

# 머신러닝을 활용한 염료 분자표현자와 폴리프로필렌 섬유 염색성 간의 상관성 분석

배영랑 · 이준헌 · 김태경<sup>†</sup>

경북대학교 공과대학 섬유시스템공학과

## Correlation Analysis Between Molecular Descriptors of Dyes and Dyeability on Polypropylene Fiber Using Machine Learning

Yeongrang Bae, Junheon Lee, and Taekyeong Kim<sup>†</sup>

Department of Textile System Engineering, Kyungpook National University, Daegu 41566, Korea

**Abstract:** Polypropylene fibers are difficult to dye because of their extreme hydrophobicity and high crystallinity. Therefore, we have developed and reported superhydrophobic dyes which have high affinity towards polypropylene fibers. In the previous studies, dyeability (K/S) was analyzed using a single molecular descriptor (logP) which effectively represented the hydrophobicity of the dyes. However, as more and more dyes were synthesized and their dyeing results obtained, the dyeability could no longer be explained by the single molecular descriptor alone. In order to numerically analyze the relationship between the dye structures and dyeability on polypropylene fibers, a machine learning approach based on multiple molecular descriptors was applied. Linear regression and random forest models were used, and model interpretation was performed using weight analysis and SHAP for explainable artificial intelligence (XAI). Both models achieved high predictive performance, indicating that the relationships between molecular descriptors and dyeability were successfully captured. Model interpretation revealed that logP, degree of alkyl substitution, and hydrogen-bond donors were consistently identified as the key molecular descriptors affecting dyeability. These results demonstrate that dyeability is determined by the combined effects of multiple molecular characteristics, providing a quantitative basis for the rational design of dyes for polypropylene fibers.

<sup>†</sup>Corresponding Author: Taekyeong Kim  
E-mail: taekyeong@knu.ac.kr

Received January 16, 2026  
Revised January 29, 2026  
Accepted January 30, 2026

© 2026 The Korean Fiber Society

**Keywords:** superhydrophobic dyes, polypropylene, molecular descriptors, dyeability, machine learning, XAI

## 1. 서 론

폴리프로필렌(polypropylene, PP) 섬유는 경량성, 내화학성, 고강도 등의 다양한 장점에도 불구하고 극도의 소수성과 높은 결정성으로 인해 기존 염료군을 사용한 일반적인 염색공정으로는 실용적인 색상 구현이 불가능했다[1-3]. 본 연구팀에서는 긴 알킬 치환기를 색소 모체에 도입하여 염료 분자의 소수성을 극대화한 초소수성 염료를 합성했으며, 폴리프로필렌 섬유에 대해 높은 친화력과 우수한 견뢰도를 나타내는 염료들을 지속적으로 보고해 왔다[4-21]. 그 결

과, 알킬 치환기의 길이가 증가함에 따라 염색성이 향상되는 경향을 관찰하였으며, 이는 염료 분자의 소수성이 극대화될수록 폴리프로필렌 섬유와의 친화력이 향상되었기 때문으로 확인되었다.

이를 토대로 알킬 치환기 도입에 따른 염색성을 분석한 초기 선행연구에서는, 당시 합성된 모든 초소수성 염료를 대상으로 염료의 소수성(logP)에 따른 염색성을 해석한 바 있다[8]. 이후 색상 다양화를 목적으로 서로 다른 염료 모체를 갖는 새로운 초소수성 염료들이 지속적으로 합성되면서, 소수성과 같은 하나의 특성만으로는 염색성 변화를 충

분히 설명하지 못하는 한계점이 확인되었다. 이후, 후속 연구에서 동일한 염료 모체 내에서 알킬 치환기 길이 증가에 따른 분자량 변화를 소수성과 함께 고려하여 염색성을 해석하려는 시도가 이루어지기도 하였다[12]. 이 경우 동일한 색소 모체를 공유하는 염료군 내에서는, 분자량을 포함한 해석이 염색성 변화 경향을 효과적으로 설명할 수 있었다. 이와 같이 분자의 고유한 물리·화학적 특성을 정량적으로 표현한 지표를 분자표현자(molecular descriptor)라 하며, 이러한 논의는 염색성이 염료의 소수성 외에도 분자량과 같이 다양한 분자표현자와의 관련성을 시사한다. 따라서 염료 모체가 상이한 초소수성 염료들을 포괄적으로 고려할 때도 염색성을 일관되게 설명하기 위해서는, 하나의 분자표현자만으로 충분한 설명력을 확보하기 어렵기 때문에 여러 개의 분자표현자를 동시에 고려하는 다차원적 관점에서의 해석을 시도할 필요성이 있다.

머신러닝(machine learning)은 인공지능의 핵심 기술로, 다수의 입력 특성을 동시에 고려해 결과를 예측하며, 이 과정에서 특성과 결과 간의 복합적인 상관관계를 학습하는

기법이다. 이러한 접근은 최근 화학 및 재료과학 분야에서 단일 지표만으로는 설명하기 어려운 분자구조 기반 물성 예측과 상관관계 분석에 활용되고 있다[22-25]. 한편, 인공지능 모델은 높은 학습 수준과 예측 성능에도 불구하고 내부 작동 원리가 직관적으로 해석되기 어렵다는 한계를 지니며, 이를 보완하기 위해 설명 가능한 인공지능(explainable artificial intelligence, XAI) 기법이 제안되었다. XAI 기법은 모델 예측 결과에 기여한 입력 특성의 상대적 중요도를 정량화함으로써, 인공지능 모델이 단순한 예측 도구를 넘어, 예측 결과의 근거를 해석할 수 있는 분석 도구로 활용될 수 있음을 보여준다[25,26].

이에 본 연구에서는 염료의 다양한 분자표현자를 동시에 고려할 수 있는 다차원적 분석 접근으로서 머신러닝 기법을 도입하고, 이를 통해 초소수성 염료의 분자 구조와 폴리프로필렌 섬유에 대한 염색성 간의 상관관계를 체계적으로 분석하고자 한다. 더 나아가 구축된 인공지능 모델에 대한 XAI 분석을 통해 염색성에 기여하는 주요 분자표현자를 확인하고자 한다.

