

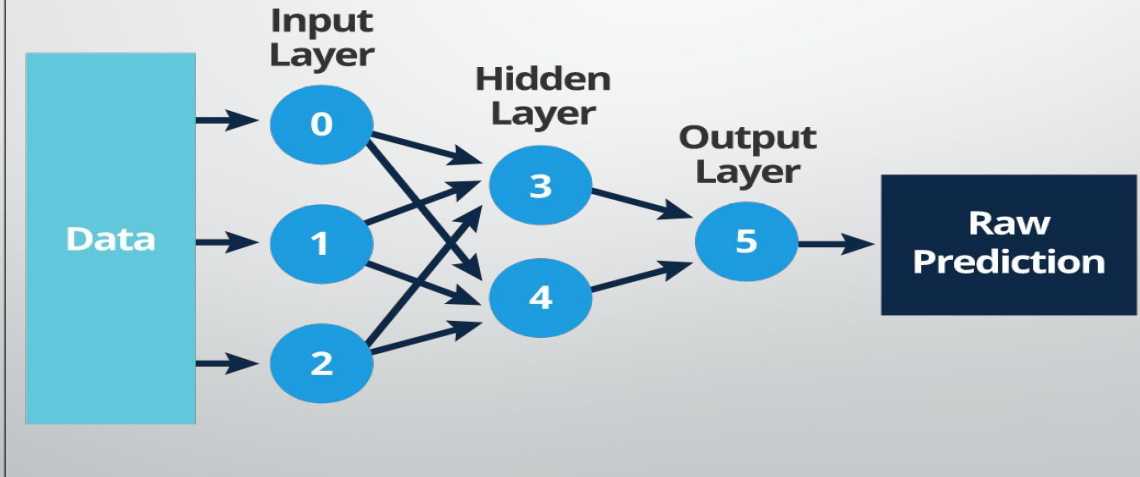


# HDL Course Project

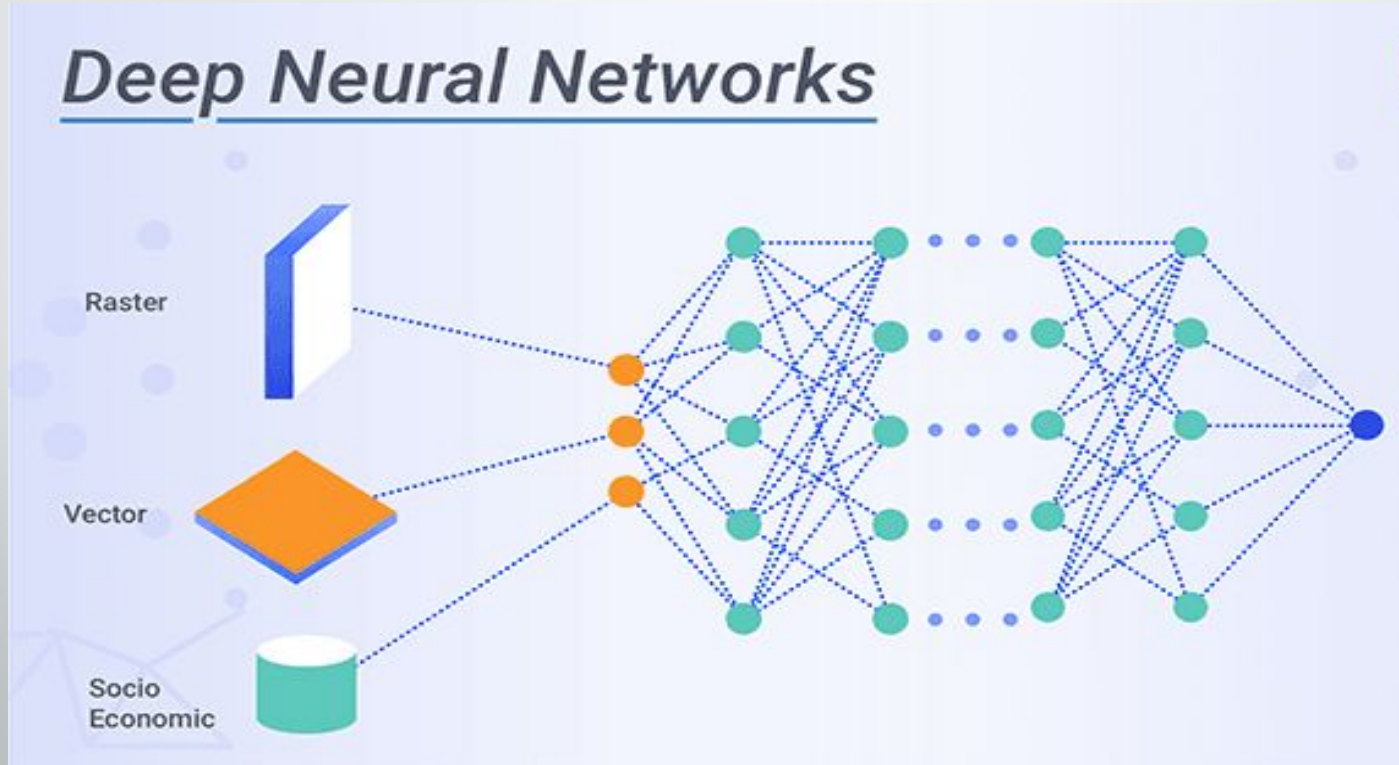
TiM-DNN: Ternary In-Memory Accelerator  
for Deep Neural Networks

Neural Network is a set of algorithms that execute to take simple decisions.

