Volume 8 No 2 December (2025)

e-ISSN: 2615-3270

p-ISSN: 2615-3599

Eigen Mathematics Journal

homepage: https://eigen.unram.ac.id/index.php/eigen

Bayesian Hyperparameter Optimization Analysis for Sustainable Innovation Performance Prediction Model

Tika Puspita[1*]

[1] *Department of Agro-industrial Technology, Universitas Brawijaya, Indonesia*
[*] *Corresponding author: tikapuspitapsb@student.ub.ac.id*

A B S T R A C T

This study examines how well the Gaussian Process Regression (GPR) model performs in interpreting the optimization outcomes achieved through Bayesian Optimization (BO) with Keras Tuner, specifically in the context of Sustainable Innovation Performance (SIP). The GPR surrogate model serves to examine the outcomes of optimization and offers valuable insights into the strategies of exploration and exploitation while seeking the most effective hyperparameters. The evaluation of the effectiveness of GPR involved calculating the Mean Absolute Error (MAE), which was bootstrapped 1000 times to establish a 95%. Confidence Interval (CI). This study's findings demonstrate the dependability of GPR in forecasting the validation loss generated by BO, characterized by minimal prediction errors and consistent confidence intervals. The results indicate that GPR serves as a dependable statistical method for assessing uncertainty in Bayesian-based optimization. Additionally, they offer valuable perspectives on how exploration and exploitation strategies can be utilized to attain optimal hyperparameter configurations. By strategically balancing exploitation and exploration, Bayesian Optimization can enhance the process of identifying optimal hyperparameter configurations within continuous innovation prediction models.

Keywords: Sustainable Innovation Performance, Bayesian Optimization, Predictive Modeling

Received	:	17-01-2025;
Revised	:	25-10-2025;
Accepted	:	03-11-2025;
Published	:	27-11-2025;

DOI: https://doi.org/10.29303/emj.v8i2.266

1. Introduction

Hyperparameter optimization represents a pivotal phase in the formulation of machine learning models, as it has a direct impact on the predictive performance of the model [1]. Conventional methods, including grid search and random search, frequently exhibit computational inefficiencies and demand significant time investment, particularly when dealing with an extensive hyperparameter space. Due to its capacity to balance exploration and exploitation, Bayesian Optimization (BO) has become a widely used alternative method for identifying optimal hyperparameters, thereby increasing its efficiency [2].

In Bayesian Optimization, intricate objective functions are approximated through a surrogate model, which is designed to probabilistically direct the exploration for hyperparameters. Gaussian

Process Regression (GPR) is a commonly utilized surrogate model. GPR is selected due to its capability to predict the value of the objective function while also offering uncertainty estimates [3]. The estimates of uncertainty hold significant relevance within the framework of Bayesian Optimization, as they facilitate the exploration of unfamiliar regions with minimal risk, while simultaneously allowing for the exploitation of regions that are already recognized for their high potential value.

The efficacy of Gaussian Process Regression as a surrogate model is contingent upon the configuration of its kernel. The Matérn kernel is frequently employed due to its adaptability in effectively modeling objective function patterns that exhibit limited smoothness [4]. Conversely, acquisition functions such as the Upper Confidence Bound (UCB) guide Bayesian Optimization (BO) in selecting the subsequent evaluation point by balancing exploration and exploitation [5].

In the realm of sustainable innovation, the effectiveness of companies engaged in sustainability-oriented innovation faces complex data dynamics that demand robust predictive modeling [6]. Sustainable Innovation Performance (SIP), which reflects a firm's ability to achieve innovation outcomes that support environmental, social, and economic sustainability, is a critical construct in this context. Bayesian Optimization (BO) with Gaussian Process Regression (GPR) has proven highly effective for hyperparameter tuning [7], and has been successfully applied in domains such as energy forecasting [8], state-of-health (SOH) prediction for lithium-ion batteries [9], and ecosystem service assessment [10]. Yet, its direct application to SIP prediction remains underexplored, despite the multidimensional complexity of sustainability indicators. This study addresses that gap by being among the first to empirically implement BO–GPR for SIP, aiming to improve accuracy, computational efficiency, and interpretability—factors essential for managerial decision-making in sustainable innovation strategies.

The research examines hyperparameter optimization through BO and evaluates GPR predictions in mapping optimization outcomes. By analyzing exploration–exploitation dynamics and validation loss, the study provides insights into GPR's reliability within the BO process. Performance is assessed using the Mean Absolute Error (MAE), with bootstrapping applied to establish a 95% confidence interval, ensuring robustness and trustworthy outcomes. In doing so, the study advances understanding of hyperparameter optimization in continuous innovation prediction models, extending the applicability of BO–GPR to the sustainability domain.

2. Literature Review

2.1. Sustainable Innovation Performance

Sustainable Innovation Performance (SIP) is gaining prominence as essential for enterprises, especially regarding the attainment of enduring sustainability objectives in conjunction with innovation. Evidence suggests that the pursuit of sustainability as a goal for innovation can markedly affect the efficiency of innovation processes, as illustrated by data envelopment analysis and regression analysis conducted within manufacturing firms [11]. Within the context of digital-based start-ups in Indonesia, research indicates that the business model has a substantial impact on sustainable performance, with innovation emerging as a more critical factor than customer participation [12].

Furthermore, the relationship between firm-specific capabilities such as absorptive capacity, intrapreneurship, and stakeholder integration and sustainable innovation has been investigated, revealing that these capabilities can strategically enhance sustainable innovation in Small and Medium-sized Enterprises (SMEs), albeit at an early [13]. Furthermore, the importance of creating new sustainability indicators that encompass economic, environmental, and social aspects has been highlighted to enhance the assessment of the sustainability of innovation processes [14].

It has also been discussed about how ambidexterity can help companies be more environmentally friendly. Companies that are ambidextrous may be better at coming up with new ideas and prolonging the life of their research and development [15]. According to [16], innovation practices emphasizing sustainability positively influence overall organizational performance, contributing to a deeper understanding of how corporate sustainability practices can be evaluated through innovation-related attributes. Overall, integrating sustainability into innovation strategies and business models, as well as strategically strengthening firm-specific capabilities, is crucial for achieving sustainable performance, particularly among SMEs. This underscores the importance of SIP and the need for its measurement across diverse organizational contexts.

