Spatial Accessibility Index and Health Facility Location Suitability Based on POI and Machine Learning in DIY Province

Fat'hul Mubin Gufron[1], Setia Pramana[1*]
[1]Department of Statistical Computing, Politeknik Statistika STIS, Jakarta, Indonesia
*Corresponding author: setia.pramana@stis.ac.id

Received: 15 June 2026
Revised: 10 August 2026
Accepted: 11 August 2026

ABSTRACT – The unequal distribution of health facilities is a major cause of inequality in accessibility, especially in hard-to-reach rural areas. Therefore, equitable development of health facilities is needed, targeting both access and location suitability. This study aims to measure spatial accessibility and identify suitable locations for health facility development to support efficient, equitable, and targeted planning. The method used includes calculating the accessibility index using Enhanced Two Step Floating Catchment Area (E2SFCA) and location suitability modeling with machine learning algorithms, as well as analysis of the relationship between the two using scatterplots. Random forest demonstrated the best performance with an accuracy of 85.71%. Identification of the relationship between accessibility measurements and suitability modeling successfully identified 85 villages with low access but high suitability, which are recommended as priority locations for health facility development. These findings are expected to form the basis for evidence-based planning to achieve equitable distribution of health services in the Special Region of Yogyakarta Province.

Keywords– Spatial Accessibility, Health Facilities, Point Of Interest, Machine Learning, Location Suitability.

I. INTRODUCTION

According to [1] health is a state of a person who is free from illness, whether physical, mental, or social. Health is a basic human right and one of the elements of well-being that must be fulfilled. This is stated in Article 28 H paragraph 1 which states "Everyone has the right to live in physical and spiritual prosperity, to have a place to live and have a good and healthy environment, and has the right to receive health services" [2].

Rapid global population growth, accompanied by increasing expectations and quality of life, demands continuous improvement in the quality of supporting sectors. The health sector is a crucial element because it is closely linked to social and economic development and community welfare [3]. In its implementation, health facilities are needed that include promotive, preventive, curative and rehabilitative services, and are supported by equal access for all levels of society in accordance with SDGs goal 3.[1].

According to [4], health facilities (faskes) consist of primary and secondary levels. Primary health facilities, namely community health centers, are health facilities that handle primary health problems. Meanwhile, advanced health facilities, namely hospitals and clinics. Advanced health facilities handle specialist and subspecialist health problems.

The government continues to encourage the development of healthcare facilities, as outlined in the 2024–2029 National Medium-Term Development Plan (RPJMN) to address the disparity in healthcare infrastructure. According to the 2023 Ministry of Health Performance Report, the provision of primary healthcare facilities has only reached 63.54%, and only 78.35% of hospitals have received full accreditation [5]. Data from CEO World, based on the 2023 World Healthcare Index, also shows that Indonesia's healthcare infrastructure remains low, with a score of 64.37.

The unequal distribution of health facilities results in limited access to services, particularly in rural areas. Rural residents must travel further than those living in urban areas. [6] This is exacerbated by geographical conditions that tend to make access to health facilities difficult. [7] As a result, many deaths occur due to delayed medical treatment. Statistics Indonesia (BPS) data from 2021 recorded that approximately 20% of deaths are caused by delays in services. A 2019 report from the Ministry of Health's Crisis Center stated that more than 18,000 deaths occur annually due to limited access to healthcare facilities. Furthermore, 70% of emergency patient deaths reportedly occur en route to the hospital.

The Special Region of Yogyakarta (DIY) still faces challenges in equitable distribution of health facilities, despite having the third-highest hospital bed ratio and the seventh-highest community health center ratio nationally according to data from the Ministry of Health [8]. However, this high ratio does not reflect equitable distribution and accessibility. This is reflected in the achievement of health indicators, where the southern regions of DIY, such as Bantul, Kulon Progo, and Gunungkidul, show lower achievements than Sleman and Yogyakarta City [9]. Human Development Index (HDI) data supports this finding, with the average HDI in the southern region in 2023 being 74.95, lower than the provincial average of 81.09. In addition, indicators such as Life Expectancy (UHH), Maternal Mortality Rate (MMR), Infant Mortality Rate (IMR), and Under-Five Mortality Rate (FMR) in the southern region are also recorded as worse than the provincial average. Studies [10] and [11] also show that health facilities are still concentrated in Yogyakarta City, while coastal areas experience limited access.

To understand the inequality in access to health services, it is necessary to pay attention to the spatial dimension, namely the availability and accessibility of facilities, as explained by Penchansky and Thomas [12]. These two dimensions form the basis for compiling the spatial accessibility index in this study [13].

In development planning, identifying areas with limited access is an important initial step[14], so that strong and accurate accessibility measurements are needed[15] which are also important in determining the location of new health services[16]. However, the development of health facilities in Indonesia is faced with complex geographical challenges. so that planning must be adaptive to regional conditions and take environmental risks into account. Various studies also emphasize the importance of a multi-criteria approach in selecting the location of health facilities, taking into account social, economic, environmental aspects and future needs to improve the effectiveness and quality of services[17]. However, to date there has not been much research that integrates spatial accessibility measurements with location suitability analysis simultaneously, so that the opportunity to produce more comprehensive and targeted priority health facility development policies is still open[16].

In practice, location suitability mapping (site suitability analysis) is generally done using conventional methods such asMulti-Criteria Decision Analysis(MCDA), which integrates various factors such as topography, population density, and environmental risks[17]. However, as technology advances, the MCDA-based approachmachine learningstarting to be applied in spatial analysis, including in location suitability assessments (site suitability analysis). Algorithms such asMulti Layer Perceptron, Random Forest, Support Vector Machine (SVM), And CatBoosthas been proven to be able to handle the complexity of spatial data and produce more accurate predictions[17], [18].

On the other hand, accessibility measurement methods have also evolved. Early approaches such as calculating straight-line distances or facility-to-population ratios, as well as gravity-based models, were deemed incapable of fully representing actual spatial conditions because they ignored aspects of service demand and the effects of distance decay [3]. Therefore, access measurement methods are needed that are not only mathematical, but also spatial and dynamic. One widely used approach is theTwo-Step Floating Catchment Area(2SFCA). However, this method still has weaknesses, namely that it does not consider distance decay and only looks at a single catchment area [3]. Therefore, [19] developed the method.Enhanced Two-Step Floating Catchment AreaE2SFCA takes into account actual travel times, spatial weighting based on distance decay, and tiered catchment zone division, thus providing a more realistic and contextual representation of accessibility.[20].

