Development of Effective Artificial Neural Network Model using Sequential Sensitivity Analysis and Randomized Training

Akshay Daydar

Abstract: As the machine learning algorithms evolve, there is a growing need of how to train the algorithm effectively for the large data with available resources in practically less time. The paper presents an idea of developing an effective model that focuses on the implementation of sequential sensitivity analysis and randomized training approach which can be one solution to this growing need. Many researchers focused on the implementation of sensitivity analysis to eliminate the insignificant features ands reduce the complexity in data selection. These sensitivity analysis methods relatively take a large time for validation through modeling and hence found impractical for large data. On the other hand, the randomized training approach was found to be the most popular approach for training the data but there is a very brief explanation available in research articles on how this training method is meaningful in getting higher accuracy. The current work focuses on the use of sequential sensitivity analysis and randomized training in an artificial neural network (ANN) for high dimensionality thermal power plant data. The sequential sensitivity analysis (SSA) technique includes the use of correlation analysis (CA), Analysis of variance (ANOVA), Akaike information criterion (AIC) in a sequential manner to reduce the validation time for all possible feature combinations. Only selected combinations are then tested against different training methods such as downward extrapolation, upward extrapolation, interpolation and randomized training in ANN. The paper also focuses on suggesting the significance of training with randomized training with comparison-based qualitative reasoning. The statistical parameters, mean square error (RMSE), Mean absolute relative difference (MARD) and R Square (R^2)were accessed for validation purposes. The research work mainly useful in the field of Ecommerce, Finance, industry and in facilities where large data is generated.

Keywords: Artificial Neural network, thermal power plant, correlation analysis, Analysis of variance, Akaike information criterion, training methods.

I. INTRODUCTION

Real-life systems such as thermal power plants are subjected to changes in operation due to fluctuation in demand over a day-time. The flexibility of the system during its operation is necessary to adopt the changes.

Modeling helps to visualize the output of the system and accordingly controlling strategy can be implemented to implement necessary changes. Thermodynamic modeling needs all the input features and this system leads to becoming a complicated system when the non-linearity is involved in output. However, the ANN modeling needs lesser inputs to predict the target value with good accuracy even if the system is non-linear and complex. The mapping of input to output to predict the target value is governed by several factors such as modeling hyper-parameters and input feature combinations. The feature selection strategies in this regard play a very crucial role to reduce redundant features. One of the approaches in this area is sensitivity analysis. The sensitivity analysis allows accessing the input feature based on their impact on the dependent feature and indicates insignificant features. Generally, the sensitivity analysis is performed with mean relative error (MRE) as an evaluation criterion [1],[3]. To simulate each combination in ANN for the determination of evaluation parameters is the most accurate and valid approach to determine the best feature combination but is very time-intensive specifically for data of high dimensionality feature space. The presented approach deals with computing such feature combination/s only by statistical tests unlike traditional sensitivity analysis then simulating only selected combinations in ANN. Further, the effective modeling term defined in the paper indicates to use of techniques that reduce overall computational time and modeling effort. The proposed method, sequential sensitivity analysis can help to save a lot of human effort and computational time by step-wise statistical tests. The tests are namely correlation analysis (CA), Analysis of variance (ANOVA) and akaike information criterion (AIC). In correlation analysis, the parameter, person's correlation coefficient (R) is used to decide the nature and strength of the relationship upon feature combinations [2]. While in ANOVA, the F and p-value quantify the significance of individual features. The AICc test helps to rank the feature combinations from correlation analysis and ANOVA based on delta AICc value. In addition, several training methods such as extrapolation and interpolation are also compared with the randomized training approach to get a detailed insight of model performance against these methods.

The objective of the study is:

1. To identify top-performing input feature combinations by SSA without performing time-intensive ANN modeling simulations.
2. To signify the effectiveness of the randomized training strategy when compared with extrapolation and interpolation strategies.

II. RELATED WORK

The application of artificial neural networks to thermal power plants has seen exponential growth since the late twenties. Specifically, the use of ANN in predicting the energy output becomes a vast area of research due to the availability of high dimensionality data. Sometimes the large data imposes a lot of burden to the limited computational resources while processing such data through layers of neural networks. Some of the general ways to reduce the dimensionality are features selection and train available samples effectively. Smrekar et.al presented one such feature selection approach named sensitivity analysis. In sensitivity analysis, the combination of input features was tested for prediction accuracies against the outputs. Each input was then removed to form a new combination, which was retrained with an identical structure and the same dataset. Input feature combination with minimum MRE was selected as input features for ANN [1]. Later they had proposed 2 stage sensitivity analysis because it was confirmed that 1st stage sensitivity analysis was insufficient in providing confidence for accurate selection. The 2nd stage sensitivity analysis was verified against MRE by adding a new input features combination [3]. Chen Shihe et.al used correlation-based feature selection to develop a model that can counteract fast cut back conditions in thermal power plants. They used the pearson correlation coefficient as a feature to know important features that need manual intervention [4]. Esref Baysal et al carried out the analysis of variance (ANOVA) and regression analysis to determine the effects of process features and optimal factor settings on exergy efficiency in the power plant [5]. Hence the statistical tool ANOVA can be used to identify the most influential features by reliable mathematical outcomes.

