ISSN 2723-6471

Optimizing The XGBoost Model with Grid Search Hyperparameter Tuning for Maximum Temperature Forecasting

Sugiarto[1,*], I Gede Susrama Mas Diyasa[2], Denisa Septalian Alhamda[3], Rangga Laksana Aryananda[4], Allan Ruhui Fatmah Sari[5], Hanifudin Sukri[6], Deshinta Arrova Dewi[7]

[1,2,3,4,5]*Faculty of Computer Science, University of Pembangunan Nasional "Veteran" Jawa Timur, Surabaya, Indonesia, 60294*

[6]*Faculty of Engineering, University of Trunojoyo Madura, Bangkalan, Jawa Timur, 69162*

[7]*Center for Data Science and Sustainable Technologies, INTI International University*

(Received: March 01, 2025; Revised: May 25, 2025; Accepted: August 10, 2025; Available online: September 11, 2025)

Abstract

This study presents a novel comparative approach to maximum temperature forecasting in Surabaya, Indonesia, by integrating Extreme Gradient Boosting (XGBoost) with Grid Search Hyperparameter Tuning and benchmarking it against Autoregressive Integrated Moving Average (ARIMA) and Neural Prophet models. The main idea is to evaluate the capability of XGBoost in capturing nonlinear patterns in environmental time series data, which traditional models often fail to address. Using 15,388 historical daily maximum temperature records from the BMKG Juanda weather station spanning 1981–2022, the objective is to identify the most accurate predictive model for short- and medium-term forecasts. The modeling process involved four stages: data acquisition, preprocessing, training, and evaluation, with performance assessed using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). The findings show that, after hyperparameter tuning, XGBoost achieved the best performance with MAE = 0.32 and RMSE = 0.65, outperforming ARIMA (MAE = 0.85, RMSE = 1.20) and Neural Prophet (MAE = 0.70, RMSE = 0.98). Prediction results for 2025 indicate peak maximum temperatures in January, October, and November, aligning with recent climate patterns. The contribution of this research lies in demonstrating the superiority of a tuned XGBoost model for complex environmental datasets, offering a practical tool for urban climate planning, agricultural scheduling, and heatwave risk mitigation. The novelty of this work is the systematic integration of Grid Search-based optimization with XGBoost for meteorological forecasting in a tropical urban context, producing higher accuracy than both classical statistical and modern hybrid time series methods. These results highlight the model's adaptability and potential for broader climate-related applications, with future research recommended to incorporate additional meteorological variables such as humidity and wind speed for even greater predictive capability.

Keywords: Forecasting, Maximum Temperature, XGBoost, Hyperparamater, Grid Search, Climate Change, Innovation

1. Introduction

Temperature is a quantity or measure of the degree of heat and cold in an area. Temperature greatly affects daily human activities [1], [2]. Humans unconsciously assess environmental conditions based on stimuli perceived through the five senses and responded to the brain to be assessed [3], [4]. Temperature is often used as an important attribute in a study so it is interesting to look deeper into other research objects [5], [6].

Based on previous research funded by ASHRAE and documented in their standards such as ASHRAE 55-1992 and ISO 7730, thermal comfort can be defined as the condition in which at least 80% of building occupants express satisfaction with the thermal environment. These standards provide a framework for assessing comfort based on both environmental and personal factors [7]. The recommended effective temperature for thermal comfort in tropical climates is 21°C–25°C, with approximately 23.4°C considered optimal and temperatures above 25.6°C leading to discomfort. To achieve this range in hot and humid conditions, passive design strategies such as solar shading and cross-ventilation are advised [8].

***Corresponding author: Sugiarto (sugiarto.if@upnjatim.ac.id)**
DOI: https://doi.org/10.47738/jads.v6i4.885

The current phenomenon of global warming, resulting in increasingly hot temperatures and a significant impact on the earth [9]. Global warming is the increase in average temperatures across the earth's surface due to the emission of large amounts of greenhouse gases that cause heat energy to be trapped in the atmosphere [10], [11]. The impacts of these phenomena contribute to climate change, which can alter long-term weather patterns, the climate system and human life as a whole [12], [13]. In addition, the impacts of global warming and climate change can be felt in Indonesia, including Surabaya City, East Java [14], [15].

A regional climate simulation using the RegCM5 model driven by ERA5 reanalysis data shows that Surabaya has experienced a rise in maximum temperatures of approximately 1.5 °C from 2020 to 2023, highlighting accelerated local warming trends [16]. This warming occurs in a city that is already the second largest in Indonesia, with a population of 2,987,863 in 2022, which continues to grow and place increasing demands on land, infrastructure, and public services [17]. Apart from the temperature, due to poor air quality with dense population and high vehicle usage, this greatly affects the effectiveness of daily activities [18].

The problems caused by high temperatures are very diverse, such as the increase in skin diseases [19], [20], [21], respiration caused by the reaction of chemical compounds with high temperatures [22], [23], excessive fatigue and dehydration [24], [25]. In agriculture, hot temperatures can be detrimental to the production of food commodities due to drought [26], [27]. A hot environment can worsen air quality which will impact overall quality of life [28], [29], including convenience [30], productivity [31], [32] and affect mood [6].

Based on these problems, the purpose of this research is to optimize the performance of the XGBoost model using Grid Search Hyperparameter Tuning and compare its forecasting accuracy against two other predictive models, namely ARIMA which is widely used due to its solid performance in linear time series. Neural Prophet is a relatively new model that has not been extensively applied for maximum temperature forecasting based on historical data. Through this comparative analysis, the study aims to identify the most effective model for temperature trend prediction in Surabaya. The results are expected to provide valuable insights to support decision-making in resource management, urban development, and extreme temperature risk mitigation, thereby improving the quality of life for the residents of Surabaya. As this study utilizes secondary meteorological data provided by BMKG and does not involve human participants or personal information, ethical approval and informed consent were not required.

2. Literature Review

The selection of XGBoost, ARIMA, and Neural Prophet in this study is based on their popularity and reported performance in time series forecasting. ARIMA is a classical linear model widely used for forecasting temperature due to its strength in modeling autoregressive and moving average patterns. Neural Prophet is a modern extension of Facebook Prophet, designed to handle time series with strong seasonality and holiday effects. Meanwhile, XGBoost is a powerful ensemble learning method known for its capability to capture complex, non-linear relationships, especially in large datasets with high dimensionality.

