GRENZE
Scientific Society

A Hybrid Approach for the Recognition of Handwritten Digits Using Machine Learning

Mrs.V.Sivasakthi M.E[1], A.Celestina Rose[2], S.Kalaimathi[3], S.Mullai[4]

Assistant Professor, Anjalai Ammal Mahalingam Engineering College, Kovilvenni, Tamilnadu[1]

B.Tech/IT., Anjalai Ammal Mahalingam Engineering College, Kovilvenni, Tamilnadu[2,3,4]

Abstract— **In this modern world, technology plays an vital role to reduce the man power. Costly manual labour is required to do a tedious job of converting the physical written data and information into digital form. Handwritten digit recognition is a methodology that automatically recognizing and detecting handwritten digital data through different Machine Learning models .In this project we use hybrid approach of machine learning algorithms to enhance the productiveness of technique and reduce the complexity of using various models.Handwritten Digit Recognition is a pivotal concern in computer vision.We mainly focused on performance and consistency of machine learning algorithms through hybrid approach.It will be more efficiency and accuracy for any digital data sets compared to the existing system.We have used the MNIST dataset for training and recognition which consist of set for hand written digits (0-9).The data consist of 70,000 images for training and testing. Each digit is represented as a 28 by 28 grey scale pixels i.e. 784 pixels intensities for better results.**

Index Terms— **Machine learning, MNIST, CNN, SVM.**

I. INTRODUCTION

Machine learning is an application of Artificial Intelligence that learns from previous experience and improves automatically through experience. Machine learning focus on the development of computer programes that can access data and use it to learn from themselves. Machine learning technology makes a machine efficient to perform pattern or text recognition. Handwritten digit recognition has become an issue of interest among researchers. There are a large number of papers and articles are being published these days about this topics..The major concerns of handwritten digit recognition is handwriting patterns differ according to the person it is normally quite difficulty to recognize even human kind handwritten digits.we use MINIST and keras . Keras is a high level deep networks library written in python and capable of running by using backend as Tensor Flow.It was developed for focusing mainly on deriving fast training. Supports both convolutional and recurrent networks and combination of both also Manually. For efficient and accuracy we use SVM and CNN. Support vector machine is a supervised machine learning algorithm which can be used for both classification and regression challenges. It is highly preferred by many as it produces significant accuracy with less computation power. Convolutional neural network is a deep learning algorithm most commonly applied to analyzing visual imagery and it successfully applied to hand written digit recognition numerous times, with very small error. Handwrittendigit acknowledgment is presently generally acclimated process in bank checks, postal locations, and so on consequently.

II. LITERATURE SURVEY

A. Learning Algorithms And Models.

In this paper,various machine learning algorithms to enhance the productiveness of technique and reduce the complexity using various models . Various machine learning algorithms such as SVM, CNN, Quantum computing,K-nearest neighbor algorithm, Deep learning used in recognition technique. Machine learning models and algorithms are very efficient to recognize or detect patterns with different writing style.According to survey by using Convolutional neural network accuracy increase to the 99.89% accuracy most among all similarly Double Q learning algorithm and Support vector machine also given high accuracy.

B. Handwritten Digits Recognition using Machine Learning Algorithms

In hand written digit recognition this project leads to development of different patterns in writing.It discusses the solution to a part of problem as we have limited the scope to only the hand written digits (0-9). We have trained a model using deep neural networks for digit recognition using Google's Machine Learning tool Tensor Flow and Python Programming language. We have used the 'MNIST DATABASE' which consist of training and test set for hand written digits (0-9) of size (28x28) pixels i.e. 784 pixel. The limitation of this model will be if digits other than (0-9) are given then the model will not be able to recognize and classify it and the model will be able to predict numbers only in black and white images.

