

2D-QSAR방법을 이용한 농약류의 무지개 송어 급성 어독성 분석 및 예측

송인식 · 차지영 · 이성광★

한남대학교 화학과

(2011. 11. 29. 접수, 2011. 12. 8. 수정, 2011. 11. 30. 승인)

Prediction and analysis of acute fish toxicity of pesticides to the rainbow trout using 2D-QSAR

In Sik Song, Ji Young Cha and Sung Kwang Lee★

Department of Chemistry, Hannam University, Daejeon, 305-811, Korea

(Received November 29, 2011; Revised December 8, 2011; Accepted November 30, 2011)

요 약: 본 연구는 농약류에 대하여 구조-활성의 정량적 관계(QSAR)를 이용하여 무지개 송어(학명: *Oncorhynchus mykiss*)의 급성 독성을 예측-분석하는 과정을 수행하였다. 모델 구현을 위해 사용된 275종의 농약류에 대한 수중 독성(96h LC₅₀) 값은 DEMETRA 프로젝트의 데이터를 사용하였다. 예측 모델에 사용된 2차원 분자 표현자는 PreADMET 프로그램으로부터 계산을 하였고, 선형 (다중 선형 회귀 방법) 모델과 비선형(서포트 벡터 머신, 인공 신경망) 학습 방법들은 실험값과 예측값의 적합도를 고려하여 최적화 되었다. 데이터 전처리 과정을 거친 뒤에, 5묶음 교차 검증과정을 포함한 모집단 기반 전진 선택법을 통해서 각 학습 방법의 최적의 표현자 집합을 결정하였다. 가장 좋은 결과는 SVM 방법 ($R^2_{CV}=0.677$, RMSECV=0.887, MSECV=0.674) 이었고, EU의 규제 기준에 따른 분류에서는 87%의 정확도를 나타내었다. MLR방법을 통해서 무지개 송어의 급성 독성에 대하여 독성을 나타내는 농약류의 구조적 특징과 지질 층과의 상호작용을 설명할 수 있었다. 개발된 모든 모델들은 5묶음 교차 검증과 Y-scrambling test 을 통해 검증되었다.

Abstract: The acute toxicity in the rainbow trout (*Oncorhynchus mykiss*) was analyzed and predicted using quantitative structure-activity relationships (QSAR). The aquatic toxicity, 96h LC₅₀ (median lethal concentration) of 275 organic pesticides, was obtained from EU-funded project DEMETRA. Prediction models were derived from 558 2D molecular descriptors, calculated in PreADMET. The linear (multiple linear regression) and non-linear (support vector machine and artificial neural network) learning methods were optimized by taking into account the statistical parameters between the experimental and predicted pLC50. After preprocessing, population based forward selection were used to select the best subsets of descriptors in the learning methods including 5-fold cross-validation procedure. The support vector machine model was used as the best model ($R^2_{CV}=0.677$, RMSECV=0.887, MSECV=0.674) and also correctly classified 87% for the training set according to EU

★ Corresponding author

Phone : +82-(0)42-629-8874 Fax : +82-(0)42-629-8811

E-mail : leesk@hnu.kr

regulation criteria. The MLR model could describe the structural characteristics of toxic chemicals and interaction with lipid membrane of fish. All the developed models were validated by 5 fold cross-validation and Y-scrambling test.

Key words: pesticide, QSAR, acute fish toxicity, rainbow trout, cross-validation

1. 서 론

EU의 신화학 물질 관리제도인 REACh (Registration, Evaluation, Authorization and Restriction of Chemicals)가 2007년부터 시행되면서, EU내에서 제조 및 수입되는 화학 물질, 혼합물, 완제품 포함 물질에 대해 등록, 평가, 허가 및 제한하도록 하였다.¹ 이를 필두로 CHINA-REACh (중국), Smart-REACh (대만), 화심법(일본), TSCA (미국)등의 국제적인 화학물질 관리제도들도 물질 규정에 대한 제정 또는 개정의 작업을 진행하고 있다. 각 제도의 등록 시에 필요한 항목은 국가 또는 연간 생산량에 따라 차이가 있으나, 주로 물리 화학적 특성(녹는점, 끓는점, 용해도 등), 환경 수명 특성(가수분해, 생분해 등), 환경 독성(어류, 박테리아 등의 급성, 만성 독성 등), 인체 유해성(유전 독성, 발암성 등)이 있다.

전세계적으로 그 생산량이 많은 농약류들도 마찬가지로 이러한 국제 화학 물질 관리제도에 따라 엄격한 등록 자료 제출을 하여야 한다. 농약류는 일반적으로 잡초, 해충에 대한 작용은 표적 유기체의 특정 대사 과정을 저해하는 메커니즘을 주로 기반으로 한다. 일반적으로 살포되는 농약류의 대부분은 물, 공기, 침전물, 생명체, 음식 등에 잔존하므로 생태계에서 독성을 유발할 가능성이 높다.² 그러므로, 위해성 평가에서 농약류들의 환경 독성 실험은 필수적으로 진행되고 있으며, 특히 수생 생태계에서는 주로 어류의 급성 또는 만성 독성을 통해서 평가한다. OECD에서는 어류의 독성 시험에 대한 지침서(guideline)³을 제정하여 표준화된 독성 시험이 진행되도록 제안하고 있다.

