HANDWRITTEN EQUATION SOLVER USING NEURAL NETWORK

Adhithya T , Akshay T A , Karun Krishnan , Omkar R
Department of Information Technology
Government Engineering College, Barton Hill, Thiruvananthapuram, Kerala.

***Abstract*: Robust handwritten character recognition is a tricky job in the area of image processing. Among all the problems, handwritten, mathematical expression recognition is one of the complicated issues in the area of computer vision research. In this project a group of handwritten arithmetic equations are recognized and a solution is made for those equations. For classification of a specific characters we apply Convolutional Neural Network. Each correct character detection and string operation is used for the solution of the equation. Finally, the experimental results show the great effectiveness of the proposed system.**

The main objective of this project is to develop a web app which calculates a Handwritten Equation and solves it by training by handwritten digits and mathematical symbols using Convolutional Neural Network with image processing techniques to achieve an accuracy of 98%.

Keywords: equation solving, optimizing, visual development

I. INTRODUCTION

Mathematics is broadly used in almost all areas of science, such as physics, engineering, medicine, economics, etc. Digital Document analysis and understanding is the major research concern today. Handwritten mathematical expression recognition is still a challenging job to do in the area of computer vision. The primary task for the recognition of mathematical expression is to segment the character and then classify those characters. An operation that seeks to decompose an image of a sequence of characters into sub images of individual symbols is called Character segmentation.

Convolutional neural network (CNN) is one of the most used classification models in the computer vision area. In the last few years, Convolutional Neural Network learning has provided outstanding performance in the field of image classification, machine learning and pattern recognition. Above all existing models, CNN is one of the most popular models and has been providing state-of-the-art recognition accuracy for object recognition, segmentation, human activity analysis, image super resolution, scene understanding, tracking, and image captioning.

For the task of image classification CNN outperforms above all the previous classification methods. CNN extracts features from the image by a series of operations. In this work we mainly focus on the recognition of linear equations and after successful detection of linear equations we apply character string operation for the solution of those equations.

In this project we consider a single equation in the form **ax + b = 0,** with any combination of its parts. Our job is to recognize the handwritten equation from the image and for each successful detection finding the solution of that equation.

Feature extraction of each individual character is the most difficult part for any handwritten equation due to its different shape and structure.

To solve this problem, we apply a convolutional neural network which doesn't require any predefined feature for classification of specific character. We use the smallest number of hidden layers in order to reduce the training time with tolerable error rate.

This project is aimed at developing software which will be helpful in recognizing handwritten characters for equation solving. It engulfs the concept of neural network. One of the primary means by which computers are endowed with humanlike abilities is through the use of a neural network. Neural networks are particularly useful for solving problems that cannot be expressed as a series of steps, such as recognizing patterns, classifying them into groups, series prediction and data mining. Pattern recognition is perhaps the most common use of neural networks. The neural network is presented with a target vector and also a vector which contains the pattern information, this could be an image or hand written data. The neural network then attempts to determine if the input data matches a pattern that the neural network has memorized. A neural network trained for classification is designed to take input samples and classify them into groups.

II. PROPOSED ENVIRONMENT AND MODELS

In this section, we introduce the overall architecture and the detailed principle of our proposed models.

In the proposed method, we are developing a web app called 'EQ-SOL' where the user can give a handwritten mathematical equation as input (image) and get the corresponding answer as output.

At first, the noise from the original input image is removed by applying binarization to it. After that we use compact horizontal projection for segmenting each line of the equation from the input image. Then we consider each part of the segmented image as a full image for further process.

For each line of the equation image we then find specific characters in the form of a connected component. Each segmented character is then provided as input to the convolutional neural network model for classification of the character. The resulting character that is the output of CNN is then used for making a character string which is similar to the original equation.

A. ARCHITECTURE OVERVIEW

In this project we use Convolutional Neural Networks (CNNs) for character detection from handwritten images to solve the equation.

