Utilizing Machine Learning to Minimize Sample Height Errors in Marshall Asphalt Mixture Design Method

Muhammed Yasin Codur
College of Engineering and Technology, American University of the Middle East, Kuwait
muhammed.codur@aum.edu.kw (corresponding author)

Halis Bahadır Kasil
Department of Research and Development, Republic of Turkey-General Directorate of Highways, Erzurum, Turkiye
halis.kasil24@erzurum.edu.tr

Emre Kuskapan
Department of Civil Engineering, Faculty of Engineering and Architecture, Erzurum Technical University, Erzurum, Turkiye
emre.kuskapan@erzurum.edu.tr

Received: 21 November 2024 | Revised: 12 December 2024, 27 January 2025, and 2 March 2025 | Accepted: 6 March 2025

 DOI: https://doi.org/10.48084/etasr.9664

ABSTRACT

In this study, a Machine Learning (ML) method was utilized to predict whether the sample heights of the briquettes prepared during the hot mix asphalt design step (Marshall method) will be within the tolerances specified in ASTM D 6927 standard. Factors affecting the sample height were analyzed using a multilayer perceptron algorithm. In the analysis process, the sample heights of the road layers consisting of base, binder, wearing, and SMA wearing layers were estimated with an accuracy of about 90% or more, demonstrating the high accuracy of the model. As a result, there is high possibility of utilizing ML to prepare asphalt specimens within the required height range.

Keywords-specimen height; optimum bitumen ratio; machine learning; pavement design

I. INTRODUCTION

The Marshall Method is widely utilized to evaluate asphalt pavement deformation and to determine the optimum bitumen content for achieving maximum performance and minimal wheel load deformation. Its simplicity, quick results, and compatibility with California Bearing Ratio (CBR) testing equipment, result in the selection of this technique [1-5].

In this method, specimens with varying bitumen contents are prepared to create performance graphs. Typically, three briquettes are produced for each bitumen ratio near the predicted optimum, with increments of 0.5%. Two ratios are chosen above and below the estimated optimum, resulting at least in 15 specimens.

The procedure for determining the optimum bitumen rate includes the following steps:

- Aggregate selection.
- Bitumen selection.
- Sample preparation (including compaction).
- Measurement of sample height.
- Density and void calculations.
- Selection of optimum bitumen ratio.

Each layer's gradation must comply with the limits specified in the Turkish Highways Technical Specifications [6], as detailed in Table I.

According to ASTM D6926, specimens should have a height of 63.5 mm, achieved by applying 50 or 75 impacts during compaction [7]. However, predicting whether the hot mix asphalt will form molds and briquettes with the desired height of 63.5 ± 2.5 mm after compaction is challenging. To control this, (1) was employed to adjust the mass of the aggregate based on the final specimen height:

$$\text{Adjusted mass of aggregate} = \frac{63.5 \times (\text{mass of aggregate})}{\text{Specimen height}} \quad (1)$$

TABLE I. BASE COURSE, BINDER COURSE, WEARING COURSE, AND SMA WEARING COURSE LIMITS [6].

Sieve sizes		Base course	Binder course	Wearing course	SMA wearing course
Inc No	mm				
1 ½"	37.5	100	-	-	-
1"	25	72-100	100	-	-
3/4"'	19.10	60-90	80-100	100	100
1/2"	12.70	50-78	58-80	88-100	90-100
3/8"	9.52	43-70	48-70	72-90	50-75
No.4	4.75	30-55	30-52	42-52	25-40
No.10	2.00	18-42	20-40	25-35	20-30
No.40	0.420	6-21	8-22	10-20	12-22
No.80	0.177	2-13	5-14	7-14	9-17
No.200	0.075	0-7	2-7	3-8	8-12
Bitumen Ratio		3.5-5.0	3.5-6.5	4.0-7.0	5.8(min.)

Air voids in an asphalt mixture are influenced by several factors including briquette weight, bitumen content, aggregate water absorption, and mixture temperature. These voids, in turn affect the final height of Marshall briquettes, which are compacted using a 101.6 mm diameter mold. If the prepared samples do not meet the required height after compaction as a result of heating and mixing with bitumen, a loss of 7 days can occur during the evaluation phase. In this case, not only time is lost, but also there is economic loss as well as material waste. Moreover, ASTM D6927 does not permit the use of specimens outside the specified height range and does not provide a correction method for such deviations [8].