## Neural Networks in Data Mining



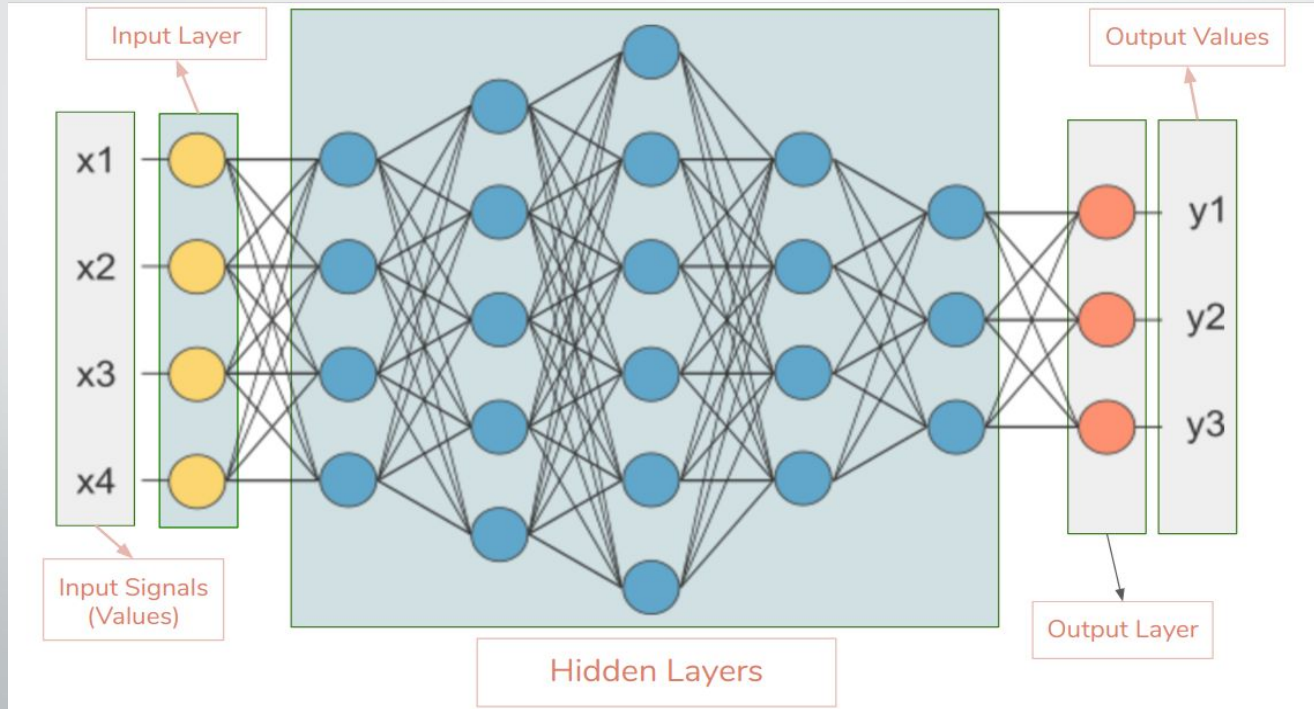
But before that, a bit about DNN



**BUT WHY**

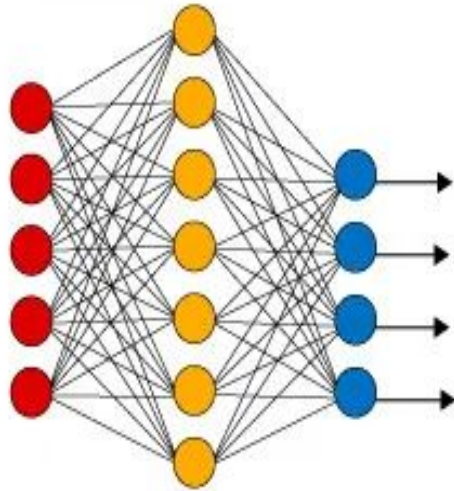
**DEEP NEURAL NETWORKS**

A Deep Neural Network is a Neural Networks that had multiple layers between input and output layers .The more layers there are the deeper the network is.