2.2. Bayesian Optimization

Bayesian Optimization (BO) is an excellent technique for hyperparameter optimization, especially suited for costly black-box functions. It functions by progressively refining a surrogate model, usually a Gaussian Process (GP), to forecast the objective function and direct the search for optimal hyperparameters [17, 18]. The surrogate model equilibrates exploration and exploitation by an acquisition function, such as the Upper Confidence Bound (UCB) [19, 20].

Gaussian Process Regression (GPR) is frequently employed in Bayesian Optimization (BO) because of its capacity to deliver uncertainty quantifications and continuous predictions [21]. The procedure entails developing a Gaussian Process model from empirical data, thereafter employing it to forecast the goal function and identify new points for evaluation [18, 21]. Keras Tuner, a hyperparameter optimization tool, may be customized with a Matern kernel and UCB acquisition function to successfully balance exploration and exploitation. This configuration is especially advantageous for noisy and stochastic issues, facilitating resilient hyperparameter optimization [22]. BO's efficacy in hyperparameter optimization arises from its capacity to leverage previous evaluations to guide further searches, minimizing redundant evaluations and rapidly converging to near-optimal solutions [21, 23]. Given its advantages over grid search and random search, it is particularly well-suited for handling complex models and big datasets [20, 21].

3. Methodology

The model is validated with K-fold cross-validation and trained using the optimal parameters. Gaussian process regression and loss evaluation ensure the model's robustness and alignment with the data. Figure 1 outlines the research workflow, starting from problem identification and hyperparameter tuning.

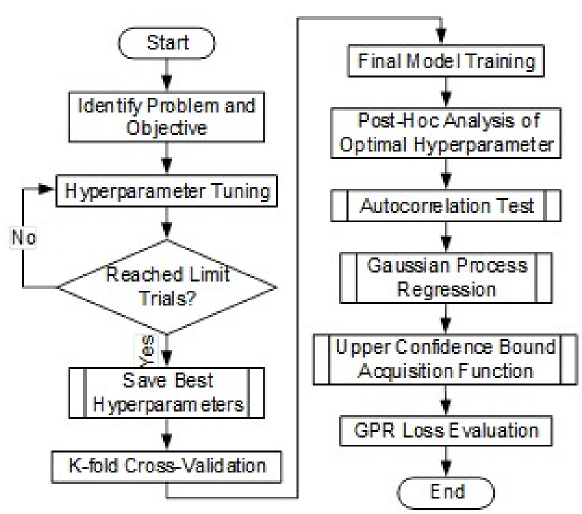

Figure 1. Flowchart of Research.

3.1. Data Collection

The distribution of surveys created using Google Forms provides the study's primary data. The study's respondents are from the food SMEs in Pekanbaru City. Partial Least Squares (PLS) model technique was then used to extract the latent score from the collected data. This latent score illustrates the link between strategic orientations and the performance of sustainable innovation (SIP). The Artificial Neural Network (ANN) model used the input as the latent score value. Bayesian Optimization (BO) is used to maximize hyperparameters thereby guiding the growth of the ANN model. There were 100 total trials conducted to find the ideal arrangement that fits the objectives of the study and the characteristics of the data. Examining how each hyperparameter affects the performance of the model helps one to optimize them. Every hyperparameter has predefined value limits meant to cover a variety of potential configurations. Table 1 lists the key hyperparameters applied in this work together with their data types.

Table 1. Hyperparameter Configuration.

No.	Hyperparameter	Data Type
1	units_input	int64
2	dropout_input	float64
3	num_layers	int64
4	units_1	int64
5	l2_regularization	float64
6	dropout_1	float64
7	learning_rate	float64
8	units_2	int64
9	dropout_2	float64
10	activation_input_relu	bool
11	activation_input_tanh	bool
12	activation_1_relu	bool
13	activation_1_tanh	bool
14	activation_2_relu	bool
15	activation_2_tanh	bool
16	optimizer_adam	bool
17	optimizer_rmsprop	bool
18	optimizer_sgd	bool

3.2. Data Analysis

3.2.1. Hyperparameter Configuration

This study intends to depict the relationship between hyperparameters applied in the artificial neural network model's optimization by means of a heatmap. Using a correlation matrix grounded on the Pearson correlation coefficient (r), which gauges the strength and direction of the linear link between two variables, correlation is computed. r ranges in value from -1 to 1; $r = 1$ denotes a perfect positive correlation, $r = -1$ denotes a perfect negative correlation, and $r = 0$ denotes no linear association.

This correlation matrix is computed using Pandas' `.corr()` function, using categorical data such as activation type or optimizer encoded using one-hot encoding to be treated as numerical variables. Using the Seaborn library, this heat map was produced with a color scale reflecting the degree of correlation: yellow represents a strong positive correlation, dark blue shows a strong negative correlation, and light blue to green indicates a weak or negligible link. Strong connections—both positive and negative—as well as relationships approaching zero are ignored in this analysis since they are deemed irrelevant using a significant criterion of $|r| > 0.5$.

3.2.2. Hyperparameter Optimization

Hyperparameter optimization seeks to minimize the loss function of the artificial neural network model by identifying the optimal combination of hyperparameters. This procedure employs Bayesian Optimization (BO), utilizing Gaussian Processes (GP) to improve predictions. The procedures for optimization are as follows.

