ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 4, Issue 6, June 2015**

# Neural Network Used For Translating the Speech to Devanagari Text Using MATLAB

Pooja Mittal [1], A.P Kapil sachdeva[2], A.P Lovely Dhawan[3]

P.G. Student, Department of ECE, IIET, Kinana, Jind Affiliated with Kurukshetra University, India[1]

Assistant Professor, Department of ECE, IIET, Kinana, Jind Affiliated with Kurukshetra University, India[2]

Assistant Professor, Department of ECE, IIET, Kinana, Jind, Affiliated with Kurukshetra University, India[3]

**ABSTRACT:** in this paper is presented an investigation of the speech recognition system that can recognize Hindi words. This paper takes a brief look at the basic building block of a speech recognition engine. It talks about implementation of different types of modules. Sound Recorder, Feature Extractor, MFCC and neural network training have been described in details. In real life, this speech recognition technology might be used in voice banking services, in medical documentation, in military etc.

**KEYWORDS:** speech recognition, MFCC, Neural network, feed forward neural network, UTF-8 UNICODE

## I. INTRODUCTION

In this far reaching hi-tech global world, development in science has stretched out its tentacles up to that optimum point where all back breaking task of human being has been completely replaced by Machines. The expeditious growth of technology in recent decades has changed the whole dimension of communication. Today people are more interested in hands free communication. [1] Considering this situation, intelligent communication between computers and humans, becomes one of the most important research fields.

Speech is one of the natural forms of human communication and it could be a useful interface to interact with machine. Unfortunately, the speech recognition process is still facing a lot of problems.

In general some of the most common are:

- *Background Noise*: a noisy environment can corrupt the signal and even the speaker himself can add noise by the way he speaks.
- *Speaker variations*: to pronounce the same word at a different time even by same the speaker is really very difficult due to dialect variation, speed of speech and emotional condition of the speaker.
- *Unsystematic break during continuous character of speech:* when we speak long sentences, we use break between words according to our choice. This makes it very hard to detect individual words.
- *Other external factors:* position and direction of the microphone with respect to speaker and many others.

**Artificial Neural Networks (ANN)** is composed of simple computational elements operating in parallel and one of the original aims was to understand and shape the functional characteristics of the brain [3]. To uniquely recognize the speech, we can train the neural network, so that particular input leads to a specific target output by comparing it against a database. In this paper we discuss the usability of Feed-Forward back- propagation neural network for speech recognition using MATLAB.

In the next chapter, a general introduction to speech recognition and its basic ideas, problems and challenges will be given. In the third chapter we focus on extracting the relevant information from the speech. The fourth chapter cover the implementation of the neural network. The conclusion remarks are given in the last chapter of the paper.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 4, Issue 6, June 2015**

## II. LITERATURE SURVEY

For a long time research on how to improve this type of communication has been done. Even in 18th century people were experimenting on speech synthesis. For example, in the late 18th century, Von Kempelen developed a machine capable of 'speaking' words and phrases and trying to create an environment in which computer can learn from the user by observing his gestures, emotions and speech instead of one where user must learn how to use the computer. Nowadays, thanks to the evolution of computational power it has become possible not only to develop, test and implement speech recognition systems, but also to have systems capable to **real-time conversion of speech into text [**2]. Unfortunately, despite the good progress made on that field, the speech recognition process is still facing a lot of problems, with most of them contributed to the fact that speech is a very subjective phenomenon.

## III. GENERAL LAYOUT AND ISSUES OF SPEECH RECOGNITION PROCESS

Speech recognition is a technology that instead of typing the keyboard or operating the buttons for the system or it can be controlled by speech itself for the better convenience. The speech recognition process can generally be divided in many different components illustrated in Fig. 1

- The First block deals with the input, consists of the acoustic environment plus a microphone connection to record the speech signal and represent it in digital form after doing amplification but additional impact of echo and reverberation because of background noise affect the quality of speech sample.


