STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
*Stat., Optim. Inf. Comput.*, Vol. **16**, August 2026, pp 1182–1201.
Published online in International Academic Press (www.IAPress.org)

# Hybrid RSM–GA Hyperparameter Tuning of Artificial Neural Networks for Academic Performance Prediction

Sudirman Rizki Ariyanto [1,*], Bambang Suprianto [2], Warju [1], Ata Syifa' Nugraha [3]

[1]*Department of Automotive Engineering Technology, Universitas Negeri Surabaya, Indonesia*
[2]*Department of Electrical Engineering, Universitas Negeri Surabaya, Indonesia*
[3]*Department of Aircraft Maintenance Technology, Universitas Sunan Gresik, Indonesia*

**Abstract** The development of artificial intelligence has encouraged the use of Artificial Neural Networks (ANNs) for academic performance prediction in competency-based education. However, in automotive vocational education, ANN hyperparameters are commonly determined through trial-and-error procedures, which may produce unstable and suboptimal models. This study proposes an integrated Central Composite Design (CCD), Response Surface Methodology (RSM), and Genetic Algorithm (GA) framework to optimize ANN hyperparameters for predicting the academic performance of Automotive Vocational Education (AVE) students at Vocational High School (VHS) NU 1 Karanggeneng, Lamongan. The quadratic RSM model explained 82.09% of the response variation and identified the learning algorithm as the most influential optimization factor. At the original optimization-response scale, the RSM–GA procedure produced an optimal ANN configuration with three hidden layers, 20 neurons, a tansig–purelin transfer-function combination, the trainlm learning algorithm, a learning rate of 0.005, and 200 epochs, achieving an MSE of 0.07 and an S/N Ratio of 22.94. In the normalized response-level benchmark, the Opt RSM–GA ANN obtained the most favorable normalized repeated MSE response, with a Mean MSE of 0.012, SD MSE of 0.0019, and S/N Ratio of 38.65. These findings indicate that the CCD–RSM–GA workflow provides a structured and reproducible approach for ANN hyperparameter optimization. Broader validation using larger, multi-school, and multi-cohort datasets is still required before practical implementation.

**Keywords** Artificial Neural Network, RSM–GA Integration, Hyperparameter Tuning, Automotive Vocational Education, Predictive Modeling

**AMS 2010 subject classifications** 68T07, 62J02, 62K20

**DOI:** 10.19139/soic-2310-5070-3638

## 1. Introduction

*The development of Artificial Intelligence* (AI) has brought significant transformations in various sectors, including education, through predictive analytics techniques capable of processing complex and multidimensional academic data [1, 2]. In vocational secondary education, academic performance prediction has become increasingly important because vocational institutions emphasize not only cognitive achievement but also technical competencies and industrial work readiness. Machine learning-based predictive models, particularly ANN, have been widely used because they can model nonlinear relationships between input and output variables and recognize complex educational data patterns [3, 4]. ANN is also relevant for competency-based curricula because it can capture gradual and hierarchical relationships among subjects in automotive engineering education [5].

However, the effectiveness of ANN models strongly depends on appropriate hyperparameter configuration. An improper network structure may reduce the model's generalization ability and increase the risk of overfitting [6, 7].

*Correspondence to: Sudirman Rizki Ariyanto (Email: sudirmanariyanto@unesa.ac.id). Department of Automotive Engineering Technology, Universitas Negeri Surabaya. Ketintang Street, Surabaya, East Java Province, Indonesia (60231).

ISSN 2310-5070 (online) ISSN 2311-004X (print)

Hyperparameters such as the number of hidden layers, number of neurons, activation functions, learning algorithm, learning rate, and training epochs directly influence convergence behavior, prediction accuracy, and model stability [8, 9]. Therefore, hyperparameter optimization is a crucial step in developing AI-based decision support systems that are accurate, stable, and reliable for vocational education environments. Without systematic optimization, ANN models may produce different prediction errors across repeated training processes, especially when the available educational dataset is relatively small [10].

VHS with AVE programs, such as VHS NU 1 Karanggeneng in Lamongan, implement a competency-based curriculum that progresses from basic engineering science to advanced automotive diagnostics. This structured curriculum requires progressive integration of competencies, where students' academic performance in the early stages may influence their success in more advanced competencies. In practice, the identification of students at academic risk is still carried out conventionally through semester grade evaluations without the support of systematic predictive analytics systems [10, 11]. This condition may delay educational intervention and limit the effectiveness of data-driven learning strategies, particularly for students who need earlier academic support before entering advanced diagnostic competencies.

In this context, academic performance prediction is not only intended to estimate students' final achievement but also to provide a methodological basis for supporting the early identification of students who may experience difficulties in advanced competency stages [2, 11]. This issue is particularly important in automotive vocational education because early academic performance, basic technical knowledge, maintenance-related competencies, and diagnostic competencies are sequentially connected. Therefore, prediction models used in this context should not only produce low prediction errors but also demonstrate stable performance across repeated training processes [3, 8]. This stability requirement becomes essential because inconsistent prediction results may reduce the reliability of AI-based academic decision support systems in vocational education settings.

Previous studies have demonstrated that ANN-based prediction models can achieve competitive performance in academic performance prediction compared with conventional statistical methods and several machine learning algorithms [3, 4, 12, 13]. Systematic reviews in educational data mining also indicate that student performance prediction has increasingly shifted from descriptive statistical analysis to machine learning and deep learning models because educational data often contain nonlinear relationships, heterogeneous predictors, and complex interaction patterns [11, 14]. However, many ANN implementations still emphasize prediction accuracy without sufficiently explaining the hyperparameter selection process, validation strategy, repeated-training stability, and comparative baseline performance. This limitation is important because ANN performance can vary substantially when network architecture, activation functions, learning algorithms, learning rate, and training epochs are not systematically optimized and compared with simpler or standard predictive models [8, 9].

Hyperparameter optimization has therefore become an important issue in machine learning-based prediction studies. Recent studies have shown that hyperparameter tuning can be performed using several strategies, including manual tuning, grid search, random search, Bayesian optimization, and metaheuristic algorithms [8, 15, 16]. Among these approaches, metaheuristic algorithms such as Genetic Algorithms (GA), Particle Swarm Optimization (PSO), and other evolutionary methods are widely used because they can explore complex and nonlinear solution spaces without requiring derivative information [17, 18, 19]. Nevertheless, purely metaheuristic search may require many evaluations when the model training process is computationally expensive. Therefore, combining statistical experimental design with evolutionary optimization can provide a more efficient alternative for identifying influential factors and searching for near-optimal hyperparameter configurations.

In this regard, Design of Experiments (DOE) approaches such as CCD and RSM are useful for reducing the number of experimental runs while still allowing the analysis of main effects, interaction effects, and quadratic effects [20, 21]. RSM has been widely applied in engineering optimization because it can construct a surrogate model that represents the relationship between input factors and performance responses [20, 22, 23]. Hybrid approaches that integrate RSM, ANN, and GA have also been reported in engineering and energy-related optimization problems, demonstrating that surrogate modeling and evolutionary search can be combined to explore nonlinear response surfaces [24, 25, 26]. However, the use of this hybrid framework for ANN hyperparameter tuning in automotive vocational education remains limited, particularly when model stability across repeated training and comparison with standard baseline models are considered.

Although RSM–GA-based optimization has been applied in several engineering and computational domains [20, 24, 25, 26], its use for ANN hyperparameter optimization in automotive vocational education remains underexplored. Therefore, this study does not claim the general integration of RSM, GA, and ANN as a fundamentally new optimization method. Instead, the novelty lies in the domain-specific adaptation, structured implementation, and empirical validation of a CCD–RSM–GA-based ANN tuning workflow for automotive vocational education data. Unlike the authors' previous study [27], which primarily demonstrated the feasibility of CCD–RSM–GA-based ANN optimization using a different vocational education dataset, the present study extends the framework through cohort-based temporal validation, leakage-aware preprocessing, robustness-oriented evaluation, GA reproducibility analysis, and expanded baseline benchmarking against Linear Regression, Random Forest, Default ANN, and Random Search ANN. Consequently, the proposed workflow is presented as a more rigorously validated and reproducible approach for ANN-based academic performance prediction in automotive vocational education rather than as a newly invented optimization algorithm.

Based on these research gaps, this study applies and evaluates a structured CCD–RSM–GA workflow to optimize ANN hyperparameters for predicting the academic performance of Automotive Vocational Education students at VHS NU 1 Karanggeneng, Lamongan. The main contributions of this study are fourfold: first, applying a CCD–RSM–GA framework for ANN hyperparameter tuning in automotive vocational education; second, evaluating ANN performance by considering both prediction response and repeated-training stability through MSE, SD MSE, and S/N Ratio; third, identifying dominant ANN hyperparameters and their interaction effects using ANOVA and response surface analysis; and fourth, evaluating the optimized ANN configuration through cohort-based temporal validation and supplementing the evaluation with a normalized repeated MSE response-level benchmark against Linear Regression, Random Forest, Default ANN, Random Search ANN, and Opt CCD ANN. Accordingly, the objectives of this study are to identify significant ANN parameters, construct a quadratic RSM model, determine the optimal ANN hyperparameter combination using GA, and evaluate the optimized configuration in improving prediction response and repeated-performance stability relative to relevant baseline models.

## 2. Method

### *2.1. Research Design and Dataset*

This study employs a quantitative experimental approach by integrating ANN, RSM, and GA to optimize ANN hyperparameters for predicting students' academic performance [20, 21, 28]. The use of ANN is relevant for academic performance prediction because this model can capture nonlinear relationships among educational variables and support predictive analytics in educational decision-making [1, 2, 3]. The research subjects were students in grades X–XII of the Automotive Engineering program at VHS NU 1 Karanggeneng, Lamongan, East Java, Indonesia. The initial academic record dataset consisted of 241 student records collected during the 2021–2023 period. After applying data completeness criteria for the required input and target variables, 153 complete student records were retained as the final modeling dataset. These complete records consisted of three academic cohorts, namely 51 students from 2021, 51 students from 2022, and 51 students from 2023. The use of a relatively small but complete educational dataset requires careful model development and validation because limited data may affect the stability and generalization ability of predictive models [10]. The data covered semesters 1 to 5 and included 11 subjects representing gradual competency stages, namely: (1) basic engineering science, (2) basic automotive technology, (3) maintenance, and (4) diagnostics.

