Accredited Ranking SINTA 2
Decree of the Director General of Higher Education, Research, and Technology, No. 158/E/KPT/2021
Validity period from Volume 5 Number 2 of 2021 to Volume 10 Number 1 of 2026

Published online on: **http://jurnal.iaii.or.id**

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 6 No. 4 (2022) 612 - 623 ISSN Media Electronic: 2580-0760

DHF Incidence Rate Prediction Based on Spatial-Time with Random Forest Extended Features

Elqi Ashok[1], Sri Suryani Prasetiyowati[2], Yuliant Sibaroni[3]
[1,2,3]Informatics, School of Computing, Telkom University
[1]elqiashok@student.telkomuniversity.ac.id, [2]srisuryani@telkomuniversity.ac.id, [3]yuliant@telkomuniversity.ac.id

Abstract

This study proposes a prediction of the classification of the spread of dengue hemorrhagic fever (DHF) with the expansion of the Random Forest (RF) feature based on spatial time. The RF classification model was developed by extending the features based on the previous 2 to 4 years. The three best RF models were obtained with an accuracy of 97%, 93%, and 93%, respectively. Meanwhile, the best kriging model was obtained with an RMSE value of 0.762 for 2022, 0.996 for 2023, and 0.953 for 2024. This model produced a prediction of the classification of dengue incidence rates (IR) with a distribution of 33% medium class and 67% high class for 2022. 2023, the medium class is predicted to decrease by 6% and cause an increase in the high class to 73%. Meanwhile, in 2024, it is predicted that there will be an increase of 10% for the medium class from 27% to 37% and the distribution of the high class is predicted to be around 63%. The contribution of this research is to provide predictive information on the classification of the spread of DHF in the Bandung area for three years with the expansion of features based on time.

Keywords: Incidence Rate, DHF, Prediction, Random Forest, Ordinary Kriging

1. Introduction

Dengue Hemorrhagic Fever (DHF) is a category of dangerous disease that can cause death for sufferers. This disease is transmitted through the bite of Aedes Aegypti and Aedes Albopictus mosquitoes. The mosquito carries the dengue virus and transmits it to humans through bites, resulting in dengue symptoms [1]. The spread of dengue cases is influenced by several factors, including rainfall [2], temperature, altitude, distribution of men [3], population mobility, population density, level of community knowledge, wind speed [4], and humidity [2][5].

DHF spreads in tropical climates such as Indonesia. One of the areas with the highest incidence rate of DHF is Bandung City. According to [6] and [7], there have been recorded fluctuations in DHF cases in Bandung City from 2017 to 2021. In 2017, 1.786 cases were recorded, in 2018 there was an increase in cases recorded at 2.826 cases [6], in 2019 the number of cases doubled from the previous year, which was recorded at 4.424 cases, then in 2020 it decreased to 2.790 cases, and again increased in 2021 to 3.743 cases [7]. The highest number of cases occurred in 2019 which was recorded at 4.424 cases, a drastic increase of 56.54% compared to the number of cases in 2018. The three sub-districts with the highest distribution of DHF cases include Arcamanik sub-district with 241 cases, Coblong sub-district with 263 cases, and Kiaracondong sub-district with 308 cases. Sub-districts with the smallest distribution of DHF cases were Sumur Bandung Sub-district, which recorded 49 cases, Bandung Wetan Sub-district, 62 cases, and Cinambo Sub-district, 70 cases [6]. This shows that DHF is a difficult disease to handle with the number of cases that always fluctuates every year and there is no optimal solution.

Therefore, the government hopes for a solution to reduce dengue cases in each sub-district. One of them is by displaying the distribution of cases in each sub-district in the form of classification prediction maps for the next few years so that the community and government can provide optimal actions and solutions to reduce the spread of dengue cases. In the field of information technology, the use of machine learning can be implemented to predict and classify dengue incidence rates based on historical data from previous years. In addition, there are other methods that can predict the incidence rate of DHF in areas where the value is unknown by kriging interpolation. One method that can be used is Ordinary Kriging. The number of

Accepted: 08-07-2022 | Received in revised: 31-07-2022 | Published: 22-08-2022

DHF cases that always fluctuates every year makes this problem even more challenging to predict the distribution of the incidence rate of DHF disease in each sub-district.

Not many studies have discussed the prediction of the classification of the incidence rate distribution of DHF based on the expansion of spatial and time-based features. Some studies usually only focus on predicting or classifying DHF, but not based on the problem of spreading cases in an area. Prediction and classification of DHF have been carried out by several researchers [8], [9], [10], and [11]. The study [8] applied the Random Forest algorithm using 10-fold cross-validation to predict dengue fever based on patient data from hospitals and laboratories. The number of trees built is 500 trees with the number of features that are tried on each splitting of 5 features. This study resulted in an accuracy of 92.34%, recall of 94.04%, and specificity of 92.19%.

Research [10] built a dengue virus diagnostic system by combining the Random Forest algorithm and Raman Spectroscopy. The data used were 100 samples collected from patients exposed to the dengue virus. Of the 100 data samples, 45 samples were labeled positive. This study reduces the dimensions of the data using the Principal Component Analysis (PCA) method and evaluates the Random Forest classification model with 5-fold cross-validation. The built diagnostic system produces 91% accuracy, 91% recall, and 91% specificity. Furthermore, the study [11] applied the Random Forest algorithm and Artificial Neural Network (ANN) to predict the clinical degree of DHF. The data used comes from patient data and laboratory data. Both models were evaluated using 5-fold and 10-fold cross-validation. This study resulted in the highest accuracy in the Random Forest model of 58% with 10-fold and the ANN model of 57% with 5-fold.

In research [9] classifying DHF disease using Support Vector Machine (SVM), Naïve Bayes, and Random Forest. This study uses secondary data from M. Syafii's research in 2006 which was taken from the medical records of patients at Dr. Hospital. Sardjito Yogyakarta on December 13-16, 2005. The data sample amounted to 213 patients with dengue fever. The features used were fever, spotting, bleeding, and tourniquet test. The performance of the SVM, Naïve Bayes, and Random Forest models was measured using accuracy and the values obtained were 44.5%, 69.8%, and 79.6%, respectively. Based on the comparison of the accuracy values of the three models, the Random Forest model has a much better accuracy value than the SVM and Naïve Bayes models in classifying DHF.

The study [12] applied the Random Forest and Logistic Regression methods with elastic-net to predict the length of stay in DHF patients and identify the most important features. The data used is data on patients with dengue fever in hospitals taken from February 2021 to September 2017. The total number of data obtained is 1148 sample data and 40 features. This study uses a data sharing scenario of 70% for training data and 30% for test data. The evaluation process of the model uses 10-fold cross-validation and the results were obtained with the Area Under the Curve (AUC) value of 0.75 for the Logistics Regression model and 0.72 for the Random Forest model.

The development of DHF disease maps has been carried out [4], [13], and [14] using Random Forest. The study [4] applied the Random Forest and K-Nearest Neighbor (KNN) algorithms with two scenarios. In the first scenario, the modeling process uses a patterned model based on data from the previous 2 years and produces the lowest Root Mean Squared Error (RMSE) value of 29.25. In the second scenario, the modeling process uses a random data model and the lowest RMSE value is 45.48. However, this research still has shortcomings, the resulting RMSE value is still quite high and the map developed is only limited to 1 year. While the map development in research [13] the features used were too few and did not use the features that caused DHF. This study resulted in a very good accuracy value, but the accuracy value was not indicated and the map developed was difficult for readers to understand.

The study [14] implemented the Random Forest algorithm to predict the transmission of dengue fever in Shenzhen City, China, and determine the most important factors. The process of mapping the risk of dengue transmission is carried out with the help of the Argis software. This study divides the data for model training by 65% and the remaining 35% is used for model testing. The results showed that the AUC value was 0.8 with the most important features being average rainfall, maximum temperature, and workplace density.

