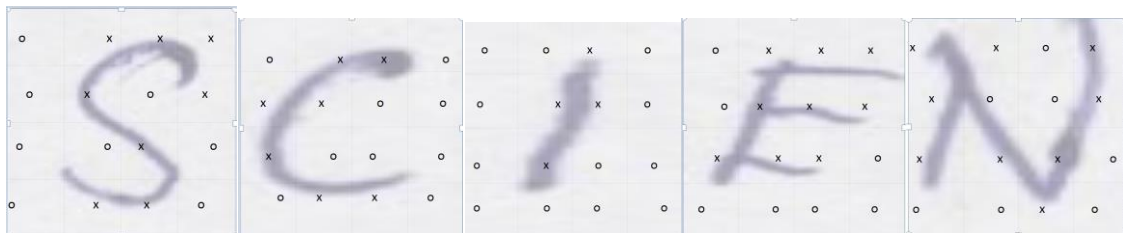
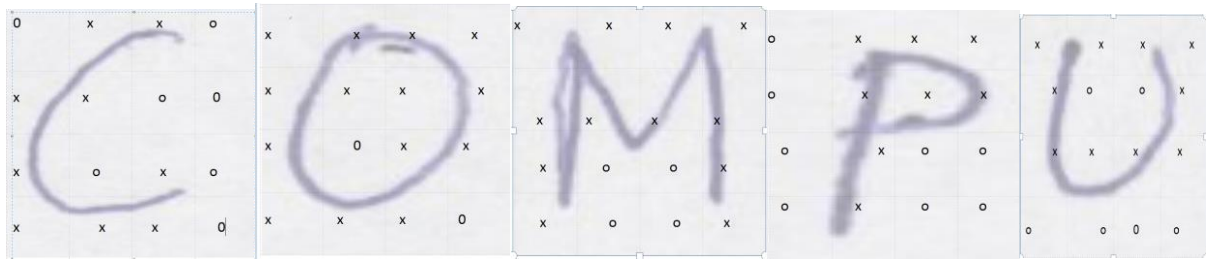


## Training data

1)



compu

ter

science

nce

computer

2)

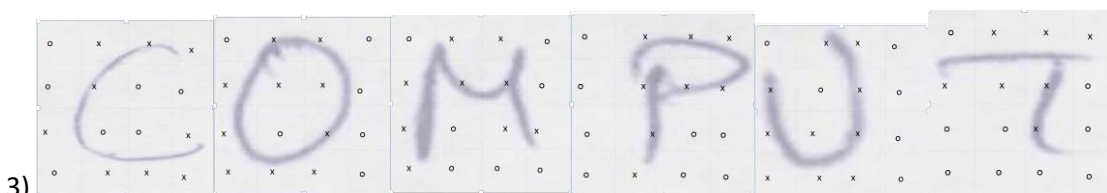
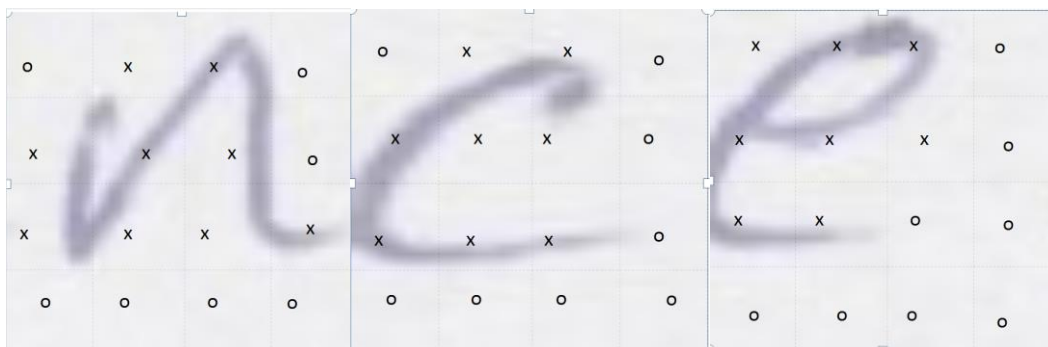
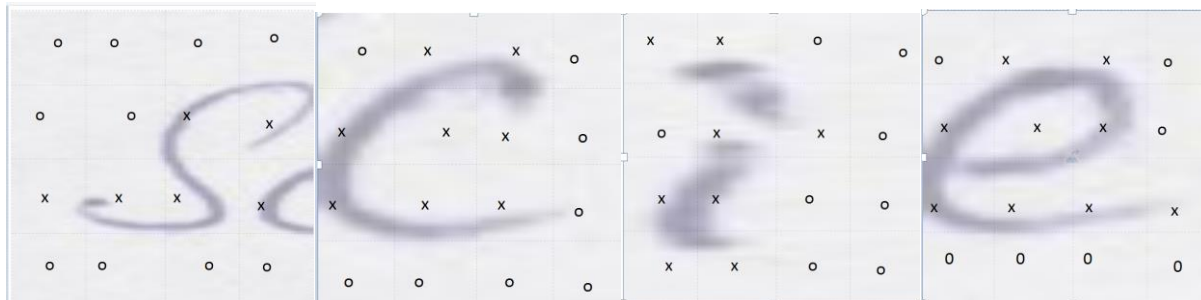
TER

SEEN

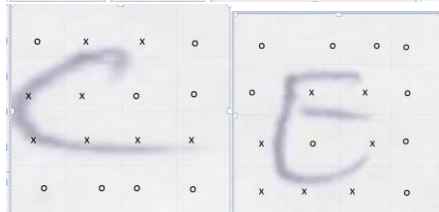
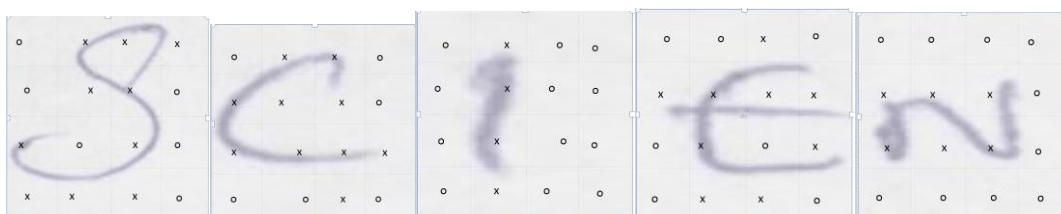
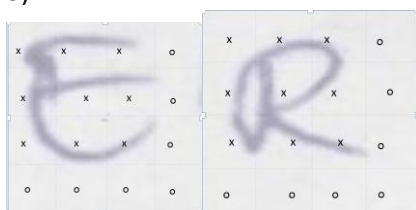
CE

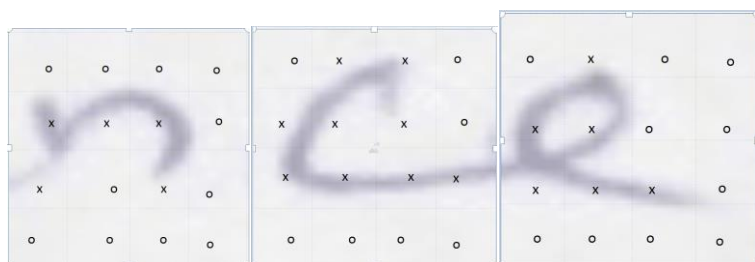
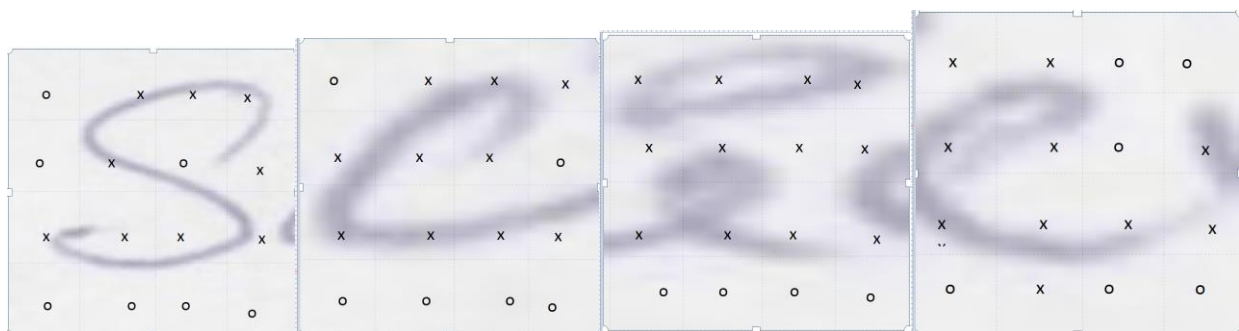
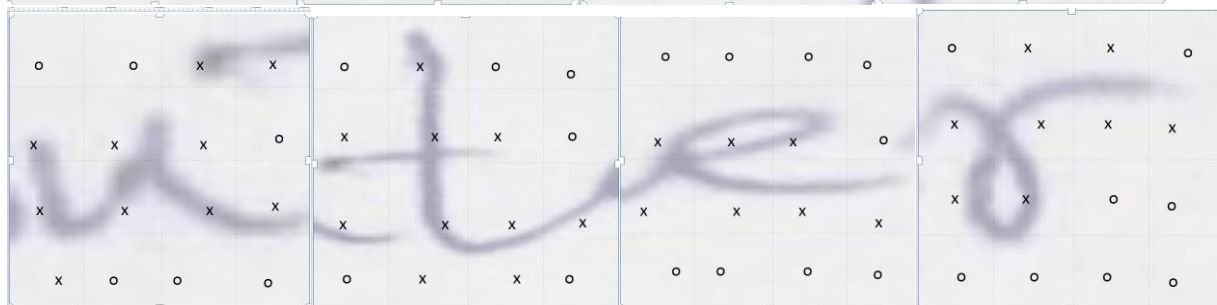
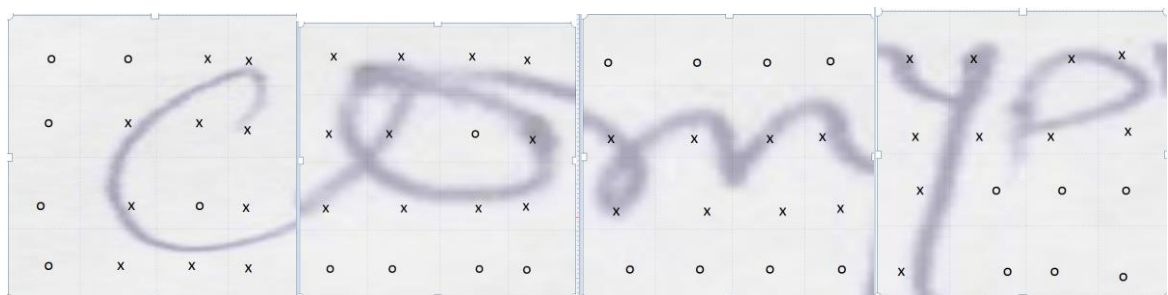
COMP

