Does Machine Learning Based Multi-Model Ensemble Methods Add Value over Existing Methods?

Nachiketa Acharya
nachiketa@iri.columbia.edu

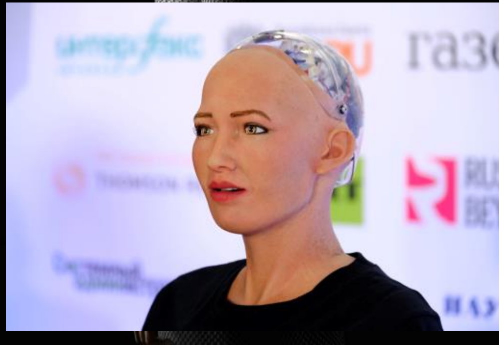
International Research Institute
for Climate and Society
EARTH INSTITUTE | COLUMBIA UNIVERSITY

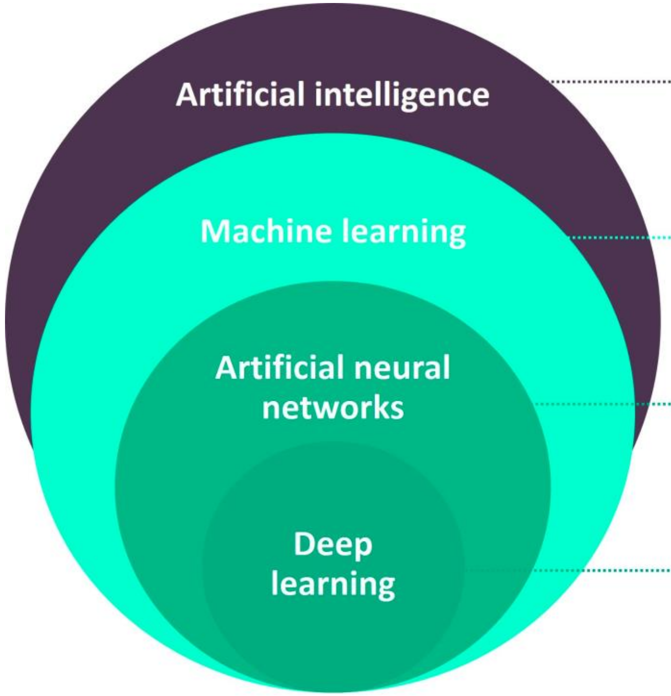
45th Annual Climate Diagnostics and Prediction Workshop
Oct,20,2020

Motivation and Goal

- The generation of multi-model ensemble (MME) is a well-accepted approach to improve the skill of forecasts from individual GCMs.
- There are two common approaches to make a MME, viz., combining the individual ensemble forecasts with equal weights, or weighted according to their prior performance.
- Irrespective of which combination method has been used, plethora of studies have shown that multi-model ensembles do increase prediction skill over single-model forecasting.
- Weighted MME: Mostly linear combination of GCM using multiple linear regression (MLR). (PCR, BMA).
- Does Machine Learning Based Multi-Model Ensemble Methods Add Value over Existing Methods?

What is Machine Learning?



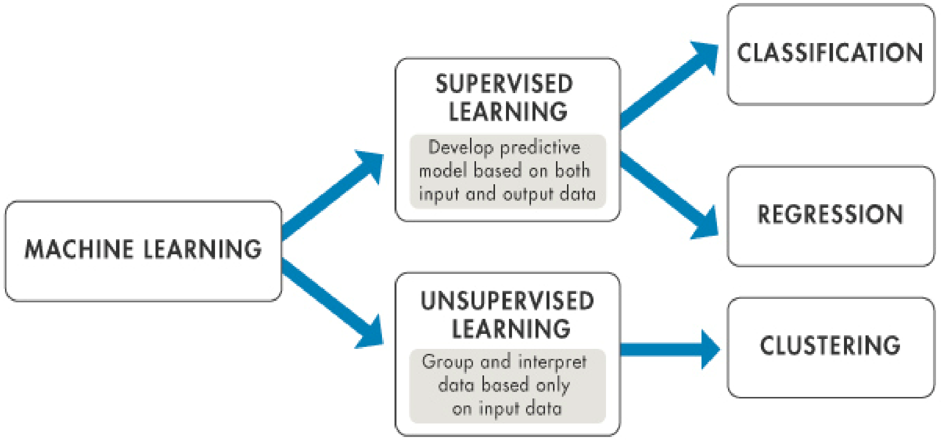


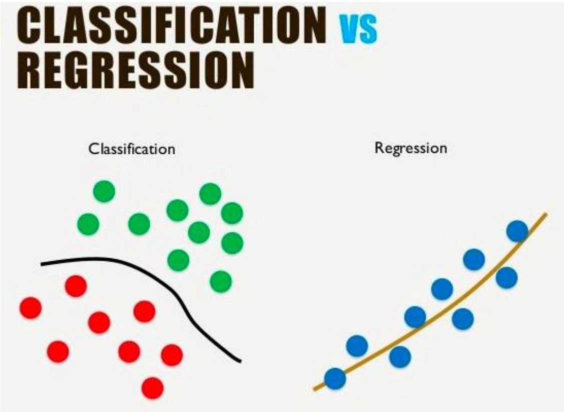
Artificial intelligence (AI)
Any techniques that enable machines to solve a task in a way like humans do

