A DE(A)NN APPROACH FOR MEASURING THE EFFICIENCY OF U.P. BOARD SECONDARY EDUCATION

Pallavi Pant
Master of Technology Research Scholar
Department of Electronics and Communication Engineering,
Rameshwaram Institute of Technology and Management, Lucknow, India.

Abstract: The research below is conducted to measure the efficiency of U.P. board secondary education. Here, the researcher has used an integrated Data envelopment analysis (DEA) and Artificial neural network (ANN) as a tool to measure and predict the efficiencies of a system. This integrated approach is commonly known as DE(A)NN.

Although DEA is a pioneering approach in the research and is used as a tool to measure the efficiency of the system, but this cannot predict the future course values. Thus, to overcome the shortcomings of the DEA system ANN is used in integration with DEA to analyze the system efficiencies as well as predict future efficiencies.

Thus, each year could demonstrate the efficiency and eliminate the inefficiency of the system. Here, a five-year time forecast is analyzed to measure the efficiency of U.P. board secondary education.

***Index Terms*: Data envelopment analysis (DEA), Artificial neural network (ANN), DE(A)NN, U.P. board secondary education**

I. Introduction

Data envelopment analysis is a principal approach for measuring and predicting benchmarks. It utilizes multiple inputs and outputs for decision making, thus is known as multi criteria decision making (MCDM) technique that analyses all DMU's(decision making units) for measuring the efficiency of the system. Various models are used for evaluating the efficiency of the system. DEA is an optimizing tool with a theoretical basis on linear programming which maximizes the efficiency of DMU. It considers ratio of weighted sum of inputs to weighted sum of outputs.
Although DEA has an ability of analyzing and benchmarking, it also has its shortcomings. DEA is a non-parametric method of calculating efficiency by minimizing inputs at a given level of output or visa-versa.

The DEA methodology helps in estimating the "relative efficiency" of a selected group of units and criteria. DEA provides the desired output by minimizing the wastage of resources.

DEA was first proposed in 1978 by by Charnes, Cooper, and Rhodes and was also named as frontier analysis. Some other researches where the DE(A)NN techniques have been used are Shokrollahpour, Lotfi and Zandieh (2016); Singh, Pant, and Goel (2018); Liu, Chen, Chiu, and Kuo (2013); Karamali,Memariani, Jahanshahloo, Malkhalifeh (2013).

II. CASE STUDY

This study presents a complementary approach to promote best performance benchmarking and performance model by using data envelopment analysis (DEA) and artificial neural network (ANN) as an adaptive decision-making tool.DEA and ANN together take advantage of optimization and prediction capabilities innate in each method. DEA is used to measure the relative efficiency of decision-making units (DMUs) which further generates test inputs for subsequent ANN prediction module. The combined modeling is effective through sequential processes which streamlines DEA analysis and ANN prediction.

Here, a DEA model is used to measure the efficiency of the UP Board secondary education result for a span of 5 years i.e. from 2015-2019. For each year various parameters that are mentioned in the methodology are taken into observation for calculating and predicting efficiencies. It further extends its capacity as a preprocessor to the subsequent neural network prediction module. Feed-Forward neural network, in combination with DEA, exhibits a promising performance by predicting efficient scores and best performance outputs for DMUs under evaluation. The study recommends an innovative performance measurement and prediction approach.

Although DEA is a pioneering approach for calculating efficiency and determining benchmarks, it is not capable in demonstrating likely future benchmarks. On comparing the most advanced future benchmarks, the benchmarks that it offers us may still be less accurate. To overcome this shortcoming, an artificial neural network is combined with DEA in this paper to calculate the relative efficiency and reliable benchmarks.

III. LITERATURE REVIEW

3.1 Data Envelopment Analysis

DEA is a popular optimization tool with linear programming as its theoretical basis. DEA recognizes best practice DMUs, measures relative efficiencies, and projects the number of variables which are necessary for making each inefficient DMU efficient. However, despite its popularity in benchmarking studies, DEA has a short coming in prediction capability and limits its usage (Mostafa, 2007; Mostafa, 2009; Wang et al., 2013).

The efficiency of the DMUs calculated using the DEA scores the efficiencies scales between zero (0) to One (1) .The envelopment surface that represents best practices can indicate how inefficient DMUs can improve to become efficient. DEA offers a complete analysis of relative efficiencies for multiple input/ multiple output situations by evaluating the performance of each DMU's. Units lying on the surface are referred to as efficient, while those that do not are inefficient. The efficient reference that includes DMU's, are the peer group for the inefficient units.