This research refines and extends prior work on time series forecasting. In 2019, a study research builds upon previous work on time series forecasting by incorporating findings from a meta-analysis of 18 studies, which showed that reducing classroom temperatures from 30 °C to 20 °C can improve cognitive performance by about 20%, with optimal performance below 22 °C—though this was validated only for temperate climates [33]. In 2021, another study applied XGBoost for retail sales prediction, integrating feature engineering and weather factors, and outperformed ARIMA, LSTM, Prophet, and GBDT, achieving RMSEs of 0.2256 and 0.0632 on two datasets, confirming its accuracy and reliability in time series prediction [34].

Then, in the ARIMA model can be used to forecast regional temperature and precipitation in the near term. Annual projections from ARIMA models integrate recent observations with long-term historical trends, estimate confidence intervals and simulate future daily temperature and rainfall. This research found that ARIMA models provide more accurate and reliable temperature and rainfall projections than other common statistical techniques. Short-term temperature and rainfall projections can be interpreted and trusted for civil and environmental engineering applications. ARIMA-based methods can be an effective alternative for obtaining near-term regional climate information using local historical observation data [35].

 ISSN 2723-6471

The performance of Neural Prophet in 2023 was evaluated by Teuku Rizky Noviandy et al. [36] in Deep Learning-Based Bitcoin Price Forecasting Using Neural Prophet, which used historical Bitcoin price data from 2014 to 2023 and achieved an RMSE of 6117.16, MAE of 4008.28, and MAPE of 1.77%, indicating strong effectiveness in capturing trends despite market volatility. Similarly, Djarot Hindarto et al. [37] in The Application of Neural Prophet Time Series in Predicting Rice Stock at Rice Stores applied the model to historical rice stock data, achieving a MAE of 12.90, RMSE of 15.80, and Loss of 0.0313. These metrics indicate reliable predictive accuracy for inventory management purposes. The results underscore the versatility of the Neural Prophet model in providing dependable forecasts not only for highly volatile financial data but also for more stable inventory time series.

In the research of analysis and comparison of methods for maximum temperature prediction in Jakarta, Indonesia has been done by Armando Jacquis F. Z., et al [38] using GRU and ANFIS methods. The results showed that both ANFIS and GRU can effectively forecast the maximum temperature with a correlation value above 0.95 and RMSE and MAPE below 2. GRU algorithm is more efficient for short-term forecasting, while the ANFIS model shows higher effectiveness in long-term forecasting. In the computation time of GRU is much shorter than that of ANFIS.

Munir et al. [39], proposed a hybrid model combining Discrete Wavelet Transform (DWT), ARIMA, and Artificial Neural Networks (ANN) for forecasting meteorological droughts using SPI and SIAP indices, yielding high accuracy with R values exceeding 0.91 and low RMSE scores. In contrast, a study [40] investigated the integration of ARIMA, SARIMAX, and Long Short-Term Memory (LSTM) models using a FUSION approach supported by IoT-based data collection. The findings indicated that individual models outperformed the integrated FUSION approach based on MAPE and MSE metrics, suggesting that stand-alone models may better capture domain-specific patterns. These results emphasize the critical importance of model selection and customization in developing accurate forecasting systems for environmental time series data.

Previous studies on temperature forecasting have commonly used ARIMA. While ARIMA handles linear patterns well, it lacks flexibility for non-linear and complex seasonal variations. Neural Prophet improves trend and seasonality modeling but can be sensitive to parameter tuning and less effective with irregular data. In this study, an optimized XGBoost model is proposed to forecast maximum temperature, using Grid Search for hyperparameter tuning. Compared to ARIMA and Neural Prophet, the optimized XGBoost demonstrates superior accuracy and better handling of non-linear relationships, highlighting its suitability for complex environmental time series.

3. Methodology

The system design in this study is shown in figure 1, with the stages consisting of four main stages, namely data acquisition, data preprocessing, data modeling, and model evaluation.

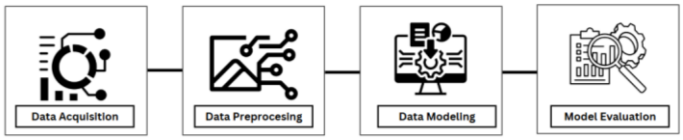

Figure 1. Flow of Key Stages in Research

This study uses daily maximum temperature data obtained from the BMKG Juanda weather station in Surabaya, covering the period from 1981 to 2022. The dataset comprises 15,388 entries with two variables—date and Tmax (°C)—sourced directly from the official records without any modification. As publicly available meteorological data, it does not involve human participants, sampling strategies, or inclusion and exclusion criteria. Once the data was collected, it underwent a preprocessing stage that included cleaning, transformation, and arrangement to ensure suitability for modeling. This process involved handling missing values, applying data normalization, and selecting relevant features. Following preprocessing, the modeling phase was carried out by developing and training three predictive models: XGBoost, ARIMA, and Neural Prophet. Each model was configured through parameter adjustments and validated to achieve optimal performance. The final stage involved evaluating the model's performance using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) as key metrics. The evaluation results served as

ISSN 2723-6471

the basis for selecting the model with the highest accuracy in forecasting maximum temperatures in Surabaya. The system implementation stage in this research is shown in figure 2.