In this study, health facility data were obtained from Point of Interest (PoI) sources due to their ability to represent spatial patterns of human activities, including public facilities and essential services. PoI data are real-time, easily accessible, large-scale, and more efficient than conventional survey methods. Google Maps was selected as the primary data source because it provides comprehensive and up-to-date information that can be accessed through an Application Programming Interface (API). This platform has been widely utilized in spatial studies, including in developing countries such as Iran, to support health service planning [21]. Furthermore, the use of big data in collecting variables related to location suitability enables the processing of spatial information at finer spatial resolutions, such as grid-based units. Big data offers advantages in terms of speed, diversity, and comprehensiveness, making it an important foundation for developing geographic information systems based on actual data to support health facility planning [22]. Therefore, this innovation is expected to provide evidence-based support for the Yogyakarta Provincial Health Office in formulating policies aimed at achieving a more equitable, targeted, and accurate distribution of health facility development. Based on this background, this study aims to calculate the spatial accessibility index of health facilities in each village in the Special Region of Yogyakarta Province, map the suitability of locations for health facility development, and identify the relationship between the spatial accessibility index of health facilities and the suitability of locations for establishing new health facilities.

II. LITERATURE REVIEW

In several previous studies the methodfloating catchment area is one of the most widely used methods for calculating the spatial accessibility index of healthcare facilities. This is because the method has undergone gradual refinement. Furthermore, the calculation takes into account not only geographic aspects but also demographics and the population ratio.supply-demand.There are several studies that use this, including [12], [23], [24] which use the 2SFCA method, in these studiesthreshold The time zone used is only 1, namely less than 30 minutes and more than 30 minutes. Meanwhile, in research [13], [19], [20], [25], [26] using the E2SFCA method which hasthreshold The time zone is divided into 3, namely 0-10 minutes, 10-20 minutes, and 20-30 minutes, each time zone has a weight. The weight used is the calculation of the results of the decay of the distance function, the weights in E2SFCA are 1, 0.68, 0.22 [19].

Meanwhile, research has been conducted several times to map and identify the suitability of health facility development locations. The methods used in mapping the suitability of health facility development locations are divided into two, namely:weigth sum model And machine learning.Research using the methodweigh sum and AHP was used in research conducted by [27][28][29]. In this research, the method used was AHP, which was carried out to determine hierarchical priorities and then continued to calculate the weight for each variable. The indicators used in this research were different, research [28] used indicators of comfort, accessibility, and disaster risk. [27] used aspects of accessibility, environment, and other public facilities. [29] used aspects of population, topography, and available health facilities. Meanwhile, in other research that usedmachine learningconducted by [17],[18],[30]. The variables used in the research vary. In research [18] the indicators used were demographics, environment, and topography. [30] used environmental, demographic, topographic, and existing health facility indicators. [17] used accessibility, demographic, topographic, and environmental indicators.

III. METHODOLOGY

This study utilizes village-level data in the Special Region of Yogyakarta Province for the year 2024 to analyze and construct a spatial accessibility index for primary and advanced healthcare services. The study area encompasses all five administrative districts/cities within the province. The data used in this research are categorized into two groups according to the study objectives, namely data for calculating the spatial accessibility index and data for mapping the suitability of locations for health facility development. Detailed information regarding the data sources and variables used in each stage of the analysis is presented in table 1 and table 2.

Table 1 Data for calculating the spatial accessibility index

Data	Source	Reference
Distance and Travel Time	OpenRouteService (*Isochrone)*	[19], [31]
Total Population	BPS and Wolrdpop	[19], [31]
Health Facility Point	Google Maps (*Point Of Interest)*and the Yogyakarta Provincial Health Office	[19], [20]
Health workers	Yogyakarta Provincial Health Office	[3], [19]

Table 2 Data for mapping the suitability of locations for health facility development

Data	Source	Reference
LST	MODIS	[18], [32]
Population Density	BPS and Worldpop	[18]
NDVI	Landsat-8	[28]
AOD	Sentinel-5p	[18], [28]
Road Distance	*OpenStreetMap*	[17], [28], [29]
River Distance	*OpenStreetMap*	[18]
Pollutan (SO2,NO2,O3,CO)	Sentinel-5p	[28], [32]
THREE	NASA SRTM/DEMNAS	[30]
TWI	NASA SRTM/DEMNAS	[30]
Slope	NASA SRTM/DEMNAS	[17]
Height	NASA SRTM/DEMNAS	[17]
Windspeed	ECMWF ERA5	[28], [33]
Rainfall	CHIRPS	[28], [33]
FRUIT	ECMWF ERA5	[28]
Health Facility Point	*Point Of Interest*	[18]
Random point (non-health facility point)	*Point Of Interest*	[18]
LST	MODIS	[18], [32]

Before further analysis is carried out, the entire dataset goes through the following stages: preprocessing to ensure quality, consistency, and suitability for modeling needs. This process is performed differently for each type of data.

In Point of Interest (PoI) data, preprocessing includes eliminating duplication due to differences in spelling of facility names, validation with official data from relevant agencies to ensure accuracy, and adding location coordinates through a manual geotagging process [34].

In official statistical data, the number of health workers is aggregated into one total capacity value at each health facility according to the classification of health workers in Law No. 17 of 2023. In addition, to estimate the 2024 population density at the 1 km × 1 km grid level, a proportional downscaling approach was employed. Since WorldPop provides gridded population data only up to 2020, while the official 2024 population data are available at the village level, the WorldPop grid values were proportionally adjusted using the ratio between the official village population totals and the corresponding WorldPop village population totals. This approach preserves the spatial distribution pattern represented by the WorldPop dataset while ensuring that the total population within each village is consistent with the official 2024 demographic statistics, thereby producing gridded population estimates suitable for subsequent spatial analysis.

Meanwhile, satellite image data is processed through several stages: (1) cloud masking to remove pixels covered by clouds; (2) cloud selection by selecting images with less than 10% cloudiness [35]; (3) reducing using the mean method to produce a stable and representative annual composite [36]; and (4) cropping or cutting the image according to the study area boundaries so that the analysis focuses on the relevant area.