utes



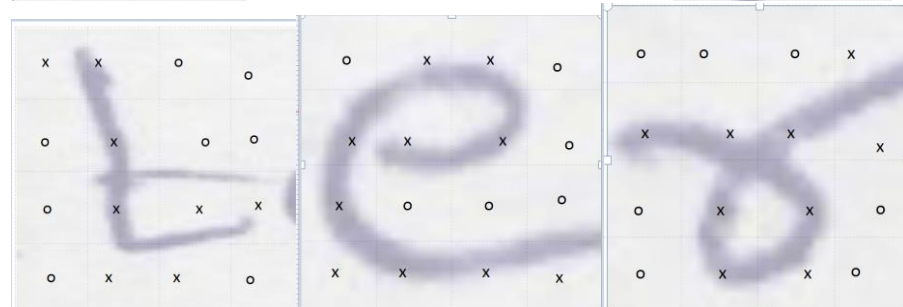
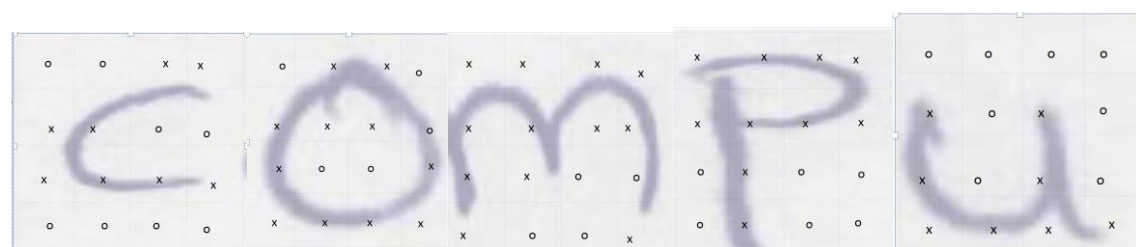
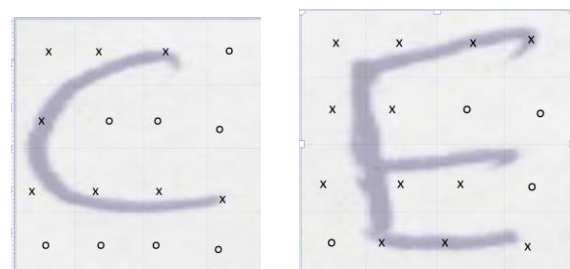
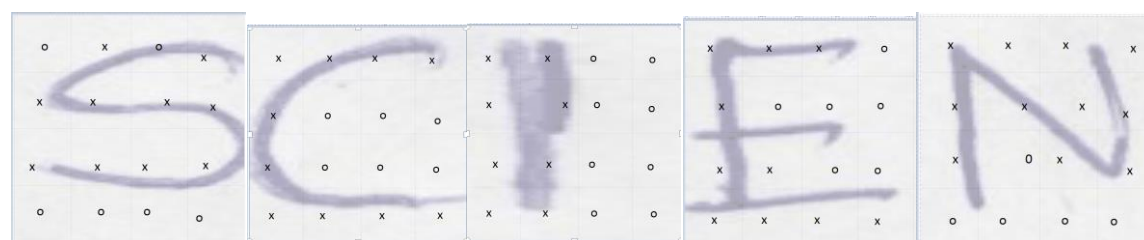
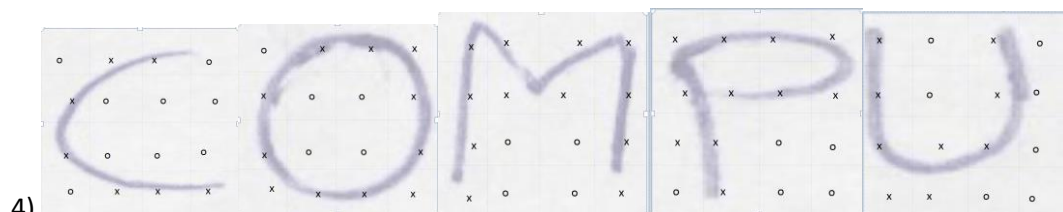
3)

