

# 아스팔트 혼합물의 골재 간극률 예측을 위한 기계학습 프레임워크

## Machine Learning Framework for Predicting Voids in the Mineral Aggregation in Asphalt Mixtures

박혜민<sup>1</sup>, 나일호<sup>2</sup>, 김현환<sup>3</sup>, 지봉준<sup>4\*</sup>

Hyemin Park<sup>1</sup>, Ilho Na<sup>2</sup>, Hyunhwan Kim<sup>3</sup>, Bongjun Ji<sup>4\*</sup>

<sup>1</sup>Non-Member, Graduate Student, Department of Regional Infrastructure Engineering, Kangwon National University, 1 Kangwondaehak-gil, Chuncheon-si, Gangwon-do 24341, Republic of Korea

<sup>2</sup>Non-Member, Team Manager, Management Planning Dept., Korea Petroleum Industries Company, 166 Ichon-ro, Yongsan-gu, Seoul, 04427, Republic of Korea

<sup>3</sup>Non-Member, Assistant Professor, Department of Engineering Technology, Texas State University, San Marcos, Texas, USA

<sup>4</sup>Member, Assistant Professor, Department of Regional Infrastructure Engineering, Kangwon National University, 1 Kangwondaehak-gil, Chuncheon-si, Gangwon-do 24341, Republic of Korea

### ABSTRACT

The Voids in the Mineral Aggregate (VMA) within asphalt mixtures play a crucial role in defining the mixture's structural integrity, durability, and resistance to environmental factors. Accurate prediction and optimization of VMA are essential for enhancing the performance and longevity of asphalt pavements, particularly in varying climatic and environmental conditions. This study introduces a novel machine learning framework leveraging ensemble machine learning model for predicting VMA in asphalt mixtures. By analyzing a comprehensive set of variables, including aggregate size distribution, binder content, and compaction levels, our framework offers a more precise prediction of VMA than traditional single-model approaches. The use of advanced machine learning techniques not only surpasses the accuracy of conventional empirical methods but also significantly reduces the reliance on extensive laboratory testing. Our findings highlight the effectiveness of a data-driven approach in the field of asphalt mixture design, showcasing a path toward more efficient and sustainable pavement engineering practices. This research contributes to the advancement of predictive modeling in construction materials, offering valuable insights for the design and optimization of asphalt mixtures with optimal void characteristics.

### 요 지

골재 간극률은 구조적 강도, 내구성, 배수 및 투수성 등 다양한 아스팔트의 특성에 직접적인 영향을 미친다. 따라서 아스팔트 포장에 사용되는 위치, 기후, 환경 등에 적절하도록 골재 간극률이 설계되어야 한다. 하지만 골재 간극률은 다양한 요인들에 의해 영향을 받으므로 그 설계가 쉽지 않다. 예를 들어 골재 입자의 크기 분포, 구성이나 아스팔트 바인더의 양, 다짐 수준 등 다양한 영향인자가 존재한다. 본 연구에서는 골재 간극률에 영향을 미치는 요인들로부터 골재 간극률을 예측하고자 하였다. 이를 위해 다양한 기계학습 모델 방법을 적용하였고 단일 기계학습 모델을 적용했을 때보다 높은 정확도로 골재 간극률을 예측할 수 있음을 보였다. 본 연구의 결과는 경험과 노동집약적인 실험에 의존하는 골재 간극률 예측에 데이터 기반의 접근방법을 적용할 수 있음을 보였으며 향후 최적 골재 간극률 설계 등에 활용 가능할 것으로 기대된다.

**Keywords** : Voids in the mineral aggregate, Pavement, Construction material, Machine learning, Ensemble

## 1. 서 론

골재 간극률(Voids in Mineral Aggregate; VMA)은 아스팔트 혼합물 전체 체적에서 골재를 제외한 부분의 체적

Received 17 Feb. 2024, Revised 4 Mar. 2024, Accepted 5 Mar. 2024

\*Corresponding author

Tel: +82-33-250-6464; Fax: +82-33-242-7355

E-mail address: [bjji@kangwon.ac.kr](mailto:bjji@kangwon.ac.kr) (B. Ji)