While research on kriging has been carried out by [15], [16], and [17]. This study [15] applied the Inverse Distances Weighted (IDW), Ordinary Kriging (OK), and Universal Kriging (UK) methods to predict the pattern of the spread of dengue hemorrhagic fever over the next three years in the city of Bandung based on the number of dengue cases interpolated at village coordinates. The data used is data on dengue cases in 152 villages in the city of Bandung from 2010 to 2015. Then the prediction results for the IDW method are obtained with an RMSE value of 18.56 for 2016, 9.53 for 2017, and 20.91 for 2018. The RMSE value with Ordinary Kriging in 2016 was 18.96, 2017 was 9.67, and 2018 was 21.26. While the RMSE Universal Kriging value in the 2016 prediction is 19.08, in 2017 it is 9.81, and in 2018 it is 22.16.

In research [16] the ordinary kriging method was used for spatial analysis of the spread of DHF mosquitoes by trapping Wolbachia bacteria. Wolbachia bacteria act as virus vectors that can inhibit the Aedes Aegypti

DOI: https://doi.org/10.29207/resti.v6i4.4268

mosquito transmission process with humans. These bacteria were distributed in three areas and observed for three years in a frequency of 4 seasons. However, the results of this study do not explain how well the semivariogram model was used in the ordinary kriging interpolation process to map the spread of DHF and Wolbachia bacteria. Meanwhile [17] used ordinary kriging to develop DHF risk distribution maps by considering entomological index and breteau index.

In studies [18], [19], and [20] discussed the application of the Ordinary Kriging method to predict the spread of DHF. In [18] a risk map for dengue transmission was developed using remote satellite imagery classified by land cover. This study used training data and test data as many as 50 samples in each class. In land cover classification, the algorithms used are cation classifier algorithms such as Maximum Likelihood, Mahalanobis Distance, and Minimum Distance. The classification results show an accuracy value of 90.6% with the Maximum Likelihood classifier.

Subsequent studies [19] predict the transmission of dengue fever in urban areas and determine the influencing environmental factors. Prediction of dengue fever transmission using Ordinary Kriging with spatial and temporal scales. While determining the influencing factors using a geostatistical additive linear model by analyzing the correlation between variables. The results of this study indicate that the features of wind speed, wind direction, and air temperature are factors that influence the transmission of dengue fever. This study has several shortcomings, including not explaining the semivariogram model used in the kriging process and the resulting map is difficult for readers to understand because there are no class labels that indicate the level of dengue fever spread.

Meanwhile, research [20] combines the Ordinary Kriging (OK) method with three spatial interpolation methods in predicting the spread of the Aedes Aegypti mosquito. The three methods are Local Polynomial Interpolation (LPI) which uses exponential kernel function, Radial Basis Function (RBF) with spline model, and Inverse Distance Weighted (IDW) method. While the model used by Ordinary Kriging is a spherical semivariogram model. This study predicts its distribution in 4 seasons and obtained RMSE values in the spring of 0.51 for IDW, 0.44 LPI, 0.47 RBF, and 0.4 OK. In summer, the RMSE values were 0.6 for IDW, 0.53 LPI, 0.51 RBF, and 0.48 OK. The RMSE values in the fall were 0.53 for IDW, 0.46 LPI, 0.48 RBF, and 0.42 OK. In winter, IDW produces an RMSE value of 0.55, LPI of 0.47, RBF of 0.48, and OK of 0.44. The results of this study indicate that OK predictions are superior to IDW, LPI, and RBF.

Based on previous research reviews with the advantages and disadvantages that have been presented, there is no research that combines the random forest and ordinary kriging algorithms. Thus, this study proposes these two methods for the prediction of the classification of the spread of the incidence of dengue fever with the expansion of features based on time. Feature expansion is carried out based on feature data of 2 years, 3 years, and 4 years previously. The data used are climate data, population, education history, and blood type. In previous studies, no one has used blood type data features. The addition of these features is carried out by considering that dengue patients are caused by mosquito bites and their blood is sucked. Therefore, this study predicts the classification with the expansion of features based on the time using the Random Forest and Ordinary Kriging algorithms. The purpose of this study was to determine the distribution of DHF in each sub-district in the next three years and to find the features that had the most influence on the spread of DHF based on the results of the most optimal feature expansion. So that the community and government can provide appropriate prevention and treatment efforts to reduce the spread of DHF in each sub-district in the city of Bandung.

2. Research Methods

The methods used are Random Forest and Ordinary Kriging. Random Forest algorithm was used to predict the classification of incidence rates in 30 sub-districts based on the expansion of features from the previous 2 to 4 years. Then, the results were interpolated with Ordinary Kriging to predict the spread of dengue in the Bandung area for the next three years. The design of the system is shown in Figure 1.

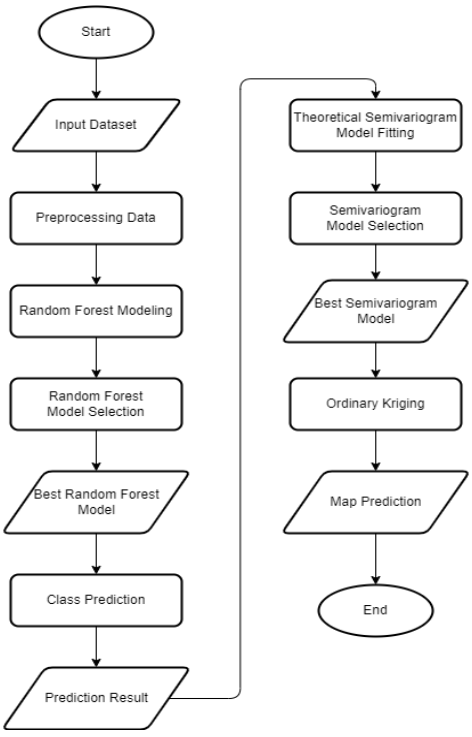

Figure 1. System Design

DOI: https://doi.org/10.29207/resti.v6i4.4268

2.1 Dataset

This study uses data on DHF cases obtained from the Bandung City Health Office, climate data from the Bandung Meteorology, Climatology and Geophysics Agency, population data, educational history data, and blood type data obtained from the Bandung City Central Statistics Agency. The data was collected based on 30 sub-districts in the city of Bandung from 2017 to 2021. Thus, the data obtained were 150 sample data and 13 features. Feature names are denoted by X1...X(n). Table 1 presents the results of the feature name notation and a description of each feature.

Table 1. Dataset

Notation	Description
X1	Total Population
X2	Proportion of Population Male
X3	Rainfall (mm)
X4	Temperature (℃)
X5	Humidity (%)
X6	Blood Type A
X7	Blood Type B
X8	Blood Type AB
X9	Blood Type O
X10	Elementary School Graduate
X11	Middle School Graduate
X12	High School Graduate
X13	College Graduate
Y	Incidence Rate (per 100.000 population)

2.2 Data Preprocessing

The dataset that has been obtained is still in the form of raw data, so it is necessary to use a data preprocessing technique. The use of preprocessing is intended so that the dataset used produces quality data and is ready to be processed to build a classification prediction model. The target variable used is the Incidence Rate (IR) with the following formula [21]:

$$IR = \frac{Case}{Population} \times 100.000 \quad (1)$$

Where Incidence Rate is a value that shows cases of DHF in a population (per 100.000 population) [21].

Table 2. Class Labeling

Class	Label Class	Range
Low	0	IR < 55
Medium	1	55 ≥ IR ≤ 100
High	2	IR > 100

Table 2 explains that an area is categorized as low if the area's IR is less than 55 per 100.000 population. If the IR number is in the range of 55 to 100 per 100.000 population, then the area is categorized as a medium, and categorized as high if the IR number is more than 100 per 100.000 population.