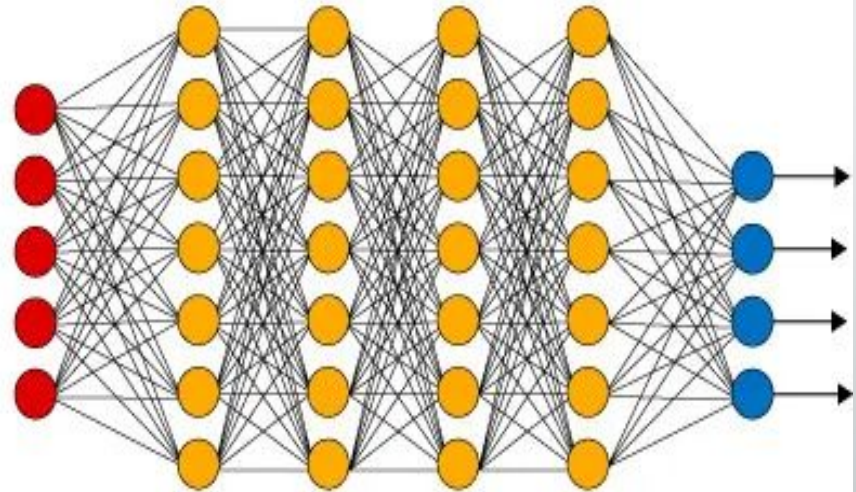


The main difference between NN and DNN is that, DNN can learn on its own and NN cannot.