을 일컫는 용어로 전체 체적에서 아스팔트와 공극이 차지하고 있는 체적의 비율을 말하며 아스팔트 혼합물의 성능에 영향을 미치는 주요 요소다(Lavin, 2003). 골재 간극률은 구조적 강도와 내구성에 영향을 미치는데 적절한 골재 간극률은 아스팔트 혼합물이 충분한 구조적 강도를 갖추고 장기간 내구성을 유지하도록 돕는다(Kandhal and Chakraborty, 1996; Li et al., 2004; Foreman, 2008). 그에 반해 너무 낮은 골재 간극률은 아스팔트 바인더가 골재를 둘러싸지 못하게 하여 혼합물의 내마모성과 방수성 저하를 야기한다. 적절한 골재 간극률은 아스팔트 혼합물의 수분 손상에 대한 저항성을 높이는데 수분이 골재와 바인더 사이에 침투하는 것을 방지하여 동결-융해 사이클과 같은 환경적 조건에 대한 저항성을 강화시킨다(Chadbourne et al., 1999). 이처럼 골재 간극률은 아스팔트 혼합물의 성능에 직접적인 영향을 미치는 주요인자이지만 다양한 요인의 복합적인 영향으로 그 값이 결정되므로 설계 시 이를 예측하는 것은 어려운 일이다(Wu et al., 2023). 골재 간극률은 골재의 크기나 분포, 모양, 표면 질감, 골재의 배합, 밀도, 그리고 혼합 과정 등 다양한 요인에 의해 영향을 받으며(Sengoz and Topai, 2007) 이러한 요인들의 상호작용이 모두 밝혀진 것이 아니기 때문에 예측을 어렵게 한다. 이러한 상호작용의 영향을 정확히 파악하고 모델링하는 것은 전통적인 경험적 방법으로는 한계가 있으며(Anderson and Bentsen, 2001; Inoue et al., 2004) 선행 연구들에서도 골재 간극률 예측은 매우 도전적인(Challenging) 분야로 남아있다고 기술하고 있다(Kandhal et al., 1998). 본 연구에서는 이러한 한계점을 극복하기 위해 아스팔트 혼합물의 골재 간극률 예측을 위한 기계학습 프레임워크를 제안한다. 기계학습 방법론은 데이터에서 숨겨진 패턴과 상관관계를 추출하는 도구로 아스팔트 바인더의 물성 예측은 물론 건전성 관리 등 토목 및 재료 분야에도 널리 사용되고 있다(Malekloo et al., 2022; Ji et al., 2023). 본 연구의 목적은 아스팔트 혼합물 설계 과정에서의 시행착오를 줄이고 최적화된 아스팔트 혼합물 설계를 통해 골재 간극률이 적절히 관리될 수 있도록 도와 아스팔트 포장의 성능과 내구성을 향상시키는 것이다. 이를 위해 기존 문헌을 분석하고 한계점을 도출하였다. 또한 공개된 실험 데이터를 수집하였으며 골재, 아스팔트, 다짐 특성 등으로부터 골재 간극률을 예측하는 다양한 기계학습 모델을 구축하였다. 사용된 기계학습 방법으로는 인공신경망, 서포트 벡터 머신과 같은 비양상블 모델, 그리고 랜덤포레스트와 XGBoost

와 같은 양상블 모델을 포함한다. 학습된 모델은 Mean Squared Error(MSE), Mean Absolute Percentage Error(MAPE), 결정계수( $R^2$ )의 세 가지 평가지표를 통해 평가하여 가장 효과적인 기계학습 모델을 선정하였다.

## 2. 문헌 연구

### 2.1 골재 간극률 예측

골재 간극률 예측의 중요도가 높은만큼 선행 연구에서는 다양한 방법을 통해 골재의 간극률을 예측하고자 하였다. Zhang(2009)은 Overall gradation prediction method를 제안함으로써 골재 간극률을 예측하고자 하였다. 종합 그라데이션 예측 방법(Overall gradation prediction method)은 골재의 체 통과 비율을 통해 골재 간극률을 예측하는 방법으로 현무암과 석회암 등의 골재를 이용하는 고온 혼합 아스팔트의 골재 간극률을 예측하기 위해 식 (1)과 같이 14개의 변수를 이용하였다.

$$\begin{aligned} \text{골재간극률} = & 10.7 - 00.0193x_1 + 0.04138x_2 - 0.06001x_3 \\ & + 0.03735x_4 - 0.04294x_5 - 0.01393x_6 \\ & - 0.05843x_7 + 0.0385x_8 + 0.2596x_9 - 0.1064x_{10} \\ & - 0.1317x_{11} + 0.0856x_{12} - 165.5x_{13} + 164.4x_{14} \end{aligned} \quad (1)$$

여기서  $x_1$ 부터  $x_{12}$ 는 26.5mm, 19mm, 16mm, 13.2mm, 9.5mm, 4.75mm, 2.36mm, 1.18mm, 0.6mm, 0.3mm, 0.15mm, 0.075mm 체를 통과한 비율을 나타내며  $x_{13}$ 과  $x_{14}$ 는 각각 골재의 골재 비중(Bulk specific density of the aggregate)와 골재의 겉보기 비중(Apparent specific density of the aggregate)를 나타낸다. 이 선행 연구는 골재의 간극률을 예측하기 위한 직관적인 방법을 제공하였으나 변수들간의 상관관계나 기타 골재의 간극률에 영향을 미치는 요인을 고려하지 못하고 있다는 한계점이 있다. 종합 그라데이션 예측 방법 외에도 점진적 그라데이션 예측 방법(Gradual gradation prediction method)도 존재하는데 점진적 그라데이션 예측 방법은 골재를 크기에 따라 여러 등급으로 나누며 특정 크기의 기본 골재를 선택한다. 그리고 더 크거나 작은 골재가 기본 골재에 추가되며 기본 골재의 공극이 변화한다고 가정한다. 즉 공극의 감소량은 두 골재 사이의 부피 간섭 관계에 따라 상이하다고 가정하는 것이다. Shen and Yu(2011)은 기본 골재의 크기를 최대 골재

크기로 사용하며, 기본 골재와 크기가 다른 골재 사이의 간섭 요인(Interference)을 정의하여 골재의 간극률을 예측하고자 하였다. 그들이 제안한 방법에 따르면 간섭계수 ( $f_v$ )는 새로 첨가된 골재의 벌크 부피에 새로운 골재크기를 더한 후 변화된 공극 부피의 비율을 뜻한다. 또한 간섭계수는 골재 입도에 영향을 받으며 골재 간극률은 식 (2)와 같이 계산할 수 있다고 주장하였다.

$$\text{골재 간극률} = \frac{\sum_{i=1}^n f_{vi} V_{ai}}{\sum_{i=1}^n (1 + f_{vi}) + V_{ai}} \quad (2)$$

