Available online at www.CivileJournal.org

Civil Engineering Journal

(E-ISSN: 2476-3055; ISSN: 2676-6957)

Vol. 10, No. 12, December, 2024

Data Mining Approach-Based Damage Identification for Asphalt Pavement Under Natural Disaster Conditions

Andri I. Rifai [1*], Muhammad Isradi [2], Joewono Prasetijo [3], Yusra A. Sari [1], Muhammad F. Zolkepli [4]

[1] Faculty of Civil Engineering and Planning, Universitas Internasional Batam, Batam, Indonesia.

[2] Department of Civil Engineering, Universitas Mercu Buana, Jakarta 11650, Indonesia.

[3] Faculty of Engineering, Universiti Tun Hussein Onn Malaysia, 84600 Panchor, Johor, Malaysia

[4] Faculty of Civil Engineering, Universiti Teknologi Malaysia, 83130 Skudai, Johor, Malaysia.

Received 13 September 2024; Revised 12 November 2024; Accepted 19 November 2024; Published 01 December 2024

Abstract

Road performance can also decline due to natural disasters such as earthquakes, often in Indonesia. Given the high risk of natural disasters in Indonesia, it is important to consider their impact. Therefore, it is necessary to prepare for road rehabilitation and reconstruction quickly and accurately. This research aims to identify potential factors causing road damage by developing an approach to obtain predictions of road damage levels due to natural disasters by utilizing the availability of historical data, developing a decision support system to rehabilitate and reconstruct roads after disasters, and developing a road damage model due to earthquakes using data mining. The data was used to assess the condition of the national road pavement in Central Sulawesi and identified the disaster events as earthquakes that originated from the USGS. Data processing uses a data mining (DM) approach, which includes three models. The results found that the SVM modeling with the DM approach had a high accuracy rate of 0.91 ± 0.01, RMSE 0.70 ± 0.02, and MAD 0.42 ± 0.01. SVM achieves the highest accuracy after 20 runs. The best hyperparameters to accomplish a fit SVM model are $\epsilon = 0.07 \pm 0.01$ and $\gamma = 0.05 \pm 0.00$. Meanwhile, for ANN, the hyperparameters are $H = 3 \pm 1$. The earthquake's magnitude (27%) and depth (24%) contribute to road damage.

Keywords: Natural Disaster; Data Mining; Asphalt Pavement; Road Maintenance; Road Damage.

1. Introduction

Road construction is conducted simultaneously in several places to achieve the set targets. In this sense, road construction organizers and all stakeholders must continue developing themselves to preserve, extend, and enhance development progress [1, 2]. However, the simultaneous implementation process frequently requires improvement. One of them is the occurrence of natural and non-natural disasters that need serious attention [3]. Failure at this stage will impact the field's overall development process [4]. Natural disasters frequently occur in Indonesia; as evidence, the Ministry of Public Works and Housing recorded 1,257 floods, 931 tornadoes, 566 landslides, 250 forest fires, 22 earthquakes, 22 tidal waves, and four droughts in 2022 alone [5]. This natural disaster has resulted in various risks of human losses, such as displacing 3,781,976 people, injuring 831 people, killing 195 people, and missing 28 people. In addition, the most visible infrastructure damage was 7,765,967 submerged buildings, 270 damaged bridges, 519

* Corresponding author: andri.irfan@uib.ac.id

http://dx.doi.org/10.28991/CEJ-2024-010-12-015

educational facilities, 320 worship facilities, 76 health facilities, and various other infrastructure [6]. Therefore, they must quickly mitigate and deal with the multiple impacts of disasters to ensure the smooth operation of government services [7].

The state is responsible for handling disaster impacts and ensuring people can do their activities safely. Controlling the effects of disasters that require attention, among other things, means maintaining the connectivity of road networks and bridges [8]. One example is the occurrence of extraordinary disasters, namely the Palu, Sigi, and Donggala earthquakes on September 28, 2018. These natural disasters caused significant damage with extensive impacts [9]. When the tragedy occurred, infrastructure development was progressing relatively rapidly. However, this incident caused a cessation of construction and maintenance of infrastructure. This earthquake is a complete phenomenon due to fault movements, tsunamis, avalanches, and catastrophic liquefaction events, the most significant global phenomenon ever [10]. Liquefaction has attracted attention from the people of Indonesia and even the world since the mudflow event during the liquefaction devastated infrastructure and dwellings on a considerable scale and generated a direct incident [11].

A disaster can reduce road performance or cause it to cease functioning abruptly [12]. Generally, the lifespan of the road pavement experiences depreciation, ranging from the time of construction and use until it is considered damaged (ends of design life) [10]. The decline in road performance is proportional to the level of disaster disturbance. Indeed, to conduct rehabilitation and reconstruction planning, it is necessary first to decrease road performance [13]. Therefore, it should model various data and the scale of disaster-induced damage to road infrastructure to estimate the level of damage in subsequent incidents.

The ability to predict road damage due to a disaster using previous damage data is essential to optimizing a road pavement management system. It should leverage the extensive data in the pavement management system to aid in developing the model [14]. DM's ability to interpret accurately is supported by a vast capacity to read available data that must be used well. Historical data from natural disaster events that resulted in suspected road damage can be used to accurately assess and forecast data for future activities. Historical data from natural disaster events that resulted in suspected road damage can be used to accurately assess and forecast data for future activities. They lose their significance if the research fails to accurately understand and predict large data sets [15, 16]. In this regard, it is required that the model can offer a reasonable approach to the interpretation process. Data mining (DM) is frequently applied for data interpretation in many fields. In using artificial intelligence (AI), DM has tremendous promise in assisting interpretation and prediction [17]. They are using AI in the transportation science scientific group to improve their approach to conducting sustainable analysis [18].

Developing an optimal model can provide alternative recommendations for maintenance, cost control, and work implementation. The optimal road maintenance pattern system is critical for planning maintenance, predicting future work, and determining its implementation costs. The research results can assist the Indonesian Ministry of Public Works and Public Housing, planning consultants, and implementing contractors in optimizing post-disaster road repairs. They can also support other studies related to this topic. Furthermore, policymakers require a specific method to develop a decision support system model and determine post-disaster pavement management steps [19]. Cross-disciplinary research development can be an option for solving these problems. Several approaches to algorithm development, like AI, DM, and MOO, can be used to create a support system for modern policymakers [20].

2. Material and Methods

The research takes place on national roads in the province of South Sulawesi. Data collection uses secondary data such as road network data, natural disaster information, traffic information, pavement conditions, maintenance activities, and road maintenance history [21]. Standard specifications can be used to determine the activity specifications. Road pavement conditions are obtained from DGH road performance data and field survey results after natural disasters [22]. The working method framework, which includes the components described above, is shown in Figure 1:

2.1. Data Source

The primary source of data on pavement condition is the IIRMS. This data is a historical record of the state of the roads, their functionality, and other pertinent details, including AADT, ESAL, potholes, cracks, and roughness. Data on road conditions from the IIRMS cover the years 2000–2022. The DM technique is applied to estimate skewed or missing data in the database; however, some data must still be filled. The IIRMS and the findings of direct measurements made at the scene of the natural catastrophe provided the data used in this investigation [23-25].

2.2. Stages of Road Performance Prediction

Researchers will use disaster data from the Indonesian island of Sulawesi to construct this model and a DM-based road performance prediction technique. Data is categorized based on disastrous occurrences for calibration, learning, testing, and validation [26]. Validation of road conditions and detailed coordinates was done by collecting data directly with Hawkeye in the last five years before the disaster. A road damage prediction model using the DM approach will be developed in this study, considering the input data shown in Table 1 without any assumptions of restrictions.