Despite these challenges, the Marshall test remains the most utilized method to estimate the optimum bitumen content [9-12]. However, as it is based on average values from briquette groups with varying bitumen ratios, it can exhibit high variability. Apart from this, the overall process -from sample preparation to the final result- is time consuming [13-16]. Therefore, enhancing this method with modern technology could enhance the specific limitations.

Although there is limited research directly examining the height of hot mix asphalt specimens, some studies have explored the impact of specimen dimensions on the properties and performance of asphalt mixtures. For example, authors in [17] investigated the effects of Portland cement and cationic slow-setting bitumen emulsion additives on road performance, with the results revealing that a mixture of 4% Portland cement and 3% bitumen emulsion was optimal for specimens with different heights. Additionally, authors in [18] assessed the impact of recycled demolition waste in bitumen emulsion on permanent deformation behavior. The findings demonstrated that specimen height affected deformation responses and that demolition waste enhanced resistance to traffic loading while reducing elastic strain.

During the hot mix asphalt design process, the height of prepared briquettes can vary depending on the mass of the sample. Due to the difficulty of predicting how the aggregate and bitumen mixture will settle and compact at a certain temperature, the final briquette height often exceeds or sometimes remains low, and thus, prolongation of the design period can be caused. This study aimed to predict the briquette height of hot mix asphalt through ML, considering factors such as aggregate mass, bitumen mass, aggregate water absorption, bitumen absorption, specific gravity, preparation temperature, and bitumen ratio.

II. MATERIALS AND DATA PROCESSING

In this study, briquettes were prepared under laboratory conditions for the base course, binder course, wearing course, and SMA wearing course layers. Examples of the sample height of each layer and the factors affecting this height are presented in Table II. The details regarding the height of the briquettes at minimum bitumen content, along with key parameters, such as aggregate mass, bitumen mass, coarse and fine aggregate, water absorption, aggregate bitumen absorption, theoretical maximum specific gravity, mixture preparation temperature, and bitumen ratio for each sample, are available and can be shared upon request.

TABLE II. DATA OF SAMPLES

Specimen height	Aggregate mass in the mixture	Bitumen mass in the mixture	Coarse aggregate water absorption	Fine aggregate water absorption	Aggregate bitumen absorption	Theoretical maximum specific gravity of the mixture	Preparation temperature of the mixture	Bitumen ratio
Data of the base course samples								
63.2	1050	31.5	1.45	1.57	0.79	2.474	145	3.0
63.2	1075	32.3	1.45	1.57	0.72	2.481	155	3.0
64.5	1150	34.5	0.75	0.84	0.41	2.569	135	3.0
Data of the binder course samples								
62.6	1050	36.8	2.01	1.64	0.71	2.450	150	3.5
65.8	1175	47.0	0.71	0.81	0.34	2.535	145	4.0
63.9	1100	38.5	1.08	1.81	0.68	2.653	150	3.5
Data of the wearing course sample.								
65.1	1150	46.0	1.09	1.59	0.57	2.614	155	4.0
64.5	1150	40.3	0.47	0.55	0.23	2.562	140	3.5
64.1	1060	42.4	1.90	1.72	0.69	2.428	145	4.0
Data of the SMA wearing course samples								
63.2	1100	60.5	0.56	0.74	0.35	2.454	134	5.5
63.7	1100	60.5	0.56	0.62	0.32	2.458	135	5.5
63.3	1100	60.5	1.23	1.35	0.65	2.571	133	5.5

The experiment considered four asphalt concrete layer designs with maximum aggregate sizes of 38.0, 25.0, and 19.0 mm, with 84 of the 102 total specimens exhibiting dense gradation while 18 demonstrated gap gradation. Among the dense-graded mixtures, 27 were for base course, 31 for binder course, 26 for wearing course, while the 18 samples with gap gradation were for SMA wearing course. Figure 1 illustrates the equipment for briquette preparation as well as the samples.