**Simple Neural Network**



**Deep Learning Neural Network**



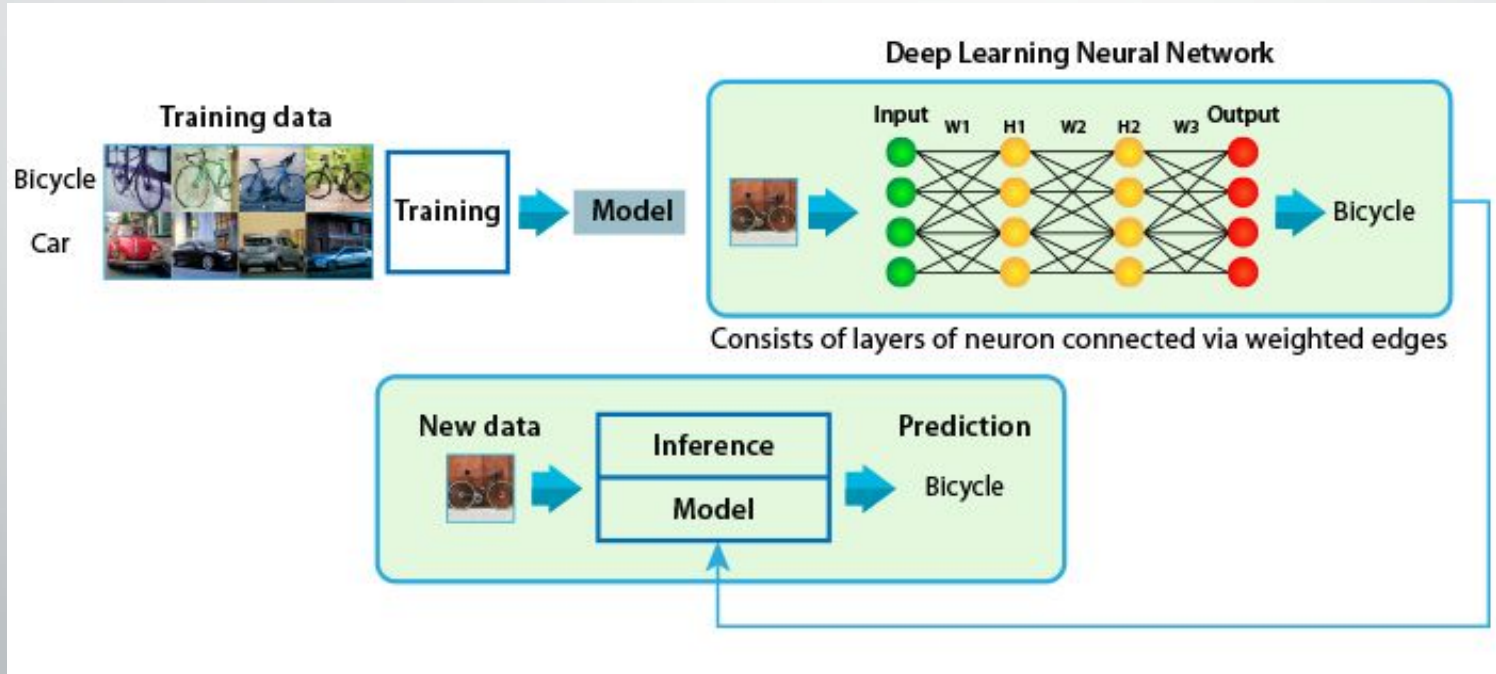
● Input Layer

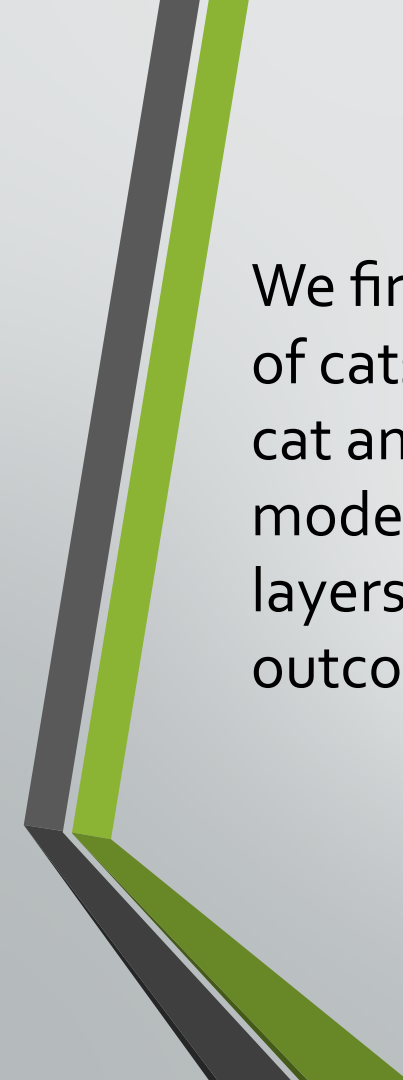
● Hidden Layer

● Output Layer



The main difference between NN and DNN is that, DNN can learn on its own and NN cannot.



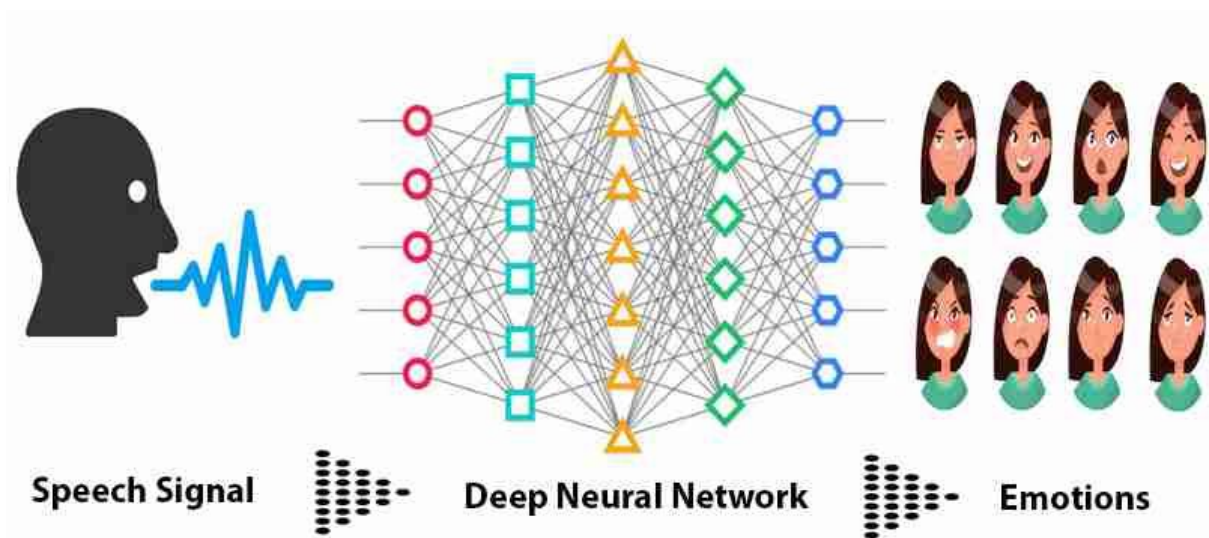
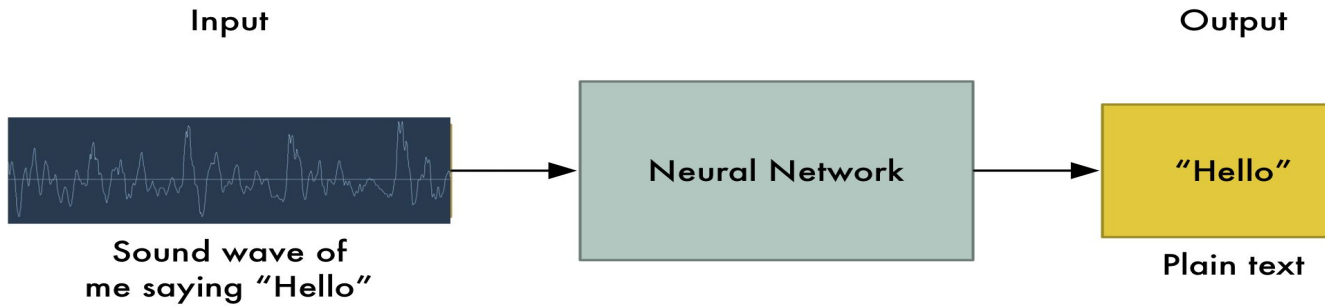