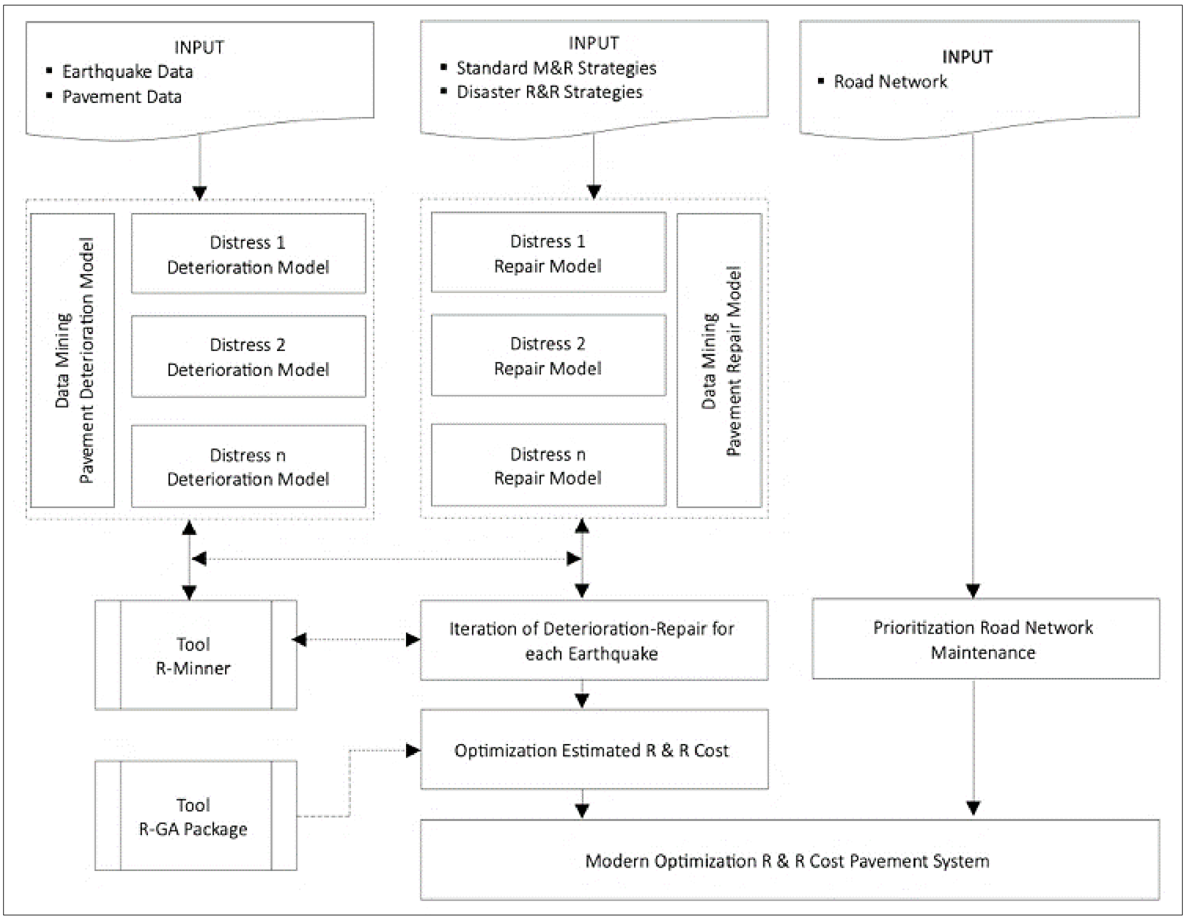

Figure 1. Methodology Framework of Optimization DSS

Table 1. Type of Data and Source

Data Type	Details	Source
Pavement Condition	▪ IRI ▪ Alligator Cracking (m^2) ▪ Longitudinal Cracking (mm) ▪ Block Cracking (m^2) ▪ Irregular Cracking (m^2) ▪ Potholes number (no/km) ▪ Rutting (m^2) ▪ Depression (mm) ▪ Deformation (m^2) ▪ Structural Patching (m^2) ▪ Seal Patching (m^2),	DGH
Reconstruction & Rehabilitation History	Resurfacing & Reconstruction schedule	DGH
Earthquakes	▪ Magnitude ▪ Epicenters ▪ Depth of seismic ▪ Time ▪ Duration ▪ Repetition	USGS DGH BMKG
Cost	Maintenance cost per km Reconstruction cost per km Rehabilitation cost per km	DGH
Performance	Performance index	DGH & literature study

The fundamental equations, the variables included in the model, and how these variables are grouped all have an impact on the model's structure. The basic equation is an equation that uses optimized parameters. The software calculation process requires other factors that the researcher must determine beforehand. For example, ANN uses several hidden layers and equation-type algorithms [27]. In some forms of modeling, data must be appropriately formatted or grouped before it can be used in model development. In addition, some models can be developed with only a tiny part of the entire database. The remaining available data will be used to evaluate the model's performance, shown in Figure 2:

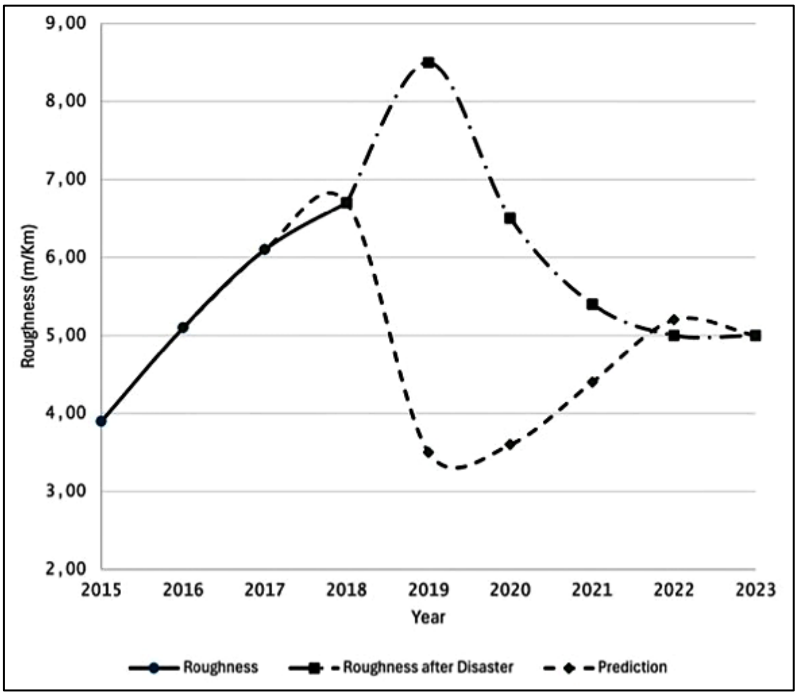

Figure 2. Snapshot of DGH Monitoring Console

Meanwhile, the R-Tool with the R-Miner Package is used for the DM model, and the interface display of the application can be seen in Figure 3:

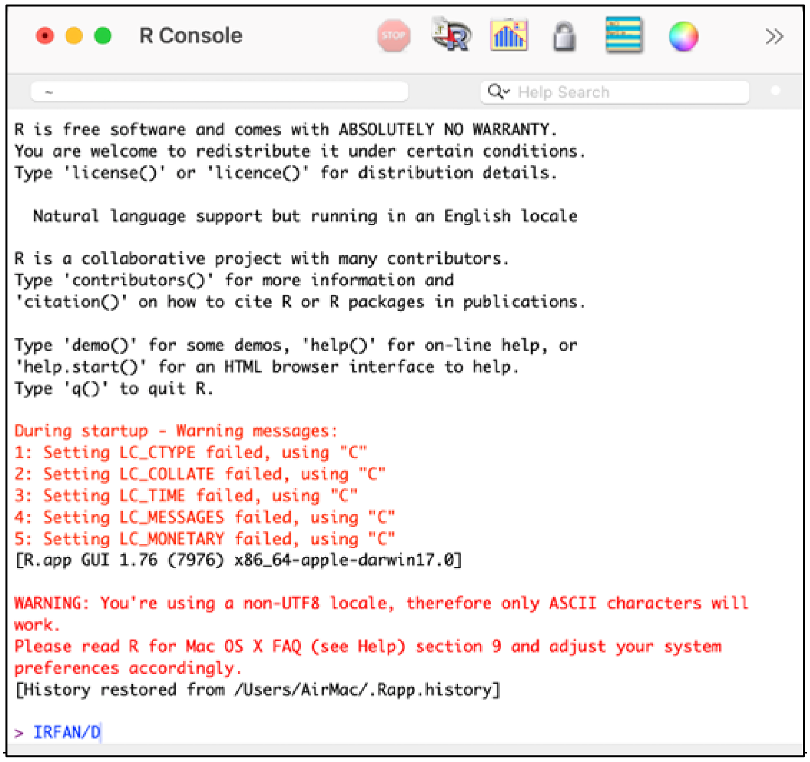

Figure 3. Snapshot of R Console

Model testing includes re-evaluating whether the model suits the stated goals. Testing can be completed by comparing the model's output with the other models' outputs using the same data. This process ensures that the model works within the scope of the road pavement management system and the expected critical value range. If the model is not considered feasible based on the evaluation results, it is necessary to re-evaluate the model type. If the evaluation still deems the model type incorrect, it becomes necessary to redevelop and alter its shape. Re-evaluating the model type is required if the evaluation findings show that it does not correspond with the currently available data [28].