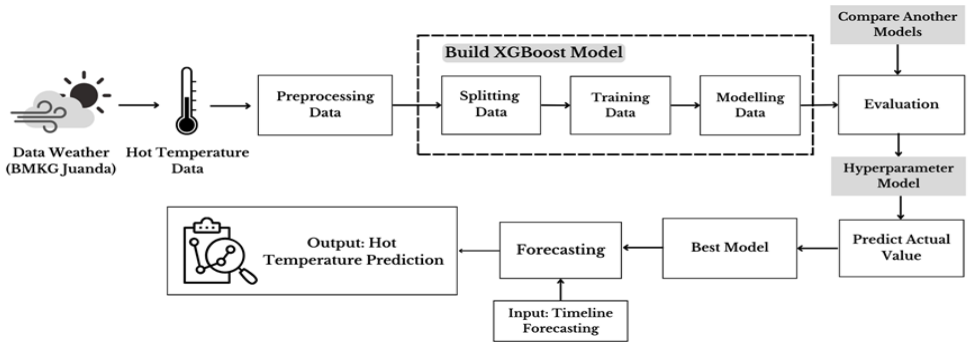

Figure 2. Model Building Implementation Flowchart

3.1. Data Acquisition

Data acquisition is a process stage used to collect and display data using a particular system [41]. In this research, data is collected from the BMKG Juanda weather dataset, which is meteorological data collected by the Meteorology, Climatology, and Geophysics Agency (BMKG) from the weather observation station in Juanda, Surabaya, East Java Province, Indonesia. This study utilizes publicly available meteorological data and does not involve human subjects or personal information. Therefore, ethical risks such as privacy or misdiagnosis are not applicable. This data includes various routinely measured weather parameters, such as air temperature, rainfall, air pressure, wind speed, and wind direction. The characteristics of the dataset used in this study, obtained from the BMKG weather observation station at Juanda, Surabaya, are presented in table 1.

Table 1. Juanda BMKG Dataset

Taverage	Tmax	Tmin	CH	QFF	ff average	most directions	dd	ff max
25.4	32.6	22.9	6.8	1009	2	B	B	9
26.6	32.0	23.4	3.6	1010	5	B	B	10
27.2	32.2	24	0.0	1009	7	B	B	12
24.7	28.6	22.9	43.0	1010	5	B	B	14
26.5	30.8	21.7	21.0	1009	8	B	B	14

3.2. Data Preprocessing

Data preprocessing is a very important initial stage in data analysis and modeling. The data preprocessing stage includes data preparation and data transformation [42]. The goal is to prepare the data so that it can be used effectively in subsequent processes, such as modeling and analysis. The dataset consists of daily maximum temperature (Tmax) and date values. As a univariate time series, only Tmax was used as the predictive feature, no additional features were engineered. Hence, feature correlation analysis was not applicable in this context. In the data preprocessing stage, several important steps are carried out, including importing datasets, feature selection, and handling missing values. The feature selection process on the dataset can improve the performance of the model. This process helps determine which variables have the most impact on the predictive model, so that only important variables are used. In addition, the feature selection stage or process can also improve the algorithm in processing data faster [43]. The data were then normalized using Min-Max scaling to ensure consistent input ranges across the models.

 ISSN 2723-6471

3.3. Algorithm Modeling

This research uses three models, namely ARIMA, XGBoost, and Neural Prophet. The process of these three time series models is shown in figure 3 below. The purpose of using the various models is to assess the performance of each and find the model that has optimal accuracy on the maximum temperature data used.

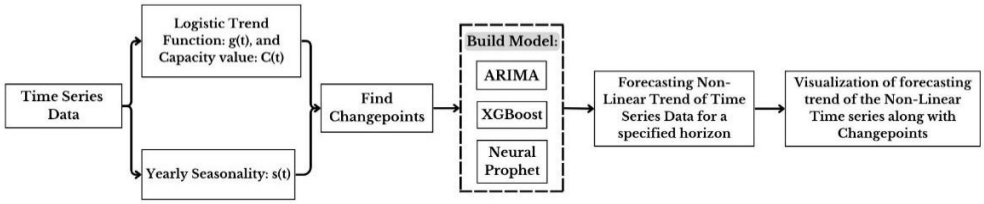

Figure 3. Modelling Stages

3.4. Extreme Gradient Boosting (XGBoost)

XGBoost is an advanced implementation of the gradient tree boosting method, designed to efficiently address various tasks such as regression, classification, and ranking. Its core principle lies in iteratively optimizing the learning process to minimize the loss function, thereby improving model accuracy. By constructing a more structured regression tree, XGBoost enhances performance while reducing model complexity to mitigate overfitting. Additionally, it incorporates several computational optimizations that accelerate training and further decrease the risk of overfitting [44]. The schematic representation of the XGBoost algorithm is illustrated in figure 4.

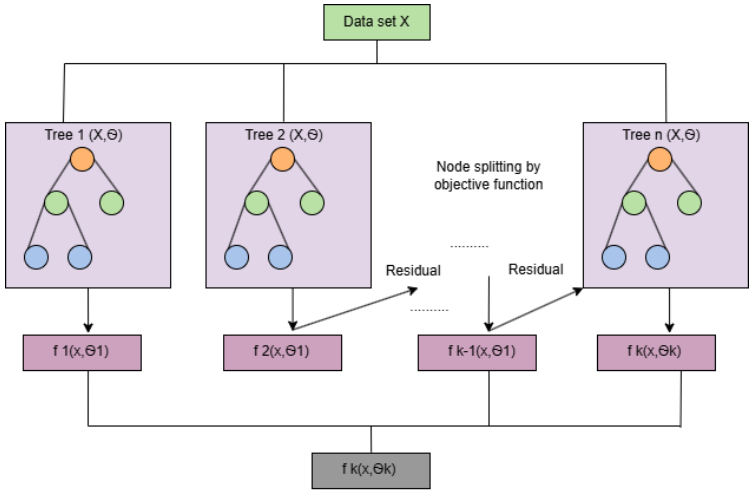

Figure 4. Schematic of XGBoost algorithm

In the XGBoost algorithm, the weights in each tree are updated incrementally. After that, all the tree weights are summed up when making predictions, then the results are entered into the function [45]. The function can be expressed as follows.

$$\hat{y}_i = \sum_{k=1}^{K} f_k\ (x_i) \tag{1}$$

Where, $\hat{y}i$ is the output of tree K; fk s the kth decision tree function-k; xi is the function trained; K is the tree created. The tree with summed weights aims to minimize the objective function. This objective function consists of two components, namely to measure the difference between the predicted value and the actual value, as well as the regularization term [46]. The objective function in XGBoost can be written as:

 ISSN 2723-6471

$$obj^{(t)} = \sum_{i=1}^{n} \left[g_i W_q (x_i) + \frac{1}{2} h_i w_q^2 (x_i) \right] + \gamma T + \frac{1}{2} \lambda \sum_{j=1}^{T} w_q^2 j \quad (2)$$

With, T is the number of leaf nodes; W is the tree nodes; γ to control the number of nodes; λ to control the number of trees. In this method, an objective function is required to assess the quality of the model generated based on the training data. This objective function consists of two important components, namely training loss value and regularization value, as shown in the following equation.

$$obj(\theta) = L(\theta) + \Omega(\theta) \quad (3)$$

L is the missing training function, and Ω is the regularization function, dan θ is the corresponding model parameter. The missing training function can be generally written as in the following equation.

$$L(\theta) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) \quad (4)$$

y_i is the true value of the data and $\hat{y}_i$ is the predicted value of the model, while n is the number of iterations of the model.