Functions of DEA*:*

1. It compares service units by evaluating all resources used and services rendered. It also identifies best practice units (branches, divisions, individuals) and the in efficient units where substantial improvement in performance is necessary. This is achieved by comparing the mix and volume of services provided and the resources used by each unit. These are compared with those of the other units, thus making DEA a very powerful benchmarking technique.

2. DEA calculates the amount and type of cost and resource savings that can be achieved by making each and every inefficient unit as efficient or most efficient or the best practice units.

Mathematically, it is a linear programming approach which maximizes relative efficiency of a DMU by considering the ratio of the weighted sum of inputs to the weighted sum of outputs. An elaborated discussion is provided in the methodology section. Therefore, DEA tries to figure out optimal use of resources to provide the desired output. Using DEA efficiently may help an organization to minimize or avoid wastage of energy, material, time, etc. and can achieve a quality output. Despite having several advantages, like other MCDM approaches DEA also have some shortcomings which makes it vulnerable under certain circumstances. The current trend in DEA research is to increase the efficiency of DEA by hybridizing it with other techniques

3.2 Artificial Neural Network

Neural Networks or an Artificial Neural network (ANN) is a computational algorithm. They intend to simulate the behavior of biological systems composed of "neurons" and are computational models inspired by central nervous systems. It not only recognizes the pattern but is also capable of machine learning. ANN represents a system of interconnected "neurons" that can compute values from inputs.

A neural network is an oriented graph which comprises of nodes that represent neurons (as in biological analogy).The nodes are connected by an arc which corresponds to dendrites and synapses. Every arc is associated with a weight, while each node receives the input and determines the activation function along with the incoming arcs, adjusted by the weight of the arc. A neural network based on the model of a human neuron, is a machine learning algorithm. . The human brain consists of millions of neurons which sends electrical and chemical signals and processes the same. These neurons are connected by a special structure known as synapses which allow neurons to pass signals.

ANN is a technique for processing information that works like a human brain. ANN includes a large number of interconnected processing units working together to process information in order to generate meaningful results.

Neural networks have application in data mining sectors. It can be used in economics, forensics, pattern recognition, etc. It is also useful for classification of large amount of data.

Structure of Neural Network

The structure of the neural network is also referred as ‘neural architecture’ or ‘neural topology’. It mainly consists of three layers known as elementary units. ANN is typically organized layered structure that is made up of various interconnected nodes containing 'activation function'. Artificial Neural network is typically organized in layers that are made up of many interconnected 'nodes' containing 'activation function'. The structure of ANN is shown in Fig.1.

A neural network contains following three layers:

a) Input layer:
The nodes of the input layer are passive and feeds the unchanged data as input to all the nodes of the hidden layer as shown in fig1. The input node receives the values as input of explanatory attributes for each observation.

b) Hidden layer:

The layer next to the input layer as shown in fig1 is known as the hidden layer. Every incoming arc from the node of the input layer is connected to the node of the hidden layer. The outgoing arcs from the hidden layer are connected to the next layer i.e. the output layer. The hidden layers consist of weights; such connections are known as the weighted connections. The incoming values from the input layer are multiplied by the weights. These weights are predefined within the program.

c) Output layer:
The functioning of the output units depends upon the hidden units and the weights between then. The hidden layers are linked to an 'output layer'. Corresponding to the prediction of the response variable the output layer returns an output value. In classification problems, there is generally only one output node. The active nodes combine and change the data to produce the output values.
A proper selection of weight is required to provide efficient data manipulation using the neural network and is different from traditional information processing.

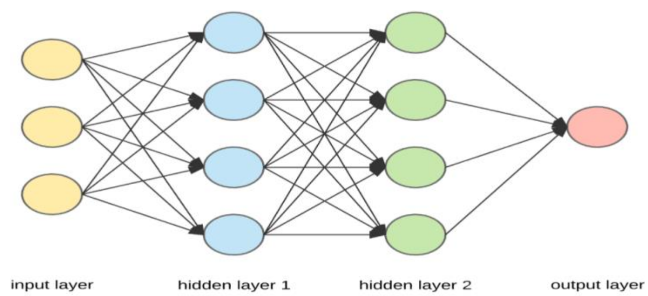

Fig 1: Structure of ANN

Neural Networks - Advantages and Disadvantages:

- It performs well with both linear and nonlinear data with a few exceptions.
- It works even if one or few units fail to respond to network. However, in order to implement large and effective neural networks a large no of processing and storage resources needs to be committed.
- It does not require reprogramming as it uses already analyzed data. They are referred to as black-box a model which provides very little insight into working of these models. The user feeds data as input and trains the neural network and waits for the output.