여기서  $f_{vi}$ 는  $i$ 번째 체 크기의  $f_v$ 값이며,  $V_{ai}$ 는  $i$ 번째 체 크기를 통과하지 못한 골재의 부피 백분율이다. 그러나 이러한 접근법의 핵심이 되는 간섭 요인은 실제 상황에서 정확도가 떨어진다는 한계점이 존재한다(Wu et al., 2023). 대표적인 두 선행 연구의 경우 골재의 입도를 기반으로 골재의 간극률을 예측하고자 하였으나 골재의 입도 외에 영향을 미치는 인자들, 예컨대 아스팔트의 등급, 아스팔트 혼합 프로세스의 변수 등을 반영하지 못하는 한계점이 존재한다. 이에 본 연구에서는 보다 다양한 인자들을 활용하여 아스팔트 혼합물의 골재 간극률을 예측하고자 한다.

## 2.2 기계학습 기법 및 앙상블 기법

기계학습 기법은 인공지능의 한 분야로 데이터에서 패턴을 자동으로 추출한 후, 새로운 데이터가 입력되었을 때 추출된 패턴을 바탕으로 예측이나 의사결정을 내리는 기법이다(Ji, 2021). 명시적으로 프로그래밍을 하지 않는다

는 점에서 프로그래밍 기반의 인공지능과 차이가 있으며 인공지능망, 서포트 벡터 머신, 의사결정 나무 등 패턴을 추출하기 위한 다양한 알고리즘들이 존재한다.

기계학습 기법은 Fig. 1과 같은 공통적인 절차를 따른다. Off-line phase는 데이터에서 패턴을 추출하는 Phase로, 이미 구축된 데이터(Historical data)에서 패턴을 추출한다. On-line phase에서는 Off-line phase에서 학습된 모델을 활용하는 Phase로, 기학습된 모델을 이용하여 학습에 이용되지 않은 새로운 데이터로부터 값을 예측한다.

대표적인 기계학습 모델로는 인공지능망(Neural network)과 서포트 벡터 머신(Support vector machine)이 있다. 인공지능망은 인간의 뇌가 정보를 처리하는 방식을 모방한 기계학습 모델로 입력층, 은닉층, 출력층으로 구성된 노드의 네트워크이다(Goodfellow et al., 2016). 각 노드는 가중치와 활성화 함수를 통해 데이터를 처리하고 네트워크를 통해 전파하는데 학습과정에서는 출력값과 실제값이 일치하도록 가중치와 편향의 값이 조정된다. 서포트 벡터 머신은 데이터를 나누는 최적의 경계(결정 경계)를 찾는 것을 목표로 하며 마진을 최적의 방식으로 학습된다(Cortes and Vapnik, 1995). 기계학습 모델의 성능을 높이기 위해 널리 사용되는 기법 중 하나가 복수의 기계학습 모델을 활용하는 방법이다. 이를 앙상블 기반의 방법론이라고 하며 앙상블 기반의 방법은 경험적으로 하나의 모델보다 더 좋은 성능을 보이는 것으로 알려져 있다(Opitz and Maclin, 1999). 앙상블 기법은 복수의 기계학습 모델을 이용하는 방법에 따라 배깅(Bootstrap aggregating; Bagging)과 부스팅(Boosting)으로 구분된다. 배깅은 부트스트랩 샘플링을 통해 원본 데이터에서 복수 개의 데이터 샘플을 추출한 후 각 데이터 샘플에서 독립적인 예측 모델을 병렬로 학습 시킨 후 이들의 결과를 종합하여 최종 예측을 수행하는 방법이다.

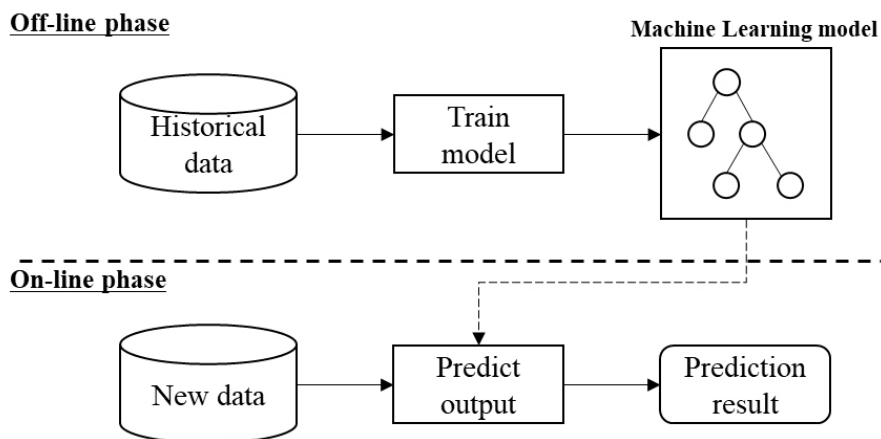


Fig. 1. Concept of machine learning model training and deployment

결과를 종합하는 방법은 회귀 문제인 경우 결과값들의 평균값을 산출하거나, 분류 문제인 경우에는 다수결로 최종값을 결정하는 방식을 취한다. 복수 개의 의사결정나무(Decision tree)를 기반 모델로 사용하는 랜덤포레스트(Random forest)가 대표적인 배깅 모델이다(Ho, 1995). 반면 부스팅은 여러 개의 약한 학습자(Weak Learner)를 순차적으로 학습하며 이전 학습자가 잘못 예측한 데이터에 더 높은 가중치를 부여하여, 새로운 학습자는 잘못 예측된 데이터를 더 잘 예측하도록 하는 방법이다. 즉, 첫 번째 모델을 학습한 후, 잘못 예측된 데이터에 대해 가중치를 증가시킨 후, 수정된 데이터에 대해 학습하는 과정을 반복하는 것이다. AdaBoost(Adaptive Boosting), Gradient Boosting, XGBoost(eXtreme Gradient Boosting)등이 대표적인 부스팅 방법(Freund and Schapire, 1997; Friedman, 2001; Chen and Guestrin, 2016)이다. 본 연구에서는 기계학습 모델 중 비양상불 모델과 양상불 모델 모두를 활용하여 성능을 비교하고, 최적의 예측모델 개발을 위한 방법을 제안하고자 한다.