Hui Peng et.al presented a modeling approach for predictive control in thermal power plants by considering a load-dependent exponential model. The selection for the better model was performed by using the akaike information criterion [6]. AIC can be effectively applied in large data sets where the models are compared using maximum likelihood estimation [7]. The proposed approach includes all these selection techniques sequentially to make a reliable selection.

While dealing with ANN, the range of data selection for training always plays a pivotal role in deciding the degree of model fit. Smrekar et al presented an approach for training data selection from the low load, medium load and high load conditions. Interpolation and extrapolation were used to test the model with different load conditions. Downward extrapolation selects a higher range of data while the upward extrapolation takes account of lower ranges to predict for lower and higher ranges respectively. Interpolation on other hand takes a mixture of higher and lower ranges to predict the intermediate range. Interpolation gave better results since it has taken the account of all the variations in the dataset for a given problem. Quality of training dataset

matters, which was further proved by the performance of upward extrapolation over downward extrapolation [3]. The paper presents compare all these approaches with randomized training. For the randomized selection strategy, the samples are selected based on the random number generated within the range of sample size.

III. METHODOLOGY

The sequential sensitivity analysis approach includes a selection of input features first from correlation analysis then with ANOVA and the final selection is made with the AIC test as indicated in figure 1.
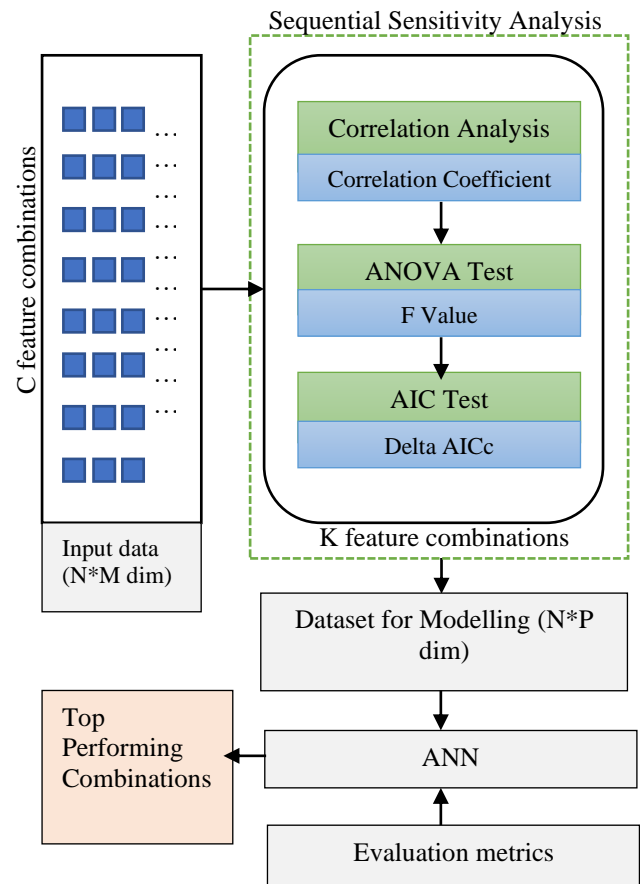

Fig. 1. The sequential sensitivity analysis method. N is the number of samples per M features. The N*P denotes the reduced dataset size after SSA per ANN simulation.

The initial C feature combinations (with a dataset of N*M dim) need not be simulated through the ANN for evaluation instead the K combinations (N*P dim) from SSA are exercised. The P (total number of features in a combination including dependent feature) is equal to M when all independent features of the N*M set are present in a combination.

Correlation Coefficient (CC) is a statistical tool used to study correlations between the set of features.

Correlation coefficient R is used to learn a relationship between two features and then the user can make a decision based on these relationships [8].

In this study, pearson's-CC was accessed for the first level of sensitivity analysis for one to one feature combinations.

The combination of type one to one indicates one dependent with one independent feature combination. Pearson's coefficient gauges the linear relationship between two features and the relationship is located between -1 and 1. The CC obtains the highest value i.e., +1, if the feature is positively correlated in a relationship and on the other hand CC, obtains -1 if the feature is negatively correlated in a relationship [9]. The relationship is given as follows:

$$R = \frac{\sum(x_i - x_m)(y_i - y_m)}{\sqrt{\sum(x_i - x_m)^2 \sum(y_i - y_m)^2}} \tag{1}$$

Where R = Correlation coefficient between x and y
x_i = Samples of first feature
y_i = Samples of second feature
x_m = Sample mean for first feature
y_m = sample mean for second feature

The second stage of feature selection was executed by ANOVA. The ANOVA is a statistical test for estimating the quantitative effect of feature combinations on the independent features. ANOVA tests whether there is a difference in means of the groups at each level of the independent feature termed as null hypothesis (H0) test [10]. The foremost assumption in the test is that all the groups (features) are distributed normally. The paper presents the 2-way ANOVA and 3-way ANOVA test for energy output as a single dependent feature.

Main notable things while performing a test for ANOVA are:

1. Degree of freedom: This indicates the total number of logically independent values in the data sample. The Df is calculated from degrees of freedom (Df_1) for the independent feature (Df_2) and the degrees of freedom for the residuals [11].