농약류들의 환경 독성 시험은 많은 인원과 시간을 필요로 하며, 이에 대한 비용은 국제 화학물질 관리제도에 따르면 수출 기업들이 부담해야 한다. 이러한 부담을 줄이기 위한 대체 방법으로 EU등 국제 기구들은 QSAR (Quantitative Structure Activity Relationships) 방법을 통해 예측되는 결과를 활용하는 방안을 법적으로 허용하고 있다.⁴ QSAR방법이란 화학 구조의 특징을 수치로 나타낸 표현자를 이용하여 활성과의 통계

적인 모델을 통해 나타내는 방법으로, 기존 화합물에 알려진 여러 화학적 특성 및 정보를 종합하여 신규 또는 실험값이 없는 화합물의 물성 및 약효, 독성 등을 예측하게 함으로써, 그 적용 범위가 의약, 환경, 식품, 신소재 개발 분야 등으로 점차 넓어지고 있다.⁵

본 연구에서는 농약류들에 대하여 어류의 급성 독성을 예측할 수 있도록 QSAR분석방법을 수행하고 모델을 검증하고자 한다. 특히 다른 어종에 비해 성장이 빠르고 번식력이 왕성하여 국제적으로 어류 독성에 대표적으로 어종으로 사용되는 무지개 송어(rainbow trout, 학명: *Onchorhynchus mykiss*)에 대하여 적용하도록 하고, QSAR방법의 검증을 위하여 교차 검증법과 Y-scrambling방법을 통해 예측 정도를 평가하도록 하였다. 마지막으로 EU등에서 규제를 목적으로 구분해 놓은 어류 급성 독성의 분류 기준에 대하여 예측모델의 결과를 통해서 분류 정확도를 확인하였다.

2. 연구 방법

Fig. 1은 본 연구에서 수행한 QSAR 과정을 도표로 나타낸 것이다. 데이터 수집, 표현자 계산, 데이터전처리, 표현자선택, 모델 성능 평가 및 검증 순으로 진행하였다.

2.1. 사용 데이터

본 연구에서는 EU가 지원한 DEMETRA project⁶⁻⁸에서 수집한 농약류 282종에 대하여 96시간동안 무지개 송어에 대하여 측정된 LC₅₀값 (mM)을 사용하였다. 이 값은 무지개 송어 개체 중 절반 가량이 치사하는 반수 치사 농도(lethal concentration 50%)를 나타내며, 이 때의 농도는 mM로 나타내었다. 이 값을 모델 구현을 위하여 종속변수로 사용할 때는 -log항을 추가하여 pLC₅₀로 전환하여 사용하였다. DEMETRA project에서는 재현성이 있는 정확한 실험 데이터를 선별하기 위하여 다음과 같은 세가지 기준을 만족하는 데이터만을 사용하였다. 1) OECD나 EPA의 지침서와 같은 표준화된 과정에 따라 얻은 데이터, 2) 우량 실험

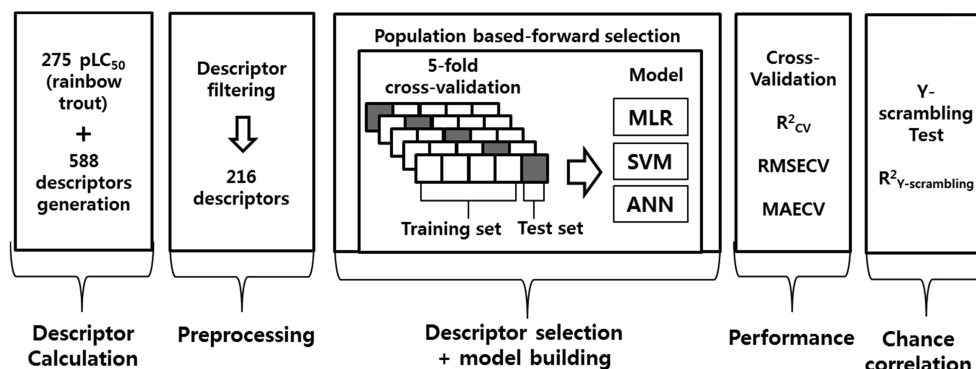


Fig. 1. Schematic overview of QSAR analysis used to developed acute fish toxicity model.

규정(Good Laboratory Practice, GLP)에 따라 얻은 데이터, 3) 실험 수행 연도, 순도, 실험 불확도를 포함한 통계 자료를 부수적인 데이터로 이용이 가능한 데이터. 282종의 농약류 중에서 분자 표현자가 계산되지 않는 7종의 무기 화합물을 제외한 뒤, 최종적으로 275종의 농약류 데이터를 모델 개발에 사용하였다.

2.2. 분자 표현자(Molecular Descriptor)

분자 표현자는 화학 분자의 구조적 특성을 정량적인 수치로 표현한 것을 말한다. 뒤에서 언급할 학습 방법에 분자 표현자를 적용하여 분자의 구조적 특성과 급성 독성과의 관계를 분석해 내고, 이를 통해 독성 분자의 구조적 공통점과 독성 메커니즘을 설명할 수 있다. 본 연구에서는 PreADMET v2.0 프로그램⁹을 이용하여 2차원 화학 구조로부터 분자 표현자들을 계산하였고, 계산된 표현자들은 143종의 구성 요소 표현자(constitutional descriptors), 79종의 정전기적 표현자(electrostatic descriptors), 24종의 기하학적 표현자(geometrical descriptors), 3종의 물리 화학적 표현자(physicochemical descriptors), 95종을 위상학적 표현자(topological descriptors), 그리고 화학적 상호 작용점(chemical feature)들간의 결합간 거리 정보를 포함하는 231종의 CATS (Chemical Advanced Template Search)¹⁰ 표현자가 사용되었다. 최종적으로 총 558종의 2차원 분자 표현자가 모델 구현에 앞서서 데이터 전처리 과정에 적용되었다.

