Detection of Asphalt Pavement Segregation Using Machine Learning Linear and Quadratic Discriminant Analyses

Arunika, Austin
Department of Metallurgical and Materials Engineering, Faculty of Engineering, Universitas Indonesia

Jaka Fajar Fatriansyah
Department of Metallurgical and Materials Engineering, Faculty of Engineering, Universitas Indonesia

Venia Andira Ramadheena
Department of Metallurgical and Materials Engineering, Faculty of Engineering, Universitas Indonesia

https://doi.org/10.5109/4774236

出版情報：Evergreen. 9 (1), pp.213-218, 2022-03. Transdisciplinary Research and Education Center for Green Technologies, Kyushu University
バージョン：

Detection of Asphalt Pavement Segregation Using Machine Learning Linear and Quadratic Discriminant Analyses

Austin Arunika[1], Jaka Fajar Fatriansyah[1,*], Venia Andira Ramadheena[1]
[1]Department of Metallurgical and Materials Engineering, Faculty of Engineering, Universitas Indonesia, Kampus UI Depok, Depok, 16424, Indonesia

*Author to whom correspondence should be addressed:
E-mail: jakafajar@ui.ac.id

(Received February 11, 2022; Revised March 22, 2022; accepted March 22, 2022).

Abstract: Segregation (hot-mix asphalt segregation) is one of the main problems affecting asphalt pavement performance. The early detection is important, but the tests are quite expensive and time-consuming. The visual examination is the cheapest method but too varied in judgement and can rise further problems. In this experiment, we developed machine learning linear and quadratic discriminant analyses to detect/classify segregated and non-segregated pavement asphalt. Six variables were employed: SD only, IR only, MAD only, IR-mean, MAD-mean, IR-mean, MAD-SD-mean and IR-SD-mean. The results showed that the complexities of information affect machine learning performance. IR-SD-mean and MAD-SD-mean parameters gave best accuracy performance for training data at 99.2% (LDA)/98.5% (QDA) and testing data at 98.33% (LDA)/95% (QDA) respectively. In general, QDA gave more accuracy performance in comparison to LDA although our data dimension is small.

Keywords: Asphalt pavement segregation; Machine Learning; segregation detection; Linear discriminant analysis; quadratic discriminant analyses

1. Introduction

Segregation (hot-mix asphalt segregation) is one of the main problems affecting asphalt pavement performance[1,2]. Segregation is defined as the separation of aggregate gradations, which are fine aggregates and coarse aggregates in the asphalt mixture. Areas with a predominance of coarse aggregate have high air voids and low asphalt content so that it can cause pavement due to exposure to moisture and low asphalt resistance because it can cause cracks and holes in the asphalt pavement. Meanwhile, in areas where the dominance of fine aggregate has low air voids and high asphalt content can trigger permanent deformation, such as bumpy roads[2]. There are a lot of factors that contribute to asphalt segregation, such as: material properties, devices/instruments handling, operational parameters, and environment conditions (humidity, temperature, pH)[3,4]. There is almost no theory on how to reduce segregation in hot mix asphalt, instead, practical approach is largely used[5,6,7,8].

Detection of segregation can be done by non-destructive testing, including visual examination, sand patch testing, nuclear density, and even infra-red tomography-based measurements[9]. The sand patch testing method, which is very common, is used to measure observations on differences in surface macrotextures. However, this method, and the other physical methods, takes a long time and costs a lot of money. Meanwhile, visual examination is often the easier method, however, visual examination is very subjective and may rise additional problems which may cause future disputes between agents and contractors.

