SHM PUBLISHER
SURYA HIJAU MANFAAT

Journal of Information System Exploration and Research

https://joiser.shmpublisher.com

p-ISSN 2964-1160 | e-ISSN 2963-6361

Temperature, Humidity, and Weather Prediction Using Random Forest and LSTM for Food Crop Cultivation Optimization in Tegal

Sarwo Edi[1*], Aji Supriyanto[2],
[1,2]Department of Informatioan Technology, Universitas Stikubank, Indonesia

DOI: https://doi.org/10.52465/joiser.v4i3.10
Received 28 May 2026; Accepted 14 July 2026; Available online 06 August 2026

Article Info

Keywords:
Temperature Humidity Index;
Classification;
Random Forest;
LSTM;
Food crops;

Abstract

The agricultural sector in the Tegal region faces uncertain climate fluctuations that directly impact food crop productivity. A crucial indicator for determining plant environmental comfort is the Temperature Humidity Index (THI). This research aims to develop a hybrid model capable of predicting and classifying future THI values to support precision decision-making for farmers. The methodology utilized historical climate data including Temperature humidity, rainfall, sunshine, and wind speed from the Tegal Maritime Meteorology Station spanning a 10-year period (2016-2025). A Long Short-Term Memory (LSTM) model was applied to forecast future THI values, while a Random Forest (RF) model was utilized to classify plant stress categories. Model performance was evaluated using Root Mean Square Error (RMSE), Accuracy, and F1-score. The results indicate that Tegal experiences comfortable ($24 \le THI < 27$), moderately comfortable ($27 \le THI < 30$), and uncomfortable ($THI \ge 30$) conditions, particularly during dry and transitional seasons. The LSTM model achieved a 98.42% prediction accuracy, and the RF model reached a 99.59% classification accuracy. In conclusion, this highly accurate model can serve as an early warning system, providing farmers with actionable recommendations for optimal planting schedules and stress mitigation.

1. Introduction

According to website https://bappeda.tegalkab.go.id (2021) [1] and supported study from Tegal Regency is one of the center production agriculture in Central Java with commodities main like rice, corn and soybeans, some of which big located Tegal part north. However, this region own vulnerability tall to variability climate, such as improvement temperature air, anomaly rainfall rain, and humidity extreme. Phenomenon the impact on drought to in the danger zone tall to very high for phenomenon

* ***Corresponding Author:***

Sarwo Edi,
Department of Informatioan Technology,
Universitas Stikubank,
Semarang, Central Java, Indonesia.
Email: sarwoedi0002@mhs.unisbank.ac.id

Drought Hydrology with the area covered reaches 13,154 hectares, changes microclimate this in a way direct influence THI value and condition agroecosystem local. In context change global climate, capabilities for predict and optimize condition microclimate become need strategic for resilience food still awake.[2]

The THI serves as an important measure of environmental thermal comfort, particularly vital for living things in densely populated areas. The THI is widely used to assess thermal comfort by utilizing temperature and humidity variability, analyzed based on observational data [3]. Growth sector agriculture in Indonesia is greatly influenced by conditions dynamic climate and weather. One of the important parameters in determine comfort environment grow plant is the THI, which represents combination between temperature air and humidity relative [4].

THI does not only functioning as indicator comfort thermal in humans and animals cattle [3] but also start lots used in analysis microclimate agriculture for evaluate level stress environment to plant food. THI values that are too high tall or too low can cause stress physiological in plants, inhibiting photosynthesis, as well as lower productivity land [5].

Security global food today This affected by several factors, such as condition pandemic, issues geopolitics, prolonged conflict, and change climate [6]. Related with change climate, then Climate Smart Agriculture (CSA) is a strategy that is needed for overcome mutual challenges related between resilience food and change climate, and in effort This needed biophysical and socioeconomic data analysis [7]. Plants food, it is hoped can resilient to challenge change frequent weather extreme and threatening success harvest [8].

As it develops technology, approach machine learning has lots used in field agriculture precision (precision agriculture). Algorithm such as machine learning and deep learning make it possible meteorological and climatological data analysis in amount big for produce more predictions accurate compared to method conventional [9]. Implementation in THI's prediction will be help farmers and stakeholders policy understand pattern change microclimate in a way more fast, adaptive, and data-driven. With thus, cultivation strategies, arrangements irrigation, as well as election time plant can optimized for minimize risk failure harvest [10]

Based on background behind said, research regarding "Prediction and Classification of Temperature Humidity Index with Random Forest and LSTM for Optimization Cultivation Food Crops in Tegal" is becoming very important. This research expected can give contribution scientific in development of predictive models at a time give benefit practical in the form of recommendation adaptive, precise and supportive agroclimate resilience food sustainable at the level local and national some models used in study this is Random Forest which is combining machine learning algorithms a number of decision tree for produce one results end [11]. Approach learning machine latest such as transfer of learning and learning strengthening methodology this promising progress significant predictions into a new era innovation and discovery. research existing benchmarks show matter. This lack like lack of generalization, highly data-driven, context and feature-driven external depending on the available data. [12].

For produce one results end according to [13] approach learning machine latest such as transfer of learning and learning strengthening, methodology this promising progress prediction rainfall rain significant into a new era innovation and discovery. research existing benchmarks show matter this lack like lack of generalization, highly data-driven, context-and feature-driven external depending on the available data [14]. This study propose federation use approach based learning in finish problem the related shortcomings in existing research [15]. Previously [16] state availability ability intelligence model forecasting artificial as used in study. This will enabling a brighter future reliable and accurate in manufacturing prediction rainfall Rain during planting season.

2. Literature Review

Previous study carried out by [3] use technique machine learning for projecting and classifying level Comfort Thermal (measured with THI) in coastal areas north of Central Java, with integration of machine learning models (Random Forest and XGBoost) for analysis climate and improvement accuracy use of Random Forest for classification (level accuracy of 99.94%) shows that ML can very accurately categorize THI levels based on temperature and humidity data historical, making it reliable tool ffor monitoring real-time although projection model accuracy XGBoost (73%) is a bit more low from classification, numbers the still considered tall for prediction climate term intermediate (5 years), shows that this model capable catch dynamics trend climate in coastal areas north western part of Central

Java (Semarang City, Kendal Regency), East Java (Pemalang, and Tegal) reported often experience slightly uncomfortable condition (THI 27–30), especially during season rain and seasons transition

According to [17], thermal comfort in the northern coastal area of western Central Java reveals a long-term warming trend, where the THI increased from 26.13 in 2018 to 27.38 in 2024 along with an increase in average temperature from 27.0°C to 28.4°C. The results of feature importance analysis using the Random Forest model show that air temperature is the most dominant factor influencing thermal comfort with a contribution of 79.4%, while air humidity only has an effect of 20.6%. Spatially, dense urban areas experiencing rapid industrialization such as Semarang City, Eastern Kendal, Pemalang, and Tegal more often experience somewhat uncomfortable conditions (THI 27–30) especially during the rainy season and transitional seasons due to the influence of the urban heat island (UHI) phenomenon and high relative humidity. On the other hand, areas such as Batang and western Kendal tend to remain comfortable (THI < 27) according to the XGBoost model projection for the 2025–2029 period, the THI index is expected to continue to experience a stable upward trend until it reaches a peak of 27.14 at the end of 2029.