Df_1= K-1 (K=1…j…. k)

Df_2= N-K (N=1…i…. n)

N is the total number of readings

K is the number of independent groups

2. Sum of square error (Sum Sq): It is calculated from the sum of squares between treatments (SSB) and residuals (SSE). The Sum Sq is the addition of SSB and SSE [11].

$$SSB = \sum n_j(x_{jm} - x_m)^2 \tag{2}$$

$$SSE = \sum \sum (x_i - x_{jm})^2 \tag{3}$$

Where, x_{jm} is mean of jth treatment, x_m is overall mean, x_i is i^{th} sample in the treatments, n_j is sample size per treatment

3. The Mean square error (Mean Sq): It is the sum of the mean of the sum of the square between treatments (MSB) and residuals (MSE). It is calculated by dividing the sum of squares by the degrees of freedom for each feature [11].

$$MSB = \frac{SSB}{DF_1} \tag{4}$$

$$MSE = \frac{SSE}{DF_2} \tag{5}$$

4. The F-value: It is the test statistic from the F test. The larger the F value, the more likely it is that the variation caused by the independent feature is real and not due to chance [11].

$$F\ Value = \frac{MSB}{MSE} \tag{6}$$

5. The Pr(>F) value: It is the P-value of the F-statistic. This shows how likely it is that the F-value calculated from the test would have occurred if the null hypothesis of no difference among group means were true [11].

The AIC test is conducted after the ANOVA to rank the selected combinations. The ranking to the combination is decided based on the AICc value or delta AICc value. The AICc score is indicative of the greatest amount of variation with the lowest possible independent features. Hence the combination with the lowest possible AICc value indicates the best combination. The delta AICc is calculated by taking the difference between the best combination (that has a maximum information content and minimum AICc value) with the other possible combinations. The AIC is the mathematical testing method to compare models even with different data distributions which makes it so effective [12]. AIC determines the relative information value of the model using the maximum likelihood estimate and the number of features (independent features) in the model as follows.

$$AICc = 2K - 2In(L) \tag{7}$$

$$Delta\ AICc = AICc_i - min\ AICc \tag{8}$$

K is the total number of independent features plus 2 as constant, while L is the log-likelihood estimate [13]. The combination with very high delta AICc was strongly rejected in the presented study.

The training methods, interpolation and extrapolation were executed by sorting the data in ascending/ descending order with the same train to test ratio. The comparison was then made with randomized training. The test and train split were made based on sample size and not on load condition as indicated by Smrekar et.al [3].

IV. EXPERIMENTATION

The Dataset exercised for analyzing the proposed method was taken from the combined cycle power plant (CCPP). The dataset contains 9568 samples with 5 features collected from a CCPP over 6 years of operation when the power plant was set to work with a full load. The measurements were taken every second [14]. The details of the dataset are indicated in Table I.

Table I: Details CCPP dataset

	Features	Unit	Min	Max	Mean
1	Energy output	MW (per Hr.)	420.26	495.76	454.36
2	Temperature	Degree Celsius	1.81	37.11	19.65
3	Exhaust vacuum	Cm Hg	25.36	81.56	54.3
4	Ambient pressure	Millibar	992.89	1033.3	1013.25
5	Relative humidity	Percentage	25.26	100.16	73.3

The Energy output was considered as a dependent feature while other features were independent once. The data cleaning task was performed to remove the non-numeric data by assigning them zero values. Then the preprocessing task was performed on the dataset which includes data scaling within the range of 0.1 to 0.9 using the min-max normalization approach. The total number of possible feature combinations for predicting energy output is indicated in Table II. From the table, there are a total of 4 one to one, 6 one to two and 3 one to three combinations. The combination of type Y to X indicates the number of dependent to the number of independent feature combinations. Y is the energy output in the present study.

Table II: Feature combinations for SSA

01	Temperature (T)
02	Exhaust vacuum (E)
03	Ambient pressure (A)
04	Relative humidity (R)
05	Temperature and Exhaust vacuum (TE)
06	Temperature and Ambient pressure (TA)
07	Temperature and Relative humidity (TR)
08	Exhaust vacuum and Ambient pressure (AR)
09	Exhaust vacuum and relative humidity (EA)
10	Ambient pressure and relative humidity (ER)
11	Temperature, Exhaust vacuum, Ambient pressure (TEA)
12	Temperature. Exhaust vacuum, Relative humidity (TER)
13	Temperature. Ambient pressure, Relative humidity (TAR)
14	Exhaust vacuum, Ambient pressure, Relative humidity (EAR)
15	Temperature, Exhaust vacuum, Ambient pressure, Relative humidity (TEAR)