- Gaussian Process Prior Distribution.
  The procedure commences with random sampling from the hyperparameter space, creating an initial dataset that serves as the Bayesian Optimization prior distribution. The primary aim is to reduce the objective function $f(\boldsymbol{h})$, where $\boldsymbol{h}$ represents the uncertain hyperparameter. The prior distribution is determined by the mean function $\mu(\boldsymbol{h})$ and the covariance function $k(\boldsymbol{h}, \boldsymbol{h}')$, as indicated in Equation 1. The covariance function employed is the Matern Kernel with $\nu = 2.5$, as delineated in Equation 2. This method aids in estimating the ideal hyperparameter values derived from the initial data.
$$f(\boldsymbol{h}) \sim GP(\mu(\boldsymbol{h}), k(\boldsymbol{h}, \boldsymbol{h}')). \tag{1}$$
$$k_{\text{Matern}}(\boldsymbol{h}, \boldsymbol{h}') = \sigma^2 \left(1 + \frac{\sqrt{5}\|\boldsymbol{h} - \boldsymbol{h}'\|}{l} + \frac{5\|\boldsymbol{h} - \boldsymbol{h}'\|^2}{3l^2}\right) \exp\left(-\frac{\sqrt{5}\|\boldsymbol{h} - \boldsymbol{h}'\|}{l}\right). \tag{2}$$
- Posterior Distribution Updates of the Gaussian Process.
  Subsequent to the preliminary assessment, the Gaussian Process (GP) model is updated using new observational data. This update yields a posterior distribution, as shown in Equation **??**, which can be used to predict unexplored hyperparameter configurations. The mean function $\mu(\boldsymbol{H})$ and the covariance function $K(\boldsymbol{H}, \boldsymbol{H})$ facilitate the continuous refinement of the objective function estimation, thereby enhancing the accuracy of subsequent hyperparameter selections.
$$p(f \mid D) = GP(\mu_{\text{posterior}}(\boldsymbol{h}), k_{\text{posterior}}(\boldsymbol{h}, \boldsymbol{h}')). \tag{3}$$
$$\begin{bmatrix} y \\ f_* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu(\boldsymbol{H}) \\ \mu(\boldsymbol{h}_*) \end{bmatrix}, \begin{bmatrix} K(\boldsymbol{H}, \boldsymbol{H}) + \sigma_I^2 & K(\boldsymbol{H}, \boldsymbol{h}_*) \\ K(\boldsymbol{h}_*, \boldsymbol{H}) & K(\boldsymbol{h}_*, \boldsymbol{h}_*) \end{bmatrix}\right). \tag{4}$$
- Upper Confidence Bound Acquisition Function (GP-UCB).
  The Upper Confidence Bound (UCB) function is employed to determine the subsequent hyperparameter, balancing exploration of uncertain regions and exploitation of well-performing areas, as articulated in Equation 5. The acquisition function is regulated by the parameter β, with a larger β value promoting exploration and a lower value emphasizing exploitation, as seen in Equation 6. The method is reiterated until the established stopping criteria are satisfied, namely after 100 iterations, at which point the optimal hyperparameters will be employed for final training.
$$a_{\text{ucb}}(\boldsymbol{h}; \beta) = \mu(\boldsymbol{h}) - \beta\sigma(\boldsymbol{h}). \tag{5}$$
$$\boldsymbol{h}_{\text{next}} = \arg\min_{\boldsymbol{h}} a_{\text{ucb}}(\boldsymbol{h}; \beta). \tag{6}$$

3.2.3. Post-Hoc Analysis of Hyperparameter Optimization

Post-hoc analysis aims to validate the results of hyperparameter optimization performed with Bayesian Optimization (BO) on Keras Tuner. This analysis uses Gaussian Process Regression (GPR) to explore patterns in the hyperparameter space and compare the predicted validation loss with the actual results from the optimization process. Several steps taken in the Post-Hoc Analysis are outlined as follows.

- Autocorrelation Test.
  The first step is to examine whether there is autocorrelation in the residual validation loss, which represents the dependence between consecutive residual values. The residual is defined as the difference between the actual value and the predicted value of the validation loss. The presence of autocorrelation may indicate that the model fails to capture data patterns or violates the assumption of residual independence, which affects the stability of predictions and the generalization of the model.
  A Durbin–Watson (DW) value close to 2 indicates that the model has met the assumption of residual independence, making the model valid for use in predictions. Conversely, if the DW value is far from 2 (approaching 0 or 4), the model needs to be re-evaluated to correct the residual pattern, for instance, by reviewing the model structure, adding parameters, or using regularization techniques to reduce overfitting. The autocorrelation test statistic is computed using Equation 7.

$$DW = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2} \tag{7}$$

- Gaussian Process Regression (GPR).
  After verifying that the residuals do not exhibit significant autocorrelation, Gaussian Process Regression (GPR) is used to model the objective function that predicts validation loss based on hyperparameters modeled as a multivariate normal distribution with mean $\mu(\boldsymbol{h})$ and covariance $k(\boldsymbol{h}, \boldsymbol{h}')$, according to Equation 1. This post hoc analysis leverages the default Bayesian Optimization (BO) setup provided by Keras Tuner, which employs a Matern kernel for flexibility in modeling both smooth and non-smooth variations in the objective function. The Matern kernel function is provided in Equation 2.
  Keras Tuner is an open-source hyperparameter tuning framework built on Keras and TensorFlow, supporting various strategies including Random Search, Hyperband, and Bayesian Optimization, and offering a define-by-run interface for constructing custom search spaces [24]. In its BO implementation, Keras Tuner models the objective function using a Gaussian Process surrogate and selects candidate hyperparameter configurations iteratively using an acquisition function—typically the Upper Confidence Bound (UCB)—to trade off exploration of uncertain regions and exploitation of known good configurations.
  To quantify the uncertainty of the surrogate model's predictions, a Confidence Interval (CI) is constructed based on the posterior distribution derived from the Gaussian Process after observing data. In this analysis, a 95% CI is used, which is computed using the predictive mean $\mu(\boldsymbol{h})$ and standard deviation $\sigma(\boldsymbol{h})$ of the Gaussian Process. This is given formally in Equation 8.

$$CI = \mu(\boldsymbol{h}) \pm 1.96 \cdot \sigma(\boldsymbol{h}) \tag{8}$$

- Upper Confidence Bound (GP-UCB) Acquisition Function in Post-hoc Analysis.

  In addition to GP predictions, the Upper Confidence Bound (UCB) is used to evaluate whether the experiment is directed by exploration or exploitation to select the next hyperparameter. UCB calculates the difference between the prediction uncertainty σ(h) and the prediction mean μ(h) as stated in Equation 4. In this analysis, $\beta = 2.6$ is used, following the default settings in Keras Tuner BO. The exploration hyperparameter selection strategy is indicated when the UCB value obtained is greater than the β value, while a smaller value indicates an exploitation strategy.
- GPR Loss Validation.
  After fitting the GPR model, predictive performance was evaluated using the bootstrap method, a statistical technique for estimating the distribution of an estimate by resampling from the existing data. In this study, 1000 bootstrap iterations were conducted to generate the empirical distribution of error values. The primary performance metric was the Mean Absolute Error (MAE), defined in Equation 9.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{9}$$

  where $\hat{y}_i$ is the predicted value from the GPR model and y_i is the observed validation loss. Bootstrap resampling yielded the empirical distribution of MAE, from which a 95% confidence interval was constructed. MAE was selected because it provides a natural and unambiguous measure of average error, unlike RMSE, which disproportionately weights larger deviations [25].