This study uses the stratified k-fold cross validation method to divide the data into 2 parts: training data and test data. The purpose of using stratified k-fold cross validation is to reduce bias in the model [22] and avoid errors caused by unbalanced classes [23]. In general, the way this method works is to divide the dataset into several folds according to the value of k, where each fold is carried out by a training process and model testing [22]. In this study, the number of k used is k=10. The selection is based on the small amount of data used in this study, so it is necessary to carry out more model training processes so that the model built is accurate and can predict well.

2.3. Random Forest

Random Forest is one of the ensemble methods that can be used for the classification of large amounts of data by building a regression tree consisting of a collection of decision trees. The decision tree was chosen randomly from the training data, then combined using the Breiman bagging method. After that, majority voting is carried out based on the decision tree to get predictive results [11].

The performance of the Random Forest model has been tested in predicting and classifying various types of datasets, even for unbalanced classes [24]. This is influenced by the use of random sampling and the principle of the ensemble technique [11]. According to [25] the Random Forest algorithm can naturally adjust to unbalanced classes by down-sampling the majority class and constructing each tree for the minority class so that the dataset becomes more balanced.

The development of the Random Forest model can be carried out in three steps. First, the data is divided into 2 parts training data and test data. The division is 2/3 of the data used as training data and the remaining 1/3 as test data used for validation of learning models on training data. Second, create a decision tree from a random data set with a bootstrap sample. The branching of each tree is determined by predictors chosen at random at the node points. Third, calculate the average value of all the results of the decision tree predictions. This average value is the result of the prediction of the random forest model. Therefore, each individual in the decision tree greatly influences the final predictive value [24]. In mathematical terms, the majority voting formula is as follows [25]:

$$m_{M,n}(\mathbf{x}; \Theta_1, \ldots, \Theta_M, \mathcal{D}_n) = \begin{cases} 1 \; if \quad \frac{1}{M}\sum_{j=1}^{M} m_n(\mathbf{x}; \Theta_j, \mathcal{D}_n) > 1/2 \\ 0 \; otherwise \end{cases} \quad (2)$$

Where $\mathcal{D}_n$ is the training data sample, M is the number of decision trees built, $m_n(\mathbf{x}; \Theta_j, \mathcal{D}_n)$ is the predicted value at point x, and $\Theta_1, \ldots, \Theta_M$ are independent random variables [25].

2.4 Random Forest Prediction Model

The Random Forest model was developed by expanding the feature column based on the features of the previous few years. From the data that has been collected for 5

DOI: https://doi.org/10.29207/resti.v6i4.4268

years, the features can be expanded based on the previous 2 years, 3 years, and 4 years. The target of the prediction model is the current year's target.

Table 3. Prediction model feature expansion scenario

Model	Label	Training Data Feature
2 years before	2A	2019, 2020
	2B	2018, 2019
	2C	2017, 2018
3 years before	3A	2018, 2019, 2020
	3B	2017, 2018, 2019
4 years before	4A	2017, 2018, 2019, 2020

Table 3 shows the scenario of feature expansion in the random forest prediction model based on feature data of 2 years, 3 years, and 4 years before. The prediction process is carried out from 2019 to 2021, for example predicting 2021 based on the feature expansion scenario of the previous 2 years, then the model uses feature column expansion in 2019, and 2020. While the target of the model is 2021. Examples of feature expansion combinations based on the previous 2 years are presented in Table 4.

Table 4. Example of a combination of feature expansion based on the previous two years

Number of Feature	Feature Combination
3	X2, X3, X7
3	X1, X3, X4
3	X1, X8, X12
3	X5, X9, X13
4	X1, X3, X4, X8
4	X2, X9, X10, X11
…	…
26	X1, X2, X3, …, X24, X25, X26

2.5 Random Forest Model Selection

The best random forest prediction model is selected based on the highest accuracy value and the most optimal number of feature extensions. Accuracy is the percentage of truth in the test data which is calculated based on the number of correct predictions divided by the total predictions [8].

$$Accuracy = \frac{TP + TN}{(TP + FP + TN + FN)} \quad (3)$$

Where TP (True Positive) is the actual class labeled positive is predicted to be true as a positive label. TN (True Negative) is the actual class labeled negative which is predicted to be true as a negative label. FP (False Positive) is the actual class labeled negative is predicted to be falsely labeled as positive. FN (False Negative) is the actual class labeled positive which is predicted to be wrongly labeled as negative [26].

Meanwhile, the feature expansion scenario is carried out by utilizing the Sklearn SelectKBest library which can improve the accuracy and performance of the prediction model [27]. The way this technique works is to select a number of k features that have the highest score, where the score is calculated using a univariate statistical analysis of each variable [28]. This study uses the f_classif score function. This function calculates criterion f using dispersion analysis based on the difference in the mean value of the features in finding dependencies on the data. The f_classif function is calculated using the following formula [29]:

$$F = \frac{\frac{1}{C-1}\sum_{i=1}^{C} N_i(\bar{x}_\iota - \bar{x})^2}{\frac{1}{C-1}\sum_{i=1}^{C}\sum_{j=1}^{N_i}(x_{i,j} - \bar{x}_\iota)^2} \quad (4)$$

Where C is the number of classes, N is the number of sample data in the dataset, N_i is the number of sample data with the label class i, $x_{i,j}$ is the feature value of class i, $\bar{x}_\iota$ is the value feature average in class i, and $\bar{x}$ is the average feature value in the data set.

2.6 Class Prediction

At this stage, incident rate class predictions are made for 2022, 2023, and 2024. The prediction process uses the best Random Forest model that has been developed based on data from the previous 2 years, the previous 3 years, and the previous 4 years. The model was selected based on the highest accuracy value and the most optimal number of features.

2.7 Theoretical Semivariogram Model

Theoretical semivariogram is a model that is used as input in the interpolation process using ordinary kriging to predict the incidence of dengue fever in 30 sub-districts and other locations whose values have not been recorded. The semivariogram model was obtained based on the parameters of the distance between 2 points, the range value, and the threshold value [30]. This study uses 3 semivariogram models, namely spherical, exponential, and gaussian models. The general form of the three models is obtained from [30] and is stated as follows.

The general form of the spherical model is shown in equation (5)

$$\gamma(h) = \begin{cases} c\left[\frac{3h}{2a} - \frac{1}{2}\left(\frac{h}{a}\right)^3\right], & h \le a \\ c, & h > a \end{cases} \quad (5)$$

Furthermore, for the exponential model, the general form is shown in equation (6)

$$\gamma(h) = c\left[1 - exp\left(-\frac{h}{a}\right)\right] \quad (6)$$

While the general form of the Gaussian model is shown in equation (7)

$$\gamma(h) = c\left[1 - exp\left(-\frac{h^2}{a^2}\right)\right] \quad (7)$$

Where $\gamma(h)$ is the theoretical semivariogram, c is the sill value, while a is the range value, and h is the distance between 2 points [30].

DOI: https://doi.org/10.29207/resti.v6i4.4268

2.8 Semivariogram Model Selection

The best semivariogram model was chosen by comparing the RMSE values in each model. The semivariogram model with the lowest RMSE value will be selected as the input model in the interpolation process using ordinary kriging. RMSE is an alternative method used to evaluate the prediction case by measuring the error rate of the prediction results of the model built [4].

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\mathrm{Y_i} - \hat{\mathrm{Y}}_\mathrm{i})^2}{n}} \quad (8)$$

Where Y is the actual value of the test data, $\hat{\mathrm{Y}}$ is the predicted value of the test data, and n is the number of test data.

2.9 Ordinary Kriging

Ordinary Kriging is a kriging technique based on stochastic interpolation [31]. This technique is most often used to estimate a value at the location point of an area based on a known variogram and use data in the surrounding environment to make predictions [32].