It is a simple mathematical model to enhance existing data analysis technologies. However, it cannot be compared with the power of the human brain, still acts as a basic building block of Artificial intelligence (AI).

3.3 DE(A)NN APPROACH

In order to overcome the shortcomings of DEA an integrated approach has been used by various researchers in the past like Misiunas, Oztekin, Yao Chen, Chandra (2015); SAR Shah (2017); Azadeh, Saberi, Moghaddam, Javanmardi (2011). These and many others have used the integrated approach to overcome the inefficiencies of DEA.

IV. PROBLEM DEFINITION

4.1 Data envelopment analysis (DEA)

According to research work done by Charnes et al. (1978), the efficiency of the DMU's can be measured as the proportion of the weighted sum of outputs to the inputs.

Efficiency = (weighted sum of inputs) / (weighted sum of outputs) (1)

It can also be written in the form given below to measure the efficiency.

Maximum E_m=

$$[(\sum_{p=1}^{O} Wp\ Output\ p, m)/(\sum_{q=1}^{i} Zp\ Input\ i, m)] \leq 1;$$

n=1, 2, 3,….,m,….N (2)

$0 \leq (\sum_{p=1}^{O} Wp\ Output\ p,m) \,/\, (\sum_{q=1}^{i} Zp\ Input\ i,m) \leq 1;$

$n = 1, 2, 3, \ldots.m, \ldots.N$ (3)

$Wp, Zp \geq 0$; for all p,q (4)

where E_m-m^{th} DMU's efficiency,
p=1 to O, q=1 to I and n=1to N,
$output_{p,m}$- p^{th} output of the m^{th} DMU,
w_p- weight of output $Output_{p,m}$,
$Input_{q,m}$-q^{th} input of m^{th} DMU,
Z_q-weight of $Input_{q,m}$, $output_{p,n}$and $input_{q,n}$are the p^{th} output and q^{th} input ^{respectively} of the n^{th} DMU, where n=1,2,3,…..m,….N.

This fractional model obtained is reduced to a linear model and is solved through a linear programming technique. If there are N numbers of DMU's, then the efficiency of each DMU is maximized relatively. The fractional model shown in equations 2, 3 and 4, is reformed as a linear program shown in equations 5,6,7 and 8.

The general form of CCR DEA model charnes et al. (1978) can be written as:

Max $E_m = \sum_{p=1}^{o} wpOutputp,m$ (5)
s.t.

$\sum_{q=1}^{I} ZpInputq,m = 1$ (6)

$\sum_{p=1}^{O} wpOutputp,n - \sum_{q=1}^{I} ZpInputq,n \ll 0;\ forallp,q$
(7)

w_p, $z_q \geq 0$; for all p,q (8)

The general form of BCC DEA model Charnes et al. (1978) can be written as:

Max $E_m = \sum_{p=1}^{o} wpOutputp,m + Z_{Oq}$ (9)
s.t.

$\sum_{q=1}^{I} ZpInputq,m = 1$ (10)

$\sum_{p=1}^{O} wpOutputp,n - \sum_{q=1}^{I} ZpInputq,n + Z0q \ll 0;\ forallp,q$ (11)
w_p, $z_q \geq 0$; for all p,q (12)

4.2 Artificial neural network

The artificial neural network utilizes feed-forward neural network. The details of the network are mentioned in the literature review section. ANN utilizes the DMU's and the efficiencies obtained from the DEA to obtain the regression plot, train state plot and perform plot. The data obtained are stimulated to obtain the future efficiencies of the DMUs which help in predicting. The software tool utilized for this process is the NN tool of MATLAB much of which is discussed in the methodology section.

V. Methodology

The methodology comprises of 4 steps as mentioned in fig2.