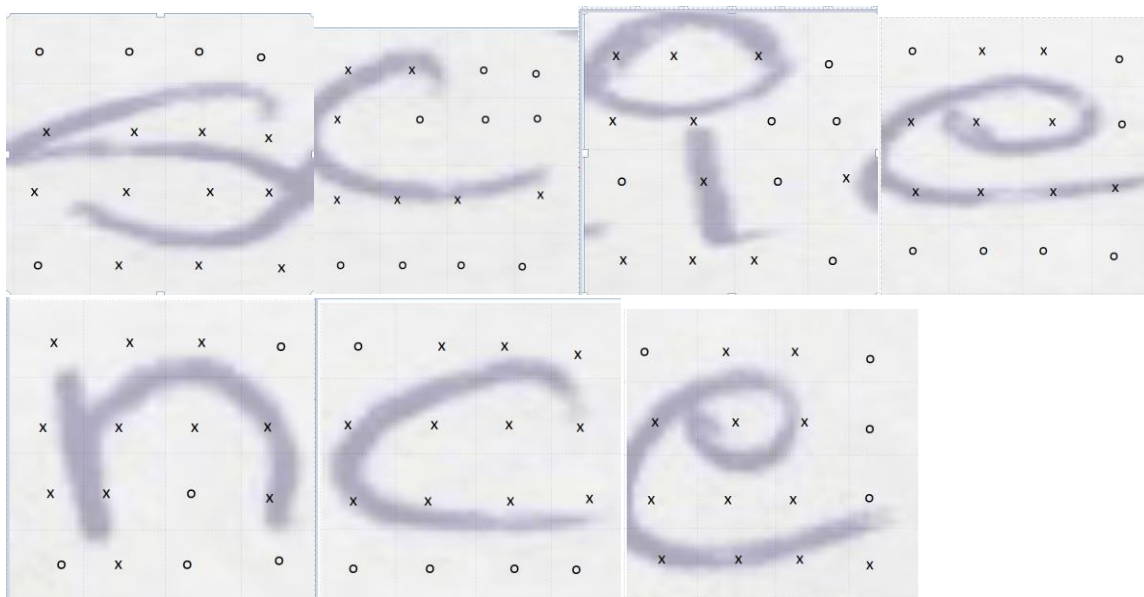




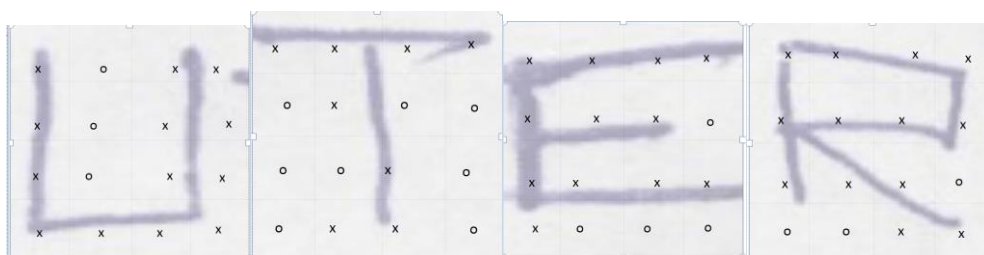


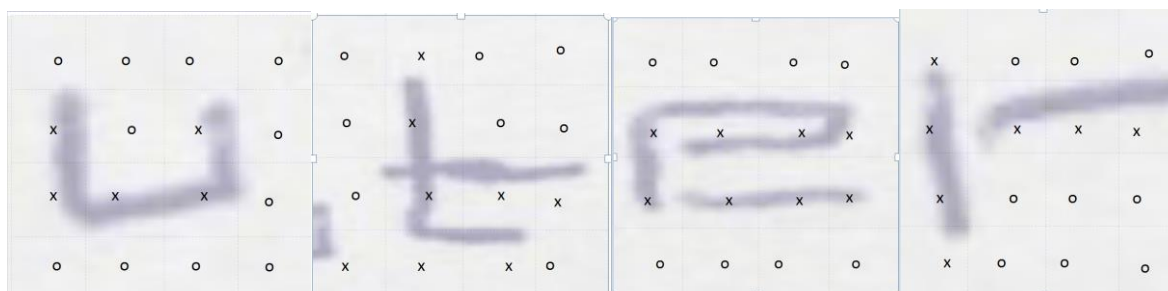
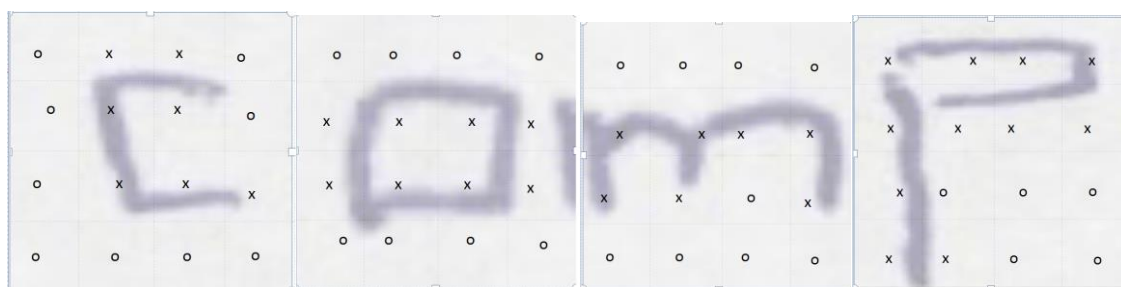
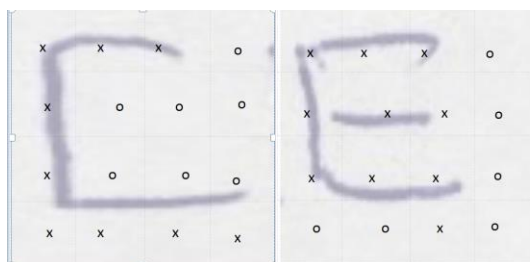
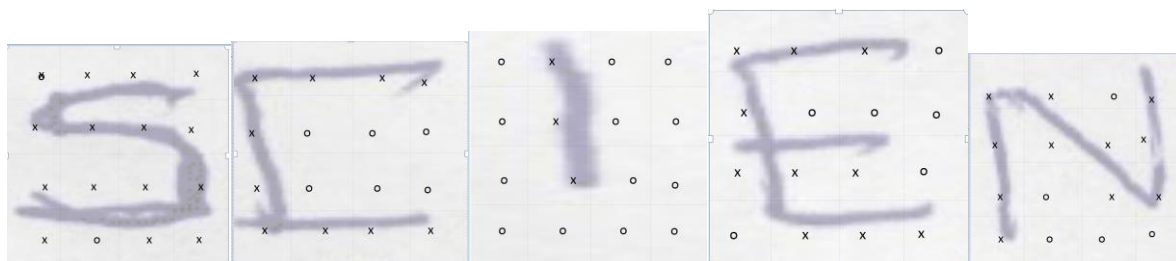
4)



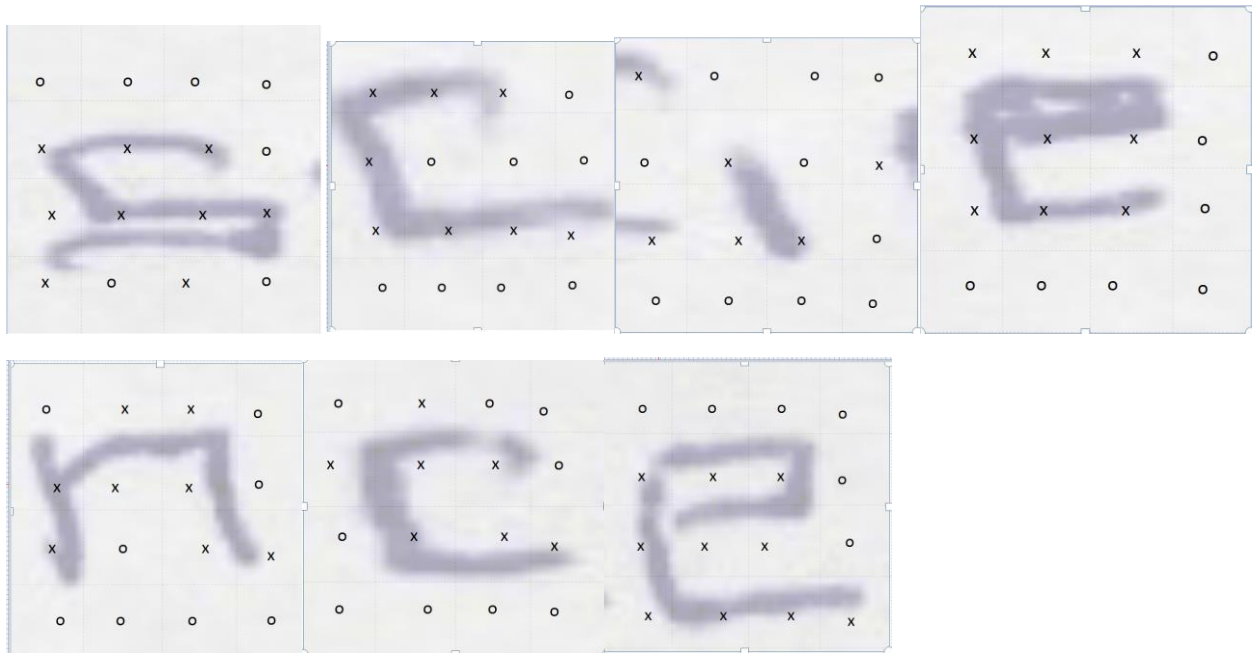


5)