The incidence rate of DHF at the point $\mathrm{X_0}$ can be predicted using the data values of *n* neighboring samples $\mathrm{X_i}$ and combining them linearly with λ_i weighting [15].

$$\hat{\mathrm{Z}}(\mathrm{X_0}) = \sum_{i=1}^{n} \lambda_\mathrm{i}\, Z(\mathrm{X_i}) \quad (9)$$

Where $\hat{\mathrm{Z}}(\mathrm{X_0})$ is the predicted value at the location point $\mathrm{X_0}$, $Z(\mathrm{X_i})$ is the IR value of DHF in each sub-district, $\mathrm{X_0}$ is the predicted sub-district location, $\mathrm{X_i}$ is the observed sub-district location, λ_i is the weighted value of the observed sub-district location, and n is the number of sample data.

Ordinary kriging is an exact interpolator which means that if $\mathrm{X_0}$ is exactly equal to the observed subdistrict location then the predicted value is exactly equal to the data value at that subdistrict location [32].

$$\hat{\mathrm{Z}}(\mathrm{X_0}) = Z(\mathrm{X_i}), \qquad if\ \mathrm{X_0} = \mathrm{X_i} \quad (10)$$

While the estimated variance of ordinary kriging can be expressed by (σ^2), where μ is the Lagrange parameter, $\gamma(\mathrm{X_0} - \mathrm{X_0})$ is the theoretical semivariogram of the point estimated with it, and $\gamma(\mathrm{X_i} - \mathrm{X_0})$ is the theoretical semivariogram of the point. which is estimated by the sample point [32].

$$\sigma^2 = \mu - \gamma(\mathrm{X_0} - \mathrm{X_0}) + \sum_{i=1}^{n} \lambda_\mathrm{i}\, \gamma(\mathrm{X_i} - \mathrm{X_0}) \quad (11)$$

3. Results and Discussions

This research uses the random forest method with a parameter experiment of the number of trees built as many as 100, 200, 300, 400, and 500 trees with the expansion of the previous 2 to 4 years of features. Meanwhile, the ordinary kriging method was experimented with by applying anisotropy to the major and minor range parameters in each semivariogram model. The best random forest model was chosen based on the highest accuracy value in the parameter experiment and the expansion of its features. In ordinary kriging, the semivariogram model with the lowest RMSE value was chosen as the best model to predict the IR distribution of DHF in the next three years.

3.1 IR DHF Classification using Random Forest

The performance of the random forest classification prediction model was measured using accuracy and the best model was selected for 2 years, 3 years, and 4 years based on the highest accuracy. Table 5, table 6, and table 7 show the accuracy value of the test results of each developed model.

Table 5. Two-year model accuracy

Model	Accuracy
Model 2A	90%
Model 2B	76.67%
Model 2C	96.67%
Model 2 Combined	83.33%

In Table 5, model 2C has higher accuracy than the other 2-year models. The accuracy obtained is 96.67%. Meanwhile, the combined 2-year model has less accuracy than models 2A and 2C. The combined 2-year model and model 2A produced the best accuracy at a hyperparameter of the number of trees built of 100 trees, while models 2B and 2C produced the best accuracy with a hyperparameter of the number of trees of 200 trees.

Table 6. Three-year model accuracy

Model	Accuracy
Model 3A	93.33%
Model 3B	76.67%
Model 3 Combined	83.33%

In Table 6, model 3A is the best model with an accuracy of 93.33%. The model uses a hyperparameter of the number of trees built as many as 500 trees. Meanwhile, the 3-year combined model has less accuracy than model 3A. However, the accuracy of the 3-year combined model is superior to model 3B. Model 3A produces the best accuracy at the hyperparameter of the number of trees built as many as 500 trees, while model 3B uses 400 trees, and the 3-year combined model uses 100 trees.

DOI: https://doi.org/10.29207/resti.v6i4.4268

Table 7. Four-year model accuracy

Model	Accuracy
Model 4A	93.33%

Furthermore, in Table 7, the 4-year model only has 1 model, namely model 4A with an accuracy of 93.33% with a hyperparameter of the number of trees built totaling 500 trees. When compared to model 3A, the resulting accuracy is the same. This is because the class studied by both models is the same, namely 2021. However, the results of the 3A feature expansion are much less, namely 5 features, while the 4A model requires 10 features. Details regarding feature expansion are presented in Figure 2, Figure 3, and Figure 4.

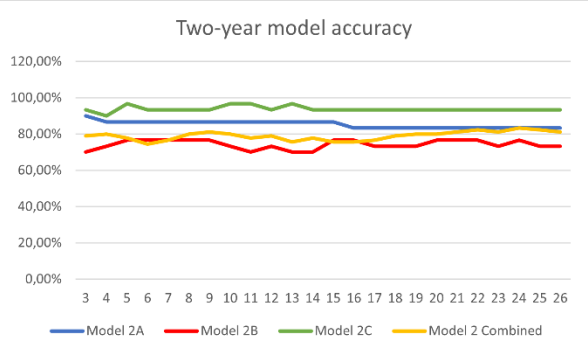

Figure 2. Two-year model accuracy based on feature expansion

Feature expansion is carried out to see the accuracy of the developed model and its effect on features several years earlier. Testing is done by comparing the accuracy of each model based on the most optimal number of features. In Figure 2, the feature expansion is carried out based on the features of the previous 2 years, testing is carried out in the range of 3 to 26 features. The 2C model is the best model when compared to other 2-year models, with an accuracy of more than 90%. On the 5 features, the resulting accuracy is 96.67%. However, the 5 selected feature sdo not represent the features of the previous 2 years and only contain the feature of 1 year of education history. Based on the comparison of the accuracy of the four 2-year models, model 2C is the best model with an accuracy of 96.67% on the expansion of 10 features. In the expansion of 5 features, the resulting accuracy is the same, but the attributes used do not include the features of the previous 2 years, while in 10 features it includes the features of the previous 2 years containing population size, proportion of male population, elementary school graduates, junior high school graduates, high school graduates, blood type B, and blood type O.

Furthermore, in Figure 3, the previous 1-year feature was added to the 3-year model, so that the features used were 39 features. The test is carried out by comparing the accuracy of each model from 3 to 39 features and selecting the model with the highest accuracy based on the most optimal feature expansion. Model 3A is the best model with an accuracy of 96.67% on 5 features. The selected features already represent the features of the previous 3 years which contain the features of a population, rainfall, blood type B, and graduation from elementary school.

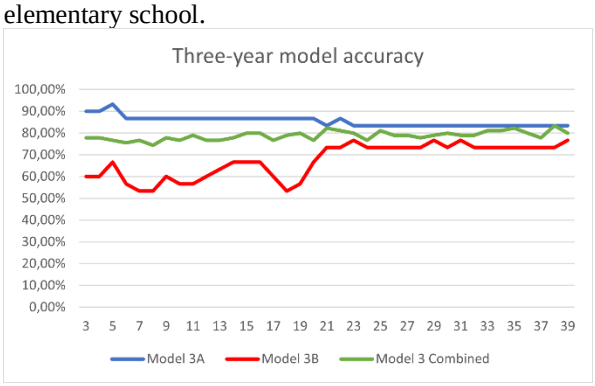

Figure 3. Three-year model accuracy based on feature expansion

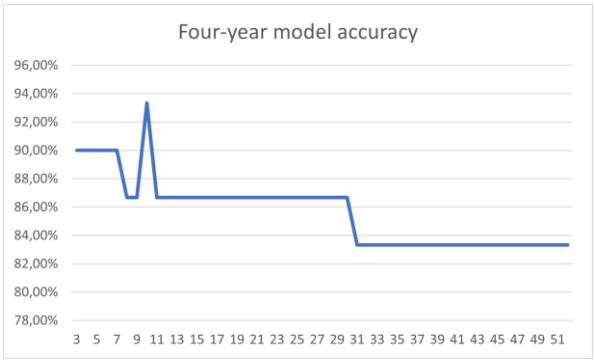

Figure 4. Four-year model accuracy based on feature expansion

In Figure 4, the 4-year model also adds features so that the features used are 52. As before, the test was carried out by comparing the accuracy of each model from 3 to 52 features and selecting the most optimal feature extension. In the expansion of 5 features, it has covered features for 4 years with an accuracy of 90%. Whereas in 10 features the accuracy produced is higher and the features have covered the previous 4 years. Therefore, the best 4-year model is in 10 features with an accuracy of 93.33%. The feature contains the population, the proportion of the male population, elementary school graduates, rainfall, temperature, humidity, blood type A, and blood type O.