The dataset of N*M dimensionality was given as input to the SSA to significantly reduce the size of the dataset. The correlation analysis was performed to select first-level combinations that constitute one dependent and one independent feature termed as one to one feature combination. As indicated in fig.2, the energy output is highly correlated with the temperature, exhaust vacuum while its correlation is moderate to weak with ambient pressure and relative humidity.
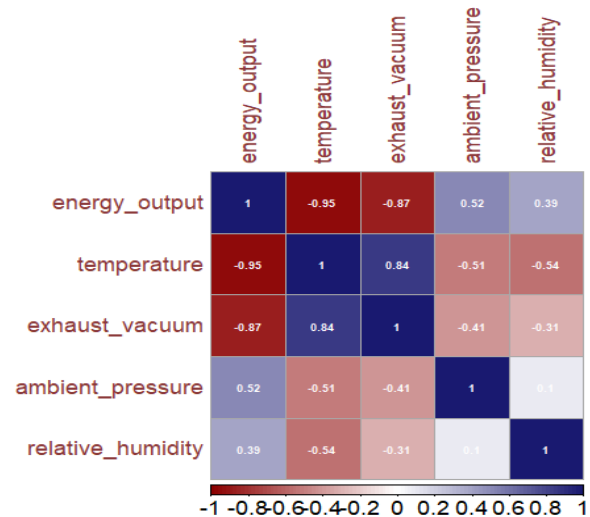
Fig. 2. Correlation plot for first stage SSA method

Hence the T and E were selected as the candidates for AIC test from one to one feature combination. Since nothing can be said about other feature combinations, we need to carry the ANOVA statistical test. The two-way and three-way ANOVA was performed on all the feature combinations except one to one feature combinations. The F value was regarded as an elimination/ selection parameter for feature combinations. From ANOVA table III (full ANOVA table is given in annexure-I) it can be observed that combinations TE, TA, TR has a high sum of F values. Hence the feature combinations TE,TA and TR were selected for the AIC test. However, from the ANOVA test, it cannot be concluded with confidence which feature combination is better than the other. Hence to fine-tune our selection the AIC test was performed to make the final selection. The AIC is used to compare all possible feature combinations even with different data distribution. For this reason, AIC is ideally suited to generalize linear modeling applications.

The ranking for combinations in AIC test was performed based on AICc and delta AICc value. The minimum value of AICc was observed for TEAR while the maximum was for E.

Table III: Combination selection based on F-values

F value	Feature1	Feature2	Feature3	Sum of F values
2-way ANOVA				
TE	102033.7	1904.797		103938.5
TA	86688.53	179.8223		86868.35
TR	108769.4	2661.979		111431.4
AR	4175.189	1794.321		5969.51
EA	33960.46	1365.038		35325.5
ER	31739.11	650.1049		32389.21
3-way ANOVA				
TEA	104816.6	1956.75	261.8824	107035.2
TER	120035.1	2240.854	1688.518	123964.5
TAR	108833	225.7577	2444.371	111503.1
EAR	36907.1	1483.478	830.9249	39221.5

15

Table IV: Combination selection based on Delta AICc values

Combinations	K	AICc	Delta AICc	AICc Wt.
TEAR	6	56188.24	0	1
TER	5	56229.25	41.01041	0
TAR	5	57166.62	978.3793	0
TR	4	57171.21	982.9703	0
TEA	5	57526.4	1338.163	0
TE	4	57782.87	1594.63	0
TA	4	59342.27	3154.034	0
T	3	59518.48	3330.241	0
E	3	67932.62	11744.38	0

Also, it can be observed from Table IV, that the TEAR has a zero delta AICc score followed by the TER (41.01). The scores for other combinations are very far from the TEAR and TER hence rejected for further consideration. However, the validation of the final combination was also performed using a randomized training strategy (see Table VI).

The step-by-step process for sequential sensitivity analysis is as follows:

1. Identify the correlation between dependent feature with each independent feature and select those feature combinations for the AIC test which shows a high Pearson's correlation coefficient R from one-to-one feature combination.
2. Perform the ANOVA test for other than one to one feature combination and select those combinations for the AIC test which combinedly give F score in *acceptable range.
3. Perform the AIC test for the selected combinations and finalize those combinations which give delta AICc value in *acceptable range.

Note- For more than one dependent feature the same procedure can be repeated.
*The choice of acceptable range is user-dependent.

Table V: Combination selection at different stages

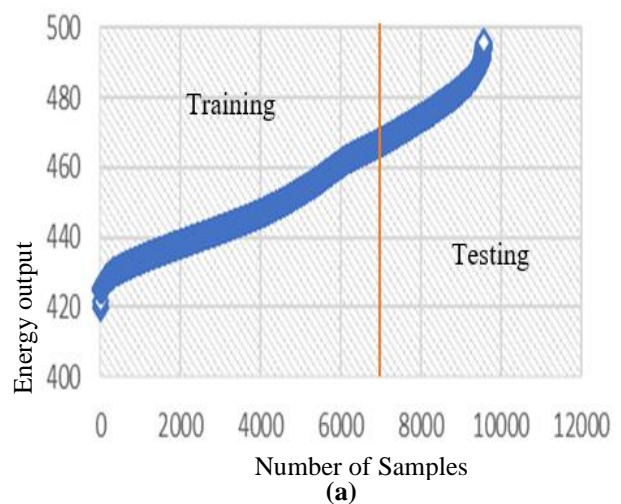| Total Possible Combinations	Correlation Analysis	SSA		ANN Modelling
		ANOVA	AIC	
T	T		⊗ T	
E	E		⊗ E	
A	⊗ A			
R	⊗ R			
TE		TE	⊗ TE	
TA		TA	⊗ TA	
TR		TR	⊗ TR	
AR		⊗ AR		
EA		⊗ EA		
ER		⊗ ER		
TEA		TEA	⊗ TEA	
TER		TER	TER	⊗ TER
TAR		TAR	⊗ TAR	
EAR		⊗ EAR		
TEAR			TEAR	TEAR