Study comprehensive carried out by [18] analyze in a way deep impact adoption practices agronomy For intensification sustainable rice-wheat crops throughout Indo-Gangetic Plains part east with show role crucial practice the in increase resilience regional food, strengthening adaptability economy farmer through improvement efficiency source power, and contribute significant to effort mitigation environment through subtraction greenhouse gas emissions glass and upgrades health land so that provide framework solid work for sustainability system agriculture in the South Asian region.

Study according to [19], this map pattern plants in the area Akarsu Irrigation District (AID) irrigation during the rainy season cold 2021 using algorithm Artificial Neural Network classification model reach level excellent average accuracy, which is above 90% for part big commodities. Based on matrix confusion matrix, accuracy for plant lettuce and potatoes reached 100% curation for plant wheat is at 89.76% because system detect existence overlapping overlap (confused) with plantation oranges (3.41%) and land empty/open (6.83%)

Besides that study applied by [20] serve methodology for determine the potential area of Green Open Space (RTH) in Bandar Lampung City using sensing data integration far, especially through analysis Index Vegetation Normalization Differential (NDVI) and THI, which is effective identify urban areas with density vegetation low and index discomfort high thermal, findings this offer guide accurate spatial for government city for prioritize location new green open space interventions to mitigate effect Urban Heat Island and in general direct increase quality environment as well as comfort thermal public.

3. Method

Figure 1. Stage flowchart study

Optimization cultivation plant food like improvement results harvest in Tegal Regency and City as impact from accurate THI predictions with use data historical average daily during 10 Year from 2016 to 2025 at BMKG Tegal, where data processing using the Long-Short Term Memory model for prediction and Random Forest model for classify THI value, in study this, implemented through three phase main input, process, and output, each of which consists of from a series step systematic and structured. Stages study seen in Figure 1.

3.1. Data processing

This stage will be discuss data processing.

3.1.1. Input

At the stage this, focus main is ensure incoming data to in quality models tall in the proper format, the data source consists of from secondary data from BMKG Station Tegal Meteorology with 10 year span (2016-2025) with data variables in the form of (temperature, humidity, rainfall, irradiation, sun and wind speed) THI research for comfort thermal area of the North coast of Tegal Regency and City, the input diagram flow begins with studies literature as step beginning that includes analysis state-of-the-art because involving collection and processing of weather data that will used for calculate THI, which is then become base in building a predictive model comfort thermal.

3.1.2. Process

In the preprocessing stage, step the first thing to do is data exploration, which aims for understand the structure and characteristics of the collected temperature and humidity data were analyzed. This exploration included examining data types, data distribution, and identifying patterns or anomalies within the data. Descriptive statistics were then presented to illustrate basic statistical summaries, such as the mean, standard deviation, and minimum and maximum values of temperature and humidity, which will help understand climate patterns in the northern coastal area of Tegal Regency.

The next step is to check for missing values, which is to determine if any data is missing. Missing data can affect the analysis results and THI calculations, so it is important to handle missing values, either by deleting rows with missing values or replacing them with estimated values such as the mean or interpolation. After that, data cleaning is performed, which includes removing blank values (NaN) and removing any possible duplicate rows. For the RF classification model, this is used. precision, recall, F1-score, and support metrics, while for the prediction model, MAE, RMSE, MAPE, and R^2 are used. measure error and performance of prediction models.

3.1.3. Output

First step in the output process is results prediction comfort thermal obtained from the model that has been built after get results predictions, done model evaluation for measure quality and accuracy The resulting predictions. At this stage, several evaluation metrics such as MAE, RMSE, MAPE, and R^2 are used to assess model performance. RMSE measures the extent to which the model predicts actual data, MAE provides information about the average prediction error, and R^2 indicates how well the model explains data variation, which is then visualized in graphical form to facilitate interpretation. This visualization forms the basis for understanding the pattern and spatial distribution of subsequent prediction results. In the model evaluation stage, the performance of the model that has been used for prediction is assessed using evaluation metrics such as MAE, RMSE, MAPE, and R^2. This evaluation aims to measure how well the model can represent actual data and ensure the accuracy of the prediction results before further use in the analysis.

The output produced then provides information value that can be directly used by the agricultural sector, such as value prediction, namely the estimation of future temperature and humidity figures and Classification of Status of Comfort Level category. comfort thermal shared become three main zones [21] consists of comfortable conditions (THI 21–24) namely conditions which are optimal for the growth of a plant, conditions which are less comfortable (THI 25–27) that is conditions where plants start to experience mild stress, this can be handled by monitoring irrigation, uncomfortable conditions (THI>28) namely conditions where there is a high risk of crop failure, emergency action is needed (extra watering/shading).

3.2. THI

THI initially developed by Thom (1959) for comfort human, index this has adapted in a way wide in field agroclimatology For measure to what extent the condition environment support or hinder optimal plant growth [22]. For plants, THI is not just number temperature, but rather representation from ability air for absorb water vapor from surface leaves (transpiration). The general THI equation used in studies agriculture tropical is shown in (1).

$$THI = T - (0.55 - 0.0055\, x\, RH)\, x\, (T - 14.5) \quad (1)$$

From the equation above is, T for temperature air (°C), RH is humidity relatively (%), ($0.55 - 0.0055xRH$) This is the Humidity Correction Factor. The number 0.55 indicates proportion humidity to temperature. The number 0.0055 works For normalize mark humidity (percentage) to be in harmony with unit temperature. increasingly high RH, then subtraction temperature (cooling) will the more small, so that final THI value will the more height (T– 14.5)

3.3. Mechine learning Random Forest (RF)

According to research conducted by [23] RF algorithm has superiority in handling complex data with non-linear pattern. RF is algorithm ensemble based that combines many models for increase accuracy classification. In its approach, RF uses gathering tree decision trees that work in a way simultaneously, so that capable produce more results stable and reduce risk of overfitting compared with a tree model single decision [24]. Equation (2) shows THI prediction.

$$\hat{Y} = \frac{1}{N}\sum_{I=1}^{N} f_{i(x)} \quad (2)$$

where $\hat{Y}$ is prediction end of THI, N is amount tree In Random Forest, $fi(x)$ is the prediction of the i-th tree for input x (temperature and humidity).

3.4. Long Short-Term Memory (LSTM)

The LSTM model is type special from Recurrent Neural Network (RNN) designed for overcome problem vanishing gradient and capturing dependence term long in series data time (time series) and increasingly big coefficient autocorrelation (ACF) then will be better [25].

$$ft = \sigma(Wf.[ht - 1, xt] + bf) \quad (3)$$

Information on Equation (3) show that f_t is the output of gate forget the time t. Its value between 0 and 1 because generated by the function sigmoid activation (σ). The value 0 has means "forget it" all information", whereas value 1 means "keep" all information". σ is function activation sigmoid function this change input become range between 0 And 1. W_f is matrix weight which studied during training for gate forget. m matrix connecting input (x_i) and hidden state previously (h_{t-1}) to gate forget. [$h_{t-1,}$ x_t] is combination from previous *hidden state* (h_{t-1}) and input at time t (x_t), b_f is bias which studied for gate forget, h_{t-1} is *hidden state* on time previously t–1, x_t is Input on time t.