Machine learning (ML)
Algorithms that allow computers to learn from examples without being explicitly programmed

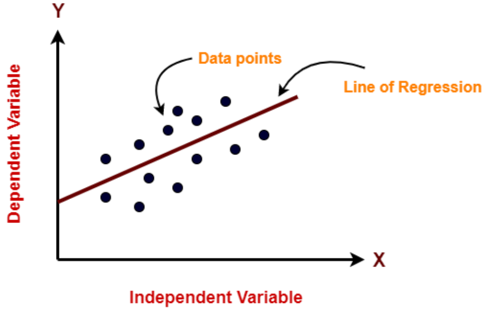
Artificial neural networks (ANN)
Brain-inspired machine learning models

Deep learning (DL)
A subset of ML which uses deep artificial neural networks as models and automatically builds a hierarchy of data representations



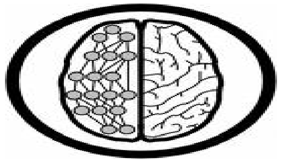


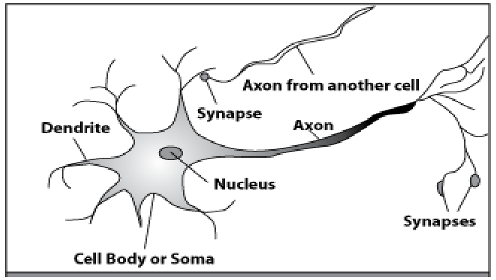
The simplest form of ML is a linear regression model!



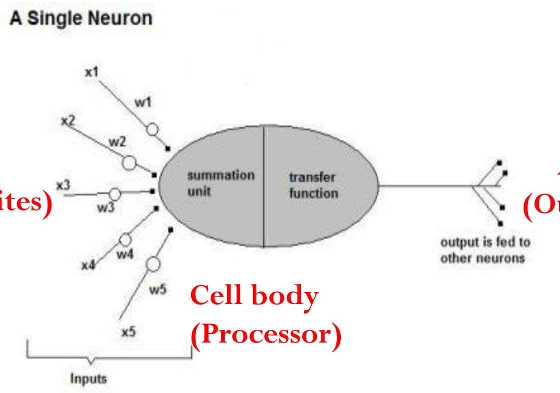
Artificial Neural Network

An Artificial Neural Network (ANN) is a computational model inspired in the networking of natural neurons



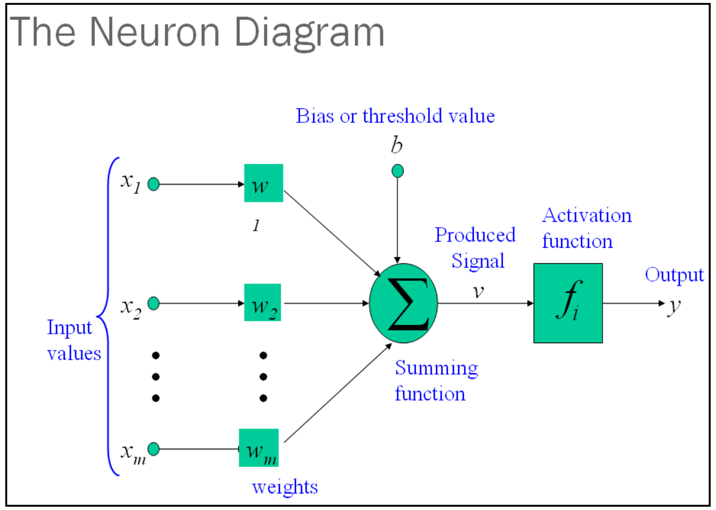


Input (Dendrites)



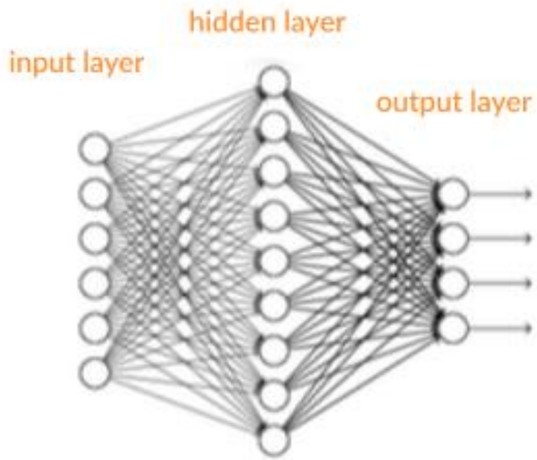
Axon (Output)

Cell body (Processor)