⊗   Indicates rejected combination

The dataset was split into train and test in the ratio of 7:3. The term test set was used in place of the validation set since a separate test set was not explicitly used to verify the performance of the model. Four approaches were exercised in the process such as upward and downward extrapolation, interpolation, randomized training.
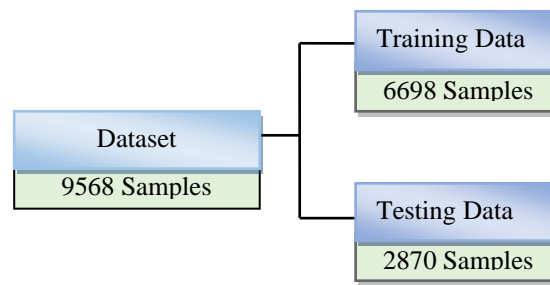
Fig. 3 Train test split for modeling

In upward extrapolation, the dataset was split to predict the data values for the upward range. At first, the dataset was arranged based on ascending order for the energy output. Then, the dataset was split in the ratio of 7:3, for training and testing. In the case of downward extrapolation, the dataset was split to predict the data values for the downward range. At first, the dataset was arranged based on descending order for the energy output. Then, the dataset was split in the ratio of 7:3, for training and testing. While in interpolation the dataset was split to predict the data values between the intermediate range. At first, the dataset was arranged based on ascending order for the energy output. Then the first and last 3348 samples (together constitute 6698 samples) were considered for training and the remaining samples (that lie in the middle range) for testing. Unlike the approaches discussed above, in the randomized training, the dataset was split randomly irrespective of its order. The split includes data from all the load ranges.

The selected combinations were used as inputs for ANN modeling. The performance was compared for eight ANN models (two selected combinations were tested for four training approaches). ANN (with single hidden layer, 5 hidden nodes, logistic activation function, backpropagation training mechanism, $1*10^{12}$ max. epochs) models with the same feature set were compared based on the RMSE, MARD and R^2 evaluation features. While assigning feature values it was considered that due to very few
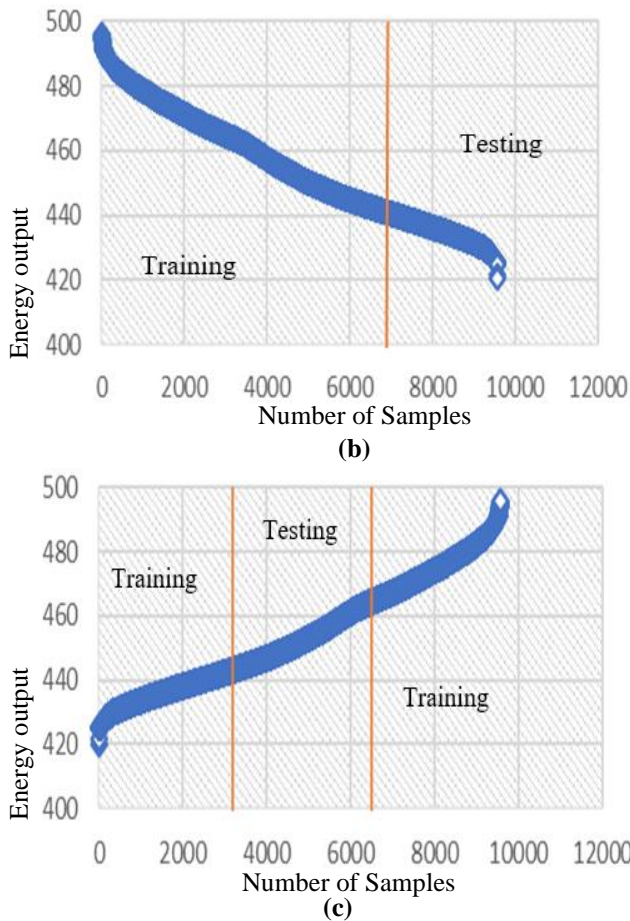

(a)

(b)

(c)

Fig. 4 Data selection with (a) upward extrapolation, (b) downward extrapolation, (c) interpolation for training

modeling features, a single hidden layer with hidden nodes equal to the number of features will be sufficient to make a generalized model. The number of epochs was taken very high indicating the frequent weight and bias assignment according to the gradient descent rule. The logistic activation function was utilized as it suits well for a time series dataset with efficient non-linear mapping.
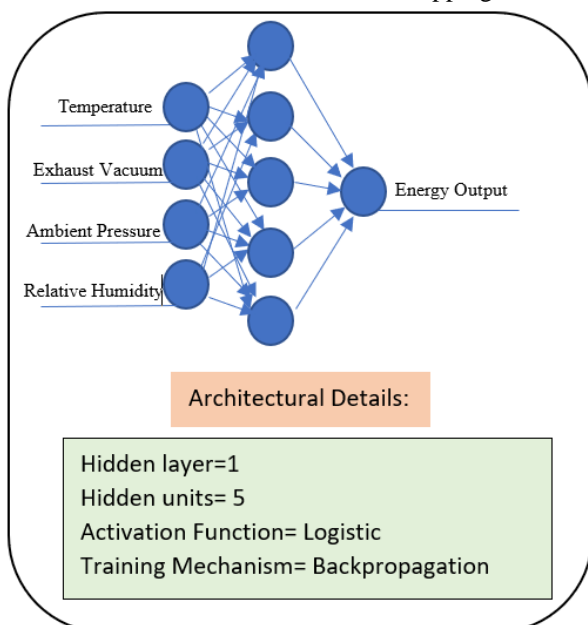