Convolutional Neural Networks, a branch of deep learning, has an impressive record for applications in image analysis and

interpretation. A convolutional neural network convolves an input image with a defined weight matrix to extract specific image features without losing spatial arrangement information.

The main advantage of CNN compared to its predecessors is that it automatically detects the important features without any human supervision. CNN is also computationally efficient. It uses special convolution and pooling operations and performs parameter sharing. This enables CNN models to run on any device, making them universally attractive.

We evaluate these architectures to determine the best performing CNN for the classification task and aim to achieve good performance levels. We then seek to train multi-class models that enhance sensitivities for the mild or early stage classes, including various methods of data pre-processing and data augmentation to both improve test accuracy as well as increase our effective dataset sample size.

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, sc, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

Architecture

The network contains nine layers with weights.

The first layer is convolutional layer with 30 filters, kernel size 5x5, with Relu activation function which is followed by max pooling layer with size (2,2). The output of the max pooling layer is given as an input to another convolutional layer of 15 filters, kernel size 3x3 and the activation function Relu, which is followed by another max pooling layer of size (2,2).

A dropout layer and flatten layer is added before connecting to fully connected layers and the remaining three are fully-connected. The output of the last fully-connected layer is fed to a SoftMax which produces a distribution over the class labels. The network maximizes the multinomial logistic regression objective, which is equivalent to maximizing the average across training cases of the log-probability of the correct label under the prediction distribution.

In this research 5x5 filter is used at the convolutional layer after the convolution of the input image with the filter for each input image a 28×28 feature vector is produced. It is concordant to apply a nonlinear layer (or activation layer) instantly after conv layer. The persistence of this layer is to introduce nonlinearity to a system. To perform down-sampling by separating the input into rectangular pooling regions, and figuring the maximum of each region a max pooling layer is used.

Pooling layer perform down-sampling operations. Calculation of the output size, O of a pooling layer with input size, I Pooling filter size, F Padding, P and stride, S has done by

$$O = (I – F + 2×P) / S + 1$$

In our method we use 2×2 pool size and the output of the pooling layer is 12×12. Random selection of input elements to zero with a given probability is done in a dropout layer. It is a simple way to prevent overfitting in the neural network. Overfitting is a stern difficulty in such networks.

Large networks are also sluggish to use, making it more challenging to contract with overfitting by combining the predictions of many different large neural nets at the test time.

The technique for addressing this problem is Dropout. The core idea is to randomly drop units (along with their connections) from the neural network during training. In our proposed method we use the probability 0.3 at the dropout layer during the training. After the convolutional and down-sampling layers one or more fully connected layers is used. The layer in which the neurons associate to all the neurons in the previous layer is fully connected.

Fully connected layers combine all the features learned by the previous layers through the image to recognize the larger patterns. To classify the images the last fully connected layer combines the features. Hence, the Output parameter in the last fully connected layer is equal to the number of classes in the target data.

In our work, the output size is 14, corresponding to the 14 classes. An activation function, Softmax, normalizes the output of the fully connected layer and is also used in the convolutional neural network model.

The softmax layer output comprises positive numbers that sum to one, in the nest that can be used as the classification probabilities at the classification layer. At the final layer, the classification layer uses the probabilities given by the softmax activation function to find the input image classes and find the loss by comparing it with the preassigned ground truth classes.