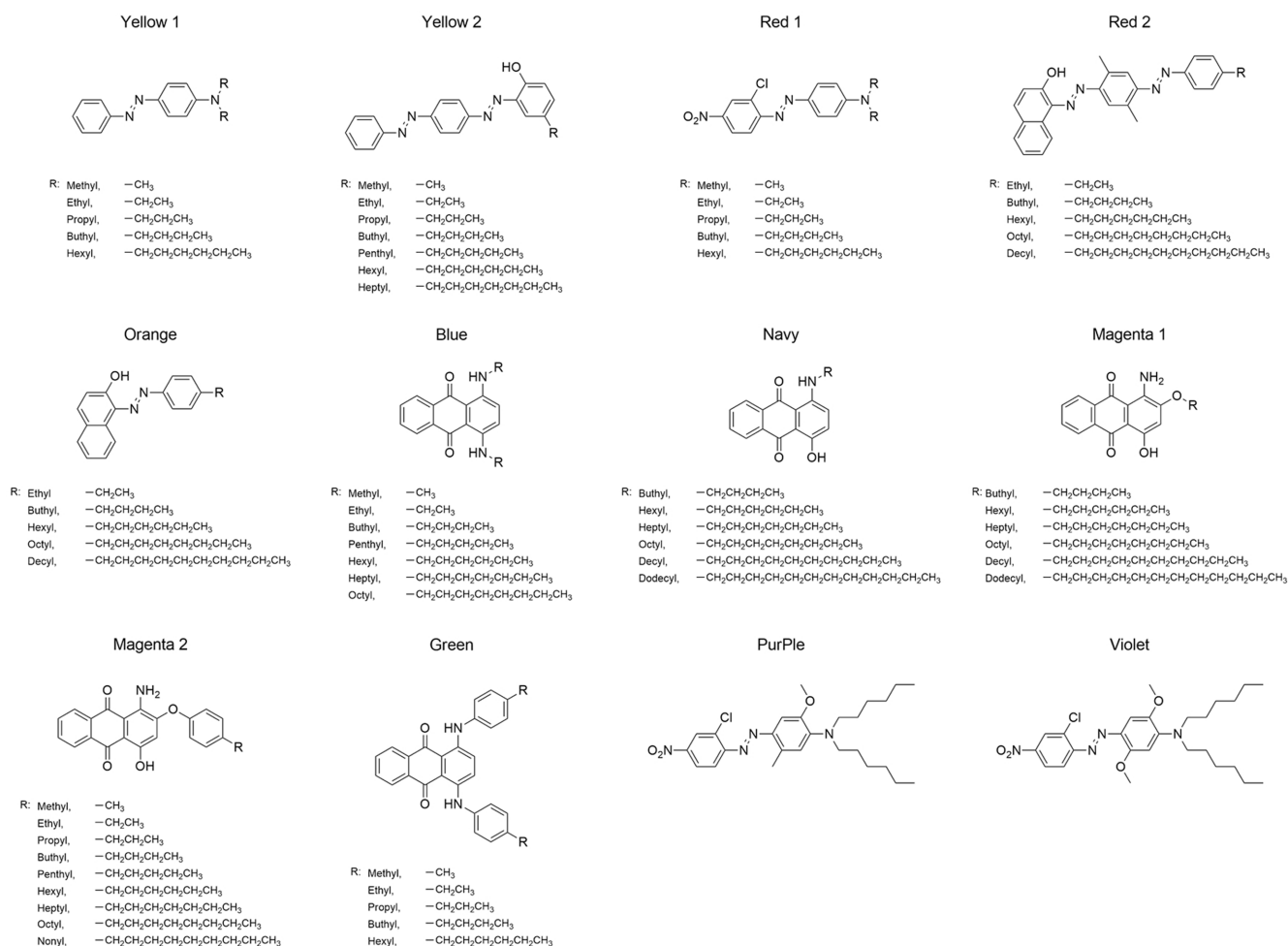


Figure 1. Chemical structures of alkyl-substituted superhydrophobic dyes [4-21].

## 2. 실험

### 2.1. 염료 및 염색

본 논문에서 사용된 섬유소재와 염료는 앞서 언급된 선행연구를 통해 이미 보고된 것으로, 개질하지 않은 순수 폴리프로필렌 섬유 편직물에 대해 두 개의 염료군으로 염색되었다[4-21]. 하나의 염료군은 선행연구에서 새롭게 합성된 초소수성 염료군으로, 극도의 소수성을 부여하기 위해 긴 알킬 치환기가 도입된 구조로 설계된 것이며(Figure 1), 다른 염료군은 비교를 위한 이미 상용화된 17종의 분산염료와 10종의 솔벤트염료를 포함한다(Figure 2).

폴리프로필렌 섬유 1.0 g을 대상으로, 순수 염료 기준 5% owf, 욕비 1:50 조건에서 130 °C에서 1시간 동안 염색하였으며, 염색을 완료한 후 일반적인 조건으로 환원 세정하였다[4-21]. 염색된 직물의 염색성을 나타내는 색상 강도

를 평가하기 위해 측색기를 이용하여 각 염료의 최대 흡수 파장에서의 표면반사율을 측정하고, Kubelka-Munk 식(1)을 적용하여 염색물의 겉보기 색농도를 나타내는 K/S 값을 산출하였다. 이용된 측색기는 Konica-Minolta CM-3600d spectrophotometer이며 표준광원 D65 조건에서 10° 관찰자 시야각을 조건으로 측정하였다[4-21].

$$K/S = \frac{(1 - R)^2}{2R} \tag{1}$$

K: absorption coefficient  
 S: scattering coefficient  
 R: reflectance

### 2.2. 염료의 분자표현자 계산

본 연구에서는 염료의 분자구조를 정량적으로 표현하기

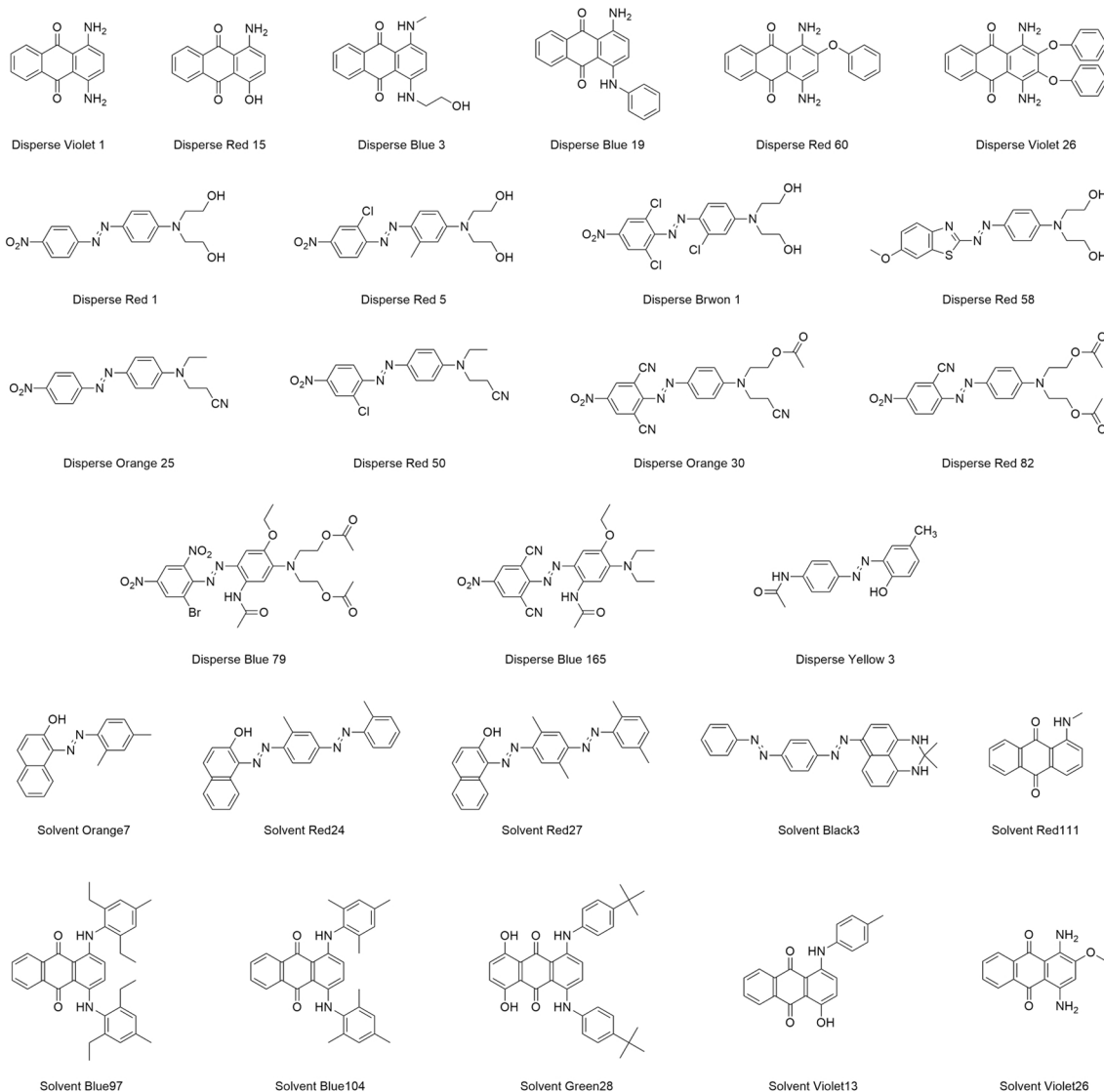


Figure 2. Chemical structures of conventional disperse and solvent dyes.

**Table 1.** Definition of molecular descriptors of the dyes used as input features for machine learning

Descriptor	Definition
C	Number of carbon atoms in the dye molecule
H	Number of hydrogen atoms in the dye molecule
N	Number of nitrogen atoms in the dye molecule
O	Number of oxygen atoms in the dye molecule
MW	Molecular weight of the dye molecule
SE_Vol	Solvent excluded volume(Å <sup>3</sup> ), representing the volume inaccessible to solvent molecules
Ovality	Shape factor describing the deviation of the molecular shape from a perfect sphere
Rot_Bond	Number of rotatable bonds, representing molecular flexibility
Access_Area	Accessible surface area(Å <sup>2</sup> ) of the molecule that is exposed to the surrounding environment and available for intermolecular interactions.
Polar_Area	Polar surface area(Å <sup>2</sup> ) associated with polar functional groups
H_bond_Acceptors	Number of hydrogen bond acceptor sites in the molecule
H_bond_Donors	Number of hydrogen bond donor sites in the molecule
Phenyl	Number of phenyl groups in the dye structure
Alkyl/1chain	Number of alkyl substituents per single alkyl chain
Total_alkyl	Total number of alkyl substituents in the dye molecule
logP	Octanol–water partition coefficient representing molecular hydrophobicity

위해 ChemDraw Ultra 10.0와 Chem3D Ultra 10.0을 사용하여 다양한 분자표현자를 계산하였다. 프로그램상에 작성된 염료 구조는 MM2 에너지 최소화(MM2 Minimize Energy) 과정을 통해 구조 최적화를 수행하였으며, 이때 minimum RMS gradient는 0.100으로 설정하였다. 구조 최적화가 완료된 후, 염료 분자의 조성, 입체적 크기, 구조적 유연성 및 소수성 특성을 정량화하기 위해 16종의 분자표현자(molecular descriptor)를 선정하고 Chem3D의 Compute Properties 메뉴를 통해 각 분자표현자 값을 계산하였다. 각 분자표현자

의 정의는 Table 1에 정리하였으며, 이들은 염색성 예측 모델의 입력 특성으로 사용되었다.