Thus, it can be concluded that feature expansion greatly affects the performance of the model and can improve accuracy. In addition, the feature expansion patterns that come out a lot and have the most influence on the spread of dengue incidence rates are population size, proportion of male population, elementary school graduation, rainfall, blood type B, and blood type O.

The three best-selected models are used as models to predict the incidence rate of DHF in 2022, 2023, and 2024. The incidence rate of DHF in 2022 is predicted using the 2C model, the incidence rate in 2023 is predicted using the 3A model, and the incidence rate in 2024 is predicted by model 4A.

DOI: https://doi.org/10.29207/resti.v6i4.4268

3.2 Theoretical Semivariogram Model

Table 8. The results of the calculation of the theoretical semivariogram parameters: nugget, major range, and minor range

Year	Model	Nugget	Major Range	Minor Range
2022	Spherical	0	20739.2	11360.9
	Exponential	0	20739.2	13143
	Gaussian	953.8	20739.2	10868.1
2023	Spherical	1293.6	14386.8	6626.3
	Exponential	629.4	14386.8	4540.5
	Gaussian	1460.8	14386.8	6824.3
2024	Spherical	1570	10717.8	6691.7
	Exponential	1559.6	14386.8	4840.5
	Gaussian	1662.2	10717.8	6824.3

Table 9. RMSE calculation results and theoretical semivariogram parameters: direction (Dir) and partial sill

Year	Model	Dir	Partial Sill	RMSE
2022	Spherical	143.2	5595.2	0.971
	Exponential	147.4	5061.6	0.762
	Gaussian	143.2	5078.8	1.075
2023	Spherical	171.3	798.4	0.996
	Exponential	169.8	1487.4	1.012
	Gaussian	171.9	677.3	0.999
2024	Spherical	171.7	487.7	0.956
	Exponential	169.9	529.9	0.961
	Gaussian	172.2	436.3	0.953

Table 8 and Table 9 are the results of the calculation of the RMSE and the theoretical semivariogram. The results of these calculations are used in the process of predicting the incidence rate of DHF with the ordinary kriging method. Table 8 shows that the data are anisotropic, indicated by the presence of major and minor range parameters. The best theoretical semivariogram model for the incidence rate in 2022, 2023, and 2024, respectively, is the Exponential, Spherical, and Gaussian model, with RMSE values of 0.762, 0.996, and 0.953, respectively. This shows that the semivariogram model for predicting incidence rates in 2022, 2023, and 2024 tends to be different. This difference is influenced by the different prediction results of the random forest model classification. The pattern for the distribution of semivariogram values is shown in Figure 5, Figure 6, and Figure 7.

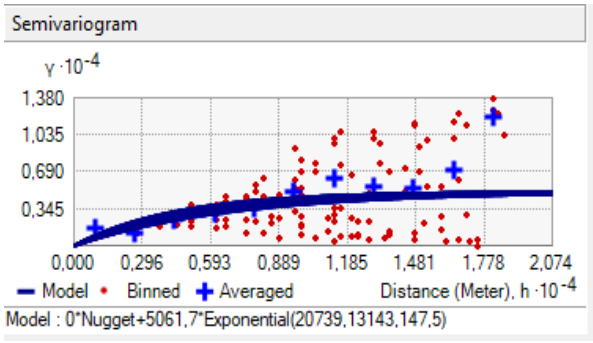

Figure 5. The distribution pattern of the exponential semivariogram incidence rate of dengue fever in 2022

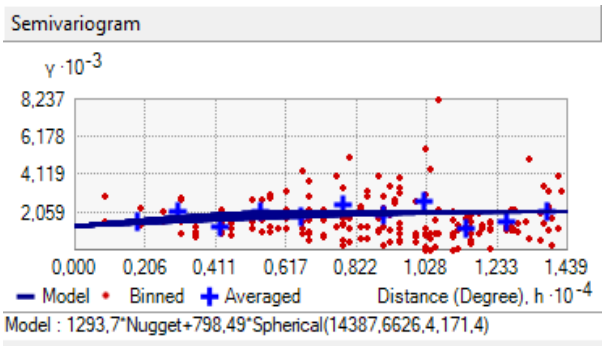

Figure 6. The distribution pattern of the spherical semivariogram incidence rate of dengue fever in 2023

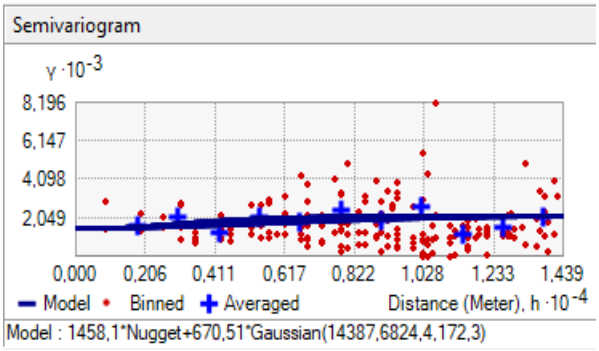

Figure 7. The distribution pattern of the gaussian semivariogram incidence rate of dengue fever in 2024

In Figure 5, it can be seen that the pattern of data distribution tends to the Southwest–Northeast with a value of 147.5. While Figure 6 and Figure 7 have a value of 171, 4, and 172,3 respectively, the data distribution pattern tends to be in the West-East direction. In addition, the incidence rate semivariogram values in Figure 6 and Figure 7 are closer to the average than in Figure 5 which tends to be stretched.

3.3 Prediction of IR DHF using Ordinary Kriging

The predicted pattern of the spread of the incidence rate of DHF is displayed in the form of a color map in Figure 8, Figure 9, and Figure 10. The color shows the interval of incident rate values in the area. There is a color gradation starting with dark blue which indicates the low incidence rate value, then followed by light blue, yellow, orange, pink, and dark red for the highest incidence rate value. Based on the lowest RMSE results in table 9, for the prediction of the incidence rate spread in 2022 the semivariogram model used is Exponential, while in 2023 using the Spherical semivariogram model, and in 2024 using the Gaussian semivariogram model.

On the contour maps of Figures 8 and 9, the area of Cimahi City and the western part of Bandung City is colored light blue to dark blue, so that the incidence rate is in the range of 71 to 149 per 100.000 population. On the other hand, the areas of West Bandung Regency, Bandung Regency, and the northern and eastern parts of Bandung City are yellow to dark red which means the

DOI: https://doi.org/10.29207/resti.v6i4.4268

incidence rate is in the range of 138 to 316 per 100.000 population.