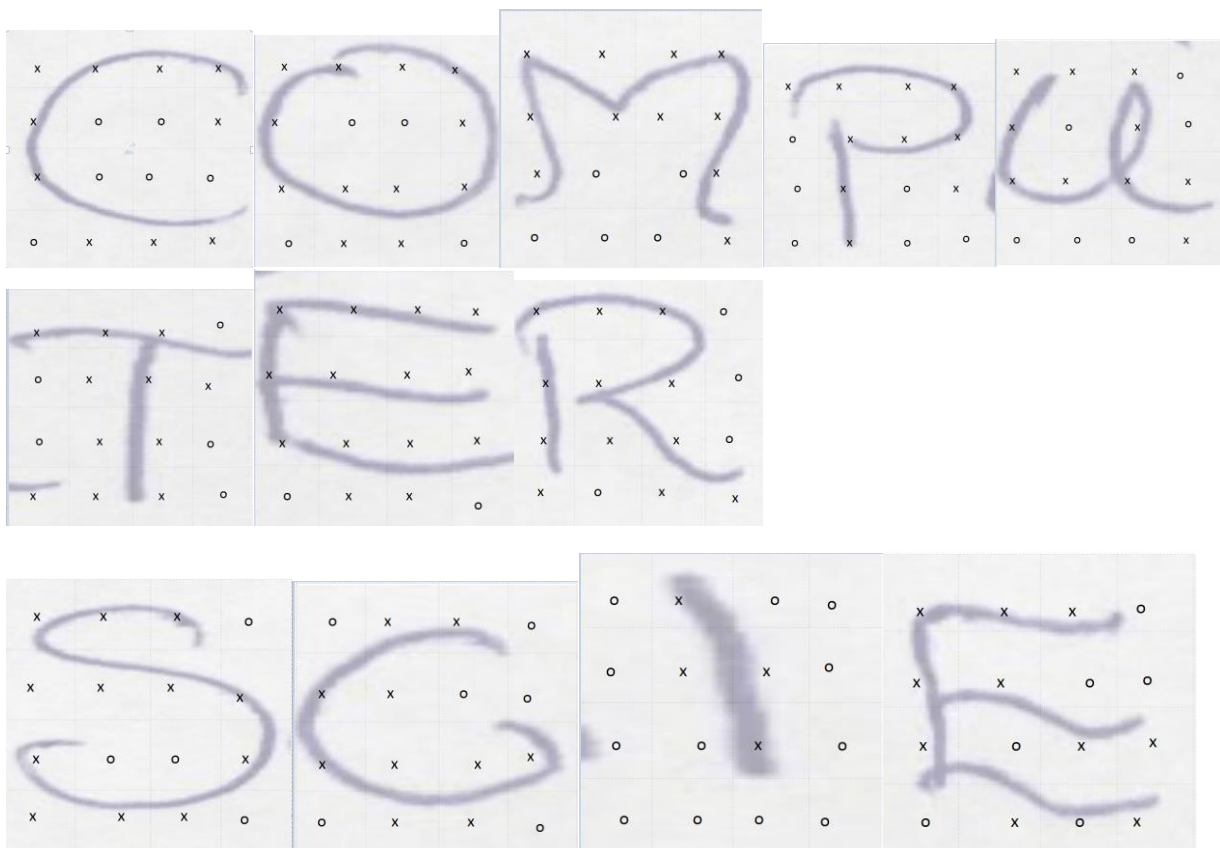


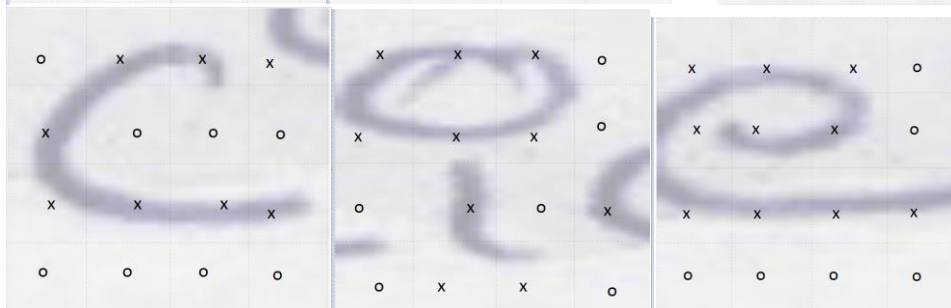
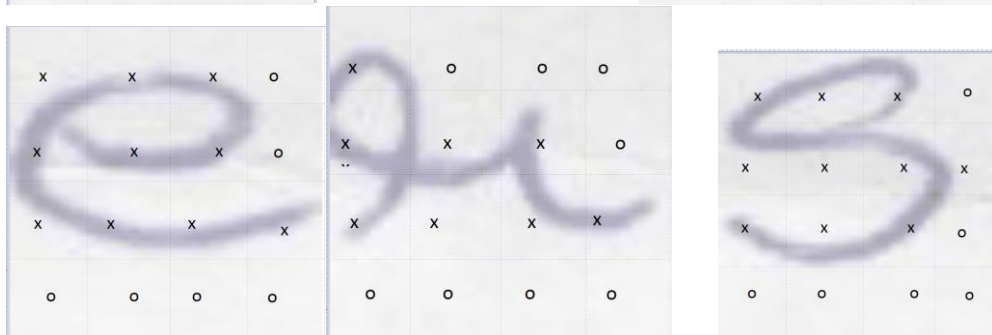
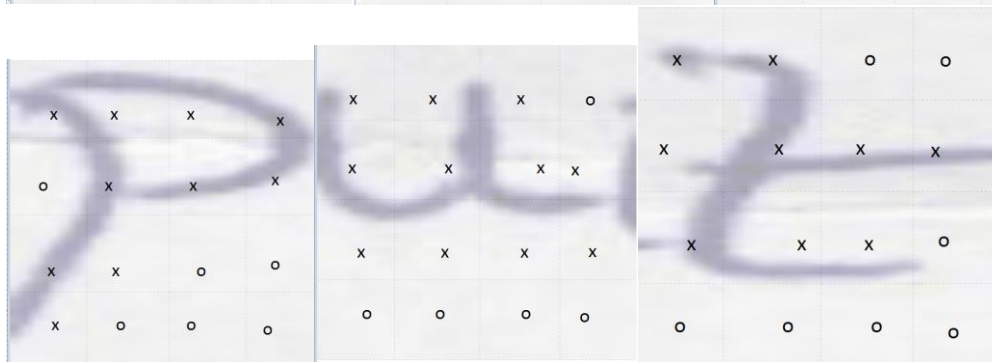
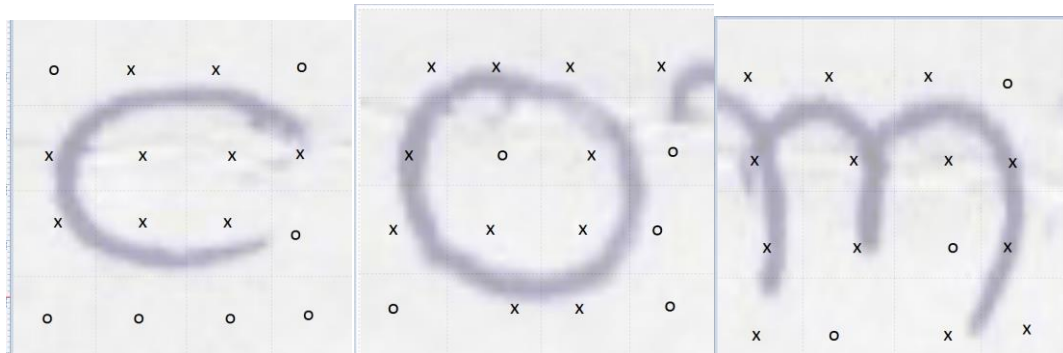
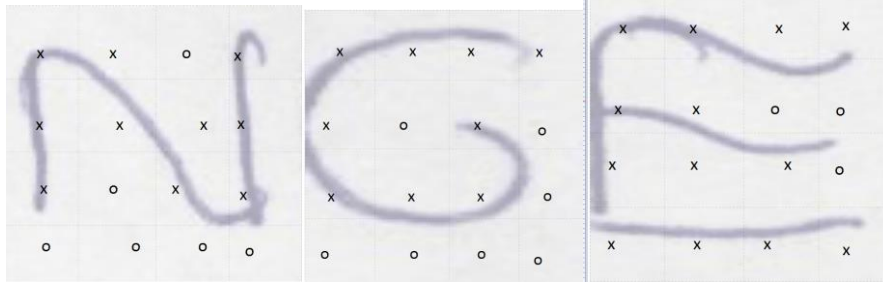


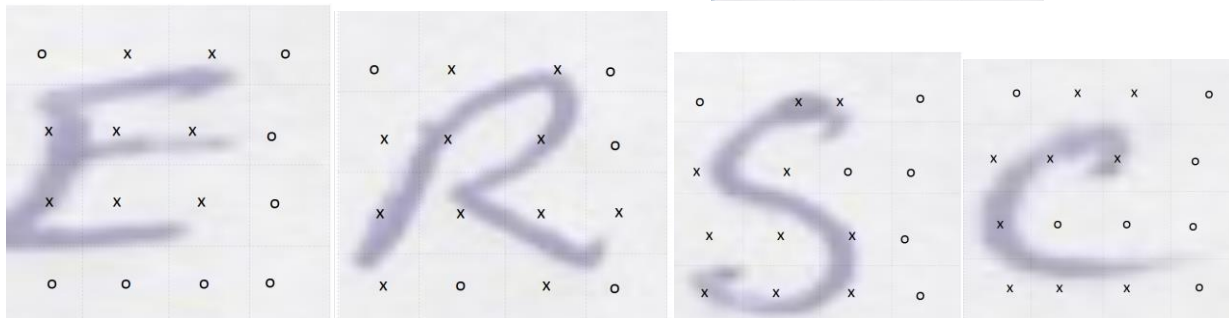
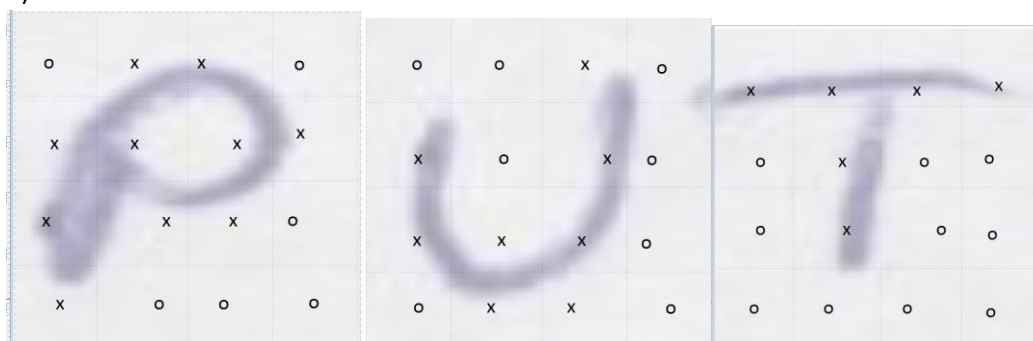
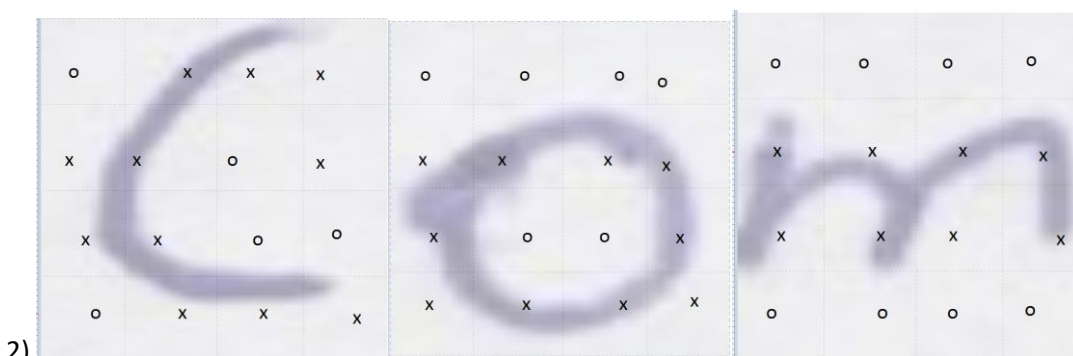
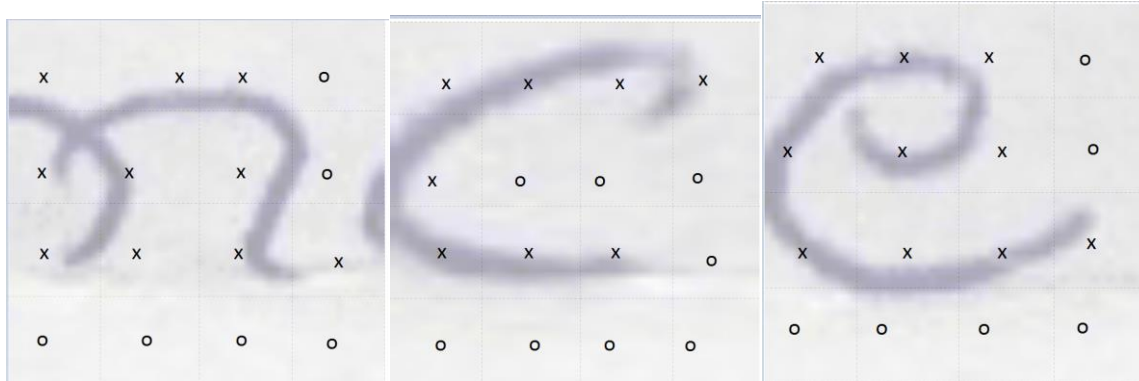