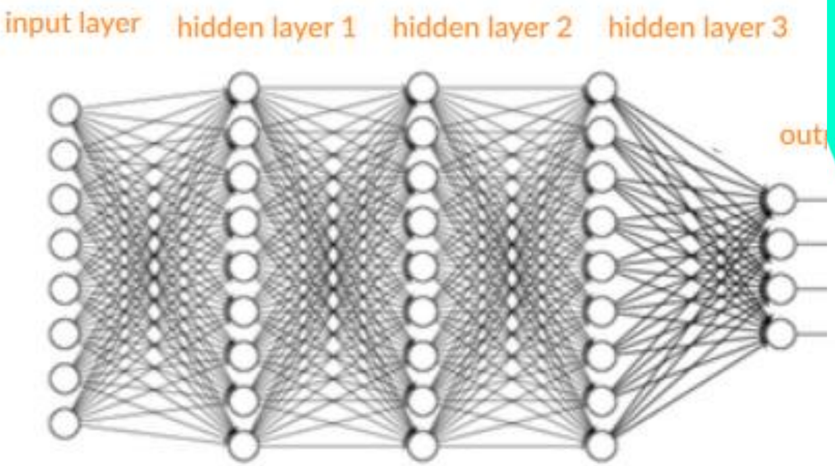


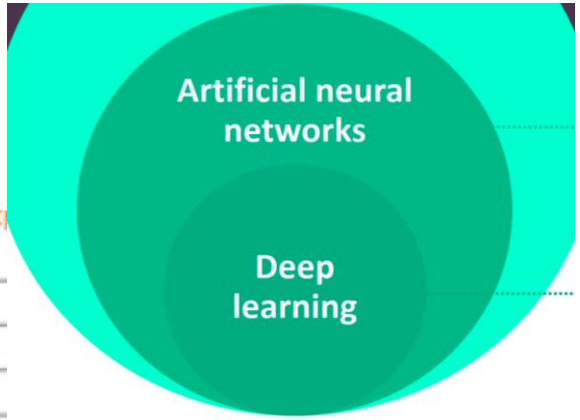
The concept of neural networks was come into being when the first model of neuron was created by two scientist, **Warren McCulloch**, a neurobiologist, and **Walter Pitts**, a statistician, in 1943.

shallow feedforward neural network

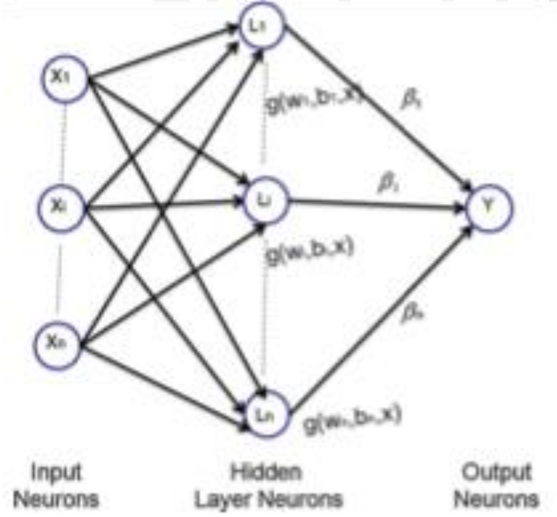


Deep neural network

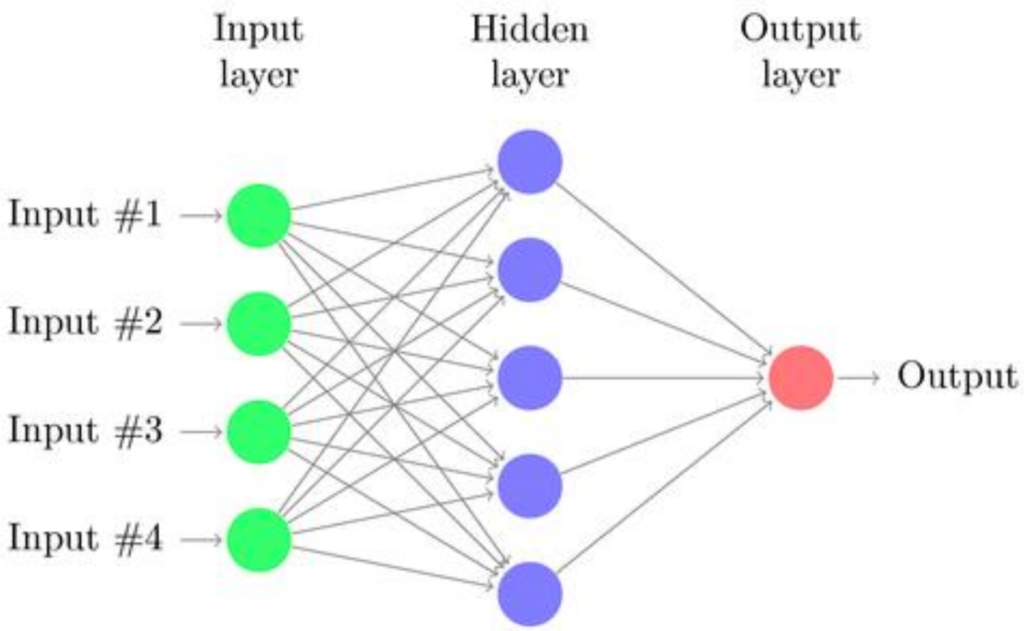




...den-layer feed forward neural networks (SLFN)



...edforward network (SLFN), as one of the ...orward neural networks, have been ...rom both theoretical and application ...arning capabilities and fault-tolerant abilities. Most of the SLFN use gradient-based learning algorithms such as the back propagation neural network (BPNN).