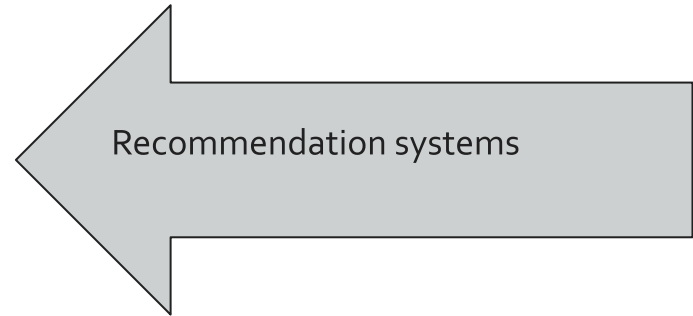
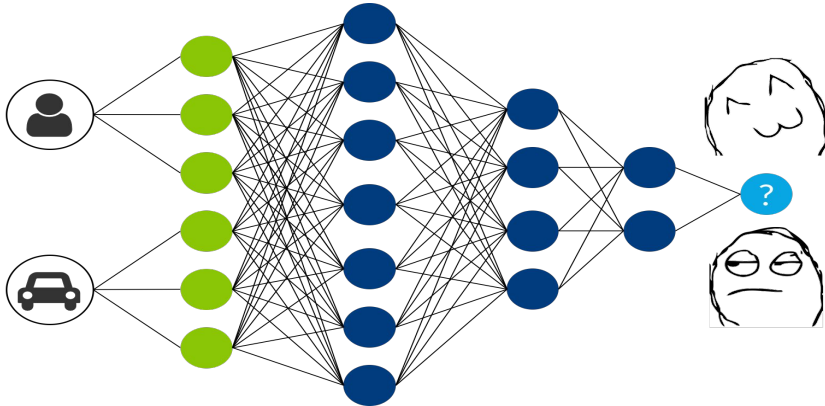
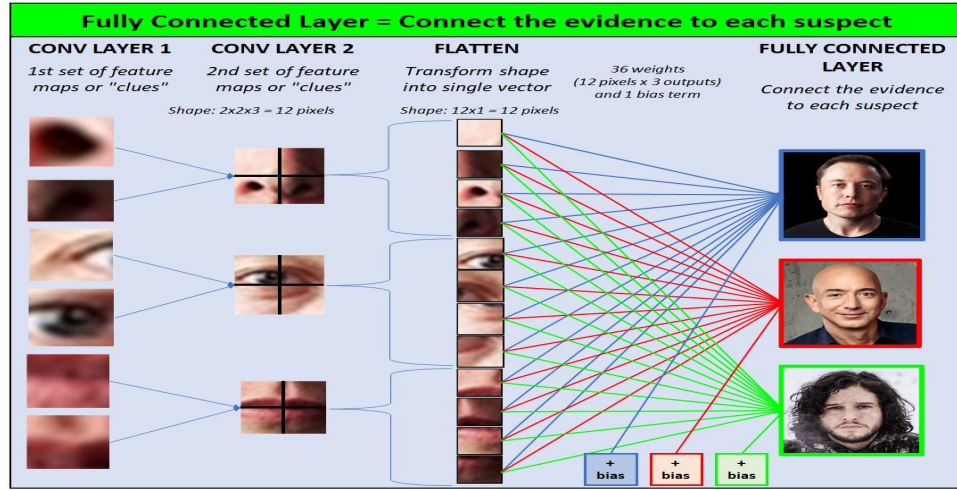
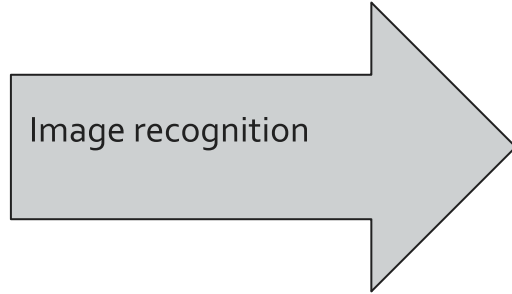
We first train these “models”. For example we feed pictures of cats and dogs and Also let it know whether it is a dog or cat and thus we train it. After that, when using the trained models, while the input moves through the layers, probability would be assigned to the possible outcomes, its all about probability here.

