UJEAS

UNIZIK JOURNAL OF ENGINEERING AND APPLIED SCIENCES

UNIZIK Journal of Engineering and Applied Sciences 6(3), September (2026), 3767-3780
Journal homepage: https://journals.unizik.edu.ng/index.php/ujeas
PRINT ISSN: 2992-4383 || ONLINE ISSN: 2992-4391

MODELING AND OPTIMIZATION OF THE NOZZLE POSITION OF A CROSS-FLOW HYDRO TURBINE USING BOX-BEHNKEN DESIGN AND MACHINE LEARNING TECHNIQUES

Obiora Clement Okafor*
Department of Mechanical Engineering, Imo State University, Owerri, Nigeria
Chinonso Hubert Achebe
Department of Mechanical Engineering, Nnamdi Azikiwe University, Awka, Nigeria
Jeremiah Chukwuneke
Department of Mechanical Engineering, Nnamdi Azikiwe University, Awka, Nigeria

**Corresponding Author's E-mail: okaforobiorac@gmail.com*

Abstract

This study employed the Box-Behnken design and machine learning techniques to model and optimize the nozzle position of a cross-flow turbine system. The Box-Behnken design was used to structure the experiment, with nozzle height (365 mm–461 mm), nozzle distance (102 mm–202 mm), and attack angle (0°–40°) as the predictors (independent factors), and efficiency, shaft power, and runner speed as the response variables. A total of 17 experimental runs were obtained from the Box-Behnken design. The experimental design was meticulously followed. The runner speed was measured using a digital tachometer and recorded for each nozzle positional setting, while the shaft power and efficiency were calculated using the appropriate formulae. The responses were optimized, modeled, and analyzed statistically using ANOVA to assess how the nozzle positional factors individually and interactively affect the responses. The results of the RSM optimization were validated using the machine learning algorithm, Sequential Least Squares Programming (SLSQP). Additionally, the responses of the cross-flow turbine were modeled using the Multilayer Perceptron (MLP) algorithm, and its prediction performance was compared to that of the RSM model. This comparison was conducted both graphically and statistically using performance indices such as mean bias error (MBE), root mean square error (RMSE), Nash-Sutcliffe efficiency (NSE), t-statistic, coefficient of correlation (r), and coefficient of determination (R^2). The study results showed that each nozzle positional factor influences the efficiency, shaft power, and runner speed of the CFT system. Additionally, it was observed that the nozzle height and attack angle interactively affect the CFT responses. Using the RSM optimization technique, the optimal nozzle position was determined to be: nozzle height = 408.4mm, nozzle distance = 102mm, and attack angle = 5°. At this position, the predicted response values were: efficiency = 91.91%, shaft power = 242.367W, and runner speed = 406rpm. The SLSQP technique validated the RSM findings by identifying a similar optimal nozzle position: nozzle height = 407.99mm, nozzle distance = 102mm, and attack angle = 4.699°. At this position, the predicted response values were: efficiency = 91.962%, shaft power = 242.599W, and runner speed = 405.99rpm. Furthermore, holistic modeling of the CFT system demonstrated that both RSM and MLP models performed well in estimating CFT responses. However, the RSM model surpassed the MLP model, as it had lower MBE, RMSE, and T-statistic values and higher NSE, r, and R^2 values. Consequently, the developed RSM model is recommended for predicting CFT responses across different nozzle positional factors.

Keywords: Crossflow turbine, Box-Behnken design, Machine learning, Nozzle position, Optimization.

1. Introduction

A cross-flow turbine (CFT) system is a type of impulse turbine that requires a low head and high volumetric flow rate application. The CFT system produces shaft power through the action of the water jets from its carefully designed nozzle on the upper and lower blade stages (Durgin & Fay, 1984; Achebe et al., 2020; Williams & Pavel, 2023). Studies have shown that 70% and 30% of water energies are extracted by the upper and lower blade stages of the CFT system (Adanta et al., 2018). Considering the high potential of hydropower on the earth's surface, which is covered with about 71% water, and the ocean has about 96.5% water (USGS, 2019; Sritram & Suntivarakorn, 2021), there is a need to maximize the efficiency of hydro-turbine systems, especially the CFT system due to its fewer design complexities and cost-effectiveness. The CFT system is classified as a small hydropower system (SHP) subdivided into Mini (<1MW), Micro (<100kW), and Pico (<10kW) energy levels (Energypedia, 2022). Moreover, the energy potentials of SHP systems are still minimally utilized as depicted in Figure 1, SHP had a global usage of just 7%.

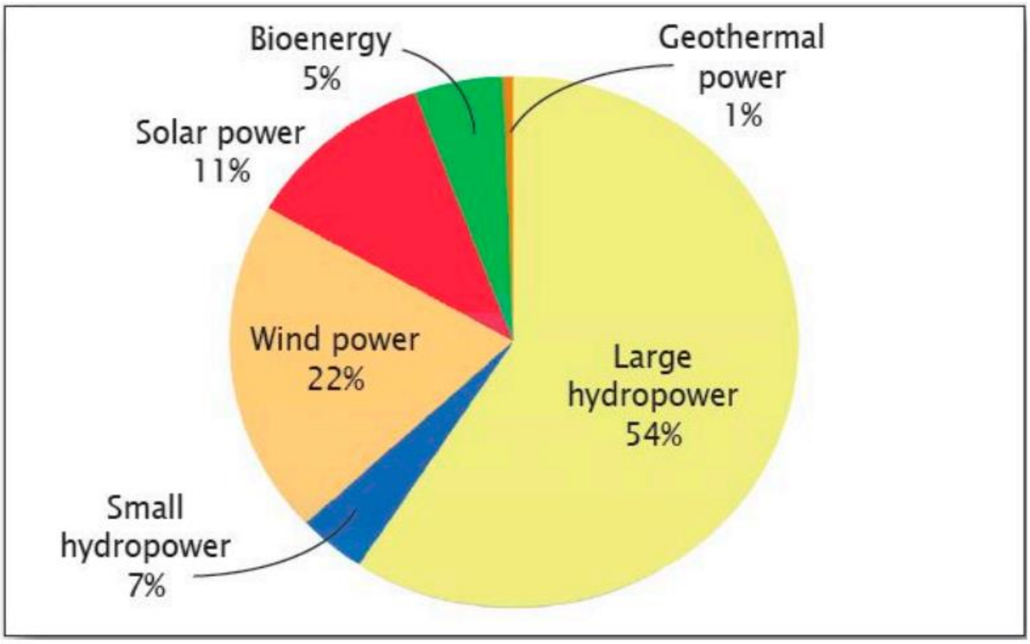

Fig. 1: Global percentage usage of various renewable energy sources (UNIDO, 2016)

In the design of an efficient CFT system, the design variables of the runner (comprising of the blades, rotor, and two wheels/discs) and the nozzle components are imperative as they significantly affect the CFT system's performance characteristics. The geometrical factors, flow characteristics, nozzle fluid properties, and the position of the nozzle with reference to the blades contribute significantly to the actualization of the cross-flow operation of the fluid through the upper blade stage to the lower stage. Several experimental and computational studies have been carried out on the optimal nozzle design for higher efficiency actualization. The effect of some positional variables of the nozzle, like the inlet angle or attack angle, nozzle tip distance from the centre shaft, and nozzle tip elevation, was investigated by Verma et al. (2017). The study achieved a maximum efficiency of 93% when positioned at an attack angle of 8°, nozzle tip distance of 4cm, and at any nozzle tip elevation, but the runner speed decreased as the nozzle tip distance was increased while having the nozzle fixed at an elevation of 13cm and 15cm from the central shaft axis. The optimal (custom) design tool which is a subset of response surface methodology (RSM) was employed to optimize the nozzle position in the study of Achebe et al. (2020), and an optimal position of the nozzle was achieved at 5°, 413mm, and 102mm respectively for attack angle, nozzle elevation from the runner shaft, and nozzle tip distance from the shaft. Also, the double-type positioned (combination of horizontal and vertical) nozzle had a higher efficiency of 81.68% compared to the horizontally and vertically placed nozzles due to the absolute flow dynamics operation of the fluid while participating in the second blade stage, which led to the maximization of water energy in the second stage (Rantererung et al., 2020). In addition, the nozzle's geometry (throat area, width, length, etc.) was found to affect the CFT's performance characteristics in the work of Oberkampf & Trucano (2002) using the computational fluid dynamics (CFD) technique. Nozzles with relatively narrow and converging inlet shapes were discovered to yield better CFT efficiency compared to other nozzle shapes investigated (Son et al., 2011). Furthermore, the effect of the nozzle's attack angle on the efficiency of the CFT system was investigated in the studies of (Desai & Aziz, 1994; Nakase et al., 1982; Fiuzat & Akerkar, 1989), and a maximum efficiency of 90% was obtained at an attack angle of 22°.