Machine learning is a smart algorithm which can offer consistent and efficient classification of a sensed or processed imagery. Various engineering problems were solved through this Artificial intelligence subset [10, 11, 12]. Processed imagery can be obtained using image processing methods which is a low cost in comparison to other asphalt segregation detection methods. Image processing is a method that manipulates the image of a photo by enhancing or extracting the image so that it will display the characteristics of the image[13]. Lin Cong et al. have used the image of the paved mixture (IPM) during construction to detect the segregation formed by detecting variations in the spatial position between the aggregates using machine learning classifier. This research classified segregation into three groups, i.e., fine aggregate segregation (FAS), non-segregation (NS) and rough aggregate segregation (CAS). The experiment conducted by Lin Cong et al. resulted in an accuracy of up to 87.5%[14]. Furthermore, Baqersad et al. have used machine learning image processing to detect asphalt segregation, which utilizes the frequency histogram data standard deviation

(SD). The result of this research is accuracy to determine the asphalt segregation is beyond 80%[15]. However, until now there is no study to detect asphalt segregation in Indonesia, which may be interesting due to varied standards in comparison to other countries. Furthermore, other deviation parameters such as Median absolute deviation (MAD) and interquartile range (IR) offer more robust and can detect outlier data in comparison to SD[16, 17, 18, 19, 20].

In this study, we use machine learning classification methods of linear discriminant analysis (LDA) and quadratic discriminant analyses (QDA) using parameters of SD, MAD as well as IR to detect asphalt segregation in asphalt pavement. We expect that the use of MAD and IR, or both yields more accuracies in comparison SD use in LDA and QDA.

2. Data and Methods

In this research, we chose pavement at the University of Indonesia, Depok as our observation place. The locations are shown in Fig 1. All the images taken contain segregated and non-segregated asphalt pavement. The images were taken using a Canon 60D digital camera with a lens specification of 18-135 mm. Ahead of taking images, the pavement surface must be cleaned and dried. The pictures were taken at almost the same time, so that all the images are shadows free and has similar illumination. The totals images are 200, 100 images belong to segregated class and non-segregated each. ImageJ software was used to extract color histogram of images after grey scaled, from RGB-color, into 256 colors mode. The samples are shown in Fig 2. As shown in Figure 2 (a) and (b), the difference between segregated and non-segregated areas can be distinguished, where the segregated asphalt pavement images tend to have more coarse aggregates and may look separate from fine aggregates in comparison to non-segregated asphalt pavement images which results in higher SD values of segregated pavement images to non-segregated one. Figure 2. (c) shown the difference between color asphalt image and (d) shown greyscaled asphalt image. Figure 3. Demonstrates the difference of histogram between segregated and non-segregated asphalt pavement images.

The color histogram data was then used to determine mean, SD, MAD, and IR. The training and testing were conducted using Machine Learning classification techniques of LDA and QDA. Machine Learning used to classify between segregation and non-segregation by the calculated MAD and IR. There are 140 and 60 of training data and 60 testing data respectively. Python, scikit-learn platform, was used in machine learning classification.

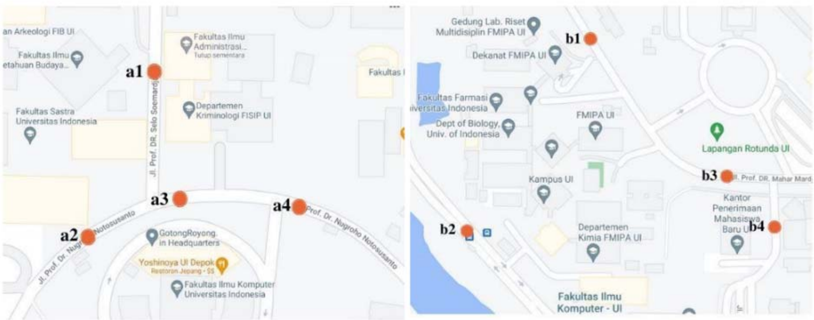


Fig. 1: The locations of observation data images. All data are located at the University of Indonesia, Depok. **a1)** -6.362974"S, 106.828526"E, **a2)** -6.364120"S, 106.828134"E, **a3)** -6.363875"S, 106.828703"E, **a4)** -6.363967"S, 106.829495"E, **b1)** -6.367520"S, 106.826864"E, **b2)** -6.369670"S, 106.825724"E, **b3)** -6.369130"S, 106.828373"E, **b4)** -6.369583"S, 106.828881"E.

