

# CLASSIFICATION ON CANCER DATASET BY USING EXTREME LEARNING MACHINE AND MAP REDUCE

Ayush Verma

SRM University, KTR campus

Chennai, Tamil Nadu,

ayush\_verma@srmuniv.edu.in

**Abstract:** Extreme learning machine (ELM) is a machine learning algorithm for any single layered feed forward neural network. It is used on computational intelligence methods that are largely used to analyze large data sets. The strategy used to play out the learning procedure is known as the learning algorithm. ELM is used for classifications and the regressions of data sets. Cancer has become one of the most dangerous illnesses now-a-days that can be treated or cured if detected early on. Cancer is a disease in which abnormal cells divide uncontrollably and destroy body tissues. Malignancy is one of the sickness in which an arrangement of cells are not ready to control their development. The newly developed Extreme Learning Machine is used for the classification problems in the cancer diagnosis area. ELM is used for the advanced and developed methodology known which are used for the cancer multi classification ELM microarray gene expression cancer diagnosis. This is used for directing the multi category classification problems in the cancer diagnosis area. ELM avoids many problems like improper learning rate and over fitting commonly faced by iterative learning methods and completes the training period very quickly. In regression (ELM) can get better generalization performance and much faster learning speed. However, the enlarging volume of datasets makes regression by ELM on very large scale datasets very slow.

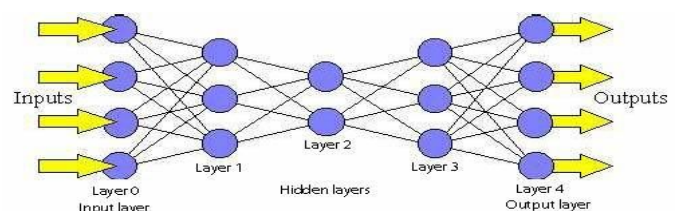
**Keywords:** ELM, multi classification, Regression.

## INTRODUCTION:

**Neural network:** Neural Network, as the name suggests, is a network of computing systems which is inspired by human

brain's neural map. Neural networks are sometimes described in terms of their depth, including how many layers they have between input and output, or the model's so-called hidden layers.

**Feed forward neural network** Feed forward neural networks are commonly used for classification problems. Perceptrons are arranged in layers, with the first layer taking in inputs, and are hence called the input layer, whereas the last layer produces outputs, and hence are called the output layers. The middle layers have no connection with the external world, and hence are called hidden layers. Data is constantly pushed forward or "fed forward" from one layer to the another layer so it is called the feed-forward neural network.



## Back propagation

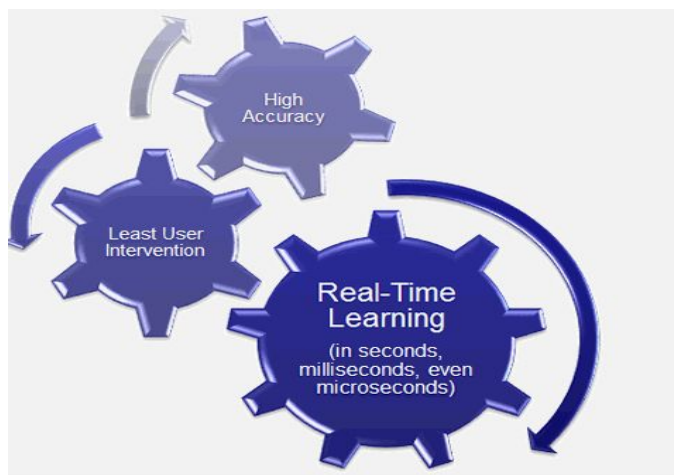
Feed forward Neural Networks such as back-propagation are slow and the reason is: Training using gradient based learning algorithms e.g. many iterative steps required to achieve better

performance. The weights are adjusted iteratively using gradient methods. Extreme Learning Machine (ELM) was proposed to overcome these issues and offer better generalization performance.