### 3. 연구 방법 및 절차

본 장에서는 연구 방법 및 절차를 소개한다. 골재의 간극률을 예측하기 위한 방법은 Fig. 2와 같은 연구 프레임워크에 따라 진행되었다.

첫 번째로 데이터를 수집한다. 수집되는 데이터는 아스팔트 침입도, 아스팔트 공용성 등급 등 아스팔트의 특성과

골재 절대간조밀도, 흡수율 등과 같은 골재의 특성, 그리고 다짐횟수, 다짐 방법과 같은 공정에 관한 특성들을 모두 포함한다. 데이터의 수집은 현장 적용 및 실험실 품질관리용 공인시험성적서를 기반으로 하여 아스팔트 바인더의 경우 제품을 취급하는 정유사 및 개질 아스팔트 공급사 등으로부터 수집하고, 아스팔트 혼합물은 국내 도로포장 현장 및 실험실, 아스콘사 등을 통해 수집되었다. Fig. 3은

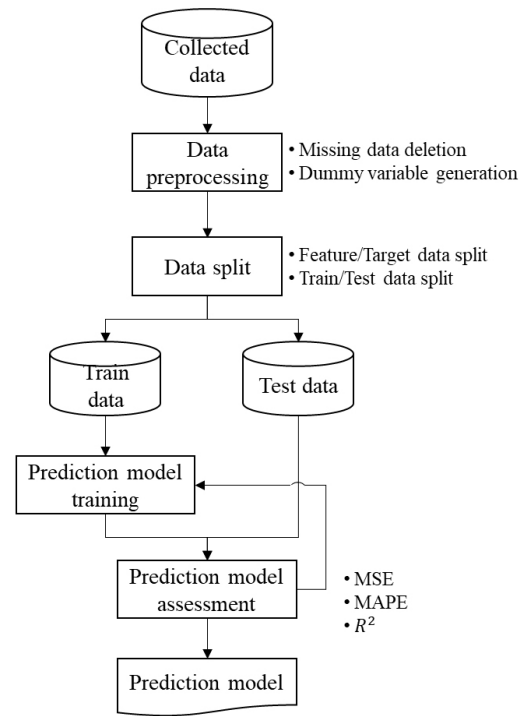


Fig. 2. Proposed framework to develop prediction model

Product name	Asphalt AP-5 PG 64-22	Product code		Date of publication	2023
Tank number		Batch number		Collection date and time	2023

Experiment items	Experiment method	Product specification	Experiment result	Note
Density, @15°C kg/m <sup>3</sup>	KSM 2002	1000 or more	1041	
Ductility (15°C) cm	KSM 2254	100 or more	140+	
Flash point, COC °C	KSM 2010	260 or more	344	
Foam on heat, 180°C	KSM 2201	None	None	
Loss on heat, 180°C	KSM 2258	0.6 or less	0.0	
Thin film heating test (penetration residue)(%)	KSM 2252	55 or more	63	
Penetration ratio after evaporation(%)	KSM 2252	110 or less	106	
Penetration ratio	KSM 2252	between 60-80	72	
Toluene soluble mass%	KSM 2202-2	99.0 or more	99.9	
Softening point °C	KSM 2250	between 44.0 – 52.0	47.5	
kinematic viscosity @120°C mm <sup>2</sup> /s	ASTM D4402		951.0	
kinematic viscosity @135°C mm <sup>2</sup> /s	ASTM D4402	To be reported	427.7	
kinematic viscosity @150°C mm <sup>2</sup> /s	ASTM D4402	To be reported	216.0	
kinematic viscosity @180°C mm <sup>2</sup> /s	ASTM D4402	To be reported	72.2	
PG-Highest design temperature (7-day average) °C	KSF 2389	To be reported	64	
PG-minimum design temperature °C	KSF 2389	64 or more	-22	
		22 or less		

Fig. 3. Sample of collected data of asphalt binder

수집된 데이터를 재가공한 예시로 아스팔트 바인더에 대한 밀도, 신도, 인화점, Foam on heat, Loss on heating, 박막가열시험, 증발 후 침입도비, 침입도, 톨루엔가용분, 연화점, 동점도, 최저/최고 설계온도 등과 같은 항목에 대한 실험 방법, 요구 제품 스펙, 실험 결과 등에 대한 정보가 제공되고 있다.

수집된 데이터는 전처리 과정을 거치며 결측치를 제거한다. 대부분의 기계학습 알고리즘은 입력된 데이터가 완전하다고 가정한다. 따라서 결측치가 존재하는 경우, 학습 알고리즘이 제대로 작동하지 않거나 오류를 발생시켜 예측 성능이 저하된다. 또한 결측치는 모델의 편향성을 유발하여 일반화 능력을 저하시킬 수 있다. 수집된 데이터의 특성상 모든 데이터 값이 존재하지 않으므로 결측치가 존재하는 경우 해당 데이터를 삭제하였다.