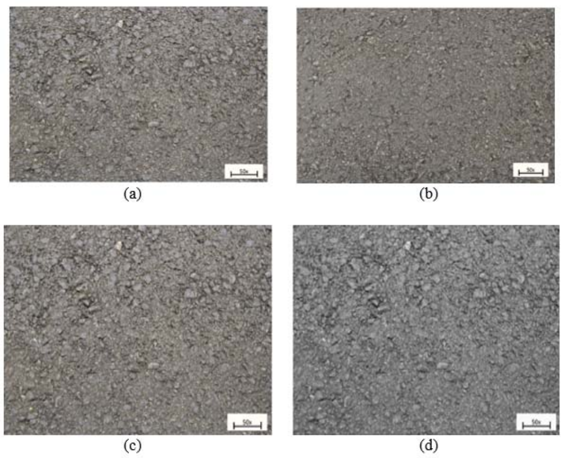
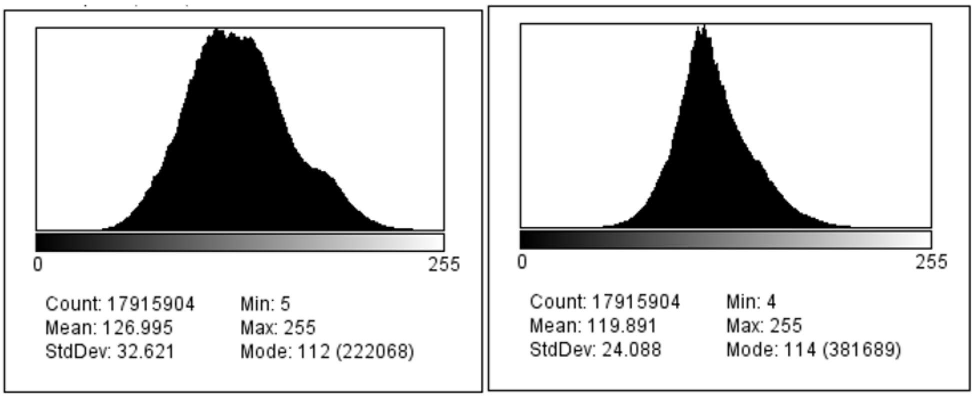


Fig. 2: (a) Segregated asphalt image, (b) Non-segregated asphalt image, (c) Colored asphalt image and (d) Greyscaled asphalt pavement image.



(a) (b)

Fig. 2: (a) The grayscale histogram of segregated (b) and non-segregated.

3. Results and Discussion

The performance of machine learning classification methods, LDA and QDA, is determined by calculating the ratio of correct answer to incorrect answer for training and testing data. 100 % value reflects perfect accuracy. Here as variables, we used SD only, IR only, MAD only, IR-mean, MAD-mean, IR- mean, MAD-SD-mean and IR-SD-mean. The results are given in Fig 4. for training and testing data accuracy.

Generally, in training data accuracy, IR and MAD parameters give more accuracy values than SD for both LDA and QDA techniques. Our SD data is quite indistinguishable between segregated and non-segregated asphalt pavement data. In addition, there is a lot of outlier's data scattered in both segregated and non-segregated asphalt pavement data samples. We did not clean data before apply machine learning method on purpose to observe that we cannot treat outlier's data as an exception. This result is supported by Leys et al., an experiment which showed that MAD gave more robust measure of dispersion and can handle outlier's data more in comparison to SD and mean[21)]. Park and Cho also demonstrated that IR can be used to handle contaminated and non-normal data in comparison to other dispersion measures[22)].