4. Results and Discussion

THI calculation, thermal comfort level classification was performed using the RF model, which was evaluated based on precision, recall, and F1-score metrics to measure the accuracy of thermal comfort zone classification. Predictive modeling of THI values using LSTM, which aims to predict future thermal comfort trends. Model evaluation was carried out with use metric Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and Coefficient of Determination (R^2) to assess how much good model in predict THI value. The results obtained from the RF and LSTM models then compared to for see the advantages and disadvantages of each, the analysis carried out in chapter This expected can give deep insight about classification comfort thermal and mitigation strategies that can implemented For increase quality plant on plants specifically the implications to sector agriculture wide especially those located on the coast north island Java.

4.1. Research location

This study carried out in coastal areas part north Tegal Regency and Tegal City (shown in Figure 2), with focus on the central area cultivation plant food (for example, rice, corn and soybeans). Determination location based on variations condition microclimate and availability of historical data cultivation use data historical average daily during 10 Year from 2016 to 2025 at BMKG Tegal which consists of from data (air temperature, humidity, rainfall, radiation the sun and speed wind).

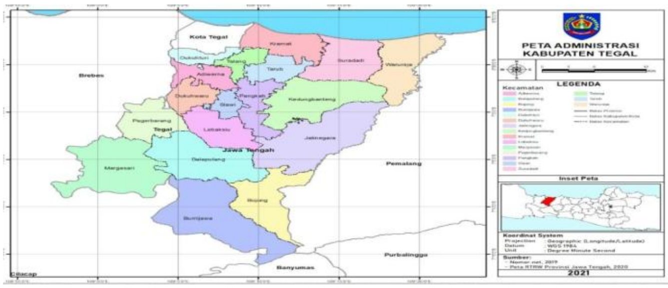

Figure 2. Research area in Tegal Regency and City

4.2. THI trend analysis year (2016-2025)

In this studies, the average temperature of location research along beach north districts and cities Tegal analyzed for identify trend change temperature during period 2016 to 2025. Visualization average temperature, as presented in graph, depicting fluctuations temperature daily in various point measurement, which is then averaged for give description comprehensive about condition temperature in the region. Figure 3 illustrates change average temperature in each location study during period 2016–2025.

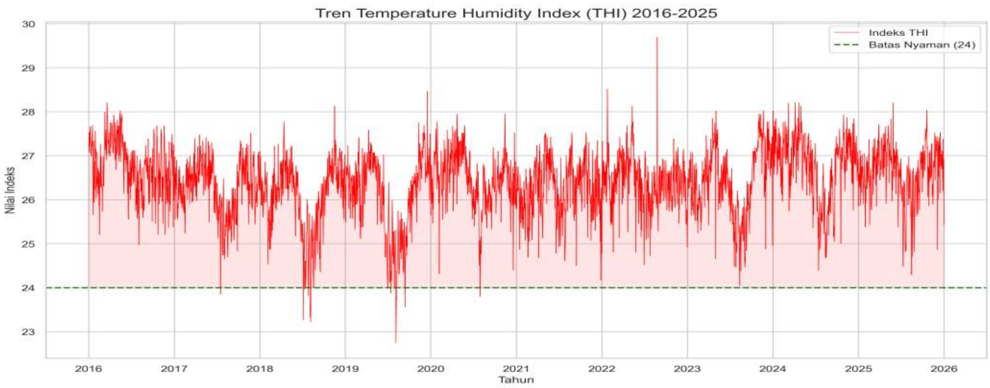

Figure 3. THI trend analysis year (2016-2025)

The graph “Trend of Temperature Humidity Index (THI) 2016–2025” shows change THI value during 10 year period based on condition temperature and humidity air. This graphic used for analyze level comfort thermal influential environment to cultivation plant food in the Tegal Red Line area show THI value of 2016–2025, green line intermittent shows the comfort limit of THI, namely around value 24. The X axis shows period time (years) Y axis shows THI value. In terms of in general, the THI value is in the range around 23 to 30 and the majority of data is above the comfort limit (THI > 24), which indicates that condition environment tend experience pressure hot or condition not enough comfortable, in the 2016–2017 period. At the beginning period observations the THI value is in the range (THI > 28). This condition show environment relatively hot and humid level comfort thermal plant start decreasing, potential stress hot light start appear. In the 2018–2019 period seen a number of THI reduction up to approaching (23–24).

This matter show there is period a better environment comfortable, possible influenced by the season rain or improvement humidity that decreases temperature effective. 2021–2025 period show relative THI condition high and stable, there is surge extremes that reach almost 30. Spike the indicates incident hot extreme, season more drought intense, or improvement temperature consequence phenomenon global climate such as El Niño.

4.3. THI trend analysis with five variables

Distribution variables microclimate in the Tegal region which consists of from temperature, humidity, rainfall rain, sunshine sun, speed wind, and THI, graph between the five variables can be seen in the graph wich shown in Figure 4

