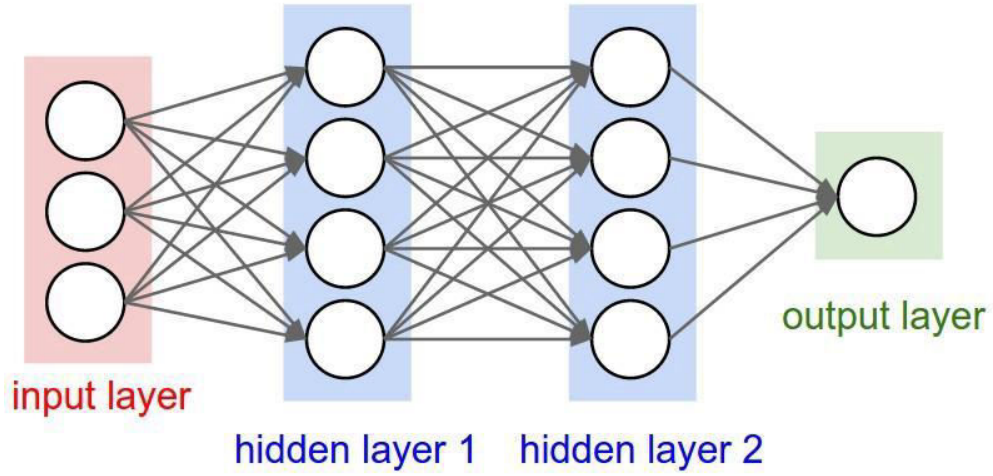


Fig.3 Structure of a Sequential Network

2. Activation functions used are:

a. ReLU - It stands for Rectified Linear Activation Function. It's a piecewise linear function in which will produce output from the input directly when positive, or the output will be zero, as shown in Fig. 4. For many different types of neural networks, it is used as the default activation function networks since it's more convenient to train a model which uses this activation function to obtain better performance.

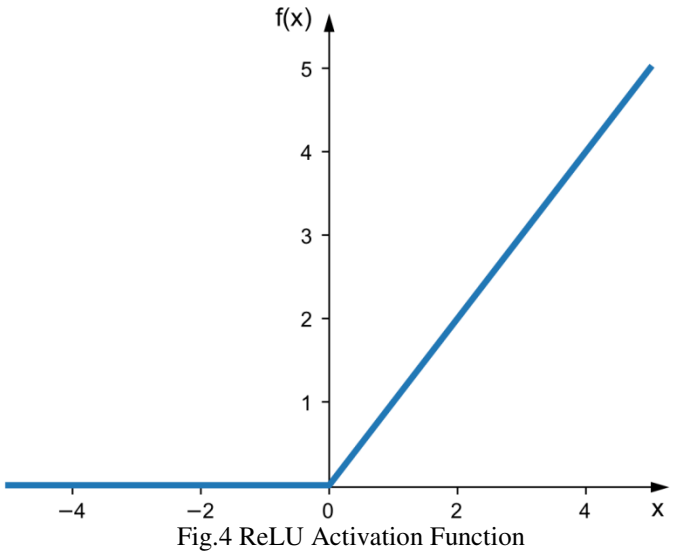


Fig.4 ReLU Activation Function

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 11, Issue 1, January 2023 ||

| DOI: 10.15680/IJIRCCE.2023.11010012 |

b.Softmax - It's used as a function which can convert a vector of numbers into a vector of probabilities (whose value lies in between 0 and 1 with 1 being the most probable outcome). The probabilities of each value are proportional to the relative scale of each value in the vector.

3. Optimizer used in this code isADAM –stands for Adaptive Moment Estimation. It is an algorithm and a technique to obtain gradient descent. This method is very effective in terms of working with large problems which also involves many parameters and data. Less memory is required and also, it's a combination of the "RMSP" algorithm and the "gradient descent with momentum" algorithm. This is used to accelerate the above algorithm by taking the weighted average of the gradients into notice. For the algorithm, making use of the averages makes the minima lean towards a faster pace.
4. The loss function used isCROSS-ENTROPY– When trying to optimize particular or any classified models, Cross entropy is best suited for loss functions. Classification is a method of supervised learning which involves predictions of class labels using one or many input variables.