### 2.3. 염료의 분자표현자 및 K/S 데이터 구축

본 연구에서는 총 92종의 서로 다른 분자구조를 갖는 염료에 대해 정량적으로 계산된 16종의 분자표현자와 실험적으로 측정된 색상 강도(K/S) 값을 포함한 데이터셋을 구축하였으며, 이를 Figure 3에 정리하였다. 본 데이터셋은 염료 구조 변화에 따른 염색성의 경향을 효율적으로 분석하기

**Figure 3.** Overview of the dataset including molecular descriptors of the dyes and color strength (K/S) on polypropylene fiber.

위해 다양한 색소 모체와 알킬 치환기 길이를 갖는 염료들로 구성하였다. 또한, 서로 다른 물리적 의미와 단위를 갖는 분자표현자를 입력 특성으로 사용하는 과정에서, 각 특성 간 수치 범위와 분포의 차이가 인공지능 모델 학습에 영향을 미칠 수 있음을 고려하여, 모든 분자표현자 값이 유사한 스케일을 갖도록 일부 정규화를 수행하였다.

#### 2.4. 머신러닝 기반 염색성 예측 모델 구축

본 연구는 염료의 여러 분자표현자와 염색성 간의 다차원적 상관관계를 반영하기 위해 머신러닝 기법을 도입하였다. 최근 염료와 섬유 간의 상호작용으로 염색성, 염료 흡착과 같은 복합적인 현상을 분자 구조 기반 머신러닝 모델로 예측하고 해석한 연구들이 보고되며, 이러한 접근의 유효성이 제시된 바 있다[27,28]. 본 연구에서는 분자표현자 입력을 기반으로 염색성(K/S)을 정량적으로 예측하는 회귀 문제로 설정하였으며, 다양한 관점에서 분자표현자의 기여양상을 비교분석하기 위해 서로 다른 학습 구조를 갖는 선형 회귀 모델과 랜덤 포레스트 모델을 적용하였다. 선형 회귀 모델은 입력된 분자 특성 지표와 이에 대응하는 가중치(weight)의 가중합과 편향(bias)을 통해 염색성을 출력값으로 예측하는 구조를 가지며, 식 (2) 같이 표현될 수 있다.

$$y = \sum_{i=1}^n w_i x_i + b \quad (2)$$

여기서  $x_i$ 는 각 분자표현자를,  $w_i$ 는 해당 특성이 염색성 예측에 기여하는 가중치를  $b$ 는 편향을 의미한다. 이러한 구조에서 모델의 예측 결과와 실제 결과 간의 오차를 최소화하도록 가중치와 편향 값을 반복적으로 갱신하며 모델의 학습이 수행되었다.

랜덤 포레스트 모델은 서로 다른 분자표현자와 염료 데이터 조합을 기반으로 다수의 결정나무(decision tree)를 생성하고, 각 결정나무의 예측 결과를 평균하여 최종 예측 결과를 도출하였다. 각 결정나무는 분자표현자 값을 기준으로 데이터를 단계적으로 분기하는 트리 알고리즘을 이용해 염색성(K/S)을 예측하였다.

모델 학습은 전체 데이터셋을 8:2의 비율로 훈련용 데이터와 테스트용 데이터로 분할하여 수행하였다. 훈련 데이터는 모델의 학습에 사용되었으며, 테스트 데이터는 학습이 완료된 모델의 예측 성능을 평가하는 데 활용하였다. 이때 인공지능 회귀 모델의 예측 성능을 정량적으로 평가하기 위해, 평균절대오차(MAE), 평균제곱근오차(RMSE), 그리고 결정계수( $R^2$ )를 평가 지표로 사용하였으며, 각 평가 지표는 다음과 같이 계산하였다.

$$MAE = \frac{1}{m} \sum_{i=1}^m |y_i - \hat{y}_i| \quad (3)$$

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2} \quad (4)$$

$$R^2 = 1 - \frac{\sum_{i=1}^m (y_i - \hat{y}_i)^2}{\sum_{i=1}^m (y_i - \bar{y})^2} \quad (5)$$

여기서  $y_i$ 는 실제값을,  $\hat{y}_i$ 는 예측값을 의미하며  $\bar{y}$ 는 실제값의 평균,  $m$ 은 평가에 사용된 전체 데이터의 개수를 나타낸다.

#### 2.5. 모델 해석(Explainable AI) 방법

본 연구에서는 인공지능 모델이 염색성 예측 과정에서 활용한 분자표현자를 해석하기 위해, 모델의 구조적 특성과 학습 방식에 따라 적절한 해석 방법을 적용하였다. 먼저 선형 회귀 계열 모델의 경우, 각 입력 특성에 대응하는 가중치를 기반으로 분자표현자와 염색성 간의 관계를 해석하였으며, 가중치의 크기와 부호를 통해 각 분자표현자가 염색성 예측에 미치는 상대적 영향과 방향성을 확인할 수 있다[29]. 반면 랜덤 포레스트 모델은 분기 구조의 트리 알고리즘을 통해 예측을 수행하므로, 선형 회귀 모델에서와 같은 명시적인 가중치가 존재하지 않는다. 이에 따라 개별 분자표현자가 염색성 예측에 미치는 영향을 직접 해석하는 것은 어렵다. 따라서 본 연구에서는 동일한 기준에서 두 모델을 비교분석하기 위해 SHAP(Shapley Additive exPlanations) 기법을 활용하였다. SHAP 기법은 각 분자표현자가 포함되었을 때와 포함되지 않았을 때의 예측값 변화가 반영된 Shapley value를 계산함으로써, 개별 예측 결과 수준에서 분자표현자의 기여도를 정량화하는 분석을 가능하게 한다[26,30]. 이와 같은 모델 해석 방법을 통해 염색성 예측에 기여하는 핵심 분자표현자를 도출하고, 모델이 학습한 분자구조와 염색성 간의 다차원적인 상관관계를 해석하였다.

### 3. 결과 및 고찰

#### 3.1. 염료의 분자표현자와 염색성(K/S) 간의 개별 상관성 분석

염료의 분자 구조적 특성과 염색성 간의 관계를 개별적으로 분석하기 위해, 16가지 분자표현자와 염색성(K/S) 간의 개별 상관성 분석을 수행하고자, 개별 산점도를 Figure 4에 나타내었다. 이는 다양한 분자표현자 간의 다차원적 관계를 분석하기 이전에, 개별 분자표현자가 염색성과 관계성을 가지는지 탐색해보기 위한 분석이다. 그 결과, 모든 산점도에서 전반적으로 각 분자표현자와 K/S 간의 관계