그 후, Fig. 4와 같은 과정을 거쳐 데이터를 분할하였다. Feature들의 값들을 통해 Target 값을 예측하는 것이 목표이므로 예측하고자 하는 아스팔트 간극률 값을 Target 값으로 설정하고, 실험 전 획득 가능한 데이터 값들을 Feature 값으로 설정하여 분리하였다. 이후 이 데이터를 학습데이터와 테스트 데이터로 나누었다. 학습 데이터는 예측 모델 학습에 사용되며, 테스트 데이터는 예측 모델의 최종 성능 평가에 사용된다. 이를 통해 학습에 활용되지 않은 데이터에 대해 얼마나 잘 작동하는지 평가할 수 있으며 학습데이터에만 과도하게 적합되는 오버피팅(Overfitting) 현상을 방지할 수 있다.

본 연구에서는 다음과 같은 세 가지 성능 평가 지표를 이용하여 간극률 예측 모델의 성능을 평가하였다. 첫 번째는 Mean Squared Error(MSE), 두 번째는 Mean Absolute Percentage Error(MAPE), 세 번째는  $R^2$ 이다. MSE는 평균

제곱오차로 관측값과 예측값 간의 차이를 제공한 후, 이를 모든 데이터 포인트에 대해 평균한 값으로 아래의 식 (3)과 같이 산출된다.

$$\text{평균제곱오차}(MSE) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3)$$

MAPE는 관측값 대비 모델 예측값의 절대 오차를 백분율로 나타낸 후, 이를 각 데이터 포인트에 대해 평균한 값으로 다음과 같은 식 (4)을 이용하여 산출된다.

$$\text{예측값의 절대오차 백분율}(MAPE) = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (4)$$

이 때  $y_i$ 는  $i$ 번째 데이터를 나타내고  $\hat{y}_i$ 는  $i$ 번째 관측치에 대한 예측값을 나타낸다.  $n$ 은 데이터의 개수를 나타낸다. 정리하자면 MAPE는 실제 데이터 대비 예측 값의 차에 대한 비율을 산출하게 된다. MSE는 오차의 크기에 민감하게 반응하여 큰 오차의 경우 MAPE보다 상대적으로 큰 패널티를 부여하는 특성이 있다. 세 번째 성능 평가 지표는  $R^2$ 로 결정계수(Coefficient of determination)이다.  $R^2$ 는 회귀 모델의 성능을 평가하기 위해 널리 사용되며 예측 모델이 데이터의 변동성을 설명하는 정도를 정량적으로 나타내는데 아래와 같은 식으로 산출된다.

$$\text{결정계수}(R^2) = 1 - \frac{SS_{res}}{SS} \quad (5)$$

여기서  $SS_{res}$ 는 식 (6)과 같은 실제값과 예측값의 차이의 제

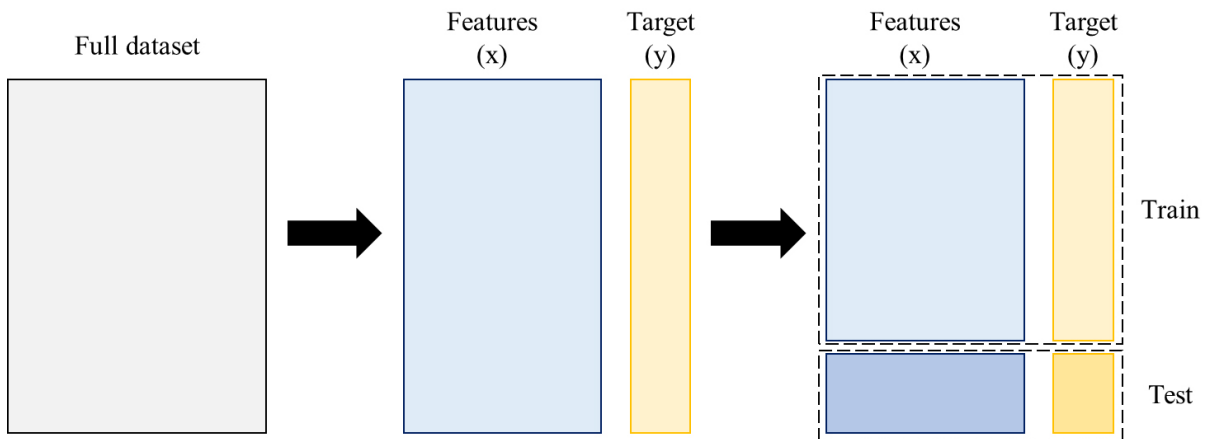


Fig. 4. Illustration of data split process

곱합,  $SS$ 는 식 (7)과 같이 실제값과 실제값들의 평균( $\bar{y}$ )간 차이의 제곱합으로  $R^2$ 가 0에 가까우면 예측모델이 데이터의 변동성을 전혀 설명하지 못한다는 의미이며 1에 가까우면 예측모델이 데이터의 변동성을 완벽하게 설명한다는 의미이다.

$$SS_{res} = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (6)$$

$$SS \quad (7)$$

본 연구에서는 각기 다른 특징을 가지는 세 가지 성능 평가지표를 활용하여 예측 모델의 성능을 다각도로 평가하고 보다 정확한 성능 평가를 수행하고자 하였다.