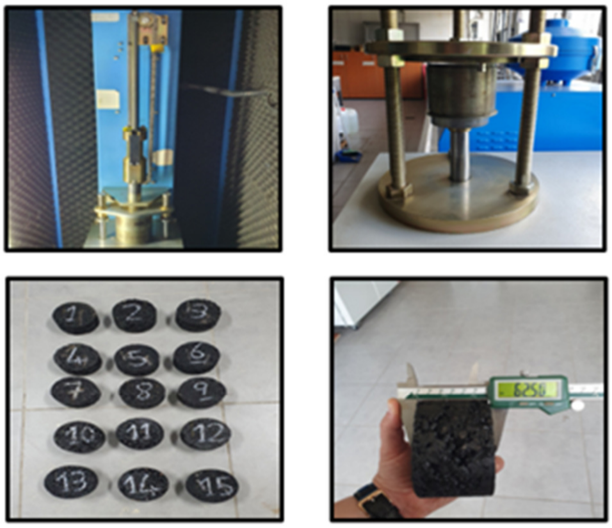

Fig. 1. Marshall experiment process.

B70/100 penetration grade bitumen from Batman Tüpraş refinery was employed, as it is easily available and commonly used in Turkey's hot mix asphalt industry. Bitumen penetration degrees were measured according to TS EN 1426 [19]. Most aggregates were of magmatic origin, with some being sedimentary. Bitumen absorption was determined following the Bituminous Mixtures Laboratory Manual [20], while water absorption and specific gravity were tested according to TS EN 1097-6 [21].

III. MACHINE LEARNING AND MULTILAYER PERCEPTRON

ML is a scientific filed focused on developing algorithms to identify data patterns and predictive models. As more data and experience are accumulated, the accuracy of these models improves [22]. ML is particularly useful when data cannot be easily encoded. There are several criteria to evaluate the results of these algorithms. Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Kappa statistics measure error rates, while precision, recall, and F-measure assess performance. MAE is expressed as the average of the difference between predicted values and actual values of all data while RMSE is calculated by taking the square root of the difference between the values estimated by the model and the actual obtained values [23]. On the other hand, Kappa value is a term for measuring the mismatch between observational values. The closer this value is to 1, the better is the agreement between observations. The kappa statistic is frequently used to test interrater reliability.

Low MAE and RMSE values, along with a high Kappa statistic, indicate a more accurate model. Likewise, high values for precision, recall, and F-measure signify strong classification performance. These metrics are derived by comparing predicted outcomes to actual outcomes using the following terms: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). Equations (2-5) and Table III detail these terms [24-26].

TABLE III. CONFUSION MATRIX

Class	TP	FP	FN	TN
Wearing	29	1	1	89
Binder	29	2	2	91
Base	23	2	2	73
SMA Wearing	17	2	2	55

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (2)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad (3)$$

$$\text{F}-\text{measure} = \frac{2\times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \quad (4)$$

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+FP+FN+TN}} \quad (5)$$

IV. RESULTS AND DISCUSSION

Three classes -ideal, medium, and critical- were created to categorize the sample heights for four different highway layers. Classification was used to simplify the data for easier interpretation. The ideal sample height of 63.5 mm was used as the reference point for classification. As the sample height deviates further from 63.5 mm, it is considered more critical rather than ideal. The classification based on sample height is depicted in Figure 2.

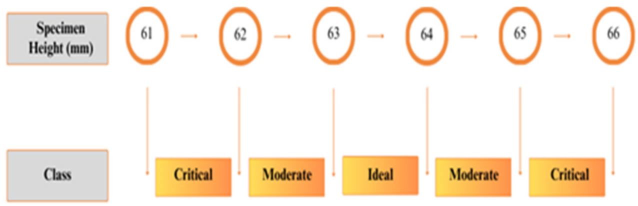

Fig. 2. Classification based on specimen height.

Using the existing classification and data for each layer, an analysis was performed with a multilayer perceptron algorithm. In this process, 70% of the dataset was employed for training, 15% for validation, and 15% for testing.

Figure 3 and Table IV display the performance and error criteria, respectively, for each road layer.