5. The metrics function used is ACCURACY– It is a function which determines the performance of the model. Metric functions are quite similar to loss functions but except for the fact that the outputs from evaluating a metric will not be used when training the model.

IV. IMPLEMENTATION

In this paper, Neural network is implemented wherein the model recognizes and predicts a handwritten digit. Initially Tensorflow and Keras are used to form the bones of the implementation. We load the datasets from both of these open-source libraries and make our model to analyze thousands of images. The model learns all the patterns, pixel placements of the greyscale images and all the neural connections.

Keras is an API (Application Programming Interface), which is designed for machine learning and deep learning. It's an open source library which has a lot of inbuilt data. It's the interface of Tensorflow library.
1. It follows the best practices available for reducing cognitive load.
2. It provides simple & consistent APIs (Application Programming Interface).
3. It has extensive documentation and has developer guides.
4. It minimizes the amount of user actions needed for common use cases, and it provides actionable & clear error messages.

TensorFlowframework is open source, it is used in machine learning and other computations on various data. It has the symbol of TF. It allows the developers to create the learning algorithms on their set of data and models, and to experiment with diverse algorithms. It allows the developers to create data flow structures, which shows how the data will movethrough a series of graphs or nodes. All the datasets which we used in this program are taken from Tensorflow. TensorFlow has a wide range of libraries containing pre-trained models that can be used in many research articles. It's called Tensorflow because it takes the input as a multi-dimensional array, in other words known as tensors. You can build a type of flowchart of operations that you want to perform on that input. The input goes in at one end from the starting, then it flows through the system and comes out from the other end as output at the end. Hence the name TensorFlow.Then we use the ADAM optimizer for training our model. It's the best optimizer to train our neural network in less time with highefficiency. An optimizer is used to update the network weights live during the training. Then we use the loss function and the metrics function to check the performance of our model. We use cross-entropy and accuracy respectively to check the performance. Once the model is trained a, we save it to our computer and comment on all the preprocessing and training code. So, whenever the code is required, we can just load the saved model since all the data is stored in it. Now the final step is to give the inputs to our model. For this we scan the digits we wrote on a sheet of paper or draw the digit in MS paint tool and download the image in PNG format. Once we upload all the images into our python script, then we use while loop to all each image and analyze it to predict the output.

V. RESULTS AND DISCUSSION

From this implementation, we are able to identify the handwritten digits asinput given to the code as shown in Fig.5, analyze it and predict the probability output as shown in Fig.6.With the code we are able to show that written data in MS paint application can be saved. Where the saved file is now loaded into the program and will run. With a while loop, we check each image and predict the value if multiple data set files are present. With this the image is analyzed by the already learnt neural connections and the grayscale pixels. It will now provide us with a prediction of the accurate output of recognized data. So, we can

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 11, Issue 1, January 2023 ||

| DOI: 10.15680/IJIRCCE.2023.11010012 |

successfully recognize and digitalize our data. We are successfully able to understand and use Sequential Model, ReLU, SoftMax, adam, Cross-Entropy, Accuracy functions for image recognition. The snips of the code used in the program are shown in the Fig.7(a) and Fig.7(b).

