**The Optimal Design of Sustainable Bio Catalytic Enzyme Applying the Combination of Response surface Method and Artificial Neural Network Modelling for Wastewater Treatment**

Laurentia Liennart1, Jennifer Caroline Wijaya2, Cherish Kumala3, Deasy Darlene Tunas4,

1. SMAK 1 BPK Penabur Bandung, Indonesia (laurentliennart@gmail.com)
2. Bina Bangsa School Semarang, Indonesia (jenniferwijaya272006@gmail.com)
3. Dwight School Seoul, Seoul (cherish.k.kjp@gmail.com)
4. Bina Bangsa School Pantai Indah Kapuk, Indonesia ([deasyy@tuta.io](mailto:deasyy@tuta.io))

***Abstract***

*The high content of food waste poses a danger to our country's garbage management, with up to 115-184 kg per individual per year. Due to the issue of food waste and bad agriculture management, the research of water pollution treatments is in great demand. We can reduce the quantity of food waste produced as well as water pollution by using sustainable eco-enzymes. The present study involving the prediction and optimization by using the combination of Response Surface Methodology (RSM) and Artificial Neural Network (ANN) modelling. The different types of enzymes activities and the days of the treatment will be the inputs to this research, while the elimination of total suspended solids (TSS), volatile suspended solids (VSS), and total ammonia nitrogen (TAN) will be the outputs of this research. In addition, the multiobjective genetic algorithm (MOGA) approach is used for optimization to maximize TSS, VSS, and TAN removal. It is interesting to note that the TSS, VSS, and TAN removal coincide satisfactorily with the values predicted by RSM and ANN with high validity regressing (R = 0.99). The output optimum of the TSS, VSS, and TAN removal by these techniques also results in the perfect agreement which are 32.87%, 51.5%, and 52.99% respectively. These findings lay the groundwork for the investigation of the output performance for water purification using machine learning techniques in combination with multiobjective optimization to estimate the appropriate eco-enzyme characteristics for maximal water pollution removal using the notion of a sustainable low-cost waste recovery method.*

***Keywords :***  *Eco enzyme, Response Surface Methodology, Artificial neural network, Multiobjective Optimization Genetic Algorithm, Water Treatment*

1. **Introduction**

Most of the Indonesia’s pollution is on water. [Not only is it ranked the world’s twentieth most polluted country](https://aqli.epic.uchicago.edu/country-spotlight/indonesia/" \l ":~:text=Indonesia%20is%20today%20the%20world's,WHO)%20guideline%20was%20permanently%20met.). [More than 70% of Indonesia's rivers are classed as "polluted,"](https://core.ac.uk/download/pdf/333894658.pdf) this is due to the vast amount of food waste that ends up in landfills [1]. [About 23-28 million tons of food is wasted per year](http://greengrowth.bappenas.go.id/en/sustainable-food-waste-management-contributes-to-low-carbon-development-in-indonesia/) in Indonesia with each person discarding almost 30 kg of food [2]. As a result, greater waste disposal costs and crucial waste management are required to tackle this issue.

There are several ways to decrease organic waste and one of the most prominent examples is green agriculture [3]. However, the green agriculture has multiple disadvantages [4]. The exorbitant cost, soil quality would most likely decrease for growing crops, negative side effects of health, and so on. This made it even harder for scientists and researchers to find what is the best for water pollution treatment. From all these difficulties, one outstanding and accessible method is using eco enzymes. Eco enzymes not only minimize the quantity of solid waste produced, particularly food waste, but also water pollution [5]. To prevent further jeopardization of our health and the environment, we need to take action by proving that eco enzymes are the most effective solution to reducing water pollution.

[The invention of eco enzymes was developed by Dr. Rosukon Poompanvong in 1984.](https://www.enzymesos.com/what-is-eco-enzyme) [With her research of more than 30 years](http://opscience.iop.org/article/10.1088/1755-1315/510/4/042015/pdf), she was able to come up with the concept of growing enzymes into organic cleansers from organic trash. This eco enzyme strategy that will be used is the fermentation of organic waste [6]. During the fermentation process, organic waste is turned into valuable enzymes. Then, the enzyme generated has cleaning properties for treating water pollution.

[Olgalizia et al. did an experiment on finding the characteristics and production of eco enzymes and the influence on aquaculture sludge](https://biointerfaceresearch.com/wp-content/uploads/2020/10/20695837113.1020510214.pdf) [7]. They were able to reduce suspended solid from the contaminated water through using different parameters such as Biological Oxygen Demand (BOD), Chemical Oxygen Demand (COD), pH, Total Dissolved Solids (TDS), Total Solid (TS), lipase, protease and amylase. The eco enzyme possesses amylase, protease, and lipase activity to be used to treat dairy waste, which contains carbohydrate, protein, and fat that those enzymes can break down. They used pH levels to determine the enzymes’ activity. Then, they used different dilution factors of the eco enzymes by using (Total Suspended Solid) TSS, (Total Ammonia Nitrogen) (TAN) and (Volatile Suspended Solid) VSS reduction. In result, the 15% dilution factor (highest concentration) showed that it is the most effective in removing TSS and VSS[. The other previous study of the eco-enzyme investigation as the water treatment was also conducted by Wen et al](https://ojs.wiserpub.com/index.php/AMTT/article/view/726) [8]. Their research on the efficacy of the eco-enzyme treatments was determined by measuring Ca2+, Na+, K+, NO3, and pH. It was then discovered that the eco-enzyme released by fruits was helpful in reducing NO3 concentration but not the rest. Instead of using parameters like TSS, COD, TDS, they used the anions.