Upon examining the performance values obtained from the algorithm, it was evident that there was a success rate exceeding 90% for each road layer. Among the performance metrics, the F-measure was detected as the most decisive variable in determining the algorithm's success. Based on this criterion, the highest overall accuracy and lowest error values were achieved in the wearing course layer.

Fig. 3. Performance criteria obtained through the multilayer perceptron algorithm.

TABLE IV. ERROR CRITERIA OBTAINED THROUGH THE MULTILAYER PERCEPTRON ALGORITHM

-	Kappa statistic	RMSE	MAE
Wearing course	0.946	0.185	0.131
Binder course	0.899	0.178	0.104
Basecourse	0.878	0.232	0.107
SMA wearing course	0.791	0.259	0.162

As for the error criteria, strong results were also observed, similar to the performance results. The Kappa statistic was highest for the wearing course, indicating strong agreement between predicted and actual values. Meanwhile, the RMSE and MAE values were the lowest in the binder course, further confirming the model's predictive strength.

V. CONCLUSIONS

This study investigated the utilization of Machine Learning (ML) in determining the optimum bitumen content in asphalt layers through Mashall method, in order to avoid time and material waste. A dataset was created considering variables affecting specimen height, and three height classes were identified: ideal, moderate, and critical. The analysis predicted the sample height classes for each road layer (base, binder, wearing, and SMA wearing courses) with an accuracy of at least 90%. Consequently, there is high possibility of enhancing the asphalt application process with ML. By enabling early prediction of specimen heights, this method helps minimize design iteration, as well as reducing reliance on manual trial-and-error. Thus, it supports more accurate determination of optimum bitumen content and contributes to a more sustainable and economically efficient asphalt design process.

ACKNOWLEDGMENT

The authors would like to gratefully acknowledge the Turkish General Directorate of Highways (GDH) for providing access to laboratory results of bitumen samples.

REFERENCES

[1] *MS-2 Asphalt Mix Design Methods*, 7th ed., Asphalt Institute, Lexiington, KY, USA, 2014.

[2] G. Bharath, M. Shukla, M. N. Nagabushana, S. Chandra, and A. Shaw, "Laboratory and field evaluation of cement grouted bituminous mixes," *Road Materials and Pavement Design*, vol. 21, no. 6, pp. 1694–1712, Aug. 2020, https://doi.org/10.1080/14680629.2019.1567375.

[3] L. Gupta and R. Kumar, "Recarpeting using cement grouted bituminous mix in urban flexible pavement: a laboratory and field evaluation," *Australian Journal of Civil Engineering*, vol. 19, no. 2, pp. 235–246, Jul. 2021, https://doi.org/10.1080/14488353.2021.1896125.

[4] M. Shukla, B. Gottumukkala, M. N. Nagabhushana, S. Chandra, A. Shaw, and S. Das, "Design and evaluation of mechanical properties of cement grouted bituminous mixes (CGBM)," *Construction and Building Materials*, vol. 269, Feb. 2021, Art. no. 121805, https://doi.org/10.1016/j.conbuildmat.2020.121805.

[5] K. A. Kaaf and V. T. Ibeabuchi, "Marshall Asphalt Mix and Superior Performance Asphalt Mix in Oman: A Comparative Study," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12258–12263, Dec. 2023, https://doi.org/10.48084/etasr.6206.

[6] *Turkish Highways Technical Specifications*, General Directorate of Highways, Ankara, Turkey, 2013.

[7] *Standard Practice for Preparation of Asphalt Mixture Specimens Using Marshall Apparatus*, ASTM D6926, 2020.

[8] *Standard Test Method for Marshall Stability and Flow of Asphalt Mixtures*, ASTM D6927, 2022.

[9] R. F. Webb, J. L. Burati, and H. S. Hill, "Effect of Specimen Thickness on Marshall Test Results," *Ratio*, vol. 2, pp. 132–140, 1985.

[10] A. E. Gomaa, "Marshall test results prediction using artificial neural network," M.S. Thesis, Arab Academy for Science and Technology, Cairo, Egypt, 2014.

[11] H. I. Ozturk, A. Saglik, B. Demir, and A. G. Gungor, "An artificial neural network base prediction model and sensitivity analysis for marshall mix design," in *6th Eurasphalt & Eurobitume Congress*, Prague, Czech Republic, Jun. 2016.