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d (Conv2D)              (None, 24, 24, 30)        780
_________________________________________________________________
max_pooling2d (MaxPooling2D) (None, 12, 12, 30)        0
_________________________________________________________________
conv2d_1 (Conv2D)            (None, 10, 10, 15)        4065
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 5, 5, 15)          0
_________________________________________________________________
dropout (Dropout)            (None, 5, 5, 15)          0
_________________________________________________________________
flatten (Flatten)            (None, 375)               0
_________________________________________________________________
dense (Dense)                (None, 136)               51136
_________________________________________________________________
dense_1 (Dense)              (None, 49)                6713
_________________________________________________________________
dense_2 (Dense)              (None, 14)                700
=================================================================
Total params: 63,394
Trainable params: 63,394
Non-trainable params: 0
```

III. IMPLEMENTATION

1. *DATASET PREPARATION*

Preparation of the dataset is the fundamental concern for this work. Characters such as English digit, alphabet and mathematical symbols all can be well defined by their edges.

For this reason we first prepared the dataset as the most precedence given to its edges that enlighten the edges. We prepared some of the dataset by ourselves and also used a modified version of the NIST dataset which is similar to the popular MINIST dataset.

Each element of the dataset is an image of dimension 28x28, which is converted into a grayscale image using OpenCV. Initially, the dataset consisted of many elements - each of them assigned to a corresponding folder - which was then reduced to the following elements, each having about 4000 data items for training the network:

- Numbers (0-9)
- Mathematical operators (+, -, *)
- The alphabet ‘x’ for solving linear equations

This was done to remove any unwanted detections later on, thereby improving the accuracy of the model. During training, we gained more than 98.4% as training accuracy.

2. *PRE-PROCESSING*

Pre-processing of the input image is the procedure which encompasses changes and modifications to the image to make it fit for recognition. The following techniques may be used for image enhancement.

1. Conversion of RGB to Gray-Scale

First of all this coloured image is transformed into a typical gray-scale image and is represented through a single matrix because the detection of characters on a coloured image is more challenging than on a gray-scale image.

2. Binarization

Binarization is the procedure of choosing a threshold value for adaptation of pixel values into 0's and 1's. In this project for horizontal projection calculation 1's represent the black pixels while 0's characterize the white pixels.

3. Segmentation

Identifying the objects or other significant information in digital images segmentation is mostly used in image processing and computer vision application, which is the process of dividing an image into multiple parts. The segmentation used is character segmentation.

3.1 Character Segmentation

Character segmentation is a procedure that looks to decompose an image of a series of characters into sub images of individual symbols. In our proposed method we use a connected component analysis method for the segmentation of specific characters from the image. Problem arises at the point of extracting the math symbol '=' which is a combination of two connected components. If two consecutive components have the minimum height and are in the same horizontal direction we can consider these two into a single component.

3. *TRAINNING DATA USING CNN*

Since convolutional neural networks work on two-dimensional data and our dataset is in the form of 785 by 1. Therefore, we need to reshape it. Firstly, assign the labels column in our dataset to variable y_train. Then drop the labels column from the dataset and then reshape the dataset to 28 by 28. Now, the dataset is prepared and ready.

a. Building Convolutional Neural Network

For building the CNN, import the necessary libraries

```
np.random.seed(1212)
from keras.models import Model
from keras.layers import *
from keras import optimizers
from keras.layers import Input,Dense
from keras.models import Sequential
from keras.layers import Dropout
from keras.layers import Flatten
from keras.layers.convolutional import Conv2D
from keras.layers.convolutional import MaxPooling2D
from keras.utils import np_utils
```

b. Creating the model

The model selected is the Keras Sequential model, which is appropriate in applications where there is exactly one input and one output.

```
model=Sequential()
model.add(Conv2D(30,(5,5),input_shape=(28,28,1),activation="relu"))
model.add(MaxPooling2D(pool_size=(2,2)))
model.add(Conv2D(15,(3,3),activation="relu"))
model.add(MaxPooling2D(pool_size=(2,2)))
model.add(Dropout(0.2))
model.add(Flatten())
model.add(Dense(136,activation="relu"))
model.add(Dense(49,activation="relu"))
model.add(Dense(14,activation="softmax"))
```

c. Fitting Model

The training model is compiled using the Adam optimiser and the Categorical Cross Entropy loss function.

```
model.compile(loss="categorical_crossentropy",optimizer="adam",metrics=["accuracy"])
model.fit(np.array(l),cat,shuffle=True,epochs=10)
```

d. Solving the equation

Simple expressions are solved using the eval() function of Python while linear equations are solved using the Sympy Python library.