Fig. 5 The ANN architecture

V. RESULTS

The results observed from experimentation are as follows:
1. The SSA was found out to be well-matched with the traditional sensitivity approach (where every combination is simulated in machine learning model for determination of evaluation parameters) in terms of determining the best feature combinations. However, the SSA did not rank feature combinations correctly (based on R^2 value) i.e. in the same order of traditional approach but it helped to determine the top-performing candidates (see Table VII).

Table VI: Result from traditional sensitivity analysis for randomized training

| Comb | TEST | | | TRAIN | | |
	RMSE	MARD	RSquare	RMSE	MARD	RSquare
T	0.042	0.112	0.91	0.041	0.110	0.912
E	0.063	0.153	0.801	0.064	0.156	0.793
A	0.123	0.308	0.287	0.122	0.312	0.303
R	0.139	0.362	0.155	0.139	0.365	0.153
TE	0.037	0.098	0.929	0.037	0.099	0.929
TA	0.041	0.107	0.915	0.040	0.105	0.918
TR	0.037	0.098	0.929	0.037	0.098	0.93
AR	0.112	0.276	0.408	0.111	0.277	0.419
EA	0.058	0.138	0.827	0.057	0.139	0.827
ER	0.061	0.149	0.814	0.062	0.152	0.807
TEA	0.036	0.095	0.931	0.035	0.092	0.936
TAR	0.037	0.096	0.931	0.037	0.097	0.930
EAR	0.054	0.130	0.846	0.054	0.131	0.844
TER	0.0357	0.092	0.936	0.036	0.094	0.936
TEAR	**0.0347**	**0.089**	**0.939**	**0.034**	**0.091**	**0.939**

Table VII Comparison of traditional Sensitivity analysis with SSA

Traditional sensitivity analysis	SSA approach
TEAR	TEAR
TER	TER
TEA	TAR
TAR	TR
TR	TEA

2. The SSA method saved a sufficient amount (80%) of time compared to the traditional sensitivity analysis approach as indicated in Table VIII. However, the claim of increased computation due to SSA is not significant as compared to an increase in time complexity due to an increase in feature size.

Table VIII: Time Comparison

Criterion	Time consumption
All combinations	583.79 sec (Approx. 10 mins)
Selected combinations	118.46 (Approx. 2 mins)
Time-saving (in %)	**79.73 (80)**

3. The randomized training strategy for TEAR was found out to be the best compared to TER as indicated in fig 6.

17

Also, the randomized training resulted superior with any other training methods as indicated in Table IX. The training results of TEAR for interpolation (RMSE=0.0312, R^2=0.963) reported better among others.
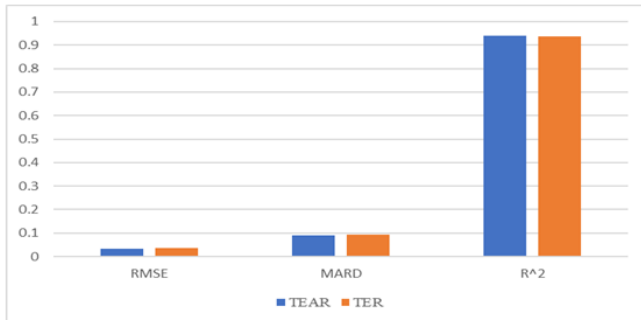

Fig. 6. The results of TEAR and TER for Randomized training

Based on the test results, TEAR with (RMSE=0.0347, MARD=0.0899, R^2=0.939) randomized training method was considered suitable for future prediction tasks.

4. Test for overfitting (positive difference between test and train result for evaluation metrics) was resulted in getting highest overfitting for downward extrapolation (RMSE difference =0.0584, MARD difference= 0.329, R^2 difference=0.821) and minimum for randomized training approach (RMSE difference=0.0002, MARD difference=0.0012, R^2 difference=0) for TEAR as shown from fig.7. Hence the randomized training was observed robust to overfitting and hence strongly recommended for training of large data sets.