[12] A. Azarhoosh and S. Pouresmaeil, "Prediction of Marshall Mix Design Parameters in Flexible Pavements Using Genetic Programming," *Arabian Journal for Science and Engineering*, vol. 45, no. 10, pp. 8427–8441, Oct. 2020, https://doi.org/10.1007/s13369-020-04776-0.

[13] N. Baldo, E. Manthos, and M. Pasetto, "Analysis of the Mechanical Behaviour of Asphalt Concretes Using Artificial Neural Networks," *Advances in Civil Engineering*, vol. 2018, no. 1, 2018, Art. no. 1650945, https://doi.org/10.1155/2018/1650945.

[14] W. F. Liu, H. M. Li, and B. P. Tian, "Research on Designing Optimum Asphalt Content of Asphalt Mixture by Calculation and Experimental Method," *Applied Mechanics and Materials*, vol. 97–98, pp. 23–27, 2011, https://doi.org/10.4028/www.scientific.net/AMM.97-98.23.

[15] O. Kaya, "Development of Neural Network-Based Asphalt Mix Design Parameters Prediction Tool," *Arabian Journal for Science and Engineering*, vol. 48, no. 10, pp. 12793–12804, Oct. 2023, https://doi.org/10.1007/s13369-022-07579-7.

[16] M. A. Çolak, E. Zorlu, M. Y. Çodur, F. İ. Baş, Ö. Yalçın, and E. Kuşkapan, "Investigation of Physical and Chemical Properties of Bitumen Modified with Waste Vegetable Oil and Waste Agricultural Ash for Use in Flexible Pavements," *Coatings*, vol. 13, no. 11, Nov. 2023, Art. no. 1866, https://doi.org/10.3390/coatings13111866.

[17] M. S. Baghini, A. B. Ismail, M. R. B. Karim, F. Shokri, and A. A. Firoozi, "Effects on engineering properties of cement-treated road base with slow setting bitumen emulsion," *International Journal of Pavement Engineering*, vol. 18, no. 3, pp. 202–215, Mar. 2017, https://doi.org/10.1080/10298436.2015.1065988.

[18] E. Yaghoubi, B. Ghorbani, M. Saberian, R. van Staden, M. Guerrieri, and S. Fragomeni, "Permanent deformation response of demolition wastes stabilised with bitumen emulsion as pavement base/subbase," *Transportation Geotechnics*, vol. 39, Mar. 2023, Art. no. 100934, https://doi.org/10.1016/j.trgeo.2023.100934.

[19] *Determination of needle penetration depth*, TS EN 1426, 2015.

[20] M. Gültekin, N. Nayır, U. Ziya, K. K. Çalışkan, A. Öztürk, S.N. Tutan and M. Komut, "Bituminous Mixtures Laboratory Manual," General directorate of highways, Ankara, Turkey. 2021.

[21] *Tests for mechanical and physical properties of aggregates - Part 6: Determination of particle density and water absorption,* EN 1-97-6, 2022.

[22] K. Goswami and A. B. Kandali, "Machine learning algorithms for predicting electrical load demand: an evaluation and comparison,"

Sādhanā, vol. 49, no. 1, Jan. 2024, Art. no. 40, https://doi.org/10.1007/s12046-023-02354-2.

[23] E. Kuşkapan, M. A. Sahraei, and M. Y. Çodur, "Classification of Aviation Accidents Using Data Mining Algorithms," *Balkan Journal of Electrical and Computer Engineering*, vol. 10, no. 1, pp. 10–15, Jan. 2022, https://doi.org/10.17694/bajece.793368.

[24] C. Choi, S. Park, and J. Kim, "Uniqueness of multilayer perceptron-based capacity prediction for contributing state-of-charge estimation in a lithium primary battery," *Ain Shams Engineering Journal*, vol. 14, no. 4, Apr. 2023, Art. no. 101936, https://doi.org/10.1016/j.asej.2022.101936.

[25] D. W. Ruck, S. K. Rogers, and M. Kabrisky, "Feature selection using a multilayer perceptron," *Journal of neural network computing*, vol. 2, no. 2, pp. 40–48, 1990.