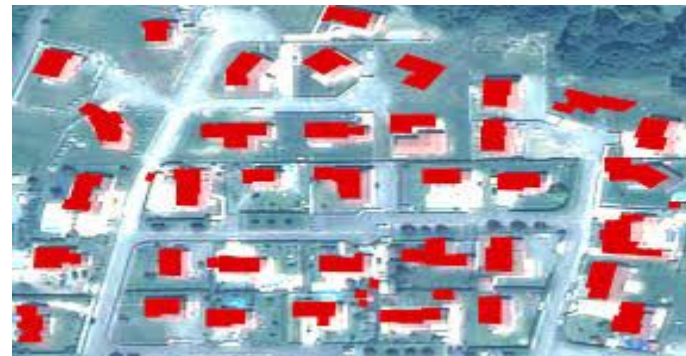
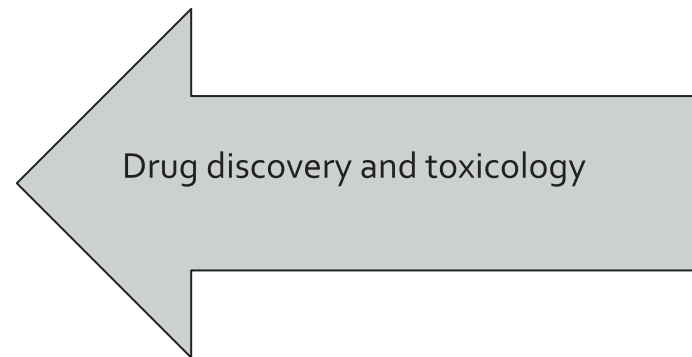
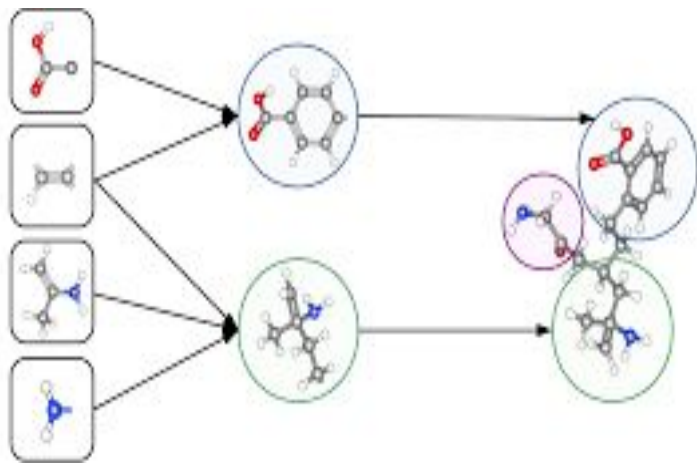


Here are some applications:

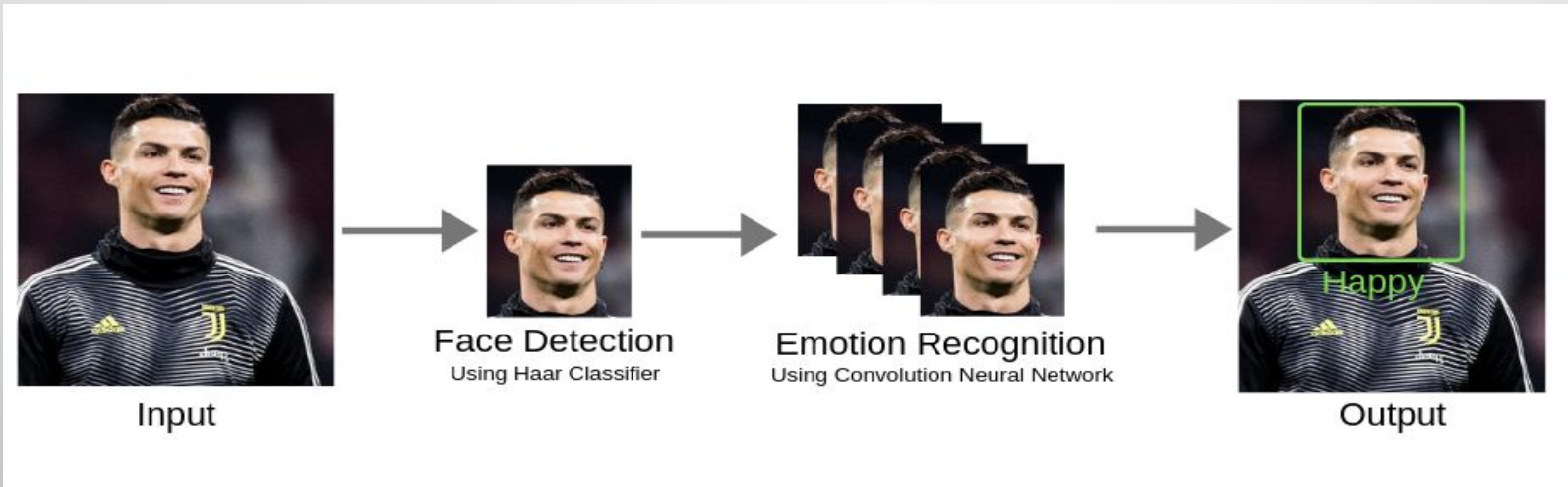
- 1.Speech Recognition
- 2.Natural Language Processing
- 3.Image recognition
- 4.Drug discovery and toxicology
- 5.Recommendation systems
- 6.Remote sensing and many more.







No doubt DNN is doing wonders in solving real world problems.