2.3. Evaluation Stage

Other alternative evaluation steps can be taken by considering classification or regression approaches. The regression evaluation process depends on the discrepancy (error values) between the estimated and actual values. The road performance prediction model performs better with a lower error value, with an error value of 0 being the optimum result. This investigation used three metrics: R^2 [Equation 3], RMSE [Equation 2], and MAD [Equation 1]. Their low MAD and RMSE values may identify high-prediction models and the unit value that is closest to R^2 values. Since RMSE applies the squared value of the distinct between measurement results and prediction model outputs, it is more easily affected by extreme values than MAD. RMSE has a higher possibility of yielding a larger substantial error number in a model than MAD. Observing the variations and calculating the error value between both models will offer an alternative viewpoint on the suggested model, enabling it to be utilized for comparison [29].

The three measurement methods can be calculated in the following way: if yk is the actual measurement result value, y^k is the predicted value of the observation to kth, and N is the number of observations, then MAD, RMSE, and R^2 can be defined as follows:

$$MAD = \frac{\sum_{i=1}^{N}|yk-\widehat{yk}|}{N} \tag{1}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}|yk-\widehat{yk}|}{N}} \tag{2}$$

$$R^2 = \left(\frac{\sum_{i=1}^{N}(yk-\bar{y}).(\widehat{yk}-y)}{\sqrt{\sum_{i=1}^{N}(yk-\bar{y})^2.\sum_{i=1}^{N}(\widehat{yk}-y)^2}}\right) \tag{3}$$

Furthermore, plotting a regression error characteristic (REC) graph, which shows the error value tolerance on the x-axis compared to the estimated % tolerance on the y-axis, makes it simple to compare the regressions of various DM models. This study also takes advantage of the model's portrayal of the feasibility level. Every output is gathered for assessment. Compiling additional scripts can facilitate R application integration with other reporting applications [30].

2.4. Sensitivities Analysis

A database-based model must meet two fundamental requirements. On the one hand, excellent predictive ability is required, while on the other hand, especially in engineering, predictive models must be comprehensible and simple, especially in engineering [31]. However, this is one of the drawbacks of database-based black-box models, such as ANN and SVM [32].

Two essential visualization techniques can be calculated using sensitivity responses to overcome this. An input importance level graph is drawn to show the influence of each input (from 0% to 100%). The sensitivity analysis explains that the higher the resulting change, the more significant the value of the information. To measure the impact of this effect, the following research from Cortez (2013) adopted the following equation [17]:

$$g_a = \sum_{j=2}^{I}\left|\hat{y}_{aj} - \hat{y}_{aj-1}\right|/(I-1) \tag{4}$$

$$R_a = g_a/\sum_{i=1}^{I} g_i.\,100(\%) \tag{5}$$

To investigate the mean impact of the input x_a in the model, we can use the Variable Effect Characteristic (VEC) graph, which depicts the value of the level 1 attribute (x-axis) with a response sensitivity analysis (y-axis). Between successive x_a and j values, VEC will perform the linear interpolation. Plotting VEC curves on the same graph can enhance the visual analysis. In that case, the x-axis will be scaled (e.g., in [0,1]) for all values of x_a. Likewise, when a pair of input data (x_{a1}, x_{a2}) simultaneously varies (F> 2), the VEC display can be described so that it shows the average response to changes that occur in pairs [33].

3. Results and Discussion

Deterioration of road performance due to natural disasters is a general term to describe how road pavement layers change their condition or service function. Pavement performance evaluation covers the study of the functional behavior with the choice of pavement length. Historical information on pavement and traffic conditions during the selected period is necessary for functional or performance [34, 35]. This can be determined using periodic observations, pavement quality measurements, and historical traffic records. Figure 4 provides the user with information on expected road performance degradation as a definition of pavement performance.

The road performance forecast model aims to dynamically predict a declining road performance trend with different algorithm options. MR, ANN, and SVM algorithms are anticipated to provide a range of methods for forecasting road performance, considering time function and road load characteristics [36]. Throughout the investigation, the outcomes of creating a road performance prediction model will be assessed and modified until the model can accurately represent the dynamics of the available data. Models for predicting road performance must be dynamic and flexible enough to adapt to changing circumstances [32, 37].

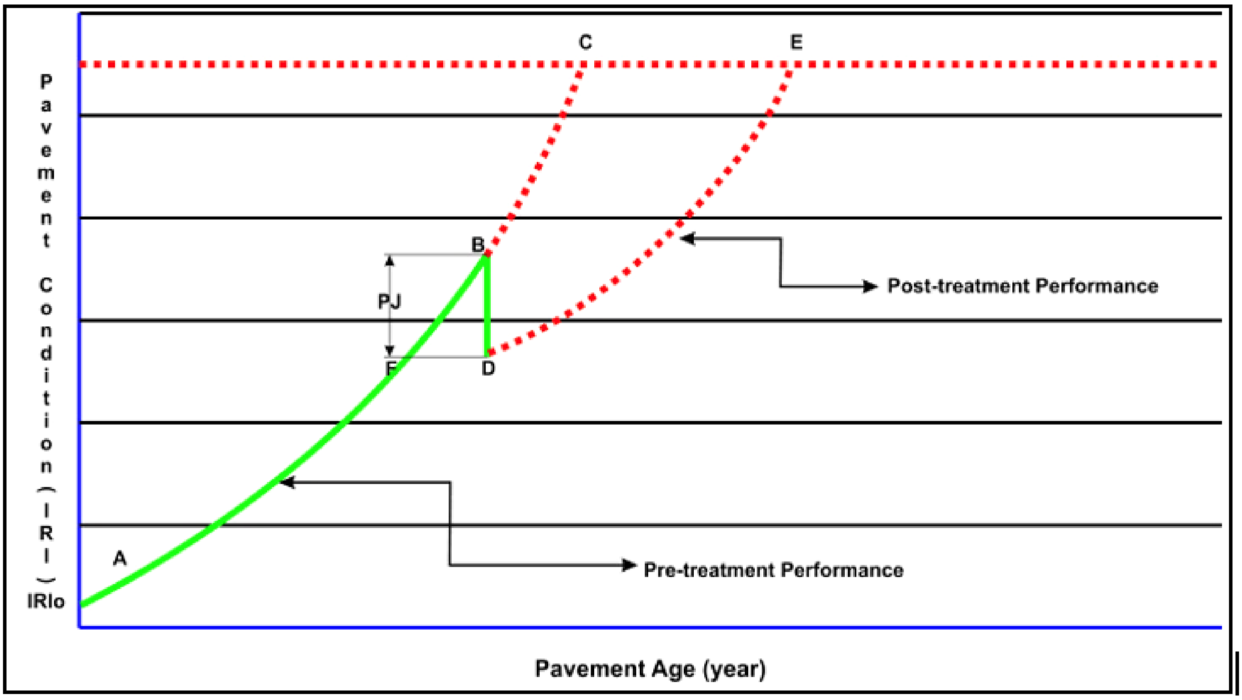

Figure 4. Decreased Road Performance (roughness)

3.1. Characteristic of Data

Data from the past that spans a respectable amount of time is utilized for validation, testing, learning, and calibration. The gathered information is tallied and organized according to the steps in the DM process [38]. Table 1 lists this investigation's sources, data collection, and descriptions. The first step in data analysis is to use statistical methods to describe the data with simple parameters. Initial data analysis used the R application with the MCDR-package to describe the main statistical parameters: the data's maximum, minimum, average, and standard deviation [39]. An overview of the data is necessary to understand its sources and distribution [40]. Various statistical approaches can be used to assess the data condition. Figure 5 shows the histogram of all input data before the disaster [41].