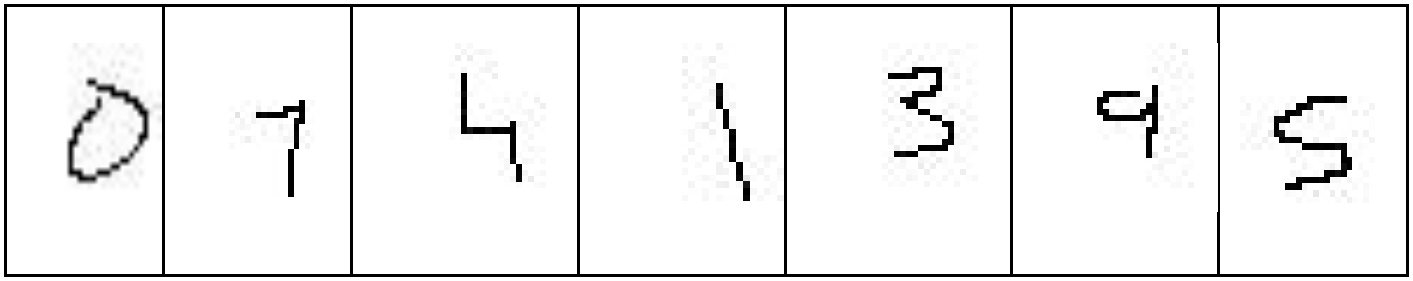


Fig.5 Handwritten digits given as input Images

```
1/1 [==============================] - 0s 363ms/step
This digit might be 0
1/1 [==============================] - 0s 24ms/step
This digit might be 7
1/1 [==============================] - 0s 22ms/step
This digit might be 4
1/1 [==============================] - 0s 21ms/step
This digit might be 1
1/1 [==============================] - 0s 35ms/step
This digit might be 3
1/1 [==============================] - 0s 26ms/step
This digit might be 9
1/1 [==============================] - 0s 47ms/step
This digit might be 5
```