4. Result and Discussion

4.1. Hyperparameter Correlation

The hyperparameter correlation heatmap analysis, which employs a significant threshold of $|r| > 0.5$, identifies numerous robust relationships that can serve as a guide for the model optimization process, as depicted in Figure 2. The activation functions in the input layer (`activation_input_relu` and `activation_input_tanh`), the first hidden layer (`activation_1_relu` and `activation_1_tanh`), and the second hidden layer (`activation_2_relu` and `activation_2_tanh`) exhibited a perfect negative correlation (-1). This relationship suggests that the selection of one activation function automatically precludes the other in the same layer, underscoring the significance of consistency in the selection of activation functions to guarantee optimal model performance.

Furthermore, a robust negative correlation ($r = -0.69$) was identified between the Adaptive Moment Estimation (Adam) optimizer and the Root Mean Square Propagation (RMSprop) optimizer, suggesting that the utilization of one optimizer will considerably diminish the utilization of the other optimizer. In model training, the choice of an optimizer is quite important since every optimizer has special properties influencing the performance of neural networks in different environments. As noted by [26], the efficacy of optimizers such as Adam, RMSProp, and Stochastic Gradient Descent (SGD) can differ markedly, as seen by the varying accuracy rates in face expression recognition tasks. Adam regularly demonstrates superior accuracy relative to SGD and RMSProp, underscoring the significance of selecting optimizers that influence model performance.

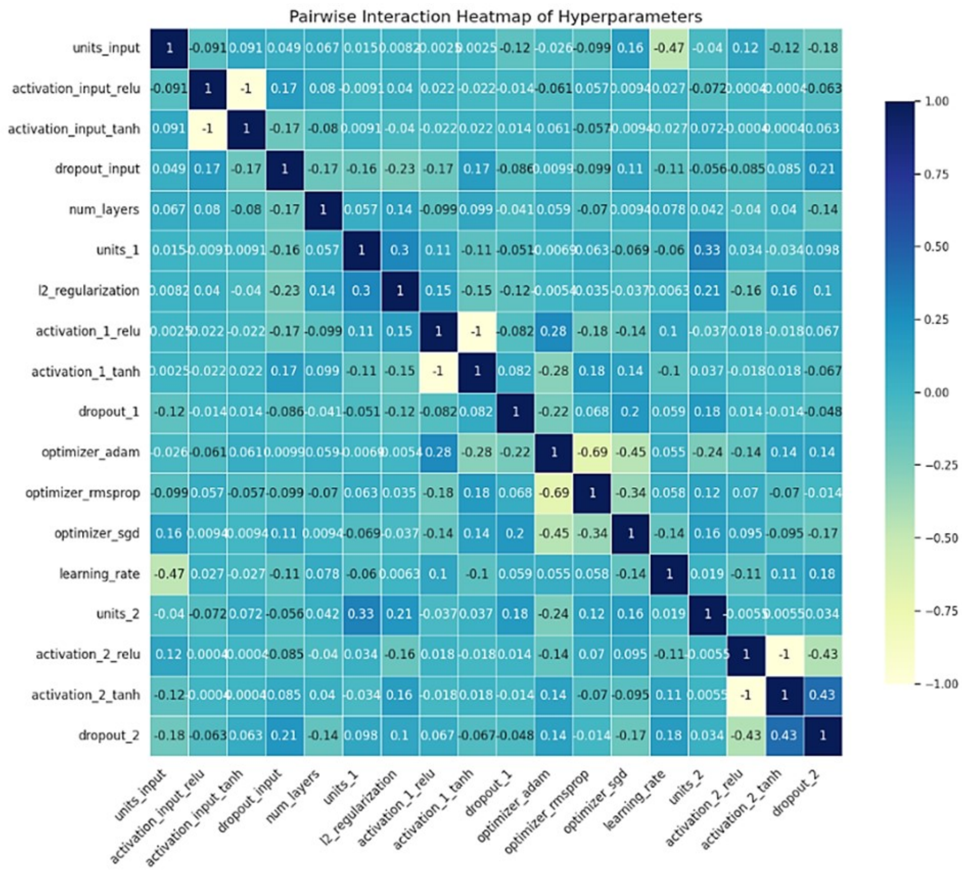

Figure 2. Hyperparameter Correlation.

4.2. Bayesian Hyperparameter Optimization

Bayesian Optimization (BO) demonstrates its efficacy in hyperparameter tuning by leveraging probabilistic models—particularly Gaussian Process Regression (GPR)—to establish the relationship between hyperparameter configurations and objective functions. As illustrated in Figure 3, the optimal point was identified with a minimum validation loss of 0.16766 at *Trial ID 68*. At the beginning of the optimization process (*Trial ID* < 40), the actual validation loss (red dots) fluctuates considerably, reflecting high uncertainty due to limited data. However, as more trials are conducted (*Trial ID* > 50), the values stabilize near zero, suggesting increased model certainty.

The mean GPR prediction (blue line) remains consistently close to the actual validation loss, illustrating the GPR model's ability to accurately describe the relationship between hyperparameters and validation loss. As the number of trials increases (*Trial ID* > 60), the model's predictive accuracy improves further. The Upper Confidence Bound (UCB) (dashed green line) reflects BO's exploration–exploitation strategy. Initially, elevated UCB values guide exploration of under-sampled regions; later, lower UCB values signify a shift towards exploiting promising configurations. As stated by [2], this method enables the model to capitalize on regions with previously recognized low validation loss predictions. This finding aligns with literature indicating that Bayesian Optimization with Gaussian Process Regression can successfully balance exploration and exploitation in hyperparameter optimization.

The width of the 95% confidence interval (CI) (purple region) represents the model's uncertainty. A wide CI during early trials suggests considerable predictive uncertainty, which progressively narrows as additional data is incorporated. After *Trial ID* > 60, the CI converges tightly around the mean prediction, indicating heightened confidence in regions with low validation loss.