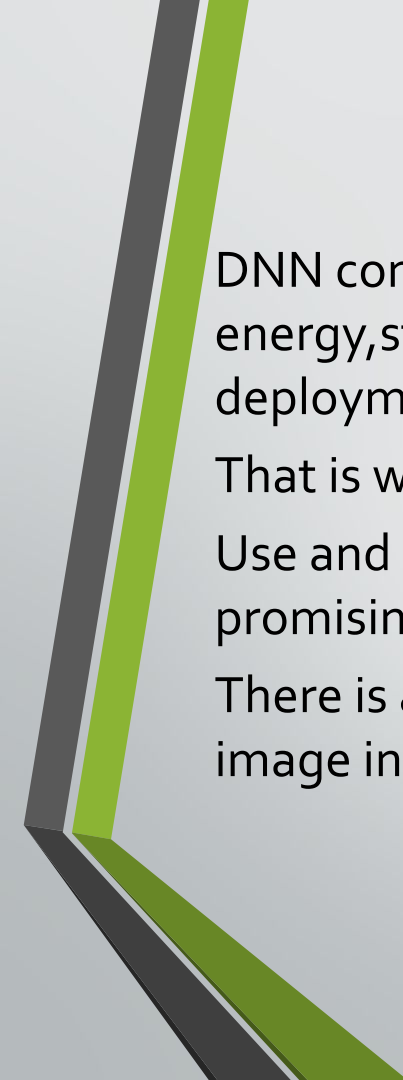
But there is a problem.

**ONE DOES NOT SIMPLY**



**GENERATE MEMES USING DEEP  
LEARNING**

[memegenerator.net](http://memegenerator.net)



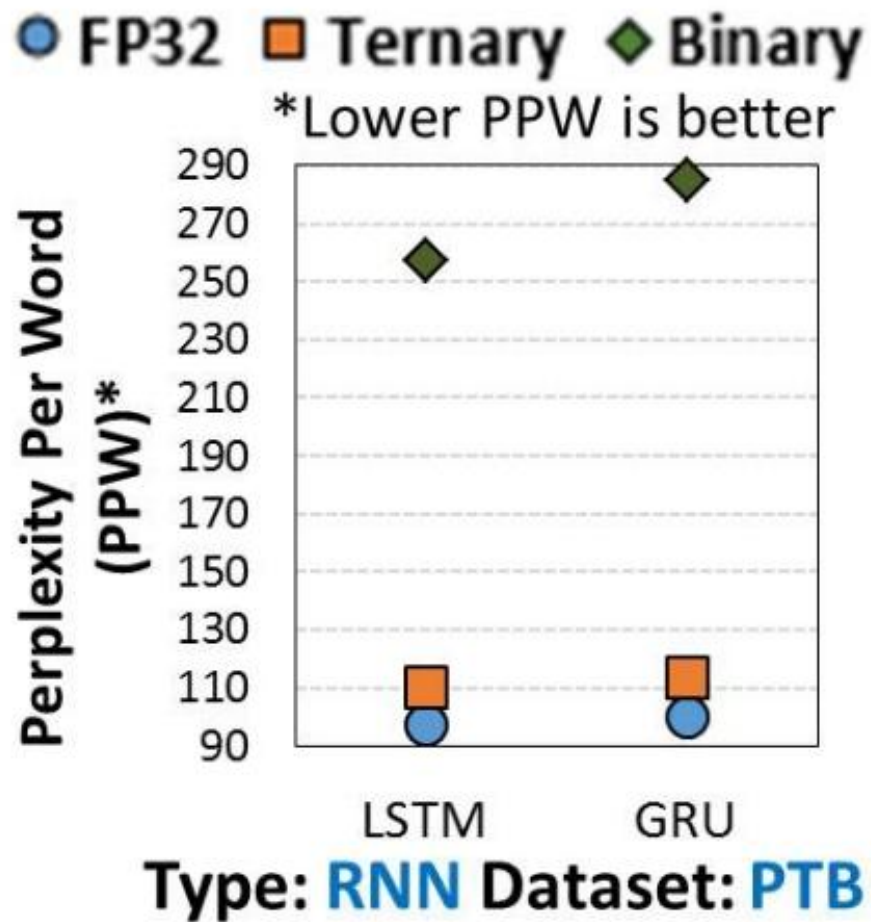
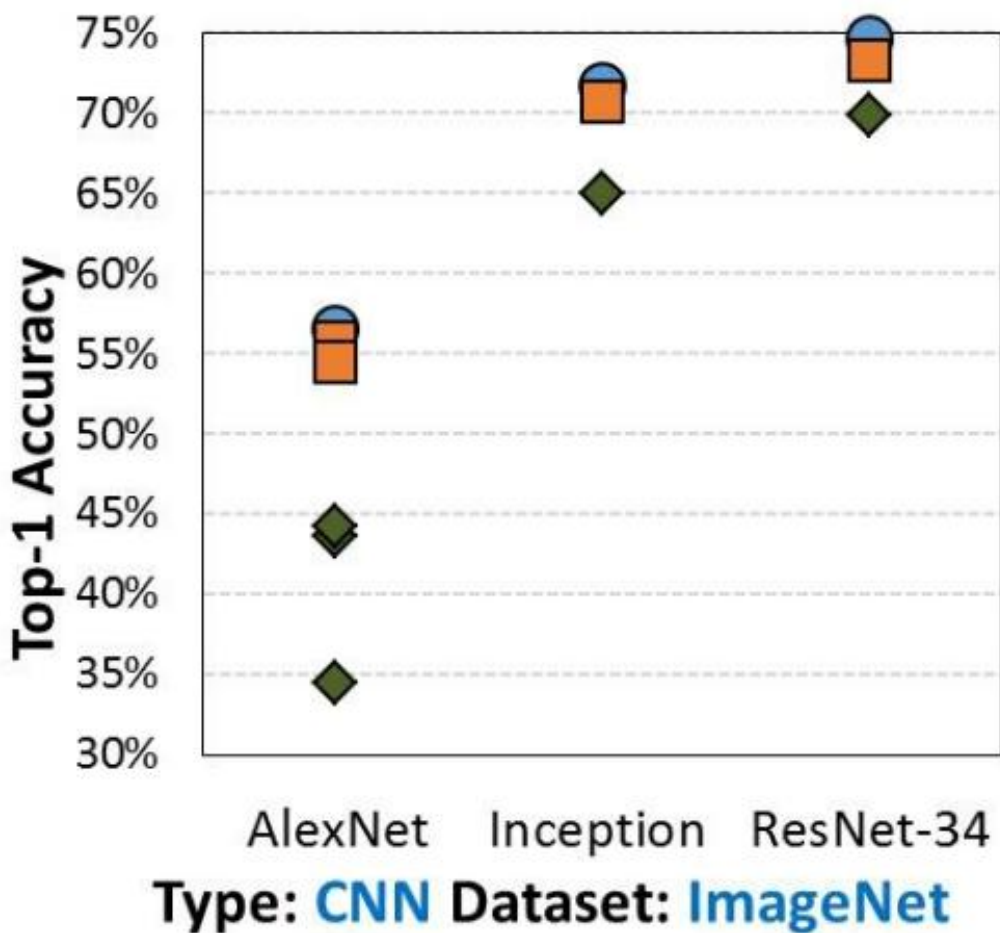
DNN consumes huge amount of resources including energy, storage, computation, data transfer, this results in DNNs limiting their deployment in energy and cost constrained devices.