Despite numerous studies on CFT optimization, there remains a critical gap in the integration of experimental design, optimization algorithms, and machine learning validation framework. Most existing works either rely solely on experimental approaches or employ computational models without rigorous cross-validation or statistical performance benchmarking. Additionally, limited attention has been given to validating RSM-based optimization results using constrained nonlinear optimization techniques such as SLSQP, particularly for nonlinear hydrodynamic systems. This study addresses these gaps by introducing a hybrid methodology that combines BBD-based experimentation, RSM optimization, SLSQP validation, and MLP-based predictive modeling. The innovative aspect of this work lies in its multi-level validation approach, where optimization results are not only statistically validated but also compared across different modeling paradigms, thereby enhancing confidence in the derived optimal conditions and improving the reliability of predictive modeling for real-world hydropower applications.

Additionally, this study presents a significant advancement in the optimization of CFT nozzle positional performance. The novelty lies in the simultaneous application of these techniques to not only optimize nozzle positional parameters but also rigorously validate and compare predictive capabilities using multiple statistical performance indices. Furthermore, this study establishes a comprehensive relationship between nozzle positional variables and turbine performance metrics (efficiency, shaft power, and runner speed), thereby providing a more holistic and reliable optimization strategy for CFT systems.

2.0 Materials and Methods

2.1 Materials/Tools

The materials/tools used in this study include a digital tachometer, flow rate meter, Design Expert 11.0 software, MATLAB 2023 version, and a test site. The Design Expert 11.0 software was used in developing the Box-Behnken experimental design. The volumetric discharge of the test site and the speed of the runner shaft were ascertained using the flow rate meter and a digital tachometer, respectively. The test site used for the experiment had a head of 6.4m and a flow rate of 0.0042m^3/s. MATLAB 2023 version was used to model the response variables- turbine efficiency, shaft power, and runner shaft speed, and also compare the estimation performance of the models with that obtained from the design expert tool. The Anaconda package, which is one of the special packages of the Python programming language, was used to validate the optimization results of the RSM technique.

2.2. Method

The methodological framework of this study is shown in Figure 2. The detailed explanations of framework elements are presented in the sub-sections.

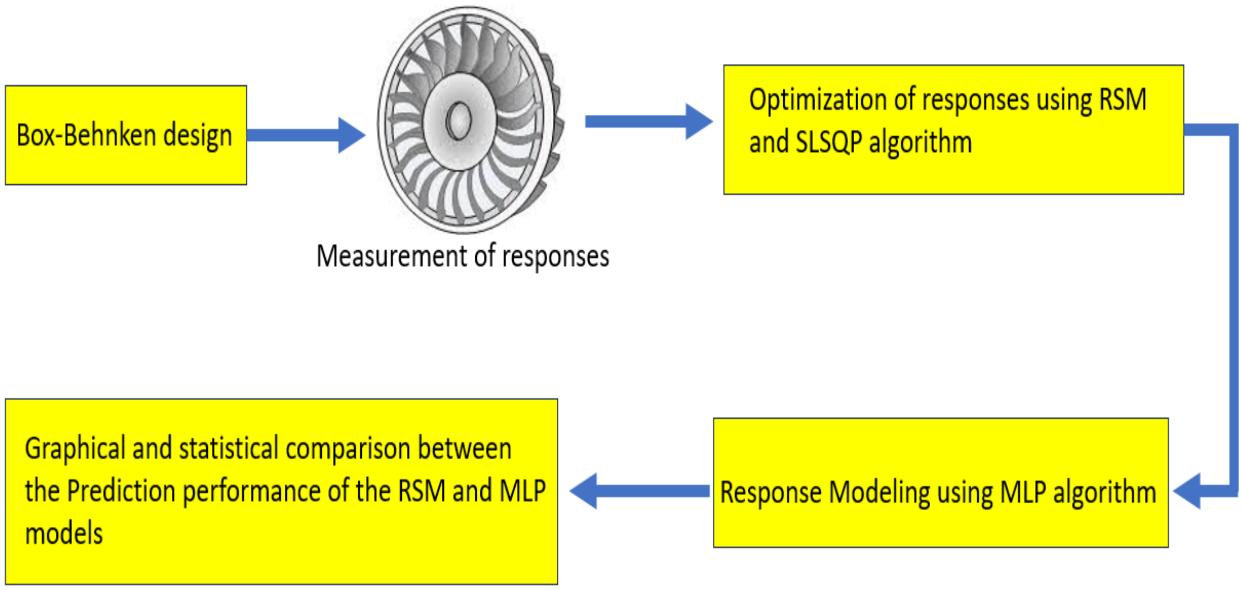

Fig. 2: Framework of research methodology

2.2.1. Box-Behnken Experimental Design

An experiment was designed using the Box-Behnken technique and was meticulously followed while measuring the various performance variables of the CFT hydro system. Box-Behnken design takes a minimum of three (3) categorical factors and duplicates the design for every combination of categorical factor levels, though only when categorical factor (s) are added. The independent variables considered in the Box-Behnken design include nozzle tip distance from the runner shaft, nozzle elevation from the runner shaft, and the attack/inlet angle, and were respectively varied at 102-202mm, 365-461mm, and 0-40°. The response variables that were used to evaluate the turbine's performance at varying nozzle positions were: turbine efficiency, shaft power, and runner shaft speed. The design outputted a total of 17 simulations/runs. Table 1 shows the experimental design.

Table 1: Box-Behnken Experimental Design

	Factor 1	**Factor 2**	**Factor 3**
Run	A: Nozzle distance Mm	B: Nozzle height Mm	C: Attack angle Degree
1	152	365	40
2	152	413	20
3	102	413	5
4	202	413	0

5	202	461	20
6	102	413	5
7	152	365	0
8	152	461	0
9	152	461	40
10	202	413	40
11	152	413	20
12	102	461	20
13	102	413	5
14	202	365	20
15	152	413	20
16	152	413	20
17	152	413	20

2.2.2. Measurement of the Response Variables of the CFT Hydro System

The CFT turbine reported in the Achebe et al (2020) study was used for the experimentation. This CFT system was designed with a flow head of 5.2m and a flowrate of 0.0015m^3/s but was used at a test site having a flow head of 6.4m and a flowrate of 0.0042m^3/s. The Box-Behnken design (Table 2) was carefully followed in the experimentation operation to accurately measure the performance variables of the hydro system. At a specific nozzle height from the shaft, nozzle distance from the shaft, and attack angle, the speed of the runner shaft was measured with a digital tachometer and recorded. The same procedure was adhered to in all 17 simulations/runs obtained from the Box-Behnken design of the experiment. Fig. 3 shows the measurement activity of the shaft's speed while the turbine is in operation.