In our experiment, we concerned about different times of image acquisition which affect color and lightning appearance. Although when we did image acquisition that the angle of camera was kept steady and the time acquisition was kept in narrow time window, the color and lightning appearance of images still quite varied which may affect machine learning capability to distinguish between segregated and non-segregated asphalt pavement data. Trying to overcome these effects, we add mean as a machine learning parameter, since mean can reflect color and lightning appearance conditions better than SD. The result for training data accuracy, IR-mean gave better performance than IR only, while MAD-mean gave similar performance to MAD only.

To strengthen accuracy performance, we add more parameters for machine learning: IR-SD-mean and MAD-SD-mean. The training data accuracy of IR-SD-mean gave highest accuracy while MAD-SD-mean did not give significant performance. The complexity, or the number of parameters in machine learning affect the accuracy or performance. The plot (Figure 5) given by Müller and Guido gave explanation that with more complex information given to the system, more accuracy was obtained for training data, but not necessary for testing data[23)]. Unfortunately, our testing data accuracy gave poor performance. Only MAD-SD-mean gave acceptable performance. Further improvement should be done.

In almost all training and testing data accuracy variables, QDA gave more accuracy than LDA. However, this is quite peculiar since LDA should give more accuracy in the case the dimension of data is small[24)] such in our case.

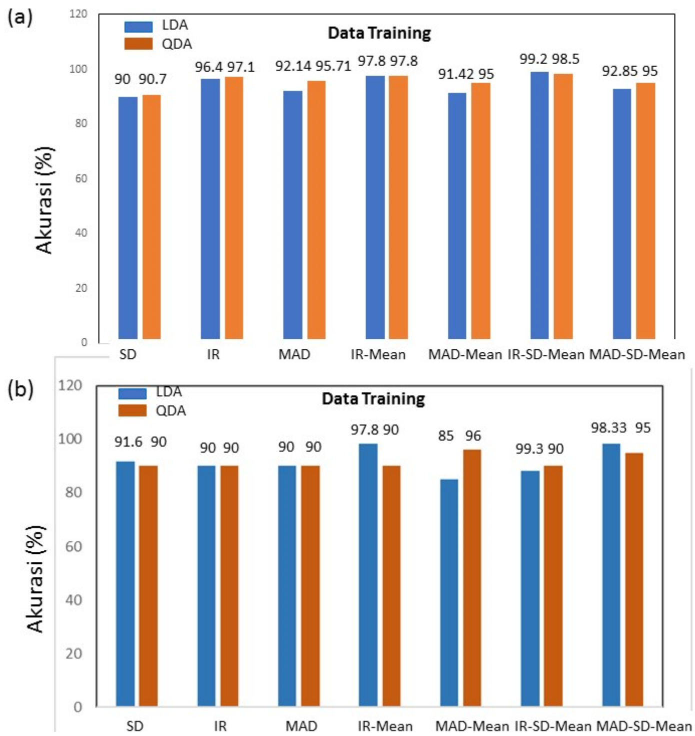


Fig. 4: Data results for (a) testing and (b) training set.

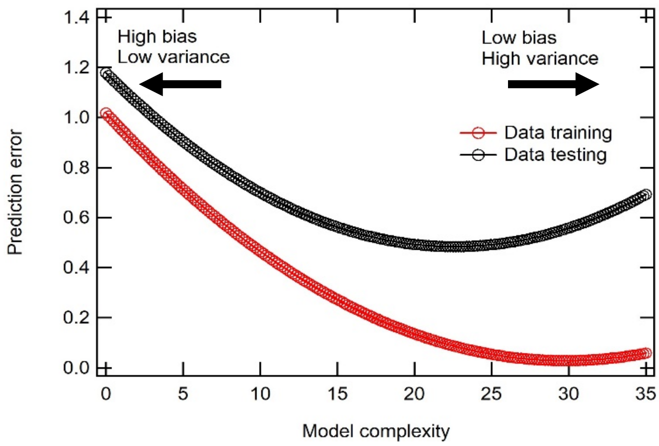


Fig. 5: The graph of the relationship between model complexity and prediction error[17].