A. Analysis Method

1) E2SFCA (Enhanced Two-Step Floating Catchment Area)

In the calculation using E2SFCA, there are two steps used. The first is calculating the ratio supply-demand for each village to health facilities that can be reached within 3 time zones (0-10 minutes, 10-20 minutes, 20-30 minutes). The following is the equation for calculating the ratio supply-demand.

$$\mathbf{Rj} = \frac{\mathbf{Sj}}{\mathbf{\Sigma k \, \epsilon\{d(k,j) \leq Dr\} Pk \, Wr}} \qquad (1)$$

Where Sj is the health worker at a health facility j, Pk is the population of village k, dkj is the travel time between k and j. Wr is the time zone weight and Dr is the boundary.threshold each time zone (r = 1,2,3). Where the weights are 1, 0.68, 0.22 [19]

The second is to do a cumulative addition for the ratio.suppy-demand to obtain the spatial accessibility index of health facilities. The second step can be expressed in the following equation.

$$\mathrm{Af} = \sum\nolimits_{j \, \epsilon\{d(k,j) \leq dr\}} \mathrm{RjWr} \qquad (2)$$

After obtaining the health facility accessibility index value, this value can be categorized into three index classes using natural breaks (jenks) which aims to minimize variation within categories and maximize differences between categories. This approach allows for a more objective distribution of index values based on actual data distribution patterns, so that the resulting categories truly reflect field conditions.

2) Spatial Autocorrelation

Spatial autocorrelation is measured using Global Moran's I. In measuring spatial autocorrelation, the following hypothesis is used.

$H0 : I = 0$ (No spatial autocorrelation)

$H1 : I \neq 0$ (There is spatial autocorrelation)

In this test reject$H0$ if the value $|Z(I)| > \mathrm{WITH}_{\alpha/2}$. The Moran index value ranges from -1 to 1, a positive Moran index value indicates positive spatial autocorrelation. Adjacent areas with similar values tend to cluster, and a negative Moran index value indicates negative spatial autocorrelation, meaning adjacent locations have different values [37]. Subsequently, Local Indicators of Spatial Association (LISA) were employed to identify local spatial clustering patterns. The statistical significance of the Local Moran's I values was evaluated at a significance level of $\alpha = 0.05$. Only statistically significant locations were classified into High–High (HH), Low–Low (LL), High–Low (HL), and Low–High (LH) clusters, whereas non-significant observations were not interpreted as meaningful local spatial clusters.

3) Mapping the Suitability of Health Facility Locations

Data used for model training consisted of 294 health facility locations obtained from Google Maps Point of Interest (POI), which were labeled as suitable locations, and an equal number of randomly generated non-health facility points labeled as unsuitable locations. The negative samples were randomly generated within areas having elevations greater than 300 m above sea level and slopes exceeding 10°, following the site suitability criteria proposed by [29] and [17]. These constraints were applied to reduce the likelihood of selecting environmentally suitable locations as negative samples while maintaining spatial realism. In addition, only one random point was generated within each spatial grid to avoid duplicate observations and representation bias, and the number of negative samples was balanced with the positive samples to mitigate class imbalance [30]. Modeling was performed at a 1 km × 1 km grid level. The optimal hyperparameters for each machine learning algorithm were determined using Grid Search. Model performance was then evaluated using stratified five-fold cross-validation to provide a consistent and balanced framework for comparing the predictive performance of all evaluated algorithms under identical experimental conditions. Although the study area exhibits significant spatial autocorrelation, conventional five-fold cross-validation was adopted to facilitate comparative model evaluation. Therefore, the reported performance should be interpreted primarily as a relative comparison among algorithms rather than as an absolute estimate of spatial generalization performance. Future studies are encouraged to employ spatial block cross-validation or other spatially explicit resampling techniques to account for spatial dependence during model validation. Algorithm machine learning which is used is Random forest, Support Vector Machine, Catboost, And Multi Layer perceptron. Random Forest is a combination of several decision trees that have the same distribution for each tree. Because it comes from several combined decision trees, random forests can reduce overfitting model from decision tree[38].SVM is a learning system that has the principle of identifyinghyperplane The best model is the boundary between the two classes. The best model is obtained when each class has the furthest distance to thehyperplane[39].Catboost one implementation of the methodgradient boosting by using binary decision treeas the basic predictor.[40]Catboost very sensitive tohyperparameter so that it can improve performance[41].

4) Scatter plot the relationship between accessibility index and suitability of health facility location.

Scatter plot can be used to see the relationship between two variables [42]. The suitability value at the 1 km × 1 km grid level is first aggregated to the village level by calculating the proportion of grids that fall into the moderate to very high category. Next, this proportion is combined with the village accessibility index value and visualized in a scatter plot. Village classification was performed by using the arithmetic mean of the accessibility index and site suitability score as the threshold to divide the scatterplot into four quadrants. Unlike the Natural Breaks (Jenks) classification, which was applied solely for cartographic visualization of the accessibility index, the quadrant analysis was intended to distinguish villages according to whether their values were above or below the overall average of each variable. Therefore, the arithmetic mean was adopted as a common reference point to facilitate comparative analysis and policy prioritization.[43], thus forming four quadrants: (1) high access–high suitability, (2) low access–high suitability, (3) high access–low suitability, and (4) low access–low suitability. Villages in the low access–high suitability quadrant are identified as development priorities, because they indicate an urgent need for health services as well as the availability of suitable locations. Meanwhile, villages with low access–low suitability need special attention from the government through more targeted interventions, either in the form of improving infrastructure or providing additional health facilities, so that service disparities can be minimized.