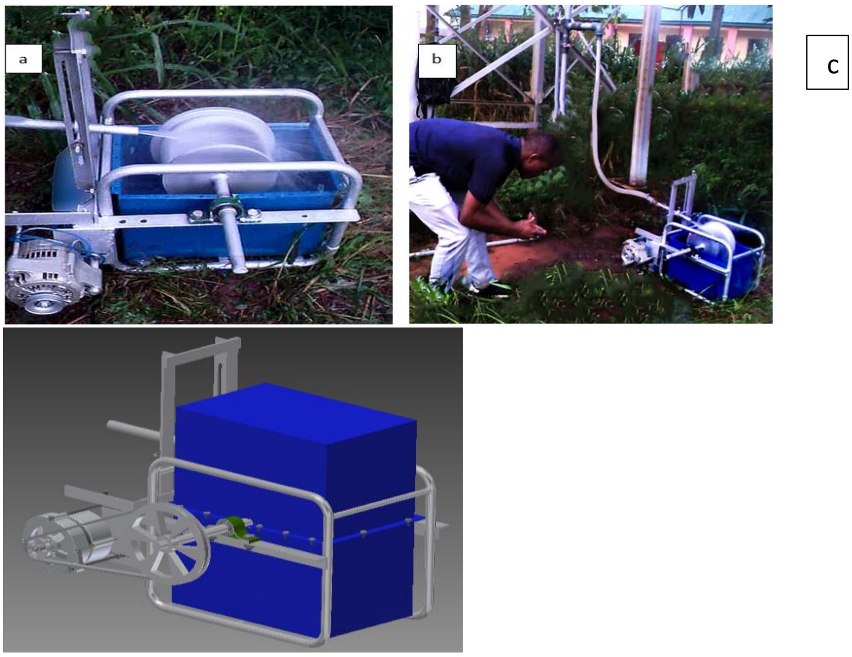

Fig. 3: (a) CFT system in operation (b) Measurement of runner shaft speed with a digital tachometer (c) Computer-aided design model of the CFT system

The shaft power and turbine efficiency were computed for each nozzle setting using equations 1-2 (Mockmore & Merryfield, 1949).

$$P_s = \left(\frac{\omega Q U_1}{g}\right)(V_1 \cos \alpha_1)\left(1 + \frac{\varphi \cos \beta_2}{\cos \beta_1}\right) = \omega . T \quad (1)$$

$$\eta_t = \frac{P_s}{\rho g Q H} \quad (2)$$

Where: φ = 0.98 (an empirical coefficient that is less than 1, used for compensating frictional effect), ω = angular velocity of the turbine shaft (rad/s), T = the torque of the turbine shaft (Nm), g = acceleration due to gravity (9.81m/s^2), α_1 = inlet/attack angle of water (degrees), P_s = shaft power (W), η_t = turbine efficiency (%), U_1 = peripheral speed of the runner, Q = flowrate of the test site (0.0042m^3/s), H = head (6.4m), β_1 = outer blade angle (28°), β_2 = inner blade angle (90°), $V_1 =$ absolute jet velocity = $C_V\sqrt{2gH}$ = 9.9m/s ($C_V = 0.98$). C_V is the velocity coefficient that accounts for water losses through the nozzle during flow operation.

2.2.3. Optimization of the Responses Using RSM and ML Techniques

The response variables were optimized to obtain the best nozzle position that would give the optimal performance characteristics of the responses. The best nozzle position was in reference to independent factors like attack angle, nozzle tip distance from the shaft, and nozzle tip elevation from the shaft. The quadratic models obtained from the RSM optimization technique for the three responses were employed in the ML system to further compare and validate the optimization outputs of RSM. The validation was done using the Sequential Least Squares Programming (SLSQP) algorithm. The RSM optimization goals are outlined in Table 2.

Table 2: RSM Optimization Goals for Each of the Factors

S/N	Factors	Goal
1	Attack angle	In range
2	Nozzle tip distance from the shaft	In range
3	Nozzle elevation from the shaft	In range
4	Efficiency	Maximize
5	Shaft power	Maximize
6	Runner speed	Maximize

The SLSQP algorithm was used because of the experimental data's nonlinearity and the need to constrain the optimization operation within the minimum and maximum range of values of both the independent and response factors.

2.2.4. Modeling Using Machine Learning Algorithms

A response prediction model based on the multilayer perceptron (MLP) neural network algorithm was developed using the MATLAB 2023 version. For the MLP system, the Levenberg-Marquardt training algorithm was used because it best suits the non-linearity pattern of the dataset. From the 17 data points, 70%, 15%, and 15% were respectively used for training, validation, and testing of the models. The MLP model had three (3) inputs, ten (10) hidden neurons, one (3) output layer, and three (3) outputs as depicted in Figure 4. The fed inputs were the inlet angle (C), nozzle tip distance from the shaft (A), and nozzle elevation from the shaft (B), while the target outputs include turbine efficiency (η_t), shaft power (P_s), and runner shaft rpm (S_r). The MLP model employed a two-layer feed-forward network with sigmoid hidden neurons and linear output neurons capable of fitting multi-dimensional mapping problems. The MLP function is expressed in equation 3.

$$y(x) = F(\sum_{i=1}^{L} w_i(t).\ x_i(t) + b) \quad (3)$$

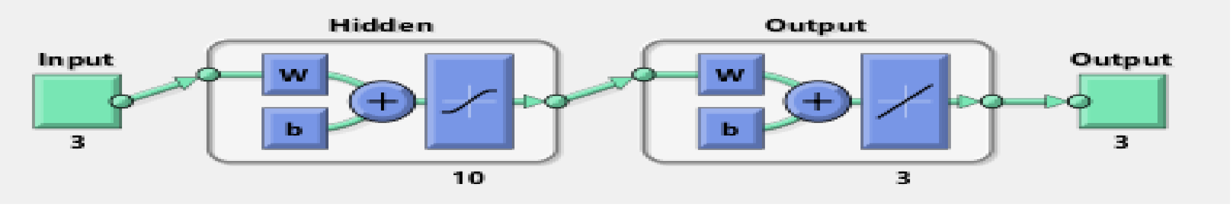

Figure 4: The architecture of the MLP model

Where: y(x) = the output, $x_i(t)$ = input variables in discrete time space t, F = the hidden transfer function, L = hidden neurons, $w_i(t)$ = the weight, b = the bias.

2.2.5. Statistical Validation of the Estimation Models

To obtain the best-performed model, the estimation performances of the RSM-based models and the MLP model were accessed statistically using the indices of mean bias error (MBE), Nash-Sutcliffe equation (NSE), root mean square error (RMSE), coefficient of correlation (r), T-statistics test, and coefficient of determination (R^2). The models used in determining these statistical indices are given in equations 4-9 (Onyeka et al., 2021; Mbah et al., 2022; Akpenyi & Okafor, 2024; Okafor et al., 2025).

$$MBE = \frac{1}{n}\sum_{i=1}^{n}(Z_{ical} - Z_{imea}) \quad (4)$$

$$RMSE = \left[\frac{1}{n}\sum_{i=1}^{n}(z_{ical} - z_{imeas})^2\right]^{\frac{1}{2}} \quad (5)$$

$$NSE = 1 - \frac{\sum_{i=1}^{n}(Z_{imea} - Z_{ical})^2}{\sum_{i=1}^{n}(Zi_{mea} - \bar{Z}_{mea})^2} \quad (6)$$

$$r = \frac{\sum XY}{\sqrt{(\sum X^2)(\sum Y^2)}} \quad (7)$$

$$t = \left[\frac{(n-1)(MBE)^2}{(RMSE)^2 - (MBE)^2}\right]^{\frac{1}{2}} \quad (8)$$

$$R^2 = \frac{\text{sum of squares (ss)}}{\text{sum of squares of residuals}} = \frac{\sum_{i=1}^{n}(Z_{imea} - \bar{Z}_{imea})^2}{\sum_{i=1}^{n}(Z_{imea} - \bar{Z}_{ical})^2} \quad (9)$$

Where: n = number of data points = 69, Z_{ical} = calculated value of each response variable, Z_{imea} = measured value of each response variable, $\bar{Z}_{mea}$ = the mean measured value of each response, X = the difference between the measured values of each response and the mean of the measured response, Y = the difference between the estimated and the mean of the estimated response variable.