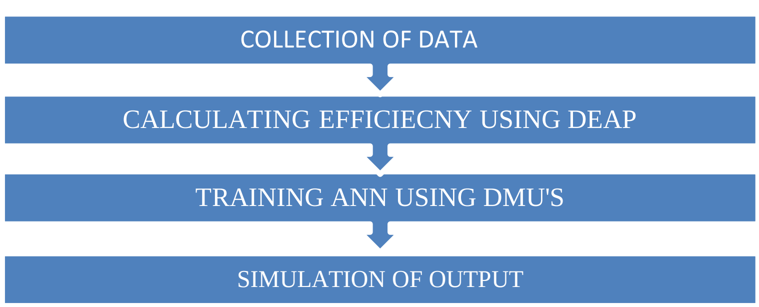

Fig 2: steps of the methodology

Below follows the methodology utilized by the researcher to measure and predicts the efficiency of the UP board secondary education result:

Steps 1: Data collection

The data collected acts as the decision making unit (DMU's) and is used first for calculating the efficiencies using DEAP software and then for training a neural network. Here, total of 9 data sets are collected for every year such that total DMU's for five years is 45. For the best result, it is advised to use data from the government organization or the authorized agency. The data in this paper is collected from the Board of High School and Intermediate Education, U.P. The data sets comprise of the number of males registered, appeared and passed along with females registered, appeared and passed; and total males, females, and students, registered, appeared and passed.

The data obtained above are divided into inputs and outputs as shown below in Table 1:

Table 1: INPUTS and OUTPUTS

INPUTS:	**OUTPUT:**
I1:Number of Males Registered	O1: Number of Males Passed
I2:Number of Females Registered	O2: Number of Females Passed
I3: Total no. of Students Registered	O3: Number of Students Passed
I4: No. of Males Appeared	
I5: No. of Females Appeared	
I6: Total Number of Students Appeared	

Step 2: Calculating efficiency using DEA

In order to calculate efficiency a computer program, known as data envelopment analysis program (DEAP), has been written to conduct data envelopment analysis (DEA). DEA uses linear programming methods to construct a non- parametric piecewise frontier over the data to be able to calculate efficiencies relative to this surface.

This program considers a variety of models. However, standard models i.e. CRS and VRS are involved in the calculation of technical and pure technical efficiencies. These methods are mentioned in Fare, Grosskopf, and Lovell (1994).

In this study, DEAP is used for calculating the CRS (constant return to scale model) and the VRS (variable return to scale model) which measures the efficiencies obtained after analyzing the input and output values.

The DEAP software is then executed to calculate the CRS value that represents the Technical Efficiency (TE) and VRS value that represents the Pure Technical Efficiency (PTE) as shown below:

```
Output orientated DEA
Scale assumption: VRS
Slacks calculated using multi-stage method

EFFICIENCY SUMMARY:
 firm  crste  vrste  scale

   1   1.000  1.000  1.000  -
   2   1.000  1.000  1.000  -
   3   0.966  0.996  0.970 irs
   4   0.884  1.000  0.884 irs
   5   0.979  1.000  0.979 irs

mean   0.966  0.999  0.966

Note: crste = technical efficiency from CRS DEA
      vrste = technical efficiency from VRS DEA
      scale = scale efficiency = crste/vrste
```

The values obtained above i.e. the TE and PTE acts as the Target value while designing ANN.

Step 3: Training neural network

For predicting the future efficiency an artificial neural network is trained using the NN tool of the MATLAB. The Artificial Neural Network can be trained using a number of available softwares. In this study, MATLAB software is employed for this purpose. In the above training, we have used the available 6 inputs and 3 outputs as an input for training the neural network and efficiencies (i.e. TE and PTE) obtained from DEAP are taken as the target value.

The Neural Network designed consists of 9 inputs and 2 output value. It also consists of a 7 neuron in the hidden layer. The numbers of neurons are selected using the formula

$$= \frac{2}{3}\text{(size of input)} + \text{no of output} \qquad (12)$$

The number of hidden neurons should be less than twice the size of the input layer. The neural network does obtained using the NN tool is shown in fig 3.