Although with these numerous experiments, scientists and researchers have not yet specified nor investigated the optimum design of the activity of bio-catalytic enzymes against the characteristics of garbage anzymes and their relation to bio-catalytic activities to stabilize the water purification. Most of the research that has been carried out only performing the investigation of the pollutant percentage reduction in the water. Modeling eco enzyme characteristics using artificial intelligence

Modelling eco enzyme characteristic uses artificial intelligence is a relatively recent strategy that provides an in-depth investigation of the factors involved and can compensate for the deficiencies of other conventional methods. The literature states that the water pollutant contents are influenced by several parameters of eco-enzyme activities such as lipase, protease, and amylase [9]. This study intends to use data-driven techniques to predict and to optimize the enzyme activities in water pollution removal. The prediction and optimization process are carried out by combining the Response Surface Method (RSM) and Artificial Neural Network (ANN) with the multi-objective optimization genetic algorithm (MOGA) method. Eco enzyme characteristics such as protease, amylase, lipase, and days of the treatment will be the inputs in this simulation. The reduction of TSS, VSS, and TAN percentage in water will be the output in this modeling. The MOGA was carried out to obtain the best input value for eco-enzyme characteristics to achieve maximum TSS, VSS, and TAN removal. The pareto graph obtained from MOGA will be determined of the most optimum variable. It is hoped that this research will pave the way for researchers and practitioners to be able to estimate the best eco-enzyme properties to obtain maximum water pollutant removal with the concept of a sustainable low-cost waste recovery strategy.

**2. Methodology**

**2.1 Production of eco enzyme**

To produce the eco enzyme, each 100 g of the food and vegetable wastes such as pineapple, orange, tomato, and mango dregs were mixed together with 300 g of brown sugar or molasses, and 3000 g water based on the following ratio of 1 : 3 : 10. The mixtures were fermented separately in air-tight containers covered with aluminum foil and placed in a dark, cool and dry place to prevent sunlight as illustrated in **Figure 1**. During the first month of the fermentation process, the formations of gasses were released daily and once a month during the next two months in order to prevent rupturing due to the built up pressure in the containers. After three months of fermentation, the mixture was filtered to obtain the enzyme solution. The data used in the optimization and prediction modeling process can be found in Rasit et al [10].

Diagram

Description automatically generated

**Figure 1** The Preparation of Eco Enzyme

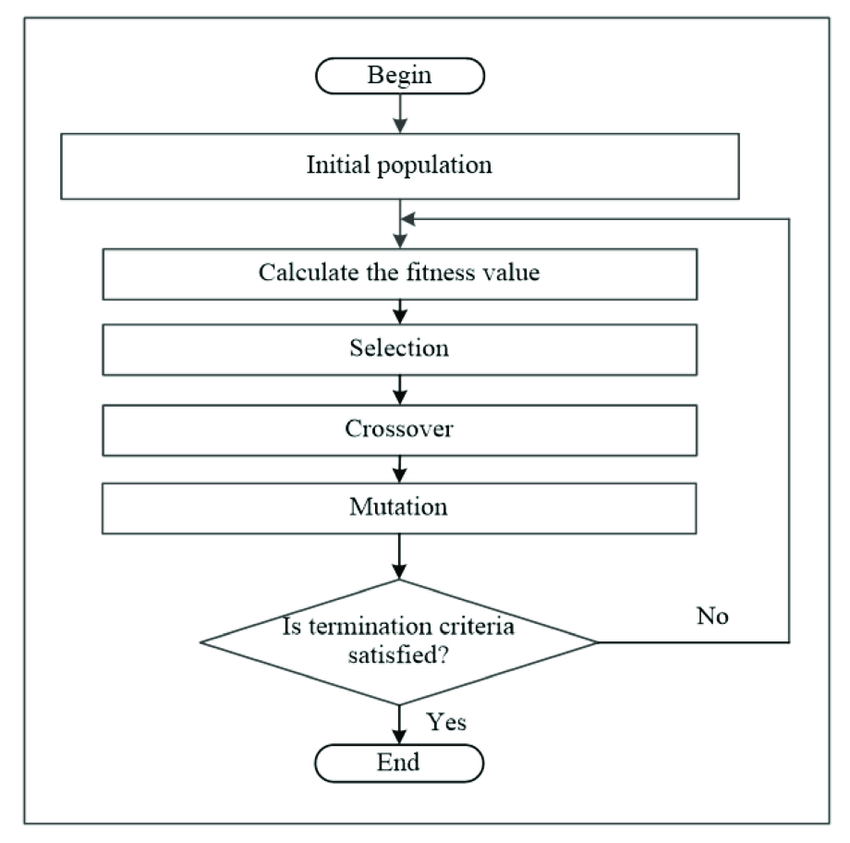
**2.2 Modeling methodology**

**2.2.1 RSM-based predictive model**

Design of Experiment (DoE) was used to figure out the causal relationship amongst the different parameters in the experiment from the dependence and interaction between them, which were shown through the outcomes of the experiments [11]. Face-centered central composite design with six center points and eight-corner points was used in the DoE analysis intending to improve the outcomes. Using the inputs and responses, DoE outcomes were gained and a verifiable connection was developed. In order to obtain the result, DoE outcomes are then statistically examined using Response Surface Method (RSM).