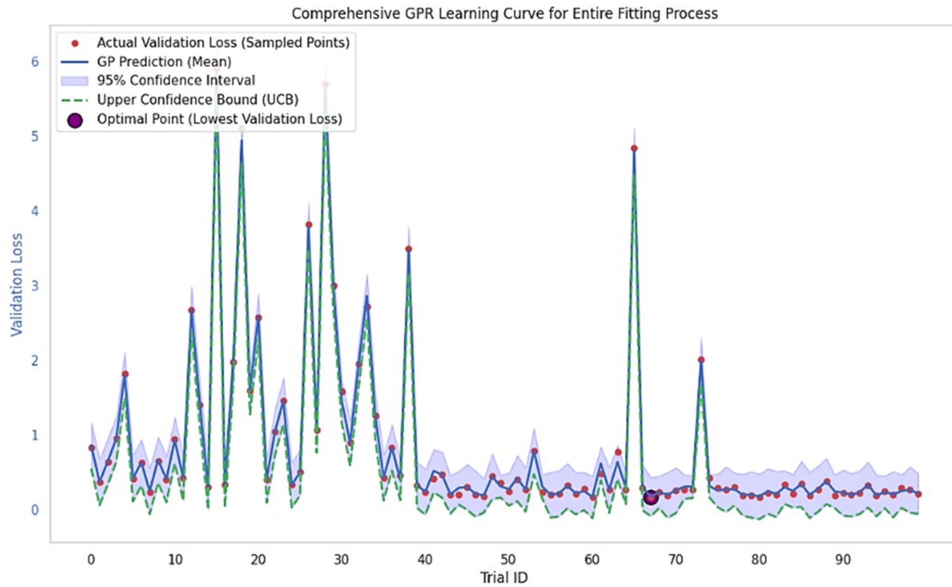

Figure 3. Bayesian Hyperparameter Optimization.

4.3. Bayesian Hyperparameter Optimization

4.3.1. Autocorrelation Test

The Durbin-Watson (DW) statistic of 1.6461 in the autocorrelation study for GPR prediction, depicted in Figure 4, signifies the absence of significant autocorrelation. A DW value near 2 often signifies that the residuals exhibit no autocorrelation, indicating the independence of the model's residuals. Given the absence of considerable autocorrelation, the residuals may be regarded as independent, so allowing the post-hoc analysis to proceed with greater assurance, free from concerns of distortions caused by residual dependency.

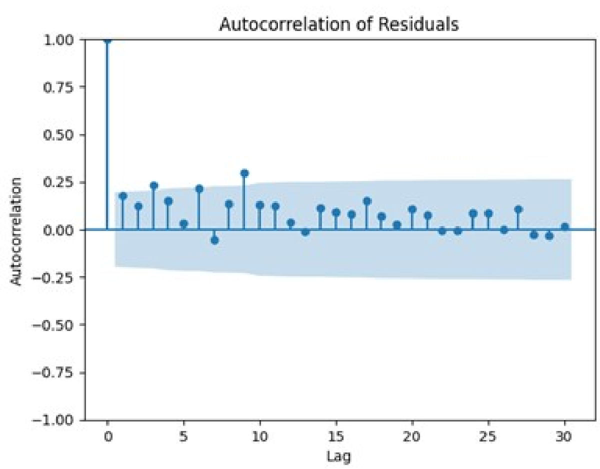

Figure 4. Autocorrelation Test.

4.3.2. Gaussian Process Regression (GPR)

Post-hoc analysis assesses the execution of exploration and exploitation tactics by Bayesian Optimization (BO) in the pursuit of optimal hyperparameter configurations through the application of Gaussian Process Regression (GPR) predictions. The analytical results demonstrate that the optimization process is mostly influenced by exploitation, whereas exploration is strategically performed in multiple trials to expand the search range. Table 2 shows the results of the GPR Analysis for all of the trials.

Table 2. GPR Analysis.

Trials	Actual Val Loss	GP Prediction	UCB	Strategy
1	0.83748	0.86085	0.55258	Exploitation
2	0.37101	0.37443	0.06558	Exploitation
3	0.64432	0.65477	0.34609	Exploitation
4	0.95278	0.95407	0.64535	Exploitation
5	1.81938	1.80635	1.49814	Exploitation
6	0.41420	0.42389	0.11524	Exploitation
7	0.63740	0.63462	0.32657	Exploitation
8	0.24012	0.25293	-0.05567	Exploitation
9	0.64931	0.67026	0.36301	Exploitation
10	0.41207	0.41283	0.10434	Exploitation
11	0.94405	0.93385	0.62639	Exploitation
12	0.42915	0.43366	0.12502	Exploitation
13	2.68014	2.67673	2.36822	Exploitation
14	1.40608	1.42157	1.11335	Exploration
15	0.30520	0.30805	-0.00013	Exploitation
16	5.90631	5.85811	5.54920	Exploitation
17	0.33596	0.34082	0.03257	Exploitation
18	1.98316	1.96841	1.65982	Exploitation
19	5.10885	4.94261	4.64298	Exploitation
20	1.59622	1.58651	1.27825	Exploration
21	2.57651	2.58781	2.28095	Exploration
22	0.40359	0.40629	0.09801	Exploration
23	1.04396	1.05043	0.74196	Exploitation
24	1.46721	1.45899	1.15030	Exploitation
25	0.33874	0.33480	0.03122	Exploitation
26	0.50514	0.50306	0.19460	Exploitation
27	3.82562	3.79802	3.49301	Exploitation
28	1.07310	1.07651	0.76774	Exploitation
29	5.69579	5.63995	5.33141	Exploration
30	2.99834	2.96089	2.65504	Exploitation
31	1.58157	1.45050	1.15909	Exploitation
32	0.90655	0.89604	0.58803	Exploitation
33	1.95181	1.93730	1.62865	Exploitation
34	2.72435	2.86359	2.56529	Exploitation
35	1.25946	1.26164	0.95322	Exploitation
36	0.42502	0.43263	0.12425	Exploitation
37	0.83293	0.83278	0.52427	Exploitation
38	0.46409	0.40261	0.13268	Exploration
39	3.49118	3.47769	3.17186	Exploitation
40	0.33240	0.33404	0.02694	Exploitation
41	0.24033	0.24109	-0.06491	Exploitation
42	0.42379	0.52250	0.23428	Exploitation
43	0.47383	0.47202	0.16528	Exploitation
44	0.20841	0.20708	-0.04845	Exploitation
45	0.20771	0.29630	0.07169	Exploitation

Table 3. GPR Analysis.