Each numerical variable was normalized to maintain the stability of the ANN training process and avoid the dominance of a particular scale in the network learning process. To prevent data leakage, the normalization parameters were estimated only from the training set and then applied consistently to the validation and testing sets. This procedure ensured that information from the validation and testing cohorts was not used during preprocessing, model training, or hyperparameter optimization [29]. The input variables consisted of subject scores from the first to fourth semesters, while the output variable was the final vehicle diagnostic competency score measured on a continuous scale ranging from 0 to 100. This score represents students' achievement in the diagnostic competency stage and was used as the prediction target in the ANN model. Because the target variable is continuous rather than

categorical, the ANN was formulated as a regression model [30]. The 2021 cohort was used for model training, the 2022 cohort for validation and hyperparameter optimization, and the 2023 cohort was retained as an unseen testing set and used only for final out-of-sample evaluation. The research design adopted a CCD-based experimental approach to evaluate the effect of ANN hyperparameter combinations on model performance responses. CCD was used to reduce the number of experimental combinations while still enabling the evaluation of main effects, interaction effects, and quadratic effects in the RSM framework [20, 21]. The general research flow is shown in Figure 1.

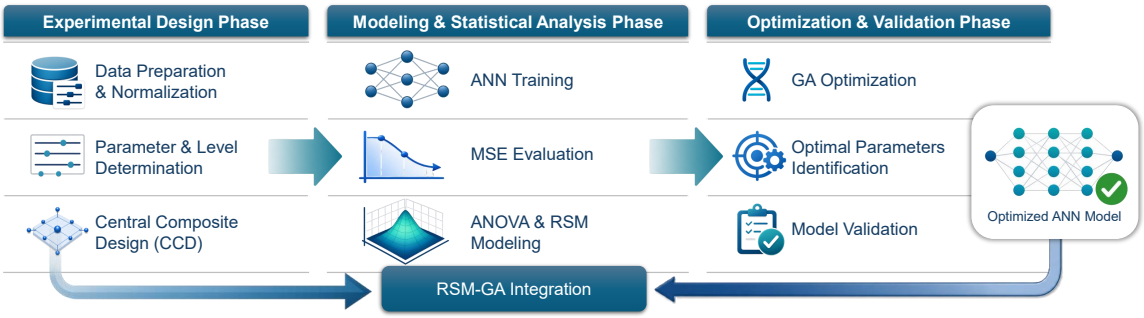



Figure 1. Research flow of the RSM–GA integration for ANN hyperparameter optimization.

### 2.2. *Data Splitting and Temporal Validation*

Following the dataset definition described in the previous subsection, this study applied a cohort-based temporal validation strategy to reduce the possibility of data leakage and to provide a more realistic evaluation of predictive performance [29, 31]. The 2021 cohort was used as the training set, the 2022 cohort was used as the validation and hyperparameter optimization set, and the 2023 cohort was retained as the unseen testing set. This temporal splitting strategy was adopted so that the model was developed using earlier academic records and evaluated using later student records that were not involved in model development.

The training set was used to fit the ANN models during the initial learning process. The validation set was used to evaluate the hyperparameter combinations generated through the CCD–RSM procedure and to guide the GA-based optimization process. In contrast, the testing set was used only for the final out-of-sample evaluation of the optimized ANN configuration. Importantly, the 2023 testing cohort was not used in ANN training, CCD experimentation, RSM model construction, GA optimization, hyperparameter selection, or model retraining. This separation was intended to ensure that the reported final performance reflects predictive ability on unseen data rather than optimization performance on the same dataset [29].

All preprocessing steps were also performed in a leakage-aware manner. The normalization parameters were estimated only from the training set and then applied consistently to the validation and testing sets. Therefore, information from the validation and testing cohorts did not influence the scaling process used during model training. This procedure is important because predictive model performance should be assessed using data that are not involved in model development, preprocessing parameter estimation, or hyperparameter optimization [29]. The temporal data splitting scheme used in this study is presented in Table 1.

### 2.3. *Artificial Neural Network Configuration and Experimental Design*

The prediction model was constructed using a multilayer feedforward Artificial Neural Network trained using a backpropagation learning algorithm [30]. Several hyperparameters were considered in the optimization process, including the number of hidden layers, number of neurons per layer, activation functions, training algorithms, learning rate, and number of training epochs [8, 9, 17]. These hyperparameters directly influence the convergence

Table 1. Cohort-based temporal data splitting scheme

| Academic Year | Number of Students | Data Subset | Purpose |
|---|---|---|---|
| 2021 | 51 | Training set | ANN model training |
| 2022 | 51 | Validation set | CCD–RSM hyperparameter evaluation |
| 2023 | 51 | Testing set | Final out-of-sample evaluation |
| Total | 153 | Complete modeling dataset | Model development and evaluation |

speed, stability, and generalization capability of the neural network model. The experimental levels and coding–decoding rules of the ANN hyperparameters are shown in Table 2.

Table 2. ANN hyperparameter levels and coding–decoding rules

| Parameter | Symbol | $-1$ | 0 | $+1$ |
|---|---|---|---|---|
| Hidden layer | HL | 2 | 3 | 4 |
| Neuron | NN | 10 | 20 | 30 |
| Hidden TF | TH | logsig | purelin | tansig |
| Output TF | TO | logsig | purelin | tansig |
| Learning algorithm | LA | trainlm | trainbr | traingdx |
| Learning rate | LR | 0.001 | 0.005 | 0.010 |
| Epoch | EP | 100 | 200 | 300 |

Table 2 presents the coded levels and actual values of the ANN hyperparameters used in the CCD–RSM and GA procedures. The coded values of $-1$, 0, and $+1$ were used to represent the predefined experimental levels of each hyperparameter. For numerical hyperparameters, such as hidden layers, neurons, learning rate, and epochs, the coded levels corresponded to specific numerical values. For categorical hyperparameters, such as hidden transfer function, output transfer function, and learning algorithm, the coded levels were used only as empirical design labels within the CCD–RSM framework. These coded values were not interpreted as physically continuous numerical quantities. Instead, they served as structured surrogate indicators to compare the performance tendency among predefined ANN categories. This treatment is consistent with the practical nature of hyperparameter optimization, where search spaces may include numerical, discrete, and categorical variables that must be encoded before optimization [32]. Therefore, any linear, quadratic, or interaction terms involving categorical factors were interpreted cautiously as empirical effects within the selected design space rather than as continuous functional relationships.

Theoretically, the full combination of three levels for seven parameters produces $3^7 = 2187$ experiments. To improve efficiency, this study applied Central Composite Design (CCD), which reduced the number of experiments to 100 parameter combinations while still allowing the estimation of main effects, interaction effects, and quadratic effects [20, 21, 31]. Each configuration was repeated five times to obtain stable performance measurements [27]. In accordance with the temporal validation protocol described in the previous subsection, the CCD-based hyperparameter evaluation was performed using the training and validation subsets only. The unseen testing set was excluded from CCD experimentation, hyperparameter evaluation, RSM modeling, and GA optimization to maintain a strict out-of-sample evaluation protocol.

Model performance during the CCD experiment was evaluated using Mean Squared Error (MSE), standard deviation of MSE, and Signal-to-Noise (S/N) ratio. MSE was used as the primary prediction error metric because it measures the average squared difference between the actual and predicted values [29]. It is calculated as

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{1}$$

where $y_i$ represents the actual value, $\hat{y}_i$ represents the predicted value, and $n$ is the number of evaluated samples. Because ANN training involves random initialization and may produce different prediction errors across repeated runs, MSE alone is not sufficient to evaluate the robustness of a hyperparameter configuration. Therefore, each ANN configuration was repeated five times, and the mean MSE and standard deviation of MSE were calculated to represent prediction accuracy and repeated-training variability. This repeated-evaluation procedure was conducted within the training–validation stage and did not involve the unseen testing set.

The S/N ratio with the Smaller-the-Better criterion was used as a robustness-oriented response in the CCD–RSM modeling and as the fitness criterion in the GA optimization stage. The use of the S/N ratio was not intended to replace MSE, but to prioritize hyperparameter configurations that produced both low prediction error and stable performance across repeated ANN training runs. This is important in small educational datasets because a configuration with low MSE in a single run may still be unreliable if its error varies substantially across repeated training processes. The S/N ratio was calculated as follows:

$$S/N = -10 \log_{10} \left( \frac{1}{r} \sum_{k=1}^{r} MSE_k^2 \right) \tag{2}$$

where $MSE_k$ is the MSE value obtained from the $k$-th repeated training run and $r$ is the number of repeated runs. In this study, $r = 5$. A higher S/N ratio indicates a more robust configuration because it reflects a smaller and more stable prediction error across repeated runs [33, 34]. Accordingly, the CCD–RSM stage used the S/N ratio to model the robustness response, while the final model evaluation still reported MSE, standard deviation of MSE, S/N ratio, correlation coefficient, $R^2$, and Adjusted $R^2$ on the unseen testing set.

### *2.4. Response Surface Modeling and Analysis of Variance*

The CCD experimental results obtained from the training–validation procedure were analyzed using Analysis of Variance (ANOVA) to identify ANN hyperparameters that significantly affected the S/N Ratio response [33, 34]. ANOVA separates the total variation into variation caused by main factors, interaction effects, quadratic effects, and experimental error. Parameters were considered significant when the p-value was less than 0.05. The unseen testing set was not involved in the ANOVA procedure, RSM model construction, or factor significance assessment. This separation was maintained to ensure that the response surface model was developed only from the hyperparameter evaluation stage and did not incorporate information from the final out-of-sample testing data [29].

Next, a second-order polynomial regression model was constructed to map the relationship between the coded ANN hyperparameters and the S/N Ratio response within the RSM framework [20, 21]:

$$Y = \beta_0 + \sum \beta_i X_i + \sum \beta_{ii} X_i^2 + \sum \beta_{ij} X_i X_j + \varepsilon \tag{3}$$

where $Y$ is the S/N Ratio response, $X_i$ represents the coded factor level, $\beta_0$ is the intercept, $\beta_i$ is the linear coefficient, $\beta_{ii}$ is the quadratic coefficient, $\beta_{ij}$ is the interaction coefficient, and $\varepsilon$ represents the experimental error.

Because several ANN hyperparameters were categorical, including hidden transfer function, output transfer function, and learning algorithm, the coded levels in the RSM equation were interpreted as empirical design labels rather than physically continuous numerical quantities. Therefore, polynomial terms involving categorical factors were used to indicate empirical performance tendencies among predefined coded categories within the CCD design space. They were not interpreted as continuous functional relationships. This treatment is consistent with mixed-variable and surrogate-assisted optimization, where numerical, discrete, and categorical variables require encoding before optimization and cautious interpretation in the fitted surrogate model [32, 35].

Model quality was evaluated using the coefficient of determination ($R^2$), Adjusted $R^2$, Predicted $R^2$, and residual analysis. The $R^2$ value was used to describe the proportion of response variation explained by the quadratic model, while Adjusted $R^2$ accounted for the number of predictors in the model [36]. Predicted $R^2$ provided an additional indication of the model's predictive capability within the experimental design space. Residual analysis was conducted to examine the normality and homoscedasticity assumptions of the fitted model.