The RSM is a reliable analysis method for examining several inputs and outputs for modeling and optimization [12]. The RSM is a robust analysis technique to analyze multiple inputs and responses for modeling and optimization. The least prominent components are used in the RSM analysis, which yields reliable findings [13]. The input parameters, response, and interaction in the first stage were displayed in a linear equation that was then used in the optimization process of the parameters in the second stage. Using RSM, a verifiable linear equation relationship between the input parameters and their responses was enhanced [14]. The analysis of variance (ANOVA) calibrates the average squared, sum of squared deviations, and degree of freedom of the model for each input value [15]. ANOVA was utilized afterwards for optimization through the application of the input parameters and significance of the model to ANOVA on linear equation.

**2.2.2 ANN-based predictive model**

Due to their capacity for universal approximation and flexible structure, which enables them to capture complicated nonlinear behaviors, ANNs are effective data-driven modeling tools that are frequently employed for the dynamic modeling and identification of nonlinear systems [16]. Such a model is made up of tiny intelligent computing units called neurons that are utilized to represent complicated nonlinear systems in accordance with the groupings of input-output data that are readily available [17]. Finding the correlations between the experiment's variables is tough since there are many conditions that must be met as well as complicated calculations that take a lot of time [18]. The ability of an artificial neural network to produce precise equations from the data given into it makes it useful for identifying these links. Artificial neural networks are useful because they can reduce costs, time, and work.

In this study, a multilayer perceptron (MLP) is utilized, which connects the networks in a feed-forward configuration to create the ideal network. The multilayer perceptron (MLP) is often regarded as the most effective method for solving nonlinear issues [19]. Rather than multiple layers of neurons, the MLP framework embodies optimized neurons and 3 interconnected layers: input layer, hidden layer, and output layer [20]. The hidden layer, which is attached to every input layer, consists of several transfer functions correlated to biological neurons. Lastly, the output layer presents the projection of the advanced model.

Figure 2 MOGA Optimization Process

The number of neurons is very significant to find the optimal network size and is usually chosen based on trial and error or a heuristic approach. Selecting the optimal quantity of neurons is a crucial step because a higher number of neurons might cause overfitting, in reverse, a lower number of neurons might cause the results to underfit. As stated earlier, to avoid overfitting due to MLP, the input data went through several essential processes: training, validation, and testing [21]. Training of neurons helps design a prime ANN model through synapses adjustments, while validation helps oversee the learning curve of neurons. Both processes can only be stopped when the overall Mean Square Error (MSE) in the training stage is higher than the ones during the validation stage. 70 % of the data sets were used to train neurons [22]. The remaining 15% of each dataset were utilized in the validation and testing process. Once the data was authenticated, the developed model was used to trigger experimental data at various input circumstances.

**2.2.3 Genetic Algorithm (GA)**

Genetic algorithm was first proposed by David Goldberg and John Holland and i[ts concept has been very active in lots of different areas](https://sci-hub.se/10.1016/j.cie.2008.09.036), proving that it has been properly and nicely designed [23]. [Genetic Algorithms are a group of similar algorithms that draw upon ideas of Darwinian evolution and genetics](https://sci-hub.se/10.1016/S0376-7361(03)80007-9) [24]. As time goes by, it continues to provide good results using this genetic algorithm.   
In GA, it is different from other types of algorithms out there. It has a population of solutions, rather than a one, single 'current' solution. Several stages in optimization process by multi-objective genetic algorithm is represented in  **Figure 2** [18]**.**

[To carry out GA, it is necessary to make a number of decisions about how to represent solutions, how to manipulate information and how the population is maintained](https://sci-hub.se/10.1016/S0376-7361(03)80007-9): [GA randomly selects the populations and applies the fitness function. Those individuals with adequate match values will be selected to enter the crossover stage, while individuals with low match values will be terminated. In addition, individuals who have been selected to the crossover stage will be mated to produce new individuals who have a combination of information from their parents. The individuals will then experience random mutations and reappear as a new population, including their parents. Up until the best result is obtained, this cycle will repeat itself.](https://users/jenniferwijaya27/Downloads/software%20for%20WSEEC/REFERENCES/WSEEC4.pdf) The optimization process is carried out in MATLAB's optimization toolbox, along with the neural function established during the setup of the artificial neural networks as the fitness function. The constraints of each variable are determined. The optimization parameters are then set until the optimization function is ready to run.

**2.2.4 Multi-objective Optimization**

This research uses ANN modelling to understand the multi-objective optimization. We had sets of data prepared before conducting this research.