Therefore, matrix elements are filled with blue values instead of just numerical values. Figure 6 shows the correlation matrix for this study's 13 attributes and target variables. In this figure, the absolute value is considered fixed. The average is depicted in white $k_{ij} \approx 0$ (*uncorrelated)* and blue, showing $k_{ij} \approx 1$ (*strongly correlated*).

To measure the level of road performance, this study applies IRI to measure road performance. IRI will be transferred from historical data and combined with other assessment data [42]. To produce a road performance prediction model, this study uses historical data IRI_0, IRI, age (years), alligator cracking (m^2), longitudinal cracking (mm), block cracking (m^2), irregular cracking (m^2), pothole number (no/km), rutting (mm), depression (mm), structural patching (m^2), seal patching (m^2), and traffic data on the national road network in central Celebes.

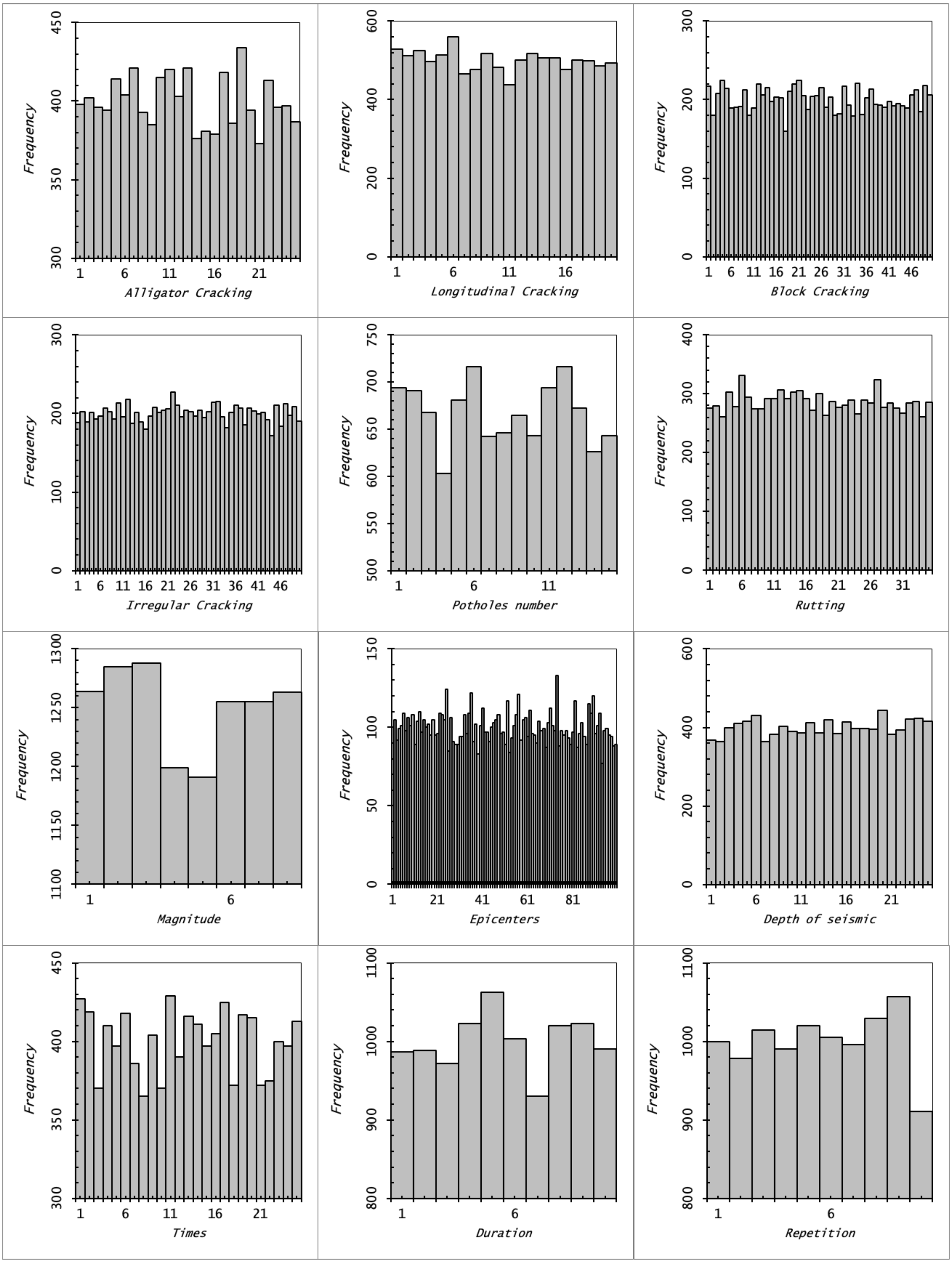

Figure 5. Histogram of all input data

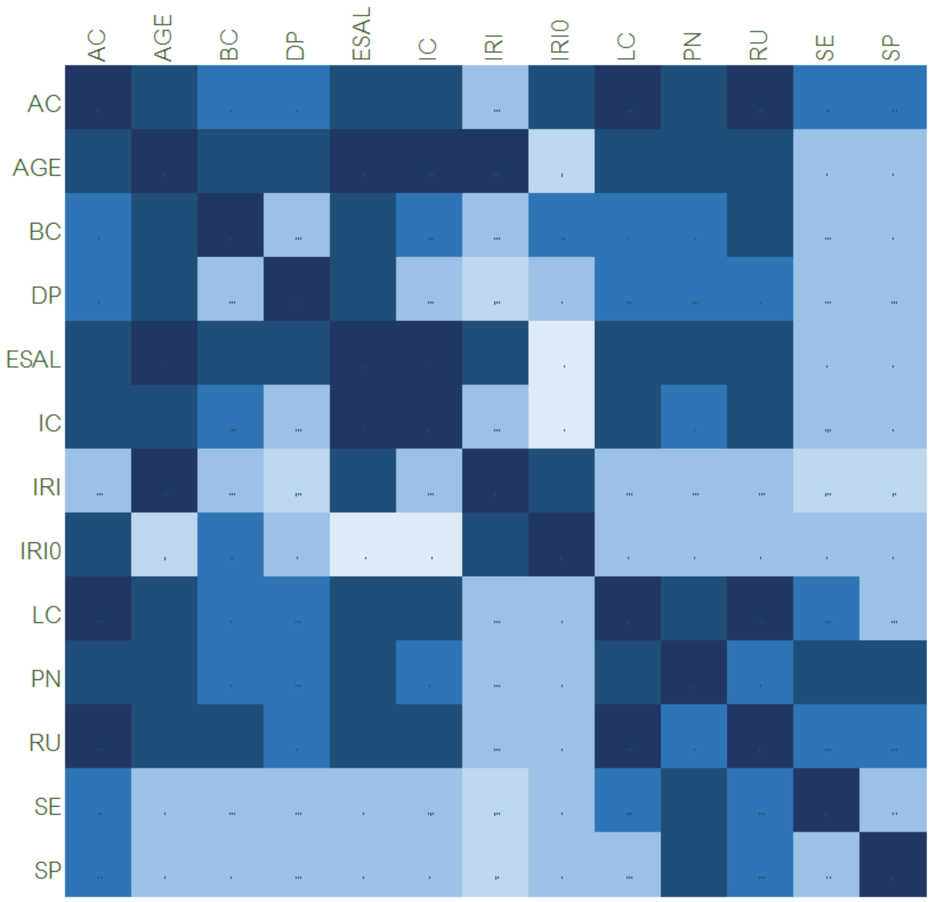

Figure 6. Correlation matrix

3.2. Model Architecture

DM modeling consists of two steps: the first is the learning stage, and the second is testing and validation with data not used in the learning stage. During the learning stage, the network uses inductive-learning principles to learn from a set of data called a training set. The input variables of the ANN model for the pavement condition prediction model are Magnitude (M), Epicentres (Ep), Depth of seismic (DS), Times (T), Duration (Dur), and Repetition (Rep). This study's pavement condition prediction modeling data set includes 11,370 samples from 68 earthquake events. Based on the earthquake *shake-map* results for the Palu area, it felt shocks of VII - IX MMI, and according to the BMKG report, 215 aftershocks measured 6.3 - 2.9 on the Richter scale. The damage included 68,451 housing units, 327 houses of worship, 265 school units, 78 office units, 362 shop units, 168 road cracks, seven bridges, etc.