**Fig 1:** Speech Recognition process

- The second part deal with the problem of background noise which affect the quality of speech signal. To reduce the leakage problem due to adjacent frequency components hamming window is used, i.e. its aim to suppress high frequency components before spectral analysis
- The third block must be capable of extracting the relevant information from pre-processed signal that are good for linguistic content and discard all other remaining irrelevant source of variations. This can be done with a variety called as MFCC, Mel Frequency Cepstral Coefficients.
- The fourth one presented as systems of interconnected neurons which can compute values from inputs and create the network for pattern recognition. Once the network is created it can be trained for a specific problem by presenting training inputs and their corresponding targets.
- The last, fifth block tries to relate the input sound to the best fitting sound in a known 'vocabulary set' of 10 words. By using "Devanagari New Normal" font this represent as a final output which can be displayed in Hindi language on command window in MATLAB.

Hence, speech Recognition is trying to understand the words being spoken and established a similarity to the human speech communication system.

## IV. THE SPEECH SIGNAL

Vocalized form of human communication is called as **speech** and its aim is to transfer the ideas. Speech is made within the speaker's brain and then, the source word sequence is performed to be delivered through her/his text generator, Travels through the Vocal tract, and Produced at speakers mouth and Gets to the listeners ear as a pressure wave [4] as shown in figure 2.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 4, Issue 6, June 2015**

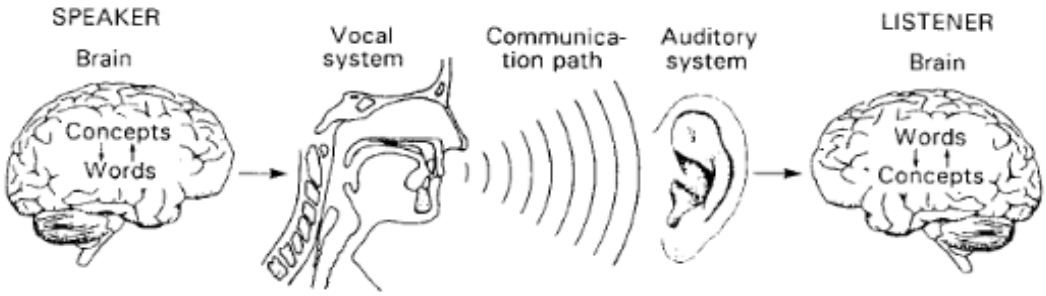


**Fig 2:** Human speech communication [5]

## V. SIGNAL CAPTURING

Signal capturing is the first step that needs to be taken for implementing the speech recognition system in MATLAB. We know that the signal from a human is an analog signal. When storing the speech to a computer, the analog signal needs to be digitized. Many data acquisition hardware devices contain one or more subsystems that convert (digitize) real-world sensor signals into numbers as shown in figure 3:

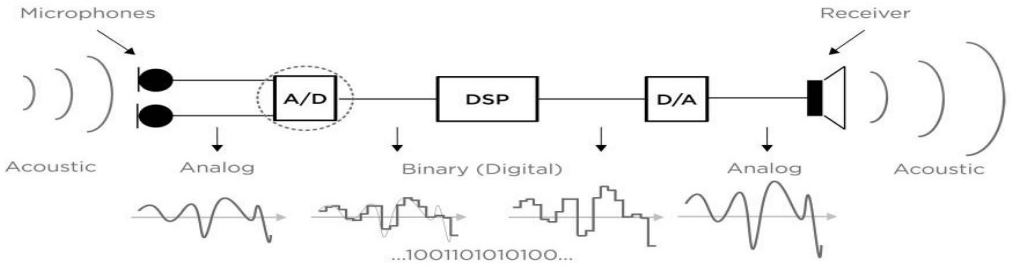


**Fig 3:** working of sound signal [6]

When people talk to a microphone, then analog signal pressurizes the air, analog to digital conversion takes place. After the real-world signal is digitized, then signals can be analyzed, stored in system memory, or store to a disk file...

### *(a) Implementation of signal Preprocessing:*

We know that audio signals are quasi periodic i.e. these are periodic on a small scale but unpredictable at some large scale as signals properties changes quite rapidly over time.