```
from sympy import solve
from sympy import sympify
from sympy.abc import F, x, y, z, a, b
from sympy.parsing.sympy_parser import parse_expr

if '*' in s:
    t = s.replace('x','')
    s=t

if "x" in s:
    try:
        equation = s.replace('--','=')
        a = []
        for i in range(0,len(equation)):
            if i==0:
                continue
            if equation[i] == 'x':
                if equation[i-1] == '+' or equation[i-1] == '-' or equation[i-1] == '=':
                    continue
                else:
                    a.append(i)

        counts = 0
        for i in a:
            if counts == 0:
                counts=1
                equation = equation[:i] + '*' + equation[i:]
            else:
                i=i+counts
                equation = equation[:i] + '*' + equation[i:]
                counts+=1

        lhs =  parse_expr(equation.split("=")[0])
        rhs =  parse_expr(equation.split("=")[1])
        solution = solve(lhs-rhs)
        decSol = round(solution[0]*1.0,2)

        print(solution[0]) #fractional solution
        print(decSol) #decimal solution
    except:
        print("invalid equation")
else:
    print(eval(s))
```

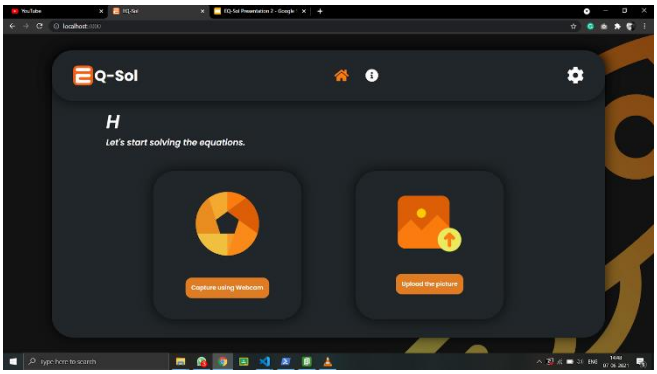
4. *FRONT-END*

On the Front-End side, we used the ReactJS framework. Also used here are NPM modules like react-router-dom, react-dropzone, reactstrap, react-image-crop and react-webcam.

The frontend has two features:

1. Camera: Here we can click a picture using from the webcam and proceed for upload
2. Upload a pic: Here we upload a picture from our computer and proceeds for upload

Also, there is an option to crop the image for selecting out the equation part of the image and proceed with that cropped image. It enhances the User Experience.



5. *INTERFACE*

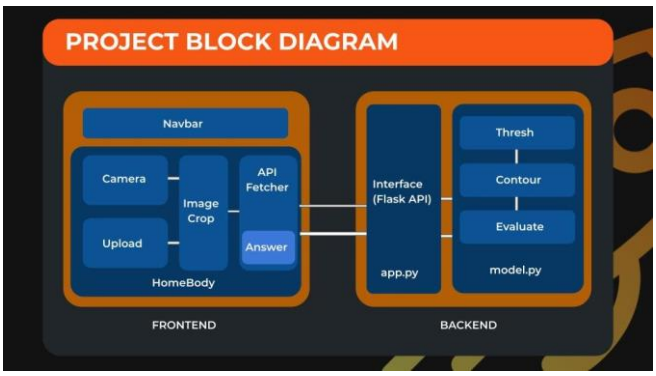
Interfacing denotes integrating Frontend with the created ML Model in the Backend. We have used Flask for implementing it along with modules like flask-cors for making the RESTful API that supports

any frontend framework. It generates an API which could be used by the frontend for data sending and receiving from the backend using standard HTTP protocol.