The model is a de facto combination of sub-models, as shown in Figure 7. The ANN approach is usually used to analyze data by connecting the input set parameters with specific pattern characteristics that the network records during the learning phase. The IIRMS database contains profiling data for 15 years, covering 11,370 observations. The network can continue to train to predict IRI up to the model with the highest level of accuracy if the learning period continues [43]. The learning process can be interrupted if needed during the iteration stage. Interception records that are made will be recorded as an algorithm refinement process.

A perceptron can be used to classify input vectors that a linear boundary can separate. However, the perceptron cannot solve simple problems with linear constraints, so separating them in vector form is necessary. The limitations of using hyperplanes can be solved by using the SVM approach. Perceptron on ANN is included in one of the simplest network forms. A perceptron is usually used to classify a specific type of pattern, often known as linear separation. A perceptron in ANN with one layer has an adjustable weight and a threshold value. The algorithm used by this perceptron rule will set its independent parameters through the learning process. The activation function is made so that there is a restriction between the positive and negative regions.

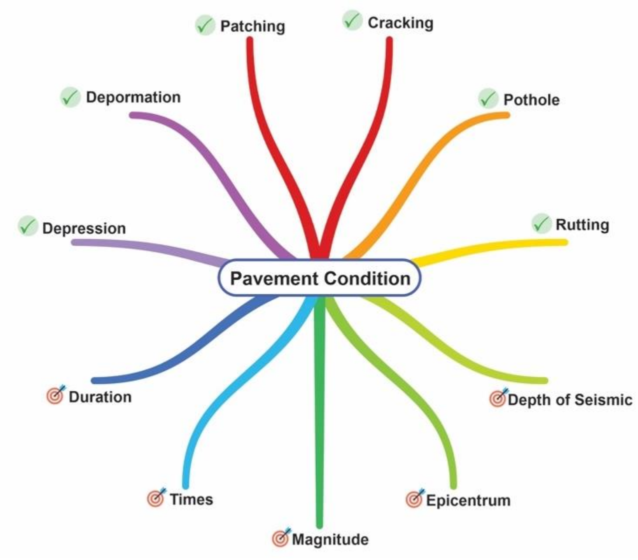

Figure 7. IRI Prediction Model Architecture

3.3. Model Interpretation

This model is built on the t-student distribution with a 95% confidence level. All DM models with ANN, SVM, and MR algorithms are trained on five input characteristics. Table 2 compares the ability of each training result model to predict IRI values using three different metrics: MAD, RMSE, and R2. This table demonstrates that the three DM models, particularly the ANN and SVM models, can accurately forecast IRI values.

Table 2. Error Metrics Model DM

Model	MAD	RMSE	R^2
MR	0.78 ± 0.00	1.12 ± 0.00	0.42 ± 0.00
ANN	0.71 ± 0.02	0.79 ± 0.03	0.81 ± 0.02
SVM	0.42 ± 0.01	0.70 ± 0.02	0.91 ± 0.01

Table 2 displays the R^2 and standard error of each constructed model. The SVM-based DM model has the most excellent R^2 value and the lowest MAD and RMSE values. With an R^2 of higher than 0.70, the prediction model, including the ANN and SVM algorithms, is deemed appropriate for estimating road performance forecasts. Using the SVM algorithm, the DM model is the road performance prediction model used in this study.

The REC curve illustrates five models: two existing and three proposed. The curve demonstrates that the SVM model has the highest accuracy, and lowest tolerance values and progresses continuously. The HDM-IV model had a high accuracy value but did not move consistently throughout each iterated tolerance setting. This REC curve, Figure 8, depicts the complete iteration process, including 20 runs on the SVM model with hyperparameters, as discussed in the preceding section. The REC curve's form can change when using different hyperparameters, and the number of iterations run varies.

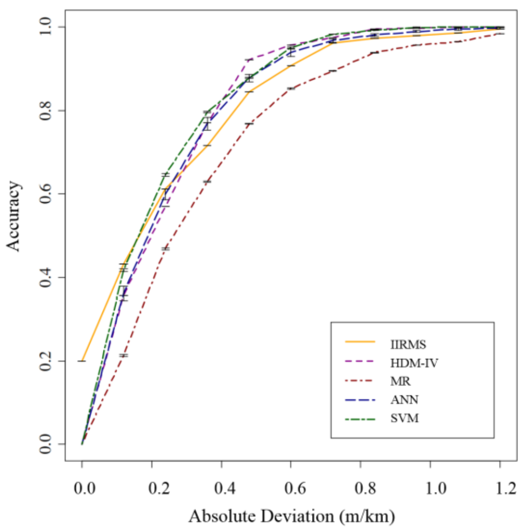

Figure 8. Iteration Process of REC Curve

The interpretation of the developed road performance prediction model yields the SVM model, which is the most accurate model. In the next part of this study, the DM model with the SVM algorithm is used as the primary model for predicting road performance, optimization, and DSS modules.

3.4. Variable Contribution

The developed DM model can evaluate the level of contribution and attributes of each variable, which serve as the input data for the model. This study's variables or characteristics included Rep, Dur, T, DS, Ep, and M. All features are grouped into three dimensions: initial performance, traffic, and distress. A parameter vector in the DM model is chosen to explain that it is a function of the variance, not the parameters, as in the parametric approach. The only requirement for a multiplier function is to generate a nonnegative definite variance matrix. Several methods can be used to estimate hyperparameter values. Mark θ can be predicted in this DM by using the cross-validation method. The hyperparameters used (H and γ) are H (2, 4, …, 10) and γ (2-15, 2-13, …, 23).

The highest accuracy is achieved in SVM, with 20 runs performed. The best hyperparameters to accomplish a fit SVM model are $\epsilon = 0.07 \pm 0.01$ and $\gamma = 0.05 \pm 0.00$. Meanwhile, the hyperparameters for ANN are $H = 3 \pm 1$. This value produces the most precise model with optimal run time. The approach can be used for further model development by trying other hyperparameter values. The contribution of each attribute and dimension is of relative importance in constructing the model. The script to get the value of this contribution is structured as follows:

mgraph (y=SVM, x = NULL, graph="IMP",

leg = c("initial","traffic","distress","patching"), xval = -1, PDF = "", PTS = -1,

size = c(5, 5), sort = TRUE, ranges = NULL, data = NULL,

digits = NULL, TC = -1, intbar = TRUE, lty = 1, col = "black",

main = "", metric = "MAE", baseline = FALSE, Grid = 0, axis = NULL)

The search results for the contribution value in the DM can be simplified and shown in Figure 9, which displays the relative importance on the x-axis for each attribute and dimensions on the y-axis, forming a road performance prediction model with the DM model approach using the SVM, ANN, and MR algorithms. The individual contribution of each attribute and dimension creating the road performance prediction model due to the earthquake is shown in this figure, relative to the importance attribute of the road condition model.