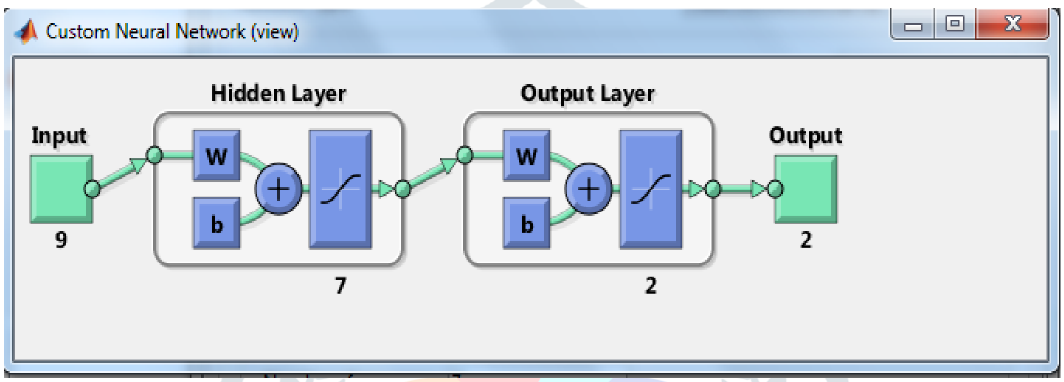

Fig 3: Neural network

After the construction of the neural network through the NN Tool in MATLAB, the next step is to train the network using the input and the outputs as mentioned in table1. Fig 4 and Fig 5 shows training information and parameters.

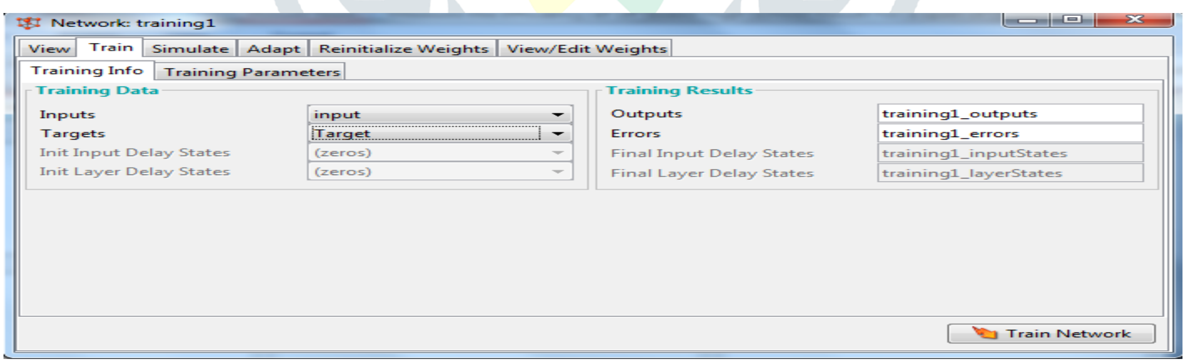

Fig 4: Network Training

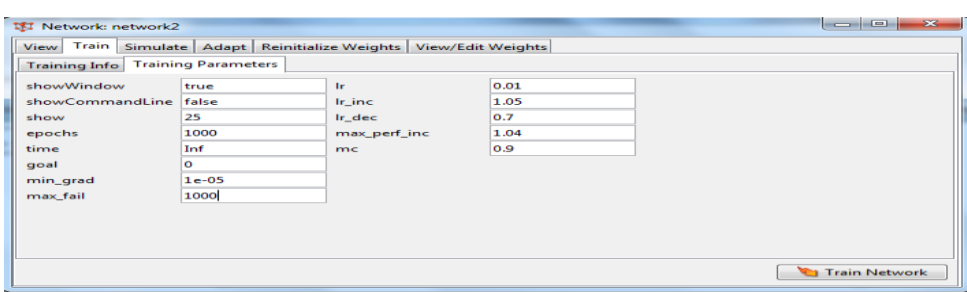

Fig 5: Training parameters

Plots:

The trained network produces the following plots that includes train state, performance plot and regression plot.