Trials	Actual Val Loss	GP Prediction	UCB	Strategy
46	0.31091	0.30645	0.00059	Exploitation
47	0.20520	0.21708	-0.08974	Exploitation
48	0.18424	0.19337	-0.03530	Exploitation
49	0.44894	0.44721	0.14127	Exploitation
50	0.35742	0.31610	0.16657	Exploitation
51	0.25032	0.27789	0.05771	Exploitation
52	0.41012	0.42416	0.11685	Exploitation
53	0.27343	0.27281	-0.02246	Exploitation
54	0.78775	0.78694	0.47809	Exploitation
55	0.24098	0.31363	0.15223	Exploitation
56	0.20112	0.20236	-0.10218	Exploitation
57	0.21130	0.22148	-0.08699	Exploitation
58	0.32348	0.32299	0.02039	Exploitation
59	0.21304	0.22153	-0.05880	Exploitation
60	0.28768	0.25216	-0.00328	Exploitation
61	0.17628	0.17408	-0.10819	Exploitation
62	0.48098	0.62179	0.39706	Exploitation
63	0.26888	0.27016	-0.03740	Exploitation
64	0.77819	0.63931	0.41458	Exploitation
65	0.27419	0.30179	0.07820	Exploitation
66	4.84735	4.80627	4.49739	Exploitation
67	0.29824	0.29924	-0.00557	Exploitation
68	0.16766	0.17661	-0.08342	Exploitation
69	0.23811	0.24060	0.02558	Exploitation
70	0.19927	0.20067	-0.10604	Exploitation
71	0.25881	0.26390	-0.04218	Exploitation
72	0.27241	0.31061	0.14772	Exploitation
73	0.27167	0.31263	0.15858	Exploitation
74	2.00952	1.99421	1.68580	Exploitation
75	0.42724	0.31930	0.16834	Exploitation
76	0.29949	0.26070	0.03634	Exploration
77	0.27235	0.27533	-0.02862	Exploitation
78	0.30307	0.27854	0.05841	Exploitation
79	0.19797	0.20146	-0.07787	Exploitation
80	0.19881	0.20608	-0.10219	Exploitation
81	0.17353	0.17962	-0.12352	Exploitation
82	0.22675	0.25244	-0.05113	Exploitation
83	0.20955	0.21307	-0.09499	Exploitation
84	0.33449	0.30538	0.08166	Exploitation
85	0.21869	0.25376	0.02977	Exploitation
86	0.35071	0.34728	0.04058	Exploitation
87	0.19567	0.20257	-0.10609	Exploitation
88	0.26857	0.28076	-0.02671	Exploitation
89	0.38503	0.38520	0.07751	Exploitation
90	0.19439	0.24926	0.02677	Exploitation
91	0.22876	0.23148	-0.07610	Exploitation
92	0.20390	0.20200	-0.08816	Exploitation
93	0.23142	0.23638	-0.05539	Exploitation
94	0.33272	0.33886	0.03221	Exploitation
95	0.19569	0.19187	-0.08658	Exploitation
96	0.24805	0.24142	0.02485	Exploitation
97	0.20628	0.20932	-0.09936	Exploitation
98	0.29054	0.25171	0.02884	Exploitation
99	0.27008	0.27045	-0.03741	Exploitation
100	0.22067	0.22001	-0.05136	Exploitation

4.3.3. Exploitation as the dominant strategy

Most trials demonstrate that BO prioritizes exploitation. (Specifically, trials 1-13 and 68-100). During this phase, the validation loss numbers are extremely close to the GPR predictions, indicating that the model can accurately map the link between hyperparameter settings and model performance. For example, in Trial 68, with the lowest validation loss of 0.16766, the GPR prediction was just slightly off at 0.17661. The accuracy of this forecast demonstrates how BO effectively uses the available information to expedite convergence to the optimal configuration.

4.3.4. Strategic Exploration to broaden the search

Although exploitation predominates, exploration is nevertheless carried out at crucial points, such as in Trials 14, 20, 22, 29, and 38, where the Upper Confidence Bound (UCB) value exceeds the GPR projection. For example, in Trial 29, the UCB hit 5.33141, showing that the model is exploring previously unexplored hyperparameter domains with considerable uncertainty. This strategy is critical to ensuring that possible untested regions are not neglected. However, the low frequency of exploration suggests a need for a better balance between exploration and exploitation.

4.3.5. Enhancement of GPR Prediction Precision

The study results indicate an enhancement in GPR prediction accuracy as the optimization process advances. During the initial phases, the disparity between the actual validation loss and the GPR forecast is typically substantial, indicating uncertainty stemming from insufficient data. As the data accumulates with each trial, the predictions attain greater precision. In Trials 60–100, the GPR forecasts closely corresponded with the actual validation loss levels, indicating the model's enhanced capacity to discern data patterns and accurately forecast hyperparameter performance.

4.4. GPR Loss Evaluation

The examination of Gaussian Process Regression (GPR) predictions using Mean Absolute Error (MAE), measured 1000 times with the bootstrap method, yields significant findings regarding the model's trustworthiness. This study produced an MAE value of 0.022474 with a 95% confidence interval of [0.016213, 0.029521]. This low MAE value indicates that the average prediction error of GPR against the actual loss value is extremely small, demonstrating the model's capability to accurately represent the relationship between hyperparameters and model performance. Moreover, the narrow confidence interval reflects the stability of GPR predictions even after 1000 recalculations with bootstrapping.

The bootstrap distribution, illustrated in Figure 5, exhibits a symmetric pattern resembling a normal distribution. Its peak aligns closely with the actual MAE value of 0.022474, supporting the credibility of the prediction results. The vertical reference lines in the figure—including the MAE baseline—indicate that GPR predictions achieve substantially higher accuracy compared to the baseline. In addition, the line representing the actual MAE lies near the apex of the bootstrap distribution, confirming that the predicted values effectively reflect the model's performance. The lower and upper bounds of the confidence interval, 0.016213 and 0.029521, ensure that the GPR prediction error remains within this interval at a 95% confidence level.

Overall, these findings demonstrate that GPR is a reliable model for forecasting loss values in Bayesian Optimization (BO). With a low prediction error and small uncertainty, GPR effectively guides both the exploration and exploitation processes in BO, offering precise and consistent estimates as the optimization approaches the optimal solution.

Furthermore, this work successfully applies Bayesian Optimization (BO) for hyperparameter tuning of Gaussian Process Regression (GPR) to predict Sustainable Innovation Performance (SIP), building on previous successful applications of BO–GPR in areas such as energy forecasting and battery state-of-health prediction. For example, [9] demonstrated the strength of BO–GPR with results including a 0.11% MAPE and R^2 = 0.9915 for lithium-ion batteries. Similarly, the BO–GPR model in this study achieves an MAE of 0.022474, with a 95% confidence interval of [0.016213, 0.029521], reflecting its robustness in modeling SIP, a complex and multidimensional domain.

