RECOGNITION OF HANDWRITTEN NUMBERS WITH DATA SETS

M. Fatih Cambek[1] and Pinar Kirci[2]

[1]Department of Computer Engineering, Faculty of Engineering, Istanbul University-Cerrahpasa, Istanbul

[2]Department of Computer Engineering, Faculty of Engineering, Bursa Uludag University, Bursa pinarkirci@uludag.edu.tr

Turkey

- In the project, with the algorithm developed over the data sets, the possible writing forms of the numbers will be taught to the machine and it will be ensured that machine can distinguish the handwritten digit combinations they will encounter.

- With the presented machine learning project, it is aimed to define numbers in handwritten documents.

- In this way, it will reduce the need for human labor. In this way, it is aimed to accelerate document processing and to increase user satisfaction.
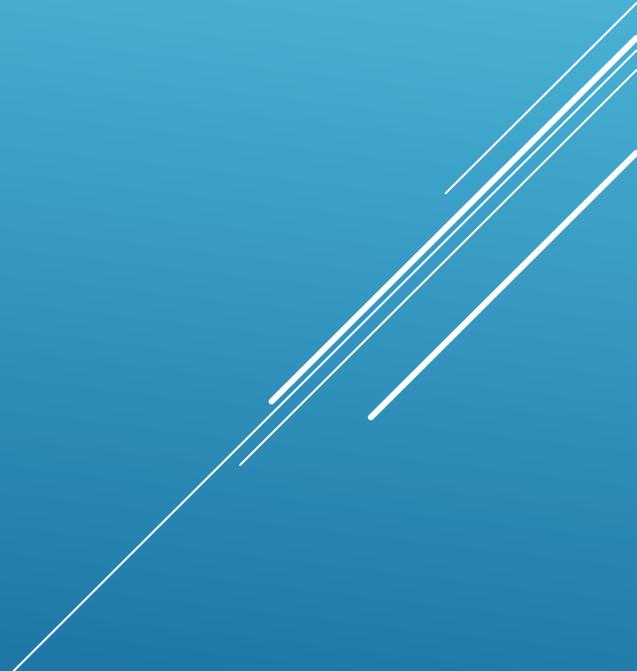
- Handwriting recognition is the recognition of handwritten letters, numbers and symbols by computer systems.

- Although it is quite easy for humans, it is a very difficult problem to perceive lines and curves on a ground as letters and numbers automatically and as meaningful words in a more advanced stage.

- Today, with the developing computer technology, documents have started to be created and filled in computer environment.
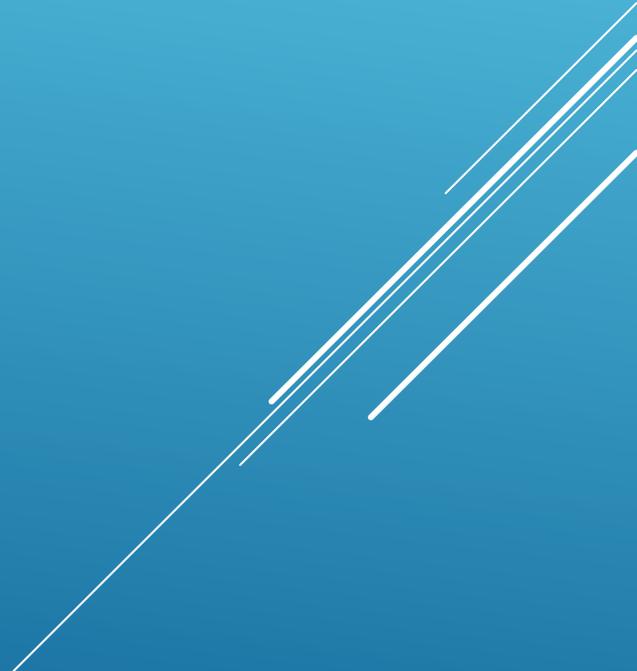
- Current technology is still at a limited level in handwriting recognition.

- Handwriting recognition is still not a fully solved problem.

- The difficulty in handwriting recognition is due to the large number of different typefaces and the fact that the letters are linked to each other, as well as differing from person to person.

- The style of writing can vary depending on the situation and the pen or paper used. Letters can be of very different shapes and sizes, depending on the writing style and speed of the people.

- Machine Learning is a method that makes inferences from existing data using mathematical and statistical methods and makes predictions about the unknown with these inferences.

- The handwritten address interpretation research began at CEDAR in 1987, funded by the United States Postal Service (USPS).

- The goal was to automate postal sorting through a system that can read a handwritten street address and postcode and encode each envelope with the destination address for machine sorting.

- This research ultimately led to the development and deployment of a system that automates mailing sorting through image analysis, number recognition, word recognition, mail directory search, and assigning a barcode that specifies the destination address.

- Since field testing began in 1996, Handwriting Address Interpretation System (HWAI) has been implemented in all USPS mail processing centers.

- Handwriting can take many different forms.

- You can use letters that are written quite properly and separately, which we can call hand print, as well as all letters can be connected to each other, or a mixture of the two can be used.

- While separated letters are much easier to recognize, recognition of compound letters is a more difficult problem to solve because it requires separation of letters.

- The graph here show the training curves on the datasets we have received.

- *Fig. 1. The X_train table is created by ordering the data from the data sets by index numbers (3822).*

- The graph here show the training curves on the datasets we have received.

- The data in the datasets contain dozens of different spelling of the numbers as shown in figure 1 and figure 2.

- *Fig. 2. Data set content*

- Our data sets consist of lines as shown in figure 2.