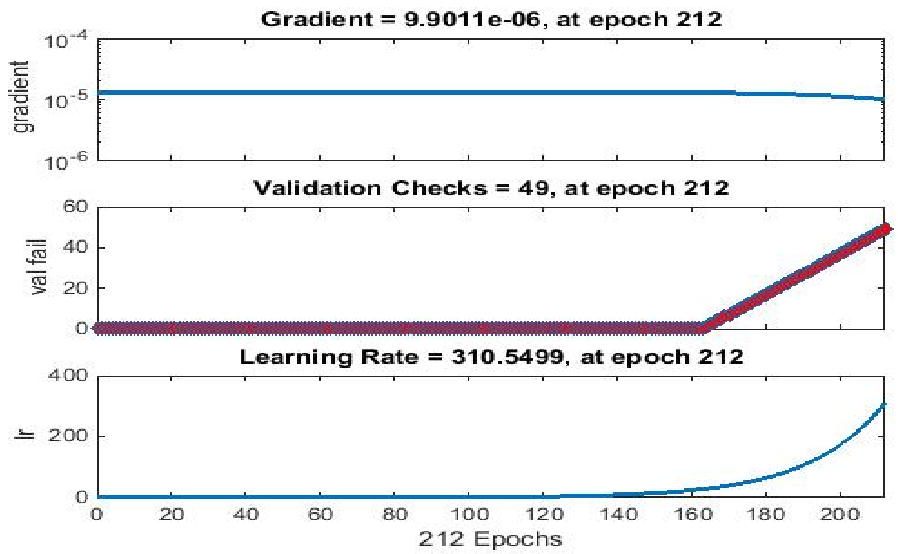

Fig 6: Plot train state

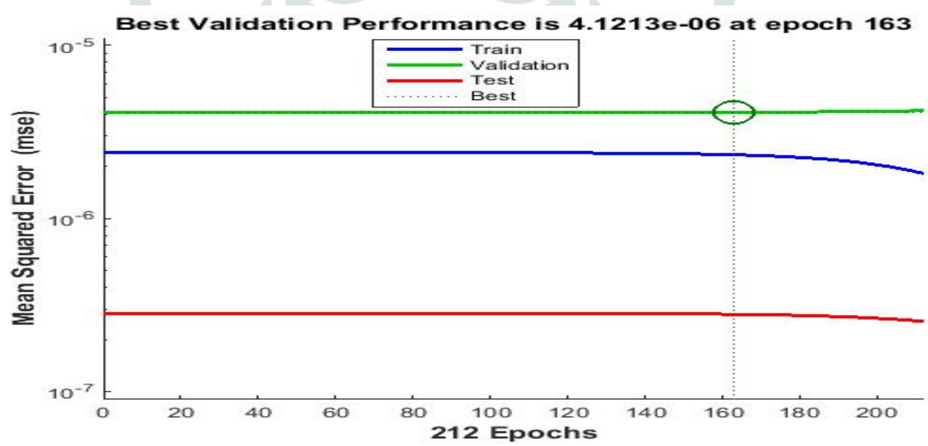

Fig 7: Plot perform

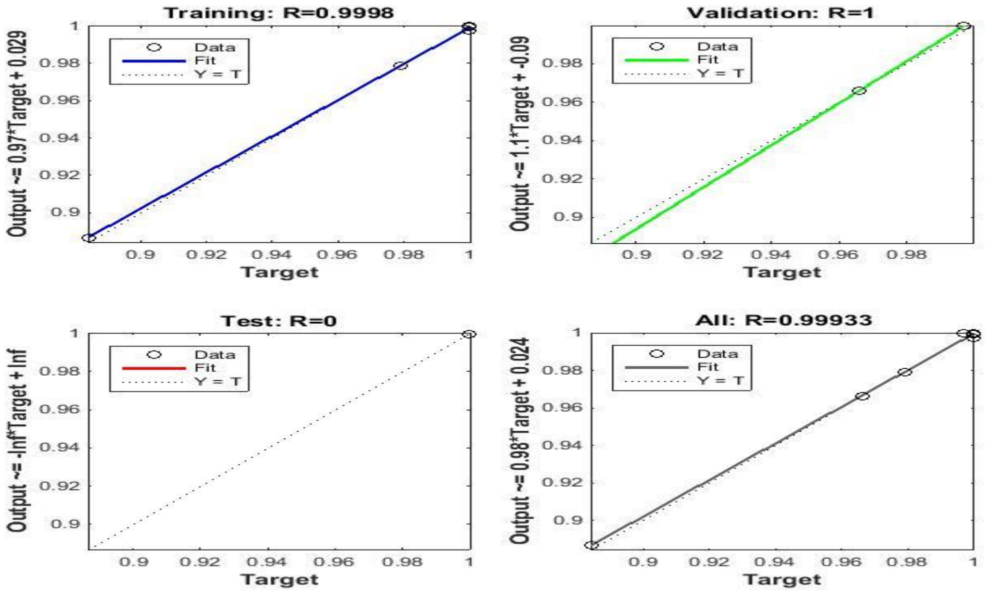

Fig 8: Plot Regression

Step 4: Simulation of outputs

Then the neural network as shown in Fig3 above is used for stimulation of output to obtain predictions which are the ANN efficiencies. The trained ANN network obtained from the above explanation is used for stimulation. The inputs of Table 1 are taken as input for stimulation while none of the value is taken as output. Table 2 contains ANN efficiencies that are obtained from the above stimulation.