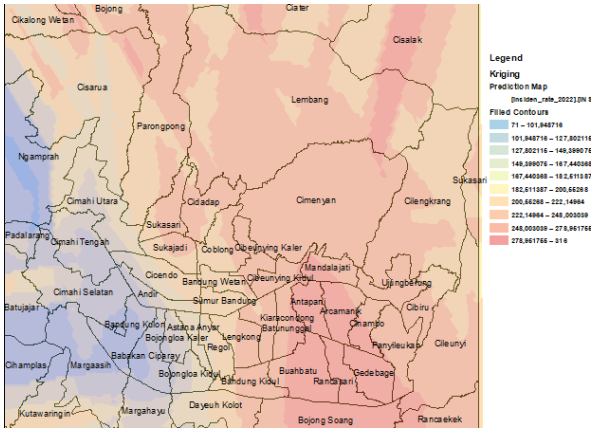

Figure 8. Prediction of the spread of dengue incidence rate in 2022

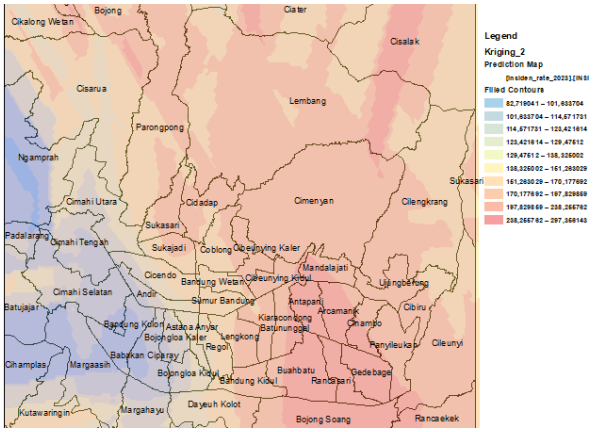

Figure 9. Prediction of the spread of dengue incidence rate in 2023

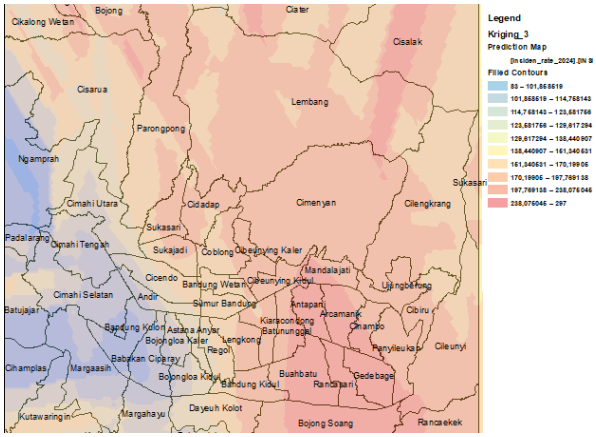

Figure 10. Prediction of the spread of dengue incidence rate in 2024

While in Figure 10, the area of Cimahi City and the western part of Bandung City are light blue and yellow, so the incidence rate is in the range of 83 to 151 per 100.000 population. Then, in the areas of West Bandung Regency, Bandung Regency, and the northern and eastern parts of Bandung City, the colors are yellow to dark red, which means that the incidence rate is in the range of 129 to 297 per 100.000 population. The results of the prediction of the incidence of dengue fever in sub-districts that have not been recorded are presented in table 10.

Table 10. Prediction results of dengue incidence rates in 2022, 2023, and 2024 in districts that have not been recorded

Region		Insiden Rate		
Regency/City	Subdistrict	2022	2023	2024
Subang Regency	Cisalak	247	143	157
Subang Regency	Ciater	226	147	189
Bandung Regency	Cilengkrang	229	150	160
Bandung Regency	Cileunyi	215	158	171
Bandung Regency	Cimenyan	191	163	154
Bandung Regency	Rancaekek	203	164	184
Bandung Regency	Bojongsoang	173	181	169
Bandung Regency	Dayeuhkolot	144	135	145
Bandung Regency	Margahayu	114	112	127
Bandung Regency	Margaasih	113	107	125
Bandung Regency	Kutawaringin	120	87	110
Bandung Regency	Cihampelas	133	108	130
Bandung Regency	Batujajar	99	104	128
Bandung Regency	Padalarang	126	103	132
Cimahi City	Cimahi Utara	120	134	142
Cimahi City	Cimahi Tengah	110	122	135
Cimahi City	Cimahi Selatan	86	115	127
West Bandung Regency	Ngamprah	144	100	139
West Bandung Regency	Cisarua	175	137	172
West Bandung Regency	Parongpong	195	147	152
West Bandung Regency	Lembang	228	165	172
West Bandung Regency	Cikalong Wetan	158	137	156
Sumedang Regency	Sukasari	222	150	142
Purwakarta Regency	Bojong	229	151	154

Based on table 10, the sub-districts in Subang Regency, Bandung Regency, West Bandung Regency, Sumedang Regency, and Purwakarta Regency predict the incidence rate in 2022 is higher when compared to 2023 and 2024. Thus, experiencing a downward trend in 2023, but an increase in 2024. Meanwhile, the predicted incidence rate in the sub-districts of Cimahi City tends to experience an upward trend from 2022 to 2024.

3.4 Discussion

Based on the results of the study, a prediction map for the classification of the distribution of the incidence rate of DHF in each sub-district in Bandung City was made using Random Forest and Ordinary Kriging with their respective advantages and disadvantages. Figure 11, figure 12, and figure 13 show maps created with Ordinary Kriging based on the best semivariogram model. Figure 11 is made with the Exponential model, Figure 12 is made with the Spherical model, and Figure 13 is made with the Gaussian model. While Figure 14, Figure 15, and Figure 16 show a map created with Random Forest based on the feature expansion in the best prediction model.

DOI: https://doi.org/10.29207/resti.v6i4.4268

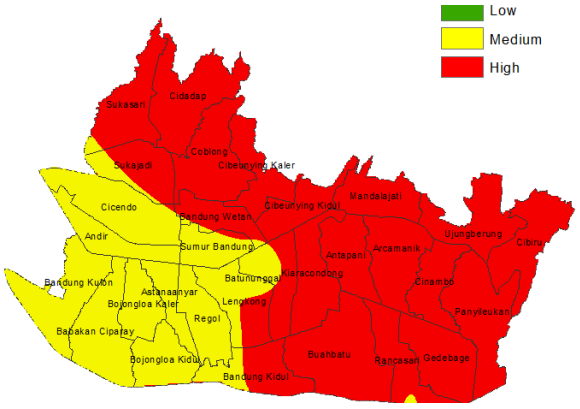

Figure 11. Prediction map for the classification of dengue incidence rates in 2022 with Ordinary Kriging

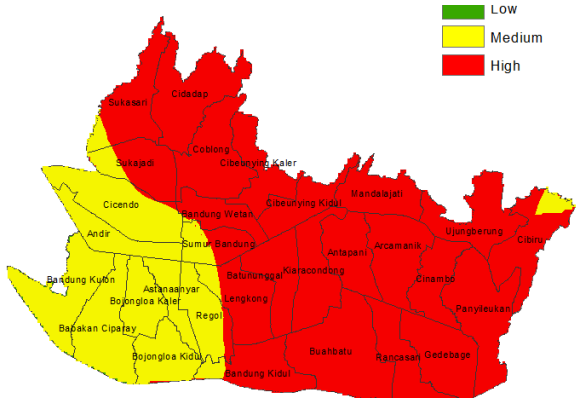

Figure 12. Prediction map for the classification of dengue incidence rates in 2023 with Ordinary Kriging

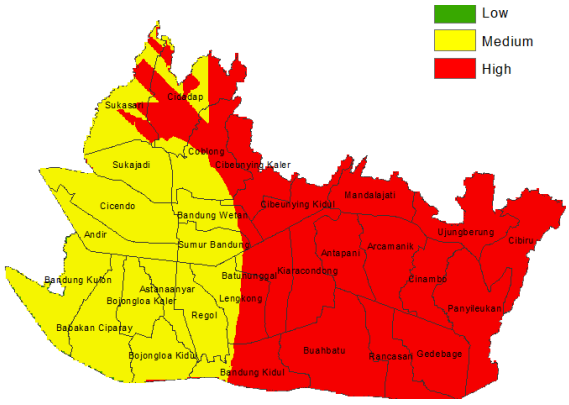

Figure 13. Prediction map for the classification of dengue incidence rates in 2024 with Ordinary Kriging

Figure 11 shows the predicted distribution of the incidence rate of dengue fever in 2022 at 33% for the medium category and 67% for the high category. Meanwhile, Figure 12 shows the distribution of the incidence rate of DHF in 2023 which is predicted to be around 27% for the medium category. This indicates a decrease in the incidence rate of 6% in the medium category and causes an increase in the high category to 73%. Then in Figure 13, the incidence rate distribution in 2024 for the medium category is predicted to experience an upward trend of 10% from 27% to 37%. Meanwhile, for the distribution of incidence rates in the high category, there is a downward trend which is predicted to be at 63%.