- The data in these lines are divided into two parts.

- The first 64 data shows how the numbers entered are formed (indicated by a red line) and the number at the end (yellow dotted display) shows with which number these 64 data correspond to the rightmost numbers (target) form our y train table.

- *Fig. 3. Shows the results of numbers in data sets*

- In our Y_train table as given in figure 3, we have the table we call target, that is, match our results.

- It allows matching the numbers that come as a result of the rows we receive in the data sets with the results.

- *Fig. 4. Top 50 Data Table of Train Data Set*

- We plotted the first 50 data in our train data set in figure 4.
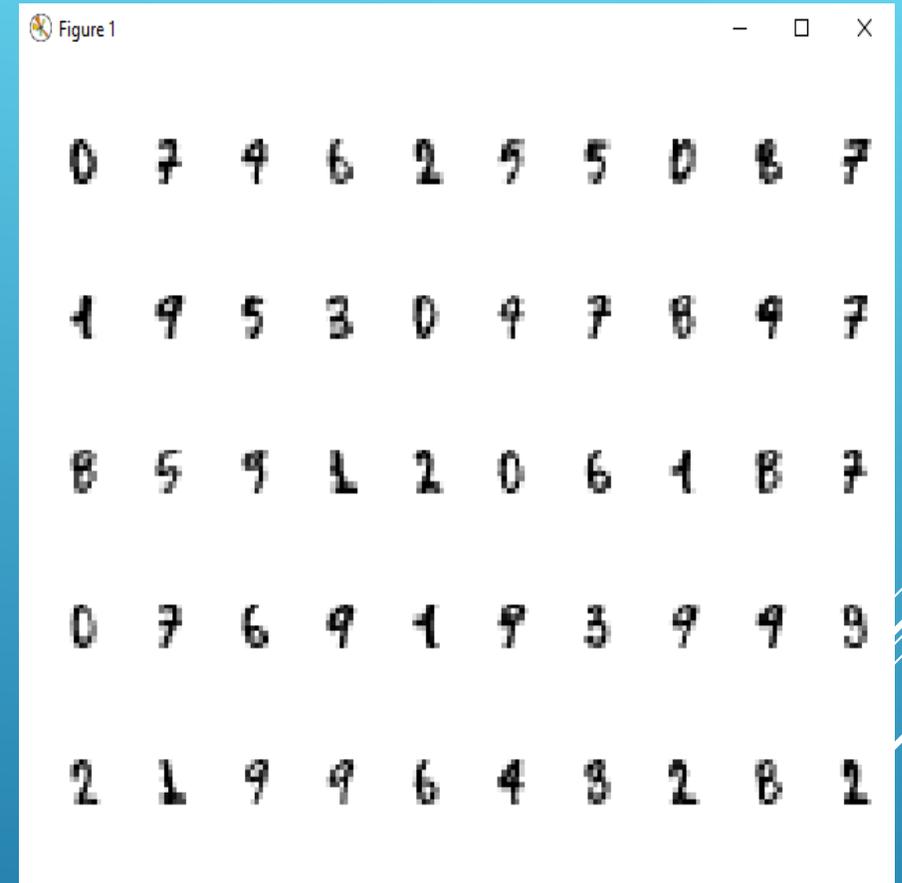
- *Fig. 5. Test results*

- In figure 5, we see the first 50 data and estimated results in the test data set.

- Each number is estimated with its labels and the estimation results are written.

- The successful predictions are shown in green, the unsuccessful ones are shown in red.
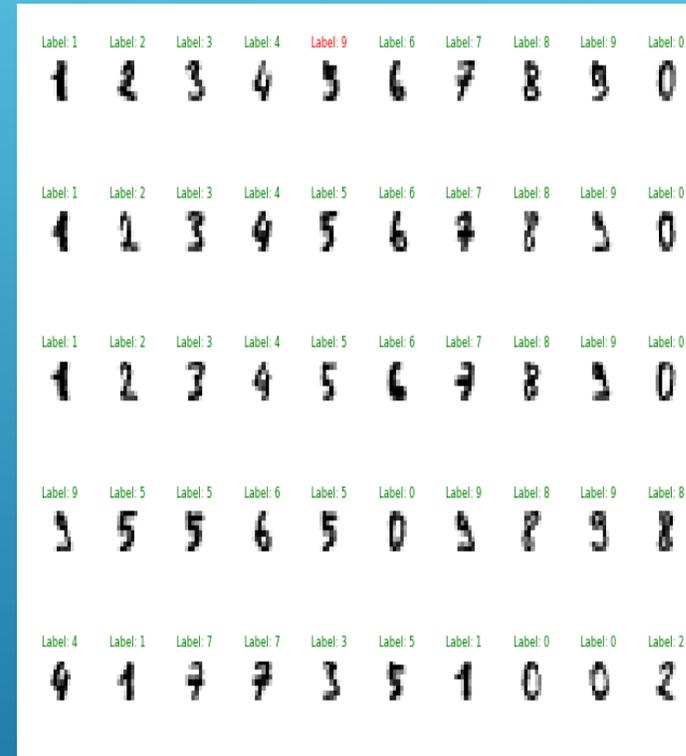
- Python platform 3.6.4 version is used in programming. The training process was carried out by using pandas library methods in the python.

- In this study, dozens of different spelling types of numbers were taught to the program with the first data sets.

- Then, the rate of matching the numbers in the data we tested with the ones entered was examined.

- In the light of the data obtained, the system has passed a certain success rate.

THANK YOU