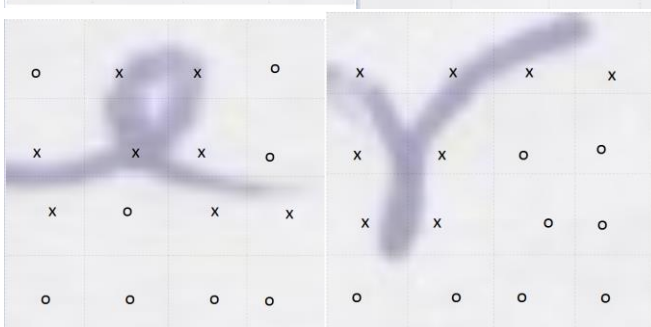
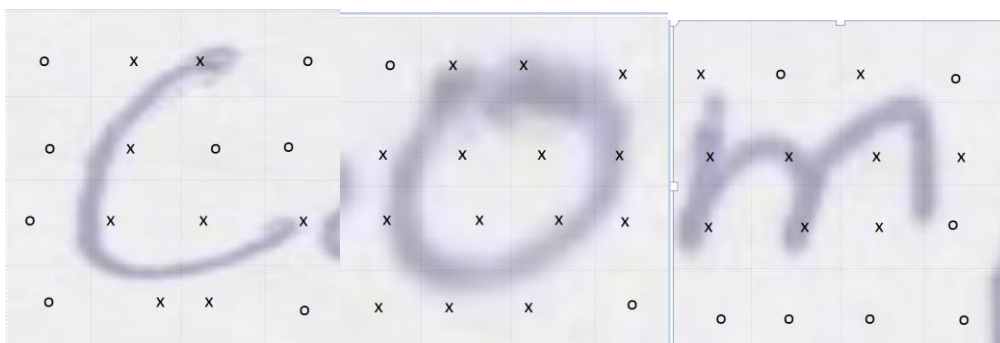
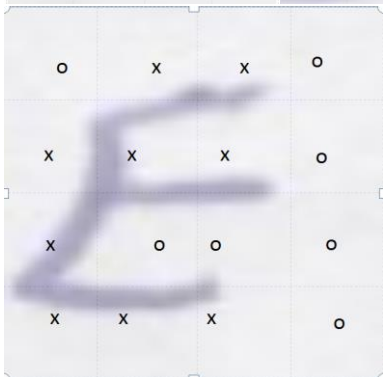
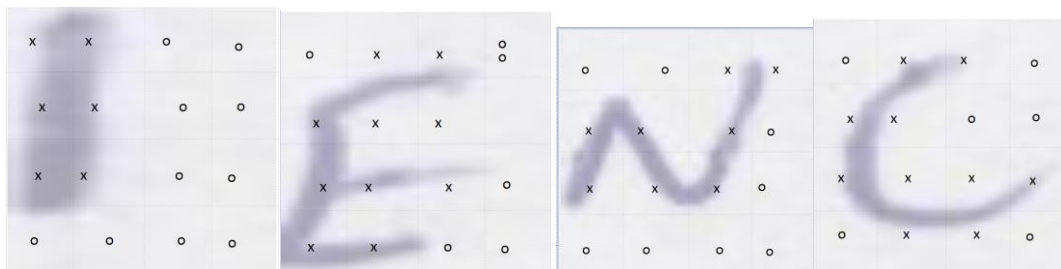
Test Data:

1)

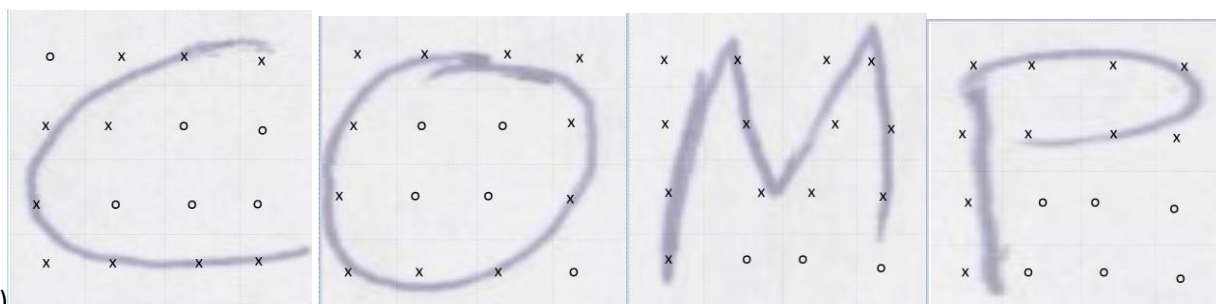
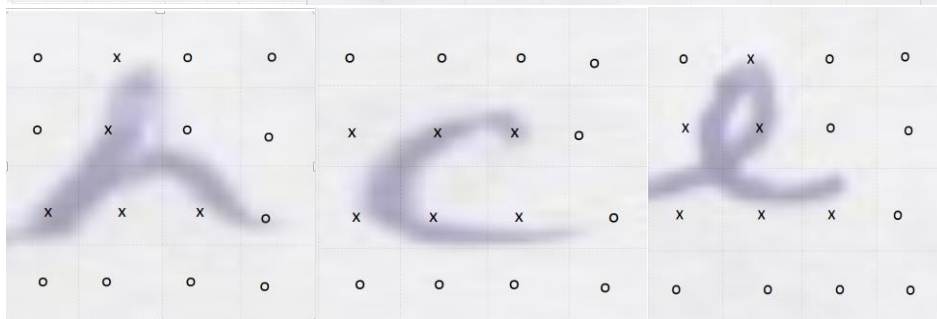
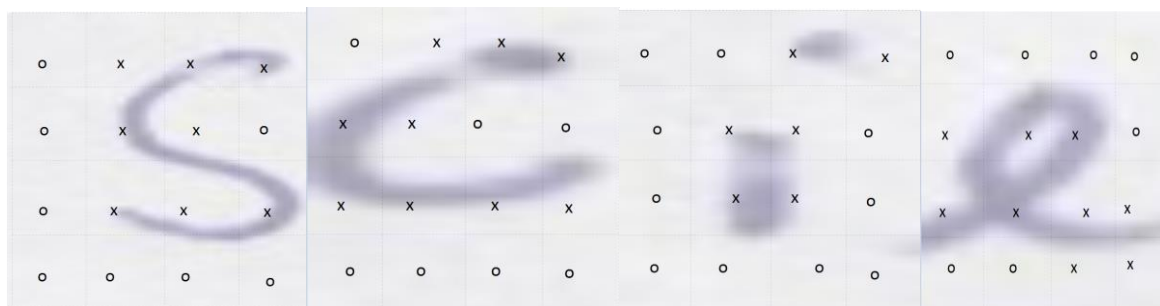




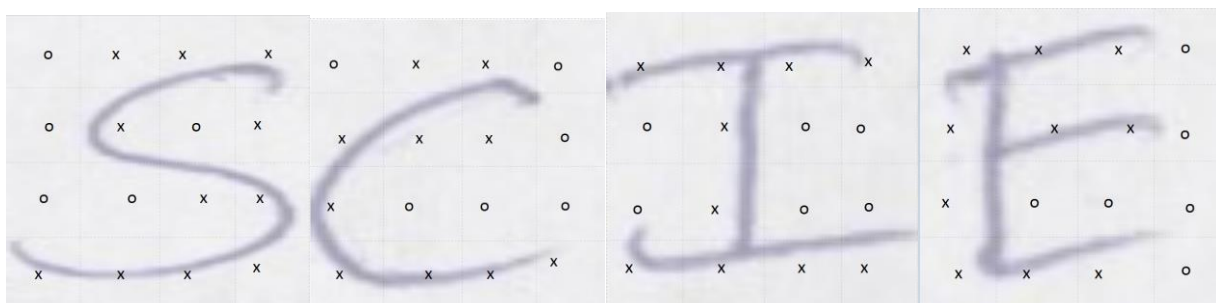
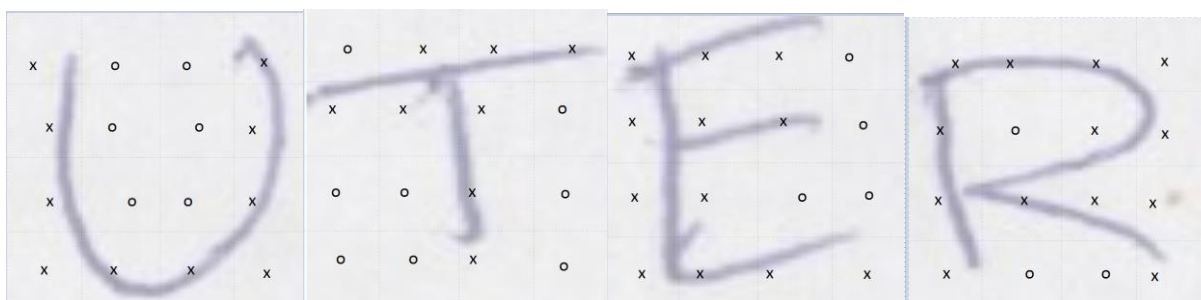