Limitations:

- ❖ Efficiency of ANN based methods is highly dependent on appropriate tuning of their adjustable parameters, e.g., the number of hidden layers, nodes, weights and transfer function. There are also several disadvantages of traditional SLFN which includes long computation time, stopping criteria, learning rate, learning epochs, local minima, and the over-tuning problems.

Extreme Learning Machine: A “Generalized” SLFN

❖To overcome such shortcomings, recently, a novel learning algorithm for single-hidden-layer feed forward neural networks (SLFN) called **extreme learning machine (ELM) has been proposed by Huang et al., (2008).**

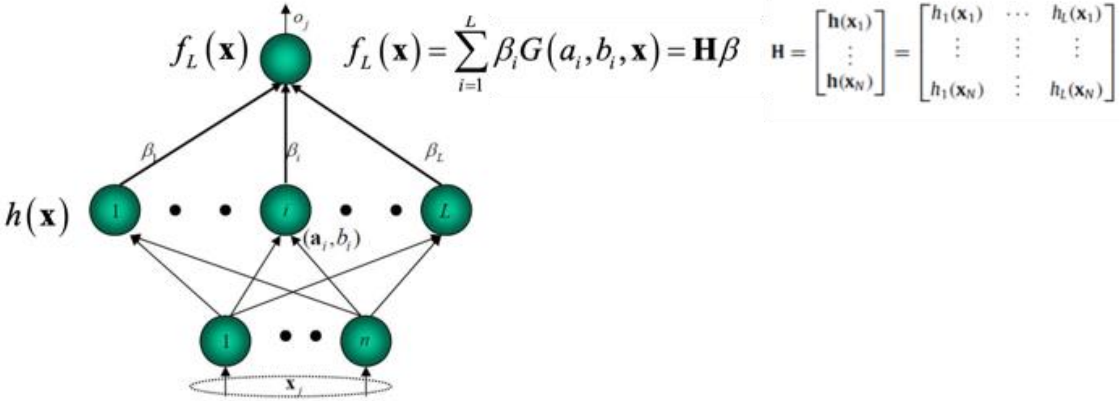
❖In the proposed algorithm, **the input weights and hidden biases are randomly chosen**. Randomly chosen input weights can efficiently learn distinct training examples with minimum error.

❖After randomly choosing the input weights and the hidden layer biases, **SLFNs can be simply considered as a linear system.** The output weights which link the hidden layer to the output layer of this linear system can now be analytically determined through **Moore- Penrose (MP) generalized inverse** of the hidden layer output matrices.

❖The basic principle which distinguishes ELM from the traditional SLFN is that **all the parameters (input weights and hidden layer biases) are not required to be tuned**.

❖This simplified approach makes ELM **thousands of times faster than that of traditional SLFN**. ELM also avoids many difficulties faced by gradient-based learning methods such as stopping criteria, learning rate, learning epochs, local minima, and the over- tuning problems.

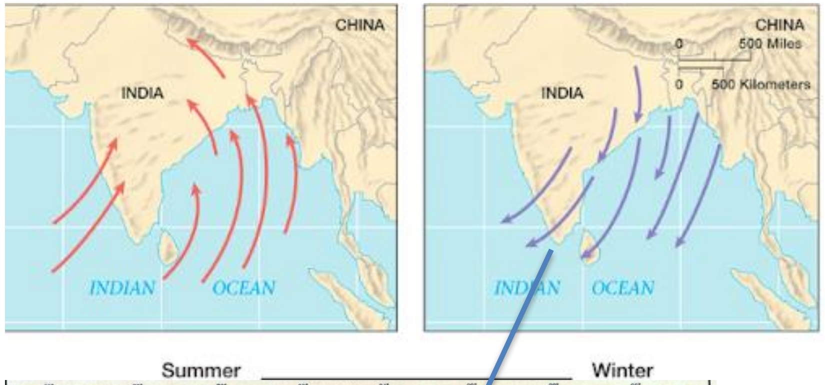
“Extreme means to move beyond conventional artificial learning techniques and to move toward brain alike learning”.-Huang,
Nanyang Technological University,