3. Results and Discussions

3.1. RSM Optimization of the Responses

Quadratic models were developed by applying the RSM technique to all the response variables employed in evaluating the CFT's performance characteristics. The models showed that each of the independent factors (A, B, and C) had a significant effect on the turbine's performance variables, with *B* and C having a combined two-factor interactive effect on the responses (η_t, P_s, and S_r). The analysis of variance (ANOVA) tables for the responses are presented in Tables 3-5.

Table 3: ANOVA for the Turbine Efficiency

Source	Sum of Squares	df.	Mean Square	F-value	p-value	
Model	11611.62	9	1290.18	14.38	0.0010	significant
A-Nozzle distance	1030.17	1	1030.17	11.48	0.0116	
B-Nozzle height	1310.48	1	1310.48	14.61	0.0065	
C-Attack angle	655.04	1	655.04	7.30	0.0306	
AB	164.95	1	164.95	1.84	0.2172	
AC	31.68	1	31.68	0.3531	0.5710	
BC	2140.91	1	2140.91	23.86	0.0018	
A^2	902.23	1	902.23	10.06	0.0157	
B^2	739.97	1	739.97	8.25	0.0239	
C^2	812.90	1	812.90	9.06	0.0197	
Residual	628.06	7	89.72			
Lack of Fit	628.06	1	628.06			
Pure Error	0.0000	6	0.0000			
Cor Total	12239.69	16				

Table 4: ANOVA for the Shaft Power of the Turbine

Source	Sum of Squares	df	Mean Square	F-value	p-value	
Model	80741.55	9	8971.28	14.37	0.0010	significant
A-Nozzle distance	7163.86	1	7163.86	11.48	0.0116	
B-Nozzle height	9113.40	1	9113.40	14.60	0.0065	
C-Attack angle	4555.66	1	4555.66	7.30	0.0306	
AB	1147.53	1	1147.53	1.84	0.2172	
AC	220.28	1	220.28	0.3529	0.5711	
BC	14887.66	1	14887.66	23.85	0.0018	
A^2	6275.54	1	6275.54	10.06	0.0157	
B^2	5141.94	1	5141.94	8.24	0.0240	
C^2	5652.70	1	5652.70	9.06	0.0197	
Residual	4368.79	7	624.11			

Lack of Fit	4368.79	1	4368.79			
Pure Error	0.0000	6	0.0000			
Cor Total	85110.34	16				

Table 5: ANOVA for the Speed of the Runner Shaft

Source	Sum of Squares	df	Mean Square	F-value	p-value	
Model	2.168E+05	9	24093.44	14.84	0.0009	significant
A-Nozzle distance	16158.90	1	16158.90	9.96	0.0160	
B-Nozzle height	20953.15	1	20953.15	12.91	0.0088	
C-Attack angle	8117.55	1	8117.55	5.00	0.0604	
AB	2051.73	1	2051.73	1.26	0.2979	
AC	1111.42	1	1111.42	0.6848	0.4352	
BC	56406.25	1	56406.25	34.75	0.0006	
A^2	14002.37	1	14002.37	8.63	0.0218	
B^2	16191.16	1	16191.16	9.98	0.0160	
C^2	20831.80	1	20831.80	12.84	0.0089	
Residual	11361.19	7	1623.03			
Lack of Fit	11361.19	1	11361.19			
Pure Error	0.0000	6	0.0000			
Cor Total	2.282E+05	16				

From Tables 3-5, using a significant value of 5% (confidence interval of 95%), it could be observed that the quadratic models for the turbine's responses- efficiency, shaft power, and runner shaft speed were all significant as their respective p-values were all less than 5%. Moreso, the independent factors- nozzle height and attack angle had a combined two-factor interaction effect on the considered turbine responses as attested by the p-values that were less than 5%. This observed interaction between nozzle height and attack angle is consistent with the findings of Nakase et al. (1982) and Fiuzat and Akerkar (1989), who demonstrated that nozzle configuration strongly affects flow distribution and energy transfer within the turbine. The quadratic models (in actual factors) for each of the turbine responses are shown in equations 10-12.

$$\text{Efficiency} = -298.91534 - 4.06637A + 4.09818B - 8.22477C + 0.003851AB - 0.004430AC + 0.024099BC + 0.007348A^2 - 0.006654B^2 - 0.041283C^2 \quad (10)$$

$$\text{Shaft power} = -786.97647 - 10.72475A + 10.80140B - 21.68924C + 0.010158AB - 0.011681AC + 0.063549BC + 0.019380A^2 - 0.017540B^2 - 0.108863C^2 \quad (11)$$

$$\text{Speed of the runner shaft} = -1855.99158 - 15.19431A + 19.84875B - 40.83977C + 0.013583AB - 0.026239AC + 0.123698BC + 0.028948A^2 - 0.031124B^2 - 0.208986C^2 \quad (12)$$

The RSM-based models shown in equations 10-12 for turbine's efficiency, shaft power, and runner speed, respectively, had R^2 values of 0.9487, 0.8827, and 0.9502. Also, the correlation coefficients (r) of the actual and model-predicted values of the responses are respectively 0.974, 0.974, and 0.975 for the efficiency, shaft power, and runner shaft speed models. The high values of R^2 and r of the models showed good model performance and suitability for response prediction at any factor level considered. Furthermore, the optimal nozzle position for improved CFT performance characteristics was obtained at an attack angle of 5°, nozzle distance of 102mm, and nozzle height of 408.4mm. The highest desirability value of 0.972 was achieved at this best nozzle positional reference. Also, at this optimal nozzle position (design point), the performance variables (efficiency, shaft power, and runner speed) of the CFT system were respectively gotten to be approximately 92%, 242.4W, and 406 rpm. Figures 4-5 display these in contour plots.

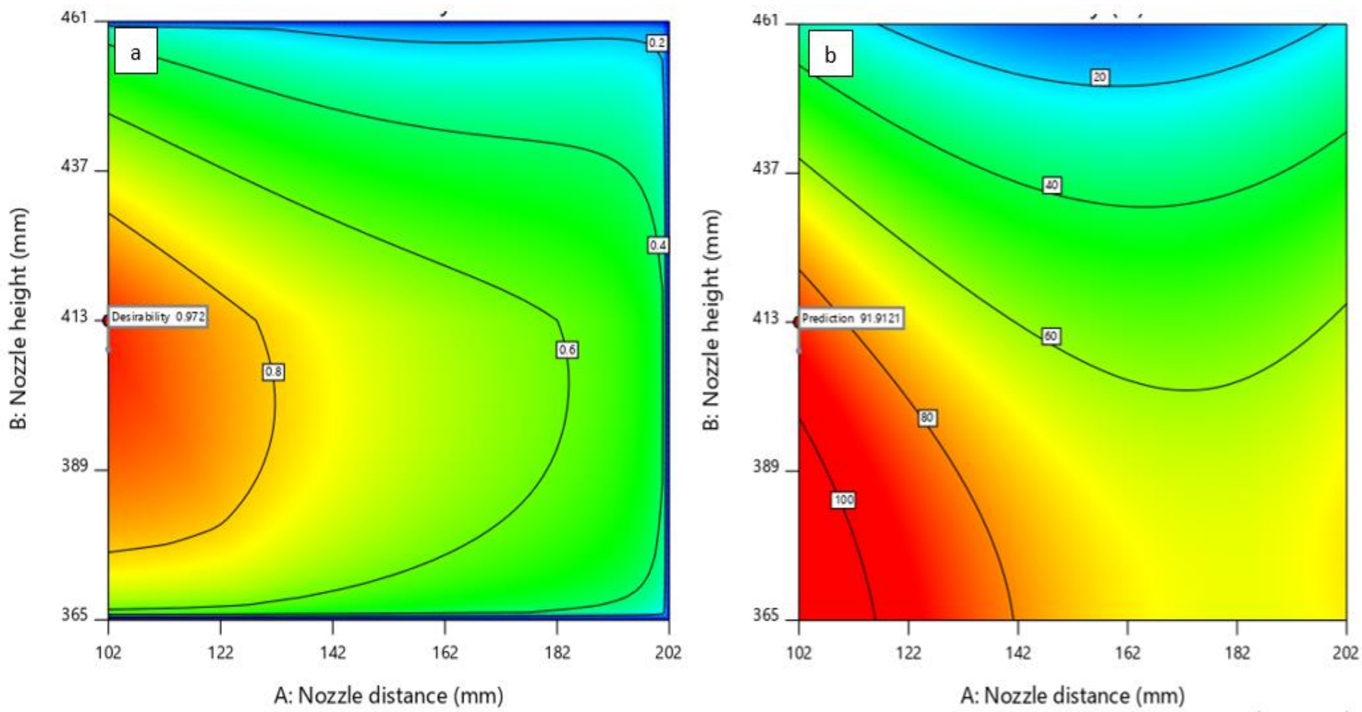

Fig. 5: Contour plot of nozzle height against nozzle distance for (a) desirability (b) efficiency