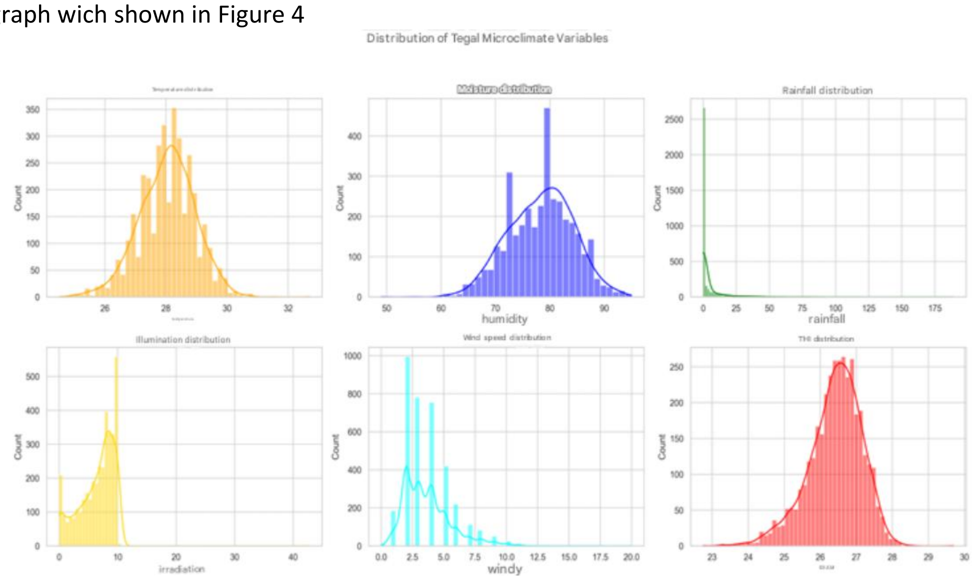

Figure 4. THI trend analysis with every variables

Temperature histogram show pattern distribution that approaches normal distribution (normal distribution). Characteristics majority temperature is in the range of 27–29°C Peak distribution is at around 28°C. This is show that the Tegal region has condition relative temperature stable, characteristic climate tropical warm and varied temperature that is not too extreme. A distribution that approaches normality indicates temperature data good enough used in machine learning modeling because no own deviation large extremes. Distribution humidity also forms pattern approaching normal. Characteristics mark humidity dominant is at (70–85%) peak distribution around (78–80%). Humidity the air in Tegal is classified as height, which is characteristics of tropical regions. Variability humidity relatively moderate so that environment tend moist throughout year especially in the season rain. Distribution rainfall, rain show right-skewed distribution pattern (skewed to right). Its characteristics part large data is at the value rainfall Rain low, there is a number of mark extreme with rainfall very heavy rain matter This show rainfall rain in Tegal is fluctuating, there are big difference between season rain and seasons drought part big day own Rain low, but in the period certain happen very heavy rain. This condition can cause drought in the season dry season, flooding in the rainy season rain and instability pattern plant.

Distribution radiation show pattern has no symmetrical. Its characteristics majority mark is in the range of 6–10 hours available mark low in some. This period show intensity radiation sun enough high, but influenced by the season rain and cover clouds. Illumination the sun is very important in the process of photosynthesis, growth vegetative plants and formation results harvest. Distribution speed wind show a pattern that tends to spread on value low majority speed wind is at 1–5 m/s. There is a number of mark extreme above 10 m/s. Speed the wind in the Tegal area is relatively low until moderate and have role important in circulation air, water evaporation, and control temperature microclimate.

Descriptive statistics are used to analyze the characteristics of data distribution, including central tendency (mean), variability (standard deviation), and range (minimum to maximum value). This can be seen in the following Table 1.

Table 1. Statistics descriptive intervariable

Descrip-tion	Tempera-ture	Humidity	Rainfall	Irradiation	Wind velocity	THI
Count	3653.0000	3653.0000	3653.00000	3653.00000	3653.00000	3653.00000
Mean	28.041007	78.317000	5.214536	6.493676	3.661648	26.414428
Std	0.911740	5.992029	14.667004	3.006059	1.933010	0.745302
Min	24.500000	49.000000	0.000000	0.000000	0.000000	22.747750

Descrip-tion	Tempera-ture	Humidity	Rainfall	Irradiation	Wind velocity	THI
25%	27.400000	74.000000	0.000000	4.500000	2.000000	25.988200
50%	28.100000	79.000000	0.000000	7.300000	3.000000	26.488750
75%	28.700000	83.000000	2.200000	8.800000	5.000000	26.919050
Max	32.700000	95.000000	189.100000	42.800000	20.000000	29.697000

Temperature in this area own climate that tends to hot and stable. The average temperature is at 28.04 °C with very small variations (standard deviation only 0.91°C). The lowest temperature ever recorded recorded is 24.5°C and the highest reaching 32.7°C. Air humidity is classified as very humid, typical of tropical regions, with an average of 78.32%. Even in its driest conditions, the humidity is Still is at 49 %. The data shows distribution rainfall very uneven rainfall (skewed). The 25% to 50% quartile value (median) is 0.00 mm, which means that in most big day no happen rain. However, the value maximum reached 189.10 mm, indicating that this region once in a while experience extreme rain or fertile in season rainfall. Standard value large deviation (14.67 mm) emphasizes fluctuations rainfall heavy rain between day, duration radiation. The average sun is around 6.49 hours per day, the wind speed blowing relatively calm until currently with an average of 3.66 knots/unit speed and maximum speed reached 20.00 knots/unit.

4.4. Correlation matrix between microclimate parameters

Following its analysis discussion short from matrix correlation (heatmap) between microclimate parameters said, this analyzed based on strength relationship (value approaching 1 or -1 means strong, approaching 0 means weak) and direction relationship (positive means compared straight negative means compared backwards).

Figure 5. Correlation matrix between microclimate parameters

In Figure 5, we can see that very strong correlation between temperature and THI (0.84) relationship both of them own correlation very strong positive by 0.84 things. This show that index discomfort thermal (THI). The more tall temperature, the THI value will soaring in a way significant.Correlation Negative Strong Enough: Humidity with Irradiation (-0.53) has connection quite upside down strong. When the intensity radiation sun height, water on the surface and air evaporate as well as temperature increased, so that humidity relative humidity of air will down drastic. Humidity and temperature (-0.39) experienced improvement temperature air followed by a decline humidity air (compared to upside down), which is characteristics general in dynamics atmosphere daily.

The correlation between rainfall and humidity (0.31) shows correlation weak positive. This means that the occurrence of rain a little lots contribute increase humidity air, but not only one factor. The determinant of the correlation between rainfall vs temperature (-0.18) and irradiation (-0.17) has correlation very weak negative. When it rains happened, cover cloud generally increase so that radiation sun decreases and temperature a little decrease.

4.5. THI prediction with LSTM.

Following is analysis discussion about chart THI projection using Long Short-Term Memory (LSTM) method for 2 years forward. Figure 6 show THI prediction with LSTM.

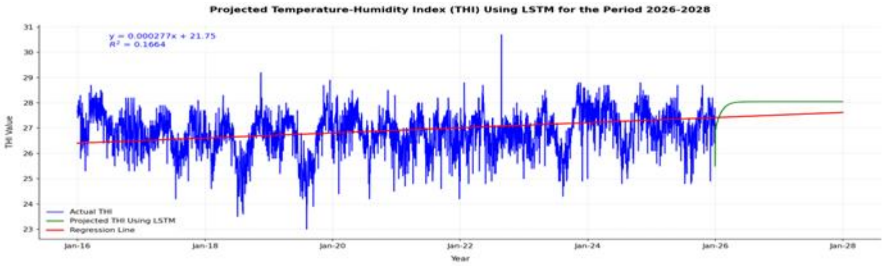

Figure 6. THI prediction with LSTM

Ascension constant colored linear regression line red own equation $y = 0.000277x + 21.75$. This positive slope show existence trend improvement long-term THI value long from 2016 to 2025. This is indicates that the area in a way gradually become the more hot and not comfortable in a way thermal along walking time strength prediction regression (R^2). The value of $R^2 = 0.1664$ is classified as low. This is normal thing for meteorological data daily Because fluctuations daily is very random (*noisy*). This number show that time (linearly) only capable explain around 16.64% of variation THI changes, while the rest influenced by dynamics weather extreme daily m start beginning in 2026, the LSTM model (green line) picks up over prediction. It looks like chart move up sharp from mark around 25.5, then slope and move constant stable (stagnant) in numbers around 28.0 of the image above can be results LSTM performance evaluation (shown in Table 2) is accuracy (1-MAPE of 98.42%), accuracy (R2 Score of 45.24%) and accuracy tolerance (< 5%) of (97.79%).