2.3. 데이터 전처리(Preprocessing)

계산된 558종의 분자 표현자로부터 우선적으로 불필요한 표현자들을 선별하여 제거하는 전처리 과정을 수행하였다. 첫 번째 과정은 모든 화합물에 대하여 계

산한 화합물의 표준 편차가 0.01이하인 표현자는 제거하였다. 이 과정은 모든 화합물의 구조적 차이에 대하여 구조적 변별력이 없는 표현자를 제거하는 과정이며, 이 방법으로 통해 96종의 표현자를 제거하였다. 두 번째 과정은 급성 독성 값(pLC₅₀)과의 상관 계수(R²)값이 0.01이하인 것은 제거하였다. 이 과정은 급성 독성과 매우 상관성이 결여되는 표현자를 제거하는 과정으로, 이 과정에서 총 176종의 표현자를 제거하게 되었다. 세 번째 과정은 표현자간의 상관 계수(R²)값이 0.9이상인 표현자들은 두 표현자들이 서로 중복으로 표현되므로, 이 중 다른 표현자와의 상관 계수들이 평균적으로 큰 표현자를 제거하였고, 이 과정으로 총 70여종의 표현자들을 제거하였다. 최종적으로 세가지 과정을 통해 기본적인 학습 모델에 적용할 표현자수를 216종으로 줄이게 되었다.

2.4. 표현자 선택 과정(descriptors selection)

데이터 전 처리 과정을 통해 선별된 표현자로 QSAR 모델을 구현할 때, 학습 방법에 적합한 표현자를 선택하는 과정이 또한 필요하다. 본 연구에서는 모집단 기반 전진 선택(population based Forward selection)방법¹¹을 통해서 각 학습 방법에 적합한 표현자를 선택하도록 하였다. 기본적으로 전진 선택 방법을 간단히 설명하면, 모델의 적합도를 가장 증가시키는 표현자를 처음 추가하고, 이 다음부터는 이전 모델의 적합도를 가장 향상시키는 표현자를 다음으로 추가하면서 가장 최적이 되는 기준값에 도달할 때 까지 추가하는 방법이다. 이 방법을 표현자의 수만큼의 개별 수식을 포함하는 모집단(population)을 통해서 최종적으로 가장 적합한 표현자 집합을 선택하게 하므로, 다른 표현자 선택 방법(유전자 알고리즘, simulated annealing 등등)

보다 빠르게 최적 표현자 집합을 선택할 수 있는 장점이 있다. 본 연구에서는 전진 선택 방법을 교차 검증 법에서 계산된 적합도를 기준으로 최적의 표현자 집합을 선택하였다.

2.5. 학습 방법

본 연구에서 QSAR모형을 구현하기 위한 학습방법으로 선형 모델인 다중 선형 회귀(Multiple Linear Regression: MLR)방법과 비선형 모델인 서포트 벡터 머신(Support Vector Machine: SVM)과 인공 신경망(Artificial Neural Network: ANN)방법을 적용하였다. 다중 선형 회귀는 종속변수(Y)와 하나 이상의 독립변수(X) 관계를 선형 회귀식으로 나타낸 것이며, 형태는 아래와 같다.

$$Y = b + a_1X_1 + a_2X_2 + \dots + a_nX_n$$

여기서 Y는 종속변수, X_n 는 독립변수(즉 분자 표현자)를 나타내고 a_n 은 독립변수의 회귀 계수로 표현되며 마지막으로 b는 절편을 나타낸다.

서포트 벡터 머신¹²방법을 이용한 회귀 모델은 data를 주어진 공간보다 교차원의 공간으로 변환하고, 교차원 공간에서 선형 회귀식을 얻는 방법이다. SVM방법의 교차원 공간 변환은 Gaussian radial basis function를 이용하였고, SVM방법내의 추가적인 매개변수를 제거할 수 있는 γ -SVM방법¹³을 적용하였다.

인공 신경망¹⁴방법은 뇌의 신경망 전파 방식을 모사한 방법으로, 다층 전진 신경망(multilayer feed forward neural network)형태의 역-전파(back-propagation) 알고리즘을 적용하였다. 본 방법은 입력층(input layer)의 뉴런은 표현자, 출력층(out layer)의 뉴런은 pLC₅₀으로 지정하고, 은닉층(hidden layer)의 뉴런 수를 최적화하였고, 최적화 과정에서 필요한 learning rate, momentum은 각각 grid search방법으로 정해진 범위 내에서 최적조건을 결정하였다.