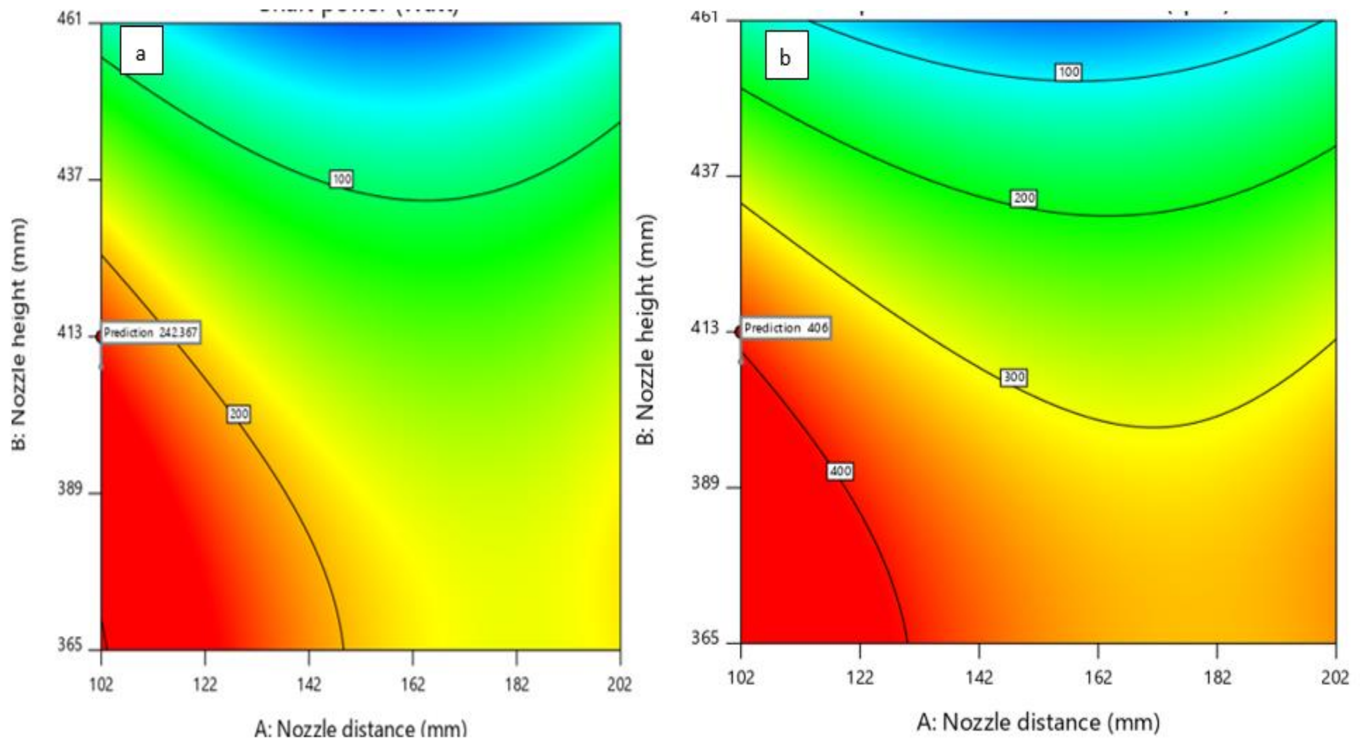

Fig. 6: Contour plot of attack angle against nozzle height for (a) shaft power (b) runner speed

The achieved optimal nozzle position for improved CFT performance characteristics agrees with the study of Achebe et al. (2020). In addition, the result of the responses agrees with the findings of Desai and Aziz (1994) and Verma et al. (2017), who reported that optimal nozzle positioning significantly enhances turbine efficiency, with peak efficiencies approaching 90–93%. Table 6 shows the summary statistics of the response predictions at the determined optimal nozzle position.

Table 6: Summary Statistics of the RSM-Based Model Predictions (Two-Sided, Confidence = 95%)

Solution 1 of 100 Response	Predicted Mean	Predicted Median	Std Dev	SE Mean	95% CI low for Mean	95% CI high for Mean	95% TI low for 99% Pop	95% TI high for 99% Pop
Efficiency	91.9121	91.9121	9.47225	5.52181	78.8551	104.969	37.6022	146.222
Shaft power	242.367	242.367	24.9823	14.5633	207.931	276.804	99.1295	385.605
Speed of the runner shaft	406	406	40.2868	23.4851	350.467	461.533	175.012	636.988

3.2. The SLSQP Model-Based Optimization of the Responses

The outputs of the SLSQP optimization program well aligned with the RSM results, though small deviations were observed. Table 4 delineates the SLSQP results, both the predicted optimal nozzle position and the predictions of the responses at this optimal position. Therefore, Table 7 validates the RSM optimization solution. Furthermore, the effectiveness of the SLSQP algorithm in validating the optimization results aligns with the work of Nocedal and Wright (2006), who highlighted its suitability for constrained nonlinear optimization problems.

Table 7: Optimal Nozzle Positions and Responses

Nozzle parameters	Values	Optimal Responses		
		η_t	P_s	S_r
Nozzle distance	102mm	91.962%	242.599W	405.99 rpm
Nozzle height	407.99mm			
Attack angle	4.699°			

3.3. Performance Results of the MLP Model

The training of the MLP model in search of an optimal solution within an epoch setting of 1-6 is depicted in Figure 7. At an epoch of 1, the best validation performance of 841.6979 was derived. At this optimal epoch point, the model's MSE output during training, validation, and testing is the same. The details of the search operation for optimal model performance point from 1-6 epochs are clearly displayed in Figure 8.