3.5. Model Testing Graph

The dataset was partitioned using a time-based split to preserve chronological order and prevent data leakage. An initial 70:30 train-validation split was used to assess the model's baseline performance, followed by an 80:20 train-test split for final evaluation to reflect realistic forecasting on unseen data. The XGBoost model was trained on these subsets, with hyperparameter tuning conducted via GridSearch to optimize predictive performance. Model evaluation was carried out using both numerical metrics and graphical analysis of the forecasting results. This process is essential to identify each model's strengths and weaknesses prior to hyperparameter tuning [47]. The primary evaluation metrics included RMSE, MAE, and MAPE, with RMSE serving as a widely used indicator of prediction error [48]. However, in this study, only RMSE and MAE were utilized to evaluate model performance, as both provide robust and interpretable measures of error in continuous temperature prediction. The RMSE formula is presented as follows [49].

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n-1} (y_i - \hat{y}_i)^2} \quad (5)$$

MAE calculates the average absolute value of the difference between the prediction and the actual value [48]. Here is the MAE formula.

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i| \quad (6)$$

3.6. Hyperparameter Tuning Model Equations

After an initial evaluation of the models used to predict maximum temperature, this study proceeds with model development through hyperparameter tuning. Hyperparameter tuning is a critical process in machine learning that aims to optimize model performance by finding the best combination of hyperparameters [50]. Hyperparameters are used to manage various aspects of machine learning that have a significant impact on performance and the resulting model [51].

3.7. Forecasting

The final stage of this study involves forecasting maximum temperature in Surabaya for the year 2025 using the best-performing model from previous evaluations. The prediction process uses the predict method to generate future values

(yhat) based on historical data, producing a forecast data frame that includes predicted values, components, and uncertainty intervals [52]. Forecasting is carried out using the most optimal model selected through comparative analysis and hyperparameter tuning. The model's accuracy was evaluated by comparing predicted values with actual observed temperatures to minimize error and validate its performance [53].

4. Results and Discussion

4.1. Predictive Model Testing

In this study, tests were conducted using several machine learning algorithms with model evaluation results with MAE and RMSE scores then compared to determine the best model [54]. The evaluation involved splitting the dataset into training and testing sets, followed by a comparative analysis of the algorithms. From the three models tested, the best-performing model was selected based on the evaluation metrics, as summarized in table 2.

Table 2. Testing Results of Selected Models

Model	Split Data Train	MAE	RMSE
ARIMA	80%	0.98	1.48
ARIMA	70%	0.98	1.40
Neural Prophet	80%	1.43	1.90
Neural Prophet	70%	1.43	1.91
XGBoost	80%	1.06	0.79
XGBoost	70%	0.92	0.70

The algorithm model evaluation results presented in table 2 demonstrate the performance of the three models used to predict high temperatures in Surabaya. Each model yields different test results due to variations in the proportion of training and test data. The ARIMA model performed quite well, with an MAE of 0.98 and an RMSE of 1.48 using 80% of the data for training, and an MAE of 0.98 and an RMSE of 1.40 using 70% of the data. These results indicate that the ARIMA model can predict heat with relatively high accuracy.

The XGBoost model also demonstrated good performance, even slightly outperforming the ARIMA model. Test results show that the highest prediction accuracy was achieved by the XGBoost model, with an MAE of 0.92 and an RMSE of 0.70 when using 70% of the data for training. These values are lower than those of the ARIMA model, indicating more accurate predictions. Meanwhile, the Neural Prophet model showed relatively lower performance compared to the other two models. With 80% training data, Neural Prophet recorded an MAE of 1.43 and an RMSE of 1.90, while with 70% training data, its MAE remained at 1.43 and RMSE slightly increased to 1.91. These values are higher than those of both the ARIMA and XGBoost models, indicating lower prediction accuracy.

The XGBoost model with 70% training data shows the best performance, with an RMSE value of only about 0.70. The ARIMA model with 70% and 80% training data also has a relatively low RMSE value, which is around 1.40-1.48. This means that the XGBoost and ARIMA models have better prediction accuracy than the Neural Prophet model. These results can be taken into consideration in choosing the best model to be used in predicting hot temperatures in Surabaya City.

The comparatively weaker performance of the Neural Prophet model may be attributed to its sensitivity to irregular or weak seasonal patterns within the dataset. Neural Prophet is optimized for structured time series with strong and periodic seasonality, which may not be dominant in Surabaya's maximum temperature data. Meanwhile, the ARIMA model, while effective in modeling linear relationships, has limited ability to capture complex nonlinear trends present in long-term climate datasets. These limitations result in lower accuracy compared to XGBoost, which is capable of learning intricate data patterns through ensemble-based tree structures and hyperparameter tuning.

4.2. Prediction Model Testing with Hyperparameter Tuning

After going through the parameter hypertuning process using 70% of the training data, the authors were able to obtain evaluation results that showed a significant improvement in the performance of the maximum temperature prediction model. The results of the model testing after the parameter hypertuning process can be seen in table 3. The evaluated models include ARIMA, Neural Prophet, and XGBoost.

Table 3. Parameter Hypertuning of each Model

Model	Boosting	Adventage
ARIMA	model = ARIMA(order=(p,d,q)), param_distributions={'order': [(p,d,q)], 'seasonal_order': [(P,D,Q,s)]}, n_iter=100, scoring='neg_mean_squared_error', cv=3, verbose=1, n_jobs=-1, random_state=42	Captures linear trends and seasonality well in time series data
Neural Prophet	model=NeuralProphet(), param_distributions={'seasonality_mode': ['additive', 'multiplicative'], 'learning_rate': [0.01, 0.1]}, n_iter=100, scoring='neg_mean_squared_error', cv=3, verbose=1, n_jobs=-1, random_state=42)	Effective for time series forecasting with irregular patterns and long-term trends
XGboost	estimator=XGBRegressor(random_state=42), param_distributions=param_distributions, n_iter=100, scoring='neg_mean_squared_error', cv=3, verbose=1, n_jobs=-1, random_state=42)	Interesting for big data application and selects optimal model parameters automatically

The results of model testing after the parameter hypertuning process can be seen in table 4. The models evaluated include ARIMA, Neural Prophet, and XGBoost.