2.6. 모델 검증 과정

QSAR모형을 구현할 때는 훈련 데이터에 대한 적합도 계산과 더불어 모델 검증을 위한 다양한 과정들이 중요하다. 본 연구에서는 두 가지의 모델 검증 과정을 수행하였다. 첫 번째 방법은 교차 검증(cross-validation) 방법으로 훈련 데이터를 정해진 수의 묶음(fold)으로 구분한 뒤, 묶음 중 한 가지를 제외한 다른 묶음들로 QSAR모형을 구현하고, 제외된 묶음을 모델에 적용하여 예측한 결과를 얻도록 한다. 이러한 과정은 한 묶

음씩 번갈아 제외하여 예측한 결과를 얻도록 하고, 제외시킨 묶음을 모두 모으면 전체 훈련 데이터의 수와 같으므로 제외된 묶음들의 예측값에 대한 적합도를 측정하여 평가하는 방법이다. 교차 검증을 통해 얻은 적합도(R^2_{cv} , RMSECV, MAECV)는 훈련 데이터만을 이용하여 얻은 적합도와(R^2 , RMSE, MAE) 구분되며, 그 계산식은 다음과 같다.

$$R^2_{cv} = 1 - \frac{\sum_{i=1}^N (y_i - y_{pre,i})^2}{\sum_{i=1}^N (y_i - \bar{y})^2}$$

$$RMSECV = \frac{\sqrt{\sum_{i=1}^N (y_i - y_{pre,i})^2}}{N}$$

$$MAECV = \frac{\sum_{i=1}^N |y_i - y_{pre,i}|}{N}$$

N은 화합물의 수를 나타내며, R^2_{cv} 는 교차 검증에 의한 결정 계수, RMSECV(root mean square error of cross-validation)와 MAECV(mean absolute error of cross-validation)는 각각 교차 검증을 통해 계산된 평균 오차와 평균 절대 오차를 나타낸다.

두 번째 방법은 Y-scrambling test로, 많은 표현자로부터 우연히 상관성을 나타냈을 가능성에 대하여 확인하는 방법이다. 종속변수인 pLC₅₀값을 무작위로 섞은 데이터를 모델 구현에 사용한 방법들을 똑같이 수행하여 얻은 적합도를 얻는데, 이러한 방법들을 여러 번 반복하여 얻은 적합도와 최적 모델의 적합도를 차이를 통해 모델 자체가 우연 상관(chance correlation)에 의한 결과가 아님을 증명하도록 하였다.

3. 결과 및 토론

3.1. 학습 방법에 적합한 표현자 선택

데이터 전처리 방법을 이용하여 216종의 분자 표현자로 축약한 뒤에, 모집단 기반 전진 선택법을 이용하여 교차 검증법의 적합도에 따라 표현자를 선택하도록 하였다. 표현자 선택, 교차 검증 등 모델 구현을 위해 사용된 모든 학습 방법들은 공개 프로그램인 Rapidminer v5.0¹⁵프로그램을 사용하여 결과를 얻었다. 표현자 선택을 위해 사용한 전진 선택법은 교차 검증의 적합도를 가장 잘 향상 시키는 표현자를 차례대로 개별 수식에 추가하게 되므로, 최종적으로 모델에 적합한 최적의 표현자 수를 결정할 필요가 있다. 5묶음(fold)로 나눈 교차 검증법을 수행하여 얻은 적합도(R^2_{cv} , RMSECV, MAECV)는 표현자의 수가 추가되면

서 모두 향상되는 결과를 나타내었다. 특히 통계적으로 의미가 있는 모델을 세우기 위하여 사용되는 기본 규칙인 화합물의 수의 1/10에 해당하는 25여종의 표현자까지 추가하여도 모든 적합도가 개선되는 결과를 나타내었기 때문에, 본 연구에는 표현자 선택을 위해서 다른 기준을 적용하여야 하였다. 이를 위해 도입된 기준값은 J_p 값으로, 그 수식은 다음과 같다.

$$J_p = \frac{(n+p) \cdot s^2}{n} = \frac{RSS}{n} \cdot \frac{(n+p)}{(n-p)} = \frac{RMSECV^2 \cdot (n+p)}{(n-p)}$$

식에서 n 은 사용된 화합물의 수, p 는 사용된 모델에 사용된 표현자의 수를 나타낸다.

각 학습 방법(MLR, SVM, ANN)의 표현자 수에 따른 5묶음 교차 검증법의 J_p 값에 대한 경향은 Fig. 2에 나타내었다. MLR 방법의 경우는 16종, SVM과 ANN 방법은 19종의 표현자를 포함할 때 가장 낮은 J_p 값을 지니므로, 각 조건을 최적의 표현자 집합으로 결정하였다. 결정된 표현자 집합에 대하여 SVM과 ANN 방법은 추가적으로 매개 변수 최적화 과정을 진행 하였다. SVM의 매개 변수는 C , γ , ν 이었으며, C 는 01, 1, 10, 100 그리고 γ 는 0.0001~0.1에 0.0001씩, 마지막으로 ν 는 0.1~0.5에 0.1씩 변환하여 모든 조건에서의 교차 검증법에 의한 RMSECV값이 가장 낮은 조건으로 결정하였다. SVM방법의 최적 매개 변수 조건은 $C=100$, $\gamma=0.0041$, $\nu=0.5$ 이었다. ANN에서 사용되는 매개 변수는 learning rate, momentum 그리고 은닉층의 뉴런 수이며, learning rate 와 momentum은 0.1~1.0 사이로 0.1씩 각각 변화하였고, 은닉층의 뉴런 수는 6개로만 고정하여 각 조건에서 마찬가지로 교차 검증에 의한 RMSECV값을 얻었다. ANN방법에서 매개 변수

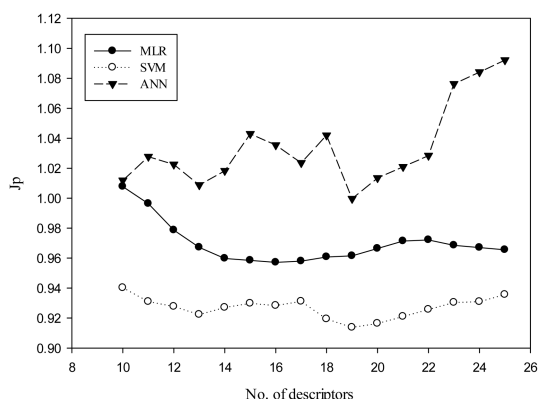


Fig. 2. J_p statistics vs. number of descriptors in the MLR, SVM and ANN model.