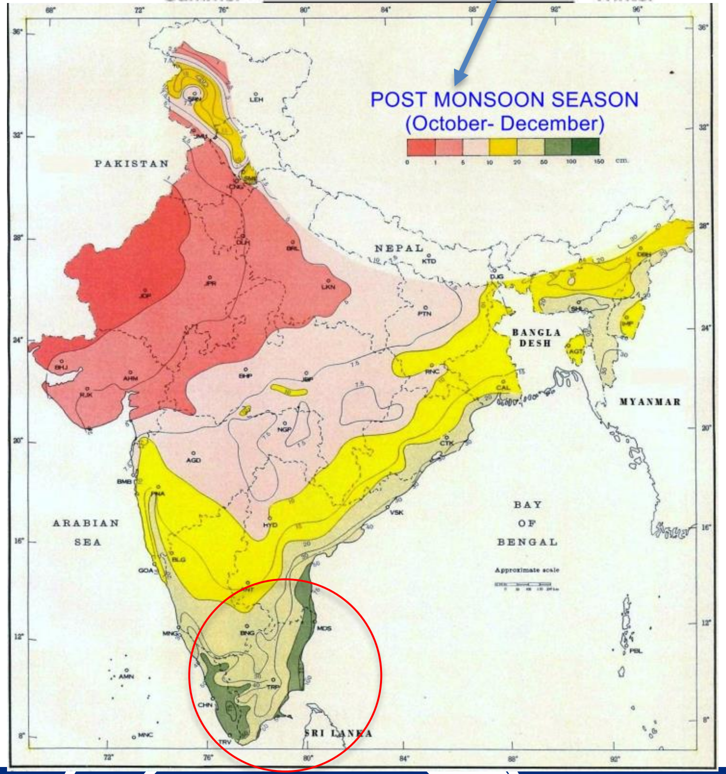


Salient Features

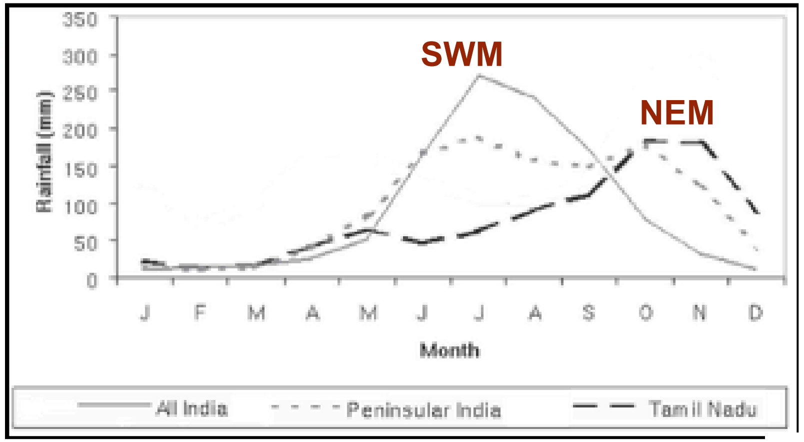
- “Simple Math is Enough.” ELM is a simple tuning-free three-step algorithm.
- The learning speed of ELM is extremely fast.

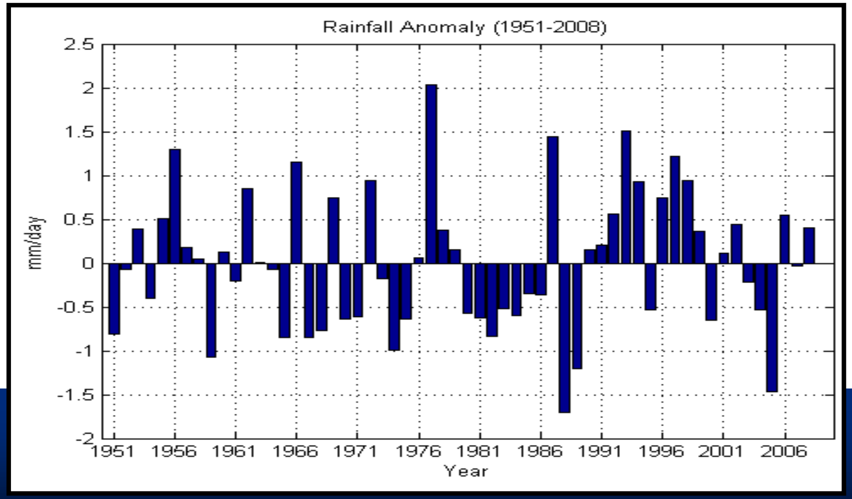
Case study: North-East (winter) Indian Monsoon



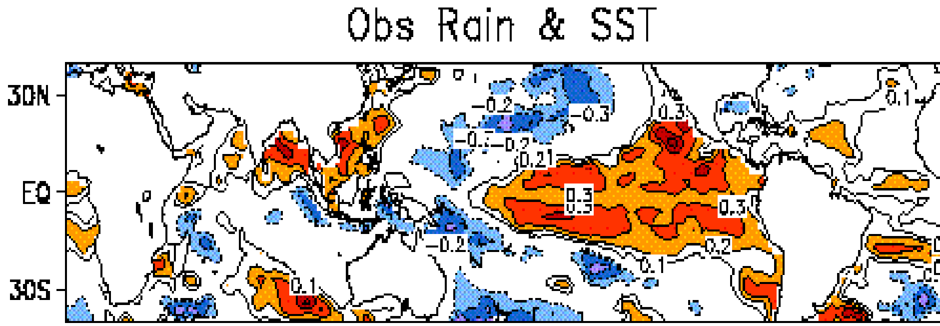


The NE monsoon season (October, November, and December) contributes to about 50% of annual rainfall in the east coast of Indian Peninsula.