**Table 1** Results of DoE based on Face Centered CCD

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Day** | **TSS Actual [%]** | **TSS Predicted [%]** | **VSS Actual [%]** | **VSS Predicted [%]** | **TAN Actual [%]** | **TAN Predicted [%]** |
| 1 | 10.02 | 13.34 | 38.03 | 40.98 | 19.52 | 17.56 |
| 2 | 18.38 | 22.28 | 51.23 | 49.5 | 19.42 | 23.39 |
| 3 | 27.22 | 31.21 | 64.32 | 58.03 | 27.81 | 29.23 |
| 4 | 46 | 40.15 | 64.57 | 66.56 | 35.03 | 35.06 |
| 5 | 54.45 | 49.09 | 72.01 | 75.09 | 44.35 | 40.89 |
| 1 | 7.95 | 4.92 | 5.56 | 2.4 | 22.15 | 23.08 |
| 2 | 14.43 | 13.86 | 17.66 | 10.93 | 27.49 | 28.91 |
| 3 | 19.47 | 22.79 | 19.4 | 19.46 | 35.24 | 34.74 |
| 4 | 29.05 | 31.73 | 22.93 | 27.99 | 39.65 | 40.57 |
| 5 | 43.06 | 40.66 | 31.74 | 36.51 | 49.18 | 46.41 |
| 1 | 8.31 | 8.16 | 21.54 | 23.31 | 41.85 | 42.08 |
| 2 | 18.98 | 17.1 | 33.37 | 31.83 | 46.45 | 47.92 |
| 3 | 26.73 | 26.03 | 41.54 | 40.36 | 55.15 | 53.75 |
| 4 | 34.2 | 34.97 | 47.83 | 48.89 | 60.07 | 59.58 |
| 5 | 41.95 | 43.91 | 57.52 | 57.41 | 65.22 | 65.41 |
| 1 | 4.87 | 0.574 | 3.55 | 9.42 | 25.5 | 19.27 |
| 2 | 10.4 | 9.51 | 11.1 | 17.94 | 22.26 | 25.1 |
| 3 | 16.91 | 18.45 | 31.98 | 26.47 | 32.32 | 30.94 |
| 4 | 26.34 | 27.38 | 37.52 | 35 | 34.5 | 36.77 |
| 5 | 33.71 | 36.32 | 48.21 | 43.53 | 40.1 | 42.6 |

Using MATLAB ANN, we enter the data and this software will release the prediction results that will be optimized by using genetic algorithm (GA). Then, we will see whether an optimal result is obtained. However, if the result of the optimization is not optimum, the decision variable's beginning value is once again set, and the optimization process is repeated until the desired outcome is reached. The following equations can be used to design a multiobjective optimization:

|  |  |  |
| --- | --- | --- |
| Find | | (1) |
| Minimizing or maximizing | | (2) |
|  | (3) | |
|  | (4) | |

Npar is the number of decision variables, fi(x) is the objective function, and Nobj is the total number of objective functions, where x is the vector of choice variables. gj(x) and hk(x), respectively, stand for the equality and inequality conditions. The quantity of equality and inequality constraints, respectively, is represented by m and n.

**3. Results and Discussion**

**3.1 Model development and verification**

The variations of three responses, TSS, VSS and TAN, were observed by changing the volume of protease, amylase and lipase. **Table 1** shows the results of percentage by using a charged-coupled device (CCD) for inputted parameters. The ANOVA results are seen from **Table 2**-**Table 4**. The findings demonstrate how the input as well as their squares and interactions, helps in analyzing the output. It also demonstrates that these characteristics are critical for evaluating the generated model. **Table 2** shows that for TSS removal with a P-value of <0.05, the corresponding F value for the constructed model is 77.06. Model terms with P-values less than 0.05 are significant, therefore these are appropriate model terms in this scenario. The results from **Table 2** demonstrate the value of the developed model in optimizing and estimating TSS removal. Eq (5) is the model's representation. The equation covers the various parameters that affect TSS removal. A represents the number of days, B is protease, C is amylase and D is lipase.

TSS = 8.94A + 55.76B - 24.88C - 43.07D + 13.38 (5)

**Table 2** ANOVA Results for TSS Removal

| **Source** | **Sum of Squares** | **df** | **Mean Square** | **F-value** | **p-value** |
| --- | --- | --- | --- | --- | --- |
| **Model** | 3628.78 | 4 | 907.2 | 77.06 | < 0.0001 |
| A-Day | 3194.08 | 1 | 3194.08 | 271.3 | < 0.0001 |
| B-Protease | 73.34 | 1 | 73.34 | 6.23 | 0.0247 |
| C-Amylase | 187.04 | 1 | 187.04 | 15.89 | 0.0012 |
| D-Lipase | 4.41 | 1 | 4.41 | 0.3743 | 0.5498 |
| **Residual** | 176.6 | 15 | 11.77 |  |  |
| **Cor Total** | 3805.38 | 19 |  |  |  |

**Table 3** ANOVA Results of VSS Removal

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **Sum of Squares** | **df** | **Mean Square** | **F-value** | **p-value** |
| **Model** | 7252.63 | 4 | 1813.16 | 86.04 | < 0.0001 |
| A-Day | 2908.56 | 1 | 2908.56 | 138.01 | < 0.0001 |
| B-Protease | 1378.59 | 1 | 1378.59 | 65.42 | < 0.0001 |
| C-Amylase | 294.74 | 1 | 294.74 | 13.99 | 0.002 |
| D-Lipase | 339.39 | 1 | 339.39 | 16.1 | 0.0011 |
| **Residual** | 316.11 | 15 | 21.07 |  |  |
| **Cor Total** | 7568.74 | 19 |  |  |  |