C. Improved Handwritten Digit Recognition Using Convolutional Neural Networks (CNN).

Traditional systems of handwriting recognition have relied on handcrafted features and an outsized amount of prior knowledge.Our aim within the proposed work is to explore the varied design options like number of layers, stride size, receptive field, kernel size, padding and dilution for CNN-based handwritten digit recognition. additionally , we aim to guage various SGD optimization algorithms in improving the performance of handwritten digit recognition. our objective is to realize comparable accuracy by employing a pure CNN architecture without ensemble architecture, as ensemble architectures introduce increased computational cost and high testing complexity

D. Handwritten Digit Recognition using CNN

Digit Recognition may be a noteworthy and important issue. because the manually written digits aren't of an identical size, thickness, position and direction, during this manner, various difficulties must be considered to work out the difficulty of handwritten digit recognition.The aim of this project is to implement a classification algorithm to acknowledge the handwritten digits. the consequences of probably the foremost broadly utilized Machine Learning Algorithms like SVM, KNN and RFC and with Deep Learning calculation like multilayer CNN utilizing Keras with Theano and Tensorflow. Utilizing these, the accuracy of 98.70% utilizing CNN (Keras + Theano) when contrasted with 97.91% utilizing SVM, 96.67% utilizing KNN, 96.89% utilizing RFC was obtained.

E. Offline Handwritten Digits Recognition Using Machine learning.

The problem of handwritten digit recognition has long been an open problem within the field of pattern classification.The main objective of this paper is to supply efficient and reliable techniques for recognition of handwritten numerals by comparing various existing classification models.

III. PROPOSED WORK

In this project we mainly focused on performance and consistency of machine Learning Algorithm through hybrid approach . It will be more efficient and accuracy for any digital Data sets compared to the Existing System. In Handwritten Digits Recognition hand written digit images are given as input. The model are often recognized by the system. an easy artificial neural network (ANN) has an input layer, an output layer and a few hidden layers between the input and output layer.Using MNIST datasets, error rates are being observed [19]. to wash blur images CNN is getting used . For this we proposed MNIST dataset. This approach reaches an accuracy of 98%. The dataset already features a well-defined train and test dataset that we will use.The MNIST handwritten digits [24] database is employed for the experiment. Out of 70,000 scanned images of handwritten digits from the MNIST database, 60,000 scanned images of digits are used for training the network and 10,000 scanned images of digits are wont to test the network. the pictures that are used for training and testing the network all are the grayscale image with a size of 28×28 pixels.MNIST datasets are the simplest and documented , and simply understood dataset within the computer vision branch and machine learning to use it as

first   dataset which we will use in our journey of deep learning. After implementing we will find that our network can classify the digits upto >97% accuracy with less training time
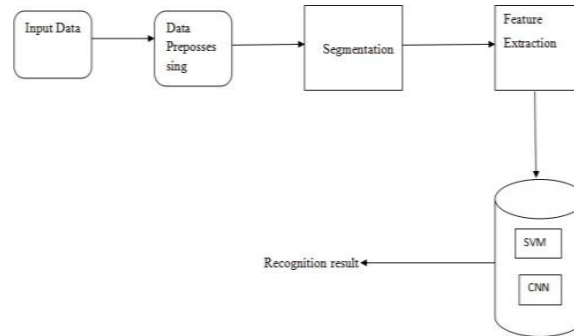

Figure 1. System Architecture

A. Modules And Descriptions Pre Processing

The role of the pre-processing step is it performs various tasks on the input image. It basically upgrades the image by making it reasonable for segmentation. For the foremost part, noise filtering, smoothing and standardization are to be wiped out this stage. The pre-processing additionally characterizes a smaller portrayal of the instance . Binarization changes over a gray scale image into a binary image.

B. Segmentation

Pre-processed digit images are segmented into a sub-image of individual   digits,   which   are assigned variety to every digit.In this a string of characters are often read and segmented into isolated characters. Segmentation is that the partition of images into set of regions. Each individual   digit   is   resized   into pixels. During this step a foothold detection technique is getting used for segmentation of dataset images. Feature extraction Feature extraction is that the pre- processed images are represented within the sort of a matrix which contains pixels of the pictures that are of very large size. It is a process of dimensionality reduction by   which an   initial   set of data is reduced to more manageable groups for processing.In the feature extraction stage redundancy from the info is removed.