IV. RESULTS AND DISCUSSIONS

1) Calculation of Spatial Accessibility Index with E2SFCA

a) Visualization of Spatial Accessibility Index of Health Facilities

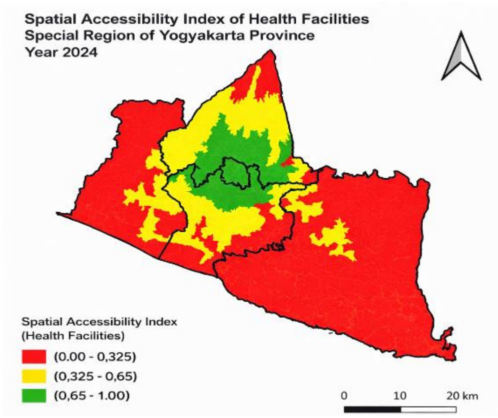

Figure 1 Visualization of Spatial Accessibility Index of Health Facilities

Figure 1 shows quite clear disparities. Green accessibility index values are concentrated in the city center, particularly Yogyakarta City and parts of Sleman and Bantul. Yellow accessibility values are spread across transitional areas surrounding urban areas, while red accessibility values dominate the outskirts, particularly Gunungkidul, Kulon Progo, and southern Bantul. These findings confirm that urban communities have relatively easy access to healthcare services, while rural areas still face significant limitations, necessitating priority policy interventions to ensure equitable healthcare distribution.

b) Visualization of Spatial Accessibility Index of Health Facilities

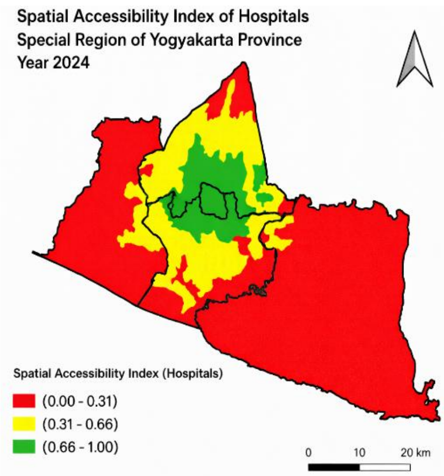

Figure 2 Visualization of Spatial Accessibility Index of Hospital

Figure 2 shows that the cities of Yogyakarta and Sleman have relatively high hospital spatial accessibility index values. Meanwhile, most areas of Yogyakarta Special Region, particularly Gunungkidul, Kulon Progo, and southern Bantul, are in the low category. This situation indicates that hospitals are still heavily concentrated in the city center, requiring residents in the outlying areas to travel longer distances and take longer to reach.

c) Visualization of Hospital Spatial Accessibility Index

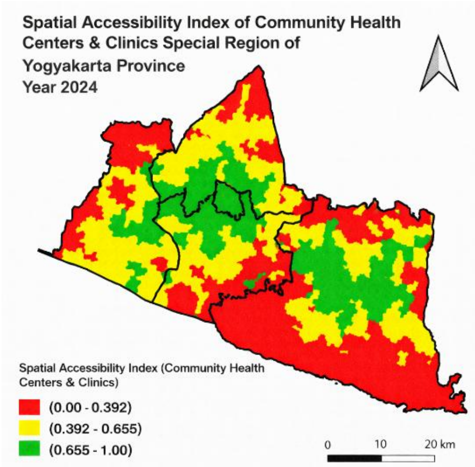

Figure 3 Visualization of Spatial Accessibility Index of Health Centers & Clinics

Figure 3 shows a relatively more even distribution compared to hospitals. Several rural areas, including parts of Gunungkidul and Kulon Progo, show moderate to high access. This can be explained by the role of community health centers and clinics, which are more widespread at the sub-district and village levels, thus serving as the main support for basic health services. However, there are still villages with low access, especially in mountainous areas and on provincial borders.

Evaluation of the index using the correlation between the average travel time to health facilities and the index score yielded a value of -0.60, indicating a strong negative relationship and indicating that the higher the accessibility score, the lower the travel time.

2) Spatial Autocorrelation

Table 3 Moran's Global Results I

	Health facilities	Hospital	Community Health Center &clinic
P- Value	0,001	0,001	0,001
Moran's Index	0.9265	0.9438	0,7416

The results of the spatial autocorrelation test using Global Moran's I show that the accessibility index of health facilities in the DIY Province in 2024 has a significant spatial pattern with a value of p-value 0.001 across all health facility categories. The high and positive Moran's I values, namely 0.9265 for combined health facilities, 0.9438 for hospitals, and 0.7416 for community health centers/clinics, confirm that the distribution of accessibility is not random, but rather forms spatial clusters. This means that villages with high accessibility tend to be close to other villages with high access, and villages with low access form clusters in peripheral areas. This pattern indicates systematic spatial inequality, where urban areas enjoy a concentration of health services, while peripheral and rural areas experience limited access. This finding underscores the importance of spatially based health policy planning, prioritizing interventions in low-access clusters to promote equitable service distribution and reduce disparities between regions. However, the global Moran's I can only describe patterns globally. Therefore, further analysis is needed with Local indicator Of Spatial Association.

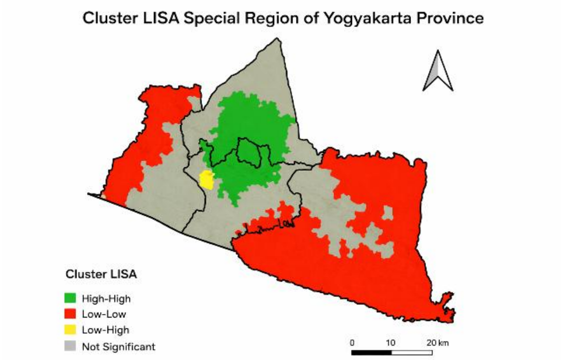

Figure 4 Significance map *Local Indicators of Spatial Association*

DOI: 10.12962/j27213862.v9i2.10083

Based on Figure 4, a significant spatial clustering pattern can be identified in the health facility accessibility index. The areas in green (High-High) shows a cluster of areas that have a high accessibility index and are surrounded by areas with high scores consisting of 105 villages. This indicates the potential for inequality because areas with high access tend to be concentrated in one area. Meanwhile, the red cluster (Low-Low) consists of 131 villages, indicating the existence of areas that are truly spatially isolated from health services, spread mainly in peripheral areas such as the south and east. This indicates spatial inequality in the distribution of health facilities.

These findings have important implications, as they suggest that efforts to equalize healthcare facilities need to be focused on areas with Low-Low clusters to prevent them from falling further behind urban areas. Meanwhile, the presence of Low-High areas suggests specific local barriers, such as limited road access or unfavorable geographic conditions, so policy interventions need to be more contextualized according to regional characteristics. Thus, LISA analysis plays a crucial role in providing a more targeted basis for prioritizing development, both through the addition of new healthcare facilities and increasing the accessibility of supporting infrastructure in low-access areas.

3) Mapping the suitability of health facility development locations

Parameter selection/hyperparameterbest done usinggrid searchwith evaluation5-fold cross validation.