4. Conclusion

The QDA and LDA techniques of machine learning to detect segregated and non-segregated pavement asphalt have been experimented. The results showed that the complexities of information affect machine learning performance. IR-SD-mean and MAD-SD-mean parameters gave best accuracy performance for training data at 99.2% (LDA)/98.5% (QDA) and testing data at 98.33% (LDA)/95% (QDA) respectively. In general, QDA gave more accuracy performance in comparison to LDA although our data dimension is small. Unfortunately, our testing data accuracy gave poor performance and further improvement should be done.

Acknowledgements

We are grateful that our work is supported by University of Indonesia through Direktorat Riset and Kementrian Pendidikan, Kebudayaaan, Riset dan Teknologi number of grant PUTI Q3 NKB-2029/UN2.RST/HKP.05.00/2020.

References

1) M. Stroup-Gardiner, and E.R. Brown, “Segregation in hot-mix asphalt pavements,” Transportation Research Board, 2000. https://trid.trb.org/view/658476
2) S.A. Cross, M.R. Hainin, and A. Ado-Osei, “Effects

of segregation on mix properties of hot mix asphalt (No. K-TRAN: KU-96-6)," University of Kansas Center for Research, Inc, 1998. https://books.google.co.id/books/about/Effect_of_Segregation_on_Hot_Mix_Asphalt.html?id=B5rwI9kVh5wC&redir_esc=y