Table IX: Comparison of TEAR and TER for test and train data

TEST DATA			
	RMSE	**MARD**	**R^2**
Upward Extrapolation			
TEAR	0.180	0.248	0.381
TER	0.183	0.250	0.481
Downward Extrapolation			
TEAR	0.090	0.390	0.098
TER	0.081	0.352	0.163
Interpolation			
TEAR	0.048	0.113	0.657
TER	0.049	0.113	0.606
Randomized			
TEAR	**0.034**	**0.089**	**0.939**
TER	0.035	0.092	0.936
TRAIN DATA			
	RMSE	**MARD**	**R^2**
Upward Extrapolation			
TEAR	0.032	0.102	0.863
TER	0.035	0.112	0.831

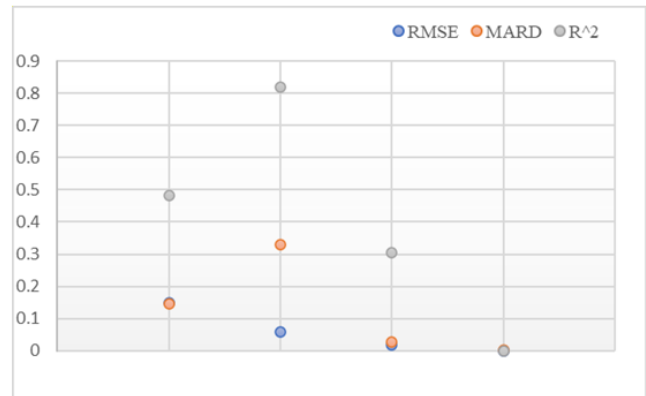| *Downward Extrapolation*			
TEAR	0.032	**0.061**	0.919
TER	0.035	0.067	0.903
Interpolation			
TEAR	**0.031**	0.085	**0.963**
TER	0.033	0.091	0.958
Randomized			
TEAR	0.034	0.091	0.939
TER	0.036	0.094	0.936

Fig. 7. The results of the difference in test and train data for evaluation metrics

VI. DISCUSSION

The article illustrates the significance of sequential sensitivity analysis in an artificial neural network with a randomized training approach. The results obtained from correlation analysis applied only to a one to one feature combination because of its inability in computing an interaction effect between different features in a combination. In the case of ANOVA, the p-value of all these results was closer to zero indicating rejection of the null hypothesis. However, the selection was made based on cumulative F value which is indicative of the significance of that factor upon combination. The AIC is a probabilistic testing method used to rank the models according to delta AICc values. The AIC not only ranks the combinations but quantifies the strength of information content as well [13]. Based on minimum delta AICc values, TEAR and TER combinations were selected for ANN simulation. All these tests assume the raw data were distributed normally. Hence the use of logistic activation function in ANN was justified since the nature of logistic probability distribution is nearly the same as the normal distribution except it has a longer tail. Multiple statistical tests were performed to assure sound selection at the end and build enough confidence for their future use. Further, after each test, the range of feature combinations was selected which provide a cushion for unexplainable behavior due to modeling.

The user-defined cutoff for the selection of feature

18

combinations in each test was purely based on a higher F value and a very low delta AICc value. However, the cutoff can be manipulated based on the user and the availability of computational resources. In terms of time management, the current SSA strategy can save a lot of computational time as compared to traditional sensitivity analysis and specifically in the areas where a lot of data is generated with very little computational facilities to process it. The results from the ANN modeling indicated that TEAR combination with randomized training approach gave the best (test) result. For TEAR, it can be observed that during the training phase the interpolation outperformed other combinations since the model learned the variations from both low and high ranges of the dataset very well. The same performance for the test set was not obtained, like randomized training. The possible reason for this problem might be overfitting. Overfitting occurs when the algorithm tries to learn the details and noise in the training data which decays its performance for new data. Hence the randomized training can be one solution to overfitting. The overfitting was also found to be less in TEAR with randomized training making it a suitable candidate for general model setting. While downward and upward extrapolation techniques generated poor results because models were trained either at lower/ higher limits only where the limited data variation anticipated hence strongly rejected for future use in modeling.

The limitation to the current study was that the results were observed for comparatively fewer features than in many real-life applications. The SSA method does not rank the feature combinations correctly, but it did well in accessing better-performing combinations. However, the selection of at least 3 to 4 combinations is recommended after each test to avoid any uncertainties. The computational efficiency was quite effective for SSA and can be tested for a large feature set in the future.

VII. CONCLUSION

The paper illustrates how the effective model can be developed using sequential sensitivity analysis and a randomized training approach for large data-driven modeling. It highlights the effectiveness of sequential sensitivity analysis in terms of reduction in the validation time and quantitative feature selection. In addition, the work emphasizes the use of randomized training strategy to train the model for better learning. However, the work can be extended to test accuracy using different modeling methods, study the interaction effect between features and test applicability for very high dimensionality feature space.