Table 2. LSTM performance accuracy results

Performance evaluation	Result
Accuracy (1-MAPE)	98.42%
Accuracy (R2 Score)	45.24%
Tolerance Accuracy (<5%)	97.79%

4.6. Seasonal Average THI (2016-2025)

The seasonal average THI for the period between 2016 to 2025 is calculated in this section, the purpose of this calculation is to understand the fluctuations in thermal comfort based on variations in.

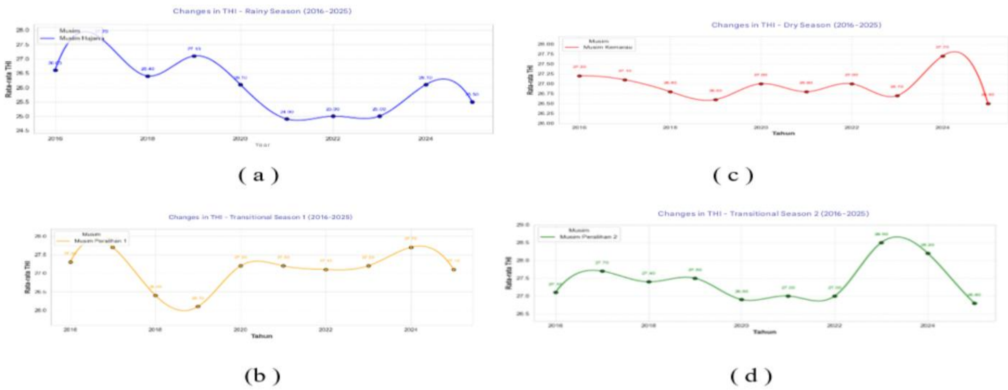

Figure 7. Seasonal Average THI (2016-2025): (a) Rainy Season, (b) Transition Season 1 (c) Dry Season, (d) Transition Season 2.

In the Figure 7, graph (a) shows THI change-rainy season (2016-2025), X- axis (horizontal): shows component year from period 2016 to 2025 and the Y axis (vertical): shows the average THI with range mark from 24.5 to 28.0. Trend chart curved line shape dynamic (spline/smooth line) colored blue which indicates fluctuations average THI value of year to year during season rain, overall average THI in the season rain in the 2016-2025 period did not stable and shows pattern dynamic fluctuations and there are two waves big increase (peak) namely in 2017 (27.70) and 2019 (27.10). High THI values (>27) indicate that although happen season rain, combination temperature warm air and high humidity trigger condition the air that feels more hot or stuffy (comfort zone) decreased. In 2021-2023, the THI value ranged from 24.90-25.00, indicating condition distant environment more cool and comfortable for creature life because level stress the heat low.

Graph in Figure 7 (b) shows a curved line graph dynamic colored orange which represents fluctuations index comfort thermal in transition season 1 (usually transitional seasons from season rain to dry season, around March–May). Analysis of transition season 1 (2016-2025) shows that index comfort thermal dominated by high values (above 27.00). This chart own characteristics fluctuations shaped the letter "U" at the beginning period (2016-2019), followed by phase long stable in the middle line (2020-2023), then closed with surge form peak twins in 2024 (27.70) before slope back in 2025.

Graph (c) shows a curved line dynamic colored red represents movement index comfort thermal especially in the dry season, it's different with chart season rain or transition that has phase stagnant, dry season (2016-2025) has characteristics constant fluctuations, this graph closed with very contrasting and extreme dynamics at the end period, namely the highest THI record in 2024 (27.70) which immediately followed by a record low THI in 2025 (26.50)

And graph (d) trend main curved line graph dynamic colored green represents movement index comfort thermal in transition season 2 (transitional season) from season drought to season rain, usually around September–November). Transition season analysis 2 (2016–2025) shown pattern the most extreme data movement among season other after had time stable and cool in 2020-2022, occurs surge index comfort very massive thermal in 2023 (28.50) due to condition very hot air, before finally down in a way drastic to point the lowest in 2025 (26.80)

4.7. THI classification with Random Forest

Following its analysis and discussion about distribution category comfort based on distribution category comfort.

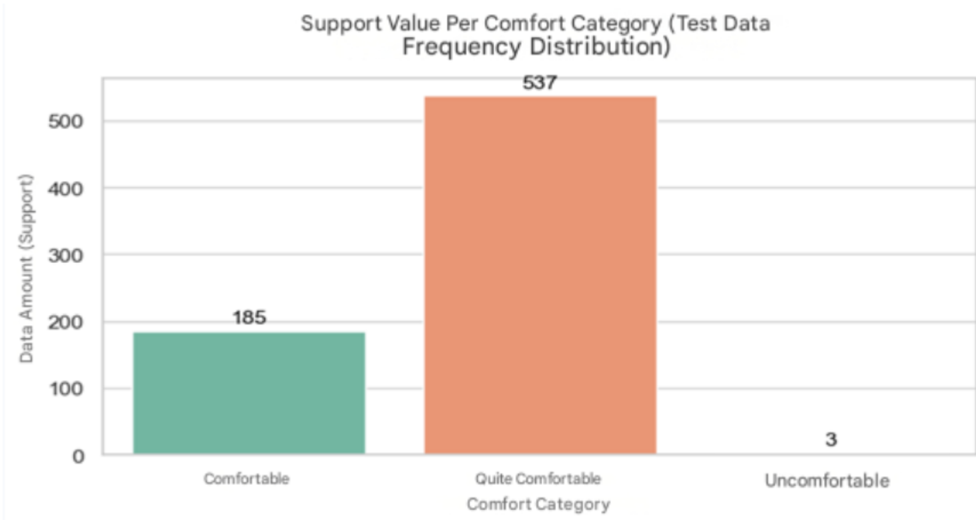

Figure 8. THI classification with Random Forest

Based on Figure 8, "Distribution Category Comfort (New Range Tropical Climate)", results adjustment range new show remote data distribution more proportional and realistic for tropical areas. Conditions environment dominated by the category As shown in Figure 9, the Random Forest model is able to predict and classify comfort categories with a consistent level of accuracy, thus strengthening the validity of the observations. “Enough Comfortable ” with amount sample highest reached 537 data, followed by the “Comfortable**”** category as many as 185 data. This is indicates that part big time observation be in the corridor index ideal and sufficient thermal comfortable for plant Because the majority of data is centered in cool areas until lukewarm. On the other hand, the “Uncomfortable” category was recorded as very minimal, namely only as many as 3 data. Scarcity of data in the category This prove that with new threshold (THI > 28), stress hot extreme (severe heat stress*)* almost no once happened at the location study during period observation. In overall, this implementation range new succeed represent climate tropical in a way accurate, where the environment tend stable at the level conducive comfort with exposure temperature air very controlled extremes.