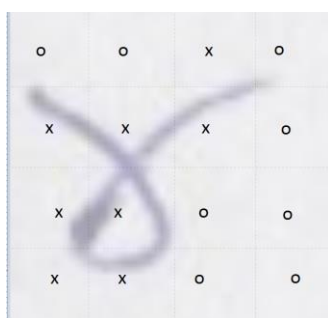
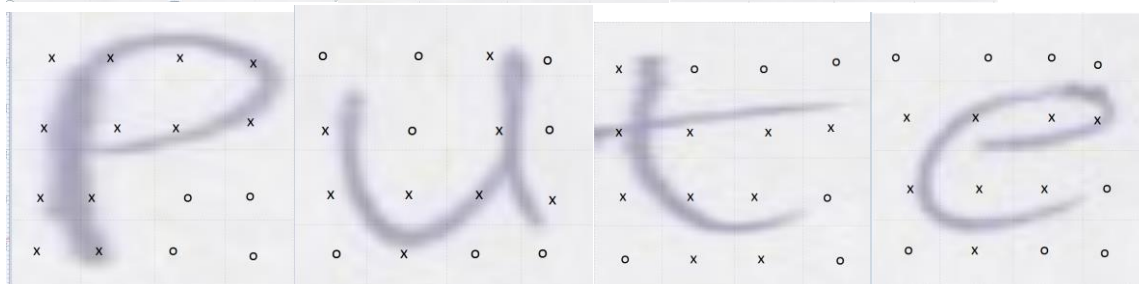
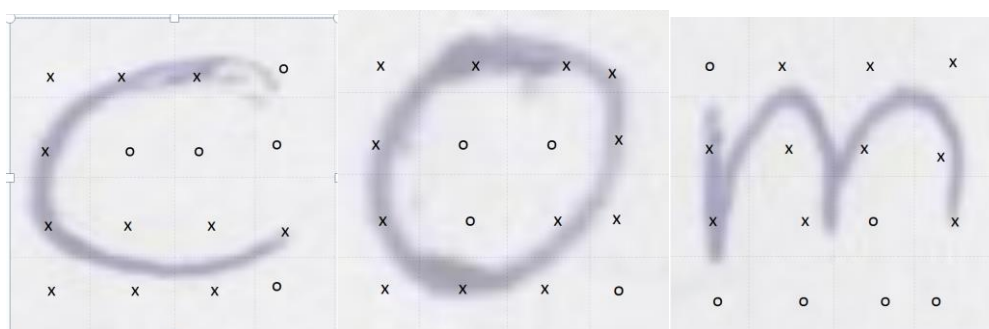
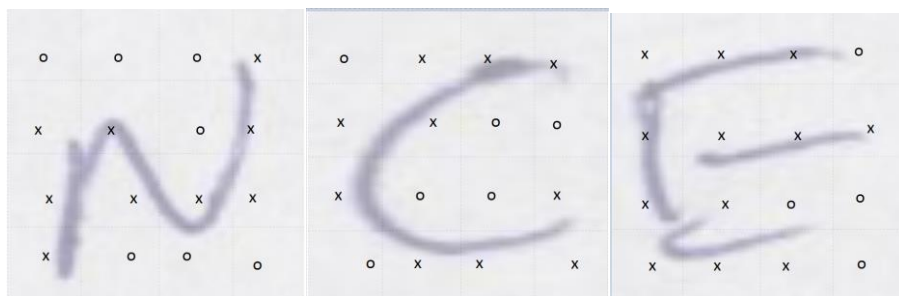


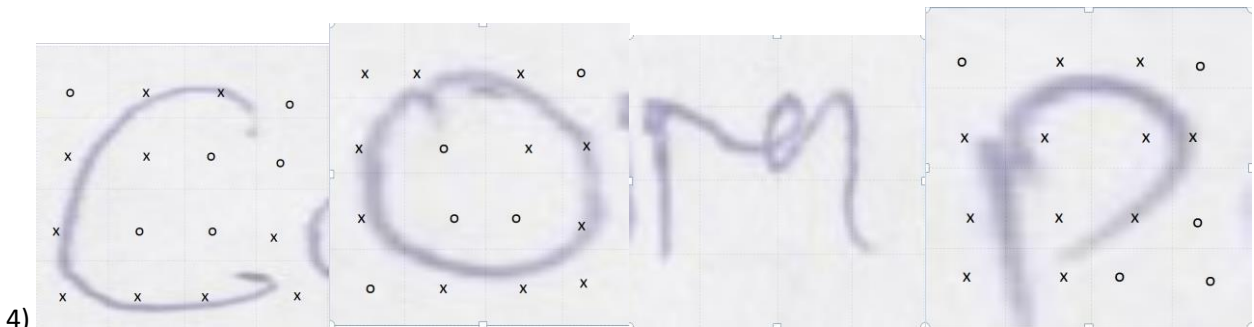
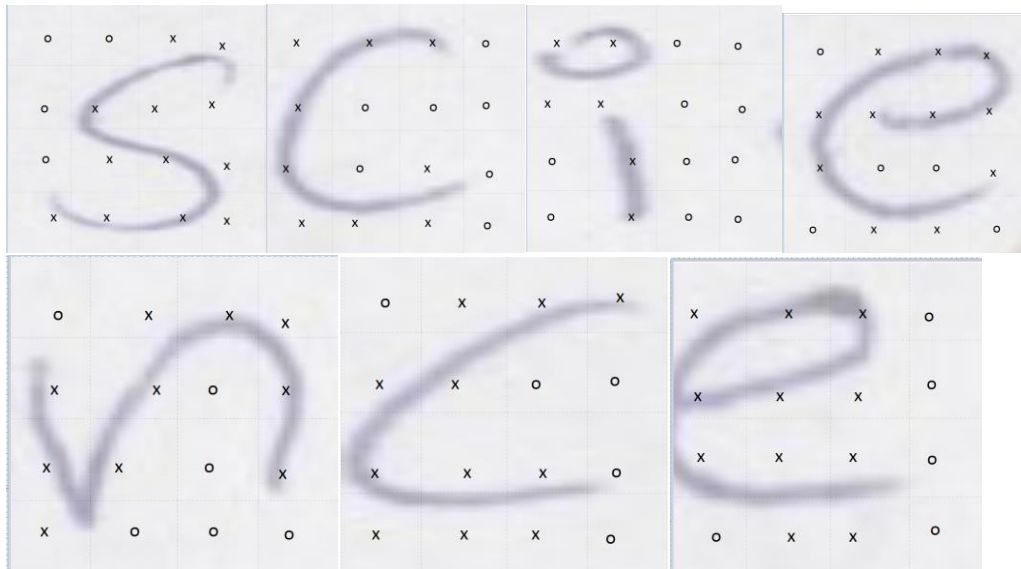


3)

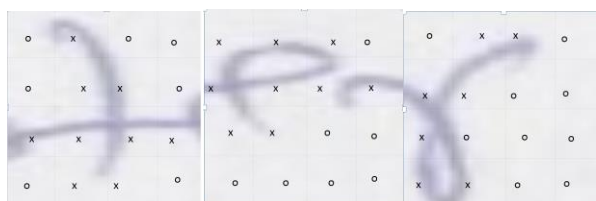
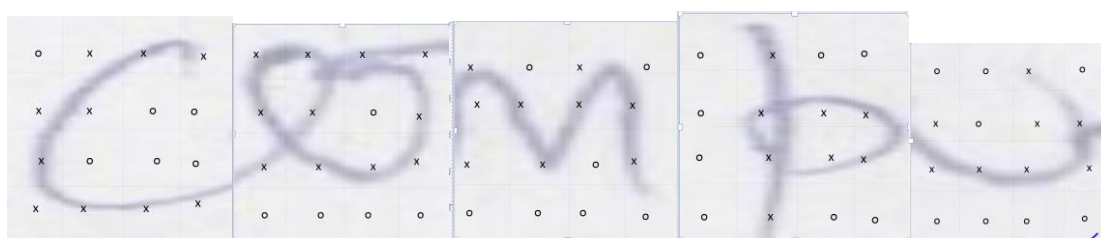
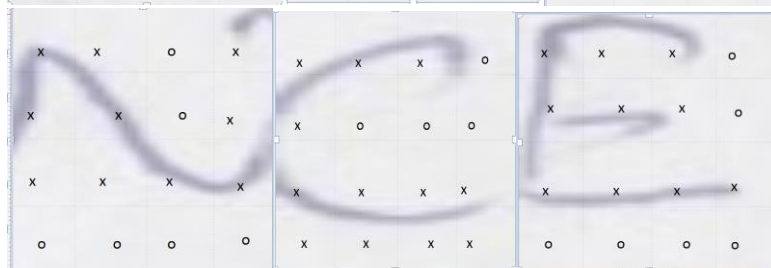
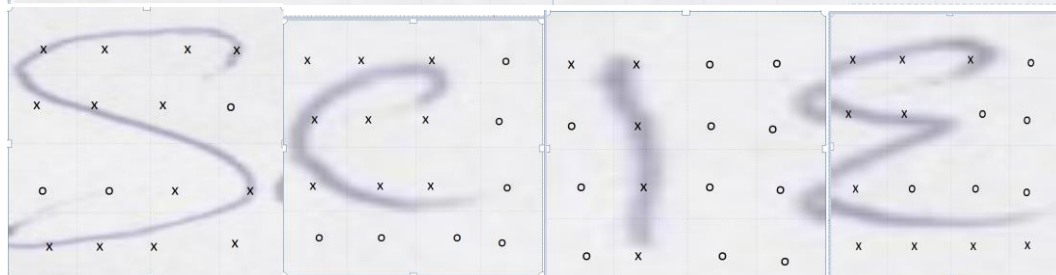
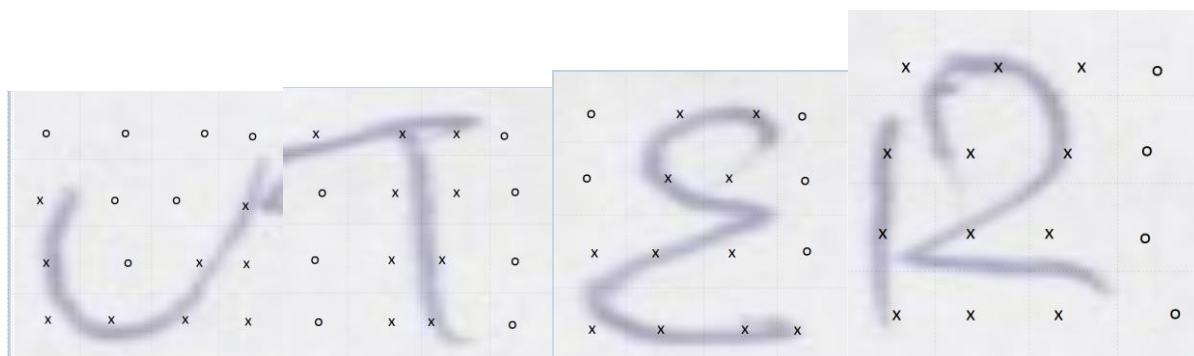