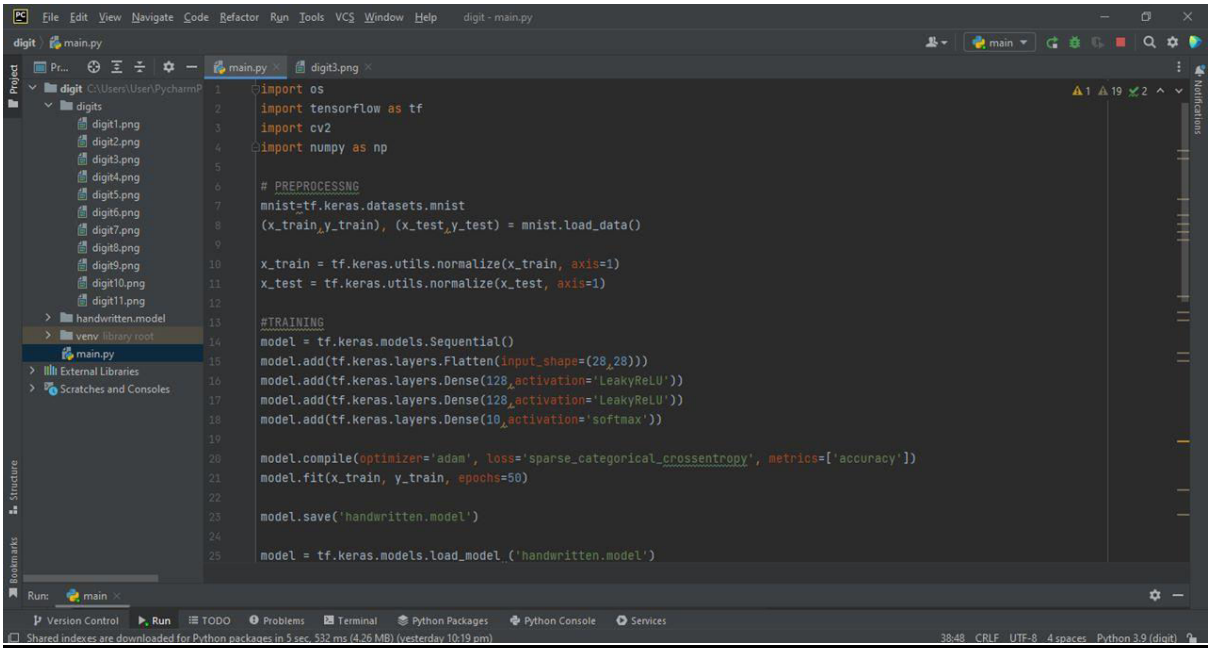
Fig.6 Outputs Obtained

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

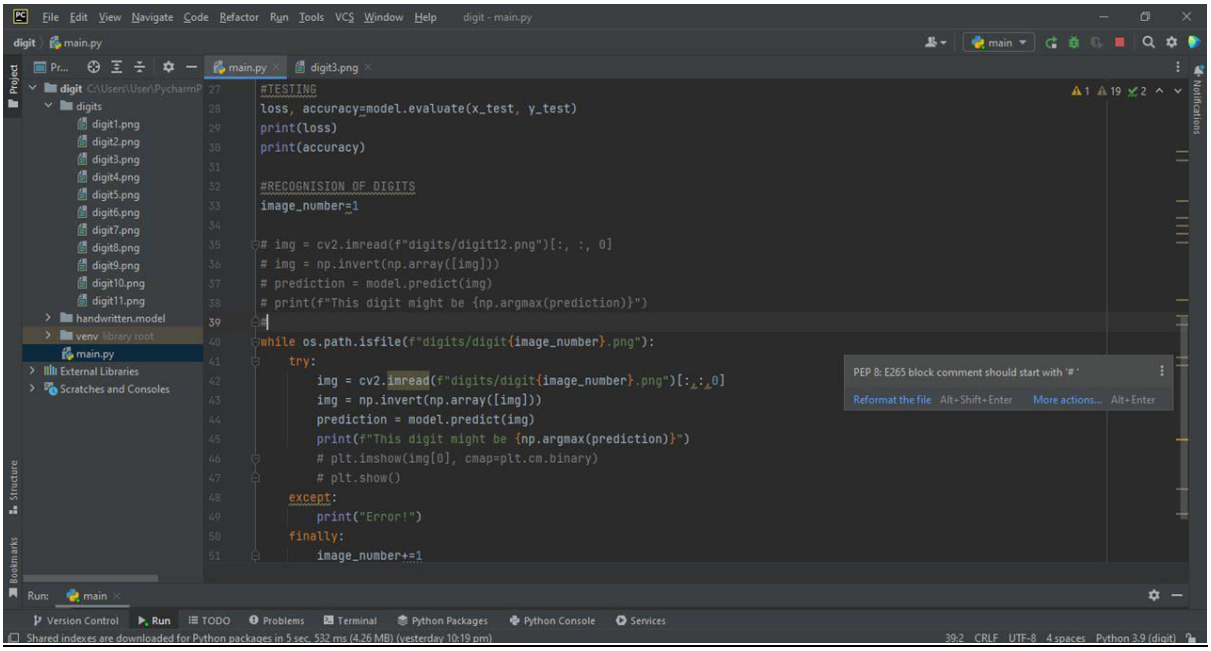
|| Volume 11, Issue 1, January 2023 ||

| DOI: 10.15680/IJIRCCE.2023.11010012 |



```
import os
import tensorflow as tf
import cv2
import numpy as np

# PREPROCESSNG
mnist=tf.keras.datasets.mnist
(x_train,y_train), (x_test,y_test) = mnist.load_data()

x_train = tf.keras.utils.normalize(x_train, axis=1)
x_test = tf.keras.utils.normalize(x_test, axis=1)

#TRAINING
model = tf.keras.models.Sequential()
model.add(tf.keras.layers.Flatten(input_shape=(28,28)))
model.add(tf.keras.layers.Dense(128,activation='LeakyReLU'))
model.add(tf.keras.layers.Dense(128,activation='LeakyReLU'))
model.add(tf.keras.layers.Dense(10,activation='softmax'))

model.compile(optimizer='adam', loss='sparse_categorical_crossentropy', metrics=['accuracy'])
model.fit(x_train, y_train, epochs=50)

model.save('handwritten.model')

model = tf.keras.models.load_model ('handwritten.model')
```

Fig.7 (a) Simulation framework of code



```
#TESTING
loss, accuracy=model.evaluate(x_test, y_test)
print(loss)
print(accuracy)

#RECOGNISION OF DIGITS
image_number=1

# img = cv2.imread(f"digits/digit12.png")[:, :, 0]
# img = np.invert(np.array([img]))
# prediction = model.predict(img)
# print(f"This digit might be {np.argmax(prediction)}")
#
while os.path.isfile(f"digits/digit{image_number}.png"):
    try:
        img = cv2.imread(f"digits/digit{image_number}.png")[:,:,0]
        img = np.invert(np.array([img]))
        prediction = model.predict(img)
        print(f"This digit might be {np.argmax(prediction)}")
        # plt.imshow(img[0], cmap=plt.cm.binary)
        # plt.show()
    except:
        print("Error!")
    finally:
        image_number+=1
```