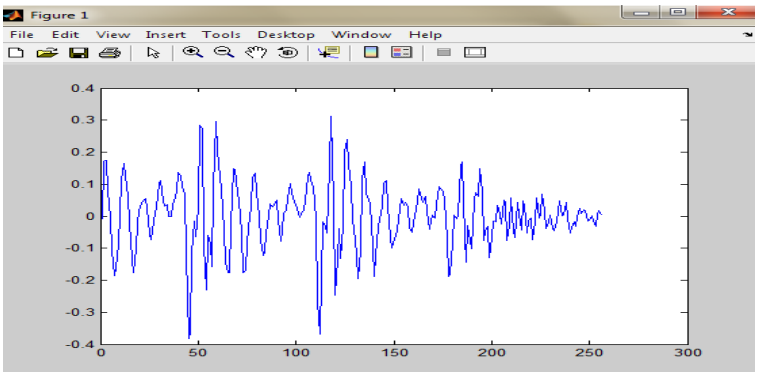


**Fig 4:** speech is divided into frames

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 4, Issue 6, June 2015**

Therefore, most of the speech signal remains invariant for a short period of time by taking short windows. These short windows of signal like this called as "FRAMES" as shown in figure 4.

To analyze how much of the signal lies within each given frequency band, Fourier transform is applied on each frame separately. Therefore, we need to perform signal pre-processing to extract only the speech related information to feed the neural network with normalized input as shown in figure 5:

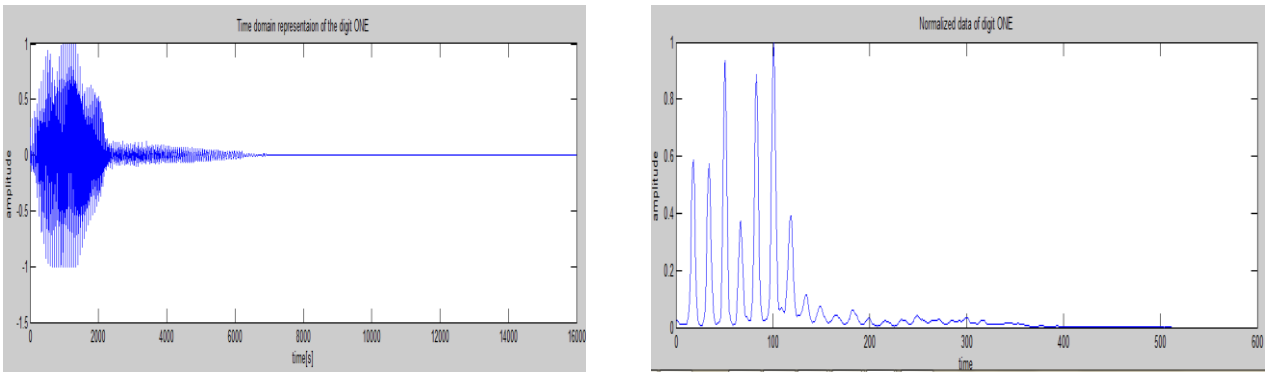


**Fig 5:** (a) Representation Of speech (b) Normalized Input

## VI. FEATURE EXTRACTION

The sounds generated by a human are filtered by the shape of the vocal tract including tongue, teeth, lungs etc. This shape determines what sound comes out. If we can determine the shape accurately, this should give us an accurate representation of the phoneme being produced [7]. MFCCs is a good idea for accurately represent this envelope. So, the development includes an extensive study of MFCC (Mel Frequency Cepstral Coefficients) which mimic the functionality of human ear. Following is the block diagram to calculate the MFCC of a speech signal as shown in figure 6:

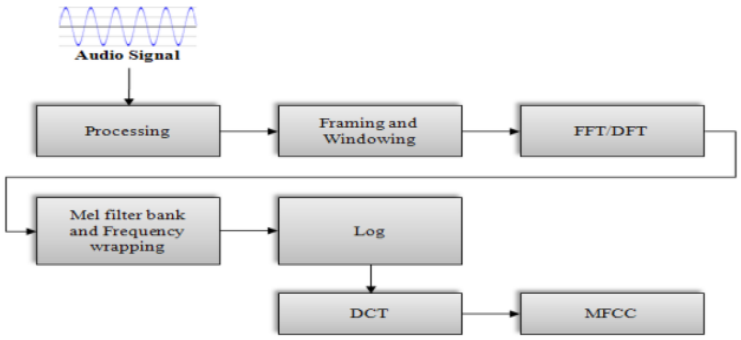


**Fig 6:** Block Diagram to calculate MFCC [8]

We know that human ear concentrate on certain region rather than using whole spectral envelope so there is need to use a scale which establish a relation between linear and logarithm and this condition is fulfilled by MFCC and extract only 20 coefficients from the spoken words which are sufficient to match the spoken words from the database. The following approximate formula is used to compute the mels for a given frequency in Hz:

$$B(f) = 2595 \log_{10}(1 + {}^{f}/_{700})$$

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 4, Issue 6, June 2015**

Mel represents the pitch (perceived frequency) of a tone as a function of its acoustics frequency and one mel is defined as one thousandth of the pitch of a 1 kHz tone. This is the inverse function of above equation:

$$f = 700\left(10^{f_{mel}/2595} - 1\right)$$

Therefore goal of feature extraction is to transform the high dimensional speech signal to relatively low dimensional features while preserving the speaker discriminative information as shown in figure 7.

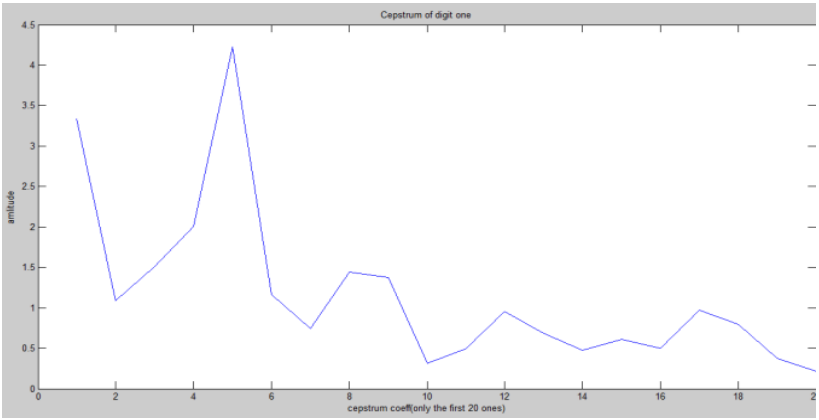


**Fig 7:** Cepstrum of digit one

## VII. FEED-FORWARD NEURAL NETWROK IMPLEMENTATION

Many authors used neural networks for speech recognition in the past. MATLAB Neural Network toolbox has been used to create, train and simulate the network. Commonly neural networks are adjusted, or trained, so that a particular input leads to a specific target output. Such a situation is shown below.
There, the network is adjusted, based on a comparison of the output and the target, until the network output matches the target. Typically many such input/target pairs are needed to train a network.

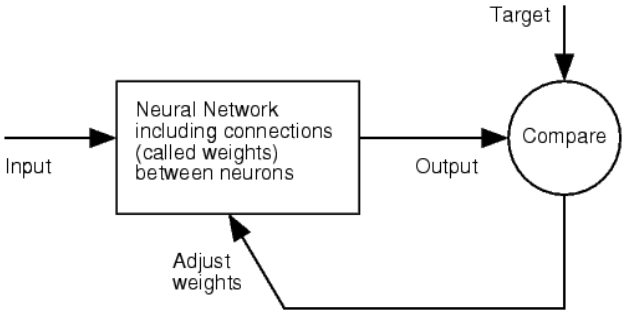


**Fig 8:** neural Network architecture

Today neural network have been trained to perform complex function in various fields, including, pattern recognition, identification and control system.