The fitted RSM model was interpreted as a surrogate approximation of the ANN hyperparameter response surface rather than as a perfect representation of the original ANN performance landscape. This interpretation is

important because the model was built from CCD experimental results and included both numerical and categorical hyperparameter settings. In surrogate-assisted optimization, approximate response models are useful for guiding the search toward promising regions, but the resulting optimum still requires final validation using the target model or independent evaluation data [29, 35, 37]. Therefore, model adequacy indicators such as $R^2$, Adjusted $R^2$, and Predicted $R^2$ were used to assess the usefulness of the surrogate model for guiding optimization, not to claim definitive out-of-sample predictive performance.

The resulting quadratic regression model was then used as the objective function for the GA optimization stage. Because the objective function was derived from the RSM surrogate model, the optimum obtained from GA was subsequently decoded into actual ANN hyperparameters and validated according to the temporal validation protocol described previously. This procedure ensured that the final reported performance was based on the unseen testing set rather than on the same data used to construct the RSM model or guide the GA search [35, 37].

### *2.5. Optimization Using Genetic Algorithms and Model Validation*

After the RSM surrogate model was constructed, the resulting second-order polynomial equation was used as the objective function in the Genetic Algorithm optimization stage. The GA was applied to search for the ANN hyperparameter combination that maximized the predicted S/N Ratio response. This procedure enabled the exploration of the hyperparameter space beyond the discrete CCD experimental points while still using the response pattern estimated from the CCD–RSM stage. Such a surrogate-guided evolutionary search is useful for reducing direct evaluation cost and exploring mixed numerical, discrete, and categorical decision spaces [35, 36, 37]. Because the GA search was conducted using the RSM surrogate model derived from the training–validation procedure, the unseen testing set was not involved in the optimization process.

The GA population consisted of 100 candidate solutions representing coded ANN hyperparameter values. Each individual encoded the seven optimized hyperparameters, namely hidden layers, number of neurons, hidden transfer function, output transfer function, learning algorithm, learning rate, and training epochs. The fitness value of each individual was calculated from the predicted S/N Ratio generated by the RSM regression equation. The GA used roulette wheel selection, a crossover rate of 0.6, a mutation rate of 0.1, and an elitism mechanism of 1% to preserve the best individual in each generation. These operators were used because GA-based optimization relies on population selection, recombination, mutation, and elitism to balance exploration and exploitation during the search process [24, 38, 39]. The final GA experiment was executed for a maximum of 1000 generations, and convergence was evaluated using the best S/N fitness, mean S/N fitness, repeated-run sensitivity analysis, and population diversity.

To improve reproducibility, the GA configuration was fixed across all optimization runs. The final configuration consisted of a population size of 100, roulette wheel selection, crossover rate of 0.6, mutation rate of 0.1, elitism of 1%, and a maximum of 1000 generations. The previous stopping statement based on 100 generations or 20 generations without significant improvement was removed to ensure consistency between the Method and Results sections. In addition, the GA was executed in five independent runs using different random seeds to examine the stability of the optimization result. The coding–decoding procedure used in this stage followed the ANN hyperparameter levels and coding–decoding rules presented in Table 2. Because GA operates in a continuous coded search space, the optimal coded solution was first mapped to the nearest valid coded level before being implemented in the ANN model. Each continuous coded value was assigned to one of the three predefined coded levels, namely $-1$, 0, or $+1$, and then converted into the corresponding actual ANN setting. This rule was applied to both numerical and categorical hyperparameters. For categorical factors, such as transfer functions and learning algorithms, the mapped coded level was used only to select the nearest predefined category and was not interpreted as a continuous numerical value [32, 35]. In the final GA solution, TO = 0.30294 was mapped to level 0 and decoded as *purelin*, while LA = $-0.93237$ was mapped to level $-1$ and decoded as *trainlm*.

After applying this mapping rule, the final GA-derived ANN configuration consisted of three hidden layers, twenty neurons, *tansig* as the hidden transfer function, *purelin* as the output transfer function, *trainlm* as the learning algorithm, a learning rate of 0.005, and 200 epochs. This procedure ensured that the final ANN model used valid implementable hyperparameter settings rather than continuous coded GA outputs. The optimized ANN configuration obtained from the RSM–GA procedure was then validated according to the temporal validation

protocol described previously. The final ANN model was trained using the training set, while the optimized hyperparameter configuration was selected through the validation and optimization stage. The unseen 2023 testing set was used only for the final out-of-sample evaluation. Therefore, the testing set was not involved in ANN training, CCD experimentation, RSM model construction, GA optimization, hyperparameter selection, sensitivity analysis, or model retraining. This separation was intended to ensure that the reported final performance represents out-of-sample predictive capability rather than performance obtained from the same data used during model optimization [29].

The final model performance was evaluated using Mean Squared Error (MSE), standard deviation of MSE, S/N Ratio, correlation coefficient ($R$), coefficient of determination ($R^2$), and Adjusted $R^2$. MSE was used to evaluate prediction error, while the standard deviation of MSE and S/N Ratio were used to assess repeated-training stability. The correlation coefficient, $R^2$, and Adjusted $R^2$ were used to describe the agreement between predicted and actual values on the unseen testing set [18, 29]. Accordingly, the final validation stage served as an independent evaluation of the optimized ANN configuration and was not used to further tune, select, or revise the hyperparameters.

In addition, a response-level benchmark was conducted to compare the optimized ANN configuration with Linear Regression, Random Forest, Default ANN, Random Search ANN, and Opt CCD ANN. The benchmark used normalized repeated MSE response summarized by Mean MSE, SD MSE, and S/N Ratio. ANN-based models were summarized using repeated best performance records during training, whereas Linear Regression and Random Forest were summarized using normalized training-performance MSE as performance-record-equivalent baselines. Therefore, this benchmark should not be interpreted as a strict final testing-set comparison for all models.

## 3. Results and Discussion

### *3.1. Experimental Results Using Central Composite Design*

The CCD-based experiment produced 100 ANN hyperparameter combinations that were evaluated using Mean Squared Error (MSE), standard deviation of MSE, and S/N Ratio with the Smaller-the-Better criterion. Based on the experimental results, the mean MSE values ranged from 0.17 to 33.78, while the S/N Ratio ranged from $-30.62$ to 12.04. This wide range indicates that the ANN hyperparameter configuration has a substantial influence on both prediction accuracy and repeated-training stability. This finding is consistent with previous studies showing that ANN performance is highly sensitive to network architecture, activation function settings, training algorithms, and learning parameters [40, 41]. To provide a clearer comparative overview, Table 3 presents the best and worst CCD configurations based on the S/N Ratio values.

Table 3. Comparison of best and worst CCD configurations

| Exp. | Configuration | Mean MSE | SD MSE | S/N Ratio |
|---|---|---|---|---|
| 70 | Best CCD | 0.17 | 0.1855 | 12.04 |
| 56 | Worst CCD | 33.78 | 3.5437 | $-30.62$ |

Table 3 shows that Experiment 70 produced the best CCD configuration, with a mean MSE of 0.17, an SD MSE of 0.1855, and an S/N Ratio of 12.04. In contrast, Experiment 56 produced the worst CCD configuration, with a mean MSE of 33.78, an SD MSE of 3.5437, and an S/N Ratio of $-30.62$. These results indicate that the best configuration not only produced a lower prediction error but also demonstrated more stable repeated-training performance. Therefore, the S/N Ratio is useful as a robustness-oriented response because it helps identify ANN configurations that combine low error with lower variability across repeated training runs.

It should be noted that Table 3 represents the CCD-stage experimental response, whereas the later baseline comparison reports a normalized repeated MSE response benchmark across baseline and optimized models. Therefore, the two tables serve different analytical purposes and should not be interpreted as identical evaluation scales.

The best configuration was obtained in Experiment 70 with a combination of three hidden layers, twenty neurons, the *tansig* activation function in the hidden layer, the *purelin* function in the output layer, the *trainbr* algorithm, a

learning rate of 0.005, and 200 epochs. The mean MSE value of 0.17 indicates a low prediction error, while the SD MSE value of 0.1855 indicates relatively stable performance across repeated training runs. In addition, the S/N Ratio value of 12.04 supports the robustness of this configuration because it reflects a favorable combination of low error and repeated-training stability. These results are consistent with previous studies indicating that the Bayesian Regularization algorithm (*trainbr*) can reduce overfitting and improve generalization capability in medium-sized datasets [15, 42].

In contrast, the worst configuration was obtained in Experiment 56, which produced a mean MSE of 33.78, an SD MSE of 3.5437, and an S/N Ratio of $-30.62$. This configuration used the *traingdx* algorithm with a high learning rate of 0.01 and a more complex architecture consisting of four hidden layers and thirty neurons. The high mean MSE indicates poor prediction accuracy, while the high SD MSE indicates unstable performance across repeated training runs. This result suggests that excessive architectural complexity combined with an unsuitable training setting may lead to convergence instability and poor generalization. This finding supports previous research highlighting the importance of balancing model complexity and generalization performance [43].

Statistically, most experiments using the *logsig* activation function produced negative S/N Ratio values below $-20$, while the *tansig–purelin* combination generated more stable results. This phenomenon can be explained by the nonlinear characteristics of academic performance data that still follow continuous regression patterns. The hyperbolic tangent function can support nonlinear representation in the hidden layer, while the linear output function maintains numerical stability for continuous prediction tasks [44, 45]. The distribution of S/N Ratio values also indicates that only a small proportion of parameter combinations produced values above the threshold of $-10$. This observation suggests that identifying an optimal configuration through trial-and-error approaches would be inefficient. Therefore, the use of systematic experimental approaches such as Design of Experiments (DOE), CCD, and RSM becomes essential for ANN hyperparameter optimization [46, 47].

### 3.2. *Analysis of Variance (ANOVA) and Model Adequacy*

After identifying performance variation through the CCD design, the next step was to evaluate the statistical significance of each factor using ANOVA. ANOVA was used to separate the total variation in the S/N Ratio response into variation due to main factors, quadratic effects, two-way interactions, lack-of-fit, pure error, and residual error. This approach is commonly used in the RSM framework to identify influential parameters and to evaluate whether the fitted model provides a statistically meaningful representation of the experimental response [25, 48]. The verified ANOVA results are presented in Table 4. The values reported in the table and the accompanying discussion were rechecked to ensure numerical consistency.