3) P. Tang, and V.M. Puri, "Methods for minimizing segregation: a review," *Particulate Science and Technology*., **22** (4) 321-337 (2004). doi.org/10.1080/02726350490501420.
4) F. Chen, D. Jelagin, and M.N. Partl, "Vibration-induced aggregate segregation in asphalt mixtures," *Materials and Structures*., **53** (2), 1-14 (2020). doi.org/10.1617/s11527-020-01459-y
5) T.N. Kennedy, M. Tahmoressi, R.J. Holmgreen, and J.N. Anagnos, "Segregation of asphalt mixtures-causes, identification and cures," 1986. https://trid.trb.org/view/576125
6) M. Stroup-Gardiner, "Methods and practices on reduction and elimination of asphalt mix segregation (No. Project 20-05, Topic 45-12)," 2015. https://www.trb.org/Publications/Blurbs/172730.aspx
7) S. Ilmiati, J. Hafiza, J.F. Fatriansyah, E. Kustiyah, and M. Chalid (2018). "Synthesis and characterization of lignin-based polyurethane as a potential compatibilizer," *Indonesian Journal of Chemistry*., **18** (3), 390-396 (2018). doi.org/10.22146/ijc.27176
8) M. Chalid, A.I. Fikri, H.H. Satrio, M. Joshua, and J.F. Fatriansyah, "An investigation of the melting temperature effect on the rate of solidification in polymer using a modified phase field model," *Int J Technol*., **7** 1321-1328 (2017). doi.org/10.14716/ijtech.v8i7.707
9) R.C. Williams, Jr.G. Duncan, and T.D. White, "Hot-mix asphalt segregation: Measurement and effects," *Transportation Research Record*., **1543** (1) 97-105 (1996). doi.org/10.1177/0361198196154300112
10) T. N., Dief and S. Yoshida, "System Identification for Quad-rotor Parameters Using Neural Network" *Evergreen Joint Journal of Novel Carbon Resource Sciences & Green Asia Strategy*, 3(01), (2016), pp.6-11.
11) A. S. Syahrul, M. G. Normah, J. T. Oh, A. Robiah, and M. Y. Yushazaziah, "Entropy generation minimization of two-phase flow in a mini channel with genetic algorithm." *Evergreen, Joint Journal of Novel Carbon Resource Sciences & Green Asia Strategy* 6, no. 01 (2019): 39-43.
12) I. Choirunisa, F. Imaduddin, E. T. Maharani, G. Priyandoko & Mazlan, S. A. "MR Damper Modeling using Gaussian and Generalized Bell of ANFIS Algorithm." (2021): 673-685.
13) Dief, Tarek N., and Shigeo Yoshida. "System Identification for Quad-rotor Parameters Using Neural Network." *Evergreen, Joint Journal of Novel Carbon Resource Sciences & Green Asia Strategy* 3, no. 01 (2016): 6-11.A.E. Maxwell, T.A. Warner, and F. Fang, "Implementation of machine-learning classification in remote sensing: An applied review," *International Journal of Remote Sensing*, **39** (9) 2784-2817 (2018). doi.org/10.1080/01431161.2018.1433343
14) C. Lin, S. Jiachen, W. Tongjing, Y. Fan, and Z. Tiantong, "A method to evaluate the segregation of compacted asphalt pavement by processing the images of paved asphalt mixture," *Construction and Building Materials*., **224** 622-629 (2019). doi.org/10.1016/j.conbuildmat.2019.07.041
15) M. Baqersad, A. Hamedi, M. Motjaba and A. Hesham, "Asphalt Mixture Segregation Detection: Digital Image Processing Approach," Advances in Materials Science and Engineering, 2017. doi.org/10.1155/2017/9493408
16) S. Walfish, "A review of statistical outlier methods," *Pharmaceutical technology*., **30** (11) 82 (2006). researchgate.net/publication/285807614_A_review_of_statistical_outlier_methods
17) Hodge, V., & Austin, J. "A survey of outlier detection methodologies" *Artificial intelligence review*, **22**(2), 85-126. (2004). doi.org/10.1023/B:AIRE.0000045502.10941.a9
18) Wang, H., Bah, M. J., & Hammad, M. "Progress in outlier detection techniques: A survey" *Ieee Access*, 7, 107964-108000. (2019).
19) Hanada, T. "Modifying the feed-in tariff system in Japan: an environmental perspective" *Evergreen: joint journal of Novel Carbon Resource Sciences & Green Asia Strategy*, 3(2), 54-58. (2016).
20) Singh, K., & Upadhyaya, S. "Outlier detection: applications and techniques" *International Journal of Computer Science Issues (IJCSI)*, 9(1), 307. (2012).
21) C. Leys, O. Klein, P. Bernard, and L. Licata (2013). "Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median," *Journal of Experimental Social Psychology*., **49** (4) 764-766 (2013). doi.org/10.1016/j.jesp.2013.03.013
22) C. Park, and B.R. Cho, "Development of robust design under contaminated and non-normal data," *Quality Engineering*., **15** (3) 463-469 (2003). doi.org/10.1081/QEN-120018045
23) A.C. Müller, and S. Guido, "Introduction to machine learning with Python: A Guide for Data Scientists (1st ed.)," O'Reilly 2017. books.google.co.id/books?id=vbQlDQAAQBAJ&hl=id
24) W. Wu, Y. Mallet, B. Walczak, W. Penninckx, D.L. Massart, S. Heuerding, and F. Erni, "Comparison of regularized discriminant analysis linear discriminant analysis and quadratic discriminant analysis applied to NIR data," *Analytica Chimica Acta* **329** (3) 257-265 (1996). doi.org/10.1016/0003-2670(96)00142-0
25) T. Hastie, R. Tibshirani, and J. Friedman, "The

Elements of Statistical Learning (Data Mining, Inference, and Prediction)," *The Mathematical Intelligencer*., **27** (2) 83-85 2009. doi.org/10.1007/BF02985802

26) Choudhary, S., Sharma, A., Gupta, S., Purohit, H., & Sachan, S." Use of RSM technology for the optimization of received signal strength for LTE signals under the influence of varying atmospheric conditions" *EVERGREEN Joint Journal of Novel Carbon Resource Sciences & Green Asia Strategy*, 7(4), 500-509. (2020)