Table 4 Model Evaluation machine Learning

Model	Accuracy	F1-Score	Recall	Precision
Multi Layer Perceptron	0.8125	0.8091	0.6909	0.9048
Support Vector Machine	0.8036	0.8013	0.7091	0.8667
Random Forest	**0.8571**	**0.8542**	**0.7273**	**0.9756**
Catboost	0.8482	0.8454	0.7273	0.9524

Based on the results in Table 4, the four algorithmsmachine learningshowed quite good performance in mapping the suitability level of health facility development locations. Among the four, the algorithmRandom Forest (RF)provide the best classification results with valuesaccuracyhighest. To strengthen the basis for classification between suitable and unsuitable locations, further evaluation was carried out through a comparison of the curves. ROC (Receiver Operating Characteristic) and AUC value (Area Under the Curve) of each model. The evaluation results show that Random Forest has the highest AUC value, namely 0.86 with threshold optimal is 0.71.

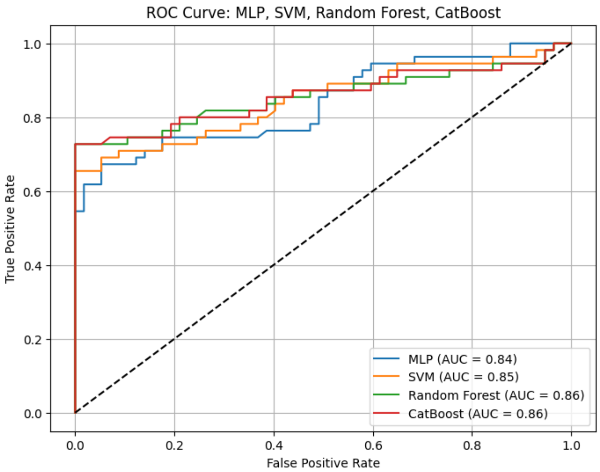

Figure 5 ROC – AUC Curve

From the modeling results *random forest* obtained *feature importance* in Figure 6.

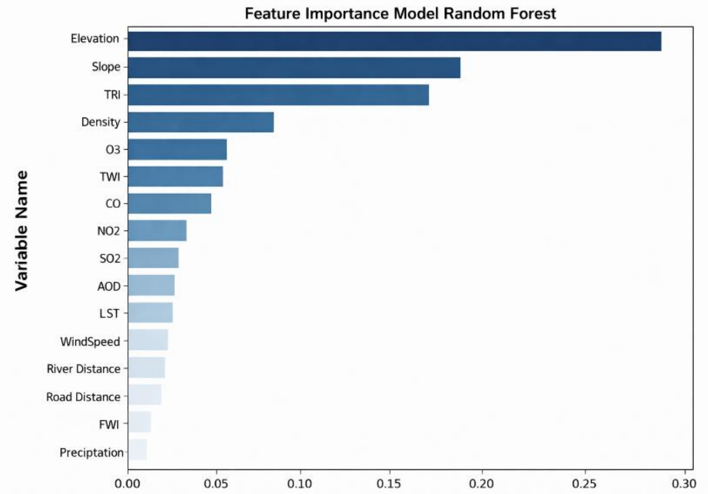

Figure 6 Feature importance model random forest

Figure 6 shows the visualization results, feature importance Random Forest model based on values Mean Decrease in Impurity(MDI) for the classification of suitability of health facility development locations. These results indicate that the variable Elevation(regional height) has the most dominant contribution in forming model decisions, followed by Slope and TRI (Topographic Roughness Index). These three are topographic indicators that play an important role in determining the physical suitability of land for development. Variables Density (population density) also stands out, emphasizing the importance of the demographic dimension in addressing health service needs.

Furthermore, environmental variables such as O_3, TWI, and NO_2 have a moderate influence, reflecting the air quality, soil moisture, and environmental stress factors taken into account. Meanwhile, variables such as precipitation, FWI, and distance to rivers and roads contribute relatively little to the model, but remain relevant as complementary variables in a spatial context.

Mapping the suitability of health facility locations is classified into five suitability categories: very low, low, medium, high, and very high. The classification method used is to divide the results probability score into 5 classes with the same number of members (quantile method) This is in accordance with research conducted by [17]. Figure 8 shows the results of mapping the suitability of health facility locations.

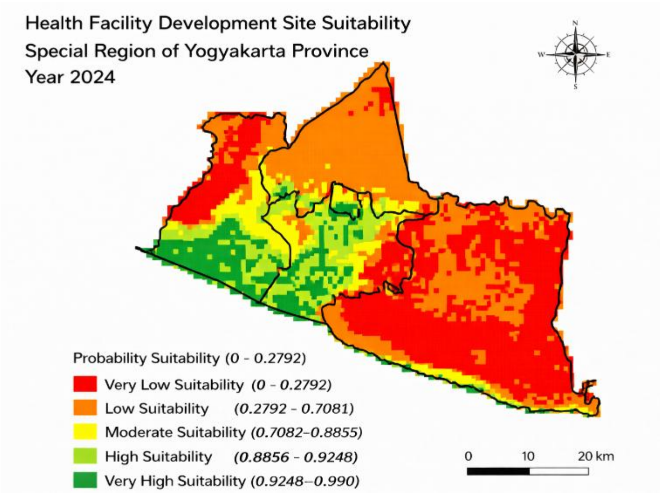

Figure 7 Feature importance model random forest

Based on the mapping results of the suitability of health facility development locations in Figure 7, the Yogyakarta Special Region (DIY) is divided into five categories, from very low to very high. The very low and low categories are generally found in the south and southeast, particularly Gunungkidul and parts of Bantul and Kulon Progo, which are characterized by steep topography, low population density, and limited basic infrastructure. Conversely, the high and very high categories are concentrated in urban areas such as Yogyakarta City, parts of Bantul, and southern Kulon Progo, which have higher population densities, good transportation access, and adequate infrastructure support. The moderate category is often found in peri-urban areas as transition zones, which still have potential but require additional intervention to optimize health facility development.

4) The relationship between the accessibility index and the suitability of the location of health facilities

Before doing *plot* First, adjustments are made for both variables. First, the accessibility index is processed *scaling* because there is inequality in the index values. Second, the suitability variables obtained at the grid level are aggregated to the village level.