Table 4. ANOVA summary for the S/N Ratio quadratic model

| Source | DF | SS | MS | F-value | Prob. > F |
|---|---|---|---|---|---|
| TH | 1 | 70.16127 | 70.16127 | 8.09888 | 0.0055 |
| TO | 1 | 989.51288 | 989.51288 | 114.22178 | < 0.0001 |
| LA | 1 | 1042.57378 | 1042.57378 | 120.34673 | < 0.0001 |
| LR | 1 | 26.28522 | 26.28522 | 3.03416 | 0.0850 |
| $TH^2$ | 1 | 506.37176 | 506.37176 | 58.45167 | < 0.0001 |
| $TO^2$ | 1 | 343.70357 | 343.70357 | 39.67451 | < 0.0001 |
| $LR^2$ | 1 | 58.78718 | 58.78718 | 6.78594 | 0.0108 |
| TH×TO | 1 | 30.79404 | 30.79404 | 3.55463 | 0.0626 |
| TO×LA | 1 | 626.04796 | 626.04796 | 72.26618 | < 0.0001 |
| LA×LR | 1 | 40.20995 | 40.20995 | 4.64153 | 0.0339 |
| Error | 89 | 771.01447 | 8.66308 | | |
| Lack-of-fit | 14 | 427.45317 | 30.53237 | 6.66527 | < 0.0001 |
| Pure Error | 75 | 343.56130 | 4.58082 | | |
| Total | 99 | 4307.02017 | | | |

The ANOVA results in Table 4 show that several ANN hyperparameters significantly affected the S/N Ratio response at the 5% significance level. The largest F-value was observed for the Learning Algorithm (LA), with $F = 120.34673$ and $p < 0.0001$, followed by the Output Transfer Function (TO), with $F = 114.22178$ and

$p < 0.0001$. These results indicate that the learning algorithm and output transfer function had the strongest empirical influence on ANN robustness within the tested design space. This finding is consistent with previous studies showing that training algorithms and output activation settings can strongly affect convergence behavior, prediction stability, and generalization in neural network regression problems [49, 50, 51].

The quadratic terms $TH^2$, $TO^2$, and $LR^2$ were also significant, with F-values of 58.45167, 39.67451, and 6.78594, respectively. These results indicate that the S/N Ratio response did not change only in a linear manner across the selected hyperparameter levels. The linear effect of LR was not significant ($p = 0.0850$), whereas its quadratic effect was significant ($p = 0.0108$), suggesting that the learning rate influenced model robustness mainly through curvature rather than a simple linear trend. Among the interaction terms, TO×LA and LA×LR were significant, while TH×TO was not significant at the 5% level ($p = 0.0626$). Therefore, interaction effects were interpreted selectively and cautiously.

Because several factors in the model were categorical, including hidden transfer function, output transfer function, and learning algorithm, the ANOVA terms involving these factors were interpreted as empirical effects among coded categories rather than as physically continuous functional relationships. Thus, the significance of LA, TO, TO×LA, or other categorical-related terms indicates performance differences among predefined ANN settings within the CCD design space. It does not imply that the categories themselves form a physically continuous scale. This cautious interpretation is important in mixed-variable and surrogate-assisted optimization, where numerical, discrete, and categorical variables must be encoded before modeling and optimization [32, 35].

Table 5. Model fit statistics for the quadratic RSM model

| Statistic | Value |
|---|---|
| $R^2$ | 0.82099 |
| Adjusted $R^2$ | 0.80087 |
| Predicted $R^2$ | 0.68569 |
| Root-MSE | 2.94331 |
| PRESS | 1353.73951 |

The model fit statistics in Table 5 show that the quadratic RSM model explained a substantial proportion of the variation in the S/N Ratio response, as indicated by an $R^2$ value of 0.82099 and an Adjusted $R^2$ value of 0.80087. However, this result should not be overinterpreted as evidence of strong out-of-sample prediction capability. The Predicted $R^2$ value of 0.68569 was lower than the Adjusted $R^2$, indicating that the model's predictive ability within the design space was weaker than its explanatory fit to the CCD data. Therefore, the RSM model was treated as a surrogate approximation rather than a final predictive model [29, 37, 52].

The lack-of-fit result was significant, with $F = 6.66527$ and $p < 0.0001$, indicating that part of the S/N Ratio variation was not fully captured by the second-order polynomial model. This result is reasonable because ANN hyperparameter performance may involve higher-order nonlinearities, stochastic training behavior, and mixed numerical-categorical interactions. A significant lack-of-fit does not necessarily invalidate the use of the RSM model for optimization, provided that the model is interpreted as an approximate surrogate rather than a complete representation of the original response landscape [37, 53, 54]. Accordingly, the RSM model was used only to guide the GA search, and the decoded optimum was subsequently evaluated using the unseen 2023 testing cohort.

### *3.3. Response Surface Analysis*

Based on the ANOVA results, a second-order polynomial regression model was constructed to represent the relationship between coded factors and the S/N Ratio response:

$$\begin{aligned} S/N = &- 16.63254 + 1.03104(TH) + 3.87203(TO) - 3.97449(LA) \\ &- 0.63108(LR) + 13.09413(TH^2) - 10.78783(TO^2) \\ &- 4.46152(LR^2) + 0.69365(TH \cdot TO) \\ &- 3.12762(TO \cdot LA) - 0.79264(LA \cdot LR) \end{aligned} \tag{4}$$

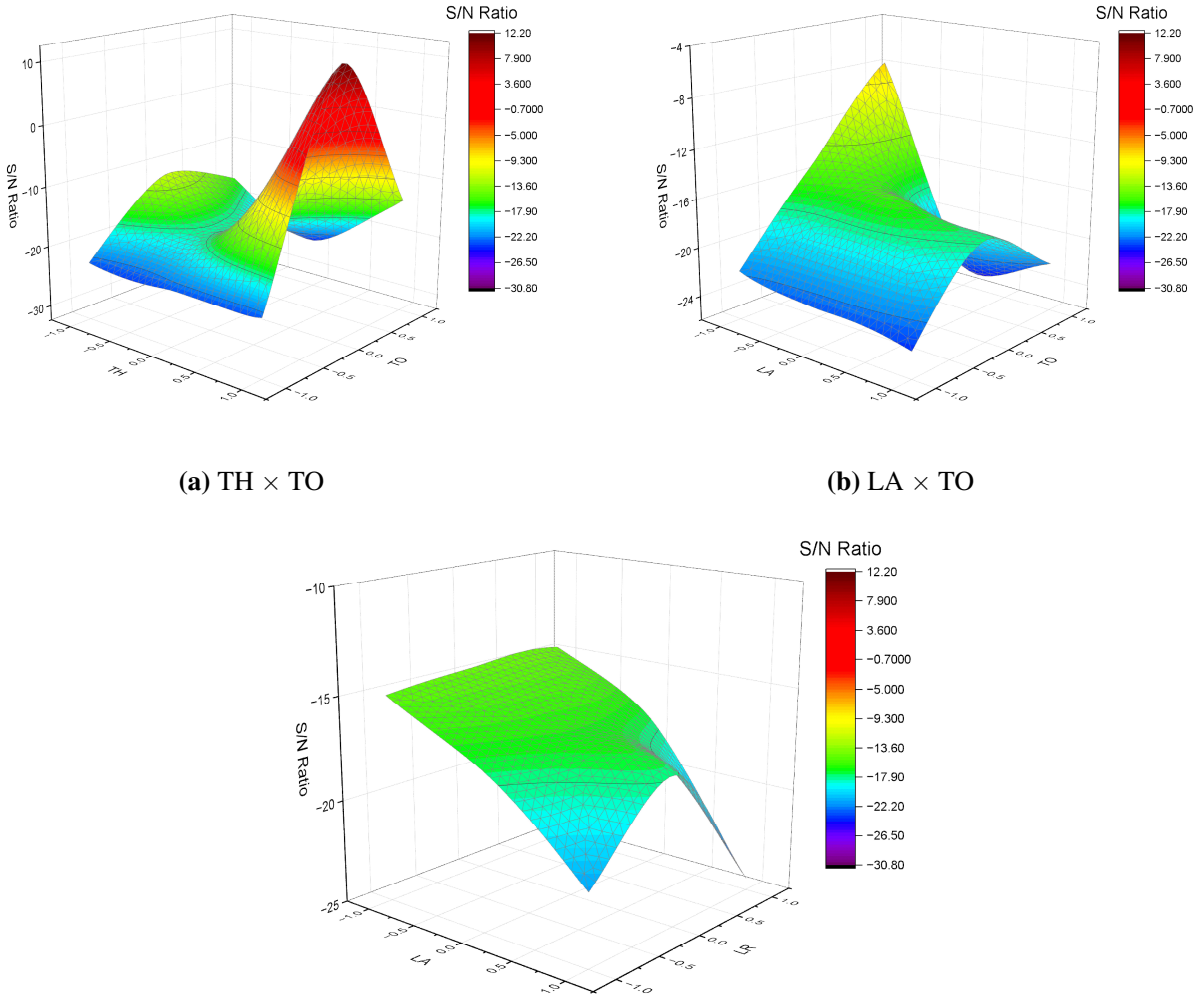
The response variable modelled in the RSM stage was the Signal-to-Noise (S/N) Ratio obtained from repeated ANN training experiments rather than the raw Mean Squared Error (MSE). Therefore, the ANOVA results, quadratic regression model, and model fit statistics reported in Table 5, including $R^2$, Adjusted $R^2$, Predicted $R^2$, Root-MSE, and PRESS, were evaluated with respect to the S/N Ratio response. In contrast, the Mean MSE and SD MSE values reported elsewhere in this study were used separately to assess prediction accuracy and repeated-training stability. Consequently, the Root-MSE value of 2.94331 and PRESS value of 1353.73951 should be interpreted as measures of prediction error and predictive adequacy for the fitted S/N Ratio model rather than for the ANN MSE values themselves. This interpretation is consistent with the use of RSM as a response modelling and optimization tool in experimental studies [55].

The equation shows that the largest linear contributions come from LA with a coefficient of $-3.97449$ and TO with a coefficient of $+3.87203$. The high absolute values of the coefficients and statistical significance ($p < 0.0001$) confirm that the selection of the training algorithm and output activation function are the main determinants of ANN model stability in predicting vocational academic performance. The negative sign on LA indicates that changes in certain algorithm levels decrease the S/N Ratio, while the positive contribution of TO indicates that certain output activation functions increase prediction stability. These findings are consistent with the theory that regularization-based algorithms such as *trainbr* have better generalization capabilities on heterogeneous education datasets [53], while linear activation functions in the output layer are more suitable for continuous regression problems [54, 56].