DATA AVAILABILITY

The CCPP (Combined cycle power plant) data is open access dataset and available at the Neural Network designer website. Link: https://www.neuraldesigner.com/learning/examples/combined-cycle-power-plant

APPENDIX

APPENDIX-I: 2-way and 3-way ANOVA on different feature combinations.
(Res- Residual)

2-way ANOVA						
Model name	Dsc.	Df	Sum Sq	Mean Sq	F value	Pr(>F)
TE	T	1	2505095	2505095	102033.7	0
	E	1	46765.93	46765.93	1904.797	0
	Res	9565	234836.6	24.55166	NA	NA
TA	T	1	2505095	2505095	86688.53	0
	A	1	5196.44	5196.442	179.8223	1.23E-40
	Res	9565	276406.1	28.89766	NA	NA
TR	T	1	2505095	2505095	108769.4	0
	R	1	61308.69	61308.69	2661.979	0
	Res	9565	220293.8	23.03124	NA	NA
EA	E	1	2108187	2108187	33960.46	0
	A	1	84738.4	84738.4	1365.038	1.91E-279
	Res	9565	593773	62.07768	NA	NA
ER	E	1	2108187	2108187	31739.11	0
	R	1	43181.51	43181.51	650.1049	8.53E-139
	Res	9565	635329.9	66.42236	NA	NA
AR	A	1	748977.1	748977.1	4175.189	0
	R	1	321878.8	321878.8	1794.321	0
	Res	9565	1715842	179.3876	NA	NA
3-way ANOVA						
Model name	Dsc.	Df	Sum Sq	Mean Sq	F value	Pr(>F)
TEA	T	1	2505095	2505095	104816.6	0
	E	1	46765.93	46765.93	1956.75	0
	A	1	6258.937	6258.937	261.8824	3.93E-58
	Res	9664	228577.7	23.8998	NA	NA
TER	T	1	2505095	2505095	120035.1	0
	E	1	46765.93	46765.93	2240.854	0
	R	1	35238.86	35238.86	1688.518	0
	Res	9564	199597.7	20.86969	NA	NA
TAR	T	1	2505095	2505095	108833	0
	A	1	5196.442	5196.442	225.7577	1.88E-50
	R	1	56264	56264	2444.371	0
	Res	9564	220142.1	23.01778	NA	NA
EAR	E	1	2108187	2108187	36907.1	0
	A	1	84738.4	84738.4	1483.478	7.61E-302
	R	1	47463.62	47463.62	830.9249	2.74E-175
	Res	9564	546309.4	57.12143	NA	NA

REFERENCES

1. Smrekar, J., D. Pandit, M. Fast, M. Assadi and Sudipta De, *"Prediction of power output of a coal-fired power plant by artificial neural network"*, Neural Computing and Applications, 19, 2009, 725-740.
2. Author: Jason Fernando, Available: Https://www.investopedia.com/terms/c/correlationcoefficient.asp
3. Smrekar J., Assadi M., Fast M., Kustrin I., De S, *"Development of artificial neural network model for a coal-fired boiler using real plant data.* Energy", 2009, 34 (2), 144-152.
4. Chen Shihe 1, Luo Jia 1, Zhang Xi 1, Huang Weijian 1,Wen Xiaoqiang2, Sun Lingfang., *"Correlation analysis and modeling of main features in thermal power unit under FCB Conditions"*, Applied Mechanics and Materials, 2014, 1611-1614.

19

5. Baysal, Eşref & Bilginsoy, A.K. & Acır, Adem, "*Parametric optimization on exergy analysis of a thermal power plant using Taguchi design method*", Energy Education Science and Technology Part A: Energy Science and Research, 29, 2012, 1313-1326.

6. Peng, Hui & Ozaki, Tohru & Haggan-Ozaki, Valerie & Toyoda, Yukihiro, "*A nonlinear exponential ARX model-based multifeature generalized predictive control strategy for thermal power plants*", Control Systems Technology, IEEE Transactions, 256 - 262.

7. Author: Alexandre Zajic, Available: https://towardsdatascience.com/introduction-to-aic-akaike-information-criterion-9c9ba1c96ced

8. Mohammed, M., A. Shawkat and T. Kevin, "*Improved C4.5 algorithm for rule-based classification. Proceedings of 9th Artificial Intelligence*", Knowledge Engineering and Database Conference AIKED'10, 2010, 296-301.

9. Elssied, Nadir & Ibrahim, Assoc Prof. Dr. Othman & Hamza Osman, Ahmed, "*A Novel Feature Selection Based on One-Way ANOVA F-Test for E-Mail Spam Classification. Research Journal of Applied Sciences*", Engineering and Technology, 7, 2014, 625-638.

10. Chandrasekharan, Sreepradha & Panda, Rames & Swaminathan, Bhuvaneswari., "*Statistical modeling of an integrated boiler for coal-fired thermal power plant*", Heliyon. 3, 2017, e00322.

11. Author: Rebecca Bevans, Available: https://www.scribbr.com/statistics/anova-in-r/

12. Cavanaugh, Joseph & Neath, Andrew, "*The Akaike information criterion: Background, derivation, properties, application, interpretation, and refinements*", Wiley Interdisciplinary Reviews: Computational Statistics, 11, 2019.

13. Author: Rebecca Bevans, Available: https://www.scribbr.com/statistics/akaike-information-criterion/

14. Kaya, Heysem & Tufekci, Pinar, "*Local And Global Learning Methods For Predicting Power Of A Combined Gas & Steam Turbine*", International Conference on Emerging Trends in Computer and Electronics Engineering, 2012, 13-18.

AUTHORS PROFILE

Akshay Daydar, is currently pursuing a Ph.D. at the Indian Institute of Technology Guwahati in the Department of Mechanical Engineering. He has got his master's degree from the National Institute of Technology, Warangal in the computer integrated manufacturing specialization. His research activity mainly focuses on machine learning, data analytics and deep learning.