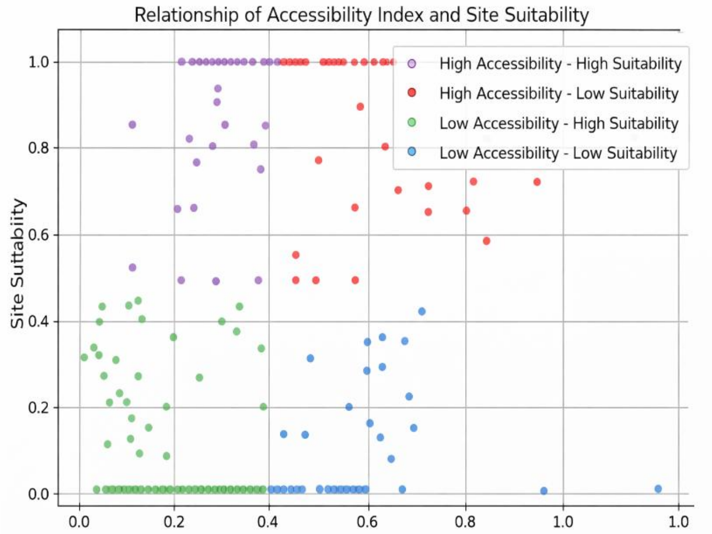

Figure 8 Scatterplot Accessibility Index and Location Suitability

Results from Figure 8 Four categories were obtained, with the accessibility and suitability index variables divided into low and high categories. The boundary between the high and low categories was based on the mean of each variable. The following is an explanation of the categories.

1. The category of villages with high accessibility and suitability reflects areas where healthcare facilities have been optimally located. There are 127 villages in this category, covering most of Yogyakarta City and the urban centers of Bantul, Kulon Progo, and Gunungkidul. This indicates that healthcare facility development in these areas has met its targets, supported by good access and favorable biophysical characteristics. Therefore, these areas are not a top priority for new healthcare facility development.
2. Villages with high accessibility but low suitability indicate that, despite easy access to health facilities, their location selection has not properly considered biophysical and environmental conditions. Forty-two villages fall into this category, including Muja Muju, Pandeyan, Rejowinangun, and Sorosutan in Yogyakarta City. This situation reflects planning that focuses on accessibility but fails to fully consider multidimensional aspects such as topography or potential risks. Therefore, risk mitigation strategies are needed to ensure the sustainability and safety of the facilities constructed.
3. Villages with low accessibility but high suitability demonstrate strategic potential for the development of new health facilities. Eighty-five villages fall into this category, including Gadingsari and Poncosari (Bantul Regency), Merdikorejo and Glagahharjo (Sleman Regency), and Semanu and Ponjong (Gunungkidul Regency). These areas possess favorable biophysical and environmental characteristics, yet still face limited access to health services. Therefore, villages in this category are recommended as a top priority in efforts to improve equitable access to health services in the Special Region of Yogyakarta.
4. Villages with low accessibility and suitability represent areas facing dual barriers: limited access and biophysical conditions that are less conducive to the development of health facilities. There are 184 villages in this category, spread across the regencies of Gunungkidul, Kulon Progo, Sleman, and Bantul. Nevertheless, the development of health facilities still needs to be considered in these areas because there are communities that continue to require services, despite geographic and environmental challenges. Possible approaches include development with specialized infrastructure that adapts to extreme conditions, such as terrain-resistant building structures, dedicated road access, and logistics distribution systems that support accessibility. Furthermore, integration with mobile services or telemedicine can complement, not replace, the need to ensure that all communities, including those living in difficult areas, maintain equal access to health services.

V. CONCLUSIONS AND SUGGESTIONS

In Based on the results and discussion in the previous section, here are some things that can be concluded:

1. Calculation of the spatial accessibility index for health facilities, taking into account demographic and spatial factors, yields three accessibility categories. In general, villages with high accessibility index values tend to be located in urban centers, while those with low accessibility values are often found in outlying areas. Calculations using Global Moran's I indicate positive spatial autocorrelation in the accessibility of health facilities, hospitals, and community health centers/clinics. This indicates that accessibility values between villages are positively correlated and form a spatial clustering pattern, where villages with similar accessibility levels tend to be geographically close. The presence of positive spatial autocorrelation confirms that health facility accessibility is not randomly distributed but rather forms a spatial clustering pattern. Villages with high accessibility levels tend to cluster in urban areas, while villages with low accessibility are also concentrated in outlying areas. This condition demonstrates that disparities in health access are systemic and regional, not just differences between individual villages. The implication is that health development interventions need to be designed within the scope of regional clusters, particularly by prioritizing areas that form low-access clusters, to reduce disparities in access to services and more effectively realize the principle of equitable health services.
2. In modeling the suitability of health facility locations using machine learning, the best model evaluation using confusion matrix and ROC AUC Curve was obtained, which was random forest which produced performance of accuracy of 85.7%, precision of 97.56%, recall of 72.73%, and F1-score of 85.42%. Analysisfeature importancerevealed that the variables of height, slope, population density (density), And (Topographic Roughness Index) (TRI) is a dominant factor in determining the suitability of a health facility location. This emphasizes the importance of combining demographic factors and the physical conditions of a region in planning the development of health services.
3. The results of the scatterplot between the spatial accessibility index of health facilities and the suitability of the location of health facilities obtained results that there are 127 villages that have high accessibility and high suitability, 184 villages have low accessibility and low suitability, 85 villages have low accessibility and high suitability, 42 villages have high accessibility and low suitability. The results of the analysis show that the distribution of health facilities is not yet even. A total of 127 villages already have high access and suitability, indicating targeted development. However, there are 85 villages with low access and high suitability, which need to be prioritized for the construction of new facilities. Meanwhile, 184 villages are in conditions of low access and suitability, so they require special approaches, such as mobile health services or increased connectivity. However, the gradual development of health facilities still needs to be carried out by considering the characteristics and geographical limitations, to ensure the sustainability and equality of health services in the region.

This study has several limitations that can serve as suggestions for further research development. The following are some suggestions for further research.

1. Expanding the scope of variables, particularly by incorporating socio-economic dimensions in greater depth, such as poverty levels, education, or health insurance participation status for calculating the spatial accessibility index.
2. Conduct further studies on the categorization of spatial accessibility indexes.
3. Calculation of distance and travel time using several modes of transportation, so that they can be compared.
4. Mapping the suitability of the location for building health facilities is adjusted to the conditions in the field, if the area is a forest area that is far from residential areas, it does not need to be considered.