The presence of significant quadratic coefficients, particularly $\text{TH}^2$ ($+13.09413$) and $\text{TO}^2$ ($-10.78783$), indicates strong nonlinear curvature in the response landscape. In the RSM framework, significant quadratic coefficients indicate the existence of local optimum points in the parameter space rather than a simple linear relationship [55, 57]. Positive values in $\text{TH}^2$ form convex curvature, while negative values in $\text{TO}^2$ form concave curvature, which geometrically produces a closed elliptical response surface. This condition confirms that ANN hyperparameter optimization in this dataset is nonlinear and requires multidimensional exploration. The interaction between TH and TO is visualized in Fig. 2(a). The contours show that the maximum response region occurs at the combination of *tansig* in the hidden layer and *purelin* in the output layer. The relatively symmetrical elliptical contour pattern indicates the dominance of quadratic effects over pure linear interactions. The hyperbolic function in the hidden layer allows modeling gradual nonlinear relationships in academic data, while the linear function in the output layer maintains the stability of numerical predictions [56, 58]. Therefore, this combination produces higher S/N Ratio values compared to other configurations.

The most significant interaction was identified between LA and TO, with a coefficient of $-3.12762$ ($p < 0.0001$). This relationship is illustrated in Fig. 2(b), which shows a sharp response gradient between the *trainbr–purelin* and *traingdx–logsig* combinations. The asymmetric contours with steep slopes indicate a strong nonlinear interaction, where a particular training algorithm becomes optimal only when combined with the appropriate output activation function. The ANN optimization literature indicates that training algorithms containing internal regularization mechanisms are highly sensitive to activation function structures and can improve convergence stability when appropriately combined [59, 60]. This explains why the CCD configuration using the *trainbr* and *purelin* combination produced the highest S/N Ratio value in the experiment. Furthermore, the interaction between LA and LR has a coefficient of $-0.79264$ with a significance of $p = 0.03391$. This relationship is visualized in Fig. 2(c). The contour plot shows that low learning rate values (0.001–0.005) in the *trainbr* algorithm produce an optimal response region, while increasing the LR to 0.01 significantly decreases the S/N Ratio. Theoretically, an excessively high learning rate can cause gradient oscillations and worsen convergence, especially in algorithms with regularization such as *trainbr* [61]. The contour pattern showing a negative slope indicates a trade-off relationship, making simultaneous optimization of both parameters crucial.

Overall, the response surface analysis shows that the ANN performance landscape in this study is dominated by quadratic effects and significant two-way interactions. No extreme saddle point patterns were observed; instead, closed elliptical curvatures indicate the existence of locally promising regions within the investigated design space. The $R^2$ value of 0.82099 indicates that the quadratic model captured a substantial proportion of the variation in the S/N Ratio response. However, considering the significant lack-of-fit result ($F = 6.66527$, $p < 0.0001$), the RSM model should be interpreted as a surrogate approximation used to guide the GA search toward promising hyperparameter regions rather than as a definitive representation of the entire ANN performance landscape.




**(a)** TH × TO

**(b)** LA × TO

**(c)** LA × LR

Figure 2. Response surface plots showing the interaction effects of ANN hyperparameters on the S/N Ratio: (a) hidden transfer function and output transfer function (TH × TO), (b) learning algorithm and output transfer function (LA × TO), and (c) learning algorithm and learning rate (LA × LR).

Therefore, the RSM–GA optimum was subsequently verified through ANN retraining and final out-of-sample evaluation using the unseen testing cohort. In the context of complex ANN-based system optimization, reduced or surrogate model-based global optimization provides a practical mechanism for reducing the search space before final validation and model assessment [62, 63, 64].

### *3.4. Optimization Using Genetic Algorithms*

After the second-order RSM surrogate model was obtained, the equation was used as the fitness function in the Genetic Algorithm optimization stage. The GA was directed to maximize the predicted S/N Ratio response, so higher S/N fitness values indicate better hyperparameter configurations. This surrogate-guided search allowed the algorithm to explore the ANN hyperparameter space beyond the original CCD experimental points while still using the response pattern estimated from the RSM model. Such an approach is commonly used in surrogate-assisted

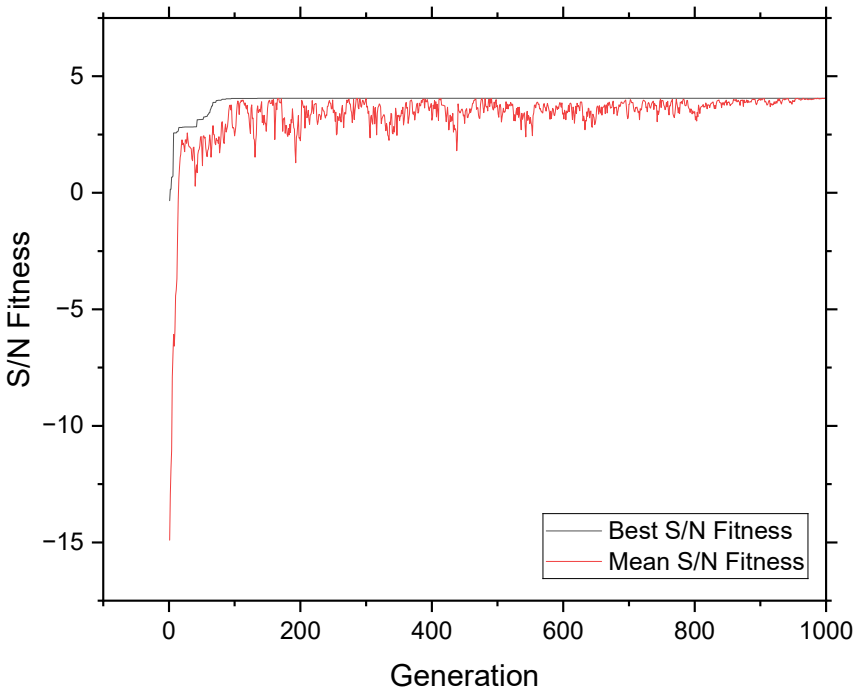



Figure 3. GA convergence curve showing the Best S/N Fitness and Mean S/N Fitness across 1000 generations. Higher S/N fitness values indicate better solutions.

evolutionary optimization because it reduces direct evaluation cost and supports the search of mixed numerical, discrete, and categorical decision spaces [21, 35, 37].

The GA convergence behavior is shown in Fig. 3. The figure presents the Best S/N Fitness and Mean S/N Fitness over 1000 generations. The Best S/N Fitness increased rapidly during the early generations and gradually stabilized in the later generations, indicating that the search process had moved toward a high-fitness region. At the 1000th generation, the Best S/N Fitness reached 4.0571, while the Mean S/N Fitness reached 4.0544. The small difference between these two values indicates that the population had become concentrated around a similar high-fitness solution. This pattern is consistent with evolutionary optimization behavior, where convergence is reflected by fitness stabilization and a narrowing gap between the best individual and the population mean [39, 65, 66].

The optimal solution obtained from the GA was interpreted through the coding–decoding rules described in Table 2. Because the GA operated in a coded search space, the continuous coded outputs were not implemented directly as ANN functions, algorithms, or numerical settings. Instead, each coded value was mapped to the nearest valid coded level and then converted into the corresponding actual ANN hyperparameter. This procedure is important in hyperparameter optimization because numerical, discrete, and categorical factors must be transformed into feasible model settings before final implementation [32, 35]. After decoding, the optimum ANN configuration consisted of three hidden layers, twenty neurons, *tansig–purelin* transfer functions, the *trainlm* learning algorithm, a learning rate of 0.005, and 200 epochs.

To evaluate the reproducibility of the GA result, five independent GA runs were conducted using different random seeds. The results are summarized in Table 6. The final Best S/N Fitness values were highly consistent across all runs, ranging only from 4.057118680 to 4.057123516. More importantly, all five runs produced the same decoded optimum configuration, namely three hidden layers, twenty neurons, *tansig–purelin* transfer functions,

the *trainlm* learning algorithm, a learning rate of 0.005, and 200 epochs. This consistency indicates that the GA optimization result was not highly sensitive to random initialization, which is important for reproducibility in stochastic evolutionary optimization [39, 67].

Table 6. GA sensitivity analysis using different random seeds

| Run | Seed | Best fit. | HL | NN | TH | TO | LA | LR | EP |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 4.057123516 | 3 | 20 | tansig | purelin | trainlm | 0.005 | 200 |
| 2 | 2 | 4.057123264 | 3 | 20 | tansig | purelin | trainlm | 0.005 | 200 |
| 3 | 3 | 4.057123414 | 3 | 20 | tansig | purelin | trainlm | 0.005 | 200 |
| 4 | 4 | 4.057123206 | 3 | 20 | tansig | purelin | trainlm | 0.005 | 200 |
| 5 | 5 | 4.057118680 | 3 | 20 | tansig | purelin | trainlm | 0.005 | 200 |

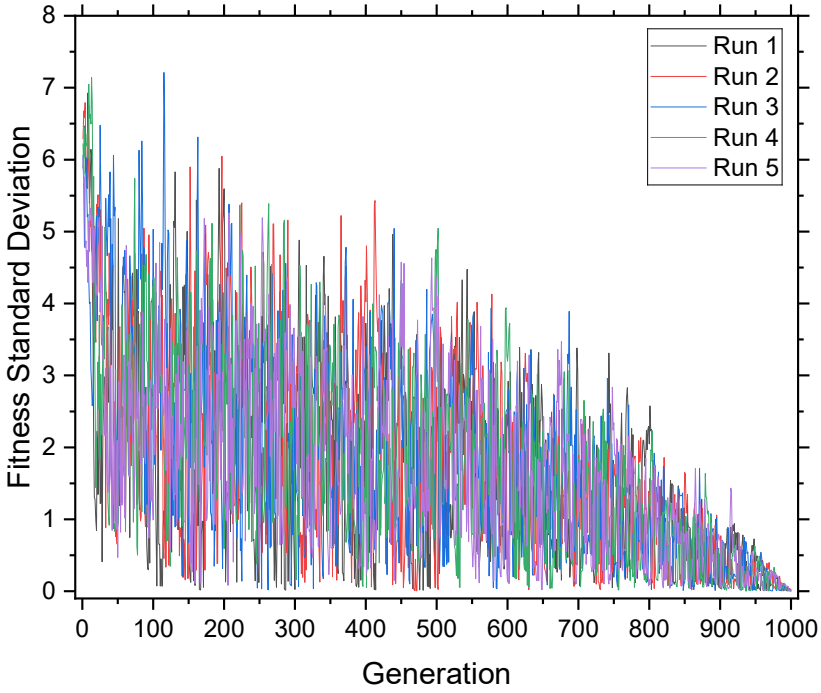



Figure 4. Population diversity during GA optimization based on the standard deviation of fitness values across generations for five independent runs.