**Table 4** ANOVA Results of TAN Removal

| **Source** | **Sum of Squares** | **df** | **Mean Square** | **F-value** | **p-value** |
| --- | --- | --- | --- | --- | --- |
| **Model** | 3274.08 | 4 | 818.52 | 111.71 | < 0.0001 |
| A-Day | 1360.61 | 1 | 1360.61 | 185.69 | < 0.0001 |
| B-Protease | 1314.15 | 1 | 1314.15 | 179.35 | < 0.0001 |
| C-Amylase | 535.95 | 1 | 535.95 | 73.14 | < 0.0001 |
| D-Lipase | 1376.83 | 1 | 1376.83 | 187.9 | < 0.0001 |
| **Residual** | 109.91 | 15 | 7.33 |  |  |
| **Cor Total** | 3383.99 | 19 |  |  |  |

**Table 3** shows the ANOVA result to estimate the VSS removal with a F value of 86.04 and P-value of <0.05. The Model F-value of 86.04 implies the model is significant. There is only a 0.01% chance that an F-value this large could occur due to noise. Eq (6) is the representation of the developed model for VSS removal estimation.

VSS = 8.53A + 241.76B - 31.23C - 377.95D + 139.01 (6)

**Table 4** shows the ANOVA result to estimate the TAN removal with an F value of 111.71 and P-value of <0.05. All the p-values here are 0.0001 which is closer to 0 making the model more significant. Eq (7) is a representation of a developed model for TAN removal estimation.

TAN = 5.83A -236.04B - 42.12C - 761.24D + 322.11 (7)

Using normal plots for residuals and outliers, the proposed model was tested for adequacy and abnormality of data. A good model should not follow any trend or sequence, and the points should be near to a straight line. **Figure 3** depicts the residual and outlier plots for the TSS removal. According to **Figure 3a**, all data points are spread randomly and do not follow any sequence, confirming the predictability of the model. **Figure 3b** shows the outlier charts for TSS removal. Data points that exceed the allowed range of ±3.67 are considered abnormal. According to **Figure 3b**  all data points fall inside this permitted range.

The residual and outlier plots for the VSS removal model are shown in **Figure 4**, with **Figure 4a** representing the residual plot and **Figure 4b** representing the outlier plot. The data points in **Figure 4a-Figure 4b** show no pattern or sequence and fall within an acceptable range of ±3.67. Lastly, **Figure 5** is the residual and outlier plots for TAN removal. With **Figure 5a** being the residual plot and **Figure 5b** being the outlier plot. It has a range from -3.67 to +3.67. Anything outside this range would also be considered abnormal.

It is determined that the produced model can be utilized to navigate the design space after analysis and verification. By altering the input parameters, a 3D surface and a contour plot may best explain the response which can be seen in **Figure *6***. TSS, VSS and TAN are respectively plotted. TAN has the most percentage (%), meaning that it is the most effective in the 19.42-65.22 range. Followed by TSS removal with a range of 4.87-54.45 and lastly, VSS removal with a range of 3.55-72.01. Based on **Figure 7**, there are seven different parametric studies in which one of them is the optimum days for the treatment for the variable A, B-D are for protease, amylase, and lipase with the three objective functions, TSS, VSS, and TAN removal. The main goal of this research is to maximize the reduction of TSS, VSS and TAN removal with the minimum days treatment to save the operation cost for water treatment. The most suitable day for the enzymes to work effectively and reach its optimum is on day 3.5. On this day, amylase would work best at a volume of 0.027ml, protease would work best at 0.073ml and lipase would be the most effective at 0.356ml. This then would give an optimum result for TSS, VSS and TAN removal which is 32.87 % (TSS), 51.5% (VSS), and 52.99% (TAN) with a desirability of 0.573.

**3.2 Artificial Neural Network (ANN) Modelling**

In this study, the ANN modeling used consisted of 4 inputs, 1 hidden later with 9 hidden neurons, and 3 target outputs. A trial-and-error method was used to find the optimal design, and the architecture with the lowest error (RMSE) and best regression coefficient was chosen. **Figure 8** depicts the architecture of the artificial neural network that was designed. The discrepancy between the neural

Chart, scatter chart

Description automatically generated

**Figure 3** The Normal Plot of Residual (a) and Outlier (b) for TSS Removal

Chart, scatter chart

Description automatically generated

**Figure 4** The Normal Plot of Residual (a) and Outlier (b) for VSS Removal

network's output and the desired data is known as the network error level [25]. By computing the weights to achieve at the best weight to employ in the testing phase, the backpropagation algorithm enables neurons to learn new information. To reduce mistakes that may arise during learning, the weights are changed repeatedly. This reverse computing technique can significantly lower the value of the mistake.