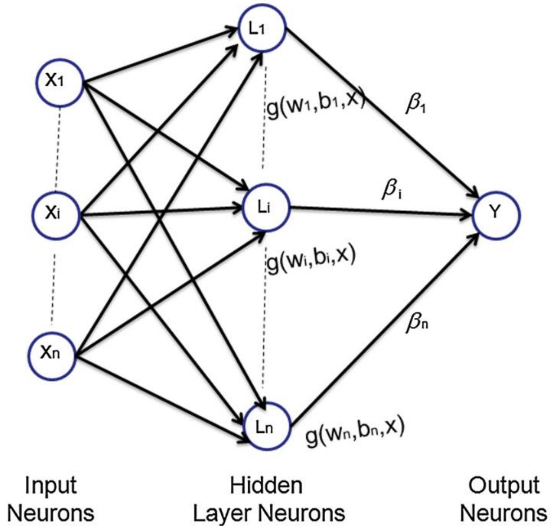




Teleconnection with ENSO



Implementation procedure of ELM for making MME



Selection of the input (Xs) and output (Y) neurons

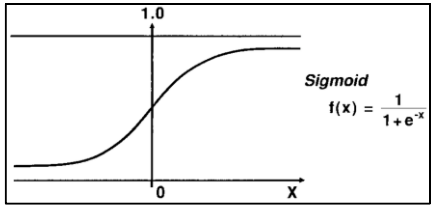
The outputs from 7 GCMs are used as input neurons and observed rainfall is treated as targeted output.

Scaling the neurons

The input neurons are scaled in the range of [-1, 1].

$$Y = Y_{\min} + (Y_{\max} - Y_{\min}) \times \left(\frac{X - X_{\min}}{X_{\max} - X_{\min}} \right)$$