Table 4. Model Testing Results with Parameter Hypertuning

Model	MAE	RMSE
ARIMA	0.85	1.2
Neural Prophet	0.70	0.98
XGboost	0.32	0.65

Based on table 4, it can be seen that each model has increased accuracy after hyperparameter tuning, which is indicated by a decrease in error values such as MAE and RMSE compared to the results before hyperparameter tuning contained in table 2. For the ARIMA model, MAE initially amounted to 0.98 down to 0.85, while RMSE decreased from 1.48 to 1.2 after hyperparameter tuning. The Neural Prophet model also showed improved performance with MAE dropping from 1.43 to 0.70 and RMSE from 1.90 to 0.98. Meanwhile, the XGBoost model experienced a significant decrease in error with MAE from 0.92 to 0.32 and RMSE from 0.70 to 0.65 after hyperparameter tuning.

These results show that the parameter optimization successfully improved the prediction accuracy of each model. Thus, the models become more capable of providing more accurate estimates of future maximum temperatures based on the training data used. From the graphs shown, it can be concluded that the XGBoost model provides the best prediction results, with lower RMSE and MAE values compared to the Neural Prophet and ARIMA models. Such parameter hypertuning process is important to optimize the performance of the prediction model, so that it can provide significant added value in practical applications that require temperature prediction with a high degree of accuracy.

4.3. Optimization of XGBoost Model Prediction Model Testing

Hyperparameter tuning is performed to maximize the prediction of the XGBoost model. GridSearchCV hyperparameters systematically search for the best combination of hyperparameters, such as learning rate, max depth, min_child_weight, subsample, and n_estimators that greatly affect prediction accuracy. This process uses cross-validation to evaluate each hyperparameter combination on multiple subsets of data, resulting in a more stable and reliable evaluation compared to methods that only use one set of training and testing data. The following are the results of parameter experiments on GridSearchCV hyperparameters.

 ISSN 2723-6471

The results shown in the table are from a series of tests of different XGBoost model parameters. The parameters tested include n_estimators, max_depth, learning_rate, subsample, and colsample bytree. Each combination of these parameters was tested to determine its effect on the model's performance in predicting weather parameters. Table 5 shows the results of three different experiments, with the evaluation metrics being RMSE and MAE. In Experiment 1, the RMSE value is 1.12 and the MAE is 0.67. In Experiment 2, the RMSE value is 1.05 and MAE is 0.78. While in Experiment 3, the best RMSE and MAE values were obtained, namely 0.65 and 0.32. These results demonstrate the variation in model performance across different parameter combinations and provide useful insights for determining the optimal configuration for weather parameter prediction. The detailed outcomes of the parameter tuning process using GridSearchCV are presented in table 5.

Table 5. GridSearhCV Experiment Table

	RMSE	MAE
Experiment 1	1.12	0.67
Experiment 2	1.05	0.78
Experiment 3	0.65	0.32

Based on the results in table 5, it can be concluded that the three experiments produced different results. The threshold parameters used are 0.5 and 0.8, showing that the higher the RMSE and MAE values, the lower the accuracy. If the values are below the green threshold, then the accuracy can be categorized as excellent. It can be concluded that the third hyperparameter tuning experiment shows the most optimal accuracy in the prediction model, with an RMSE value of 0.65 and MAE of 0.32. The visualization of prediction graph of Maximum Heat Temperature in 2025 is shown in figure 5.

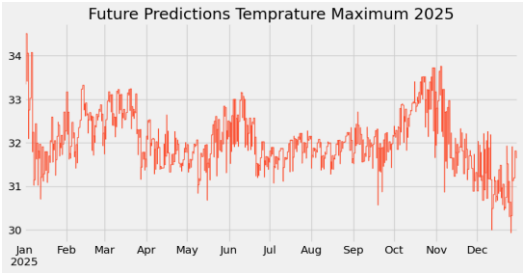

Figure 5. Prediction Graph of Maximum Heat Temperature in 2025

The heat trend graph in figure 5 above is the result of predicting the heat temperature in 2025 every month for 1 year. With a model that has very low RMSE and MAE values, the results of the trend are quite accurate and can be tested for validation. in the trend graph, the maximum heat prediction occurs in the January, October to November period.

5. Discussion

The visualization of temperature prediction with XGBoost is shown in figure 6.

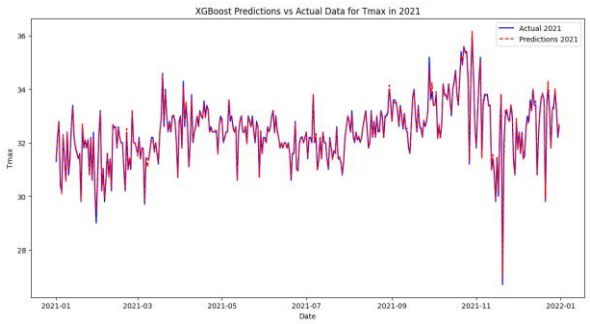

Figure 6. Temperature Prediction with XGBoost

This study makes a significant contribution to weather prediction and machine learning in Surabaya. After hyperparameter tuning, the XGBoost model outperformed ARIMA and Neural Prophet in forecasting maximum temperature, and the results closely align with actual 2021 data, as illustrated in figure 6. This demonstrates its effectiveness as a reliable model for weather forecasting, particularly for Surabaya, a city increasingly affected by global warming. Accurate temperature predictions support the city in multiple sectors, providing essential guidance for sustainable development and risk management. In the area of city planning, such forecasts aid in designing green spaces and heat-resilient infrastructure to mitigate the effects of extreme temperatures [9]. For resource management, they help optimize the use of water and energy, particularly during periods of peak demand. In the public health sector, accurate predictions enable early responses to heat-related illnesses through preventive measures and public awareness campaigns [55]. In agriculture, they assist farmers in improving planting and harvesting schedules, thereby increasing productivity and reducing losses [56]. Furthermore, reliable temperature forecasts play a crucial role in disaster risk mitigation by enhancing preparedness and response to extreme events such as heatwaves.

