**HANDWRITTEN TEXT RECOGNITION**

**OCR (optical character recognition)**

OCR (optical character recognition) is the recognition of printed or written [text](https://whatis.techtarget.com/definition/text) [character](https://whatis.techtarget.com/definition/character)s by a computer. This involves photo scanning of the text character-by-character, analysis of the scanned-in image, and then translation of the character image into character codes, such as ASCII, commonly used in data processing.

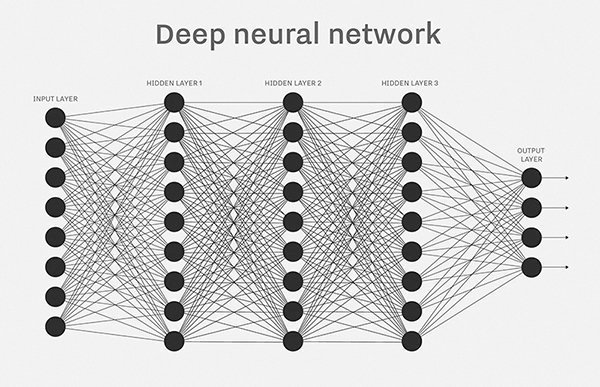
Types:

* **Optical character recognition (OCR)** – targets typewritten text, one [glyph](https://en.wikipedia.org/wiki/Glyph) or [character](https://en.wikipedia.org/wiki/Character_(symbol)) at a time.
* **Optical word recognition** – targets typewritten text, one word at a time (for languages that use a [space](https://en.wikipedia.org/wiki/Space_(punctuation)) as a [word divider](https://en.wikipedia.org/wiki/Word_divider)). (Usually just called "OCR".)
* [**Intelligent character recognition**](https://en.wikipedia.org/wiki/Intelligent_character_recognition)**(ICR)** – also targets handwritten [printscript](https://en.wikipedia.org/wiki/Printscript" \o "Printscript) or [cursive](https://en.wikipedia.org/wiki/Cursive) text one glyph or character at a time, usually involving [machine learning](https://en.wikipedia.org/wiki/Machine_learning).
* [**Intelligent word recognition**](https://en.wikipedia.org/wiki/Intelligent_word_recognition)**(IWR)** – also targets handwritten [printscript](https://en.wikipedia.org/wiki/Printscript" \o "Printscript) or [cursive](https://en.wikipedia.org/wiki/Cursive) text, one word at a time. This is especially useful for languages where glyphs are not separated in cursive script.

**Neural network**

A neural network is a system of hardware and/or software patterned after the operation of neurons in the human brain. Neural networks -- also called artificial neural networks -- are a variety of [deep learning](https://searchenterpriseai.techtarget.com/definition/deep-learning-deep-neural-network) technologies.

The network is composed of a large number of highly interconnected processing elements (neurons) working in parallel to solve a specific problem. Neural networks learn by example. They cannot be programmed to perform a specific task. The examples must be selected carefully otherwise useful time is wasted or even worse the network might be functioning incorrectly. The problem is that there is no way of knowing if the system is faulty or not, unless an error occurs.



**Handwritten text recognition using java**

Step1:

Install Java JDK1.8 ->Make sure JDK should have AWT or Swing, Need OCR.jar file for running this code.

Step 2:

Need training set to recognize character (I took A to Z) -> Save following byte format as “sample.dat”.

A:00110001100111001010111111100110001

B:11111100011000111111100111000110111

C:11111100001000010000100001100001111

D:11111100011000110000100011000111111

E:11111100001000011111100001000011111

F:11111100001000011110100001000010000

G:01110110001000010111100011000111111

H:10001100001000111001111111000110001

I:11111001000010000100001000010000111

J:11111001000010000100101001010011100

K:10001100111111011010100101001110011

L:10000100001000010000100001000011111

M:11111101111011110100101001010110101

N:11111110111000110001100001000110001

O:11111100011000110000100011000111111

P:11111100011001111110100001000010000

Q:01111110011000110001100111101101111

R:11111100011000111011111101001110001

S:01111110001100001111000010000111111

T:11111001000010000100001000010000100

U:10001100001000110001100011001111110

V:10001100011101101011011100111000110

W:10101101011010110101101011011111111

X:10011110100111001100111001011010010

Y:10001110110111001100010000100001000

Z:11111000110011001100110001000011111

Step3:

Write java code for train the sample characters (the set of characters that were inputted by the user). This code includes input, output, classification and training set count.

Step4:

Write java code to provide a small component that the user can draw handwritten letters into.

This class also contains the routines necessary to crop and down sample the written character.

Step5:

Write java code for displaying down sample image. This code used to draw a sample image of a character.

|  |
| --- |
|  |

Step6:

Write java code for classifying the input character based on trained set. Network and, Kohonen Network used to find the weights of the output neurons base on the input from the input neurons.

Kohonen Network takes input as 3 parameters given as follow,

* inputCount Number of input neurons
* outputCount Number of output neurons
* Owner The owner object, for updates. - This @param sends to Main entry java file.