Selection of the activation function

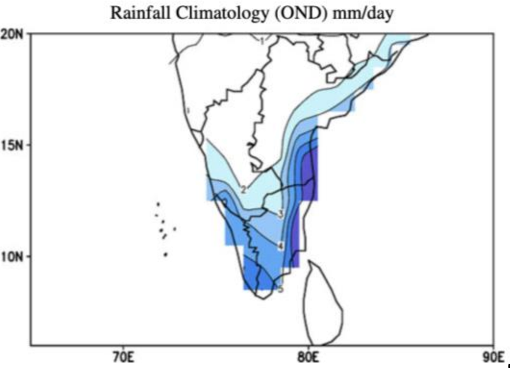


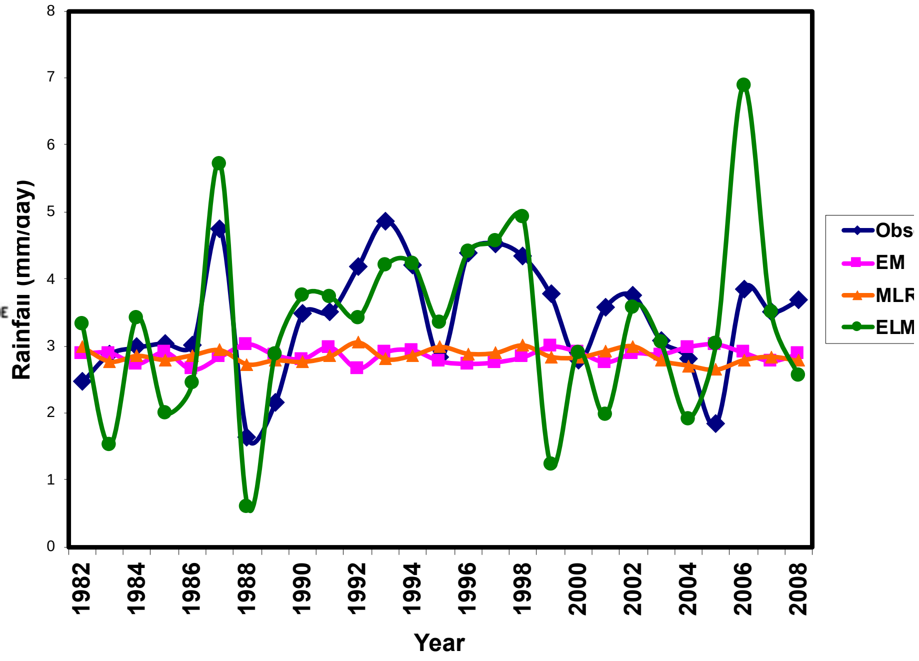
Training and testing of the model

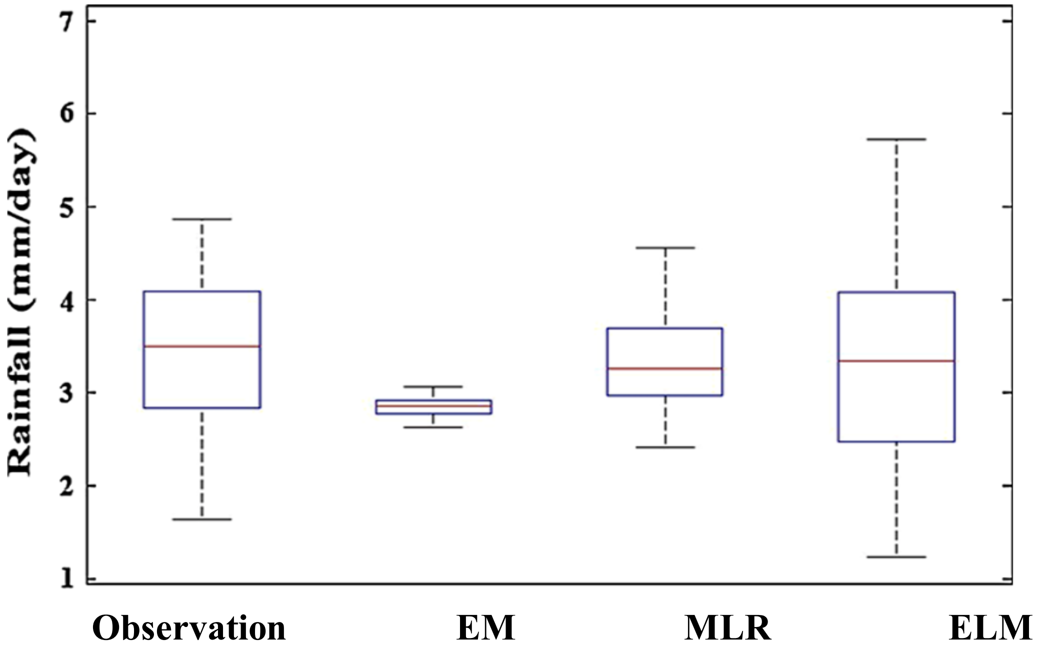
A leave-one-out cross- validation method has been used.
Data period:1982-2008

Final structure of the ELM with 7-neurons as input, 25 nodes in the hidden layer and 1 output neuron (7-25-1)

Performance of ELM-MME compare to standard MME







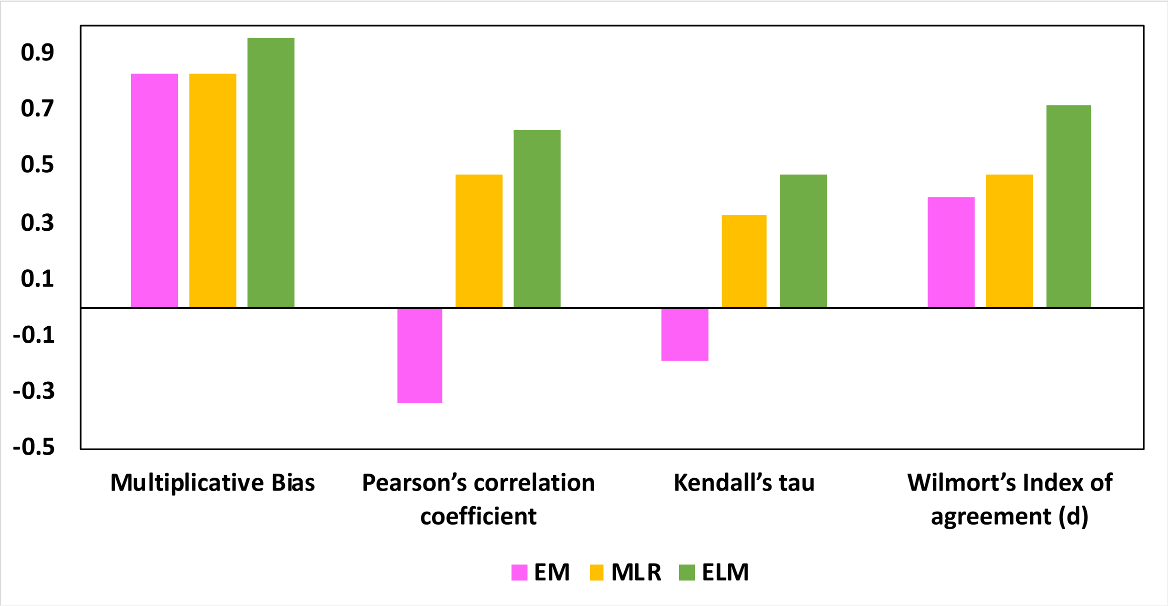
EM= Mean of all GCM.