The novelty of this research lies in applying hyperparameter-optimized BO–GPR to SIP, demonstrating its potential to deliver accurate, efficient, and interpretable predictions across environmental, social, and economic sustainability dimensions. These results provide valuable insights for sustainability-oriented innovation.

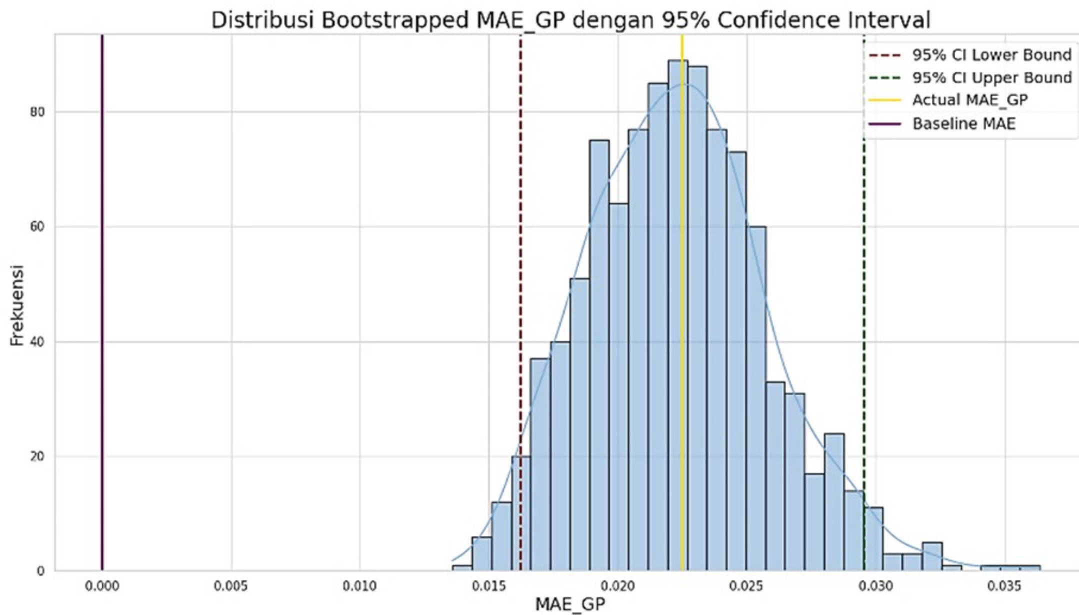

Figure 5. Bootstrapped GPR Loss Distribution.

5. Conclusion

Bayesian Optimization (BO) results have been effectively analyzed using Gaussian Process Regression (GPR), which provides valuable insights into exploration and exploitation strategies. GPR demonstrates strong capability in accurately predicting the validation loss produced by BO, supported by a stable 95% confidence interval and relatively small prediction errors. These findings indicate that GPR can be reliably used to interpret and refine BO outcomes, particularly within hyperparameter optimization settings relevant to sustainable innovation applications. Overall, this study highlights the reliability of GPR in guiding the exploration–exploitation mechanism of BO and in enhancing the performance of predictive models.

Appendices

$f(\boldsymbol{h})$	Objective function to be minimized or target function in Gaussian process
GP	Gaussian process
$\mu(\boldsymbol{h})$	Mean function in Gaussian process
$k(\boldsymbol{h}, \boldsymbol{h}')$	Kernel (covariance) function between two hyperparameters
$p(f \mid D)$	Posterior probability of the function given the hyperparameter dataset
$\mu_{\text{posterior}}(\boldsymbol{h})$	Posterior mean in Gaussian process
$k_{\text{posterior}}(\boldsymbol{h}, \boldsymbol{h}')$	Posterior covariance in Gaussian process
$k_{\text{Matern}}(\boldsymbol{h}, \boldsymbol{h}')$	Matérn covariance function
σ^2	Process variance or noise variance
l	Length-scale parameter
$\|\boldsymbol{h} - \boldsymbol{h}'\|$	Euclidean distance between two hyperparameter vectors
exp	Exponential function
$\boldsymbol{y}$	Observation output vector
f_*	Function prediction at a new point
$\mathcal{N}$	Multivariate normal distribution
$\mu(H)$	Mean function at observation data
$\mu(\boldsymbol{h}_*)$	Mean function at prediction point
$K(H, H)$	Covariance matrix between observation data
σ_I^2	Noise variance multiplied by the identity matrix
$K(H, \boldsymbol{h}_*)$	Covariance between observation data and prediction point
$K(\boldsymbol{h}_*, H)$	Covariance between prediction point and observation data
$K(\boldsymbol{h}_*, \boldsymbol{h}_*)$	Covariance at prediction point
$a_{\text{ucb}}(\boldsymbol{h}; \beta)$	UCB acquisition function value at hyperparameter $\boldsymbol{h}$
β	Acquisition function parameter
$\sigma(\boldsymbol{h})$	Prediction standard deviation at $\boldsymbol{h}$
$\boldsymbol{h}_{\text{next}}$	Next hyperparameter to evaluate
$\arg\min$	Argument that minimizes a function
DW	Durbin–Watson statistic
e_t	Residual at time t
CI	Confidence interval

REFERENCES

[1] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *Journal ofMachine Learning Research*, vol. 13, pp. 281–305, 2012. http://dx.doi.org/10.1145/3575882.3575900.

[2] J. Snoek, H. Larochelle, and R. P. Adams, "Practical bayesian optimization of machine learning algorithms," *Advances in Neural Information Processing Systems*, vol. 25, pp. 2960–2968, 2012. http://dx.doi.org/10.1163/15685292-12341254.

[3] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning.* MIT Press, 2006. www.GaussianProcess.org/gpml.

[4] Z. Y. Shnain, A. K. Mageed, H. S. Majdi, M. Mohammadi, A. A. AbdulRazak, and M. F. Abid, "Investigating the effect of tio2-based nanofluids in the stability of crude oil flow: parametric analysis and gaussian process regression modeling," *Journal of Petroleum Exploration and Production Technology*, vol. 12, no. 9, pp. 2429–2439, 2022. https://dx.doi.org/10.1007/s13202-022-01473-6.