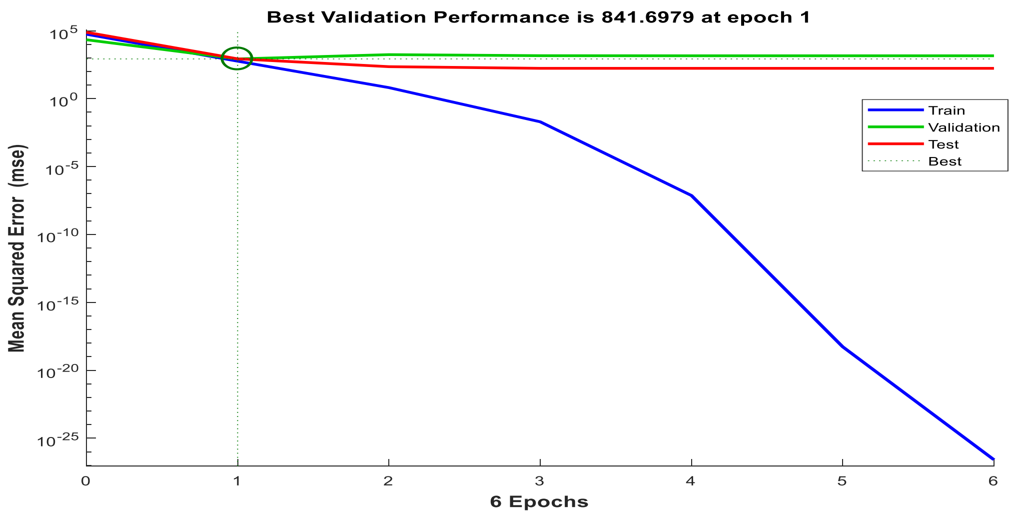

Figure 7: Optimal solution point of the MLP model

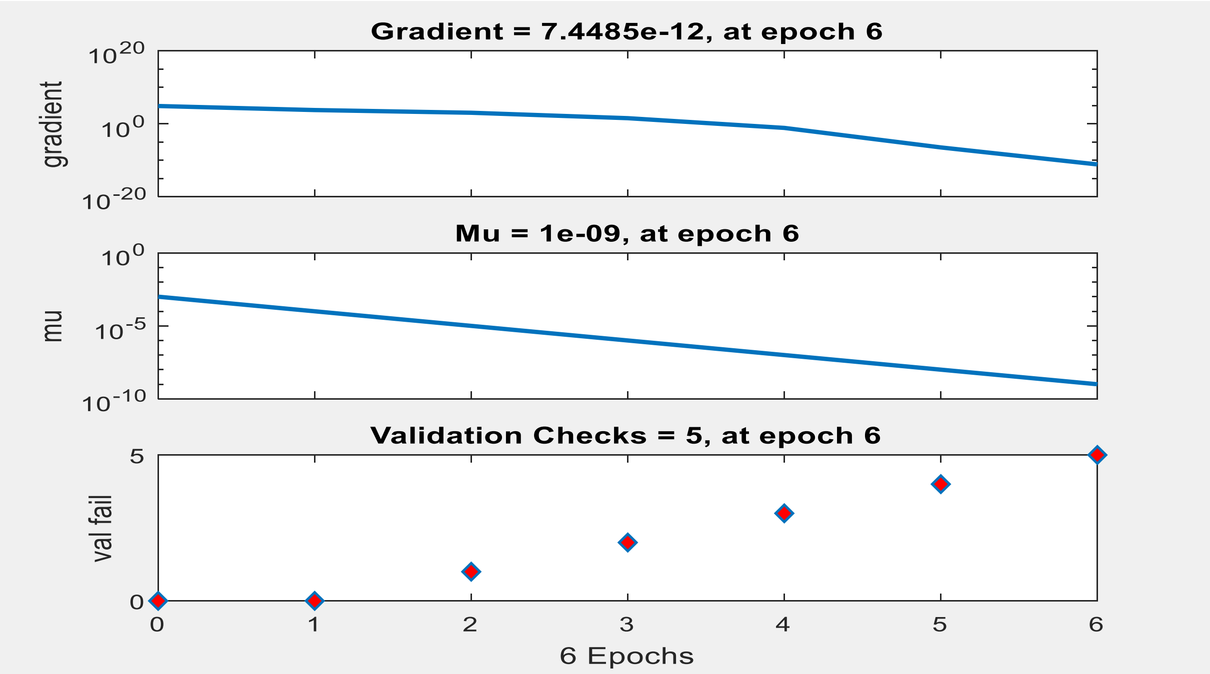

Figure 8: Plot of the iteration process

Figure 8 reveals the different iteration processes within the epoch points of 1-6, with the gradient value of 7.4485e-12 quantifying the decline from the epoch point of 1 to 6 during the MLP model's validation performance check. Figure 9 presents a clearer picture of the model's behaviour during the training, validation, testing, and overall modeling phases.

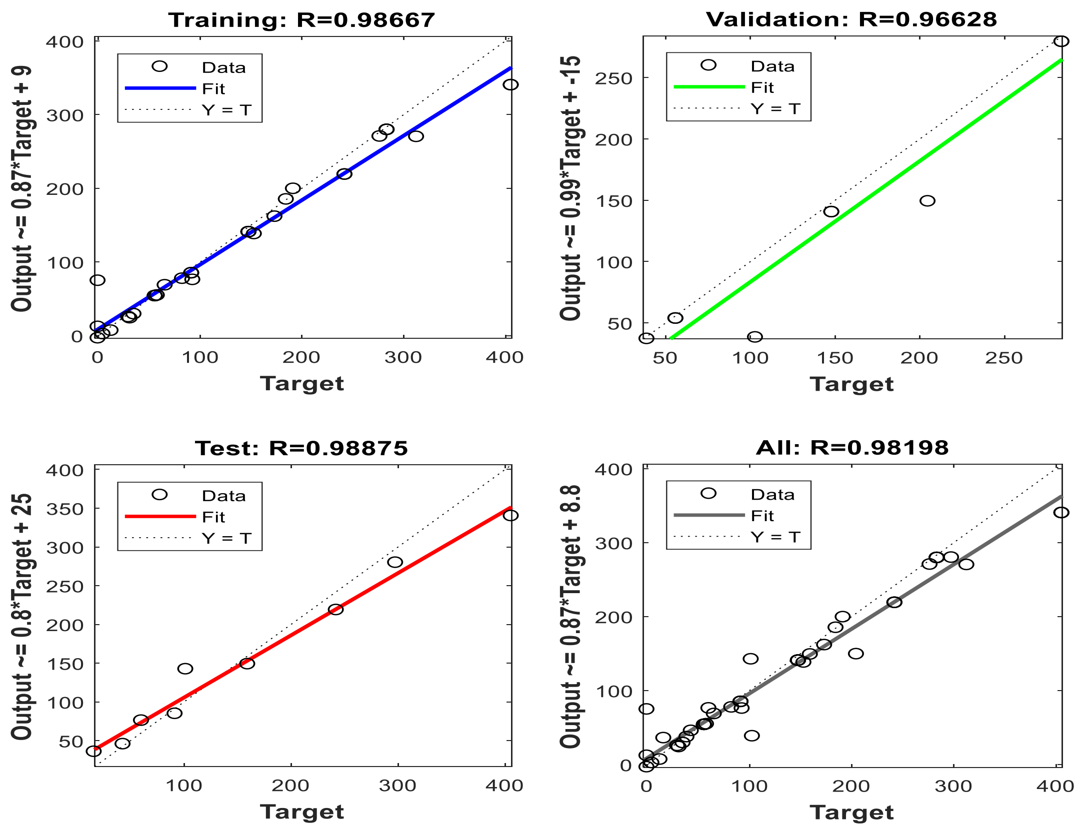

Figure 9: Graphical display of the validation, training, test, and overall behaviour of the MLP model

From Figure 9, it can be observed that the MLP model performed well in fitting the data to the regression line. The predicted response values are closer to the actual/measured values, thus attesting to good prediction performance of the MLP model. In addition, this observation is also buttressed by the fact that the correlation coefficient R of the training, validation, testing, and overall process are all near unity, thus proving that the actual and predicted responses of the CFT system are positively correlated.

3.4. Performance Comparison Between the RSM and MLP Models

Figures 10-12 depict the graphical visualizations of the RSM and MLP response predictions in comparison with the measured response values.