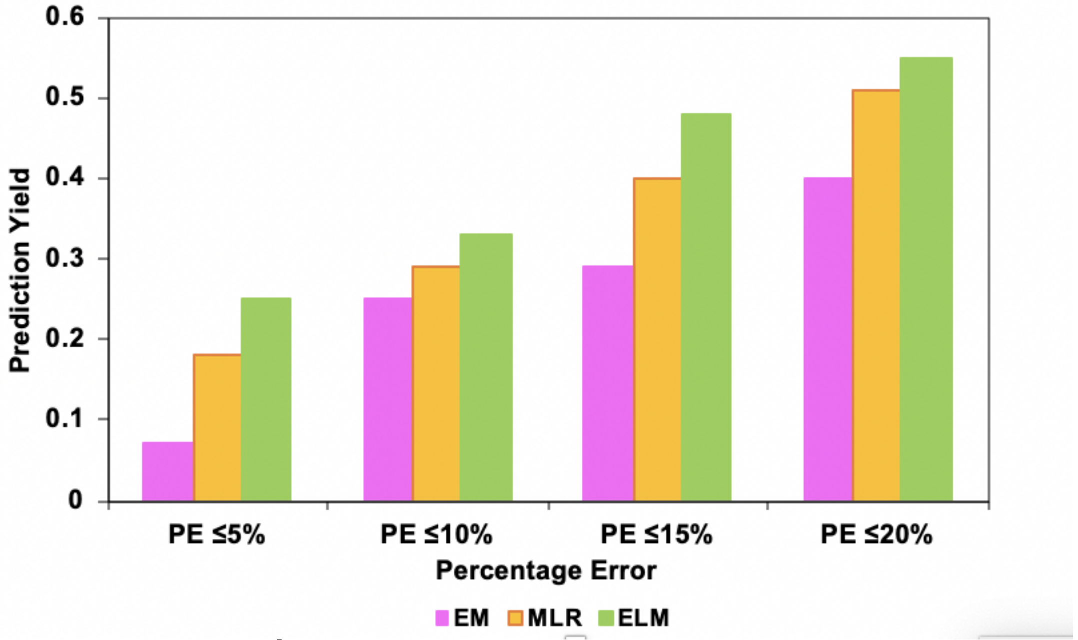
$$S_t = \overline{O} + \frac{1}{N}\sum_{i=1}^{N}\left(\frac{F_{i,t} - \overline{F}_i}{\sigma_{F_i}}\right)\sigma_O$$

MLR= Multiple linear regression between GCMs and Observation (“Superensemble”)

$$S_t = \overline{O} + a_i \sum_{i=1}^{N}\left(\frac{F_{i,t} - \overline{F}_i}{\sigma_{F_i}}\right)\sigma_O$$

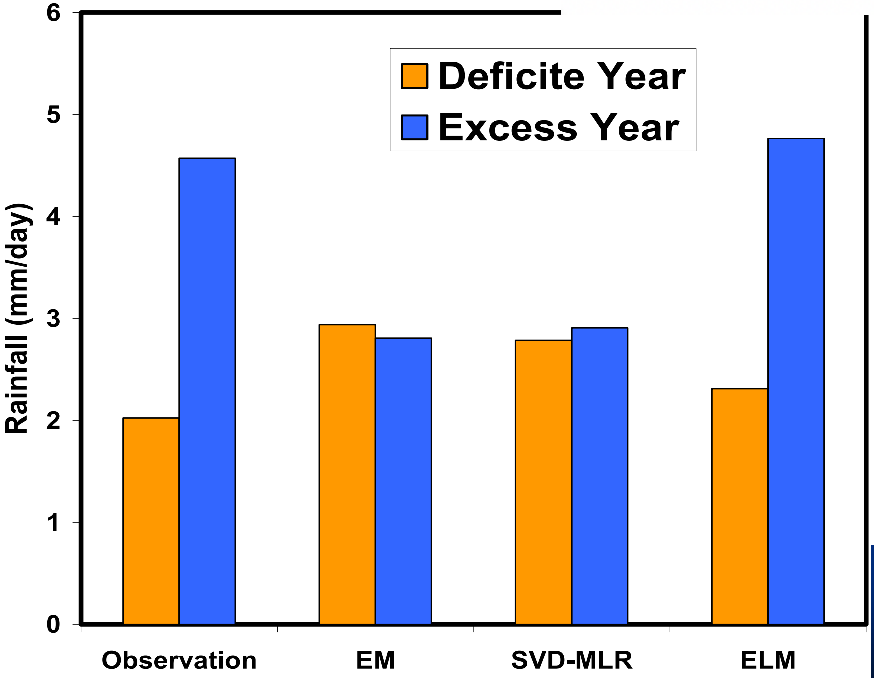
Performance of ELM-MME compare to standard MME





Index of Agreement (d)

$$d = 1 - \left[\frac{\sum_i (f_i - O_i)^2}{\sum_i \left(\left| f_i - \overline{O} \right| + \left| O_i - \overline{O} \right| \right)^2} \right] \quad (0 \leq d \leq 1).$$



Deficit year: 1982,1988,1989,2005;

Excess year: 1987,1993,1996,1997,1998.

Concluding Remark

- ELM is simplified and “generalized SLFN” which make it thousands of times faster than that of traditional SLFN.
- In ELM, parameters (input weights and hidden layer biases) need not be tuned.
- There is a significant improvement by ELM compare to existing MME methods in terms of skill scores.
- Especially, ELM capture the inter-annual variability of observed rainfall over other MME schemes.
- ELM also capture the “extreme” year and discriminate between wet and dry year.
- Scope: Need more data to set up a more robust network for ELM.

Thank you!

Reference: Acharya N, Srivastava N.A., Panigrahi B.K. and Mohanty U.C. (2013): Artificial Neural Network based Multi-model ensemble to improve prediction of northeast monsoon rainfall over South Peninsular India: an application of Extreme Learning Machine. *Climate Dynamics* DOI: 10.1007/s00382-013-1942-2.

Any Feedback?