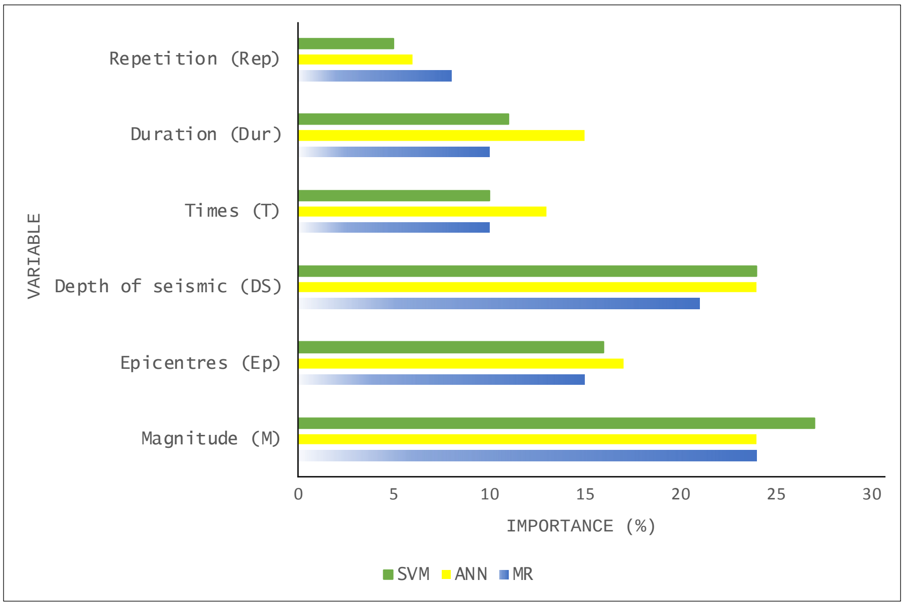

Figure 9. Relative Importance of Each Attribute

The subsequent model analysis is to develop an algorithm to select the main dimensions that influence the road performance prediction model and analyze the supporting variables that affect the models that are not accommodated in this model. The algorithm is compiled using a script that is entered into the package R-Miner command in the form of the VEC menu, which is as follows:

mgraph(y=SVM, x = NULL, graph="VEC", leg = NULL, xval = -1, PDF = "", PTS = -1,

size = c(5, 5), sort = TRUE, ranges = NULL, data = NULL,

digits = NULL, TC = -1, intbar = TRUE, lty = 1,

col = "black", "red", "green", "blue"

main = "", metric = "MAE", baseline = FALSE, Grid = 0,

axis = NULL)

The VEC analysis's results illustrate the influence of the main attributes that move dynamically in the earthquake-induced road performance prediction model with this SVM model in the form of magnitude. The decrease in IRI follows the increase in magnitude and depth of seismic activity; conversely, IRI does not have much effect due to epicenter and duration. Figure 10 provides a complete description of the changes in IRI values in the prediction model for road performance due to earthquakes.

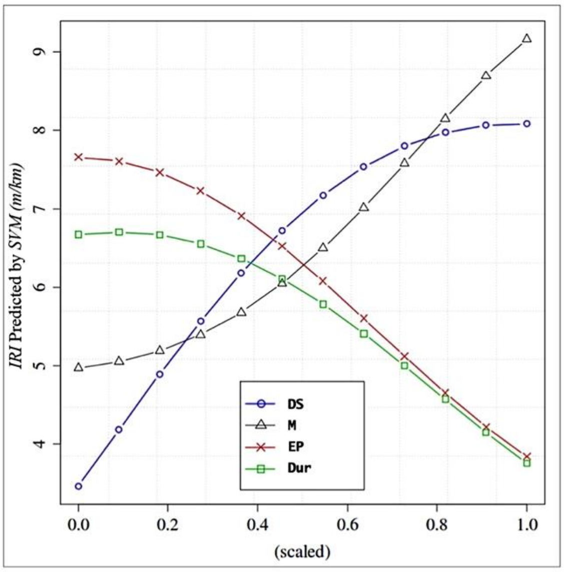

Figure 10. VEC Curve for Prediction Changes IRI Values

Through the resulting VEC curve, the magnitude, the most influential dimension group, has a naturally readable dynamic movement when IRI reaches a scale of 5.0 m/km, which then moves quickly even beyond the depth of the seismic attribute. In comparison, the depth of the seismic attribute begins to affect the decrease in IRI from the start (initial performance) and then decreases when the attribute scale reaches 0.6. The DM model with the SVM algorithm used in preparing the VEC curve shows that almost all attributes have a non-linear relationship with IRI. As empirical data in this model shows, all variables impact the level of road performance.

3.5. Sensitivity Analysis

This sensitivity analysis finds the effect of changes in input attributes on changes in IRI values. The possible consequences of these changes can be identified and anticipated by conducting a sensitivity analysis. The following is the script used to develop the sensitive analysis.

```
#---- apply a sensitivity analysis ----
Mdls[[1]]$IMP <- Importance(Mdls[[1]]$FM,
    data = Mdls[[1]]$bd, # your data just with the input variables
    RealL = 12,
    method = "sensg",
    measure = "gradient",
    sampling = "regular",
    baseline = "mean",
    responses = TRUE,
    outindex = NULL,
    task = "default",
    PRED = NULL,
```

Figure 10 shows the magnitude, time (t), and IRI relationship. The high magnitude causes a decrease in IRI in only a short time. This formula provides various conclusions. One of them involves a large-scale road that necessitates special handling beyond the scope of a standard approach. The package used is Global Sensitivity Analysis (GSA).

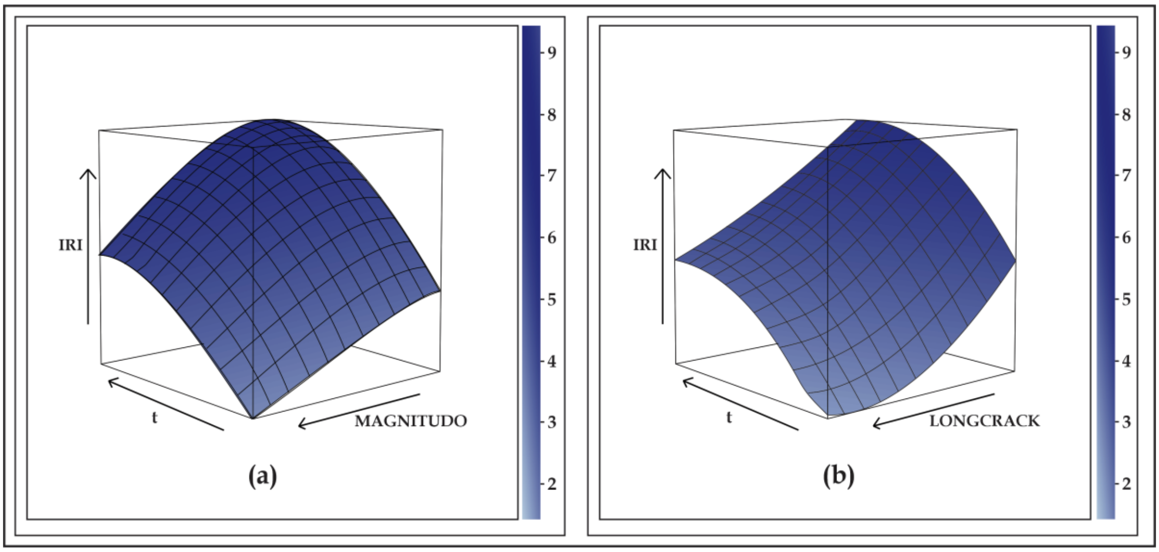

Figure 11. VEC Interaction (a) Magnitude with t (b) Longitudinal Crack with t

While the behavior of the Depth of Seismic conditions that interact with t has its characteristics, it can also be handled using the damage growth character, which tends to be close to linear. Through the sensitivity analysis approach using the SA and GSA algorithms, changes in the effect of each attribute are obtained in each iteration. SA and GSA can be applied to determine the nature of the modeling in the form of a behavior model for the level of road performance in actual conditions.