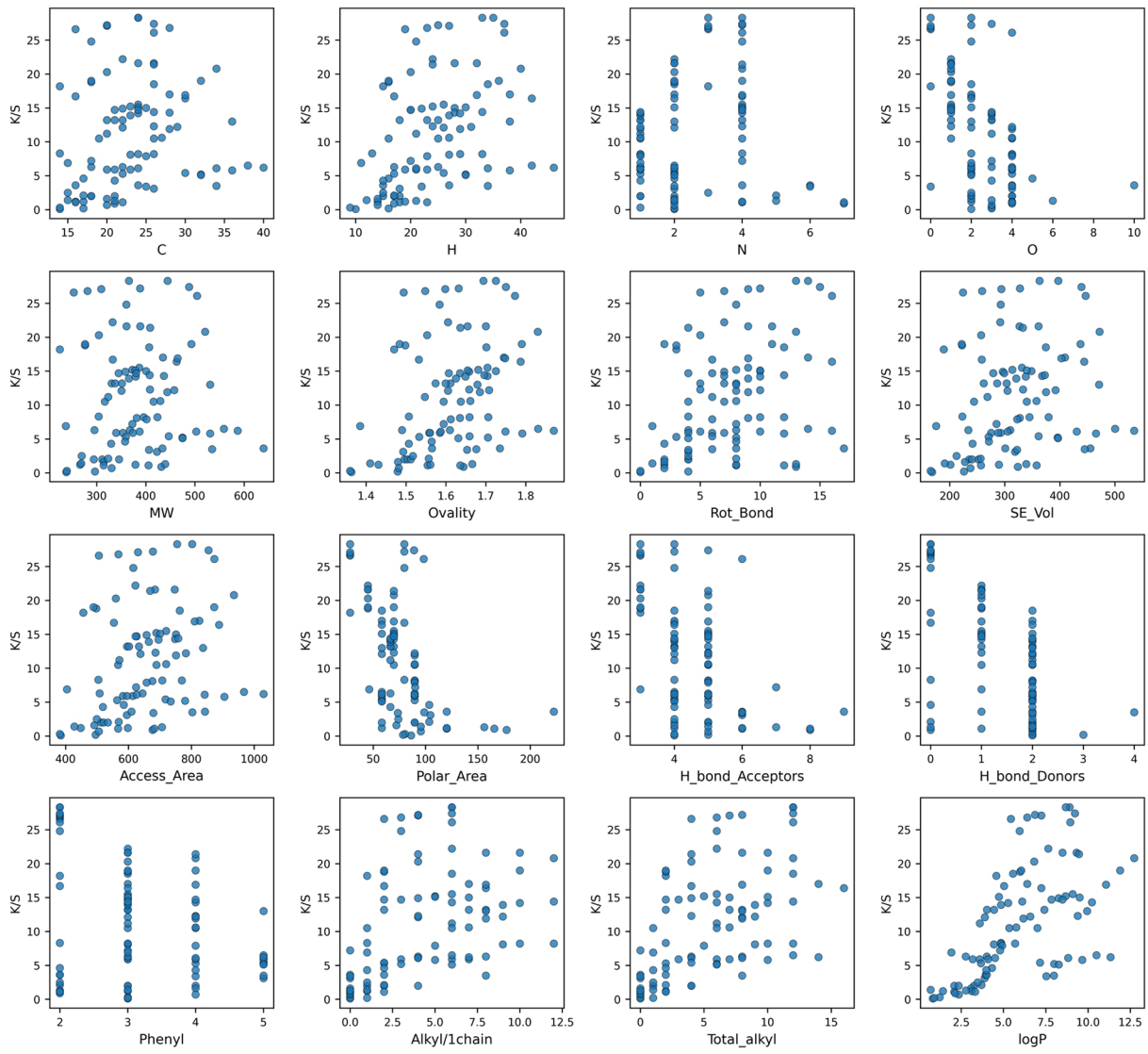


Figure 4. Scatter plots of individual molecular descriptors versus K/S.

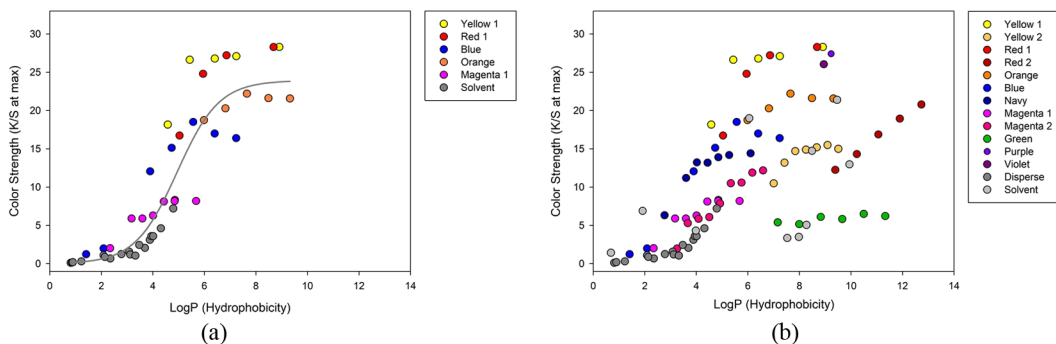


Figure 5. LogP vs. K/S scatter plots showing (a) six dye groups with a good relationship and (b) all dye groups.

가 넓게 분산된 형태로 나타나 명확한 상관관계를 가진다고 설명하기 어려운 것으로 확인되었다. 일부 분자표현자 산점도에서 상대적으로 조밀하게 분포하는 부분이 관찰되어 경향성이 있을 수도 있으나, 제한적인 범위에서만 국소

적으로 나타날 뿐, 특정한 단일 분자표현자가 염색성과 명확한 상관관계를 가진다고 해석하기에는 한계가 있다.

한편, 선행연구에서 염료의 소수성을 나타내는 logP와 염색성 간의 상관관계를 기반으로 염색성을 해석하고 그 결

과를 보고한 바 있으며, 동일한 색소 모체를 갖거나 제한된 염료군을 대상으로 분석을 수행하였었다[8,12]. 이러한 경우, 제시된 Figure 5(a)와 같이 logP 증가에 따라 K/S가 선형적으로 증가하는 비교적 명확한 경향이 관찰되었다. 또한, 비교를 위해 함께 제시된 기존 상용 분산염료의 경우, 일반적으로 소수성 염료로 분류되에도 불구하고 전반적으로 낮은 K/S를 나타내었다. 이는 분산염료의 소수성이 폴리프로필렌 섬유에 대한 충분한 친화성을 확보하지 못하였음을 의미한다. 이에 반해 선행연구에서 합성된 초소수성 염료들은 logP 값의 증가와 함께 염색성이 현저히 향상되었으며, 기존 분산염료 대비 뚜렷한 염색성 개선 효과를 보였다. 이는 염료 분자의 소수성이 극단적으로 강화될 때 폴리프로필렌 섬유와의 친화성이 효과적으로 향상된다는 점을 명확히 보여준다.

이러한 접근을 바탕으로, 본 연구에서는 색상 다양화를 목적으로 합성된 서로 다른 색소 모체의 모든 초소수성 염료들을 대상으로 logP와 K/S 간의 관계를 분석하였으며, 염료 종류별로 구분하여 Figure 5(b)에 나타내었다. 그 결과, 동일한 색소 모체 내에서는 logP 증가에 따라 K/S가 증가하는 경향이 동일하게 관찰되었으나, 서로 다른 색소 모체를 포함한 전체 염료군을 포괄적으로 고려할 경우 logP와 K/S 관계가 뚜렷한 선형성을 보이지 않고 넓게 분산된 양상을 나타냈다. 이는 logP와 같은 단일 분자표현자만으로

는 다양한 구조를 갖는 초소수성 염료의 염색성 차이를 일관된 관계로 설명하기 어렵다는 점을 의미한다.

### 3.2. 분자표현자 기반 염색성 예측 모델 구축 및 학습양상

본 연구에서 분자표현자와 염색성(K/S) 간의 정량적 관계를 학습하기 위해 선형 회귀 기반 모델과 랜덤 포레스트 모델을 구축하였다. 선형 회귀 모델은 분자표현자와 염색성(K/S) 간의 관계를 학습하고, 각 분자표현자의 기여도를 직접적으로 해석할 수 있도록 단일층 구조로 설계하였다. 입력층(input)에 16종의 분자표현자를 입력한 뒤, 활성화 함수로 linear를 사용하는 하나의 출력층(dense)을 통해 K/S 값을 예측하도록 구성하였다. 이와 같이 입력과 출력을 직접 연결 구조에서는 중간 은닉층을 거치지 않으므로, 학습된 가중치가 각 분자표현자의 기여도와 방향성을 직접적으로 반영한다는 장점이 있다. 모델 학습은 학습률을 0.05로 설정한 Adam optimizer를 사용하여 안정적으로 최적화하였고, 손실함수로는 평균제곱오차(MSE)를 적용하여 예측 오차를 최소화하였다. 선형 회귀 모델 학습 과정의 안정성과 양상을 확인하기 위해 학습에 따른 가중치(weight) 및 편향(bias)의 분포 변화를 모니터링하였다(Figure 6(a), 6(b)). 학습이 진행됨에 따라 각 파라미터들이 최적화되었으며, 이는 모델이 입력된 분자표현자 정보를 반영하여 파라미터를 점진적으로 조정해나가는 학습 과정으로 해석할 수 있다.