Table2: ANN Efficiencies obtained for 5 years

YEAR	ANN 1	ANN 2
2015	0.9974	0.9997
2016	0.9995	0.9995
2017	0.9660	0.9999
2018	0.8866	1.0000
2019	0.9789	0.9994

VI. CONCLUSION:

Table 3 shows the comparison of efficiencies obtained through and ANN. After comparing the efficiencies we reach to a conclusion that the most efficient years of all the taken 5 years is that of the year 2016 followed by 2015.

Thus we see that the result obtained from the above processes gives us a clear idea of the year which is most efficient. Also, we see that the efficiency of the year is not dependent on the percentage of students passed but various factors as mention in Table 1 are also equally important.

Table3: Comparison of DEA and ANN Efficiencies

YEARS	TE	PTE	ANN efficiency for Output 1	ANN efficiency for Output1
2015	1	1	0.997	0.999
2016	1	1	0.999	0.999
2017	0.966	0.997	0.966	0.999
2018	0.884	1	0.887	0.999
2019	0.979	1	0.979	0.999

REFERENCES:

1. Charnes, Abraham, William W. Cooper, and Edwardo Rhodes. "Measuring the efficiency of decision making units." European journal of operational research 2.6 (1978): 429-444.https://personal.utdallas.edu/~ryoung/phdseminar/CCR1978.pdf
2. Shokrollahpour, Elsa, FarhadHosseinzadehLotfi, and Mostafa Zandieh. "An integrated data envelopment analysis–artificial neural network approach for benchmarking of bank branches." Journal of Industrial Engineering International 12.2 (2016): 137-143, DOI: https://doi.org/10.1007/s40092-015-0125-7
3. Natthan Singh & Millie Pant & Amit Goel, (2018). "ANN embedded data envelopment analysis approach for measuring the efficiency of state boards in India," International Journal of System Assurance Engineering and Management, Springer, 2018, DOI : https://doi.org/10.1007/s13198-018-0743-8
4. Hsiang-Hsi Liu*, Tser-Yieth Chen, Yung-Ho Chiu, Fu-Hsiang Kuo(2013) "A Comparison of Three-Stage DEA and Artificial Neural Network on the Operational Efficiency of Semi-Conductor Firms in Taiwan"https://m.scirp.org/papers/abstract/27494
5. Leila Karamali, AzizollahMemariani , Gholam Reza Jahanshahloo , Mohsen Rostamy-malkhalifeh (2013)
6. "Benchmarking by an Integrated Data Envelopment Analysis-Artificial Neural Network Algorithm" https://www.semanticscholar.org/paper/Benchmarking-by-an-Integrated-Data-Envelopment-Karamali-Memariani/18c3b1a5febaec3fca5573cf7fb3fdff27c8d212
7. Mostafa, Mohamed. "Benchmarking top Arab banks' efficiency through efficient frontier analysis." Industrial Management & Data Systems (2007).https://doi.org/10.1155/2018/7625025
8. Mostafa, Mohamed M. "Modeling the efficiency of top Arab banks: A DEA–neural network approach." Expert systems with applications 36.1 (2009): 309-320.
9. Wang et al., 2013
10. Misiunas, Nicholas, et al. "DEANN: A healthcare analytic methodology of data envelopment analysis and artificial neural networks for the prediction of organ recipient functional status." Omega 58 (2016): 46-54. https://doi.org/10.1016/j.omega.2015.03.010
11. Shah, Syyed Adnan Raheel, et al. "Road safety risk evaluation using gis-based data envelopment analysis—artificial neural networks approach" Applied Sciences 7.9 (2017): 886. https://doi.org/10.3390/app7090886
12. Azadeh, Ali, et al. "An integrated data envelopment analysis–artificial neural network–rough set algorithm for assessment of personnel efficiency." Expert Systems with Applications 38.3 (2011): 1364-1373. https://doi.org/10.1016/j.eswa.2010.07.033