[5] C. Bian, X. Wang, C. Liu, X. Xie, and L. Haitao, "Impact of exploration-exploitation trade-off on ucb-based bayesian optimization," *IOP Conference Series: Materials Science and Engineering*, vol. 1081, no. 1, p. 012023, 2021. https://dx.doi.org/10.1088/1757-899x/1081/1/012023.

[6] S. Liu, Y. Xiang, and H. Zhou, "A deep learning-based approach for high-dimensional industrial steam consumption prediction to enhance sustainability management," *Sustainability (Switzerland)*, vol. 16, no. 22, 2024. http://dx.doi.org/10.3390/su16229631.

[7] C. Ji, "Predicting fuel consumptions and exhaust gas emissions for lng carriers via machine learning with hyperparameter optimization," in *SNAME 26th Offshore Symposium, TOS 2021*, OnePetro, 2021. https://dx.doi.org/10.5957/TOS-2021-09.

[8] Y. Hamoudi, H. Amimeur, D. Aouzellag, M. G. M. Abdolrasol, and T. S. Ustun, "Hyperparameter bayesian optimization of gaussian process regression applied in speed-sensorless predictive torque control of an autonomous wind energy conversion system," *Energies*, vol. 16, no. 12, p. 4738, 2023. https://dx.doi.org/10.3390/en16124738.

[9] Q. Liu, X. Bao, D. Guo, and L. Li, "Bayesian optimization algorithm-based gaussian process regression for in situ state of health prediction of minorly deformed lithium-ion battery," *Energy Science & Engineering*, vol. 12, no. 4, pp. 1472–1485, 2024. https://dx.doi.org/10.1002/ese3.1678.

[10] Y. Liu, T. Yang, L. Tian, B. Huang, J. Yang, and Z. Zeng, "Ada-xg-catboost: A combined forecasting model for gross ecosystem product (gep) prediction," *Sustainability (Switzerland)*, vol. 16, 2024. https://dx.doi.org/10.3390/su16167203.

[11] J. Shin, C. Kim, and H. Yang, "The effect of sustainability as innovation objectives on innovation efficiency," *Sustainability*, vol. 10, no. 6, p. 1966, 2018. http://dx.doi.org/10.3390/su10061966.

[12] P. A. Danarahmanto, I. Primiana, Y. Azis, and U. Kaltum, "The sustainable performance of the digital start-up company based on customer participation, innovation, and business model," *Business: Theory and Practice*, vol. 21, no. 1, pp. 115–124, 2020. https://dx.doi.org/10.3846/btp.2020.11053.

[13] A. Widya-Hasuti, A. Mardani, D. Streimikiene, A. Sharifara, and F. Cavallaro, "The role of process innovation between firm-specific capabilities and sustainable innovation in smes: Empirical evidence from indonesia," *Sustainability*, vol. 10, no. 7, p. 2244, 2018. http://dx.doi.org/10.3390/su10072244.

[14] A. Melane-Lavado and Álvarez-Herranz, Agustín, "Different ways to access knowledge for sustainability-oriented innovation. the effect of foreign direct investment," *Sustainability (Switzerland)*, vol. 10, no. 11, 2018. http://dx.doi.org/10.3390/su10114206.

[15] F. Michelino, A. Cammarano, A. Celone, and M. Caputo, "The linkage between sustainability and innovation performance in it hardware sector," 2019. https://dx.doi.org/10.3390/su11164275.

[16] Maletič, Matjaž and Maletič, Damjan and Dahlgaard, Jens J. and Dahlgaard-Park, Su Mi and Gomišček, Boštjan, "The relationship between sustainability– oriented innovation practices and organizational performance: Empirical evidence from slovenian organizations," *Organizacija*, vol. 47, no. 1, pp. 3–13, 2014. http://dx.doi.org/10.2478/orga-2014-0001.

[17] S. Putatunda and K. Rama, "A modified bayesian optimization based hyper-parameter tuning approach for extreme gradient boosting," *2019 15th International Conference on Information Processing: Internet of Things, ICINPRO 2019 - Proceedings*, pp. 22–27, 2019. http://dx.doi.org/10.1109/ICInPro47689.2019.9092025.

[18] M. Zhang, H. Li, and S. Su, "High dimensional bayesian optimization via supervised dimension reduction," *IJCAI International Joint Conference on Artificial Intelligence*, pp. 4292–4298, 2019. https://dx.doi.org/10.24963/ijcai.2019/596.

[19] R. Moriconi, K. S. Kumar, and M. P. Deisenroth, "High-dimensional bayesian optimization with projections using quantile gaussian processes," *Optimization Letters*, vol. 14, no. 1, pp. 51–64, 2020. http://dx.doi.org/10.1007/s11590-019-01433-w.

[20] T. T. Joy, S. Rana, S. Gupta, and S. Venkatesh, "Fast hyperparameter tuning using bayesian optimization with directional derivatives," *Knowledge-Based Systems*, vol. 205, 2020. https://dx.doi.org/10.1016/j.knosys.2020.106247.

[21] M. Feurer and F. Hutter, *Hyperparameter Optimization*. 2019. http://dx.doi.org/10.1007/978-3-030-05318-5_1.

[22] M. Parsa, J. P. Mitchell, C. D. Schuman, R. M. Patton, T. E. Potok, and K. Roy, "Bayesian multi-objective hyperparameter optimization for accurate, fast, and efficient neural network accelerator design," *Frontiers in Neuroscience*, vol. 14, no. July, pp. 1–16, 2020. http://dx.doi.org/10.3389/fnins.2020.00667.

[23] L. Yang and A. Shami, "On hyperparameter optimization of machine learning algorithms: Theory and practice," *Neurocomputing*, vol. 415, pp. 295–316, 2020. http://dx.doi.org/10.1016/j.neucom.2020.07.061.

[24] T. O'Malley, E. Bursztein, J. Long, F. Chollet, H. Jin, and L. Invernizzi, "Keras tuner," 2019. https://github.com/keras-team/keras-tuner.

[25] C. Willmott and K. Matsuura, "Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance," *Climate Research*, vol. 30, no. 1, pp. 79–82, 2005. https://dx.doi.org/10.3354/cr030079.

[26] D. Anggara, N. Suarna, and Y. Arie Wijaya, "Performance comparison analysis of optimizer adam, sgd, and rmsprop on the h5 model," *Jurnal Ilmiah NERO*, vol. 8, no. 1, p. 2023, 2023. https://dx.doi.org/10.21107/nero.v8i1.19226.