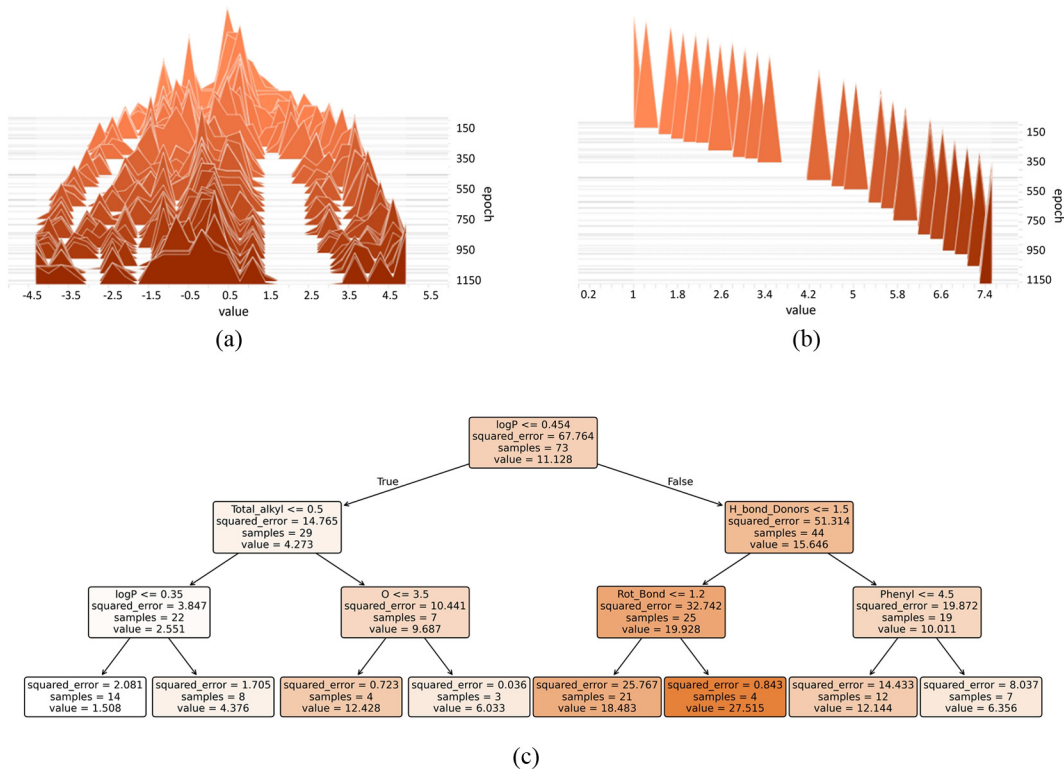


Figure 6. Visualization of training behavior of the weight (a) and bias (b) parameters in the linear regression model, (c) example decision tree from the trained random forest.

랜덤 포레스트 모델은 scikit-learn의 RandomForestRegressor를 이용하여 구축하였으며, n\_estimators는 100으로 설정하였다. 랜덤 포레스트는 다수의 결정트리를 기반으로 한 앙상블 모델로, 각 결정트리는 입력된 분자표현자를 기준으로 데이터를 반복적으로 분할하면서 예측값을 산출하는 구조를 가진다. 이 과정에서 선형 회귀 모델과 달리 각 분자표현자에 곱해지는 명시적인 가중치(weight)는 존재하지 않으며, 대신 분자표현자의 특정 임계값을 기준으로 데이터를 분기하는 과정이 반복적으로 최적화된다. Figure 6(c)는 랜덤 포레스트를 구성하는 결정트리 중 하나의 예시를 나타낸 것으로, logP와 같은 소수성 관련 분자표현자를 기준으로 1차 분기가 이루어진 후, 알킬 치환기 수, 수소 결합 특성 등 서로 다른 분자표현자들이 순차적으로 분기 기준으로 사용되어 데이터가 세분화되는 과정을 확인할 수 있다. 이러한 분기 구조를 통해 랜덤 포레스트 모델은 여러 분자표현자들이 조합하여 염색성(K/S)에 영향을 미치는 비선형적 관계와 상호작용 효과를 자연스럽게 학습할 수 있다.

**3.3. 분자표현자-염색성 관계를 학습한 예측 모델의 성능 평가**

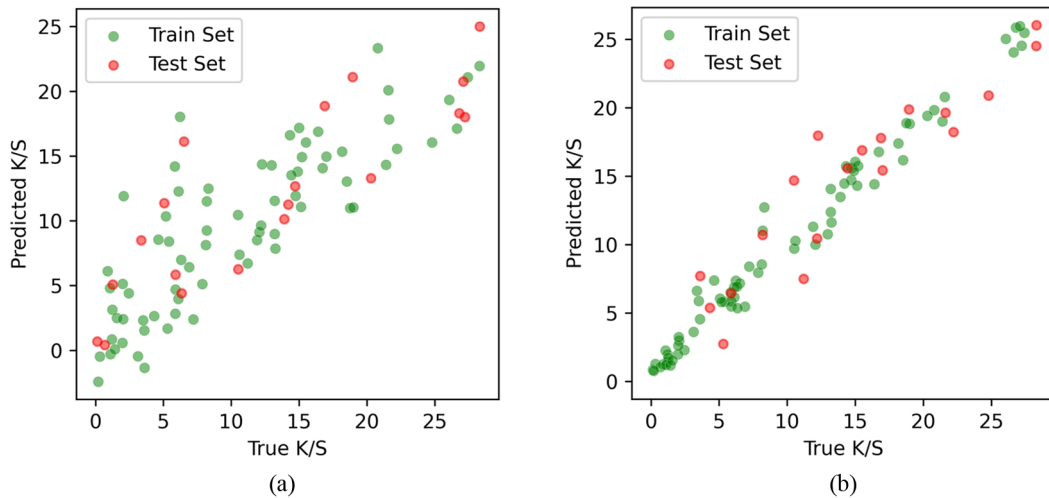
본 연구에서 구축한 모델이 여러 분자 특성 지표와 염색성 간의 관계를 잘 학습하였는지 확인하기 위해, 분자표현자를 기반으로 염색성을 얼마나 정확하게 예측하는지 평가하였다. 선형 회귀 모델과 랜덤 포레스트 모델을 이용하여 예측한 K/S 값과 실험적으로 측정된 K/S 값의 산점도로 나타내어 Figure 7에 제시하였다. 그래프에서 초록색 점은 인공지능 모델이 머신러닝 과정에서 실제로 학습에 사용된 훈련 데이터의 결과이고, 빨간색 점은 학습에 사용되지 않은 테스트 데이터에 대한 예측 결과이다.

먼저 훈련 데이터에 대한 예측 결과를 보면, 두 모델 모두 예측값이 실측값 인근에 분포하며 전반적인 선형적 대응 관계를 형성하고 있어, 분자표현자와 염색성 간의 관계

를 적절히 학습하였음을 확인할 수 있다. 이는 모델이 입력된 분자표현자 정보를 기반으로 염색성 변화를 합리적으로 설명할 수 있는 예측 구조를 형성했음을 의미한다. 여기서 중요한 점은 모델 학습에 사용되지 않은 테스트 데이터에 대한 예측 결과이다. 테스트 데이터는 모델이 처음 접하는 새로운 염료임에도 불구하고, 두 모델 모두에서 학습에 사용되지 않은 데이터들의 예측값이 이미 학습에 사용된 훈련 데이터의 분포와 거의 정확히 일치하는 경향을 유지하였다. 이러한 결과는 모델이 단순히 학습 데이터를 기억한 것이 아니라, 분자표현자와 염색성 간에 내재된 관계를 학습하고 이를 새로운 염료 구조에 대해서도 적용할 수 있음을 시사한다.

모델별 예측 결과를 비교하면, Figure 7(a)의 선형 회귀 모델은 실측값의 전반적인 경향을 재현하였으나 예측값의 분산이 비교적 크게 분포하며 예측 오차가 상대적으로 크게 나타났다. 반면, Figure 7(b)의 랜덤 포레스트 모델은 낮은 K/S 영역부터 높은 K/S 영역까지 전 범위에서 예측값이 실측값 인근에 조밀하게 분포하며, 훈련 데이터와 테스트 데이터 모두에서 일관된 예측 성능을 나타냈다. 이러한 차이는 선형 회귀 모델이 모든 분자표현자를 동시에 고려하여 평균적인 경향을 반영하는 구조인 반면, 랜덤 포레스트 모델은 분자표현자를 기준으로 데이터를 단계적으로 분기하며 보다 복합적인 관계를 반영하는 구조적 차이에 따른 것으로 해석할 수 있으며, 랜덤 포레스트 모델이 염색성 예측에 더 적합한 모델로 작용한 것으로 판단된다.

종합적으로, 이러한 결과는 본 연구에서 구축한 예측 모델들이 분자표현자와 염색성(K/S) 간의 관계를 효과적으로 학습하였음을 시각적으로 보여주며, 특히 랜덤 포레스트 모델에서 상대적으로 더 정밀한 예측 성능이 나타났다. 이와 같은 예측 성능을 정량적으로 비교하고자 평균절대오차(MAE), 평균제곱근오차(RMSE), 그리고 결정계수(R<sup>2</sup>)를 산



**Figure 7.** Comparison of predicted and experimental true K/S values using (a) linear regression and (b) random forest models.

**Table 2.** Prediction performance of linear regression and random forest models evaluated using MAE, RMSE, and R<sup>2</sup>

Model	Linear regression		Random forest	
	Train	Test	Train	Test
MAE	3.40	4.17	0.99	2.53
RMSE	4.28	5.07	1.32	2.90
R <sup>2</sup>	0.68	0.71	0.97	0.85

출하였으며, Table 2에 정리하였다.