의 최적 조건은 learning rate=0.5, momentum=0.6 이었다.

각 학습 방법의 최적 모델은 Table 2에 나타내었고, 각 학습 방법의 예측값은 실험값과 비교하여 Table 1에 나타내었다. 각 학습 방법의 통계 결과를 비교하면, SVM방법이 교차 검증의 적합도 뿐만 아니라, 훈련 데이터에서의 적합도에서도 다른 방법과 비교하여 가장 우수한 것으로 나타났고, 실험값과 예측값과의 비교는 Fig. 3에 나타내었다. 교차 검증에 의한 R^2_{cv} 값을 3가지 학습 방법에 대하여 모두 0.6이상이므로, 기본적으로 QSAR에서 제안하는 0.5인 값¹⁶보다 크기 때문에 기본적으로 모델 예측에 모든 방법이 적합하다고 볼 수 있다.

3.2. MLR방법에 의한 QSAR 모델 분석

일반적으로 선형 모델인 MLR방법은 표현자들의 선형적인 관계를 수식으로 표현하므로, 모델 자체가 black box로 여겨 지는 SVM, ANN방법보다 표현자를 통해서 모델을 분석하고 해석하기가 용이하다. MLR 방법에서 가장 좋은 적합도를 나타낸 16종의 표현자에 대한 설명과 회귀 계수 및 t-ratio값을 Table 3에 나타내었다. MLR수식에서 회귀 계수의 부호는 급성 독성과의 상관관계를 나타내고, t-ratio 값은 회귀 계수와 표준 오차와의 비율을 t-test한 결과를 나타내며, 이 값의 절대값이 크면 상대적으로 수식에서 해당 표현자의 기여도가 크다고 평가할 수 있다. 이 수식에서 가장 기여도 높은 표현자는 No_{ring} 3로 분자 구조 안에서 삼각 고리의 수를 나타낸다. 삼각고리를 지니고 있는 분자는 매우 높은 어독성을 지닌 것으로 나타내는데 이것은 상대적으로 ring strain에 의한 구조 불안정성으로 인해 어류 내에서의 반응성이 증가하기 때문으로 해석된다. 다음으로 중요한 $AlogP98$ ^{17, 18}는 옥탄올/물 분배 계수를 Ghose가 제안한 원자 기여도 추가 방법에 의해 계산한 값으로, 분자의 소수성 또는 친지성 성질을 표현하는 표현자이다. 여기서는 주로 어류 표피의 지질층을 투과하는 성질과 깊은 관련이 있다. $AlogP98$ 의 회귀 계수 부호가 (+)를 나타내므로, $AlogP98$ 값이 큰 분자는 분자의 친지성이 커서 어류 표피의 지질막을 통과하기가 용이하여 어류 독성이 증가하는 것으로 설명할 수 있다. $AlogP98$ 값이 높을수록 Octanol에 분포 하며 비극성의 특징을 나타내고 비극성인 물고기의 피부 지질층을 통과하여 독성을 일으킨다. 하지만 너무 높은 $AlogP98$ 값은 큰 비극성을 띄어서 물에 잘 녹지 않아 독성 물질이 지질 층까

Table 1. 275 pesticides used in the study with experimental and predicted pLC₅₀ values