4. Conclusion

The model development with the DM approach begins by using the initial data in the learning stage, then the test stage, and finally, the validation stage. DM techniques, especially the SVM and ANN algorithms, are powerful tools for predicting and interpreting post-disaster IRI values. This approach can conduct the learning stages accurately and describe the complex relationship among changes in IRI values, distress, and traffic conditions as contributing factors. At this stage, SVM achieves high performance, with R^2 reaching 0.91 as a performance indicator and error values obtained of MAD 0.42 and RMSE 0.70. The SVM-based DM model has the most excellent R^2 value and the lowest MAD and RMSE values. With an R2 of higher than 0.70, the prediction model, including the ANN and SVM algorithms, is deemed appropriate for estimating road performance forecasts.

Based on the GSA algorithm, it is shown that the most significant contribution is a disaster. This proves that changes in IRI values are strongly influenced by a disaster, contributing to 27.00% [magnitude]. The IRI impairment process is consistent and closely related to the initial IRI. All distress is associated with each other, especially the crack, pothole, and patching methods. Based on VEC, the DM approach is in the form of MR, ANN, and SVM algorithms, each of which has a different level of accuracy. The highest accuracy is achieved in SVM, with 20 runs performed. The best hyperparameters to accomplish a fit SVM model are $\epsilon = 0.07 \pm 0.01$ and $\gamma = 0.05 \pm 0.00$. Meanwhile, the hyperparameters for ANN are $H = 3 \pm 1$.

A critical methodology that complements DM is the application of GSA in the learning stage. This methodology can make a valuable contribution to model interpretation and provide improved comprehension of the results of data analysis. DM has high learning ability and flexibility, even when solving high-dimensional problems. Structured database support with essential data can strengthen DM performance. However, on the other hand, the main weakness associated with using DM techniques in solving complex problems is the interpretation model in the form of high-level mathematical language according to the implemented algorithm. This study applied the GSA approach during the learning, testing, and validation processes to overcome these weaknesses.

This research is limited to flexible pavement structures. The same approach can be taken for rigid pavements with different characteristics and behaviors by changing the approach pattern to variables and algorithms. The DM capabilities of the R-Miner package are high and wide. Various prediction models and interpretation needs can be developed with the DM approach. A deep understanding of computer language and mathematical algorithms is needed to get more profound research results. As mentioned throughout the system development description, fully integrating all technologies and tools will enhance capabilities and user-friendliness. In this study, integrating new geographic information system technology at the concept and initial level simulation stages highlights the need for the following research version in the form of migration capabilities and system expansion to various models.

5. Declarations

5.1. Author Contributions

Conceptualization, A.I.R., M.I., and J.P.; methodology, A.I.R., M.I., and J.P.; formal analysis, A.I.R., Y.A.S., M.I., and J.P.; investigation, A.I.R., M.F.Z., and M.I.; data collection, A.I.R. and M.I.; writing—original draft preparation, Y.A.S., A.I.R., M.F.Z., and M.I.; writing—review and editing, J.P., Y.A.S., M.I., and A.I.R. All authors have read and agreed to the published version of the manuscript.

5.2. Data Availability Statement

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

5.3. Funding

The authors would like to thank the Research and Community Service - Universitas Internasional Batam, for funding this research.

5.4. Acknowledgments

This research is part of an international collaboration program supported by Universitas International Batam, Universitas Mercu Buana, Universiti Tun Hussein Onn Malaysia, and Universiti Teknologi Malaysia. We are also grateful to our research colleagues and all the reviewers who have contributed insights and expertise that have immensely helped this research.

5.5. Conflicts of Interest

The authors declare no conflict of interest.

6. References

[1] Liu, H., Tatano, H., Kajitani, Y., & Yang, Y. (2022). Analysis of the influencing factors on industrial resilience to flood disasters using a semi-markov recovery model: A case study of the Heavy Rain Event of July 2018 in Japan. International Journal of Disaster Risk Reduction, 82, 103384. doi:10.1016/j.ijdrr.2022.103384.

[2] Isradi, M., Dwiatmoko, H., Prasetijo, J., Rifai, A. I., Zainal, Z. F., Zhang, G., & Firdaus, H. Y. (2024). Identification of hazardous road sites: a comparison of blackspot methodology of Narogong Road Bekasi and Johor Federal Roads. Sinergi (Indonesia), 28(2), 347–354. doi:10.22441/sinergi.2024.2.014.

[3] Souza Almeida, L., Goerlandt, F., & Pelot, R. (2022). Trends and gaps in the literature of road network repair and restoration in the context of disaster response operations. Socio-Economic Planning Sciences, 84, 101398. doi:10.1016/j.seps.2022.101398.

[4] Zamanifar, M., & Hartmann, T. (2021). Decision attributes for disaster recovery planning of transportation networks; A case study. Transportation Research Part D: Transport and Environment, 93, 102771. doi:10.1016/j.trd.2021.102771.

[5] Statistik. (2024).Data Statistik Bencana SITABA, Ministry of Public Works and Housing, Jakarta, Indonesia.

[6] Iskaputri, A., Razak, A., & Arifin, M. A. (2020). Logistics Management of the Regional Disaster Management Agency of South Sulawesi Province. Hasanuddin Journal of Public Health, 1(1), 41–50. doi:10.30597/hjph.v1i1.9511.

[7] Hedriyanti, G., & Syamsuddin, A. B. (2021). The role of women against disaster management in the social service of southern Sulawesi province. Journal of Social Welfaism, 44(2), 21-37. (In Indonesian).

[8] García-Alviz, J., Galindo, G., Arellana, J., & Yie-Pinedo, R. (2021). Planning road network restoration and relief distribution under heterogeneous road disruptions. OR Spectrum, 43(4), 941–981. doi:10.1007/s00291-021-00644-x.

[9] Mao, X., Zhou, J., Yuan, C., & Liu, D. (2021). Resilience-based optimization of postdisaster restoration strategy for road networks. Journal of Advanced Transportation, 2021(1), 8871876. doi:10.1155/2021/8871876.

[10] Liu, H., Tatano, H., Kajitani, Y., & Yang, Y. (2022). Modeling post-disaster business recovery under partially observed states: A case study of the 2011 great East Japan earthquake. Journal of Cleaner Production, 374, 133870. doi:10.1016/j.jclepro.2022.133870.

[11] Zamanifar, M., & Hartmann, T. (2020). Optimization-based decision-making models for disaster recovery and reconstruction planning of transportation networks. Natural Hazards, 104(1), 1–25. doi:10.1007/s11069-020-04192-5.

[12] Özerol Özman, G., Arslan Selçuk, S., & Arslan, A. (2024). Image classification on Post-Earthquake damage assessment: A case of the 2023 Kahramanmaraş earthquake. Engineering Science and Technology, an International Journal, 56, 101780. doi:10.1016/j.jestch.2024.101780.

[13] Liu, H., Tatano, H., & Samaddar, S. (2023). Analysis of post-disaster business recovery: Differences in industrial sectors and impacts of production inputs. International Journal of Disaster Risk Reduction, 87, 103577. doi:10.1016/j.ijdrr.2023.103577.

[14] Reis, H. C., Turk, V., Karacur, S., & Kurt, A. M. (2024). Integration of a CNN-based model and ensemble learning for detecting post-earthquake road cracks with deep features. Structures, 62, 106179. doi:10.1016/j.istruc.2024.106179.

[15] Karimi, S., & Mirza, O. (2023). Damage identification in bridge structures: review of available methods and case studies. Australian Journal of Structural Engineering, 24(2), 89–119. doi:10.1080/13287982.2022.2120239.

[16] Hoang, N. D., & Bui, D. T. (2018). Predicting earthquake-induced soil liquefaction based on a hybridization of kernel Fisher discriminant analysis and a least squares support vector machine: a multi-dataset study. Bulletin of Engineering Geology and the Environment, 77(1), 191–204. doi:10.1007/s10064-016-0924-0.