Fig.7(b) Simulation framework of code

VI. PERFORMANCE ANALYSIS

We use the loss function and the metrics to get the loss and accuracy respectively. The loss function used for the model isCross-Entropy. It tells us the amount of loss was obtained by training our model. The metrics used for our model is Accuracy. It tells us how accurate our model is after training it. The Accuracy should be close to 1 and the Loss should be close to 0. Then and only

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 11, Issue 1, January 2023 ||

| DOI: 10.15680/IJIRCCE.2023.11010012 |

then, ourmodel is close to perfection. If the loss is high compared to 0 and accuracy compared to 1, then the model is not perfect, and we need to train the model with different parameters and a greater number of epochs. The loss we got after training our model is around 3%The accuracy we got after the model is trained is around 97%, as shown in Fig. 8.

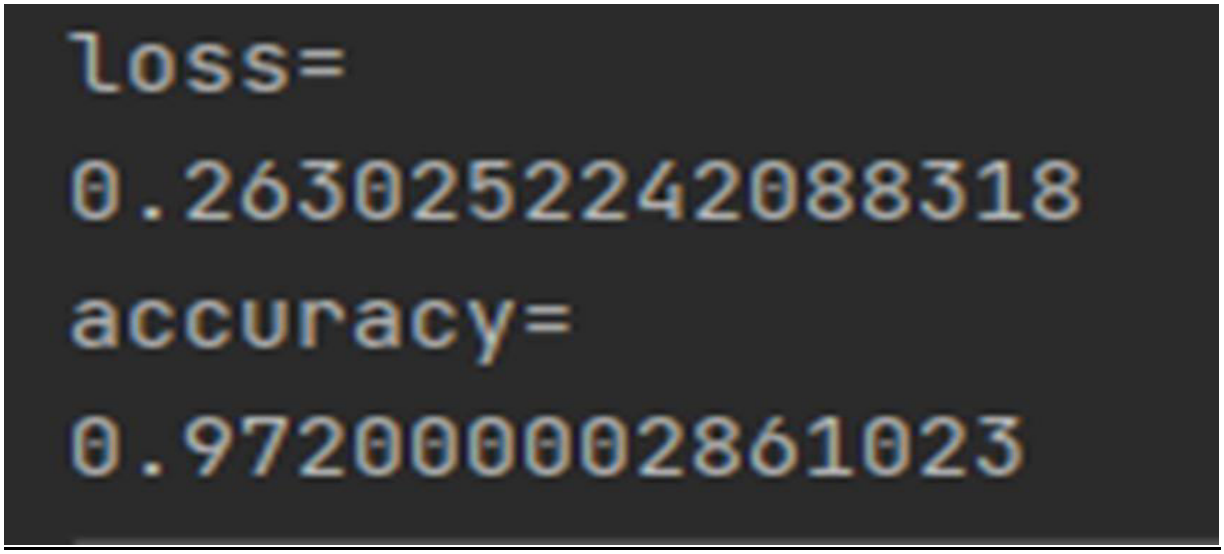


Fig.8 Performance Analysis- Loss and Accuracy of testing model

VII. CONCLUSION

In this implementation, we created a neural network to recognize a handwritten digit using TensorFlow and Keras. using the MNISTdataset, our model canrecognize handwritten digits which are being input to the model. The performance of CNN for thehandwritten digit is accurate. The method works well, and the loss percentage is less with all those training sessions. The only difficulty here is the noise in the image, but with the training it has, it tries to achieve the best possible output. The model's performance and accuracy were tested after training the model for 50 epochs. With the results given from this work we are more confident in finding other ways to make this better and to make it easier for complex data like converting handwritten paragraphs into text. Through this research work we understood all the mechanisms used to identify handwritten data. We understand the importance of hand recognition as it is easy for the user to write data on paper and use handwritten data recognition to convert it into text instead of the typing it on keyboard. Further it is recommended to implement on edge computing platforms like Raspberry Pi 4 system for actual usage.