RANDOM FOREST PERFORMANCE TEST REPORT (TEST DATA)

Global Model Accuracy on New Data: 99.59%

Classification Report Model Evaluation:

	precision	recall	f1-score	support
Comfortable	0.99	1.00	0.99	185
Quite Comfortable	1.00	1.00	1.00	537
Uncomfortable	1.00	0.67	0.80	3
accuracy			1.00	725
macro avg	1.00	0.89	0.93	725
weighted avg	1.00	1.00	1.00	725

Figure 9. Random forest performance test

4.8. Confusion matrix Random Forest evaluation

Matrix used For evaluate performance of the Random Forest (RF) model in classify level comfort thermal based on Index Temperature Humidity (THI). of explanation in on can can be in Figure 10.

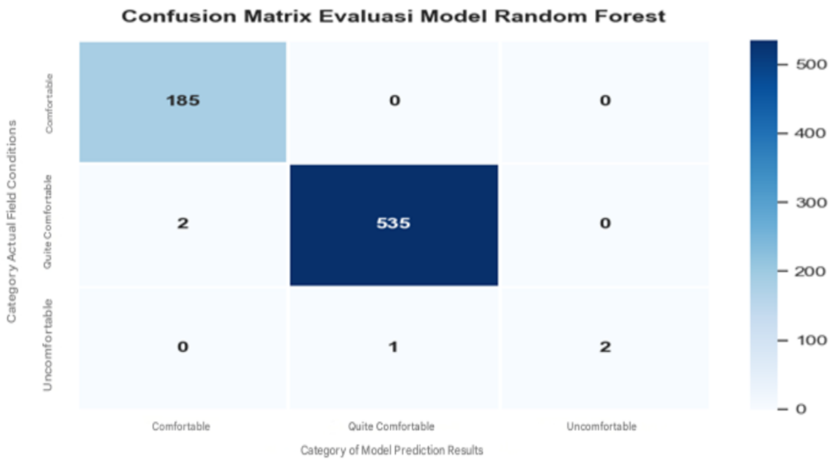

Figure 10. THI confusion matrix with Random Forest

Based on results mapping on the classification scatter plot graph comfort thermal, the data distribution shows very clear and neat grouping (clustering) based on combination temperature air and humidity daily (RH). Most of the test data is centered on the category Enough Comfortable (color orange) which dominates range temperature lukewarm typical of tropical regions, followed by the category Comfortable (color green) which is formed at the time temperature air more low and humidity moderate. The decision boundaries formed by the Random Forest model are visible consistent follow rules Thom's physical formula, where the increase temperature accompanied by humidity tall will linearly direct shift position point to direction Uncomfortable cluster (color red).

Effectiveness the performance of this Random Forest model proven by the level accuracy very high testing, namely reached 99.59%. Figure 10 show that algorithm machine learning capable recognize, study, and map pattern spatial from climate parameters micro in a way almost perfect without experience overlapping overlapping which means inter-category. With structure precise clusters This model proved to be very reliable. For used in predict as well as monitor dynamics index comfort environment tropical in a way automatically in the future.

4.9. Graphics evaluation THI prediction

Based on results evaluation prediction THI (Temperature-Humidity Index) value using the Random Forest Regressor model, the model performance shows level very high accuracy and precision on test data. Result shown in Figure 11. This proven by the value coefficient determination (R^2 Score) which reached 99.92%, which means variation variables independent in the form of temperature air (T) and humidity (RH) are capable of explain change actual THI value in a way almost perfect. Error value predictions were also recorded as very low, with The Mean Absolute Error (MAE) is 0.0062 and the Root Mean Squared Error (RMSE) is 0.0206. The error values are close to zero This signify that distance or deviation between THI values predicted by the model with mark actual in the field very much small and not significant.

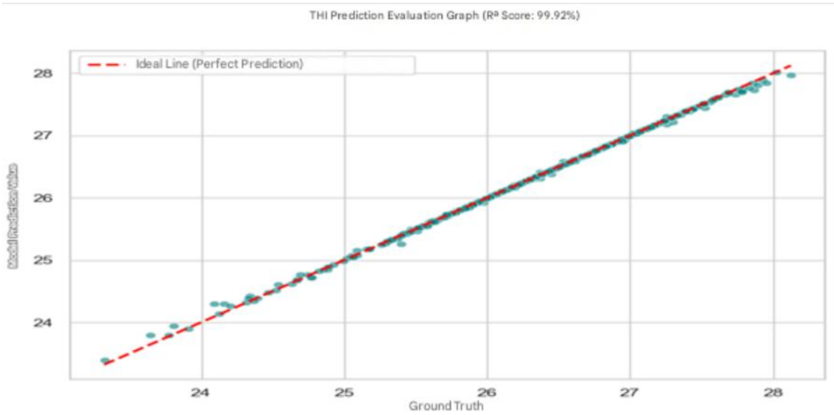

Figure 11. Graph evaluation prediction

The reliability of this model emphasized by the value Mean Absolute Percentage Error (MAPE) which is only by 0.0240 %. scientifically, the MAPE value is far below 10% threshold categorizes performance this model forecasting to in "Very Accurate " level. Low consistent RMSE values with MAE also showing that the model does not produce error extreme predictions (outliers) on new test data. Through combination all over metric regression said, can concluded that algorithm Random Forest is very effective and valid for used as system estimate automatic mark index comfort thermal tropical based on weather parameters daily.

4.10. THI prediction table (2016-2025)

The results of testing the Random Forest Classifier model show performance very precise and optimal classification on test data. Most of the data was successful predicted in a way right on its main diagonal element, where the category “Enough Comfortable" dominate with amount classification very high, followed by the “Comfortable” and “Uncomfortable” categories. The high numbers on the main diagonal This reflect that the model has ability differentiate characteristics physique temperature inter-class air (T) and humidity (RH) in a way strong, so that capable minimize the occurrence of misplaced labels, such as seen in Table 3.

Table 3. THI prediction table results

	T	RH	THI_ Calculation	Current_Category	Prediksi_Category	Prediction_Status
508	28.1	80	2.660.400	Enough Comfortable	Enough Comfortable	Appropriate
2780	28.5	73	2.642.100	Enough Comfortable	Enough Comfortable	Appropriate
416	28.5	75	2.657.500	Enough Comfortable	Enough Comfortable	Appropriate
3475	27.7	77	2.603.020	Enough Comfortable	Enough Comfortable	Appropriate
2107	28.8	81	2.730.565	Enough Comfortable	Enough Comfortable	Appropriate
3097	27.7	73	2.573.980	Enough Comfortable	Enough Comfortable	Appropriate
2306	25.1	73	2.480.850	Comfortable	Comfortable	Appropriate
1221	27.6	73	2.565.465	Enough Comfortable	Enough Comfortable	Appropriate
2437	27.5	82	2.621.300	Enough Comfortable	Enough Comfortable	Appropriate
795	28.6	76	2.673.880	Enough Comfortable	Enough Comfortable	Appropriate
2587	27.2	83	2.601.255	Enough Comfortable	Enough Comfortable	Appropriate
456	28.8	80	2.722.700	Enough Comfortable	Enough Comfortable	Appropriate
2795	28.7	67	2.612.270	Enough Comfortable	Enough Comfortable	Appropriate
1115	27.6	83	2.637.515	Enough Comfortable	Enough Comfortable	Appropriate
1777	26.1	87	2.527.060	Enough Comfortable	Enough Comfortable	Appropriate

Although global performance of the model is very high, there is A little error or missed the boundary data area between adjacent categories. This minor error in the form of slight shift in data between category Comfortable and Quite Comfortable, and between Enough Comfortable and Uncomfortable. Phenomenon this is very normal in a way statistics Because mark decimal actual THI index which is right on the numbers critical separator (such as close to 25.00 or 28.00) has characteristics distribution very similar features, so creating a slight bias at the branching tree decision tree. overall, Confusion Matrix This prove reliability of the Random Forest model in guard balance predictions at every comfort level thermal climate tropical.