No.	Name	Exp. pLC ₅₀ (mM)	Pre. Log 1/LC ₅₀ (mM)			No.	Name	Exp. pLC ₅₀ (mM)	Pre. Log 1/LC ₅₀ (mM)		
			MLR	SVM	ANN				MLR	SVM	ANN
1	Fenoxycarb	2.275	2.275	2.841	1.932	34	Imazethapyr	-0.075	-0.075	-0.287	0.267
2	Diclotophos	1.576	1.576	1.461	0.683	35	Uniconazole	1.295	1.295	1.638	0.953
3	Fosamine	-0.346	-0.346	-0.458	-0.004	36	Dimethenamid	2.026	2.026	1.497	2.101
4	Difenzoquat	-0.295	-0.295	1.724	1.837	37	Tridiphane	2.781	2.781	4.326	3.288
5	Clopyralid	0.388	0.388	1.046	0.807	38	Fenoxaprop-ethyl	2.896	2.896	3.583	2.855
6	Dowicil	1.088	1.088	0.697	0.819	39	Chloroethoxyfos	3.577	3.577	3.430	3.234
7	3-Chloro-p-toluidine	1.264	1.264	0.668	1.498	40	Dichlorprop	2.672	2.672	1.225	0.563
8	Oleic acid	1.543	1.543	1.487	1.166	41	Fluvalinate	5.239	5.239	4.423	4.897
9	perfluorooctane sulfonate	2.081	2.081	2.132	2.877	42	Pacloutrazol	1.024	1.024	1.336	1.295
10	DDAC	2.112	2.112	2.281	2.003	43	Imazaquin	0.046	0.046	0.040	0.238
11	2-Mercaptobenzothiazole	2.414	2.414	2.326	2.757	44	Mecoprop	0.236	0.236	0.839	0.322
12	DTEA	2.518	2.518	1.280	2.176	45	Flurprimidol	1.225	1.225	2.328	1.992
13	Dichloroisocyanuric acid	2.944	2.944	2.202	1.977	46	Napropamide	1.46	1.46	2.055	1.802
14	Niclosamide	3.89	3.89	4.462	2.805	47	Fenarimol	2.198	2.198	2.499	2.054
15	fluoroacetic acid	0.268	0.268	0.226	0.521	48	Fonofos	4.091	4.091	3.113	3.749
16	Propionic acid	0.162	0.162	-0.160	0.572	49	Fenpropathrin	5.182	5.182	5.211	5.525
17	Pelargonic acid	0.24	0.24	0.435	1.796	50	Triadimenol	1.325	1.325	1.322	1.312
18	2,4-D Acid	0.303	0.303	0.980	0.488	51	Triadimefon	1.855	1.855	2.126	2.197
19	Triclopyr	0.341	0.341	1.104	0.720	52	Beta-Cyfluthrin	5.805	5.805	5.642	6.046
20	Naptalam	0.615	0.615	0.491	0.957	53	Difenoconazole	2.7	2.585	3.042	2.934
21	1-Naphthylacetic acid	0.823	0.823	0.600	1.166	54	Difethialone	4.024	4.460	3.893	4.220
22	Isopropanol	-2.335	-2.335	-0.670	-1.993	55	Resmethrin	5.616	4.711	5.256	5.285
23	4-(2-nitrobutyl) morpholine	2.4	2.4	1.204	1.173	56	1,3-Dichloropropene	1.45	1.847	1.108	1.712
24	Metalaxyl	0.332	0.332	1.588	1.169	57	Propetamphos	2.034	2.167	1.909	2.266
25	Sethoxydim	0.285	0.285	1.479	0.653	58	Tris (hydroxymethyl) nitromethane	-0.433	-0.372	-0.775	-0.368
26	Methomyl	2.276	2.276	1.354	1.779	59	Ethylene dichloride	-0.357	1.712	0.529	1.750
27	2,4-D 2-butoxymethylethyl ester	2.548	2.548	2.576	2.871	60	Dodine	-0.297	0.871	0.758	-0.261
28	Bromoxynil butyrate	4.078	4.078	3.417	2.699	61	Oxazolidine E	-0.188	0.707	0.439	1.119
29	Antimycin A	7.739	7.739	7.167	7.397	62	Z-11-Hexadecenal	-0.128	2.485	0.215	1.950
30	Nerolidol	2.092	2.092	1.629	2.182	63	Triflusulfuron-methyl	-0.119	0.574	-0.461	0.380
31	Imazamethabenz-methyl	0.013	0.013	1.465	0.755	64	3,5 Dimethyl-1-(hydroxymethyl) pyrazole	0.011	0.114	0.353	0.735
32	Farnesol	2.092	2.092	1.398	1.749	65	Oxadine A	0.026	-0.047	-0.075	0.664
33	Spinosad	1.387	1.387	2.977	1.054	66	Daminozide	0.031	-1.085	0.026	-0.308