### (a) *Multi Layer Feed Forward Neural Network:*

Neural networks are good at fitting functions and recognizing patterns. The feed forward neural network was the first and simplest type of artificial neural network. In this network, the information moves in only one direction, forward,

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 4, Issue 6, June 2015**

from the input nodes, through the hidden nodes (if any) and to the output node[9] as shown in figure 9: After feature extraction we left with only twenty values to represent a digit, then a set of a Mel frequency Cepstrum Coefficients of 10 digits act as input for the neural network which produce a single output.
Usually the sum of each node are weighted and passed through a non-linear function known as a transfer function or activation function. Therefore, the choice of transfer functions in neural networks is of crucial importance to their performance. In this paper a database of one to ten digits is created and 70% data is used for training while other 30% is used to test the network (as these not included in the training set) as shown in figure 10:

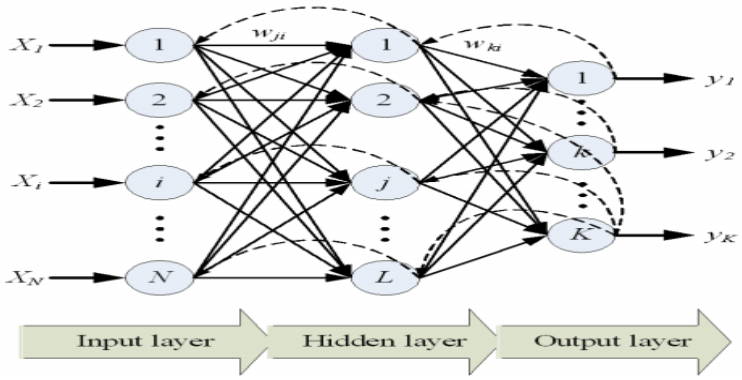


**Fig 9:** Back propagation ANN [10]

Log-sigmoid-Transfer Function commonly used in back propagation networks, as it is differentiable and its input values lies in the range of -5 up to 1.5. But the trained network can also be tested with real time input from a microphone.
The hidden layer consist this activation function using the **'tansig'** MATLAB NN Toolbox function. The choice of hidden layers depends on some factors like the amount of input and output layer neuron number, the capacity of the network and the size of the training set.

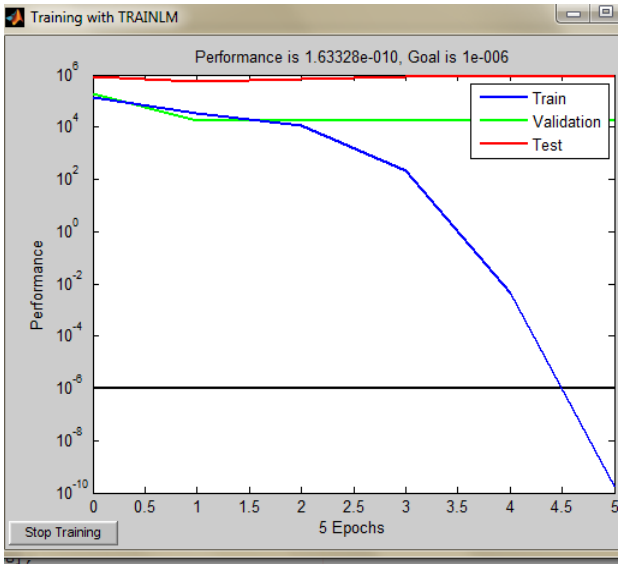


**Fig 10:** Training the feed-forward back-propagation network

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 4, Issue 6, June 2015**

There are some rules of thumb for tuning the number of hidden layer neurons for a neural network:

- The number of hidden neurons should be less than twice the size of the input layer.
- The number of hidden neurons should be 2/3 the size of the input layer, plus the size of the output layer.
- The number of hidden neurons should be between the size of the input layer and the size of the output layer[11]

The Performance of the network is mainly dependent on the quality of the signal pre-processing. Therefore, increasing the number of hidden layer causes the training time to grow and make the system more complex. The neural network performs very well with MFCCs as input having more than 70% successful classification rate.

## VIII. DISPLAY THE FINAL OUTPUT WITH UTF-8 CODING

We know that computers can only deal with binary numbers. It stores letters and other characters by assigning a unique number for each one. Before Unicode was invented, there were hundreds of different encoding systems but no one contain enough characters for assigning these numbers. Unicode provides a unique number for every character, no matter what the platform, no matter what the program, no matter what the language.
UTF-8 stands for "Universal Character Set Transformation Format-8bit is a character encoding capable of encoding all possible characters [12].
There are so many dialects are present whole over the world but for Indians Hindi is the most majestic and prominent language to express our thoughts, feelings and expression.