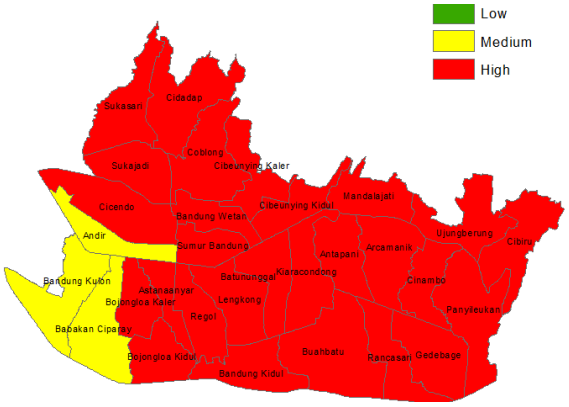

Figure 14. Prediction map of dengue incidence rate classification in 2022 with Random Forest

Figure 14 shows a map of the predicted distribution of dengue incidence rates in 2022 with Random Forest. Of the 30 sub-districts, 3 of them are categorized as a medium class, the remaining 27 sub-districts are categorized as high class. Three sub-districts with the medium class category are the Andir sub-district, Bandung Kulon sub-district, and Babakan Ciparay sub-district. Thus, the prediction of the incidence rate distribution of DHF in 2022 is 10% for the medium class and 90% for the high class. This shows that the distribution of DHF is higher in the Northwest-Southeast to the eastern part of Bandung City.

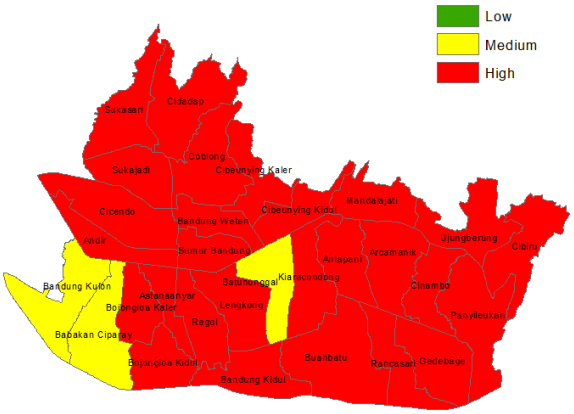

Figure 15. Prediction map of dengue incidence rate classification in 2023 with Random Forest

Based on Figure 14, Figure 15, and Figure 16 show a map of the distribution of incidence rates predicted by Random Forest. The three maps show the results of the incidence rate prediction with a distribution pattern that tends to be the same. The prediction results for the moderate category incidence rate are at 10% and 70 for the distribution of the high category incident rate in 2022, 2023, and 2024. However, when comparing the prediction maps of random forests in 2022 with 2023

DOI: https://doi.org/10.29207/resti.v6i4.4268

and 2024, there are differences in the pattern of regional distribution. In 2022 the incidence rate in Andir District is predicted to be 79 which belongs to the medium category, then in 2023 and 2024 it is predicted that around 124 are included in the high category. This causes an increase in the incidence rate of DHF in Andir District by 49%. Meanwhile, in Batununggal District, the incidence rate decreased by 64%, which is predicted to be in the moderate category in 2023 and 2024.

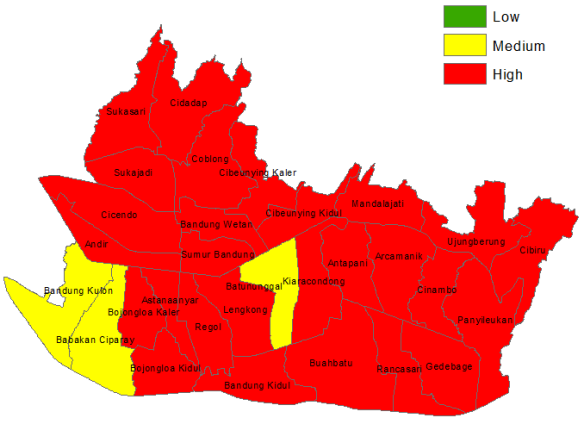

Figure 16. Prediction map of dengue incidence rate classification in 2024 with Random Forest

When comparing the results of the prediction map for the classification of random forest and ordinary kriging, the distribution of the incident rate is moderate in the western and southwest areas of Bandung City and the distribution of the high incidence rate occurs in the eastern area of Bandung City. In addition, for the medium and high categories, ordinary kriging has a lower distribution pattern than random forest. However, the advantages of ordinary kriging can be used to predict the incidence of DHF rates in areas whose values have not been recorded.

The prediction results of random forest classification and ordinary kriging are good enough to display in the form of a map. The random forest classification prediction model developed in this study has better performance than studies [8], [9], [10], and [11]. This is because the random forest model developed in this study applies feature expansion based on several previous years and obtained an accuracy value of 97% in model testing. While the model evaluation results in [8], [9], [10], and [11] have an accuracy value of less than 97%. Thus, feature expansion greatly affects the performance of the random forest model and can increase accuracy. In addition, the map produced by this research is better than research [4], [13], and [15]. This study combines random forest and ordinary kriging methods to produce a prediction map for the distribution of dengue incidence rates for the next three years. Whereas studies [4] and [13] using the random forest method produced prediction maps for one year only. While research [15] developed a map for the next few years using the kriging method, but the resulting map is not based on classification, so it does not know which areas have a DHF incidence rate in the low, medium, and high categories. This study produces a map that predicts the classification of the distribution of dengue incidence rates in the low, medium, and high categories.

4. Conclusion

Based on the research results, it can be concluded that the expansion of attributes based on time in the process of developing a classification prediction model with random forest affects the accuracy produced. The best classification prediction model for DHF with the random forest is based on the previous 2 years, 3 years, and 4 years with the resulting accuracy of 97%, 93%, and 93%, respectively. The model produced a prediction of the classification of the incidence rate of DHF with a moderate class distribution of 10% and a high-class distribution of 90% for 2022, 2023, and 2024. Furthermore, Ordinary Kriging predicted the distribution of incident rates in other locations and 30 sub-districts with RMSE values of 0.762 for 2022, 0.996 for 2023, and 0.953 for 2024. Meanwhile, the most influential features on the spread of dengue disease obtained by expanding features based on time are population, the proportion of the male population, rainfall, blood type B, blood type O, and elementary school graduation. Overall, this research can be used as a reference to reduce the spread of DHF. So that related parties can provide optimal solutions by utilizing the most influential causal factors based on the results of this study to reduce the incidence rate of DHF in each sub-district in Bandung City. For further research, prediction of the spread of DHF can be done by adding datasets, especially per village and other factors that cause DHF, and using other methods as a comparison in prediction and classification to get better accuracy results.

Acknowledgment

The authors would like to thank Telkom University, Bandung City Health Office, Bandung City Central Statistics Agency, and Bandung City Meteorology, Climatology and Geophysics Agency for the support of facilities and infrastructure in providing data and information so that this research can be completed.

Reference

[1] Direktorat Promosi Kesehatan dan Pemberdayaan Masyarakat Kementrian Kesehatan RI, “Demam Berdarah,” *Direktorat Promosi Kesehatan dan Pemberdayaan Masyarakat Kementrian Kesehatan RI*, Jakarta, 2016.

[2] C. Li, X. Wang, X. Wu, J. Liu, D. Ji, and J. Du, “Modeling and projection of dengue fever cases in Guangzhou based on variation of weather factors,” *Sci. Total Environ.*, vol. 605–606, pp. 867–873, Dec. 2017, doi: 10.1016/J.SCITOTENV.2017.06.181.

DOI: https://doi.org/10.29207/resti.v6i4.4268

[3] S. Yuliant, P. Sri Suryani, and S. Iqbal Bahari, "Determination of dengue hemorrhagic fever disease factors using neural network and genetic algorithms/Yuliant Sibaroni, Sri Suryani Prasetiyowati and Iqbal Bahari Sudrajat," *Math. Sci. Informatics J.*, vol. 1, no. 2, pp. 77–86, 2020.