Table 1. Continued

No.	Name	Exp. pLC ₅₀ (mM)	Pre. Log 1/LC ₅₀ (mM)			No.	Name	Exp. pLC ₅₀ (mM)	Pre. Log 1/LC ₅₀ (mM)		
			MLR	SVM	ANN				MLR	SVM	ANN
67	Fenridazone	0.032	0.893	0.266	0.365	103	Dimethoate	1.568	2.143	1.911	2.756
68	Imidacloprid	0.048	1.914	2.213	1.697	104	Chlorpropham	1.574	1.888	1.966	1.912
69	Glyphosate	0.101	0.103	0.443	-0.265	105	Norflurazon	1.574	2.171	1.641	1.726
70	Bentazone	0.14	1.047	0.999	-0.094	106	Trimethacarb	1.614	1.872	1.978	1.886
71	Tebuthiuron	0.203	1.164	1.830	1.140	107	2,4-DP, 2-EH Ester	1.643	3.584	3.679	3.611
72	Acephate	0.222	0.134	1.542	0.754	108	Cycloate	1.68	2.000	2.023	2.059
73	Limonene	0.231	1.362	1.051	1.538	109	Atrazine	1.681	1.485	1.639	1.402
74	Diflubenzuron	0.346	2.415	2.182	1.476	110	Chlorimuron ethyl	1.694	0.544	1.352	1.002
75	Primisulfuron-methyl	0.348	0.951	0.690	0.473	111	Chlorthal-dimethyl	1.702	3.439	2.245	3.349
76	Mesotrione (AMBA)	0.355	1.966	0.698	1.565	112	Oxamyl	1.718	1.331	1.270	1.424
77	Methyl chloroform	0.409	2.079	0.751	2.411	113	Chloroneb	1.748	2.016	1.304	2.178
78	N-methylneodecanamide	0.411	1.375	1.481	1.528	114	Methyl nonyl ketone	1.754	1.665	1.412	1.696
79	Diethyltoluamide	0.429	1.431	1.318	1.425	115	Vernolate	1.764	1.985	2.051	1.929
80	Simazine	0.456	1.444	1.841	1.353	116	Formetanate	1.768	1.333	1.425	1.312
81	Hexadecadienol acetate	0.471	2.890	0.814	2.897	117	Flutolanil	1.777	2.185	2.263	2.147
82	Cymoxanil	0.512	0.430	0.780	-0.067	118	O-Phenylphenol	1.792	1.717	1.608	1.436
83	Diflufenzopyr	0.526	0.318	0.798	0.434	119	Flufenacet	1.794	2.765	2.137	2.803
84	Dimethipin	0.575	0.615	0.917	0.690	120	Naphthalene	1.807	1.970	1.465	1.450
85	Cimectecarb	0.584	2.946	0.335	1.318	121	2-Hydroxyethyl octyl sulfide	1.817	1.396	2.159	1.784
86	Terbacil	0.671	1.127	0.742	0.817	122	Terbutylazine	1.83	1.493	1.740	1.403
87	Bronopol	0.678	0.912	0.525	0.809	123	acetaldehyde	1.84	1.618	1.497	1.538
88	Metribuzin	0.708	0.942	1.050	1.138	124	Tebuconazole	1.845	1.374	1.503	1.868
89	Aldicarb sulfone	0.724	0.874	1.118	0.765	125	Ametryne	1.852	1.621	1.768	1.499
90	Methyl anthralinate	0.775	0.767	1.118	0.646	126	Metolachlor	1.862	1.690	2.178	2.074
91	Bromacil	0.861	1.513	1.168	1.148	127	Diphenylamine	1.886	1.865	1.554	1.614
92	1,3,5-Triethylhexahydro-s-triazine	0.866	0.519	0.523	0.700	128	Fluridone	1.889	2.576	2.231	2.305
93	Pirimicarb	0.915	1.448	1.443	1.702	129	Iprodione	1.896	1.877	1.664	1.999
94	EPTC	0.977	1.770	1.873	1.699	130	Linuron	1.919	2.038	1.960	2.215
95	Ethoxyquin	1.053	0.853	1.395	1.838	131	Prometryn	1.92	1.688	1.870	1.592
96	Prometon	1.061	1.272	1.385	1.259	132	Cyprodinil	1.971	3.251	1.629	2.711
97	Clomazone	1.101	2.117	2.141	1.862	133	Benzisothiazolin-3-one	1.975	1.944	2.355	1.699
98	Pebulate	1.439	1.951	1.960	1.941	134	Propanil	1.977	2.036	2.284	2.141
99	Siduron	1.458	1.655	1.833	1.370	135	Parathion-methyl	1.981	2.827	2.323	2.873
100	Paranitrophenol	1.49	1.610	1.582	1.628	136	Chlorflurenol methyl	1.987	2.368	1.749	2.160
101	Dichlobenil	1.543	2.604	2.302	2.331	137	Oryzalin	2.026	1.466	2.013	1.625
102	Nitrapyrin	1.551	2.936	1.937	2.880	138	Diethatyl ethyl	2.039	2.629	2.211	2.226