This   paper   compares   the performance of machine learning classifier models namely Neural Network, K-Nearest Neighbor (K-NN), Random Forest, Decision Tree and Bagging with gradient boost The results of this paper shows that K- NN has equally high accuracy of 96.7% compared to Neural Network of 96.8%, but K-NN achieves a processing speed with almost 10 times faster.

C. Digit Recognition Using Learning Algorithm

Support vector Machine - SVM is a supervised Learning method. A Support vector machine is a machine learning algorithm that analyses data for classification and regression analysis. It is   trained with a series of data ,the task of an SVM is to determined which category a new datapoint belongs in.

D. Convolutional Neural Network

Convolutional Neural Network is a Neural network that has one or more convolutional layers and are used mainly for image processing,classification,segmentati on and so on. A covolution can be thought as "looking at a function's surrounding to make better/ accurate predictions of its outcome".

E. Classifies Digit from the training set

The recognition phase next undergoes training. The more the training set the better the system learns. The classifiers of the system are trained in this phase and are now ready for actual prediction. It is initial set of data used to help a program understand how to apply neural networks and produce results. It is then complimented by validation and testing dataset. The initial approach to the training set images that are to be processed in order to reduce the data, by thresholding them into a binary image. As the training data provides the correct accuracy of hand written digits of the training examples, using large datasets(MINIST)will ensure more accurate .

IV. IMPLEMENTATION AND ANALYSIS

In our implementation we will be using python because we have our keras deep learning library built in python . So by using keras models we can implement our network and create a driver program to call the network to take inputs from the data set. The driver program also has learning algorithm , training and testing datasets. MNIST datasets are the best and well known , and easily understood dataset in the computer vision branch and machine learning to use it as first dataset which we can use in our journey of deep learning. After implementing we can find that our network can classify the digits upto >98% accuracy with less training time The objectives of handwritten character recognition that we have seen in real time is it has reduced manual work so that the old literatures can be converted into digitized for manually Our network has been trained with and tested with 70,000 datasets.

V. CONCLUSIONS

Performance of a network depends on many factors like low memory requirements, low run time and better accuracy , although in this paper it is primarily focused on getting better accuracy rate for classification. .In fact,the time complexity of current algorithms or models is very high because of contiguous accumulation and incremental development of handwritten digit sample collection of precision of identification.After analysing the results of each model, Each classifier has its own exactness and time utilization. It is remarkable that their accuracy depends not only of the sloppiness of the writing but also of the classes of numerals. It is clear that SVM has given the least error rate of about 2.25% on the trained data set. we obtained an accuracy of > 97% which is good enough to test our classification implementation
We have given a learning rate of 0.01 to our algorithm and obtained good classification results , everytime after training we are taking random inputs from our testing dataset and calculating the efficiency each time it is excecuted.

REFERENCES

[1] Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, volume 1, pages 886–893.IEEE, 2005.
[2] Srivastava, Shobhit &Kalani, Sanjana& Hani, Umme & Chakraborty, Sayak. (2017). Recognition of Handwritten Digitsusing Machine Learning Techniques. International Journal of Engineering Research and. V6. 10.17577/IJERTV6IS050456.
[3] Leticia M. Seijas and Enrique C. Segura,"A Wavelet-based Descriptor for Handwritten Numeral Classification," 2012 International Conference on Frontiers in Handwriting Recognition.
[4] Saeed Al-Mansoori paper on "Intelligent Digit Recognition using Artificial Neural Networks Vol 5, Issue 5, (Part-3) May 2015 , pp 46-51
[5] Yann LeCun, Leon Bottou,Yoshua Bengio and Patrick Haffner , paper on "Gradient Based Learning Applied to Document Recognition" , Proc of the IEEE , NOVEMBER 1998