## 4. 간극률 예측 결과 및 분석

### 4.1 간극률 예측 모델 학습

본 연구에서는 Table 1과 같은 환경에서 간극률 예측 모델을 개발하고 평가하였다. 하드웨어 환경은 CPU는 i7-12700K, Memory는 64.0GB, GPU는 Geforce RTX 3070 Ti로 구성되어 있으며 소프트웨어 환경은 Window 10에서 Python 3.10.10버전을 사용하여 Jupyter Notebook에서 코드를 작성하였다. 사용된 주요 라이브러리는 Numpy 1.23.5, Pandas 2.0.3, Scikit-learn 1.2.2 버전이다. Scikit-learn은 다양한 기계학습 모델 개발을 위해 필요한 함수를 제공한다.

예측 모델 학습에 사용된 Feature는 앞선 장에서 언급하였듯이 결측치가 있는 Feature를 제외하고 다음과 같다; 아

스팔트 침입도(25°C), 아스팔트 공용성 등급, 절대 건조 밀도( $g/m^3$ ), 잔골재 흡수율(%), 잔골재 안정성(%), 잔골재(0.3mm), 잔골재(0.15mm), 잔골재(0.08mm), 채움재 박리 저항성, 채움재 소성지수(%), 채움재 흐름시험(%), 채움재 비중, 채움재(0.6mm), 채움재(0.3mm), 채움재(0.15mm), 채움재(0.08mm), 배합비(합성입도 13mm), 배합비(합성입도 10mm), 배합비(합성입도 5mm), 배합비(합성입도 2.5mm), 배합비(합성입도 0.3mm), 배합비(합성입도 0.15mm), 배합비(합성입도 0.08mm), 다짐횟수, 다짐방법.

기계학습 알고리즘 별로 특성이 상이하므로 골재의 간극률을 예측하기 위해 적합한 기계학습 알고리즘을 선택하기 위해 인공신경망, 서포트 벡터 머신, 랜덤 포레스트, XGBoost 등 다양한 기계학습 알고리즘을 구현하고 적용하였다.

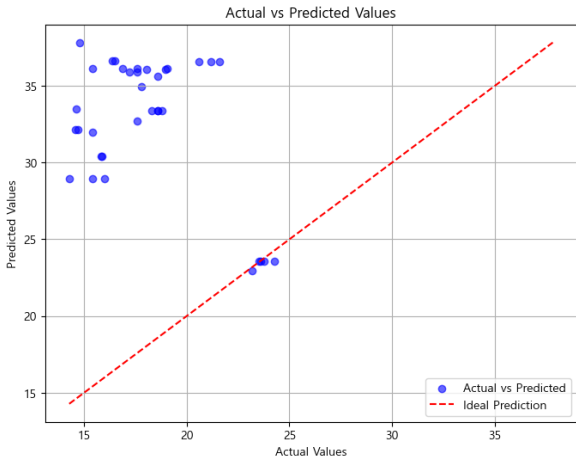
### 4.2 간극률 예측 모델 결과 분석

동일한 학습 데이터로 인공신경망, 서포트 벡터 머신, 랜덤 포레스트, XGBoost의 네 가지 예측 모델을 학습한 후, 마찬가지로 동일한 테스트 데이터로 데이터를 통해 각 모델의 예측 성능을 확인하였다. Fig. 5는 각 모델 별로 실제값과 예측값 간의 관계를 나타내는 산점도(Scatter plot)이다. 산점도에서 각 점은 개별 테스트 데이터를 나타내며 점의 x축은 실제값에 해당하며 y축은 예측 모델에 의해 예측된 값을 나타낸다.

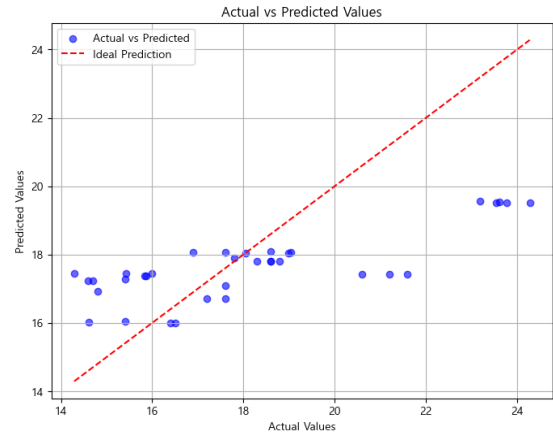
빨간색 대각선은 이상적인 예측선을 나타내며 파란색 점들이 빨간색 대각선에 가까울수록 예측 모델의 예측값이 실제 값에 가깝다는 것을 의미한다. 즉, 예측 모델의 성능이 높은 것을 나타낸다. 반대로 산점도 내에 대각선으로부터 멀리 떨어진 점들은 예측값과 실측값의 오차가 큰 데이터를 나타낸다. 이러한 점들은 예측모델의 골재 간극률 예측이 부정확했음을 보여준다. 정리하자면, 대각선에 가깝게 점들이 밀집해있을수록 예측모델의 예측 정확도가 높으며, 예측 모델이 정확하게 골재 간극률을 예측했다는 것을 의미한다. 앙상블 모델이 아닌 인공신경망과 서포트 벡터 머신의 경우 대각선에서 멀리 떨어진 점들이 다수 존재한다. 인공신경망의 경우 예측값이 실제값보다 과대하게 나타났으며 28에서 38사이로 간극률을 예측하는 것으로 나타났다. 서포트 벡터 머신의 경우 골재 간극률이 16에서 19 사이의 값을 예측하며 실제값의 변동성을 반영하지 못하는 것으로 보인다. 반면 앙상블 기반의 방법