In addition to convergence and repeated-run consistency, population diversity was analyzed to examine how the GA population evolved during the search process. As shown in Fig. 4, population diversity was measured using the standard deviation of fitness values across individuals in each generation. The diversity values were relatively high and fluctuated during the early generations, indicating broad exploration of the search space. As the generations progressed, the diversity gradually decreased and approached zero. At the final generation, the diversity values for Runs 1–5 were 0.004222, 0.022657, 0.008018, 0.019800, and 0.008275, with a mean final diversity of 0.012595. This decreasing trend indicates convergence toward narrow high-fitness regions [65, 66].

Overall, the GA optimization results show that the proposed RSM–GA framework produced a stable and reproducible ANN hyperparameter configuration. Figure 3 confirms convergence through the stabilization of the

Best S/N Fitness and Mean S/N Fitness, Table 6 shows that different random seeds produced the same decoded configuration, and Fig. 4 supports the convergence interpretation through decreasing population diversity. These three pieces of evidence indicate that the GA search was stable within the RSM surrogate landscape. However, because the optimization was guided by a surrogate model, the decoded optimum was not treated as final evidence of generalization until it was evaluated using the unseen testing cohort, consistent with surrogate-assisted validation principles [29, 37].

### *3.5. Model Validation and Baseline Comparative Analysis*

The model validation and baseline comparative analysis were conducted to address the reviewer's concern regarding the limited comparison between the Opt CCD ANN and Opt RSM–GA ANN configurations. In addition to the two optimized ANN configurations, this study included Linear Regression, Random Forest, Default ANN, and Random Search ANN as baseline models. The comparison was designed as a response-level benchmark based on normalized repeated MSE response. For ANN-based models, the reported values were obtained from repeated best performance records during model training. For Linear Regression and Random Forest, normalized training-performance MSE was used as a performance-record-equivalent baseline. Therefore, this comparison should be interpreted as a normalized MSE response-level benchmark rather than as a strict final testing-set comparison for all models. This strategy provides supplementary response-level evidence for assessing the response behavior and repeated-performance stability of the proposed CCD–RSM–GA framework relative to conventional statistical, ensemble-based, untuned ANN, and stochastic hyperparameter-search baselines. The inclusion of baseline and optimized models is important because predictive learning analytics studies should not only report prediction accuracy but also examine model errors, robustness, and practical reliability for educational decision-making [68, 69].

Table 7. Benchmark comparison of baseline models and optimized ANN configurations based on normalized repeated MSE response

| Model | Run 1 | Run 2 | Run 3 | Run 4 | Run 5 | Mean MSE | SD MSE | S/N Ratio |
|---|---|---|---|---|---|---|---|---|
| Linear Regression | 0.198 | 0.198 | 0.198 | 0.198 | 0.198 | 0.198 | 0.0000 | 14.05 |
| Random Forest | 0.124 | 0.123 | 0.122 | 0.120 | 0.124 | 0.122 | 0.0016 | 18.24 |
| Default ANN | 0.051 | 0.078 | 0.096 | 0.093 | 0.065 | 0.077 | 0.0171 | 22.10 |
| Random Search ANN | 0.269 | 0.276 | 0.415 | 0.310 | 0.243 | 0.302 | 0.0600 | 10.22 |
| Opt CCD ANN | 0.341 | 0.340 | 0.340 | 0.340 | 0.341 | 0.340 | 0.0005 | 9.36 |
| Opt RSM–GA ANN | 0.013 | 0.013 | 0.008 | 0.011 | 0.013 | 0.012 | 0.0019 | 38.65 |

Note: The reported values represent normalized repeated MSE responses. ANN-based models were summarized using repeated best performance records during model training, whereas Linear Regression and Random Forest were summarized using normalized training-performance MSE as performance-record-equivalent baselines. Therefore, this table should be interpreted as a response-level benchmark rather than as a strict final testing-set comparison.

As shown in Table 7, within this normalized response-level benchmark, the Opt RSM–GA ANN configuration produced the lowest Mean MSE and the highest S/N Ratio among the evaluated approaches. It should be noted that the Mean MSE value of 0.012 refers to the normalized repeated MSE response used in the benchmark, whereas the MSE value of 0.07 reported earlier refers to the original optimization-response scale. Under this benchmark setting, the Mean MSE of the Opt RSM–GA ANN was lower than those of Linear Regression (0.198), Random Forest (0.122), Default ANN (0.077), Random Search ANN (0.302), and Opt CCD ANN (0.340). In addition, the Opt RSM–GA ANN produced a low SD MSE of 0.0019 and the highest S/N Ratio of 38.65. These results provide additional response-level evidence that the RSM–GA-based optimization stage identified a favorable and stable ANN configuration within the evaluated benchmark.

The inclusion of Linear Regression and Random Forest provides additional comparative context because both models represent commonly used non-ANN baselines in educational data mining and academic performance prediction. Linear Regression provides a simple statistical benchmark, whereas Random Forest represents a nonlinear ensemble model that can capture more complex relationships among academic variables. Recent

studies have emphasized that interpretable baseline models and ensemble learners should be included in student performance prediction studies because they may provide competitive reference points and help justify the use of more complex ANN-based models [71, 72]. In addition, predictive models in education are often linked to the early identification of academic risk, dropout tendency, or student success; therefore, baseline comparison is useful for examining whether a proposed model provides additional response-level evidence beyond methodological complexity [68, 70]. In this study, Random Forest achieved a lower Mean MSE than Linear Regression; however, within the normalized response-level benchmark, both models showed less favorable normalized response values than the Opt RSM–GA ANN.

The Default ANN and Random Search ANN were included to clarify whether the proposed framework provides improvement over untuned and stochastic hyperparameter-search ANN baselines. Although the Default ANN produced a relatively low Mean MSE, its SD MSE was higher than that of the Opt RSM–GA ANN, indicating lower repeated-performance stability. Meanwhile, Random Search ANN produced a higher Mean MSE than Linear Regression, Random Forest, Default ANN, and Opt RSM–GA ANN. This result suggests that random exploration of the hyperparameter space does not necessarily guarantee a favorable response, particularly when the search space is broad and the dataset size is limited. These findings support previous studies showing that ANN performance depends strongly on architecture, activation function, training algorithm, and tuning strategy [13, 16, 73, 74].

The comparison between Opt CCD ANN and Opt RSM–GA ANN further clarifies the contribution of the proposed optimization framework within the evaluated response-level benchmark. Although both configurations use three hidden layers, twenty neurons, the *tansig* activation function in the hidden layer, the *purelin* function in the output layer, the same learning rate, and the same number of epochs, the training algorithm differs. The Opt CCD ANN uses the *trainbr* algorithm, whereas the Opt RSM–GA ANN uses the *trainlm* algorithm. The lower Mean MSE and higher S/N Ratio of the Opt RSM–GA ANN provide additional response-level evidence that the integration of RSM and GA was able to refine the promising solution region identified by CCD and select a more favorable ANN configuration. Overall, the normalized response-level benchmark supports the methodological usefulness of the proposed CCD–RSM–GA framework in improving repeated-performance stability within the evaluated benchmark setting. However, because this benchmark was not designed as a strict final testing-set comparison for all models, broader validation across different schools, cohorts, and curriculum contexts remains necessary before making stronger generalization, practical deployment, or early-warning claims.

## 4. Conclusion

This study developed and evaluated a systematic CCD–RSM–GA framework for ANN hyperparameter optimization to improve the response quality and stability of academic performance prediction in Automotive Vocational Education (AVE). The quadratic RSM model explained 82.09% of the response variation and identified the learning algorithm as the most influential optimization factor. The subsequent GA optimization refined the promising CCD search region and produced the optimal ANN configuration consisting of three hidden layers, 20 neurons, a *tansig–purelin* transfer-function combination, the *trainlm* learning algorithm, a learning rate of 0.005, and 200 epochs. At the original optimization-response scale, this configuration achieved an MSE of 0.07 and an S/N Ratio of 22.94, indicating improved response quality and repeated-performance stability compared with the best CCD-selected configuration.

The revised response-level benchmark further supported the methodological usefulness of the proposed framework. By including Linear Regression, Random Forest, Default ANN, Random Search ANN, Opt CCD ANN, and Opt RSM–GA ANN, the study provided supplementary response-level evidence that, within the normalized response-level benchmark, the Opt RSM–GA ANN produced the most favorable normalized repeated MSE response among the evaluated approaches, with a Mean MSE of 0.012, SD MSE of 0.0019, and S/N Ratio of 38.65. The MSE of 0.07 and the Mean MSE of 0.012 refer to different evaluation scales and should not be interpreted as directly interchangeable performance measures. These findings indicate that the CCD–RSM–GA workflow provides a structured and reproducible approach to ANN hyperparameter optimization compared with conventional trial-and-error tuning, untuned ANN modeling, and unstructured stochastic hyperparameter search.

Nevertheless, the findings should be interpreted within the scope of the available cohort-based dataset and response-level benchmarking design. Broader validation using larger, multi-school, and multi-cohort datasets remains necessary before the optimized model can be recommended for wider practical implementation. Future studies may incorporate non-academic variables, additional curriculum-related predictors, and alternative optimization strategies to further examine model robustness and generalizability.

## Acknowledgement

The authors would like to express their gratitude and appreciation to the Institute for Research and Community Service through the Basic Research scheme, based on Contract Number B/53820/UN38.III.1/LK.04.00/2025, Universitas Negeri Surabaya, Indonesia, for the funding support provided so that this research could be carried out properly. The authors also extend their highest appreciation to VHS NU 1 Karanggeneng, Lamongan, East Java, Indonesia, for its willingness to serve as the research location and for its administrative and technical support throughout the data collection process.

## Data Availability Statement

The anonymized experimental results supporting the CCD–RSM–GA analysis are provided as supplementary material. Student-level raw academic records are not publicly shared due to privacy and institutional restrictions.