4.11. Graphics mapping results THI classification using Random Forest

Based on the heatmap of the Random Forest model performance test matrix, the value metric evaluations including precision, recall, and F1-score show level success very high classification in every category comfort like seen in the Figure 11.

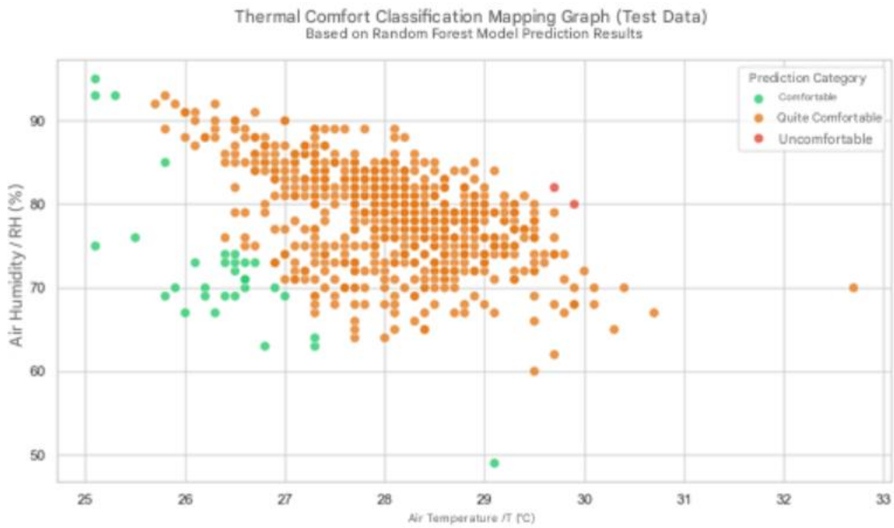

Figure 11. THI mapping results graph

Category Enough Comfortable get the most optimal and closest score perfect (PRECISION: 0.9942, RECALL: 0.9985) because supported by availability abundant data samples (support) in the climate dataset tropical this. Meanwhile, category “Comfort” also shows very stable performance with F1-score value of 0.9375, indicating that the model is capable detect condition cool in a way precise and balanced without any significant bias.

On the other hand, the “Uncomfortable” category succeed record perfect precision value of 1.0000, which means all data that was predicted as "Uncomfortable" by the model was proven 100 % correct condition current field. Although recall value in class is lower (0.6667) due to influence amount very few extreme data samples (imbalanced data), the global F1-score value still endure solid at 0.8000. Supported by the achievement Global Accuracy of 99.31%, matrix performance This confirm that algorithm Random Forest is very reliable, stable, and valid for used as system standardization monitoring index comfort thermal in a way automatic.

5. Conclusion

Following is conclusion comprehensive from all over series data analysis and machine learning modeling done, arranged formally to the topic "Prediction and Classification of Temperature Humidity Index (THI) with Random Forest and LSTM is based on analysis 10-year trend (2016–2025) at Tegal BMKG Station, index comfort thermal (THI) on all seasons (Rainy, Dry, Transition 1, and Transition 2) show pattern very dynamic fluctuations with trend high THI value (dominated by range 25.5 to 27.5). Characteristics climate tropical Tegal which has temperature daily warm and humid tall in a way automatic shift condition average environment to direction “Enough Comfortable" to “Uncomfortable” for a number of type vegetation sensitive. Anomaly weather extreme events that occurred in 2023 and 2024 (especially during the rainy season) dry season and transition 2) had time jack up THI value up to touch record the highest 28.50, which indicates existence risk stress heat (severe heat stress) for plant food if not mitigated.

Classification Model (Random Forest Classifier) applies threshold based climate tropical (Comfortable) ≤ 25, Enough Comfortable 25- 28, Uncomfortable ≥ 28), the model is successful classify level comfort thermal with very high global accuracy, namely 99.31 %. Matrix performance (Precision, Recall, and F1-score) proves that the model is capable separate the decision boundaries of the Temperature (T) and Humidity (RH) parameters precision, although there is inequality amount of data (imbalanced data) in the category extreme

Prediction Model LSTM THI value produces level outstanding reliability normal with R 2 Score reached 99.92%. Very low mark error (Mean Absolute Error /MAE: 0.0062; Root Mean Squared Error /RMSE: 0.0206) and mark The Mean Absolute Percentage Error (MAPE) of only 0.0240% confirms this model combination to in " Highly Accurate " category for projecting fluctuations mark future index, results prediction series time (time-series) of LSTM can made into references scientific For determine timetable irrigation adaptive, settings shade / trap water vapor, up to adjustment time planting to minimize risk decline results harvest in a way significant or damage permanent consequence change local climate.

Credit Authorship Contribution Statement

All Author: Writer responsible answer on conceptualization research, secondary data curation from BMKG, formalization analysis mathematics using the methodology formula machine learning model development, implementation algorithm Random Forest and LSTM using Python, processing visualization graph (evaluation matrix, confusion matrix, and graph classification), as well as do review critical (review*)* and final editing to script For optimization implementation in the field cultivation plant food.

Declaration of Competing Interests

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.

Data Availability

Data will be made available on request.

Use of Artificial Intelligence

AI technology was used in this manuscript solely for grammatical editing, sentence structure refinement, and text outlining. All core ideas, data collection, analysis, and conclusions are strictly the authors' original work. The authors retain full responsibility for the entire content..

References

[1] B. D. Primadita *et al.*, "STUDI KOMPARASI PETA BAHAYA KEKERINGAN METEOROLOGI DAN HIDROLOGI WS PEMALI-COMAL," vol. 3, no. 2, pp. 141–152, 2023, doi: 10.56860/jtsda.v3i2.85.

[2] A. B. Møller, V. L. Mulder, G. B. M. Heuvelink, and N. M. Jacobsen, "Can We Use Machine Learning for Agricultural Land Suitability Assessment ?," 2021, doi: 10.3390/agronomy11040703

[3] Afandi and A. Supriyanto, "Thermal Comfort Projection on Northern Coast of Central Java Using Machine Learning," *J. RESTI*, vol. 9, no. 4, pp. 865–878, 2025, doi: 10.29207/resti.v9i4.6537.

