# Wave Farm Energy Prediction using Multi-Layer Perceptron Neural Networks

1st Mohammad Jameel Jibreel Mamogkat
*BS Computer Engineering*
*Mapúa Malayan Colleges Mindanao*
Davao City, Philippines
mjjmamogkat@mcm.edu.ph

2nd Neil Genesis Cunanan
*BS Computer Engineering*
*Mapúa Malayan Colleges Mindanao*
Davao City, Philippines
email address or ORCID

***Abstract*—This study explores the use of a multi-layer perceptron (MLP) neural network to predict the total power output of large-scale wave energy converter (WEC) farms based solely on buoy layout configuration. Using simulation datasets from Perth and Sydney, we trained and evaluated a model based on 49-buoy layouts containing only (X, Y) coordinates as input features. The model achieved a mean absolute error (MAE) of 25,062 W and percent error of 0.63% on the Perth dataset, demonstrating high accuracy. When tested on unseen Sydney layouts, the error increased to 166,534 W (4.14%), showing moderate generalization capability. A sample Philippine layout was also tested, yielding a predicted output of 5,350,615.65 W. Comparative experiments with a 100-buoy dataset showed higher errors (up to 10.18%) due to limited training data. These results support the viability of layout-only models for early-stage design evaluation of WEC farms and align with the goals of UN SDG 7 by aiding renewable energy planning in developing regions.**

***Index Terms*—Wave Energy, Neural Networks, MLP, Renewable Energy, SDG 7, Machine Learning, Buoy Layout Prediction**

## I. Introduction

Wave energy is one of the most promising sources of renewable power, offering consistent generation potential in coastal regions. As global efforts toward clean energy intensify, wave energy converter (WEC) farms have emerged as a viable solution for countries with extensive shorelines. Designing efficient WEC layouts, however, remains a computationally intensive task due to the complex interactions between buoy positions, wave interference, and energy capture.

Traditional approaches to layout optimization often rely on simulations that incorporate detailed wave conditions such as height, direction, and frequency. While accurate, these simulations are time-consuming and require substantial computational resources. To address this limitation, this study proposes a machine learning approach that predicts the total power output of WEC farms using only the spatial layout of buoys as input.

By leveraging a multi-layer perceptron (MLP) neural network, we aim to model the relationship between buoy positioning and energy output without the need for wave-specific environmental data. This approach enables rapid evaluation of various layouts during early design stages, potentially accelerating deployment in developing regions.

The model was trained using the publicly available Large-Scale Wave Energy Farm dataset, focusing on 49-buoy layouts simulated under Perth and Sydney wave regimes. Results show that the model achieved high accuracy on the training region and demonstrated moderate generalization to unseen layouts. A simulated Philippine layout was also tested to assess the model's applicability in a general context. However, it is important to note that the model performs predictions as if the site experiences Perth-like wave conditions.

This work aligns with the United Nations Sustainable Development Goal 7 (Affordable and Clean Energy) by contributing to the development of accessible tools for renewable energy planning and assessment.

## II. Related Work

Wave energy converter (WEC) farms have long posed challenges in layout optimization due to the complex hydrodynamic interactions between devices. Traditional methods often involve high-fidelity simulations that account for wave interference, energy capture, and device spacing, but these are computationally expensive and time-consuming.

Neshat et al. [1] introduced a large-scale simulated dataset of WEC layouts, generated using a frequency-domain model under realistic sea states based on data from Perth and Sydney, Australia. The dataset includes thousands of 49- and 100-buoy layouts with corresponding absorbed power values, making it suitable for benchmarking optimization and machine learning approaches.

Their work focused on applying multi-strategy evolutionary algorithms to optimize layout configurations for energy maximization. In contrast, our study applies a supervised learning approach using a multi-layer perceptron (MLP) neural network to directly predict energy output from layout coordinates, without incorporating wave-specific parameters. This allows for rapid evaluation of potential layouts, particularly during the early planning stages in data-scarce regions.

## III. Methodology

### *A. Dataset and Features*

This study uses a dataset introduced by Neshat et al. [1], which simulates wave energy converter (WEC) layouts under real-world sea conditions in Perth, Australia. The primary dataset used is the 49-buoy layout file (`WEC_Perth_49.csv`), consisting of 36,043 unique configurations. Each sample represents a buoy layout as 98 numerical features—specifically the $X$ and $Y$ coordinates for 49 point absorber buoys. The target variable is `Total_Power`, representing the absorbed energy in watts.