6. Declarations

6.1. Author Contributions

Conceptualization: S., I.G.S.M.D., D.S.A., R.L.A., A.R.F.S., H.S., D.A.D.; Methodology: I.G.S.M.D.; Software: S.; Validation: S., I.G.S.M.D., and D.A.D.; Formal Analysis: S., I.G.S.M.D., and D.A.D.; Investigation: S.; Resources: I.G.S.M.D.; Data Curation: I.G.S.M.D.; Writing Original Draft Preparation: S., I.G.S.M.D., and D.A.D.; Writing Review and Editing: I.G.S.M.D., S., and D.A.D.; Visualization: S. All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] S. Fischer., "Emerging Effects of Temperature on Human Cognition, Affect, and Behaviour," *Biol. Psychol.*, vol. 189, no. 1, pp. 1-21, May 2024, doi: 10.1016/j.biopsycho.2024.108791.

[2] G. Sullivan and M. Spencer, "Heat and temperature," *BJA Educ.*, vol. 22, no. 9, pp. 350–356, Sep. 2022, doi: 10.1016/j.bjae.2022.06.002.

[3] C. Song, H. Ikei, and Y. Miyazaki, "Physiological Effects of Forest-Related Visual, Olfactory, and Combined Stimuli on Humans: An Additive Combined Effect," *Urban For. Urban Green.*, vol. 44, no. 1, pp. 12-37, Aug. 2019, doi: 10.1016/j.ufug.2019.126437.

[4] D. S. Bari, M. N. S. Rammoo, H. Y. Y. Aldosky, M. K. Jaqsi, and Ø. G. Martinsen, "The Five Basic Human Senses Evoke Electrodermal Activity," *Sensors*, vol. 23, no. 19, pp. 8181-8192, Sep. 2023, doi: 10.3390/s23198181.

[5] G. Zheng, K. Li, and Y. Wang, "The Effects of High-Temperature Weather on Human Sleep Quality and Appetite," *Int. J. Environ. Res. Public. Health*, vol. 16, no. 2, pp. 270-282, Jan. 2019, doi: 10.3390/ijerph16020270

[6] F. Barbosa Escobar, C. Velasco, K. Motoki, D. V. Byrne, and Q. J. Wang, "The Temperature of Emotions," *PLOS ONE*, vol. 16, no. 6, pp. 1-21, Jun. 2021, doi: 10.1371/journal.pone.0252408.

[7] A. Albatayneh, D. Alterman, A. Page, and B. Moghtaderi, "The Impact of the Thermal Comfort Models on the Prediction of Building Energy Consumption," *Sustainability*, vol. 10, no. 10, pp. 3609-3618, Oct. 2018, doi: 10.3390/su10103609.

[8] A. Ospino, C. Robles, and I. Tovar, "Strategies to Improve Thermal Comfort and Energy Efficiency of Social Interest Housing in Tropical Areas," *J. Sustain. Dev. Energy Water Environ. Syst.*, vol. 12, no. 4, pp. 1–18, Dec. 2024, doi: 10.13044/j.sdewes.d12.0526.

[9] S. Bolan *et al.*, "Impacts of Climate Change on the Fate of Contaminants Through Extreme Weather Events," *Sci. Total Environ.*, vol. 909, no. 1, pp. 1-18, Jan. 2024, doi: 10.1016/j.scitotenv.2023.168388.

[10] A. Pandey, "Global Warming and Increase of Global Temperature: Model Based Study," *J. Nonlinear Anal. Optim.*, vol. 15, no. 4, pp. 117–127, 2024, doi: 10.2139/ssrn.4880990.

[11] L. Wang, L. Wang, Y. Li, and J. Wang, "A Century-Long Analysis of Global Warming and Earth Temperature Using a Random Walk with Drift Approach," *Decis. Anal. J.*, vol. 7, no. 1, pp. 1-17, Jun. 2023, doi: 10.1016/j.dajour.2023.100237.

[12] K. Furtak and A. Wolińska, "The Impact of Extreme Weather Events as a Consequence of Climate Change on the Soil Moisture and on the Quality of the Soil Environment and Agriculture – a Review," *CATENA*, vol. 231, no. 1, pp. 1-18, Oct. 2023, doi: 10.1016/j.catena.2023.107378.

[13] K. Abbass, M. Z. Qasim, H. Song, M. Murshed, H. Mahmood, and I. Younis, "A Review of the Global Climate Change Impacts, Adaptation, and Sustainable Mitigation Measures," *Environ. Sci. Pollut. Res.*, vol. 29, no. 28, pp. 42539–42559, Jun. 2022, doi: 10.1007/s11356-022-19718-6.

[14] J. Ekawati, H. Sofari, W. Rahmawati, S. I. Permata, and E. Setiawan, "Mitigating Climate Change Towards Livable City (Case: Bandung City, West Java)," *J. Archit. Des. Urban.*, vol. 6, no. 1, pp. 36–50, Jan. 2024, doi: 10.14710/jadu.v6i1.21612.

[15] B. S. S. Wibawa *et al.*, "Effects of Ambient Temperature, Relative Humidity, and Precipitation on Diarrhea Incidence in Surabaya," *Int. J. Environ. Res. Public. Health*, vol. 20, no. 3, pp. 2313-2325, Jan. 2023, doi: 10.3390/ijerph20032313.

[16] A. M. Dary, M. A. Mardyanto, J. Hermana, and C. Imron, "Climate Change and Its Effect on Temperature and Precipitation Trends: Case Study in Surabaya Using RegCM5," *Int. J. Comput. Sci. Appl. Math.*, vol. 11, no. 1, pp. 33–37, 2025

[17] Y. D. Anggraini, S. J. Sukmana, D. Sulistyowaty, F. Rohmiah, and H. Rasidi, "The Urban Planning of Surabaya City with The Aim of Creating a Green City," *J. Civ. Eng. Plan. Des.*, vol. 3, no. 1, pp. 1–5, Jul. 2024, doi: 10.31284/j.jcepd.2024.v3i1.5277.