Table 1. Development environment used for the experiment

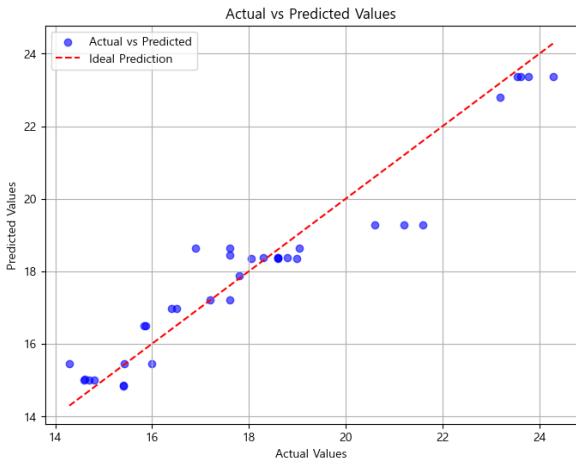
Component		Specification
Hardware	Processor	i7-12700K
	Memory	64.0GB
	GPU	Geforce RTX 3070 Ti
Software	Operating System	Window 10
	Programming Language	Python 3.10.10
	Development Tools	Jupyter Notebook
	Libraries	Numpy 1.23.5 Pandas 2.0.3 Scikit-learn 1.2.2



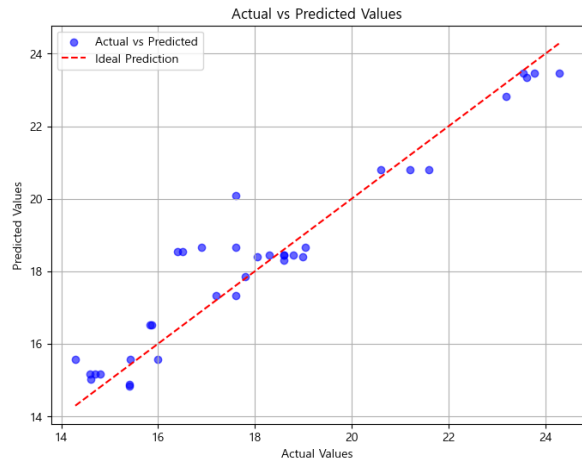
(a) Neural network



(b) Support vector machine



(c) Random forest



(d) XGBoost

Fig. 5. Visualization of predicted values by machine learning algorithm with actual values

Table 2. Prediction performance by machine learning model

Model	Ensemble Status	Performance metric		
		MSE	MAPE	$R^2$
Neural network	Non-Ensemble	0.86	248.99	-28.7
Support vector machine	Non-Ensemble	0.10	5.18	0.38
Random forest	Ensemble	0.03	0.64	0.92
XGBoost	Ensemble	0.04	0.75	0.91

인 Random forest와 XGBoost의 경우 대부분의 데이터 포인트가 대각선에 가깝게 위치하고 있음을 알 수 있다. 즉 대부분의 경우 예측값이 실측값과 유사하다. 이는 MSE, MAPE,  $R^2$ 를 통해 성능 평가한 결과에도 명확히 드러난다(Table 2).

비앙상블 모델인 인공신경망과 서포트 벡터 머신은 MSE와 MAPE에서 모두 높은 오차를 보인다. 인공신경망은 MSE가 0.86, MAPE가 248.99로 예측성능이 매우 낮음을 나타낸다. 특히  $R^2$ 는 0보다 작은 값으로 모든 예측값을

관측치의 평균으로 예측했을때보다 낮은 예측 정확도를 보임을 알 수 있다. 서포트 벡터 머신은 인공신경망에 비하여 낮은 MSE와 MAPE를 보이며 높은  $R^2$ 를 보였지만 MAPE를 통해 해석 가능하듯이 예측값은 평균 5%의 오차가 존재하기에 절대적인 예측 성능이 높다고 볼 수 없다. 반면 앙상블 기반의 방법인 Random forest와 XGBoost는 모든 성능지표에서 비앙상블 방법인 인공신경망과 서포트 벡터 머신에 비해 높은 예측 성능을 보였다. 특히 Bagging 방법 기반의 Random forest는 가장 높은 성능을 보였다. 이를 통해 Random forest 기반의 골재 간극률 예측 모델이 가장 높은 정확도를 보인다는 것을 알 수 있다. Random forest가 가장 높은 정확도를 보인 이유는 골재 간극률 예측과 같이 여러 요소들이 복합적으로 작용할 때 Random forest가 데이터 내 존재하는 비선형 패턴과 상호작용을 효과적으로 포착하였으며 여러 결정 나무를 이용했기 때문에 과적합의 위험을 줄였기 때문으로 판단된다.

## 5. 결론

본 연구에서는 골재 간극률 예측을 위해 기계학습 기반의 접근방안을 제안하고 실제 아스팔트 혼합물의 데이터를 통해 제안된 접근방안을 검증하였으며 그 결과를 요약하면 다음과 같다.

- (1) 기존의 연구는 골재의 입도만을 고려하였으며 예측 방법 또한 단순 선형 회귀식이나 간단한 수식에 그쳤다. 본 연구에서는 기존 연구와 달리 골재의 입도 외의 다양한 요소를 고려하였으며 요소간의 상호작용이나 비선형적인 관계도 반영할 수 있도록 기계학습 기반의 접근방안을 통해 아스팔트 혼합물의 골재 간극률을 예측하였다.
- (2) 예측 모델의 성능 평가는 평가 방법에 따라 성능 해석 결과가 상이할 수 있다. 본 연구에서는 예측모델 평가 알고리즘의 평가 방법 중 MSE, MAPE,  $R^2$ 를 이용하여 다각도로 예측모델의 성능 평가를 수행하였으며 이는 예측 모델의 예측 성능 평가가 편향되지 않도록 돕는다.
- (3) 전체적으로 여러 모델을 학습하여 사용하는 앙상블 모델이 단일 모델을 사용하는 비앙상블 모델에 비해 높은 예측 성능을 보였으며 이 중 Random forest 방법이 가장 높은 성능을 보였다.