To build the feature matrix $\mathbf{X}$, all columns starting with `X` or `Y` were retained. The label vector $\mathbf{y}$ was set to the `Total_Power` column. The auxiliary column `qW` was discarded as it was not relevant for this prediction task.

### B. Data Preprocessing

*1) Train-Test Split:* The dataset was randomly divided into training and testing sets using an 80:20 ratio with a fixed seed for reproducibility. This split ensures that model performance is evaluated on unseen layouts.

*2) Feature Scaling:* To improve model convergence and performance, all features were normalized using `StandardScaler` from scikit-learn. The scaler was fitted only on the training set and applied to both training and test sets to prevent data leakage.

### C. Model Architecture and Training

*1) MLP Configuration:* We used a supervised learning approach by training a feedforward neural network via scikit-learn's `MLPRegressor`. The final architecture had two hidden layers with 128 and 64 neurons, respectively. The ReLU activation function yielded the best performance among tested variants. The optimizer used was Adam, with a learning rate of 0.0005 and a training cap of 1000 iterations.

Earlier versions of the model used fewer epochs (300 iterations) and a higher learning rate (0.001), which led to underfitting and reduced performance. Tuning these hyperparameters significantly improved prediction accuracy.

We also compared multiple activation functions: ReLU, tanh, and logistic. As shown in Table II, ReLU achieved a mean absolute error (MAE) of 25,062.32 W and percent error of 0.64%. In contrast, both tanh and logistic activations resulted in errors exceeding 99%, with MAE values around 3.9 million. These results confirmed that ReLU was the most suitable activation function for this task.

*2) Model Evaluation Metrics:* Model performance was evaluated using three key metrics:

- **Mean Absolute Error (MAE)**
- **Mean Squared Error (MSE)**
- **Mean Percent Error (MPE)** relative to average output

The MAE is defined as:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i|$$

The MPE is defined as:

$$\mathrm{MPE} = \frac{100}{n}\sum_{i=1}^{n} \left|\frac{y_i - \hat{y}_i}{\bar{y}}\right|$$

The MSE is defined as:

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

where $y_i$ is the actual total power and $\hat{y}_i$ is the model's prediction.

*3) Visualization:* We plotted the predicted vs actual power output to visually assess the regression fit. A dashed identity line $y = x$ was included to help visualize deviation from ideal predictions.

### D. Generalization Tests and Additional Experiments

*1) Sydney Dataset:* To evaluate generalization, the model was tested on a separate set of 49-buoy layouts simulated under Sydney wave conditions. The same Perth-trained scaler was used for consistency. Performance metrics were computed in the same manner.

*2) Activation Function Comparison:* To evaluate sensitivity to activation choice, the model was re-trained using different activation functions: ReLU, tanh, and logistic. Each variant used the same architecture, learning rate, and training iterations. MAE, MSE, and percent error were logged for comparison.

*3) Philippine Layout Test:* A custom layout based on a hypothetical Philippine wave farm configuration was also tested. The layout followed a 7×7 grid spacing and was processed using the existing Perth-trained model and scaler to estimate its potential energy output.

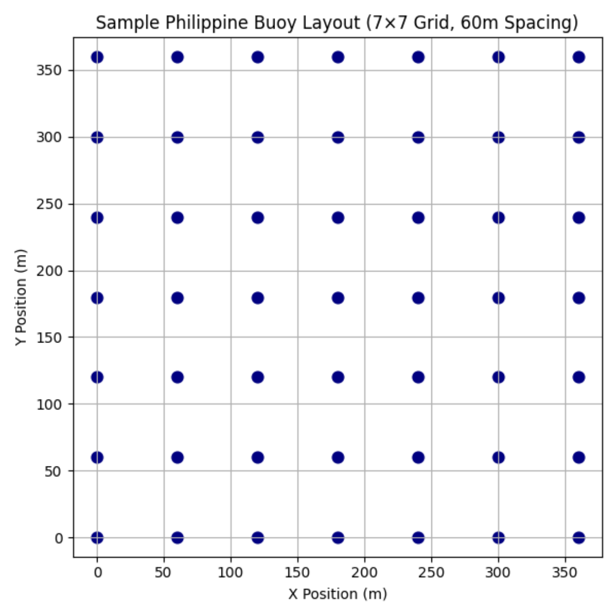


Fig. 1. Sample Philippine Buoy Layout (7×7 Grid, 60m Spacing).