From the frontend part, we use the fetch function to fetch the API by POST method with image as a base64 encoded string to backend. In flask we declare the flask app with method POST. First, access the parameter using request and convert that base64 image string to normal image and save it in localhost directory. Then the url of the created image is passed to the ML model function.

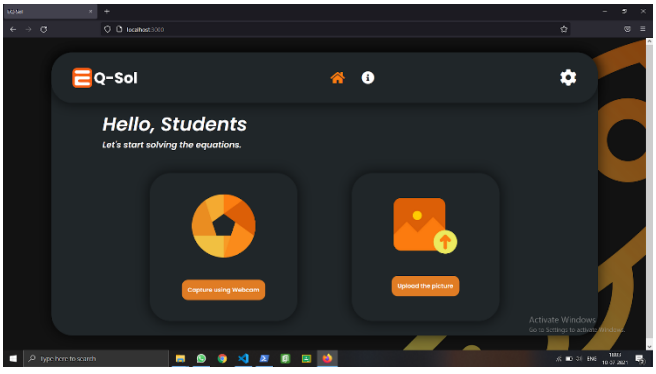
When it reaches the ML model, the backend code works and returns the result as a JSON object which contains the detected equation and its corresponding answer. Same result is passed to frontend by returning the result from the ML model function. In frontend the fetch function returns an object from the backend which is converted to normal form and displayed by accessing its attributes.

The website offers two options to scan the handwritten equation. It can either take a live photo using the webcam or a pic can be uploaded from the computer. It is then sent to the backend via Flask API. The characters/digits are then recognized and extracted. It undergoes several procedures like converting the image into grayscale, contour detection, reshaping etc and is fed to the model.It is then evaluated and sent to the frontend via Flask API where the answer, along with the type of equation, is displayed.

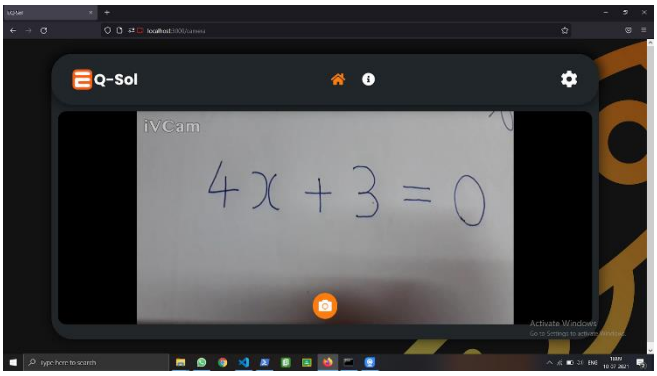


IV. *EXPERIMENTAL* RESULTS

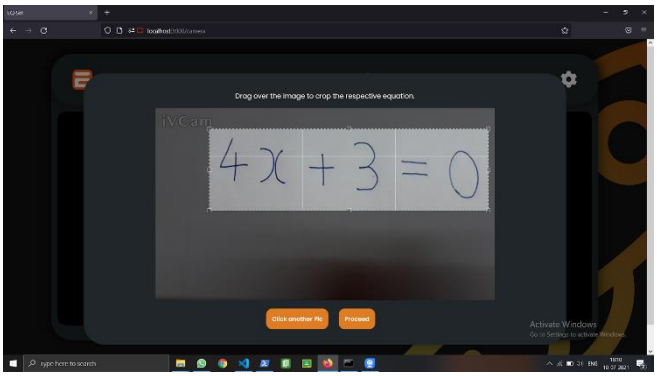
We have added screenshots which show the working of the application. It is able to detect the equation in the picture given both as upload a picture or capture a picture through the webcam and also able to show the solution of it.



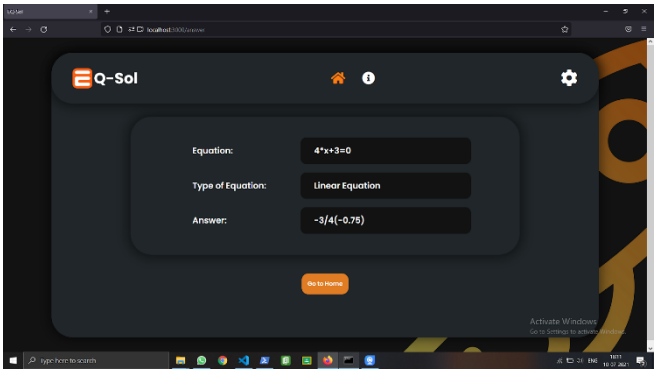
Home page



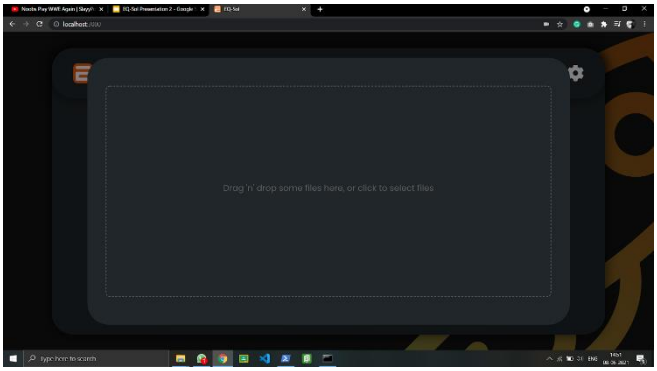
Captured image using webcam



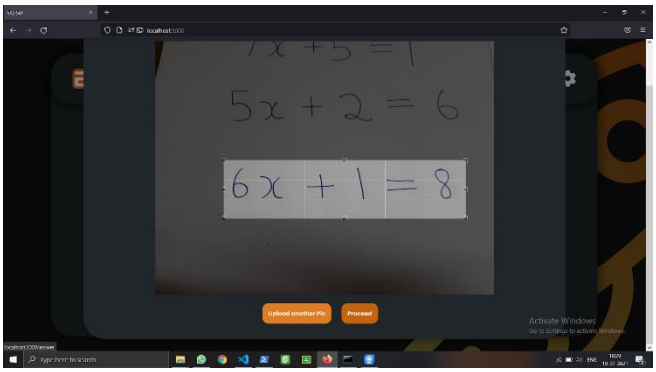
Cropped the image



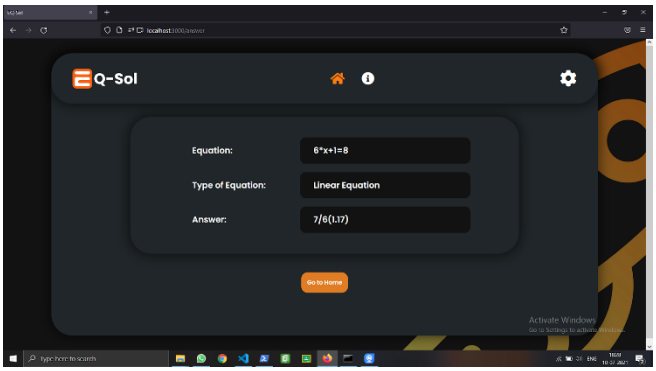
Answer



Drag and Drop the image



Cropped the image



Answer

V. EVALUATION

1. Training accuracy

The accuracy of our model based on examples it was constructed on : (using model.fit() function).