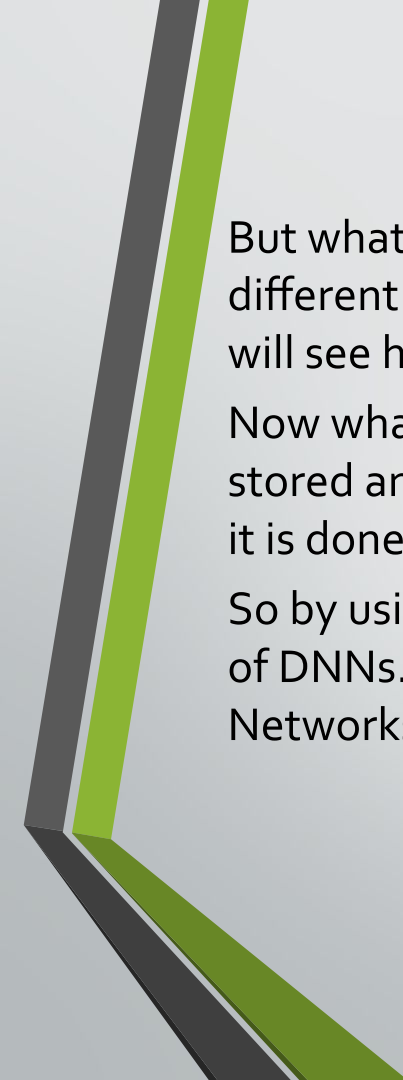
That is where Tiny DNN accelerator comes in play.

Use of lower precision to represent the weights and activations in DNNs is a promising technique for improving the efficiency.

There is a sweet spot in trade off between efficiency and accuracy. Let us see image in next slide.







But what is ternary representation, just like we use 0 and 1 in binary, we use three different ones in ternary and for balance we use -1, 0, 1 to represent the data. We will see how ternary is better than binary in more detail later.

Now what is In-Memory, The processing will be more efficient if the memory stored and processing unit are as near to each other. Find it increases even more if it is done inside at the memory unit itself.

So by using In-Memory and Ternary representation we can "Accelerate" one work of DNNs. So, this comes the name Ternary In-Memory Accelerator for Deep Neural Networks.

---

We shall go into details of  
circuits in the next review.







# Thank You!

Team:

D.Y.D.Pradeep(AP18110020221)

D. Vinayak Kumar (AP18110020267)

P. Srikari (AP18110020259)