학습 데이터에 대한 예측 성능 평가 결과, 선형 회귀 모델은 낮은 수준의 MAE와 RMSE가 3.40과 4.28로 나타나 일정 수준의 예측 오차를 보였으며, 실제값의 변화를 얼마나 잘 설명하는지 나타내는 결정계수 R<sup>2</sup>는 0.68로 나타나 염색성 변화를 부분적으로 설명하였다. 반면, 랜덤 포레스트 모델은 MAE와 RMSE가 각각 0.99와 1.32로 선형 회귀 모델에 비해 현저히 낮은 오차를 보였고, R<sup>2</sup> 또한 0.97로 매우 높은 예측 정확도를 보였다. 테스트 데이터에 대한 예측 성능 평가 결과, 선형 회귀 모델의 MAE와 RMSE는 각각 4.17과 5.07로 나타났으며, R<sup>2</sup>는 0.71로 학습 데이터 대비 유사한 수준의 설명력을 유지하였다. 랜덤 포레스트 모델은 테스트 데이터에서도 MAE 2.53, RMSE 2.90으로 선형 회귀 모델보다 낮은 예측 오차를 보였고, R<sup>2</sup> 역시 0.85로 더 높은 값을 나타내어 새로운 염료 데이터에 대해서도 우수한 예측 성능을 유지하였다. 학습에 사용되지 않은 새로운 염료 데이터에 대한 안정적인 예측 성능은 분자표현자와 염색성 간의 일반화된 상관관계를 학습하였음을 시사하며, 본 연구에서 제안한 분자표현자 기반 회귀 모델이 염색성(K/S)을 정량적으로 예측하는 데 있어 충분한 신뢰성과 활용 가능성을 지님을 정량적으로 뒷받침한다.

### 3.4. 염색성에 영향을 미치는 분자표현자의 개별 중요도 분석

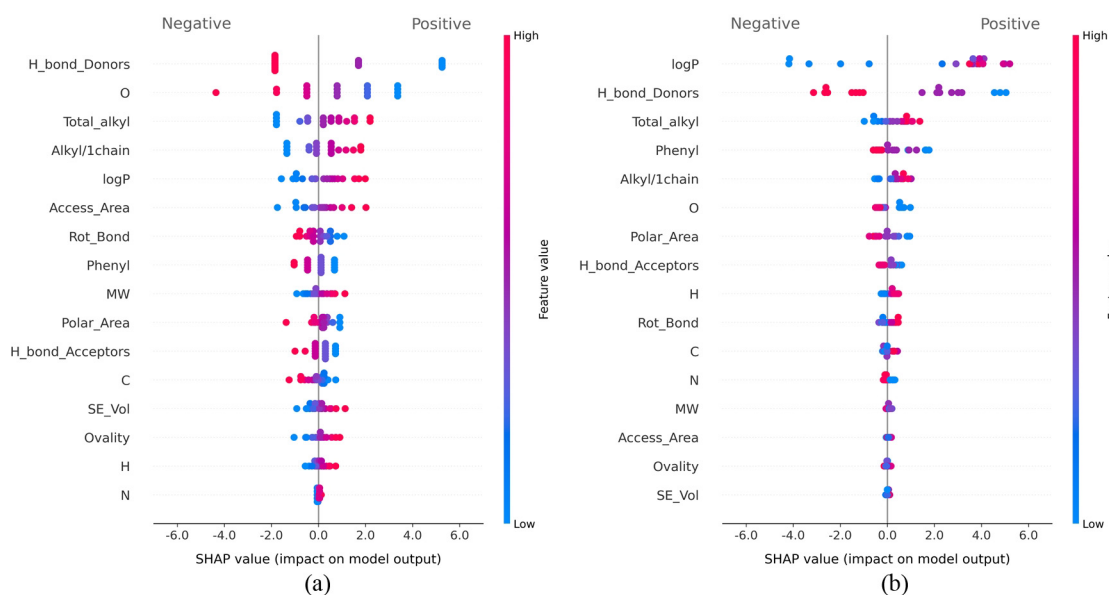
본 연구에서는 분자표현자와 염색성 간의 관계를 학습한 예측 모델을 해석함으로써, 염색성 예측 과정에서 각 분자표현자가 어떠한 역할을 수행했는지 분석하고자 한다. 앞서 식 (2)에 나타난 바와 같이, 선형 회귀 모델은 각 분자표현자  $x_i$ 는 염료 분자의 구조적 특징을 나타내는 지표로, 본 연구에서는 총 16개의 분자표현자가 사용되었다. 이에 대응하는 가중치  $w_i$ 는 해당 분자표현자가 염색성 예측에 미치는 영향의 크기와 방향을 나타내는 계수로, 학습 과정에서 예측 오차를 최소화하여 정확한 예측을 하도록 최적화된다. 가중치의 절댓값이 클수록 해당 분자표현자가 염색성 예측에 미치는 영향이 크다는 것을 의미하며, 양의 부호를 갖는 경우 염색성을 증가시키는 방향으로, 음의 부호를 갖는 경우 염색성을 감소시키는 방향으로 작용함을 나타낸다. 학습이 완료된 이후 최종적으로 최적화된 각 분자표현자에 대한 가중치 값은 Table 3에 정리하였다.

**Table 3.** Weights and bias of the linear regression model

Descriptor	Weight	
Ovality	4.158064	Positive ↑
Total_alkyl	3.315855	
logP	3.213423	
Alkyl/1chain	3.140659	
Access_Area	0.646156	
MW	0.638094	
SE_Vol	0.619332	
H	0.403834	
N	0.034308	
H_bond_Acceptors	-0.430209	
Phenyl	-0.572007	
C	-0.824683	
O	-1.286996	
Rot_Bond	-1.445437	
Polar_Area	-1.779099	
H_bond_Donors	-3.550233	Negative

이 중 Ovality, 총 알킬 치환기를 나타내는 Total\_alkyl, logP, 알킬 치환기 당 사슬길이를 나타내는 Alkyl/1chain은 상대적으로 큰 양의 가중치를 나타내며, 염색성 예측에 주요하게 기여한 특성으로 확인되었다. 이러한 양의 가중치는 해당 분자표현자의 값이 클수록 모델이 염색성을 우수하게 예측하도록 학습하였음을 보여준다. 반면, H\_bond\_Donor, 분자구조 내 극성기의 표면적을 의미하는 Polar\_Area, 회전이 가능한 단일결합의 수를 의미하는 Rot\_bond는 상대적으로 큰 음의 가중치를 보이며, 해당 분자표현자의 값이 클수록 염색성을 낮게 예측하는 경향을 보여준다. 그 외 접근 가능한 표면적을 나타내는 Access\_Area, 분자량, 용매 배제 부피를 나타내는 SE\_Vol, 방향족 고리의 갯수를 의미하는 Phenyl, H\_bond\_Acceptors, C, H, O, N과 같은 분자표현자들은 가중치의 절댓값이 상대적으로 작아, 염색성 예측에 있어 영향력이 작은 보조적인 역할을 하는 특성으로 나타났다. 전반적으로 양의 가중치를 보인 특성들은 알킬 치환기와 이에 따른 소수성, 구조적 변화에 관련된 분자표현자인 반면, 음의 가중치를 보인 특성들은 소수성 섬유와의 상호작용을 저해하는 극성 표현자 또는 구조적 특성과 연관된 분자표현자로 나타났다. 이와 같은 가중치 산출을 통해 선형 회귀 모델에서는 분자 특성별로 염색성에 미치는 영향의 방향성을 확인할 수 있다.

그러나, 이러한 해석은 가중치 개념이 존재하지 않는 랜덤 포레스트 모델에서 시행할 수 없으므로, 본 연구에서는 SHAP 분석을 활용하였다. SHAP 분석은 각 입력 특성이 포함되었을 때와 포함되지 않았을 때의 예측값 변화를 기반으로 영향력을 평가하는 방법으로, 모델의 내부 구조와 무관하게 입력 특성의 기여도를 분석할 수 있어, 서로 다른



**Figure 8.** SHAP-based interpretation of molecular descriptor contributions to K/S prediction based on (a) linear regression and (b) random forest.

모델 간 해석 결과를 비교하는 데 유용하다. Figure 8은 분자표현자와 염색성(K/S) 간의 관계를 학습한 선형 회귀 모델과 랜덤 포레스트 모델의 SHAP 분석 결과를 나타낸다.