본 연구는 아스팔트 혼합물을 디자인하는 단계에서 활용될 수 있다. 전통적으로 아스팔트 혼합물의 디자인은 전문가의 역량에 의존하거나 시행착오 기반의 접근 방법이었으나 과학적이고 체계적인 방법으로 골재 간극률을 예측하는데 활용될 수 있을 것으로 기대된다.

## Acknowledgement

This work was supported by the Technology development Program(RS-2023-00256583) funded by the Ministry of SMEs and Startups(MSS, Korea)

## References

1. Anderson, R. M. and Bentsen, R. A. (2001), "Influence of

- voids in the mineral aggregate (VMA) on the mechanical properties of coarse and fine asphalt mixtures", *Journal of the Association of Asphalt Paving Technologists*, Vol.70.
2. Chadbourn, B. A., Skok Jr, E. L., Newcomb, D. E., Crow, B. L. and Spindle, S. (1999), *The effect of voids in mineral aggregate (VMA) on hot-mix asphalt pavements*, Minnesota Department of Transportation.
3. Chen, T. and Guestrin, C. (2016), "Xgboost: A scalable tree boosting system", *In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp.785-794.
4. Cortes, C. and Vapnik, V. (1995), "Support-vector networks", *Machine Learning*, Vol.20, pp.273-297.
5. Foreman, J. K. (2008), *Effect of voids in the mineral aggregate on laboratory rutting behavior of asphalt mixtures*, University of Arkansas.
6. Freund, Y. and Schapire, R. E. (1997), "A decision-theoretic generalization of on-line learning and an application to boosting", *Journal of Computer and System Sciences*, Vol.55, No.1, pp.119-139.
7. Friedman, J. H. (2001), "Greedy function approximation: a gradient boosting machine", *Annals of Statistics*, pp.1189-1232.
8. Goodfellow, I., Bengio, Y. and Courville, A. (2016), *Deep learning*. MIT press.
9. Ho, T. K. (1995), "Random decision forests", *In Proceedings of 3rd International Conference on Document Analysis and Recognition*, Vol.1, pp.278-282.
10. Inoue, T., Gunji, Y. and Akagi, H. (2004), "Rational design method of hot mix asphalt based on calculated VMA", *In Proceedings of the 3rd Eurasphalt and Eurobitume Congress held in Vienna*, Vol.2.
11. Ji, B. (2021), "Machine Learning-based Concrete Crack Detection Framework for Facility Maintenance", *Journal of the Korean Geo-Environmental Society*, Vol.22, No.10, pp.5-12.
12. Ji, B., Bhattarai, S. S., Na, I. H. and Kim, H. (2023), "A Bayesian deep learning approach for rheological properties prediction of asphalt binders considering uncertainty of output", *Construction and Building Materials*, Vol.408, pp.133671.
13. Kandhal, P. S. and Chakraborty, S. (1996), *Evaluation of voids in the mineral aggregate for HMA paving mixtures*, Report No.9, National Center for Asphalt Technology.
14. Kandhal, P. S., Foo, K. Y. and Mallick, R. B. (1998), "Critical review of voids in mineral aggregate requirements in Superpave", *Transportation Research Record*, No.1609, pp.21-27.
15. Lavin, P. (2003), *Asphalt pavements: a practical guide to design, production and maintenance for engineers and architects*, CRC Press.
16. Li, X. F., Doh, Y. S., and Kim, K. W. (2004), "Estimation of Rutting based on Volumetric Properties of Asphalt Mixture", *International Journal of Highway Engineering*, Vol.6, No.3, pp.79-90.



17. Malekloo, A., Ozer, E., AlHamaydeh, M. and Girolami, M. (2022), "Machine learning and structural health monitoring overview with emerging technology and high-dimensional data source highlights", *Structural Health Monitoring*, Vol.21, No.4, pp.1906-1955.
18. Opitz, D. and Maclin, R. (1999), "Popular ensemble methods: An empirical study", *Journal of Artificial Intelligence Research*, Vol.11, pp.169-198.
19. Sengoz, B. and Topal, A. (2007), "Minimum voids in mineral aggregate in hot-mix asphalt based on asphalt film thickness", *Building and Environment*, Vol.42, No.10, pp.3629-3635.
20. Shen, S. and Yu, H. (2011), "Analysis of aggregate gradation and packing for easy estimation of hot-mix-asphalt voids in mineral aggregate", *Journal of Materials in Civil Engineering, ASCE*, Vol.23, No.5, pp.664-672.
21. Wu, B., Pei, Z., Luo, C., Xia, J., Chen, C. and Wang, M. (2023), "Review and evaluation of the prediction methods for voids in the mineral aggregate in asphalt mixtures", *Journal of Materials in Civil Engineering, ASCE*, Vol.35, No.3, pp.04022455.
22. Zhang, H. (2009), *Asphalt mixture volume index and anti-rutting performance unified prediction model and its application*, Master's thesis, School of Civil Engineering, Shandong University