[18] A. A. Assayuti, N. Ani, Y. Pujowati, A. T. Abeng, and D. M. Kamal, "Impact of Air Pollution, Population Density, Land Use, and Transportation on Public Health in Jakarta," *J. Geosains West Sci.*, vol. 1, no. 02, pp. 35–43, Jun. 2023, doi: 10.58812/jgws.v1i02.391.

[19] M. L. Williams, "Global Warming, Heat-Related Illnesses, and the Dermatologist," *Int. J. Womens Dermatol.*, vol. 7, no. 1, pp. 70–84, Jan. 2021, doi: 10.1016/j.ijwd.2020.08.007.

[20] T. Mifsud, C. Modestini, A. Mizzi, O. Falzon, K. Cassar, and S. Mizzi, "The Effects of Skin Temperature Changes on the Integrity of Skin Tissue: A Systematic Review," *Adv. Skin Wound Care*, vol. 35, no. 10, pp. 555–565, Oct. 2022, doi: 10.1097/01.ASW.0000833612.84272.da.

[21] K. J. J. K. Jetly *et al.*, "Risk Factors For Scabies In School Children: A Systematic Review," *Vopr. Prakt. Pediatr.*, vol. 17, no. 2, pp. 117–125, 2022, doi: 10.20953/1817-7646-2022-2-117-125.

[22] M. Amoadu, E. W. Ansah, J. O. Sarfo, and T. Hormenu, "Impact of Climate Change and Heat Stress on Workers' Health and Productivity: A Scoping Review," *J. Clim. Change Health*, vol. 12, no. 1, pp. 1-19, Jul. 2023, doi: 10.1016/j.joclim.2023.100249.

[23] F. S. Arsad ,"The Impact of Heatwaves on Mortality and Morbidity and the Associated Vulnerability Factors: A Systematic Review," *Int. J. Environ. Res. Public. Health*, vol. 19, no. 23, pp. 1-16, Dec. 2022, doi: 10.3390/ijerph192316356.

[24] B. Yin, W. Fang, L. Liu, Y. Guo, X. Ma, and Q. Di, "Effect of Extreme High Temperature on Cognitive Function at Different Time Scales: A National Difference-in-Differences Analysis," *Ecotoxicol. Environ. Saf.*, vol. 275, no. 1, pp. 1-18, Apr. 2024, doi: 10.1016/j.ecoenv.2024.116238.

[25] S. D. S. Pramesti, H. M. Denny, and Y. Setyaningsih, "The Relationship Between Fluid Intake and Heat Sress with the Hydration Status of Workers: A Scooping Review," *Indones. J. Glob. Health Res.*, vol. 6, no. 5, pp. 2737–2746, 2024

[26] S. Shrestha, J. Mahat, J. Shrestha, M. K.C., and K. Paudel, "Influence of High-Temperature Stress on Rice Growth and Development: A Review," *Heliyon*, vol. 8, no. 12, p. e12651, Dec. 2022, doi: 10.1016/j.heliyon.2022.e12651.

[27] M. Li, H. Y. Wang, A. A. K. Najm, B. A. Othman, and D. Law, "Effects of Molybdenum on Growth and Fruit Quality of Small Fruit Melon (cucumis Melo L.) Cultivated Under High-Temperature Stress," *Acta Sci. Pol. Hortorum Cultus*, vol. 23, no. 4, pp. 41–54, Sep. 2024, doi: 10.24326/asphc.2024.5345.

[28] I. Manisalidis, E. Stavropoulou, A. Stavropoulos, and E. Bezirtzoglou, "Environmental and Health Impacts of Air Pollution: A Review," *Front. Public Health*, vol. 8, p. 14, Feb. 2020, doi: 10.3389/fpubh.2020.00014.

[29] J. D. Périard, T. M. H. Eijsvogels, and H. A. M. Daanen, "Exercise Under Heat Stress: Thermoregulation, Hydration, Performance Implications, and Mitigation Strategies," *Physiol. Rev.*, vol. 101, no. 4, pp. 1873–1979, Oct. 2021, doi: 10.1152/physrev.00038.2020.

[30] M. Özdamar and F. Umaroğullari, "Thermal Comfort and Indoor Air Quality," *Int. J. Sci. Res. Innov. Technol.*, vol. 5, no. 3, pp. 90–109, 2018.

[31] M. Levi, T. Kjellstrom, and A. Baldasseroni, "Impact of Climate Change on Occupational Health and Productivity: A Systematic Literature Review Focusing on Workplace Heat," *Med. Lav.*, vol. 109, no. 3, pp. 163–179, Apr. 2018, doi: 10.23749/mdl.v109i3.6851.

[32] P. Shafigh, M. A. Hafez, Z. Che Muda, S. Beddu, A. Zakaria, and Z. Almkahal, "Influence of Different Ambient Temperatures on the Thermal Properties of Fiber-Reinforced Structural Lightweight Aggregate Concrete," *Buildings*, vol. 12, no. 6, p. 771, Jun. 2022, doi: 10.3390/buildings12060771.

[33] P. Wargocki, J. A. Porras-Salazar, and S. Contreras-Espinoza, "The Relationship Between Classroom Temperature and Children's Performance in School," *Build. Environ.*, vol. 157, no. 1, pp. 197–204, 2019.

[34] L. Zhang, W. Bian, W. Qu, L. Tuo, and Y. Wang, "Time Series Forecast of Sales Volume Based on Xgboost," *J. Phys. Conf. Ser.*, vol. 1873, no. 1, pp. 1-12, Apr. 2021, doi: 10.1088/1742-6596/1873/1/012067.

[35] Y. Lai and D. A. Dzombak, "Use of the Autoregressive Integrated Moving Average (ARIMA) Model to Forecast Near-Term Regional Temperature and Precipitation," *Weather Forecast.*, vol. 35, no. 3, pp. 959–976, Jun. 2020, doi: 10.1175/WAF-D-19-0158.1.

[36] T. R. Noviandy, "Deep Learning-Based Bitcoin Price Forecasting Using Neural Prophet," *Ekon. J. Econ.*, vol. 1, no. 1, pp. 19–25, Jul. 2023, doi: 10.60084/eje.v1i1.51.