Figure 8에서 각 점은 단일 염료의 개별 분자표현자를 의미하며, 각 분자표현자가 해당 염료의 염색성 예측 결과에 기여한 정도를 나타낸다. 점의 색상은 분자표현자 값의 상대적인 크기를 의미하며, 붉은색은 높은 값, 푸른색은 낮은 값을 나타내며, x축의 SHAP value는 예측에 대한 기여 방향과 크기를 의미한다. 양의 값은 염색성을 크게 예측하도록, 음의 값은 염색성을 작게 예측하도록 작용했음을 나타내며, SHAP value의 절댓값이 클수록 예측에 미치는 영향이 크다. 그래프 상단에 위치한 분자표현자일수록 SHAP value의 절댓값이 크고 분포 범위가 넓게 나타나 전반적인 중요도가 높은 것으로 해석된다.

Figure 8a에 제시된 선형 회귀 모델의 SHAP 분석 결과, H\_bond\_Donor의 분포 폭이 가장 넓게 나타나 염색성 예측에 가장 핵심적인 역할을 하는 분자표현자로 확인되었으며, 값이 증가함에 따라 염색성을 낮게 예측하는 경향성을 확인할 수 있다. 이어서 중간 수준의 분포를 보이는 분자 특성지표들 중 총 알킬 치환기 수를 나타내는 Total\_alkyl, 알킬 치환기 당 사슬 길이를 나타내는 Alkyl/1chain, logP와 접근 가능 표면적을 나타내는 Access\_Area는 특성값이 높을수록 염색성을 우수하게 예측하는 경향을 보였으며, 반면 산소 원자 수(O), 회전 가능한 결합 수를 나타내는 Rot\_bond, 방향족 고리 수를 나타내는 Phenyl은 특성값에 따라 염색성을 낮게 예측하는 경향을 나타냈다. 그 외, 분자량, 극성 표면적을 의미하는 Polar\_Area, H\_bond\_Acceptors, C, H, N, 용매 배제 부피를 나타내는 SE\_Vol과 Ovality는

SHAP 값 분포가 상대적으로 좁게 나타나 중요도가 약한 것으로 분석된다.

한편, Figure 8b에 제시된 랜덤 포레스트 모델의 SHAP 분석 결과, logP, H\_bond\_Donor는 상대적으로 넓은 분포로 나타나 염색성 예측에 중요한 영향을 미친 분자표현자로 확인되었다. 여기서 logP는 값이 클수록 염색성이 우수하게 예측되는 경향을 보인 반면, H\_bond\_Donor는 값이 클수록 염색성을 낮게 예측하였다. 이어서 중간 수준의 분포를 보이는 분자표현자들 중 총 알킬 치환기 수를 나타내는 Total\_alkyl, 알킬 치환기 당 사슬 길이를 나타내는 Alkyl/1chain은 값이 클수록 염색성을 우수하게 예측하는 경향을 보였으며, 방향족 고리 수를 나타내는 Phenyl, 산소 원자 수(O), 극성 표면적을 나타내는 Polar\_Area는 특성값에 따라 염색성을 낮게 예측하였다. 그 외, H\_bond\_Acceptors, Rot\_bond, C, H, N, 분자량, Access\_Area, Ovality 및 SE\_Vol은 SHAP 값 분포가 상대적으로 좁게 나타나 중요도가 약한 것으로 분석된다.

비록 모델별 해석 결과에는 차이가 존재하지만, 두 모델에서 공통적으로 높은 기여도와 동일한 방향성을 보이는 분자표현자들이 도출되었으며, 이는 염색성(K/S) 예측에 공통적으로 작용하는 주요 분자 특성이 있음을 보여준다. 가장 먼저 logP, 총 알킬 치환기 수를 나타내는 Total\_alkyl, 알킬 치환기 당 사슬 길이를 나타내는 Alkyl/1chain은 선형 회귀 모델과 랜덤 포레스트 모델 모두에서 SHAP value 분포 폭이 넓게 나타나며 큰 영향력을 갖는 핵심 특성으로 도출되었고, 값이 클수록 염색성을 우수하게 예측하는 경향이 나타났다. 이들은 모두 알킬 치환기 도입과 연관된 분자특성지표로, 긴 알킬 치환기 도입으로 인한 소수성 증대

가 폴리프로필렌 섬유와의 친화력을 개선한 것으로 해석된다. H\_bond\_Donor 역시 두 모델 모두에서 높은 영향력을 보였으며, H\_bond\_Donor의 증가가 염색성을 저해하는 방향으로 학습하였다. 이는 H\_bond\_Donor가 염료 분자의 극성을 증가시켜 소수성 환경과의 친화적 상호작용을 약화시키기 때문으로 해석되며, 모델의 구조와 관계없이 염색성 발현에 있어 염료 분자의 소수성과 극성 간의 균형이 중요함을 시사한다. H\_bond\_Donor와 함께 극성 특성을 반영하는 산소 원자 수(O)와 Polar\_Area 또한 염색성을 저해하는 방향으로 유의미한 영향력을 보였으며, 염색성 예측에 있어 보조적인 기여를 한 것으로 해석된다.

또한 방향족 고리 수를 나타내는 phenyl은 두 모델 모두에서 유의미한 기여도를 보였으며, 값이 증가할수록 염색성을 낮게 예측하는 경향을 나타냈다. 방향족 고리는 염료 모체를 구성하는 요소로 발색 특성과 밀접한 관련이 있으며, 구조상 분자에 강직성과 평면성을 부여하여 섬유 내부로의 침투 과정에서 불리하게 작용한 것으로 해석된다. 이외에도 접근 가능 표면적(access area), 회전 가능한 결합 수(rotatable bonds)는 선형 회귀 모델에서 유의미한 중요도를 보였으나, 예측 성능이 보다 우수한 랜덤 포레스트 모델에서는 미약한 중요도를 보여, 염색성 예측에 있어 보조적인 역할을 수행한 것으로 해석된다.

본 연구에서 염색성(K/S)은 단일 분자표현자에 의해 설명되기보다 소수성, 알킬 치환 정도, 극성 작용기, 구조적 특성 등 다양한 분자 특성들이 복합적으로 작용한 결과로 해석될 수 있다. 이러한 해석은 단순한 수치적 상관관계를 나열하는 데 그치지 않고, 각 분자표현자가 갖는 물리·화학적 의미를 바탕으로 염색성에 미치는 영향을 이해하고자 하였다. 이와 같이 다차원적 관계를 반영한 분석은 기존 단일 분자표현자 중심의 염색성 해석이 갖는 한계를 보완하며, 분자의 특성을 종합적으로 고려한 염색성 예측 모델의 필요성을 뒷받침한다.

#### 4. 결 론

본 연구에서는 머신러닝 기법을 적용하여 분자표현자 기반 선형 회귀 모델과 랜덤 포레스트 모델을 구축해 폴리프로필렌 섬유에 대한 염색성(K/S)을 예측하였다. 또한 모델 해석을 위해 SHAP 분석을 하고 결과를 종합함으로써 단일 분자표현자 중심의 기존 염색성 해석을 넘어, 다양한 분자 표현자와 염색성 간의 다차원적인 상관관계를 규명하였다.

본 연구에서 구축한 분자표현자 기반 예측 모델들은 학습 데이터뿐만 아니라 학습에 사용되지 않은 새로운 염료 데이터에 대해서도 안정적인 예측 성능을 보였다. 이는 모델이 단순히 개별 데이터의 패턴을 기억한 것이 아니라, 분

자표현자와 염색성(K/S) 간에 존재하는 내재된 상관관계를 효과적으로 학습하였음을 의미하며, 염색성 예측을 위한 정량적 모델로서 충분한 신뢰성과 활용 가능성을 지남을 보여준다.

또한, 본 연구에서는 가중치 분석과 함께 SHAP 분석 같은 모델 해석 기법을 활용하여 분자표현자 기반 예측 모델의 예측 근거를 정량적으로 분석하였으며, 폴리프로필렌 섬유에 대한 염색성이 단일 분자특성에 의해 결정되기보다는 서로 다른 물리·화학적 성격을 갖는 다차원적인 분자특성들의 균형적 작용에 의해 형성됨을 확인하였다. 특히 소수성을 반영하는 logP와 알킬 치환 관련 분자표현자들은 염료-섬유 간 친화성을 결정하는 핵심 요인으로 작용한 반면, H\_bond\_Donors와 같은 극성 관련 특성은 소수성 기반의 염색 과정을 방해하는 방향으로 영향을 미쳤다. 이는 폴리프로필렌 섬유 염색에서 소수성 증대가 중요한 요인임과 동시에, 염료 분자에 내재된 극성 특성과 같은 구조적 요소들이 염색성에 복합적으로 작용함을 의미한다. 또한 방향족 고리 수와 같이 구조적 형상과 관련된 분자표현자들도 핵심 특성을 보완하는 요인으로 작용하여, 염색성이 특정 단일 분자 특성에 의해 결정되기보다, 다차원적 상호작용이 이뤄진 결과임을 보여준다. 이는 트리 기반 분기 구조를 통해 복합적인 특성 관계를 반영할 수 있는 랜덤 포레스트 모델에서 효과적으로 학습되었으며, 본 연구에서 랜덤 포레스트 모델이 선형 회귀 모델에 비해 우수한 예측 성능을 보인 결과와도 일관된 결과를 나타낸다.