```
model.fit(np.array(l),cat,shuffle=True,epochs=10)

Epoch 1/10
1626/1626 [==============================] - 35s 16ms/step - loss: 1.7619 - accuracy: 0.6431
Epoch 2/10
1626/1626 [==============================] - 28s 17ms/step - loss: 0.2238 - accuracy: 0.9330
Epoch 3/10
1626/1626 [==============================] - 30s 18ms/step - loss: 0.1399 - accuracy: 0.9578
Epoch 4/10
1626/1626 [==============================] - 29s 18ms/step - loss: 0.1091 - accuracy: 0.9671
Epoch 5/10
1626/1626 [==============================] - 29s 18ms/step - loss: 0.0899 - accuracy: 0.9733
Epoch 6/10
1626/1626 [==============================] - 29s 18ms/step - loss: 0.0744 - accuracy: 0.9765
Epoch 7/10
1626/1626 [==============================] - 30s 18ms/step - loss: 0.0690 - accuracy: 0.9793
Epoch 8/10
1626/1626 [==============================] - 29s 18ms/step - loss: 0.0626 - accuracy: 0.9803
Epoch 9/10
1626/1626 [==============================] - 29s 18ms/step - loss: 0.0574 - accuracy: 0.9823
Epoch 10/10
1626/1626 [==============================] - 30s 19ms/step - loss: 0.0503 - accuracy: 0.9846
<tensorflow.python.keras.callbacks.History at 0x216a8787580>
```

2. Testing Accuracy

We divided the dataset into two - training and testing dataset in the ratio 7:3. On evaluating, we got the accuracy as 98.4% for the testing dataset.