[17] Cortez, P. (2010). Data Mining with Neural Networks and Support Vector Machines Using the R/rminer Tool. Advances in Data Mining. Applications and Theoretical Aspects. ICDM 2010. Lecture Notes in Computer Science, 6171. Springer, Berlin, Germany. doi:10.1007/978-3-642-14400-4_44.

[18] Irfan, A. I., Pereira, P., Hadiwardoyo, S. P., Correia, A. G., & Cortez, P. (2015). Implementation of Data Mining to Support Road Pavement Management System in Indonesia. Jurnal HPJI, 1(2), 93-104.

[19] Sui, Q. R., Chen, Q. H., Wang, D. D., & Tao, Z. G. (2023). Application of machine learning to the Vs-based soil liquefaction potential assessment. Journal of Mountain Science, 20(8), 2197–2213. doi:10.1007/s11629-022-7809-4.

[20] Kuwabara, K., Xiu, H., Shinohara, T., & Matsuoka, M. (2020). Evaluation of Liquefaction Occurrence for Large Area: A Machine Learning Approach. Proceedings of 17^{th} World Conference on Earthquake Engineering, 1-12.

[21] Isradi, M., Prasetijo, J., Rifai, A. I., Andraiko, H., & Zhang, G. (2024). The Prediction of Road Condition Value during Maintenance Based on Markov Process. International Journal on Advanced Science, Engineering and Information Technology, 14(3), 1083–1090. doi:10.18517/ijaseit.14.3.19475.

[22] DGH. (2020). Road Condition Book 2020. Director of PUPR. Available online: https://www.azores.gov.pt/NR/rdonlyres/D21CF49B-EF59-4E76-88BD-5D0EEC3A2D4F/1098978/PlanoARPLAlcool.pdf (accessed on November 2024).

[23] Isradi, M., Prasetijo, J., Prasetyo, Y. D., Hartatik, N., & Rifai, A. I. (2023). Prediction of Service Life Base on Relationship between PSI and IRI for Flexible Pavement. Proceedings on Engineering Sciences, 5(2), 267–274. doi:10.24874/PES05.02.009.

[24] Xu, S., Liu, Q., Bo, Y., Chen, Z., & Wang, C. (2024). Estimating the International Roughness Index of asphalt concrete pavement by response-based testing equipment and intelligent algorithms. Construction and Building Materials, 433, 136659. doi:10.1016/j.conbuildmat.2024.136659.

[25] Kaloop, M. R., El-Badawy, S. M., Hu, J. W., & Abd El-Hakim, R. T. (2023). International Roughness Index prediction for flexible pavements using novel machine learning techniques. Engineering Applications of Artificial Intelligence, 122, 106007. doi:10.1016/j.engappai.2023.106007.

[26] Basnet, K. S., Shrestha, J. K., & Shrestha, R. N. (2023). Pavement performance model for road maintenance and repair planning: a review of predictive techniques. Digital Transportation and Safety, 2(4), 253-267. doi:10.48130/DTS-2023-0021.

[27] Abedi, M., Shayanfar, J., & Al-Jabri, K. (2023). Infrastructure damage assessment via machine learning approaches: a systematic review. Asian Journal of Civil Engineering, 24(8), 3823–3852. doi:10.1007/s42107-023-00748-5.

[28] Yilmaz, M., Yalcin, E., Kifah, S., Demir, F., Sengur, A., Demir, R., & Mehmood, R. M. (2024). Improving the Classification Performance of Asphalt Cracks after Earthquake with a New Feature Selection Algorithm. IEEE Access, 12, 6604–6614. doi:10.1109/ACCESS.2023.3343619.

[29] Trucchia, A., Izadgoshasb, H., Isnardi, S., Fiorucci, P., & Tonini, M. (2022). Machine-Learning Applications in Geosciences: Comparison of Different Algorithms and Vegetation Classes' Importance Ranking in Wildfire Susceptibility. Geosciences, 12(11), 424. doi:10.3390/geosciences12110424.

[30] Elhadidy, A. A., Elbeltagi, E. E., & Ammar, M. A. (2015). Optimum analysis of pavement maintenance using multi-objective genetic algorithms. HBRC Journal, 11(1), 107–113. doi:10.1016/j.hbrcj.2014.02.008.

[31] Elhadidy, A. A., Elbeltagi, E. E., & El-Badawy, S. M. (2020). Network-Based Optimization System for Pavement Maintenance Using a Probabilistic Simulation-Based Genetic Algorithm Approach. Journal of Transportation Engineering, Part B: Pavements, 146(4), 4020069. doi:10.1061/jpeodx.0000237.

[32] Rifai, A. I., Hendra, & Prasetyo, E. (2020). Data mining applied for liquefaction mapping and prediction learn from Palu earthquakes. Civil Engineering and Architecture, 8(4), 507–514. doi:10.13189/cea.2020.080414.

[33] Kuhn, K. D. (2012). Pavement Network Maintenance Optimization Considering Multidimensional Condition Data. Journal of Infrastructure Systems, 18(4), 270–277. doi:10.1061/(asce)is.1943-555x.0000077.

[34] Abaza, K. A. (2023). Stochastic-based pavement rehabilitation model at the network level with prediction uncertainty considerations. Road Materials and Pavement Design, 24(11), 2680–2698. doi:10.1080/14680629.2022.2164330.

[35] Abaza, K. A. (2006). Iterative linear approach for nonlinear nonhomogenous stochastic pavement management models. Journal of Transportation Engineering, 132(3), 244–256. doi:10.1061/(ASCE)0733-947X(2006)132:3(244).

[36] Preethaa, S., Natarajan, Y., Rathinakumar, A. P., Lee, D. E., Choi, Y., Park, Y. J., & Yi, C. Y. (2022). A Stacked Generalization Model to Enhance Prediction of Earthquake-Induced Soil Liquefaction. Sensors, 22(19), 7292. doi:10.3390/s22197292.

[37] Rifai, A. I., Hadiwardoyo, S. P., Correia, A. G., Pereira, P., & Cortez, P. (2015). The data mining applied for the prediction of highway roughness due to overloaded trucks. International Journal of Technology, 6(5), 751–761. doi:10.14716/ijtech.v6i5.1186.

[38] Guo, H., Zhuang, X., Chen, J., & Zhu, H. (2022). Predicting Earthquake-Induced Soil Liquefaction Based on Machine Learning Classifiers: A Comparative Multi-Dataset Study. International Journal of Computational Methods, 19(8), 2142004. doi:10.1142/S0219876221420044.

[39] Ji, A., Xue, X., Wang, Y., Luo, X., & Zhang, M. (2020). An integrated multi-objectives optimization approach on modelling pavement maintenance strategies for pavement sustainability. Journal of Civil Engineering and Management, 26(8), 717–732. doi:10.3846/jcem.2020.13751.

[40] Sindi, W., & Agbelie, B. (2020). Assignments of Pavement Treatment Options: Genetic Algorithms versus Mixed-Integer Programming. Journal of Transportation Engineering, Part B: Pavements, 146(2), 04020008. doi:10.1061/jpeodx.0000163.

[41] Karimai, K., Liu, W., & Maruyama, Y. (2024). Prediction and Factor Analysis of Liquefaction Ground Subsidence Based on Machine-Learning Techniques. Applied Sciences (Switzerland), 14(7), 2713. doi:10.3390/app14072713.

[42] Ramachandran, S., Rajendran, C., Veeraragavan, A., & Ramya, R. (2017). A Framework for Maintenance Management of Pavement Networks under Performance-Based Multi-Objective Optimization. Airfield and Highway Pavements 2017, 209–221. doi:10.1061/9780784480922.019.

[43] Yu, B., Gu, X., Ni, F., & Guo, R. (2015). Multi-objective optimization for asphalt pavement maintenance plans at project level: Integrating performance, cost and environment. Transportation Research Part D: Transport and Environment, 41, 64–74. doi:10.1016/j.trd.2015.09.016.