이러한 다차원적 분자표현자 기반의 모델 해석은 기존의 단일 분자표현자 중심 염색성 해석이 갖는 한계를 보완하고, 분자의 구조적 특징이 염색성에 미치는 영향을 정량적으로 이해할 수 있는 근거를 제공한다. 나아가 예측 성과 해석 결과를 연계함으로써, 폴리프로필렌 섬유에 적합한 염료 분자 설계 방향을 제시하고, 고기능성 염료 개발을 위한 합리적인 분자 설계 지침으로 활용될 수 있을 것으로 기대된다.

**감사의 글:** 본 논문은 2025년 산업통상자원부의 재원으로 한국산업기술진흥원의 지원을 받아 수행된 산업혁신인재성장지원사업의 결과물입니다. (과제명 : 섬유패션산업 DX 전문인력양성, 과제번호 : RS-2024-00410875).

#### References

1. I. S. Cho, D. S. Jang, B. Y. Seo, S. J. Park, K. M. Cho, and Y. B. Shim, "Practical Dyeing", Hyungseul Publish Co., Seoul, 2000, pp.397-399.
2. M. Ahmed, "Polypropylene Fibers - Science and Technology", Elsevier Scientific Publishing Co., New York, 1982, pp.462-501.

3. E. R. Trotman, "Dyeing and Chemical Technology of Textile Fibres", Wiley-Interscience, New York, 1984, pp.497-498.
4. T. K. Kim, S. H. Yoon, J. P. Hong, H. J. Kim, and J. S. Bae, "Coloration of Pure Polypropylene Fiber with Super Hydrophobic Dyes", *Textile Coloration and Finishing*, 2006, **18**, 30-34.
5. T. K. Kim, J. S. Jung, S. H. Yoon, M. K. Kim, and Y. A. Son, "The Coloration Properties of Alkylsubstituted Anthraquinoid Dyes for Pure Polypropylene Fiber", *Textile Coloration and Finishing*, 2007, **19**, 28-34.
6. T. K. Kim, J. S. Jung, S. I. Son, S. H. Yoon, M. Kim, and J. S. Bae, "Synthesis and Application of Alkyl-substituted Disazo Yellow Dyes for Unmodified Polypropylene Fiber", *Fiber. Polym.*, 2008, **9**, 538-543.
7. T. K. Kim, J. S. Jung, K. J. Jang, S. H. Yoon, and M. K. Kim, "Synthesis and Application of Alkyl-substituted High Chroma Yellow Dyes for Unmodified Polypropylene Fiber", *Fiber. Polym.*, 2009, **10**, 148-153.
8. T. K. Kim, K. J. Jang, and S. H. Jeon, "Calculation and Analysis of Hydrophobicity of the Dyes Synthesized for Unmodified Polypropylene Fibers Using Molecular Descriptors", *Textile Coloration and Finishing*, 2009, **21**, 21-26.
9. T. K. Kim, K. J. Jang, and S. H. Jeon, "Synthesis of Red Disperse Dyes with Various Diazo Components and Coloration of Unmodified Pure Polypropylene Fibers", *Textile Coloration and Finishing*, 2010, **22**, 1-7.
10. T. K. Kim, K. J. Jang, and S. H. Jeon, "Synthesis and Application of Alkyl-substituted Red Dyes for Unmodified Polypropylene Fiber", *Fiber. Polym.*, 2011, **12**, 174-179.
11. D. S. Kwak and T. K. Kim, "Synthesis of Diamino-anthraquinoid Blue Disperse Dyes Having Alkyl Substituents Longer Than Pentyl Group and Their Dyeability Toward Pure Polypropylene Fibers", *Textile Coloration and Finishing*, 2012, **24**, 106-112.
12. D. S. Kwak and T. K. Kim, "Dyeing of Ultra High Molecular Weight Polyethylene Fibers with Diamino-anthraquinoid Blue Disperse Dyes Having Linear Alkyl Substituents", *Textile Coloration and Finishing*, 2012, **24**, 180-188.
13. T. K. Kim and S. H. Jeon, "Coloration of Ultra High Molecular Weight Polyethylene Fibers Using Alkyl-substituted Monoazo Yellow and Red Dyes", *Fiber. Polym.*, 2013, **14**, 105-109.
14. T. K. Kim and Y. R. Chae, "Synthesis and Application of Anthraquinoid Magenta Dyes for Pure Polypropylene Fibers", *Textile Coloration and Finishing*, 2013, **25**, 102-109.
15. T. K. Kim and C. H. Lee, "Synthesis of a Super Hydrophobic Violet Dye for Pure Polyolefin (PP/UHMWPE) Fibers", *Textile Coloration and Finishing*, 2013, **25**, 165-171.
16. T. K. Kim and Y. R. Chae, "Synthesis and Application of Novel High Light Fastness Red Dyes for Ultra High Molecular Weight Polyethylene Fibers", *Fiber. Polym.*, 2014, **15**, 248-253.
17. T. K. Kim and M. H. Ryu, "Synthesis of Super Hydrophobic Orange Dyes Having Maximum Absorption at 450-500 nm for Pure Polyolefin Fibers", *Textile Coloration and Finishing*, 2014, **26**, 165-172.
18. N. Jo, J. H. Lee, and T. K. Kim, "Synthesis of Bluish-green Dyes for Pure Polyolefin Fibers", *Textile Coloration and Finishing*, 2016, **28**, 156-163.
19. J. H. Lee, H. J. Ma, and T. K. Kim, "Synthesis of Hydrophobic Anthraquinoid Magenta Dyes Having Linear Alkyl Substituents", *Fiber. Polym.*, 2017, **18**, 1691-1696.
20. T. K. Kim and H. J. Ma, "Dyeing of Ultra High Molecular Weight Polyethylene Fiber Using Anthraquinoid Superhydrophobic Navy Dyes", *Textile Coloration and Finishing*, 2019, **31**, 98-106.
21. H. J. Ma and T. K. Kim, "Synthesis of Super Hydrophobic Disperse Dyes Having Long Alkyl Substituents and Their Dyeability onto Unmodified Polypropylene Fibers", *Fiber. Polym.*, 2020, **21**, 1767-1772.
22. K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, and A. Walsh, "Machine Learning for Molecular and Materials Science", *Nature*, 2018, **559**, 547-555.
23. J. Schmidt, M. R. G. Marques, S. Botti, and M. A. L. Marques, "Recent Advances and Applications of Machine Learning in Solid-state Materials Science", *npj Comput. Mater.*, 2019, **5**, Article No. 83.
24. A. Jain, "Machine Learning in Materials Research: Developments over the Last Decade and Challenges for the Future", *Curr. Opin. Solid State Mater. Sci.*, 2024, **28**, Article Number: 101189.
25. Y. Chen, S. Zhang, Y. Wen, Z. Lai, and T. Liu, "Prediction of Material Stability of Two-dimensional Semiconductors: An Interpretable Machine Learning Perspective", *APL Mater.*, 2024, **12**, 091120.
26. S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions", *Adv. Neural Inf. Process. Syst.*, 2017.
27. J. Qin, S. Habib, X. Shao, S. M. Billah, Y. Wang, M. Cai, H. Li, Y. Wu, J. Liang, B. Cao, and S. Mahmud, "AI-Assisted Dyeing Optimization of Polyimide Fibers via Backpropagation Neural Networks", *Prog. Org. Coat.*, 2025, **209**, 109611.
28. S. Liu, D. Chen, F. Zhang, Q. Zhao, J. He, and X. Dong, "Machine Learning Assisted Prediction of Disperse Dye Exhaustion on Polylactic Acid Fiber with Interpretable Model", *Dyes and Pigments*, 2025, **237**, 11269.
29. TensorFlow Documentation, [https://www.tensorflow.org/api\\_docs/python/tf/keras/layers/Dense](https://www.tensorflow.org/api_docs/python/tf/keras/layers/Dense) (Accessed January 2, 2026).
30. L. S. Shapley, "A Value for n-Person Games", in "Contributions to the Theory of Games", H. W. Kuhn and A. W. Tucker, Eds., Princeton University Press, Princeton, NJ, 1953, Vol. 2, pp.307-317.