[37] D. Hindarto, F. Hendrata, and M. Hariadi, "The Application of Neural Prophet Time Series in Predicting Rice Stock at Rice Stores," *J. Comput. Netw. Archit. High Perform. Comput.*, vol. 5, no. 2, pp. 668–681, Aug. 2023, doi: 10.47709/cnahpc.v5i2.2725.

[38] A. J. F. Zamelina, D. Adytia, and A. W. Ramadhan, "Forecasting of Maximum Temperature by using ANFIS and GRU Algorithms: Case Study in Jakarta, Indonesia," *3rd Int. Conf. Intell. Cybern. Technol. Appl. ICICyTA*, vol. 2023, no. 1, pp. 49–54, 2023, doi: 10.1109/ICoICT55009.2022.9914885.

[39] Md. M. H. Khan, N. S. Muhammad, and A. El-Shafie, "Wavelet Based Hybrid ANN-ARIMA Models for Meteorological Drought Forecasting," *J. Hydrol.*, vol. 590, no. 1, pp. 1-20, Nov. 2020, doi: 10.1016/j.jhydrol.2020.125380.

[40] S. S. Kumar, A. Kumar, S. Agarwal, M. Syafrullah, and K. Adiyarta, "Forecasting Indoor Temperature for Smart Buildings with ARIMA, SARIMAX, and LSTM: A Fusion Approach," *9th Int. Conf. Electr. Eng. Comput. Sci. Inform. EECSI*, vol. 2022, no. 1, pp. 186–192, 2022, doi: 10.23919/EECSI56542.2022.9946498.

[41] D. A. Devi, T. Satya, and S. Sugun.L, "Design and Implementation of Real Time Data Acquisition System using Reconfigurable SoC," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 9, pp. 1-12, 2020, doi: 10.14569/IJACSA.2020.0110938.

[42] S. A. Alasadi and W. S. Bhaya, "Review of Data Preprocessing Techniques in Data Mining," *J. Eng. Appl. Sci.*, vol. 12, no. 16, pp. 4102–4107, 2017

[43] N. Pudjihartono, T. Fadason, A. W. Kempa-Liehr, and J. M. O'Sullivan, "A Review of Feature Selection Methods for Machine Learning-Based Disease Risk Prediction," *Front. Bioinforma.*, vol. 2, no. 1, pp. 1-20, Jun. 2022, doi: 10.3389/fbinf.2022.927312.

[44] Z. Arif Ali, Z. H. Abduljabbar, H. A. Tahir, A. Bibo Sallow, and S. M. Almufti, "eXtreme Gradient Boosting Algorithm with Machine Learning: a Review," *Acad. J. Nawroz Univ.*, vol. 12, no. 2, pp. 320–334, May 2023, doi: 10.25007/ajnu.v12n2a1612.

[45] Y. Zou, M. Xia, and X. Lan, "Interpretable Credit Scoring Based on an Additive Extreme Gradient Boosting," *Chaos Solitons Fractals*, vol. 194, no. 116216, pp. 1-12, 2025

[46] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco California USA: ACM, Aug. vol. 2016, no. 1, pp. 785–794, 2016. doi: 10.1145/2939672.2939785.

[47] B. Azari, K. Hassan, J. Pierce, and S. Ebrahimi, "Evaluation of Machine Learning Methods Application in Temperature Prediction," *Comput. Res. Prog. Appl. Sci. Eng.*, vol. 8, no. 1, pp. 1–12, 2022, doi: 10.52547/crpase.8.1.2747.

[48] N. W. Azizah, E. Y. Puspaningrum, and I. G. S. S. Mas Diyasa, "Analyzing the Relationship Between Meteorological Parameters and Electric Energy Consumption Using Support Vector Machine and Cooling Degree Days Algorithm," *J. Inf. Syst. Inform.*, vol. 6, no. 2, pp. 729–750, Jun. 2024, doi: 10.51519/journalisi.v6i2.719.

[49] T. O. Hodson, "Root-Mean-Square Error (RMSE) or Mean Absolute Error (MAE): When to Use Them or Not," *Geosci. Model Dev.*, vol. 15, no. 14, pp. 5481–5487, Jul. 2022, doi: 10.5194/gmd-15-5481-2022.

[50] W. Nugraha and A. Sasongko, "Hyperparameter Tuning on Classification Algorithm with Grid Search," *SISTEMASI*, vol. 11, no. 2, pp. 391-409, May 2022, doi: 10.32520/stmsi.v11i2.1750.

[51] C. M. Liyew, E. Di Nardo, S. Ferraris, and R. Meo, "Hyperparameter Optimization of Machine Learning Models for Predicting Actual Evapotranspiration," *Mach. Learn. Appl.*, vol. 20, p. 100661, Jun. 2025, doi: 10.1016/j.mlwa.2025.100661.

[52] P. Yuvarani, P. Bharani, B. Dharun, and P. Dinesh, "Time Series Forecasting of Ethereum Price by FB-Prophet," *2023 4th Int. Conf. Signal Process. Commun. ICSPC*, pp. 272–277, 2023, doi: 10.1109/ICSPC57692.2023.10125661.

[53] R. Widiono, B. Z. Cirgon, S. N. Aji, Subhiyanto, and A.Riyadi, "Prediction of Students Passedon Time with Classification Technique Using Have Naïve Bayes Algorithm," *Int. J. Comput. Tech.*, vol. 6, no. 6, pp. 1–7, 2019.

[54] R. M. X. Wu., "Comparative Study of Ten Machine Learning Algorithms for Short-Term Forecasting in Gas Warning Systems," *Sci. Rep.*, vol. 14, no. 1, pp. 21-39, Sep. 2024, doi: 10.1038/s41598-024-67283-4.

[55] T. B. W. Hartono, "Physiological Responses of Workers' Vital Signs in High Temperature Environments at The Tofu Home Industry Kedung Tarukan Surabaya," *J. Kesehat. Lingkung.*, vol. 11, no. 3, pp. 242-256, Jul. 2019, doi: 10.20473/jkl.v11i3.2019.242-251.

[56] S. Mishra, K. Spaccarotella, J. Gido, I. Samanta, and G. Chowdhary, "Effects of Heat Stress on Plant-Nutrient Relations: An Update on Nutrient Uptake, Transport, and Assimilation," *Int. J. Mol. Sci.*, vol. 24, no. 21, pp. 1-20, Oct. 2023, doi: 10.3390/ijms242115670.