[4] J. Avicena, Z. Zyeraf, M. S. Lubis, and U. Tanjungpura, "TINGKAT KENYAMANAN TERMAL TAMAN SRIWEDARI BERDASARKAN ANALISIS TEMPERATURE HUMIDITY INDEX DAN PREDICTED MEAN VOTE," pp. 1–8, 2023, doi: 10.26418/jelast.v11i2.84656.

[5] F. Fauzi, I. Kharisudin, R. Wasono, T. W. Utami, and I. W. Harmoko, "THERMAL STRESS PROJECTION BASED ON TEMPERATURE- HUMIDITY INDEX (THI) UNDER CLIMATE CHANGE SCENARIO," no. 18, pp. 65–73, 2022.

[6] Ernawati and Guntur, "Dampak Krisis Global terhadap Ketahanan Pangan Nasional di

Indonesia," vol. 2, no. 2 pp. 37–41, 2024, doi: 10.70716/emis.v2i2.288.

[7] P. Pandya and N. K. Gontia, "Early crop yield prediction for agricultural drought monitoring using drought indices, remote sensing, and machine learning techniques Parthsarthi Pandya," vol. 14, no. 12, pp. 4729–4746, 2023, doi: 10.2166/wcc.2023.386.

[8] Budiarto I. et al, "Performance comparison of manual and automatic rain prediction using machine learning," *Pros. Semin. Nas. Fis.*, vol. 6, no. 1, pp. 794–803, 2025, doi: 10.33122/ejeset.v6i1.955.

[9] A. Y. Barrera-animas *et al.*, "Machine Learning with Applications Rainfall prediction : A comparative analysis of modern machine learning algorithms for time-series forecasting," *Mach. Learn. with Appl.*, vol. 7, no. August 2021, p. 100204, 2022, doi: 10.1016/j.mlwa.2021.100204.

[10] D. Castro, A. Sanchez, N. Gutierrez, H. Mauricio, S. Alejandra, and B. Sergio, "A framework for predicting greenhouse microclimate in tomato production using XGBoost, LSTM and random forest," pp. 1–17, 2026, doi: 10.1007/s00704-026-06044-x.

[11] D. Candra, R. Novitasari, P. K. Intan, and G. G. Wisnawa, "Penerapan Metode Principal Component Analysis (PCA) dan Long Short-Term Memory (LSTM) dalam Memprediksi Prediksi Curah Hujan Harian," vol. 5, no. 1, pp. 1–11, 2023, doi: 10.47065/bits.v5i1.3114.

[12] S. Iskandar, L. Sebastian, and E. Rusdiyanto, "Analisis Tingkat Kenyamanan Thermal Menggunakan Metode Discomfort Index di Kecamatan Putussibau Selatan Kabupaten Kapuas Hulu," vol. 13, no. 1, pp. 224–230, 2025, doi: 10.26418/jtllb.v13i1.87198.

[13] J. Nithyashri, R. Kumar, S. Balakrishnan, M. A. Kumar, P. Prabu, and S. Nandhini, "Measurement : Sensors IOT based prediction of rainfall forecast in coastal regions using deep reinforcement model," *Meas. Sensors*, vol. 29, no. August, p. 100877, 2023, doi: 10.1016/j.measen.2023.100877.

[14] G. Patel, S. Verma, and M. Kant, "Temporal Disaggregation and Short - Term Forecasting of Hourly Rainfall Data : A Case Study of Raipur City Chhattisgarh, India," *Iran. J. Sci. Technol. Trans. Civ. Eng.*, no. 2022, 2025, doi: 10.1007/s40996-025-01901-7.

[15] J. Hern, S. Bonachela, and M. Dolores, "Computing air temperature and humidity for reference crop evapotranspiration calculation in passive Mediterranean greenhouses," vol. 302, no. May, 2024, doi: 10.1016/j.agwat.2024.108991.

[16] A. Harmin, M. Isra, M. Bravikawati, and A. Naziah, "Analisis Temperature Humidity Index (THI) terhadap Kenyamanan Termal dengan Penambahan Tumbuhan Hijau pada Gedung Bertingkat," vol. 9, no. 1, pp. 294–298, 2025.

[17] Afandi and Supriyanto, A., "Thermal Comfort Projection on Northern Coast of Central Java," vol. 9, no. 4, pp. 865–878, 2025, doi: 10.29207/resti.v9i4.6537.

[18] J. S. Mishra *et al.*, "An impact of agronomic practices of sustainable rice-wheat crop intensification on food security, economic adaptability, and environmental mitigation across eastern Indo-Gangetic Plains," *F. Crop. Res.*, vol. 267, no. September 2020, p. 108164, 2021, doi: 10.1016/j.fcr.2021.108164.

[19] M. Cetin and O. Alsenjar, "Estimation of crop water stress index and leaf area index based on remote sensing data," vol. 00, no. 0, pp. 1–15, 2023, doi: 10.2166/ws.2023.051.

[20] V. Rilansari, C. Rahmah, M. F. Rajabi, and M. Gi. D. Saputra, "Penentuan Kawasan Potensial Ruang Terbuka Hijau Berdasarkan Kerapatan Vegetasi Normalized Difference Vegetation Index (NDVI) dan Indeks Kenyamanan Temperature Humidity Index (THI) di Kota Bandar Lampung," *J. Desa Kota*, vol. 7, no. 2, pp. 140–148, 2025, [Online]. Available: http://jurnal.uns.ac.id/jdk/article/view/97376.

[21] J. Sitepu, "Kajian Kenyamanan Ruang Terbuka Hijau Lingkungan di Fakultas Pertanian Universitas Brawijaya Study of Microclimate Situation and Green Open Space Environments Comfort at Agriculture Faculty of Brawijaya University," vol. 9, no. 9, pp. 539–545, 2021.

[22] C. C. Ekine-Dzivenu *et al.*, "Evaluating the impact of heat stress as measured by temperature-humidity index (THI) on test-day milk yield of small holder dairy cattle in a sub-Sahara African climate," *Livest. Sci.*, vol. 242, no. January, p. 104314, 2020, doi: 10.1016/j.livsci.2020.104314.

[23] A. Supriyanto, D. A. Diartonor, B. Hartono, A. Jananto, and A. Afandi, "Classification Of Sea Wave Heights On The North Coast Of Central Java Using Random Forest," *J. Tek. Inform.*, vol. 6, no. 4, pp. 2263–2280, 2025, doi: 10.52436/1.jutif.2025.6.4.5108

[24] C. Chen *et al.*, "Mechanics Forecast of rainfall distribution based on fixed sliding window long short-term memory," vol. 2060, 2022, doi: 10.1080/19942060.2021.2009374.

[25] A. Wijayanto, A. Sugiharto, R. Santoso, U. Diponegoro, and P. Korespondensi, “Identifikasi Dini Curah Hujan Berpotensi Banjir Menggunakan Algoritma Long Short-Term Memory (Lstm) Dan Isolation Forest,” vol. 11, no. 3, 2024, doi: 10.25126/jtiik.938718.