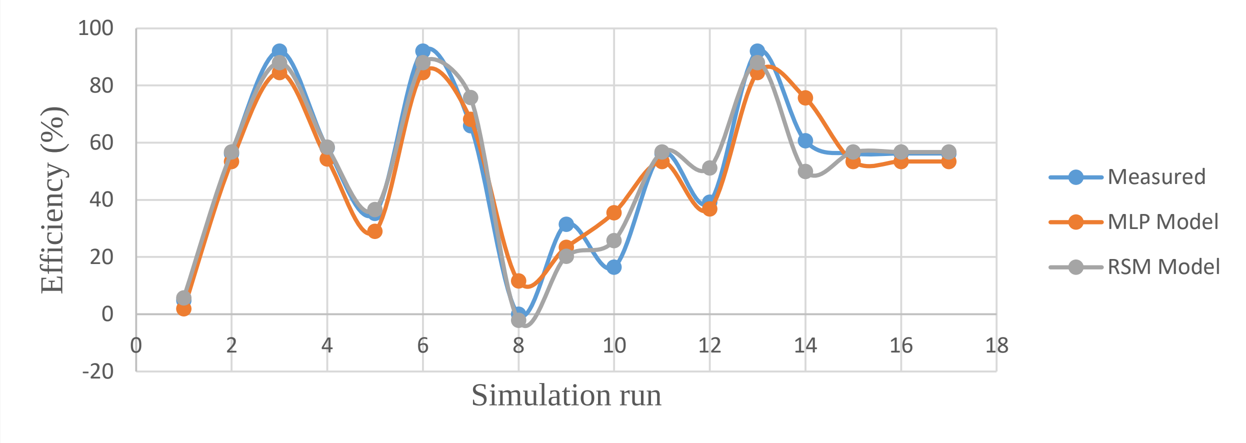

Figure 10: Comparison between the MLP and RSM model performances for CFT's efficiency

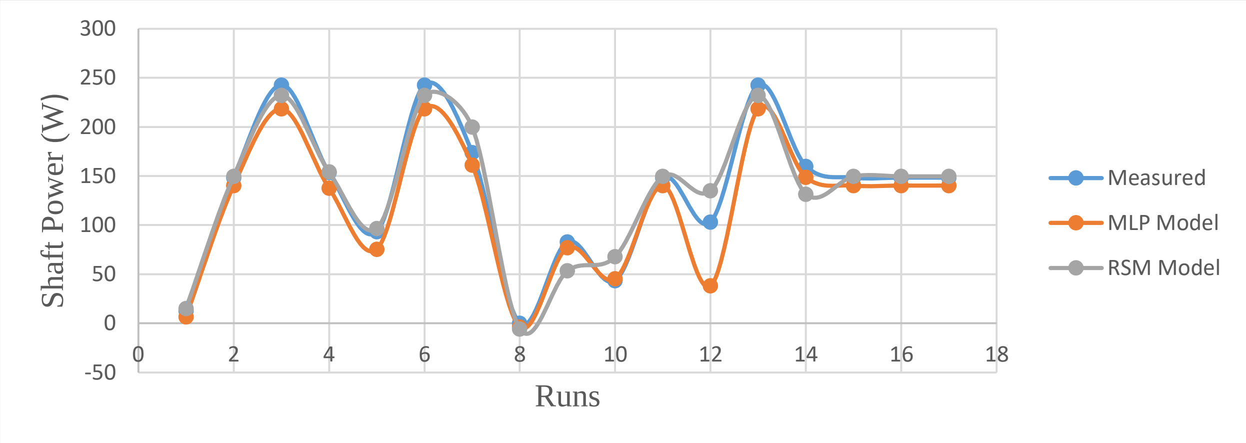

Figure 11: Comparison between the MLP and RSM model performances for CFT's shaft power

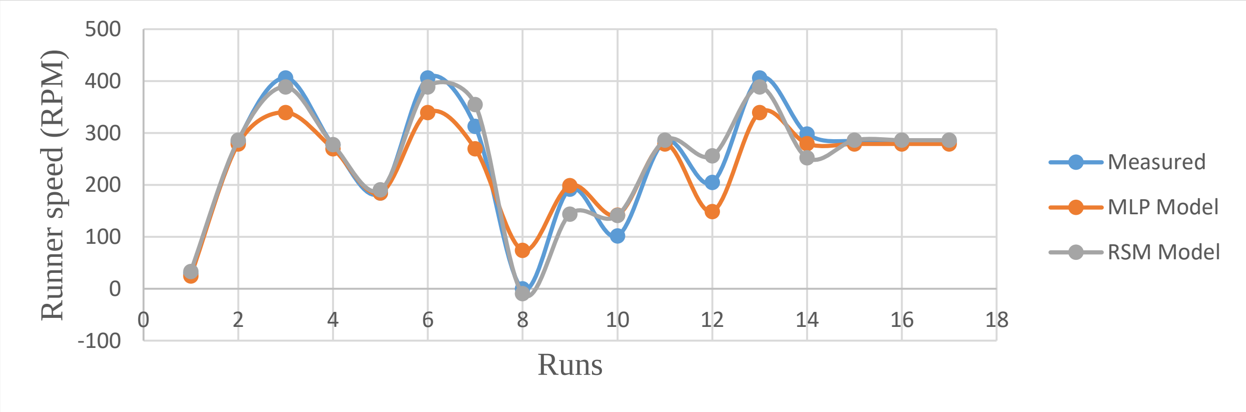

Figure 12: Comparison between the MLP and RSM model performances for CFT's runner speed

From figures 10-12, it can be deduced that both the MLP and RSM models have good estimation performances as the predicted efficiency, shaft power, and runner speed of the models aligned with the measured values, though some deviations were observed. However, the RSM-based models appeared to outperform the MLP model as their predicted response values better fit the measured response values. Table 8 shows the results of the statistical analysis employed in evaluating the model estimation performances. These statistical indices helped with the conclusive deductions of the best-performing model for CFT's response prediction at any factor level considered.

Table 8: Statistical Evaluation of the RSM and MLP Models

Statistical tool	RSM Model			MLP Model		
	η_t	P_s	S_r	η_t	P_s	S_r
MBE	0.00118	-0.00000	0.00000	-0.70473	-14.65465	-13.66488
RMSE	6.07608	16.03002	25.85204	7.94157	20.65957	39.02850
NSE	0.94872	0.94867	0.95021	0.91240	0.91475	0.88653
r	0.97402	0.97400	0.97479	0.95715	0.97879	0.96860
T-statistics	0.00077	-0.00000	0.00000	-0.35636	-4.02538	-1.49514
R^2	0.9487	0.8827	0.9502	0.91240	0.91475	0.88653

From Table 8, both models have good prediction performances, but the RSM model outperformed the MLP model because it gave lower MBE, RMSE, and T-statistic values, and higher values of NSE, r, and R^2. The high coefficient of determination (R^2) obtained in this study confirms the robustness of RSM in modeling nonlinear systems, as also reported by Myers et al. (2016). In addition, although machine learning models provide strong predictive capabilities, their performance may depend on dataset size and structure, as noted in previous studies (Yadav & Chandel, 2014; Pang et al., 2020; Bhansali et al., 2023), hence the reason why the RSM outperformed the ML-model.

4. Conclusions

The optimization of CFT systems remains a critical research focus in enhancing the efficiency and viability of small-scale hydropower generation. Given the strong dependence of turbine performance on nozzle positioning parameters, a systematic and integrated optimization approach is essential for achieving improved energy conversion efficiency. In this study, a hybrid methodology combining BBD experimental design, RSM-optimization, Sequential Least Squares Programming (SLSQP), and multilayer perceptron (MLP) modeling was employed to investigate and optimize the influence of nozzle height, nozzle distance, and attack angle on key turbine performance indicators, efficiency, shaft power, and runner rpm. The major findings of the study are summarized as follows:

1. The nozzle positional parameters (nozzle height, nozzle distance, and attack angle) significantly affect the turbine's efficiency, shaft power, and runner speed.

2. A strong interaction effect between nozzle height and attack angle was observed, highlighting their combined influence on the hydrodynamic behavior within the turbine.

3. The optimal nozzle configuration obtained via RSM was: nozzle height = 408.4 mm, nozzle distance = 102 mm, and attack angle = 5°, resulting in efficiency = 91.91%, shaft power = 242.367 W, and runner speed = 406 rpm.

4. The SLSQP optimization algorithm closely validated the RSM results by yielding: nozzle distance =102mm, nozzle height = 407.99mm, and attack angle = 4.699°, confirming the robustness and reliability of the optimization framework.

5. Both RSM and MLP models demonstrated good predictive capability; however, the RSM model outperformed the MLP model based on lower error metrics (MBE, RMSE, and T-statistic) and higher correlation-based indices (NSE, r, and R^2).