*4) 100-Buoy Dataset (Extension):* We also experimented with the 100-buoy dataset (`WEC_Perth_100.csv`), which contains 7,277 layouts. This larger configuration has 200 spatial inputs (X1–Y100). Due to reduced sample size, model performance declined and generalization was limited. Detailed results are discussed in the next section.

## IV. Results and Discussion

### A. Model Performance on Perth 49-Buoy Dataset

The ReLU-activated MLP model achieved strong performance on the Perth 49-buoy training and test set. The final evaluation metrics were:

- **MAE:** 25,062.32 W
- **MSE:** 30,783,45603.22 $W^2$
- **Percent Error:** 0.64%

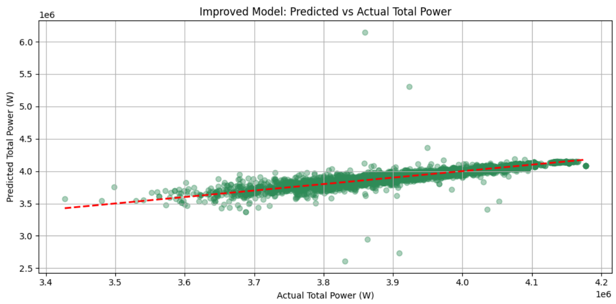


Fig. 2. Predicted vs Actual Power Output for Perth 49-Buoy Model

Fig. 2 shows the regression fit. Most points align closely to the $y = x$ line, indicating accurate power predictions.

### B. Effect of Training Configuration

Early experimentation used a higher learning rate and fewer epochs, which led to poor model performance. Reducing the learning rate and increasing the number of training iterations significantly improved results.

TABLE I
MODEL PERFORMANCE BEFORE AND AFTER TUNING

| Configuration | MAE | Percent Error |
|---|---|---|
| Original (lr=0.001, epochs=300) | 43,576.12 | 1.12% |
| Improved (lr=0.0005, epochs=1000) | 25,062.32 | 0.64% |

### C. Effect of Activation Functions

Alternate activation functions were tested. Both `tanh` and `logistic` yielded extremely poor results, with MAEs above 3.9 million and percent errors near 99.88%. ReLU significantly outperformed these:

TABLE II
ACTIVATION FUNCTION COMPARISON

| Activation | MAE | MSE | Percent Error |
|---|---|---|---|
| ReLU | 25,062.32 | 30,783,45603.22 | 0.64% |
| tanh | 3,923,935.90 | 154,827,231,2302.09 | 99.88% |
| logistic | 3,932,990.46 | 154,831,522,73511.07 | 99.88% |

This suggests that ReLU is more suitable for predicting total power output because, unlike `tanh` and `logistic` functions which squash their outputs into fixed ranges (e.g., between 0 and 1 or $-1$ and 1), ReLU allows outputs to grow freely. This flexibility is especially important in regression problems like this one, where the target values (power in watts) can be very large. The saturation of `tanh` and `logistic` makes it harder for the model to learn such wide-ranging values, leading to poor performance.

### D. Generalization on Sydney Layouts

The trained model was evaluated on the 49-buoy Sydney dataset. Performance degraded but remained acceptable, demonstrating the model's ability to generalize to different sea conditions:

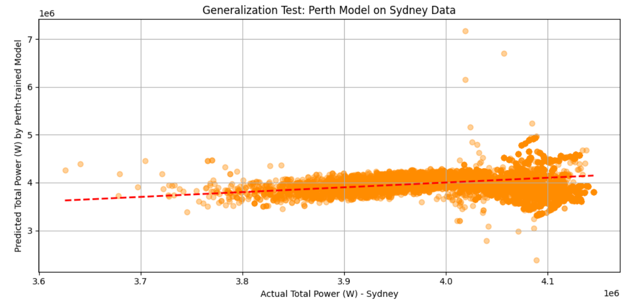


Fig. 3. Model Trained on Perth, Tested on Sydney 49-Buoy Layout

As shown in Fig. 3, although there is a noticeable spread, predictions remain within a reasonable range, suggesting partial generalizability.

### E. Testing on Sample Philippine Layout

A custom 7×7 grid buoy layout with 60-meter spacing was created and tested using the Perth-trained model to estimate potential energy output. This layout, shown in Fig. 1, was chosen for simplicity and generalizability. The model predicted a total power output of approximately **5.35 megawatts (5,350,615.65 watts)**.