```
In [23]: model.evaluate(np.array(m),tac)
         432/432 [==============================] - 3s 6ms/step - loss: 0.0720 - accuracy: 0.9841
Out[23]: [0.0720311626791954, 0.9840545058250427]
```

VI. CONCLUSION

In this paper we are focusing on recognizing handwritten Mathematical expressions and equations. Feature extraction was the most complicated part of classification. Convolutional Neural Network the most powerful classification model is used in the classification part. Finally we obtained a state of art performance in the detection phases and also in the solution phases. To simplify the math, the main task is the feature extraction from the image and recognition with the help of the CNN model. Any person can easily use this as the process is easy and does not require any advanced knowledge of the technology.

In this paper we mainly focused on recognizing handwritten Mathematical expressions and equations. Projection analysis specially compact horizontal projection is used for the segmentation of each line of quadratics. Connected components which have a very high success rate are used for character segmentation.

Moreover with some predefined features about handwriting it is difficult to recognize handwriting. Convolutional Neural Network the most powerful classification model is used in the classification part.

VII. REFERENCES

[1]. https://www.researchgate.net/publication/326710549_Recognition_and_Solution_for_Handwritten_Equation_Using_Convolutional_Neural_Network
[2]. https://www.nist.gov/itl/products-and-services/emnist-dataset
[3]. https://aihubprojects.com/handwriting-recognition-using-cnn-ai-projects/
[4]. https://www.geeksforgeeks.org/handwritten-equation-solver-in-python/
[5]. Mandal, S Shahnawazuddin, Rohit Sinha, S. R. Mahadeva Prasanna and Suresh Sundaram, "Exploring Sparse Representation for Improved Online Handwriting Recognition", 2018 16th International Conference on Frontiers in Handwriting Recognition, 2018.
[6]. S.E Benita Galaxy and S. Selvin Ebenzer, "Enhancement of segmentation and zoning to improve the accuracy of handwritten character recognition", International Conference on Electrical Electronics and Optimization Techniques (ICEEOT), 2016.
[7]. R. Sethi and I. Kaushik, "Hand Written Digit Recognition using Machine Learning," 2020 IEEE 9th International Conference on Communication Systems and Network Technologies (CSNT), 2020, pp. 49-54, doi: 10.1109/CSNT48778.2020.9115746.
[8]. Catherine Lu and Karanveer Mohan, "Recognition of Online Handwritten Mathematical Expressions Using Convolutional Neural Networks", cs231n project report stanford, 2015.
[9]. Martin Riedmiller and Heinrich Braun, "A Direct Adaptive Method for Faster Backpropagation Learning: The RPROP Algorithm", IEEE international conference on neural networks, 1993.
[10]. Braun Riedmiller, "A direct adaptive method for faster backpropagation learning: The rprop algorithm", Neural Networks, 1993.
[11]. Ahmad-Montaser Awal, Harold Mouchère and Christian Viard-Gaudin, "Towards Handwritten Mathematical Expression Recognition", 2009 10th International Conference on Document Analysis and Recognition.

[12]. Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever and Ruslan Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting", Journal of Machine Learning Research, vol. 15, pp. 1929-1958, 2014.

[13]. S. N. Shuvo, F. Hasan, S. A. Hossain and S. Abujar, "Handwritten Polynomial Equation Recognition and Simplification Using Convolutional Neural Network," 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2020, pp. 1-6, doi: 10.1109/ICCCNT49239.2020.9225587..