Step7:

Write a java code to load sample data and train the data for testing.

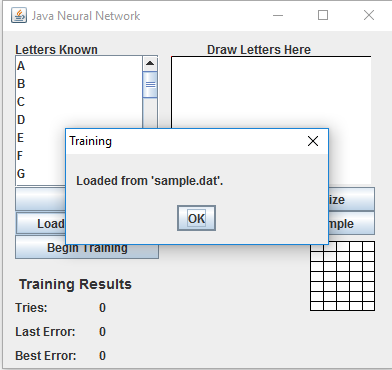
Step 8:

Run –Method:

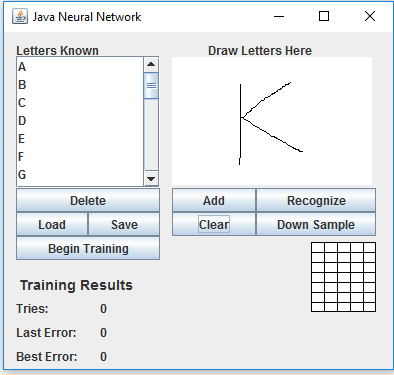
* Run MainEntry java file.
* Click Load button - Sample data loading into panel.
* Click Begin training button for train the sample data.
* Draw the Letter into “Draw Letters here” panel.
* Click Recognize button for finding Handwritten text.
* If you want to add new letters to Sample data, Click add button and add necessary character. -> Click save button to save new character into sample data. - > Again have to Click Begin training button for train the sample data.

Demo:

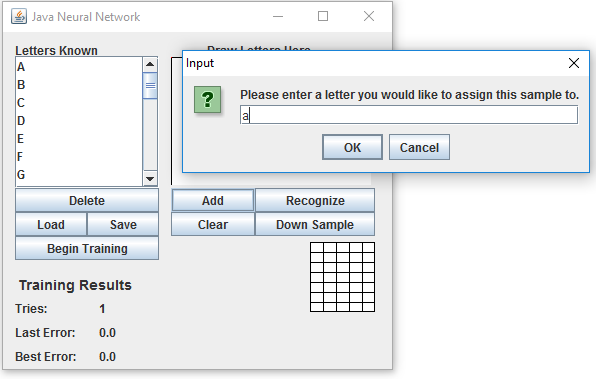
1. Load sample data



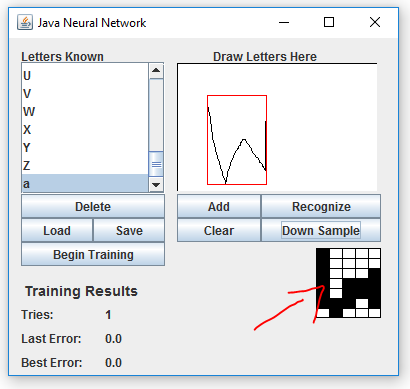
2. Draw letter to recognize



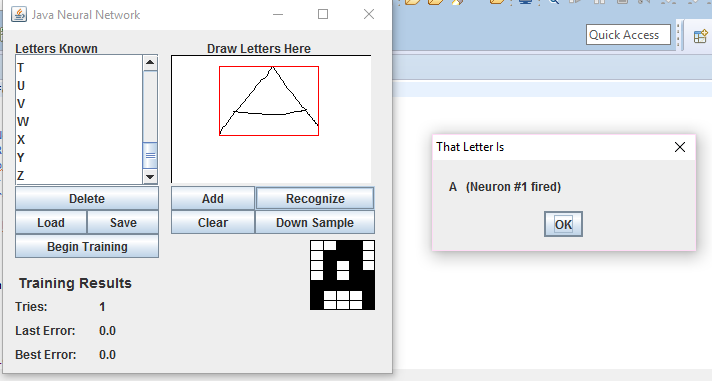
3. Add new letter if you want

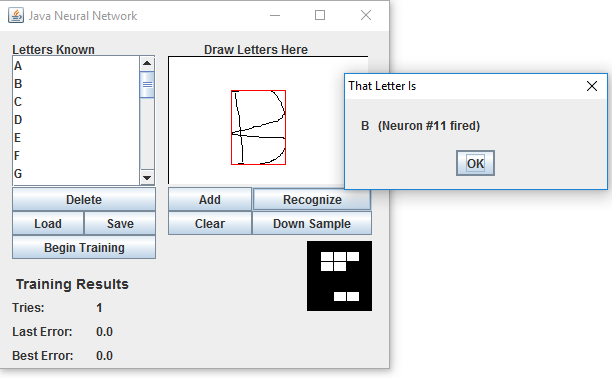


4. Check downsample image using Downsample button.



5. Recognize input letter





References:

<https://www.youtube.com/watch?v=aircAruvnKk>

<http://www2.sys-con.com/itsg/virtualcd/java/archives/0705/heaton/index.html>