REFERENCES

[1] Undang Undang, “Undang-Undang Republik Indonesia No 17 Tahun 2023 tentang Kesehatan,” no. 187315, pp. 1–300, 2023.

[2] A. A. Pradana, C. Casman, R. Rohayati, and M. Kamal, “Program Universal Health Coverage (Uhc) Di Indonesia,” *J. Endur.*, vol. 7, no. 2, pp. 462–473, 2022, doi: 10.22216/jen.v7i2.1363.

[3] H. Faraji Sabokbar, H. Mohammadi, S. Tahmasbi, Y. Rafii, and A. Hosseini, “Measuring spatial accessibility and equity to healthcare services using fuzzy inference system,” *Appl. Geogr.*, vol. 136, no. October, p. 102584, 2021, doi: 10.1016/j.apgeog.2021.102584.

[4] Permenkes No 24 Tahun 2020, “Peraturan Menteri Kesehatan No 24 Tahun 2020,” vol. 1969, no. 1, pp. 1–24, 2008.

[5] Kemenkes RI, “Laporan Kinerja Kementerian Kesehatan Tahun 2023,” *J. GEEJ*, vol. 7, no. 2, 2023.

[6] T. I. Shah, S. Milosavljevic, and B. Bath, “Measuring geographical accessibility to rural and remote health care services: Challenges and considerations,” *Spat. Spatiotemporal. Epidemiol.*, vol. 21, pp. 87–96, 2017, doi: 10.1016/j.sste.2017.04.002.

[7] S. Mishra, P. K. Sahu, A. K. Sarkar, B. Mehran, and S. Sharma, “Geo-spatial site suitability analysis for development of health care units in rural India: Effects on habitation accessibility, facility utilization and zonal equity in facility distribution,” *J. Transp. Geogr.*, vol. 78, no. October 2018, pp. 135–149, 2019, doi: 10.1016/j.jtrangeo.2019.05.017.

[8] Kementrian Kesehatan, *Profil Kesehatan*. 2023.

[9] D. Dinas Kesehatan DIY, “Profil Kesehatan Provinsi Daerah Istimewa Yogyakarta,” *Buku Data 2023*, p. 41, 2024.

[10] M. I. Sadali, M. A. F. Alfana, Z. Hadijah, E. L. Rosewidiadari, and R. Andika, “Dominasi kota sebagai konsentrasi fasilitas kesehatan (Studi kasus: Daerah Istimewa Yogyakarta),” *Reg. J. Pembang. Wil. dan Perenc. Partisipatif*, vol. 17, no. 1, p. 136, 2022, doi: 10.20961/region.v17i1.44948.

[11] N. Yatiman and A. Pujiyono, “Analisis Efisiensi Teknis Anggaran Belanja Sektor Kesehatan Pemerintah Daerah

Kabupaten/Kota Di Provinsi Daerah Istimewa Yogyakarta Tahun 2008-2010," *Diponegoro J. Econ.*, vol. 2, pp. 1–13, 2013.

[12] M. S. Kilinc, A. B. Milburn, and J. L. Heier Stamm, "Measuring potential spatial accessibility of home healthcare services," *Socioecon. Plann. Sci.*, vol. 59, pp. 13–25, 2017, doi: 10.1016/j.seps.2016.09.007.

[13] M. Husna and J. S. Tarigan, "Aksesibilitas Spasial Ke Pusat Kesehatan Masyarakat(Puskesmas) Dengan Metode Enhance Two-Step Floating Catchment Area (E2Sfca) Studi Kasus Di Kota Medan," *Kumpul. J. Ilmu Komput.*, vol. 08, no. 3, pp. 338–351, 2021.

[14] N. Wan, F. B. Zhan, B. Zou, and E. Chow, "A relative spatial access assessment approach for analyzing potential spatial access to colorectal cancer services in Texas," *Appl. Geogr.*, vol. 32, no. 2, pp. 291–299, 2012, doi: 10.1016/j.apgeog.2011.05.001.

[15] S. C. G. Paul L Delamater, Joseph P Messina, Ashton M Shortridge, "Measuring geographic access to health care: raster and network-based methods," *Int. J. Health Geogr.*, pp. 1029–1030, 2012.

[16] B. C. Lin *et al.*, "Spatial decision on allocating automated external defibrillators (AED) in communities by multi-criterion two-step floating catchment area (MC2SFCA)," *Int. J. Health Geogr.*, vol. 15, no. 1, pp. 1–14, 2016, doi: 10.1186/s12942-016-0046-8.

[17] K. Y. Almansi, A. Rashid, M. Shariff, and A. F. Abdullah, "applied sciences Learning Approaches : Evidence from the Gaza Strip in Palestine," pp. 1–22, 2021.

[18] R. Virtriana *et al.*, "Development of location suitability prediction for health facilities using random forest machine learning in 2030 integrating remote sensing and GIS in West Java, Indonesia," *Environ. Adv.*, vol. 19, no. August 2024, p. 100604, 2025, doi: 10.1016/j.envadv.2024.100604.

[19] W. Luo and Y. Qi, "An enhanced two-step floating catchment area (E2SFCA) method for measuring spatial accessibility to primary care physicians," *Heal. Place*, vol. 15, no. 4, pp. 1100–1107, 2009, doi: 10.1016/j.healthplace.2009.06.002.

[20] A. Al-Omari, N. Shatnawi, and A. Al-Mashaqbeh, "Use of an E2SFCA method to assess healthcare resources in Jordan during COVID-19 pandemic," *Egypt. J. Remote Sens. Sp. Sci.*, vol. 25, no. 4, pp. 1057–1068, 2022, doi: 10.1016/j.ejrs.2022.11.007.

[21] F. Rajabi, F. Hosseinali, and H. Rabiei-Dastjerdi, "An examination and analysis of the clustering of healthcare centers and their spatial accessibility in Tehran metropolis: Insights from Google POI data," *Sustain. Cities Soc.*, vol. 117, p. 105845, 2024, doi: 10.1016/j.scs.2024.105845.

[22] S. Dash, S. K. Shakyawar, M. Sharma, and S. Kaushik, "Big data in healthcare: management, analysis and future prospects," *J. Big Data*, vol. 6, no. 1, 2019, doi: 10.1186/s40537-019-0217-0.