REFERENCES

1. Wei Lu, Zhijian Li and Bingxue Shi, "Handwritten Digits Recognition with Neural Networks and Fuzzy Logic", IEEE International Conference on Neural Networks, 1995.
2. D. Castro, B. L. D. Bezerra and M. Valena, "Boosting the deep multidimensional long-short-term memory network for handwritten recognition systems", ICFHR, 2018.
3. Viswanatha, V., R. K. Chandana, and A. C. Ramachandra. "Real Time Object Detection System with YOLO and CNN Models: A Review." (2022).
4. P. Voigtlaender, P. Doetsch and H. Ney, "Handwriting recognition with large multidimensional long short-term memory recurrent neural networks", ICFHR, 2016.
5. Gideon Maillette de Buy Wenniger, Lambert Schomaker and Andy Way, "No Padding Please: Efficient Neural Handwriting Recognition", ICDAR, 2019.
6. Rohan Vaidya, Darshan Trivedi, Sagar Satra and Prof. Mrunalini Pimpale, "Handwritten Character Recognition Using Deep-Learning", ICICCT, 2018.
7. Viswanatha, V., R. K. Chandana, and A. C. Ramachandra. "IoT Based Smart Mirror Using Raspberry Pi 4 and YOLO Algorithm: A Novel Framework for Interactive Display." Indian Journal of Science and Technology 15.39 (2022): 2011-2020.
8. Herleen Kour and Naveen Kumar Gondhi, "Machine Learning approaches for Nastaliq style Urdu handwritten recognition: A

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 11, Issue 1, January 2023 ||

| DOI: 10.15680/IJIRCCE.2023.11010012 |

survey", ICACCS, 2020.

9. Shahbaz Hassan, Ayesha Irfan, Ali Mirza and Imran Siddiqi, "Cursive Handwritten Text Recognition using Bi-Directional LSTMs: A case study on Urdu Handwriting", Deep-ML, 2019.
10. Viswanatha, V., Ashwini Kumari, and B. M. Sathisha. "Implementation of IoT in Agriculture: A Scientific Approach for Smart Irrigation." 2022 IEEE 2nd Mysore Sub Section International Conference (MysuruCon). IEEE, 2022.
11. Siham Tabik, Daniel Peralta, Andrs Herrera-Poyatos and Francisco Herrera, "A snapshot of image Pre-Processing for convolutional neural networks: Case study of MNIST", International Journal of Computational Intelligence Systems 10, vol. 555, no. 1, January 2017.
12. J. Wang and X. Hu, "Gated recurrent convolution neural network for ocr", NIPS, 2017.
13. Viswanatha, V., B. M. Sathisha, and Ashwini Kumari. "Custom Hardware and Software Integration: Bluetooth Based Wireless Thermal Printer for Restaurant and Hospital Management." 2022 IEEE 2nd Mysore Sub Section International Conference (MysuruCon). IEEE, 2022.
14. M. M. Al-Taee, S. B. H. Neji and M. Frikha, "Handwritten Recognition: A survey," 2020 IEEE 4th International Conference on Image Processing, Applications and Systems (IPAS), Genova, Italy, 2020, pp. 199-205, doi: 10.1109/IPAS50080.2020.9334936.
15. R. R. Ingle, Y. Fujii, T. Deselaers, J. Baccash and A. C. Popat, "A Scalable Handwritten Text Recognition System," 2019 International Conference on Document Analysis and Recognition (ICDAR), Sydney, NSW, Australia, 2019, pp. 17-24, doi: 10.1109/ICDAR.2019.00013.
16. Viswanatha, V., Ashwini Kumari, and Pradeep Kumar. "Internet of things (IoT) based multilevel drunken driving detection and prevention system using Raspberry Pi 3." International Journal of Internet of Things and Web Services 6 (2021).