**Fig 11:** Finally speech is converted into text

So, with the help of MATLAB (the language of technical computing) we got 70% success to translate the spoken words into "Devanagari" text with the help of UNICODE characters.
Lets suppose digit one is used for comparing against database of 10 digits. 50 upto 100 epoches are enough to train the network sufficently so that network error reaches almost zero. finally the output can be visualised on the screen as shown in figure 12 :. here digit 'one' is written in hindi in MATLAB by training the inputs.

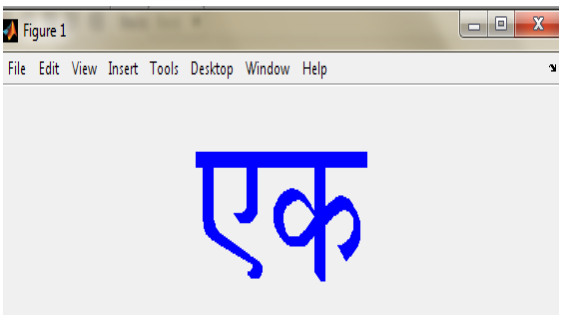


**Fig 12:** Final Output is displayed in MATLAB
**'090F' and '0915' codes are used for letter (E) and ( k).**

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 4, Issue 6, June 2015**

## IX. CONCLUSION

This paper describe the implementation of automatic speech recognition system which can be easily synthesized and recovered from neural network classifier and MFCC is a very reliable tool for signal preprocessing stage. This implementation try to build a text-dependent system or something that could actually determine what word was being spoken to the system by the user and provide significant help for the people with disability and require low operational cost. So in order to improve the designed systems to make it work better, design some analog and digital filters for processing the input signals which can reduce the effects of the input noise and to establish the large data base of the speech signals for different words.

## X. ACKNOWLEDGMENT

"Say a word of gratitude and splendor, in a moment it is gone, but there are a hundred ripples, circling on & on. I place on record, my sincere thanks to Dr. Rohit Garg, Principal of IIET,Kinana for providing me with all the necessary facilities for the research. It is truly a matter of great pleasure me to express my thanks Mr. Kapil Sachdeva for his guidance, understanding, patience, and giving me right direction during my whole research. Finally, and most importantly, I would like to thank my parents for showing the trust on me every time.

## REFERNCES

1. implementation of an acoustic echo canceller using MATLAB pdf
2. http://www.doiserbia.nb.rs/img/doi/1450-9903/2010/1450-99031001001G.pdf
3. http://www.doiserbia.nb.rs/img/doi/1450-9903/2010/1450-99031001001G.pdf
4. http://cs.haifa.ac.il/~nimrod/Compression/Speech/S1Basics2010.pdf
5. http://www.diva-portal.org/smash/get/diva2:347957/FULLTEXT01.pdf
6. https://www.google.co.in/search?q=speech+recognition&biw=1366&bih=643&source=lnms&tbm=isch&sa=X&ei=E3lQVc-nENLt8AXT3IH4Cw&ved=0CAcQ_AUoAg#tbm=isch&q=analog+to+digital+converter+in+speech.
7. http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/
8. https://www.google.co.in/search?q=block+diagram+of+mfcc&biw=1366&bih=643&tbm=isch&tbo=u&source=univ&sa=X&ei=gw5SVafOKYmB8gWS5YDYBA&ved=0CBwQsAQ
9. http://www.doiserbia.nb.rs/img/doi/1450-9903/2010/1450-99031001001G.pdf
10. https://www.google.co.in/search?q=neural+network+architecture&biw=1366&bih=643&tbm=isch&tbo=u&source=univ&sa=X&ei=6w9SVcz9J4XemAWInoG4Ag&sqi=2&ved=0CCcQsAQ
11. http://en.wikipedia.org/wiki/Feedforward_neural_network
12. http://en.wikipedia.org/wiki/UTF-8