### Extreme machine learning:

Extreme machine learning is considered as a two stage multiple layer Feed Forward Neural Network – wherein the first stage would be the stage where the learning systems gives the connections within the hidden neurons and second stage would be the adjustment of those connections with the output neurons, along with the learning algorithm that completes the training fairly fast. ELM classifier transfers the training of FNN into linear problem in which the connections only as output neurons need adjusting. The reason why ELM is faster than the other machine learning classifiers is that the learning parameters of the nodes are randomly assigned. The expression level of genes holds the solutions to overcome basic drawbacks related to prevention and treatment of cancer. The microarray gene data must be pre processed for classification with good accuracy using the classifier Extreme Learning Machine is currently popular neural network architecture based on the random projections. The training, prediction is fast compared with the other non linear methods. ELM, as ELM is fully automated and in theory almost no intervention is needed from the user.

### Structure of Extreme Learning Machine:



### ELM Algorithm:

- Given a training set  $\mathbb{N} = \{(Y_i, t_i) \mid Y_i \in \mathbb{R}^n, t_i \in \mathbb{R}^m, i = 1, 2, \dots, N\}$ , activation function  $g(x)$  and hidden node number –  $L$ ,
- Step – 1: Assign random hidden nodes by randomly generating learning parameters  $(u_i, v_i)$ ,  $i = 1, 2, \dots, L$
- Step – 2: Calculate the hidden layer output matrix  $H$
- Step – 3: Calculate the output weight  $\beta : \beta = \hat{H}^+ T$ .
- Where  $\hat{H}^+$  is the Moore Penrose generalized inverse of the hidden layer matrix –  $H$ .
- The number of hidden nodes in the hidden layer of SLFN is obtained by trial and the error method

**ELM in Large Data Applications:** For Large Data Applications ELM can have the property of low computational complexity due to which ELM which attracts extensive attention from research communities. The reduced computational burden makes ELM and its variants like Error-minimized ELM, Symmetric ELM more feasible in dealing with large data applications than the conventional iterative algorithms. ELM not only learns thousands of times faster than conventional popular learning algorithm for Single Layered Feed Forward Neural Network but also produces good generalization performance.

**ELM in matlab:** The reliable performance and quick learning speed, image recognition and objective detection using ELM have attracted increasing attention in the recent years. Among these applications, face recognition based on ELM and its developments is one of the top topics which has been widely researched by many researchers by face recognition performance using the ELM classifier

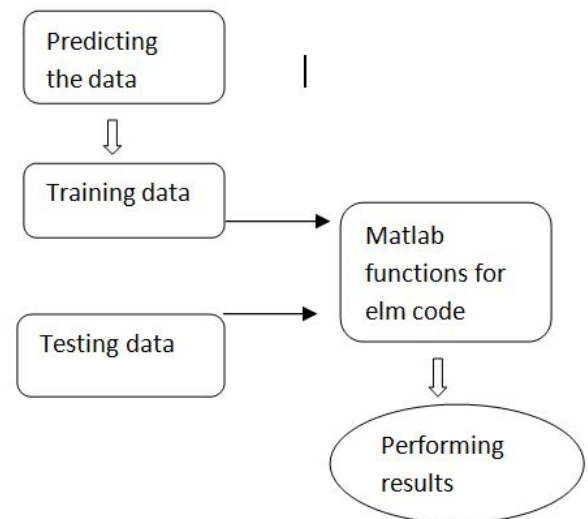
**LITERATURE SURVEY:** Regression is the first basic problems in data mining. For regression (ELM) can get better generalization performance at a much faster learning speed. However, the enlarged volume of datasets makes regression by ELM on very huge type of datasets with a challenging task via analyzing the mechanism of ELM algorithm, an efficient parallel ELM for regression is developed and implemented based on Map Reduce framework.

(ELM) classification is quick and efficient methods which are used for a multi-category cancer diagnosis problem depends

on microarray data. Its performance has been compared with other methods such as the SVM algorithms ELM avoids many problems like local minima, improper learning rate and over fitting commonly faced by iterative learning methods and finishes the training very quickly. SVM for multi category classifications is done by modifying the binary classification method of SVM .This inevitably involves more classifiers, greater system complexities and computational burden, and a longer training time.