It is important to note that this prediction reflects performance under Perth-like environmental conditions, not actual Philippine wave climates. Nonetheless, the high output suggests the model's ability to generalize to unseen spatial configurations. This supports the model's utility in early-stage spatial planning where site-specific data is not yet available.

For more reliable predictions tailored to local conditions, future work should follow approaches like those in Neshat et al. [1], where buoy layouts and energy outputs were generated using optimization algorithms and real wave climate models. Similar methods, applied to Philippine maritime data, would produce more actionable estimates for deployment.

### F. Scaling to 100-Buoy Layouts

The model was retrained using the larger 100-buoy dataset, which consists of 7,277 samples. Although this configuration introduced more input features (X1–Y100), the limited dataset size led to degraded performance. Table III summarizes the resulting errors for both training on Perth and testing on either Perth or Sydney layouts.

As shown in Table III and Figs. 4 and 5, the model achieved higher prediction errors compared to lower-dimensional cases.

TABLE III
100-BUOY MODEL GENERALIZATION PERFORMANCE

| Model | Train Data | Test Layout | % Error |
|---|---|---|---|
| 100-Buoy | Perth (7,277) | Perth | 3.67% |
| 100-Buoy | Perth | Sydney (100-buoy) | 10.18% |

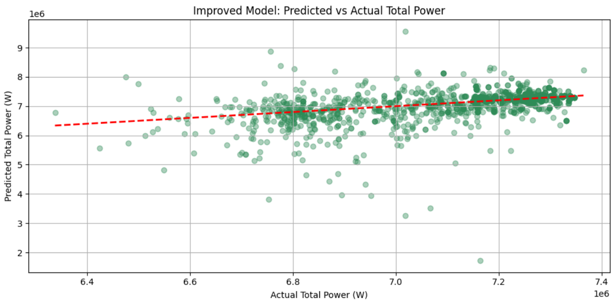


Fig. 4. Model Trained and Tested on Perth 100-Buoy Layout

This demonstrates the challenge of learning accurate representations in higher-dimensional input spaces when constrained by limited training data. The increased complexity demands more training examples to maintain generalizability.

## V. CONCLUSION

This study demonstrated the potential of using a Multi-Layer Perceptron (MLP) model to predict the total energy output of wave energy converter (WEC) farm layouts based solely on spatial coordinates. The model achieved high accuracy when trained and tested on the 49-buoy Perth dataset, particularly after tuning the training configuration and selecting the ReLU activation function.

Evaluation on a custom 7×7 buoy layout, used as a sample for a potential layout in Philippine setting, yielded a predicted output of approximately 5.35 megawatts. While promising, this result should not be interpreted as a real-world estimate for Philippine waters, as the model was trained under Perth-specific wave conditions. Nonetheless, it highlights the model's capacity to generalize to unseen spatial configurations.

Scaling to 100-buoy layouts introduced challenges due to the limited number of available training samples. Generalization performance declined significantly, especially when tested on layouts from different wave climates, emphasizing the difficulty of higher-dimensional spatial prediction with sparse data.

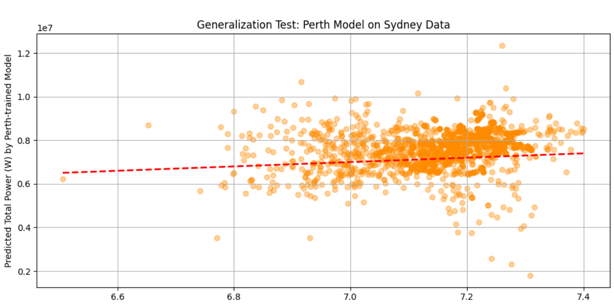


Fig. 5. Perth 100-Buoy Model Tested on Sydney 100-Buoy Layout

For future work, incorporating real site-specific environmental data and adopting methods such as optimization-based layout generation informed by local wave climate models—as seen in prior studies—would improve the reliability and applicability of predictions for actual deployment scenarios.

## REFERENCES

[1] Mehdi Neshat, Nataliia Y. Sergiienko, Bradley Alexander, and Markus Wagner. Optimisation of large wave farms using a multi-strategy evolutionary framework. *arXiv preprint arXiv:2003.09594*, 2020.

[2] UCI Machine Learning Repository, "Large Scale Wave Energy Farm Layout Dataset." [Online]. Available: https://archive.ics.uci.edu/dataset/882/large+scale+wave+energy+farm, [Accessed: Jun. 11, 2025].