[4] A. Salam, S. S. Prasetiyowati, and Y. Sibaroni, "Prediction Vulnerability Level of Dengue Fever Using KNN and Random Forest," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 4, no. 3, pp. 531–536, Jun. 2020, doi: 10.29207/RESTI.V4I3.1926.

[5] I. Alkhaldy, "Modelling the association of dengue fever cases with temperature and relative humidity in Jeddah, Saudi Arabia—A generalised linear model with break-point analysis," *Acta Trop.*, vol. 168, pp. 9–15, Apr. 2017, doi: 10.1016/J.ACTATROPICA.2016.12.034.

[6] Dinas Kesehatan Kota Bandung, "Profil Kesehatan Kota Bandung Tahun 2019," Bandung, 2019.

[7] Dinas Kesehatan Kota Bandung, "Profil Kesehatan Kota Bandung Tahun 2020," Bandung, 2021.

[8] A. S. Fathima and D. Manimeglai, "Analysis of significant factors for dengue infection prognosis using the random forest classifier," *Int. J. Adv. Comput. Sci. Appl.*, vol. 6, no. 2, pp. 240–245, 2015.

[9] R. Arafiyah, F. Hermin, I. R. Kartika, A. Alimuddin, and I. Saraswati, "Classification of Dengue Haemorrhagic Fever (DHF) using SVM, naive bayes and random forest," in *IOP Conference Series: Materials Science and Engineering*, 2018, vol. 434, no. 1, p. 12070.

[10] S. Khan *et al.*, "Random Forest-Based Evaluation of Raman Spectroscopy for Dengue Fever Analysis," *Appl. Spectrosc.*, pp. 1–7, 2017, doi: https://doi.org/10.1177/0003702817695571.

[11] P. Silitonga, B. E. Dewi, A. Bustamam, and H. S. Al-Ash, "Evaluation of Dengue Model Performances Developed Using Artificial Neural Network and Random Forest Classifiers," *Procedia Comput. Sci.*, vol. 179, pp. 135–143, 2021, doi: https://doi.org/10.1016/j.procs.2020.12.018.

[12] M. Shahid Ansari *et al.*, "Identification of predictors and model for predicting prolonged length of stay in dengue patients," *Health Care Manag. Sci.*, vol. 24, no. 4, pp. 786–798, Dec. 2021, doi: 10.1007/S10729-021-09571-3/TABLES/4.

[13] J. Ong *et al.*, "Mapping dengue risk in Singapore using Random Forest," *PLoS Negl. Trop. Dis.*, vol. 12, no. 6, p. e0006587, Jun. 2018, doi: 10.1371/JOURNAL.PNTD.0006587.

[14] L. Mao, L. Yin, X. Song, and S. Mei, "Mapping intra-urban transmission risk of dengue fever with big hourly cellphone data," *Acta Trop.*, vol. 162, pp. 188–195, Oct. 2016, doi: 10.1016/J.ACTATROPICA.2016.06.029.

[15] S. S. Prasetiyowati and Y. Sibaroni, "Prediction of DHF disease spreading patterns using inverse distances weighted (IDW), ordinary and universal kriging," in *Journal of Physics: Conference Series*, 2018, vol. 971, no. 1, p. 12010.

[16] T. L. Schmidt *et al.*, "Local introduction and heterogeneous spatial spread of dengue-suppressing Wolbachia through an urban population of Aedes aegypti," *PLOS Biol.*, vol. 15, no. 5, p. e2001894, May 2017, doi: 10.1371/JOURNAL.PBIO.2001894.

[17] M. C. P. Parra *et al.*, "Using adult Aedes aegypti females to predict areas at risk for dengue transmission: A spatial case-control study," *Acta Trop.*, vol. 182, pp. 43–53, Jun. 2018, doi: 10.1016/J.ACTATROPICA.2018.02.018.

[18] C. Lorenz *et al.*, "Remote sensing for risk mapping of Aedes aegypti infestations: Is this a practical task?," *Acta Trop.*, vol. 205, p. 105398, May 2020, doi: 10.1016/J.ACTATROPICA.2020.105398.

[19] L. Sedda *et al.*, "The spatial and temporal scales of local dengue virus transmission in natural settings: A retrospective analysis," *Parasites and Vectors*, vol. 11, no. 1, pp. 1–14, Feb. 2018, doi: 10.1186/S13071-018-2662-6/TABLES/2.

[20] P. J. Tsai, T. H. Lin, H. J. Teng, and H. C. Yeh, "Critical low temperature for the survival of Aedes aegypti in Taiwan," *Parasites and Vectors*, vol. 11, no. 1, pp. 1–14, Jan. 2018, doi: 10.1186/S13071-017-2606-6/TABLES/2.

[21] Direktorat Jenderal Pengendalian Penyakit dan Penyehatan Lingkungan, *Pedoman Pencegahan dan Pengendalian Demam Berdarah Dengue di Indonesia*. Jakarta: Kementrian Kesehatan RI, 2017.

[22] Y. Zeng, K. Jiang, and J. Chen, "Automatic seismic salt interpretation with deep convolutional neural networks," *ACM Int. Conf. Proceeding Ser.*, pp. 16–20, Apr. 2019, doi: 10.1145/3325917.3325926.

[23] E. G. Adagbasa, S. A. Adelabu, and T. W. Okello, "Application of deep learning with stratified K-fold for vegetation species discrimation in a protected mountainous region using Sentinel-2 image," *https://doi.org/10.1080/10106049.2019.1704070*, vol. 37, no. 1, pp. 142–162, 2019, doi: 10.1080/10106049.2019.1704070.

[24] C. M. Yeşilkanat, "Spatio-temporal estimation of the daily cases of COVID-19 in worldwide using random forest machine learning algorithm," *Chaos, Solitons & Fractals*, vol. 140, p. 110210, 2020.

[25] G. Biau and E. Scornet, "A random forest guided tour," *Test*, vol. 25, no. 2, pp. 197–227, Jun. 2016, doi: 10.1007/S11749-016-0481-7.

[26] R. Irmanita, S. S. Prasetiyowati, and Y. Sibaroni, "Classification of Malaria Complication Using CART (Classification and Regression Tree) and Naïve Bayes," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 1, pp. 10–16, Feb. 2021, doi: 10.29207/RESTI.V5I1.2770.

[27] B. George, "A study of the effect of random projection and other dimensionality reduction techniques on different classification methods," *Baselius Res.*, p. 201769, 2017.

[28] T. Desyani, A. Saifudin, and Y. Yulianti, "Feature Selection Based on Naive Bayes for Caesarean Section Prediction," in *IOP Conference Series: Materials Science and Engineering*, 2020, vol. 879, no. 1, p. 12091.

[29] O. I. Sheluhin and V. P. Ivannikova, "Comparative analysis of informative features quantity and composition selection methods for the computer attacks classification using the unsw-nb15 dataset," *Т-Comm-Телекоммуникации и Транспорт*, vol. 14, no. 10, 2020.

[30] S. S. Prasetiyowati, M. Imrona, I. Ummah, and Y. Sibaroni, "Prediction of public transportation occupation based on several crowd spots using ordinary Kriging method," *J. Innov. Technol. Educ.*, vol. 3, no. 1, pp. 93–104, 2016.

[31] S. K. Adhikary, N. Muttil, and A. G. Yilmaz, "Genetic Programming-Based Ordinary Kriging for Spatial Interpolation of Rainfall," *J. Hydrol. Eng.*, vol. 21, no. 2, p. 04015062, Sep. 2015, doi: 10.1061/(ASCE)HE.1943-5584.0001300.

[32] H. Wackernagel, "Ordinary Kriging," *Multivar. Geostatistics*, pp. 79–88, 2003, doi: 10.1007/978-3-662-05294-5_11.

DOI: https://doi.org/10.29207/resti.v6i4.4268