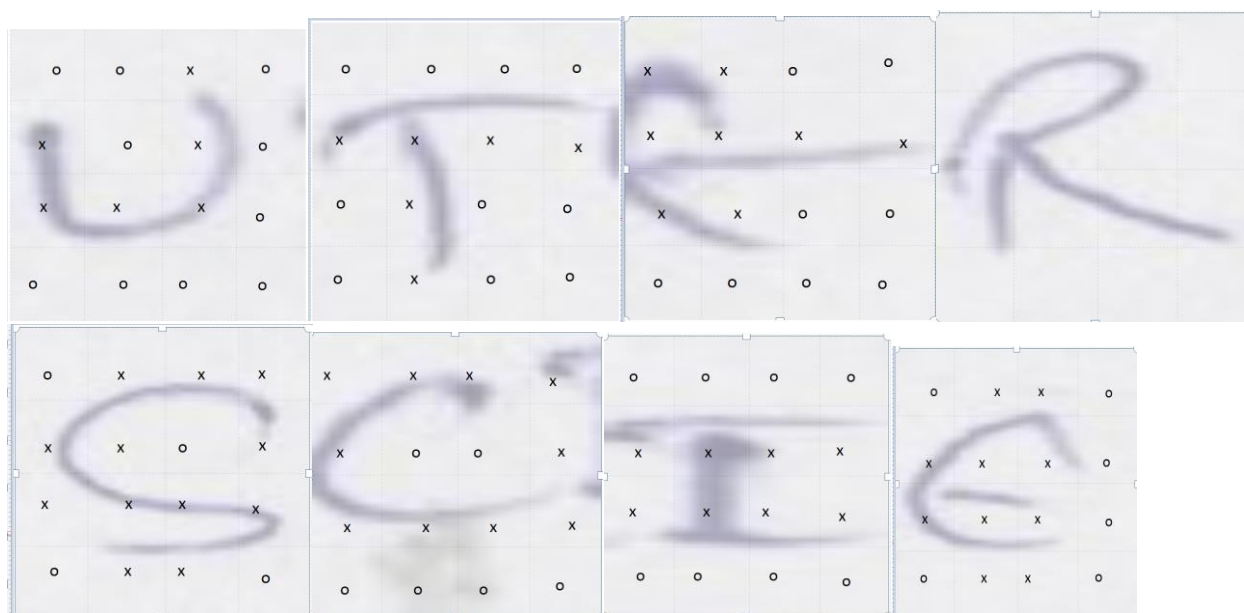
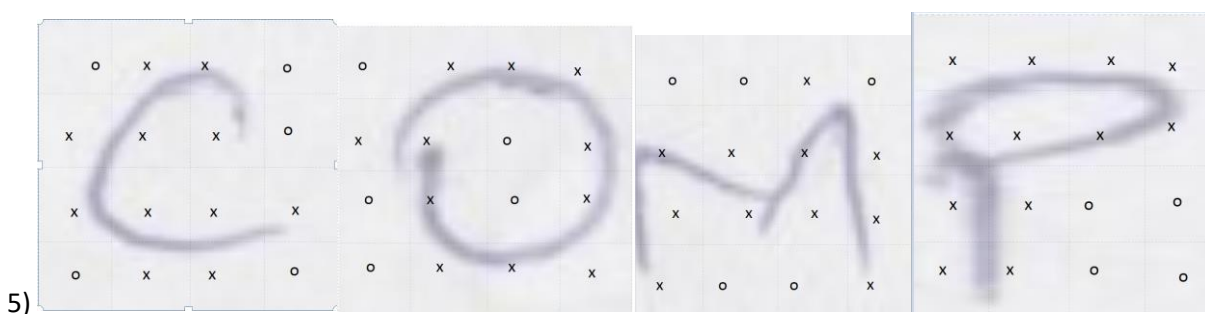
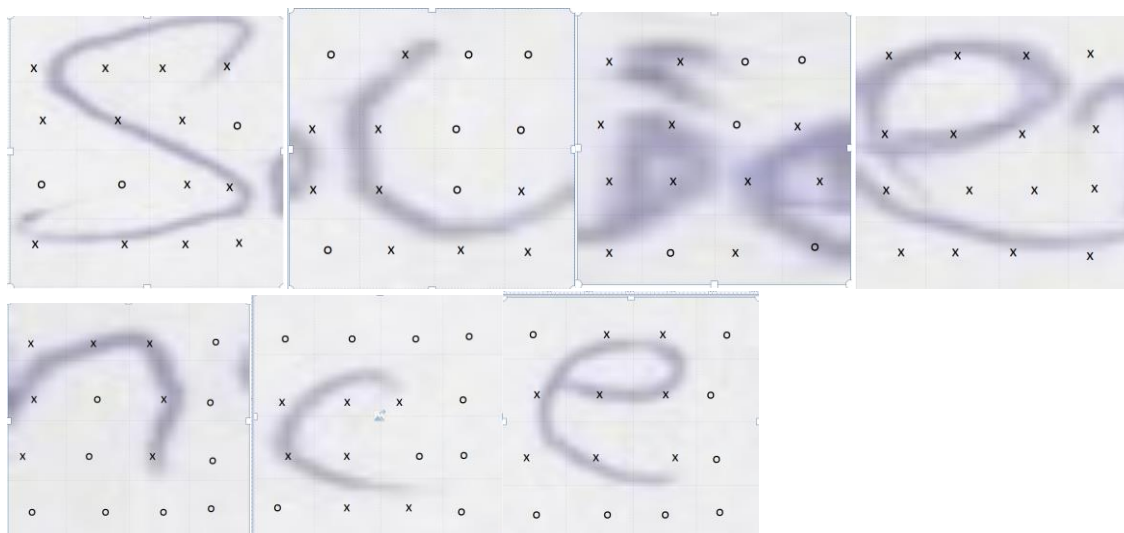