## REFERENCES

1. Y. Wang, *Artificial intelligence in student management systems to enhance academic performance monitoring and intervention*, Sci. Rep., vol. 15, no. 1, p. 35122, Oct. 2025, doi: 10.1038/s41598-025-19159-4.
2. W. Ahmed, M. A. Wani, P. Plawiak, S. Meshoul, A. Mahmoud, and M. Hammad, *Machine learning-based academic performance prediction with explainability for enhanced decision-making in educational institutions*, Sci. Rep., vol. 15, p. 26879, 2025, doi: 10.1038/s41598-025-12353-4.
3. C. F. Rodr'iguez-Hern'andez, M. Musso, E. Kyndt, and E. Cascallar, *Artificial neural networks in academic performance prediction: Systematic implementation and predictor evaluation*, Comput. Educ. Artif. Intell., vol. 2, p. 100018, 2021, doi: 10.1016/j.caeai.2021.100018.
4. H. Chavez, B. Chavez-Arias, S. Contreras-Rosas, J. M. Alvarez-Rodr'iguez, and C. Raymundo, *Artificial neural network model to predict student performance using nonpersonal information*, Front. Educ., vol. 8, p. 1106679, Feb. 2023, doi: 10.3389/feduc.2023.1106679.
5. A. Rivas, A. Gonz'alez-Briones, G. Hern'andez, J. Prieto, and P. Chamoso, *Artificial neural network analysis of the academic performance of students in virtual learning environments*, Neurocomputing, vol. 423, pp. 713–720, Jan. 2021, doi: 10.1016/j.neucom.2020.02.125.
6. M. S. Advani, A. M. Saxe, and H. Sompolinsky, *High-dimensional dynamics of generalization error in neural networks*, Neural Networks, vol. 132, pp. 428–446, Dec. 2020, doi: 10.1016/j.neunet.2020.08.022.
7. A. P. Piotrowski and J. J. Napiorkowski, *A comparison of methods to avoid overfitting in neural networks training in the case of catchment runoff modelling*, J. Hydrol., vol. 476, pp. 97–111, Jan. 2013, doi: 10.1016/j.jhydrol.2012.10.019.
8. M. A. K. Raiaan et al., *A systematic review of hyperparameter optimization techniques in Convolutional Neural Networks*, Decis. Anal. J., vol. 11, p. 100470, Jun. 2024, doi: 10.1016/j.dajour.2024.100470.
9. S. Mezzah and A. Tari, *Practical hyperparameters tuning of convolutional neural networks for EEG emotional features classification*, Intell. Syst. with Appl., vol. 18, p. 200212, May 2023, doi: 10.1016/j.iswa.2023.200212.
10. L. M. Abu Zohair, *Prediction of Student's performance by modelling small dataset size*, Int. J. Educ. Technol. High. Educ., vol. 16, no. 1, p. 27, Dec. 2019, doi: 10.1186/s41239-019-0160-3.
11. M. H. Bin Roslan and C. J. Chen, *Educational Data Mining for Student Performance Prediction: A Systematic Literature Review (2015–2021)*, Int. J. Emerg. Technol. Learn., vol. 17, no. 05, pp. 147–179, 2022, doi: 10.3991/ijet.v17i05.27685.
12. B. Alnasyan, M. Basheri, and M. Alassafi, *The power of Deep Learning techniques for predicting student performance in Virtual Learning Environments: A systematic literature review*, Comput. Educ. Artif. Intell., vol. 6, p. 100231, Jun. 2024, doi: 10.1016/j.caeai.2024.100231.
13. Y. Baashar et al., *Toward Predicting Student's Academic Performance Using Artificial Neural Networks (ANNs)*, Appl. Sci., vol. 12, no. 3, p. 1289, Jan. 2022, doi: 10.3390/app12031289.
14. I. Issah, O. Appiah, P. Appiahene, and F. Inusah, *A systematic review of the literature on machine learning application of determining the attributes influencing academic performance*, Decis. Anal. J., vol. 7, p. 100204, Jun. 2023, doi: 10.1016/j.dajour.2023.100204.

15. P. Pravin, J. Z. M. Tan, K. S. Yap, and Z. Wu, *Hyperparameter optimization strategies for machine learning-based stochastic energy efficient scheduling in cyber-physical production systems*, Digit. Chem. Eng., vol. 4, p. 100047, Sep. 2022, doi: 10.1016/j.dche.2022.100047.
16. N. H. Tiep et al., *A New Hyperparameter Tuning Framework for Regression Tasks in Deep Neural Network: Combined-Sampling Algorithm to Search the Optimized Hyperparameters*, Mathematics, vol. 12, no. 24, p. 3892, Dec. 2024, doi: 10.3390/math12243892.
17. R. Mehdi and M. Nachouki, *Genetic Algorithm-Based Approach for Predicting Student Academic Success*, in 2023 24th International Arab Conference on Information Technology (ACIT), IEEE, Dec. 2023, pp. 1–5, doi: 10.1109/ACIT58888.2023.10453789.
18. P. Zhang, J. Wang, Z. Tian, S. Sun, J. Li, and J. Yang, *A genetic algorithm with jumping gene and heuristic operators for traveling salesman problem*, Appl. Soft Comput., vol. 127, p. 109339, Sep. 2022, doi: 10.1016/j.asoc.2022.109339.
19. M. Kaveh and M. S. Mesgari, *Application of Meta-Heuristic Algorithms for Training Neural Networks and Deep Learning Architectures: A Comprehensive Review*, Neural Process. Lett., vol. 55, no. 4, pp. 4519–4622, Aug. 2023, doi: 10.1007/s11063-022-11055-6.
20. I. Veza, M. Spraggon, I. M. R. Fattah, and M. Idris, *Response surface methodology (RSM) for optimizing engine performance and emissions fueled with biofuel: Review of RSM for sustainability energy transition*, Results Eng., vol. 18, p. 101213, Jun. 2023, doi: 10.1016/j.rineng.2023.101213.
21. L. G. de Oliveira, A. P. de Paiva, P. P. Balestrassi, J. R. Ferreira, S. C. da Costa, and P. H. da Silva Campos, *Response surface methodology for advanced manufacturing technology optimization: theoretical fundamentals, practical guidelines, and survey literature review*, Int. J. Adv. Manuf. Technol., vol. 104, no. 5–8, pp. 1785–1837, Oct. 2019, doi: 10.1007/s00170-019-03809-9.
22. A. Reza, L. Chen, and X. Mao, *Response surface methodology for process optimization in livestock wastewater treatment: A review*, Heliyon, vol. 10, no. 9, p. e30326, May 2024, doi: 10.1016/j.heliyon.2024.e30326.
23. S. Habchi, N. Lahboubi, B. Sallek, and H. El Bari, *Response surface methodology for anaerobic digestion of waste from poultry slaughterhouse: Optimization of load and hydraulic retention time*, Results Eng., vol. 18, p. 101215, Jun. 2023, doi: 10.1016/j.rineng.2023.101215.
24. Y. Zhan and J. Zhu, *Response surface methodology and artificial neural network-genetic algorithm for modeling and optimization of bioenergy production from biochar-improved anaerobic digestion*, Appl. Energy, vol. 355, p. 122336, Feb. 2024, doi: 10.1016/j.apenergy.2023.122336.
25. S. Kumari, J. Chowdhry, P. Sharma, S. Agarwal, and M. C. Garg, *Integrating artificial neural networks and response surface methodology for predictive modeling and mechanistic insights into the detoxification of hazardous MB and CV dyes using Saccharum officinarum L. biomass*, Chemosphere, vol. 344, p. 140262, Dec. 2023, doi: 10.1016/j.chemosphere.2023.140262.
26. M. Gavrilescu, S.-A. Floria, F. Leon, and S. Curteanu, *A Hybrid Competitive Evolutionary Neural Network Optimization Algorithm for a Regression Problem in Chemical Engineering*, Mathematics, vol. 10, no. 19, p. 3581, Sep. 2022, doi: 10.3390/math10193581.
27. S. R. Ariyanto, B. Suprianto, Warju, R. Suhartini, M. Samani, and K. R. Haratama, *Hyperparameter Optimization of ANN for Students' Performance Prediction Using Response Surface Methodology and Genetic Algorithm*, JOIV Int. J. Informatics Vis., vol. 10, no. 1, p. 358, Jan. 2026, doi: 10.62527/joiv.10.1.3973.
28. A. Mahdaddi, S. Meshoul, and M. Belguidoum, *EA-based hyperparameter optimization of hybrid deep learning models for effective drug-target interactions prediction*, Expert Syst. Appl., vol. 185, p. 115525, Dec. 2021, doi: 10.1016/j.eswa.2021.115525.
29. A. E. Ivanescu et al., *The importance of prediction model validation and assessment in obesity and nutrition research*, Int. J. Obes., vol. 40, no. 6, pp. 887–894, Jun. 2016, doi: 10.1038/ijo.2015.214.
30. H. Li, *Machine learning optimization for vocational literacy education evaluation: A big data-powered decision support system*, Alexandria Eng. J., vol. 129, pp. 1258–1271, Oct. 2025, doi: 10.1016/j.aej.2025.08.029.
31. M. F. Musso, C. F. R. Hern'andez, and E. C. Cascallar, *Predicting key educational outcomes in academic trajectories: a machine-learning approach*, High. Educ., vol. 80, no. 5, pp. 875–894, 2020, doi: 10.1007/s10734-020-00520-7.
32. A. Caparrini, *mloptimizer: Genetic algorithm-based hyperparameter optimization in Python*, SoftwareX, vol. 31, p. 102265, 2026, doi: 10.1016/j.softx.2026.102567.
33. C. Chanakyan et al., *Optimization of FSP Process Parameters on AA5052 Employing the S/N Ratio and ANOVA Method*, Adv. Mater. Sci. Eng., vol. 2021, no. 1, Jan. 2021, doi: 10.1155/2021/6450251.
34. N. Q. Mahmood, Y. F. Tahir, M. Hikmat, M. S. Abdulsatar, and P. Baumli, *Experimental Investigation of the Surface Roughness for Aluminum Alloy AA6061 in Milling Operation by Taguchi Method with the ANOVA Technique*, J. Eng., vol. 30, no. 03, pp. 1–14, Mar. 2024, doi: 10.31026/j.eng.2024.03.01.
35. Y. Liu and H. Wang, *Surrogate-assisted hybrid evolutionary algorithm with local estimation of distribution for expensive mixed-variable optimization problems*, Appl. Soft Comput., vol. 133, p. 109957, Jan. 2023, doi: 10.1016/j.asoc.2022.109957.
36. H. Piepho, *An adjusted coefficient of determination ($R^2$) for generalized linear mixed models in one go*, Biometrical J., vol. 65, no. 7, Oct. 2023, doi: 10.1002/bimj.202200290.
37. H. A. Pedrozo, M. A. Zamarripa, A. Uribe-Rodr'iguez, G. Panagakos, M. S. Diaz, and L. T. Biegler, *Surrogate model optimization: A comparison case study with pooling problems of $CO_2$ point sources*, Comput. Chem. Eng., vol. 200, p. 109199, 2025, doi: 10.1016/j.compchemeng.2025.109199.
38. R. BV and A. R. Annigeri, *Optimization of Process Parameters of 3D Printed Polylactic Acid Parts Using Genetic Algorithm*, ASEAN Eng. J., vol. 15, no. 2, pp. 149–155, May 2025, doi: 10.11113/aej.v15.22422.
39. C. Kim and I. Joe, *A Balanced Approach of Rapid Genetic Exploration and Surrogate Exploitation for Hyperparameter Optimization*, IEEE Access, vol. 12, pp. 192184–192194, 2024, doi: 10.1109/ACCESS.2024.3508269.
40. K. Borhani and R. T. K. Wong, *An artificial neural network for exploring the relationship between learning activities and students' performance*, Decis. Anal. J., vol. 9, p. 100332, Dec. 2023, doi: 10.1016/j.dajour.2023.100332.
41. S. Gyparakis, I. Trichakis, T. Daras, and E. Diamadopoulos, *Artificial Neural Networks (ANNs) and Multiple Linear Regression (MLR) Analysis Modelling for Predicting Chemical Dosages of a Water Treatment Plant (WTP) of Drinking Water*, Water, vol. 17, no. 2, p. 227, Jan. 2025, doi: 10.3390/w17020227.
42. X. Wang, X. Sun, Y. Wu, F. Gao, and Y. Yang, *Optimizing reverse osmosis desalination from brackish waters: Predictive approach employing response surface methodology and artificial neural network models*, J. Memb. Sci., vol. 704, p. 122883, Jun. 2024, doi:

10.1016/j.memsci.2024.122883.
43. U. A. Dodo et al., *Comparative study of different training algorithms in backpropagation neural networks for generalized biomass higher heating value prediction*, Green Energy Resour., vol. 2, no. 1, p. 100060, Mar. 2024, doi: 10.1016/j.gerr.2024.100060.
44. L. D. Silvi, A. R. Shanmugam, and K. S. Park, *An artificial neural network model for prediction of wall temperature in vertical tube under subcritical and supercritical flow*, Eng. Appl. Artif. Intell., vol. 162, p. 112694, Dec. 2025, doi: 10.1016/j.engappai.2025.112694.
45. X. Guan and H. Burton, *Bias-variance tradeoff in machine learning: Theoretical formulation and implications to structural engineering applications*, Structures, vol. 46, pp. 17–30, Dec. 2022, doi: 10.1016/j.istruc.2022.10.004.
46. A. Talaei-Khoei and L. Motiwalla, *A new method for improving prediction performance in neural networks with insufficient data*, Decis. Anal. J., vol. 6, p. 100172, Mar. 2023, doi: 10.1016/j.dajour.2023.100172.
47. A. A. Stepashkin and N. Y. Nikitin, *Statistical analysis, regression, and neural network modeling of the tensile strength of thermoplastic unidirectional carbon fiber-polysulfone composites*, Carbon Trends, vol. 15, p. 100368, Jun. 2024, doi: 10.1016/j.cartre.2024.100368.
48. M. Karimipourfard et al., *A Taguchi-optimized Pix2pix generative adversarial network for internal dosimetry in $^{18}$F-FDG PET/CT*, Radiat. Phys. Chem., vol. 218, p. 111532, May 2024, doi: 10.1016/j.radphyschem.2024.111532.
49. I. Sabry, A. M. Hewidy, M. Alkhedher, and A.-H. I. Mourad, *Analysis of variance and grey relational analysis application methods for the selection and optimization problem in 6061-T6 flange friction stir welding process parameters*, Int. J. Light. Mater. Manuf., vol. 7, no. 6, pp. 773–792, Nov. 2024, doi: 10.1016/j.ijlmm.2024.06.006.
50. Y. Nouri, M. A. Ghanbari, and P. Fakharian, *An integrated optimization and ANOVA approach for reinforcing concrete beams with glass fiber polymer*, Decis. Anal. J., vol. 11, p. 100479, Jun. 2024, doi: 10.1016/j.dajour.2024.100479.
51. S. Pereira, P. Canhoto, and R. Salgado, *Development and assessment of artificial neural network models for direct normal solar irradiance forecasting using operational numerical weather prediction data*, Energy AI, vol. 15, p. 100314, Jan. 2024, doi: 10.1016/j.egyai.2023.100314.
52. A. Gupta, T. S. Stead, and L. Ganti, *Determining a Meaningful R-squared Value in Clinical Medicine*, Acad. Med. Surg., Oct. 2024, doi: 10.62186/001c.125154.
53. C. Dong, G. Li, and X. Feng, *Lack-of-Fit Tests for Quantile Regression Models*, J. R. Stat. Soc. Ser. B Stat. Methodol., vol. 81, no. 3, pp. 629–648, Jul. 2019, doi: 10.1111/rssb.12321.
54. S. J. Teran Hidalgo, M. C. Wu, S. M. Engel, and M. R. Kosorok, *Goodness-of-fit test for nonparametric regression models: Smoothing spline ANOVA models as example*, Comput. Stat. Data Anal., vol. 122, pp. 135–155, Jun. 2018, doi: 10.1016/j.csda.2018.01.004.
55. R. Mart'inez-Castro, J. Fl'orez-Santiago, R. Valle-Molinares, J. Cabrera-Barraza, and F. Espitia-Almeida, *Optimized microwave-assisted azadirachtin extraction using response surface methodology*, Heliyon, vol. 10, no. 10, p. e31504, May 2024, doi: 10.1016/j.heliyon.2024.e31504.
56. B. Ramadevi, V. R. Kasi, and K. Bingi, *Fractional ordering of activation functions for neural networks: A case study on Texas wind turbine*, Eng. Appl. Artif. Intell., vol. 127, p. 107308, Jan. 2024, doi: 10.1016/j.engappai.2023.107308.
57. R. Yan et al., *Label-Efficient Self-Supervised Federated Learning for Tackling Data Heterogeneity in Medical Imaging*, IEEE Trans. Med. Imaging, vol. 42, no. 7, pp. 1932–1943, Jul. 2023, doi: 10.1109/TMI.2022.3233574.
58. R. Parhi and R. D. Nowak, *The Role of Neural Network Activation Functions*, IEEE Signal Process. Lett., vol. 27, pp. 1779–1783, 2020, doi: 10.1109/LSP.2020.3027517.
59. M. Mansouri, M. Shayanmehr, and A. Ghaemi, *Exploring the adsorption desulfurization efficiency using RSM and ANN methodologies*, Sci. Rep., vol. 15, no. 1, p. 20869, Jul. 2025, doi: 10.1038/s41598-025-05688-5.
60. S. R. Dubey, S. K. Singh, and B. B. Chaudhuri, *Activation functions in deep learning: A comprehensive survey and benchmark*, Neurocomputing, vol. 503, pp. 92–108, Sep. 2022, doi: 10.1016/j.neucom.2022.06.111.
61. C. F. G. Dos Santos and J. P. Papa, *Avoiding Overfitting: A Survey on Regularization Methods for Convolutional Neural Networks*, ACM Comput. Surv., vol. 54, no. 10s, pp. 1–25, Jan. 2022, doi: 10.1145/3510413.
62. M. Tao, J. Li, and C. Theodoropoulos, *Reduced model-based global optimisation of large-scale steady state nonlinear systems*, in *Proceedings of the 29th European Symposium on Computer Aided Process Engineering*, Eindhoven, Netherlands, Jun. 16–19, 2019, pp. 1039–1044, doi: 10.1016/B978-0-12-818634-3.50174-0.
63. C. Chen, J. Yang, and Y. Zhou, *Neural Network Training Techniques Regularize Optimization Trajectory: An Empirical Study*, in 2020 IEEE International Conference on Big Data (Big Data), IEEE, Dec. 2020, pp. 141–146, doi: 10.1109/BigData50022.2020.9378359.
64. K. Wang, Y. Dou, T. Sun, P. Qiao, and D. Wen, *An automatic learning rate decay strategy for stochastic gradient descent optimization methods in neural networks*, Int. J. Intell. Syst., vol. 37, no. 10, pp. 7334–7355, Oct. 2022, doi: 10.1002/int.22883.
65. Z. Sabir, S. Saoud, M. A. Z. Raja, H. A. Wahab, and A. Arbi, *Heuristic computing technique for numerical solutions of nonlinear fourth order Emden–Fowler equation*, Math. Comput. Simul., vol. 178, pp. 534–548, Dec. 2020, doi: 10.1016/j.matcom.2020.06.021.
66. S. E. Avramenko, T. A. Zheldak, and L. S. Koriashkina, *Guided Hybrid Genetic Algorithm for Solving Global Optimization Problems*, Radio Electron. Comput. Sci. Control, no. 2, pp. 174–188, Jul. 2021, doi: 10.15588/1607-3274-2021-2-18.
67. H. Maaranen, K. Miettinen, and A. Penttinen, *On initial populations of a genetic algorithm for continuous optimization problems*, J. Glob. Optim., vol. 37, no. 3, pp. 405–436, Jan. 2007, doi: 10.1007/s10898-006-9056-6.
68. M. Hlosta, C. Herodotou, T. Papathoma, A. Gillespie, and P. Bergamin, *Predictive learning analytics in online education: A deeper understanding through explaining algorithmic errors*, Comput. Educ. Artif. Intell., vol. 3, p. 100108, 2022, doi: 10.1016/j.caeai.2022.100108.
69. P. Jiao, F. Ouyang, Q. Zhang, and A. H. Alavi, *Artificial intelligence-enabled prediction model of student academic performance in online engineering education*, Artif. Intell. Rev., vol. 55, no. 8, pp. 6321–6344, Dec. 2022, doi: 10.1007/s10462-022-10155-y.
70. I. A. Siddiqui, S. M. Zahari, N. A. M. Ghani, M. S. AlHareky, J. AlHumaid, M. A. AlGhamdi, and A. Rasheed, *Survival time in higher education program after a dropout using modified survival function: A retrospective study to predict average graduation time and factors leading to early dropout*, Stat. Optim. Inf. Comput., vol. 15, no. 3, pp. 1445–1462, Nov. 2025, doi: 10.19139/soic-2310-5070-2499.

71. F. E. Ar'evalo-Cordovilla and M. Pe na, *Comparative analysis of machine learning models for predicting student success in online programming courses: A study based on LMS data and external factors*, Mathematics, vol. 12, no. 20, p. 3272, Oct. 2024, doi: 10.3390/math12203272.
72. T.-T. Huynh-Cam, L.-S. Chen, and H. Le, *Using decision trees and random forest algorithms to predict and determine factors contributing to first-year university students' learning performance*, Algorithms, vol. 14, no. 11, p. 318, Oct. 2021, doi: 10.3390/a14110318.
73. Y. A. Alsariera, Y. Baashar, G. Alkawsi, A. Mustafa, A. A. Alkahtani, and N. Ali, *Assessment and Evaluation of Different Machine Learning Algorithms for Predicting Student Performance*, Comput. Intell. Neurosci., vol. 2022, pp. 1–11, May 2022, doi: 10.1155/2022/4151487.
74. W. Pannakkong, K. Thiwa-Anont, K. Singthong, P. Parthanadee, and J. Buddhakulsomsiri, *Hyperparameter tuning of machine learning algorithms using response surface methodology: A case study of ANN, SVM, and DBN*, Math. Probl. Eng., vol. 2022, p. 8513719, Jan. 2022, doi: 10.1155/2022/8513719.