The significance of these findings lies in their contribution to advancing the design and optimization of CFTs through a multi-method integrated framework. The strong agreement between RSM and SLSQP results demonstrates the reliability of combining statistical and numerical optimization techniques for solving nonlinear engineering problems. Furthermore, the superior performance of the RSM model over the MLP model suggests that, for relatively small and structured experimental datasets, statistically derived models can provide more stable and interpretable predictions than data-driven approaches. The identification of optimal nozzle parameters and the quantified interaction effects provide deeper insight into the internal flow dynamics and energy transfer mechanisms within the turbine. These outcomes not only enhance the theoretical understanding of CFT operation but also offer practical guidelines for the design and deployment of efficient small hydropower systems, particularly in resource-constrained and decentralized energy environments.

5. Recommendations

Future research should focus on expanding the dataset and incorporating advanced machine learning techniques such as deep neural networks and ensemble learning methods to improve predictive performance under more complex operating conditions. Additionally, integrating computational fluid dynamics (CFD) simulations with experimental and data-driven models would enable a more detailed analysis of flow behavior and turbulence effects within the turbine. Further investigations may also consider additional design and operational parameters, including blade geometry, flow variability, and multi-nozzle configurations, to develop a more comprehensive and globally optimized CFT design framework.

Author contributions: Obiora Clement Okafor contributed to the conceptualization, methodology, software development, validation, formal analysis, investigation, resource provision, original draft preparation, and review and editing of the manuscript. Chinonso Hubert Achebe was involved in the investigation, visualization, methodology, review and editing, and supervision. Jeremiah Chukwuneke participated in the methodology, data curation, review and editing, and supervision.

Data availability statement: The data that supports this study is available upon request.

Conflict of interest: The authors declare no conflict of interest.

Ethics: There are no ethical issues with the publication of this manuscript.

References

Achebe, C. H., Okafor, O. C., & Obika, E. N. (2020). Design and implementation of a crossflow turbine for Pico hydropower electricity generation. Heliyon, 6(7), 1–13. https://doi.org/10.1016/j.heliyon.2020.e04523

Adanta, D., Hindami, R., Budiarso, Warjito, & Siswantara, A. I. (2018). Blade depth investigation on cross-flow turbine by numerical method. *2018 4th International Conference on Science and Technology (ICST)*. IEEE. Retrieved from http://dx.doi.org/10.1109/icstc.2018.8528291

Akpenyi-Aboh, O.N., and Okafor, O.C. (2024). Modeling the Solar Intensity of Asaba Town in Nigeria Using Response Surface Methodology and Machine Learning Techniques. American Journal of Mechanical and Industrial Engineering, 9(4), 63-74. https://doi.org/10.11648/j.ajmie.20240904.11

Bhansali, A., Narasimhulu, N., Pérez de Prado, R., Divakarachari, P. B., & Narayan, D. L. (2023). A review on sustainable energy sources using machine learning and deep learning models. Energies, 16(17), 6236. https://doi.org/10.3390/en16176236

Desai, V.R, & Aziz, N.M. (1994). Parametric Evaluation of Cross-flow Turbine Performance. American Society of Civil Engineers (ASCE) Journal of Energy Engineering, 120(1), 17–34.

Durgin, W., and Fay, W. (1984). Some fluid flow characteristics of a cross-flow type hydraulic turbine (77–83). In Small Hydro Power Fluid Machinery; ASME: New York, NY, USA.

Energypedia. (2022, March 22). Hydro power basics. Retrieved January 3, 2024, from energypedia website: https://energypedia.info/wiki/Hydro_Power_Basics#cite_note-1

Fiuzat, A., and Akerkar, B. (1989). The Use of Interior Guide Tube in Cross Flow Turbines. In Waterpower'89, Proceedings of the International Conference on Hydropower, Niagara Falls, NY, USA, 23–25 August 1989; 1111–1119. American Society of Civil Engineers: New York, NY, USA.

Mbah, O.M., Madueke, C.I., Umunakwe, R. and Okafor, O.C. Machine Learning Approach for Solar Irradiance Estimation on Tilted Surfaces in Comparison with Sky Models Prediction, Journal of Engineering Sciences, vol. 9, no. 2, pp. G1–G6, 2022, https://doi.org/10.21272/jes.2022.9(2).g1

Mockmore, C. A., & Merryfield, F. (1949). The Banki Water Turbine (25, 5–27). Oregon State University, Corvallis, USA: Engineering Bullentin Series.

Myers, R. H., Montgomery, D. C., & Anderson-Cook, C. M. (2016). Response surface methodology: Process and product optimization using designed experiments. John Wiley & Sons.

Nakase, Y., Fukutomi, J., Watanabe, T., Suetsugu, T., Kubota, T., and Kushimoto, S. (1982). A study of Cross-Flow turbine (Effects of nozzle shape on its performance). 1419, 14–19. In Proceedings of the Winter Annual Meeting ASME, Phoenix, AZ, USA.

Nocedal, J., & Wright, S. (2006). Numerical optimization. Springer Science & Business Media.

Oberkampf, W. L., & Trucano, T. G. (2002). Verification and validation in computational fluid dynamics. Progress in Aerospace Sciences, 38(3), 209–272. https://doi.org/10.1016/s0376-0421(02)00005-2

Okafor, O.C., Achebe, C.H. and Mbaka, U.O. (2025). Application of statistical thermodynamics to modeling fluid discharge through a circular orifice system. American Journal of Mechanical and Industrial Engineering, 10(1): 1-11. https://doi.org/10.11648/j.ajmie.20251001.11

Onyeka, V.O., Nwobi-Okoye, C.C., Okafor, O.C., Madu, K.E., and Mbah, O.M. (2021). Estimation of global solar radiation using empirical models. Journal of Engineering Sciences, Vol. 8(2), pp. 11-19, https://doi.org/10.21272/jes.2021.8(2).g2

Pang, Z., Niu, F., & O'Neill, Z. (2020). Solar radiation prediction using recurrent neural network and artificial neural network: A case study with comparisons. Renewable Energy, 156, 279–289. https://doi.org/10.1016/j.renene.2020.04.042

Rantererung, C. L., Soeparman, S., Soenoko, R., & Wahyudi, S. (2020). A double nozzle cross flow turbine fluid flow dynamic. Journal of Southwest Jiaotong University, 55(4), 1–11. https://doi.org/10.35741/issn.0258-2724.55.4.49

Son, S.W, Inagaki, M., Han, C.M, and Choi, Y.D. (2011). Effect of nozzle shape and draft tube on the performance of cross-flow turbine for small hydropower. In: Proceedings of the 11th Asian International Conference on Fluid Machinery and the 3rd Fluid Power Technology Exhibition. IIT Madras, Chennai, India; 2011.

Sritram, P., & Suntivarakorn, R. (2021). The efficiency comparison of hydro turbines for micro power plant from free vortex. Energies, 14(23), 7961. https://doi.org/10.3390/en14237961

United Nations Industrial Development Organization (UNIDO, 2016). World small hydropower development report. Retrieved from: https://sustainabledevelopment.un.org [Accessed on: 12th January, 2024].

USGS. (2019, November 13). How Much Water is There on Earth? Retrieved January 3, 2024, from U.S. Geological Survey website: https://www.usgs.gov/special-topics/water-science-school/science/how-much-water-there-earth

Verma, V., Gaba, V. K., & Bhowmick, S. (2017). An experimental investigation of the performance of cross-flow hydro turbines. Energy Procedia, 141, 630–634. https://doi.org/10.1016/j.egypro.2017.11.084

Williams S.E., & Pavel Y.T. (2023). Deployment of design sequence for crossflow turbine functionality enhancement. Journal of Advanced Research in Fluid Mechanics and Thermal Sciences, *103*(2), 118–140. https://doi.org/10.37934/arfmts.103.2.118140

Yadav, A. K., & Chandel, S. S. (2014). Solar radiation prediction using Artificial Neural Network techniques: A review. Renewable and Sustainable Energy Reviews, 33, 772–781. https://doi.org/10.1016/j.rser.2013.08.055