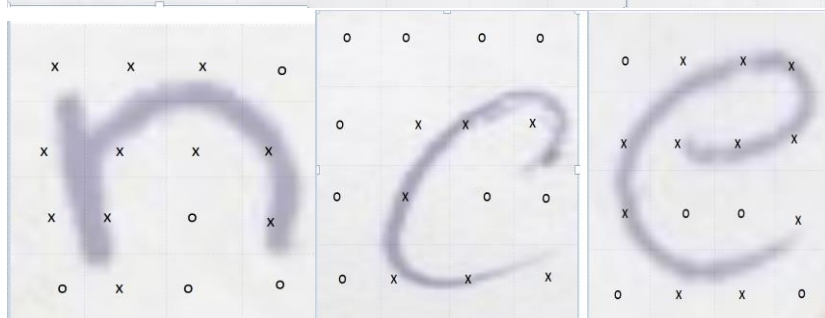
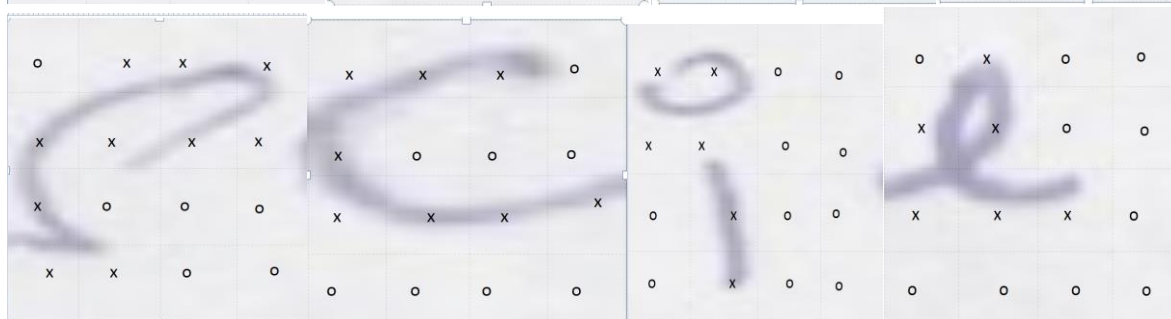
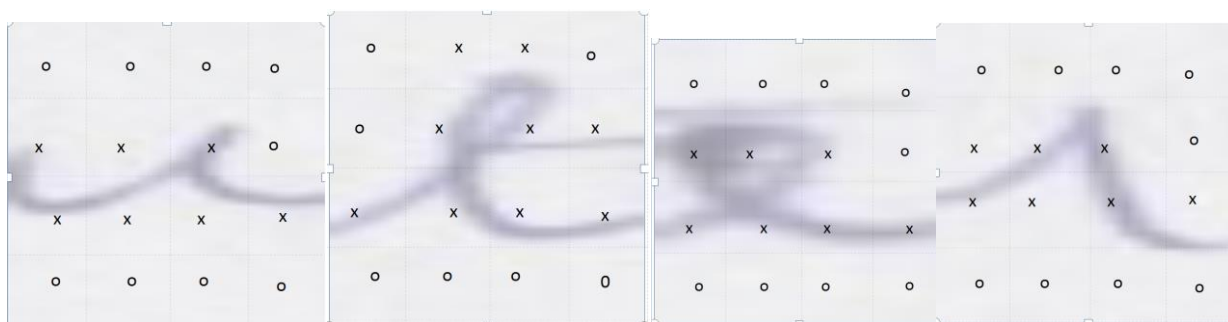
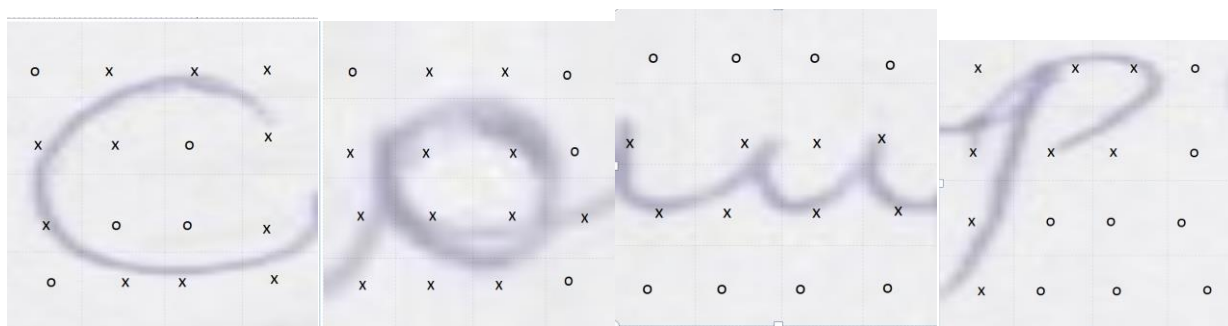




4)









Analysing Data 1:

MUSIC  
SCIENCE

2)

POMP  
OUS

3)

RENT  
rent

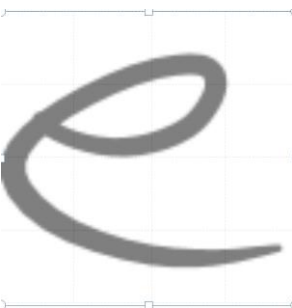
4)

POMP  
OVS

5) 



6) 



Data		Matches in training and test data
M	111111111111001	Training data 2
U	0110011001100110	Training data 4
S	0111011100110111	Test data 3
I	1111010001111110	Training data 3
C	011001001000110	Matches 92 % with training data 2
S	0111111100110110	Matches with test data 5
C	0111110011001110	Training data 3

E	1111111111001111	Matches 90% with training data 4
N	0100111111110011	Matches 95% with training data 1
E	1110111011111111	Matches with test data 2

Accuracy:87.7

Data	Matching data in training and test data
P 0000011101000100	Training data 3
O 000001110111 0000	92% matching with training data 5
M 000011111111 0000	Matches 80% with training data 3
P 000001110100 0100	Matches 78 % with test data 4
O 000011101110 0000	Matches 80% with test data 5
U 000011111110 0110	Matches 90% with training data 3
S 01111110111 1110	Matches 78% with training data 5

Accuracy:85%

Data	Matching data in training and test data
R 0110 0110 0110 0110	Training 82% data 5
E 0000 0110 0110 1100	Matches 90% with training data 4
N 0000 0111 0111 0111	Matches 80% with test data 4
T 0000 1111 0100 0100	Matches 75%with test data 3

Accuracy:79%

	Matching data in training and test data
r 0110 0100 0100 0000	Matches 79% with the training data 3
e 0110 1110 1111 0000	Matches 88% with training data 5
n 0111 0111 0111 0000	matches 92% with test data 4
t 0100 1111 0110 0000	Matches 80% with training data 3

Accuracy:82.3%

	Matching data in training and test data
P 0111 0111 1100 1000	matches with test data 4
O 0110 1111 1011 1110	Matches with training data 5
m 1110 1111 1001 0000	Matches 87% with training data 5
p 0000 1111 1111 1100	Matches with test data 3
O 0111 1101 1111 0110	Matches 90% with training data 4
u 0000 1111 1111 0110	Matches 87% with training data 5
s 0011 0110 0111 1100	Matches 80% with test data 3

Accuracy: 79.2

data	Matching data in training and test data
m 1101 1111 1111 0001	Matches 90% with the training data 4
u 0101 1111 1111 1110	Matches 79% with training data 5
s 0011 1110 0010 0110	Matches with test data 1
i 0110 0110 0110 0110	Matches 90% with the test data 3
c 0000 1111 1111 0000	Matches with test data 4
S 0111 0110 0011 0110	Matches 87% with training data 5
c 0000 1111 1111 0000	Matches 88% with test data 3
e 0110 0110 0111 0000	Matches with test data 3
n 0010 0111 0110 0010	Matches 88% with test data 5
e 0110 1110 1111 0111	Matches 78 % with training data 4

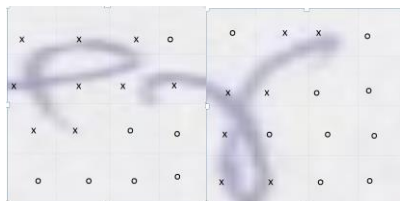
Accuracy : 75.6%

### Analysis:

#### Things contributing to the lower accuracy level:

Well there are many reasons contributing to the low level of accuracy some of the major reasons could be:

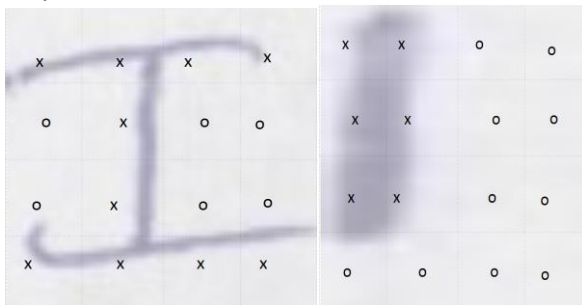
- Inability to completely separate the cursive/ running letters and place them in a grid.Because in the cases where a person writes continuously the letters may overlap in such cases when we place a letter into a grid for encoding some part of the neighboring letters may be accidentally placed inside the grid(like the one shown below) when we use this to train the machine for a particular letter it may result in poor accuracy.



- Alignment Of the letters:
- The alignment of the letters while writing also plays a huge role in determining the accuracy of the machine (e.g)Some people will write in a slanting manner in which the letters lean towards to a particular side.
- In some cases alignment of the paper while writing or scanning also play a huge role in training the data.



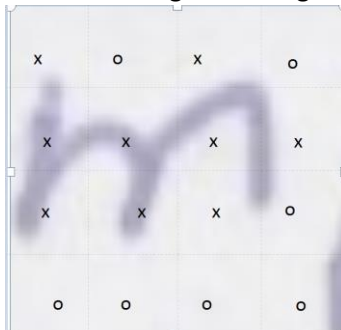
- The type of paper (ruled or unruled) does also matters since people tend to write in slanting manner in an unruled paper than in a ruled one.
- Since we manually stretch each letters to fit into the grid the stretching may contribute to the error rate since a letter may be stretched in either ways and if we choose different ways to stretch each letter then the resulting training model may be trained in different ways for the same letter and thus costs the accuracy of the machine.
- Different people have different handwriting styles.EG: the letter I can be written in two different ways



#### Methods to improve:

##### Introduce a proper preprocessing method:

A proper pre processing method must be introduced inorder to remove the irrelevant information.eg: removing any discontinuous characters in a grid



In the above example there is a junk in the bottom right corner of the grid.it must be removed.However this might not be useful in all cases since in some cases the discontinuous characters may help in defining a letter as shown in the example below



##### Correcting the alignment of slant letters:

We can improve the accuracy by drawing an imaginary line on the input sample and tilting the letters which fall away from the line. However this might not be the best approach in case of small letters.however some people's slant writing may even vary in same word thus making it difficult to change the slant angle even within human intervention.