[23] Z. Tao, Y. Cheng, and J. Liu, "Hierarchical two-step floating catchment area (2SFCA) method: Measuring the spatial accessibility to hierarchical healthcare facilities in Shenzhen, China," *Int. J. Equity Health*, vol. 19, no. 1, pp. 1–16, 2020, doi: 10.1186/s12939-020-01280-7.

[24] A. Vo, M. Plachkinova, and R. Bhaskar, "Assessing healthcare accessibility algorithms: A comprehensive investigation of two-step floating catchment methodologies family," *2015 Am. Conf. Inf. Syst. AMCIS 2015*, no. September, 2015.

[25] C. Drake *et al.*, "A comparison of methods for measuring spatial access to health care," *Health Serv. Res.*, vol. 56, no. 5, pp. 777–787, 2021, doi: 10.1111/1475-6773.13700.

[26] N. L. Leyso and M. Umezaki, "Spatial inequality in the accessibility of healthcare services in the Philippines," *GeoJournal*, vol. 89, no. 3, pp. 1–14, 2024, doi: 10.1007/s10708-024-11098-3.

[27] B. Halder, J. Bandyopadhyay, and P. Banik, "Assessment of hospital sites' suitability by spatial information technologies using AHP and GIS-based multi-criteria approach of Rajpur–Sonarpur Municipality," *Model. Earth Syst. Environ.*, vol. 6, no. 4, pp. 2581–2596, 2020, doi: 10.1007/s40808-020-00852-4.

[28] T. S. Anggraini, N. M. K. Anantri, R. P. Pratama, M. A. E. Rahadianto, A. Syetiawan, and A. D. Sakti, "MULTI-SCENARIO SPATIAL MODELING of HEALTH FACILITIES DEVELOPMENT to SUPPORT the ACHIEVEMENT of UNIVERSAL HEALTH COVERAGE," *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch.*, vol. 46, no. M-2–2022, pp. 7–13, 2022, doi: 10.5194/isprs-Archives-XLVI-M-2-2022-7-2022.

[29] Q. A. Nsaif, S. M. Khaleel, and A. H. Khateeb, "Integration of GIS and remote sensing technique for hospital site selection in Baquba District," *J. Eng. Sci. Technol.*, vol. 15, no. 3, pp. 1492–1505, 2020.

[30] K. Y. Almansi, A. R. M. Shariff, B. Kalantar, A. F. Abdullah, S. N. S. Ismail, and N. Ueda, "Performance Evaluation of Hospital Site Suitability Using Multilayer Perceptron (MLP) and Analytical Hierarchy Process (AHP) Models in Malacca, Malaysia," *Sustain.*, vol. 14, no. 7, 2022, doi: 10.3390/su14073731.

[31] H. Widayanti, "Keterjangkauan Spasial Puskesmas Di Pusat Kota Semarang Menggunakan Isochrone," *J. Litbang Sukowati Media Penelit. dan Pengemb.*, vol. 6, no. 1, pp. 54–67, 2022, doi: 10.32630/sukowati.v6i1.323.

[32] B. S. Alotaibi, S. Lo, E. Southwood, and D. Coley, "Evaluating the suitability of standard thermal comfort approaches for hospital patients in air-conditioned environments in hot climates," *Build. Environ.*, vol. 169, p. 106561, 2020, doi: 10.1016/j.buildenv.2019.106561.

[33] A. D. Sakti *et al.*, "Spatial Prioritization for Wildfire Mitigation by Integrating Heterogeneous Spatial Data: A New Multi-Dimensional Approach for Tropical Rainforests," *Remote Sens.*, vol. 14, no. 3, 2022, doi: 10.3390/rs14030543.

[34] K. Patroumpas, D. Skoutas, G. Mandilaras, G. Giannopoulos, and S. Athanasiou, "Exposing points of interest as linked geospatial data," *ACM Int. Conf. Proceeding Ser.*, pp. 21–30, 2019, doi: 10.1145/3340964.3340976.

[35] X. Gao, H. Chi, J. Huang, Y. Han, Y. Li, and F. Ling, "Comparison of Cloud-Mask Algorithms and Machine-Learning Methods Using Sentinel-2 Imagery for Mapping Paddy Rice in Jianghan Plain," *Remote Sens.*, vol. 16, no. 7, 2024, doi: 10.3390/rs16071305.

[36] N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore, "Google Earth Engine: Planetary-scale geospatial analysis for everyone," *Remote Sens. Environ.*, vol. 202, pp. 18–27, 2017, doi: 10.1016/j.rse.2017.06.031.

[37] R. Pratiwi, "Pengujian Hasil Dependensi Menggunakan Moran's I dan Geary's C pada Pemodelan Geographycally Weighted Regression (Studi Kasus pada Upah Minimum di Provinsi Jawa Timur)," pp. 1–53, 2019.

[38] L. Breiman, "Random Forests," *Kluwer Acad. Publ.*, vol. 45, pp. 5–32, 2001, doi: 10.1007/978-3-030-62008-0_35.

[39] C. Zhang, Y. Liu, and N. Tie, "Forest Land Resource Information Acquisition with Sentinel-2 Image Utilizing Support Vector Machine, K-Nearest Neighbor, Random Forest, Decision Trees and Multi-Layer Perceptron," *Forests*, vol. 14, no. 2, 2023, doi: 10.3390/f14020254.

[40] A. A. Ibrahim, R. L. Ridwan, M. M. Muhammed, R. O. Abdulaziz, and G. A. Saheed, "Comparison of the CatBoost Classifier with other Machine Learning Methods," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 11, pp. 738–748, 2020, doi: 10.14569/IJACSA.2020.0111190.

[41] J. T. Hancock and T. M. Khoshgoftaar, "CatBoost for big data: an interdisciplinary review," *J. Big Data*, vol. 7, no. 1, 2020, doi: 10.1186/s40537-020-00369-8.

[42] M. Friendly and D. Denis, "The early origins and development of the scatterplot," *J. Hist. Behav. Sci.*, vol. 41, no. 2, pp. 103–130, 2005, doi: 10.1002/jhbs.20078.

[43] Fu, Y., Zhao, G., Yu, C., Li, Z., Li, C., Shan, J., Gao, Z., Huang, Z., Guo, G., Liu, S., & Liu, H. (2025). Study on strategies to improve clinical fellow training satisfaction based on importance-performance analysis (IPA). *BMC medical education*, *25*(1), 1137. https://doi.org/10.1186/s12909-025-07625-6