Root mean square error (RMSE), mean absolute error (MAE), and mean bias error (MBE) were used to determine the errors for each neuron and each data category [26].The closer the regression of coefficient (R) to 1 is, the more accurate the trendline is. From **Figure 9**, we can say that the blue line (training state) has the most accurate trendline as the R-value is 0.99. It also shows that it is touching every dot on the graph. The regression coefficient (R) values for the trendlines in the data are R = 0.99 for the training data, R = 0.99 for the validation data, and R = 0.99 for the testing data and R = 0.99 for all data combined, as illustrated in **Figure 9a.** The ANN regression **Figure 9a**,shows several correlation coefficients and comparisons between regression model and ANN during training, validation, and testing. **Figure 9b-Figure 9d** illustrate the error histogram, training performance, and training state, respectively. The error histogram is a histogram of the differences between the target and predicted values after the feedforward neural network has been trained. These error numbers show how the expected and goal values differ. The performance graph depicts the network's best validation performance. The variation in the gradient

Chart, scatter chart

Description automatically generated

**Figure 5** The Normal Plot of Residual (a) and Outlier (b) for VSS Removal

Chart, surface chart

Description automatically generated

**Figure 6** 3D Surface Contour Plot by RSM Correlated for (a) TSS, (b) VSS, and (c) TAN Removal

Diagram

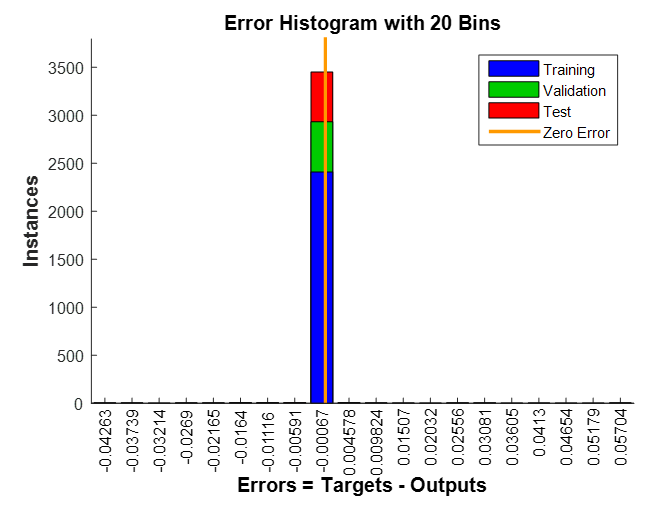
Description automatically generated

**Figure 7** Ramp Function of RSM Optimization

Diagram, schematic

Description automatically generated

**Figure 8** The Architecture of ANN Modelling

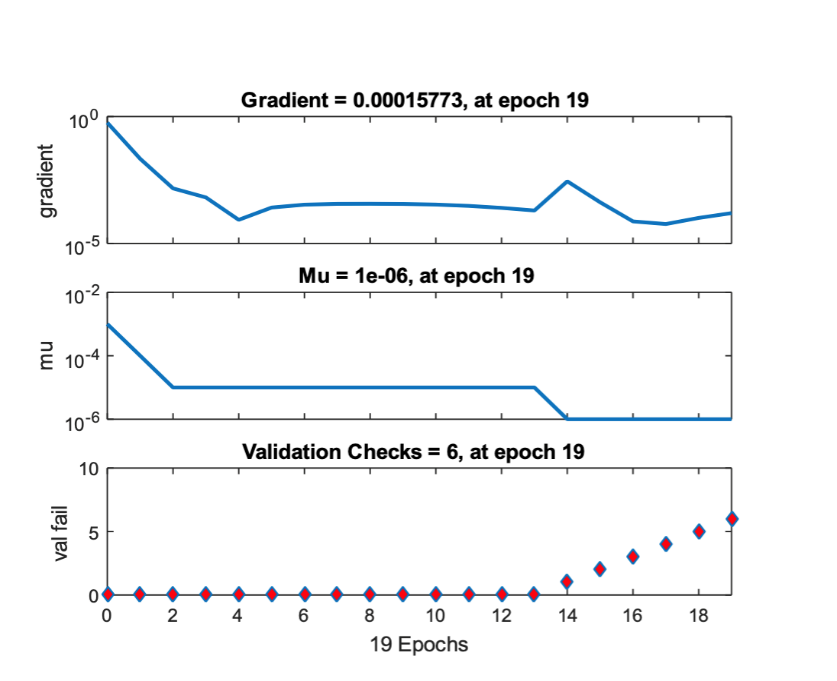
**

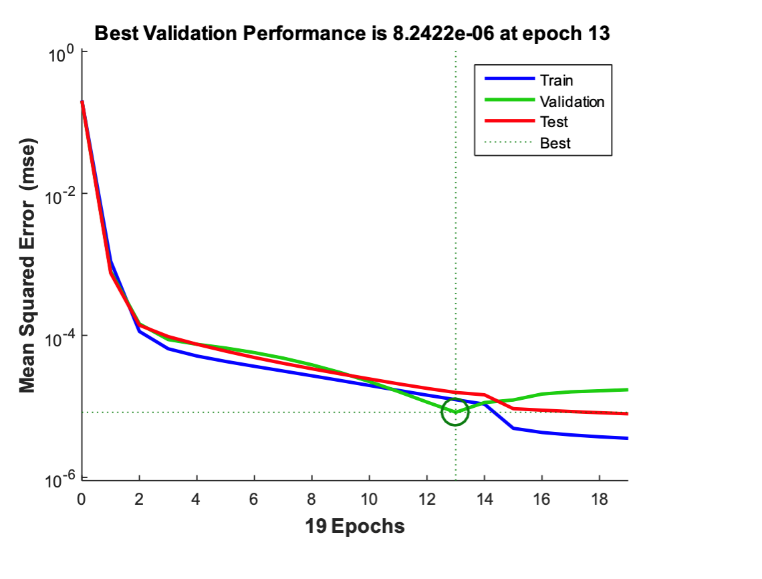
(c)

(a)

(b)

(a)





(d)

**Figure 9 .**  (a) ANN regression data, (b) error histogram of network, (c) the validation performance, (d) the training state.