- Extreme Learning Machine of A novel Distributed based on Map Reduce framework, named ELM which cover the shortage of traditional ELM whose learning is weak to the huge dataset. Firstly, after adequately analyzing the property of traditional ELM.
- The learning process that which output connections are tuned via minimizing the cost functions via linear system. The computational bare of ELM has been reduced as the cost is solving in a linear system. The complexity attracted a great deal of attention from the research community with low computational, especially for high dimensional and large datasets
- The feature selection procedure was applied to the datasets. Rather than randomly assigning ELM-RBF centers and width. The results are compared to several classification methods like OPELM in terms of Accuracy and minimum Sensitivity
- The newly developed Extreme Learning Machine (ELM) is used for directing multi category classification problems of the cancer diagnosis area. Which avoids problems like local minima

## IMPLEMENTATION WORKFLOW:



## METHODOLOGY:

Microarray analysis is not easy because of the huge number of genes that are done simultaneously. in several factors of interest (for time instance and different kinds of treatments) in the experimental design by, the interpretation of the data will become even more complicated. The factors of interest should be separated from one other to draw output from the data analysis. To overcome these problems, a new methodology called ELM is developed. Nothing like traditional implementations and learning theory, from function approximation point of view, ELM theory shows that the hidden node parameters can be completely independent from the training data based upon these training and the testing data .the ELM packages are installed in the matlab for the inbuilt functions to perform the results like training time and the accuracy of the particular data.

## Training model in Matlab:

```
[>> m = load('elm_model')

m =

    TrainingAccuracy: 0.5000
           Elm_Type: 1
    ActivationFunction: 'sig'
                T: [2x4 double]
                P: [2x4 double]
    NumberofInputNeurons: 2
                label: [0 1]
    NumberofOutputNeurons: 2
                InputWeight: [3x2 double]
    BiasofHiddenNeurons: [3x1 double]
                tempH: [3x4 double]
                H: [3x4 double]
                OutputWeight: [3x2 double]
                Y: [2x4 double]
                output: [1 0 0 1]
```

## Conclusion:

The cancer classification for ELM algorithm in which the problem is presented. It is observed that the algorithm finds a small gene subset that provides high classification accuracy compared with different techniques and implementations for this algorithm which is used for classification. The feature selection procedure was applied to the datasets and ELM is found to obtain the best performance results for these datasets.

## References:

- [1] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, "[Extreme Learning Machine for Regression and Multiclass Classification](#)," *IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics*, vol. 42, no. 2, pp. 513-529, 2012.
- [2] Z. Bai, G.-B. Huang, D. Wang, H. Wang and M. B. Westover, "[Sparse Extreme Learning Machine for Classification](#)," *IEEE Transactions on Cybernetics*, vol. 44, no. 10, pp. 1858-1870, 2014.
- [3] G.-B. Huang, X. Ding, and H. Zhou, "[Optimization Method Based Extreme Learning Machine for Classification](#)", *Neuro computing*, vol. 74, pp. 155-163, 2010
- [4] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "[Extreme Learning Machine: A New Learning Scheme of Feedforward Neural Networks](#)," 2004 *International Joint Conference on Neural Networks (IJCNN'2004)*, (Budapest, Hungary), July 25-29, 2004.
5. [Extreme Learning Machines on High Dimensional and Large Data Applications: IEEE Transactions on Systems](#) A SurJiuwenCao1 and ZhipingLin2
6. [Multi category Classification Using an Extreme Learning Machine for Microarray Gene Expression Cancer Diagnosis](#) Narasimhan Sundararajan, and P. Saratchandran *IEEE Transactions on Systems* Year : 2007
7. [Extreme Learning Machine RBF Network case](#) *Transactions on Systems* Year: 2014 G.-B. Huang and C.-K. Siew.
8. [parallel extreme learning machine for regression based on Map Reduce](#) *IEEE Transactions* in 2015 Tianfeng, fuizhen hung
9. [distributed extreme learning machine with Map Reduce](#) *IEEE Transactions* in 2012 by Zhiqiong Wang Chen Chen Linlin Ding Guoren Wang Yuhai Zhao
10. [Extreme Learning Machines on High Dimensional and Large Data Applications](#) by *IEEE Transactions* in 2009 by JiuwenCao1 and ZhipingLin