.

coefficient with respect to the number of epochs is the training state.

**3.3 MOGA Optimization**

**3.3.1 DoE MOGA Optimization**

MOGA analysis was performed using MATLAB software obtaining Pareto graphs for the three objective functions of TSS, VSS, and TAN removal percentage. The optimization process is carried out using the equation that has been generated from DoE, namely Eq, (8)-(10). These three objective values are maximized to get excellent water purification. The input range values of days of treatment, lipase, amylase, and protease were determined based on the previous experimental data.

The results of the optimization with the DoE equation can be seen in **Figure 10**. The Pareto graph obtained here shows that the maximum TSS, VSS and TAN removal values reached to 55.63%, 99.37%, 77.64% with the minimum optimal values for water purification were 31.64%, 34.06%, and 24.08%, respectively. The optimum value generated by MOGA technique is still in the same range as the optimization with DoE as described in **Table 5**.

Chart, scatter chart

Description automatically generated

**Figure 10** DoE MOGA Optimization

**Table 5** Optimal Results of DoE MOGA Optimization

| **TSS %** | **VSS %** | **TAN %** | **Protease [unit/ml]** | **Amylase [unit/ml]** | **Lipase [unit/ml]** |
| --- | --- | --- | --- | --- | --- |
| 55.6 | 44.0 | 57.4 | 0.07 | 0.03 | 0.37 |
| 45.4 | 52.3 | 77.6 | 0.02 | 0.03 | 0.36 |
| 46.1 | 59.2 | 20.4 | 0.13 | 0.16 | 0.39 |
| 49.1 | 34.2 | 69.6 | 0.03 | 0.03 | 0.37 |
| 39.8 | 90.6 | 31.0 | 0.13 | 0.18 | 0.37 |
| 53.1 | 47.8 | 61.3 | 0.06 | 0.03 | 0.36 |
| 43.8 | 74.8 | 24.7 | 0.13 | 0.17 | 0.38 |
| 34.4 | 99.3 | 33.5 | 0.13 | 0.21 | 0.37 |
| 46.9 | 52.6 | 74.9 | 0.03 | 0.03 | 0.36 |
| 49.3 | 34.0 | 65.1 | 0.04 | 0.04 | 0.37 |
| 31.6 | 84.6 | 50.0 | 0.11 | 0.16 | 0.36 |
| 49.5 | 57.3 | 47.8 | 0.11 | 0.05 | 0.37 |
| 42.8 | 79.46 | 25.46 | 0.13 | 0.18 | 0.38 |

**3.3.2 ANN Moga Optimization**

The developed network function is found to be accurate in predicting the percentage of TSS, VSS, and  TAN removal. The function created by the ANN procedure is then employed as a fitness function in the optimization phase. Following the selection of the population and the optimization operations, a Pareto graph is formed, as illustrated in **Figure 11**. The Pareto frontier is achieved by solving the optimization model with the LM method.

Chart, scatter chart

Description automatically generated

**Figure 11** ANN MOGA Optimization

From the optimization results with the prediction equation by ANN, the optimal values for TSS, VSS and TAN removal were maximum 55.35%, 99.61%, and 75.96% with the minimum optimal values of 28.75%, 40.09%, and 626.35 reaching. The results of the optimization with ANN have the same range as the optimization with the equations obtained from DoE.

**4. Conclusion**

Bio catalytic enzyme activities and the days of the treatment of the eco-enzyme for water purification have been effectively predicted and optimized in this work using data-driven methodologies. To determine the perfect optimization of these 3 enzymes (amylase, lipase and protease) is costly and time-consuming in experimental works. It is challenging for even researchers to find the exact value of a great ratio of enzymes to make the most effective eco-enzymes to treat water pollution. Therefore, in order to improve and to optimize the characteristics of eco-enzyme, future research should concentrate on machine learning techniques and multi-objective optimization using the combination of response surface methodology and artificial neural network modelling. Through this techniques, we were able to find out the most optimum volume of enzymes and get maximal results for parameters such as TSS, VSS, and TAN removal in water. From all the data we gathered from the DoE optimization, DoE MOGA Optimization, and ANN Moga Optimization where the value has the same ranges, the optimal values of the removal of TSS, VSS and TAN are 32.87 %, 51.5%, and 52.99% respectively. To obtain this, the values for the input parameters such as days, protease, amylase, and lipase should be in the range 3.5, 0.07, 0.027, and 0.36, respectively. The combined optimization study of these two methods resulted in the perfect combination of the optimal values. These results show the appropriate estimates of eco-enzyme characteristics for the maximal water pollution removal using the notion of a sustainable low-cost waste recovery method.

**5. Acknowledgements**

The authors acknowledge the supports from ALL-in Eduspace in the completion of this research article

**6. References**

1. Gómez-Sanabria, A., et al., *Sustainable wastewater management in Indonesia's fish processing industry: Bringing governance into scenario analysis.* Journal of Environmental Management, 2020. **275**: p. 111241.

2. Seberini, A. *Economic, social and environmental world impacts of food waste on society and zero waste as a global approach to their elimination*. in *SHS Web of Conferences*. 2020. EDP Sciences.

3. Dordas, C., *Role of nutrients in controlling plant diseases in sustainable agriculture. A review.* Agronomy for Sustainable Development, 2008. **28**(1): p. 33-46.

4. Padma, K., K. Don, and P. Josthna. *MODERN FARMING UNDER GREEN REVOLUTION-ITS ADVANTAGES AND DISADVANTAGES*. in *ABOUT CONGRESS*.

5. Janarthanan, M., K. Mani, and S.R.S. Raja, *Purification of Contaminated Water Using Eco Enzyme.* IOP Conference Series: Materials Science and Engineering, 2020. **955**(1): p. 012098.

6. Arun, C. and P. Sivashanmugam, *Investigation of biocatalytic potential of garbage enzyme and its influence on stabilization of industrial waste activated sludge.* Process Safety and Environmental Protection, 2015. **94**: p. 471-478.

7. Galintin, O., N. Rasit, and S.J.B.R.i.A.C. Hamzah, *Production and characterization of eco enzyme produced from fruit and vegetable wastes and its influence on the aquaculture sludge.* 2021. **11**(3): p. 10205-10214.

8. Wen, L.C., et al., *Effective Microorganisms in Producing Eco-Enzyme from Food Waste for Wastewater Treatment.* 2021: p. 28-36.

9. ВСЕУКРАЇНСЬКОЇ, М.X. and С.А.Т. ВИКЛАДАЧІВ, *To Make the World Smarter and Safer.*

10. Rasit, N. and F.S.J.I.J.S.T. Mohammad, *Production and characterization of bio catalytic enzyme produced from fermentation of fruit and vegetable wastes and its influence on aquaculture sludge.* 2018. **4**: p. 12-26.

11. Jankovic, A., G. Chaudhary, and F. Goia, *Designing the design of experiments (DOE) – An investigation on the influence of different factorial designs on the characterization of complex systems.* Energy and Buildings, 2021. **250**: p. 111298.

12. Khuri, A.I. and S.J.W.I.R.C.S. Mukhopadhyay, *Response surface methodology.* 2010. **2**(2): p. 128-149.

13. Danish, M., et al., *Thermal analysis during turning of AZ31 magnesium alloy under dry and cryogenic conditions.* The International Journal of Advanced Manufacturing Technology, 2017. **91**(5): p. 2855-2868.

14. Gupta, A.K.J.I.J.o.P.R., *Predictive modelling of turning operations using response surface methodology, artificial neural networks and support vector regression.* 2010. **48**(3): p. 763-778.

15. Neamat, S., M.J.J.o.A.S. Hassan, and T. Trends, *A Review on Using ANOVA and RSM Modelling in The Glass Powder Replacement of The Concrete Ingredients.* 2021. **2**(02): p. 72-77.

16. Nagy, Z.K.J.C.E.J., *Model based control of a yeast fermentation bioreactor using optimally designed artificial neural networks.* 2007. **127**: p. 95-109.

17. Mago, V.K., *Cross-Disciplinary Applications of Artificial Intelligence and Pattern Recognition: Advancing Technologies: Advancing Technologies*. 2011: IGI Global.

18. Yulia, F., et al., *Multi-objective genetic algorithm optimization with an artificial neural network for CO2/CH4 adsorption prediction in metal–organic framework.* 2021. **25**: p. 100967.

19. Garg, S., et al., *Experimental data, thermodynamic and neural network modeling of CO2 solubility in aqueous sodium salt of l-phenylalanine.* 2017. **19**: p. 146-156.

20. Ramchoun, H., et al., *New modeling of multilayer perceptron architecture optimization with regularization: an application to pattern classification.* 2017. **44**(3): p. 261-269.

21. Benyekhlef, A., et al., *Application of artificial neural network (ANN-MLP) for the prediction of fouling resistance in heat exchanger to MgO-water and CuO-water nanofluids.* 2021. **84**(3): p. 538-551.

22. Haykin, S., *Neural networks and learning machines, 3/E*. 2009: Pearson Education India.

23. Wang, D., et al., *An immune-genetic algorithm for introduction planning of new products.* 2009. **56**(3): p. 902-917.

24. Mitchell, M., *An introduction to genetic algorithms*. 1998: MIT press.

25. Zupan, J.J.A.C.S., *Introduction to artificial neural network (ANN) methods: what they are and how to use them.* 1994. **41**: p. 327-327.

26. Kato, T., *Chapter 4 - Prediction of photovoltaic* power *generation output and network operation*, in *Integration of Distributed Energy Resources in Power Systems*, T. Funabashi, Editor. 2016, Academic Press. p. 77-108.