Table 1. Continued

No.	Name	Exp. pLC ₅₀ (mM) ^a	Pre. Log 1/LC ₅₀ (mM)			No.	Name	Exp. pLC ₅₀ (mM) ^a	Pre. Log 1/LC ₅₀ (mM)		
			MLR	SVM	ANN				MLR	SVM	MLR
139	Trichloromelamine	2.06	1.830	2.403	2.781	176	Etridiazole	2.507	2.565	2.408	2.920
140	Thiazopyr	2.067	3.393	2.675	3.298	177	Sulfotep	2.508	3.486	3.323	3.461
141	Carboxin	2.071	1.680	1.728	0.973	178	Chlorhexidine	2.524	1.989	2.866	2.528
142	Diuron	2.078	1.795	2.018	2.104	179	Oxydemeton-methyl	2.528	2.127	2.350	2.178
143	Diphacinone	2.082	3.248	2.424	2.896	180	Irgarol	2.529	3.188	2.210	2.270
144	Bromoxynil	2.122	2.557	2.308	2.526	181	Aldicarb	2.531	1.491	1.708	1.597
145	Temephos	2.126	4.722	2.468	4.290	182	Dicloran	2.568	2.125	2.175	2.714
146	Thiodicarb	2.126	3.081	1.784	2.522	183	Parathion	2.572	2.844	2.633	2.885
147	Tebupirimfos	2.157	2.351	2.550	2.861	184	2,4-D Isopropyl Ester	2.581	2.365	2.289	2.251
148	Strychnine	2.163	1.422	1.821	2.226	185	Isopropalin	2.586	3.336	3.181	2.742
149	1,2-Dibromo-2,4-dicyanobutane	2.182	2.493	2.503	2.426	186	Flumiclorac pentyl	2.586	2.934	3.751	3.315
150	2,4-D butoxyethyl ester	2.187	2.534	2.529	2.587	187	Bifenazate	2.597	2.412	2.315	2.141
151	Flumioxazin (V-53482)	2.188	2.138	2.408	2.589	188	Amitraz	2.598	1.891	2.184	2.200
152	Ryanodine	2.188	1.746	2.531	1.612	189	Chloroxylenol	2.639	1.272	1.800	1.715
153	Parachlorometacresol	2.192	1.315	2.172	1.536	190	Tribufos	2.678	4.213	3.020	4.144
154	Paradichlorobenzene	2.223	2.379	2.565	2.331	191	Fenthion	2.704	2.976	3.149	2.975
155	Carbaril	2.225	2.310	1.925	1.711	192	Fludioxonil	2.723	1.903	2.893	2.280
156	MTI	2.242	1.626	1.587	1.522	193	Triclopyr BEE	2.739	2.843	3.027	3.051
157	Bendiocarb	2.27	1.638	1.707	1.572	194	Bensulide	2.742	1.939	2.885	3.187
158	Piperonyl butoxide	2.274	1.915	1.948	2.301	195	Carbofuran	2.765	1.566	2.000	1.737
159	Isofenphos	2.283	2.700	2.059	2.839	196	Triflumizole	2.775	2.278	2.189	2.646
160	MGK-264	2.294	2.261	1.952	2.279	197	2,4-D butyl ester	2.841	2.602	2.498	2.568
161	Imazalil	2.303	2.260	2.424	2.255	198	Acetochlor	2.851	1.640	2.147	2.020
162	Phenmedipham	2.328	2.360	2.113	2.402	199	Pirimiphos methyl	2.878	2.300	2.723	2.771
163	1,3-Dichloro-5,5-dimethylhydantoin	2.335	1.622	1.182	2.100	200	Ethion	2.886	4.317	3.229	4.062
164	Bromuconazole	2.346	2.428	2.975	2.703	201	Methyl isothiocyanate	2.89	1.332	0.619	1.578
165	Fenbuconazole	2.351	2.511	2.693	2.605	202	Butralin	2.902	3.064	2.825	2.094
166	Fenhexamid	2.353	3.367	2.649	2.745	203	Azoxystrobin	2.934	2.984	3.097	2.897
167	Acibenzolar-s-methyl	2.378	2.856	2.491	2.664	204	Dithiopyr	2.941	3.758	2.995	3.411
168	DBNPA	2.384	1.919	2.042	1.919	205	Oxyfluorfen	2.946	3.057	3.643	3.289
169	Thiobencarb	2.39	2.517	2.587	2.326	206	PCNB	2.965	4.297	3.384	3.982
170	Dipropyl isocinchomeronate	2.4	2.372	1.665	1.827	207	Triclosan	3.002	2.807	3.335	2.926
171	Tri-allate	2.405	2.607	2.209	2.700	208	Pindone	3.04	2.346	1.557	1.865
172	Carfentrazone-ethyl (F8246)	2.411	2.784	2.789	3.066	209	Clodinafop-propargyl	3.067	3.090	2.930	3.130
173	Alachlor	2.431	1.732	2.088	2.074	210	Propachlor	3.095	1.727	1.626	1.564
174	Oxadiazon	2.459	2.660	3.470	2.922	211	Methylene bis(thiocyanate)	3.165	3.017	2.823	2.639
175	Clorophene	2.482	2.114	2.265	2.088	212	Chlorophacinone	3.172	3.499	3.094	3.259

Table 1. Continued

No.	Name	Exp. pLC ₅₀ (mM)	Pre. Log 1/LC ₅₀ (mM)			No.	Name	Exp. pLC ₅₀ (mM)	Pre. Log 1/LC ₅₀ (mM)		
			MLR	SVM	ANN				MLR	SVM	ANN
213	EPN	3.187	3.214	3.328	3.174	245	4,5-Dichloro-1,2-dithiol-3-one	4.126	3.704	3.784	3.665
214	Kresoxim methyl	3.217	2.730	2.321	2.604	246	Trifluralin	4.183	3.524	3.837	2.967
215	Benomyl	3.232	1.621	2.011	1.932	247	gamma-HCH	4.208	3.419	3.866	3.845
216	Chinomethionat	3.263	3.355	2.920	3.008	248	Chlordecone	4.228	5.108	3.885	4.360
217	Thiram	3.267	3.549	3.531	3.164	249	Flumetralin	4.245	4.163	3.903	3.614
218	1,2-Benzenedicarboxaldehyde	3.27	2.003	1.655	1.261	250	Profenofos	4.25	3.691	4.548	3.481
219	Anilazine	3.294	2.810	2.748	2.647	251	Folpet	4.296	3.351	3.954	3.269
220	Pendimethalin	3.309	3.007	2.849	2.212	252	Phorate	4.302	2.913	4.644	3.290
221	Zoxamide	3.334	2.615	2.728	2.856	253	Brodifacoum	4.321	4.066	4.196	4.278
222	Dichlorvos	3.344	2.743	2.284	2.347	254	Methidathion	4.334	3.460	3.992	3.543
223	Naled	3.377	3.404	2.928	3.517	255	Prallethrin	4.399	4.547	3.725	3.954
224	Propargite	3.473	2.403	3.131	2.660	256	Chlorpyrifos-methyl	4.408	3.440	4.066	3.569
225	Dicofol	3.475	3.157	3.701	3.260	257	Trifloxystrobin	4.465	3.054	4.123	3.166
226	Phosmet	3.48	3.109	3.159	3.253	258	Terbufos	4.579	3.004	3.852	3.385
227	Diiodomethyl p-tolyl sulfone	3.511	2.477	2.637	2.727	259	Heptachlor	4.727	3.721	5.069	4.149
228	Hydramethylnon	3.518	3.131	3.860	3.436	260	Chlorfenapyr (Pirate)	4.739	2.935	3.837	3.327
229	Bromoxynil octanoate	3.59	3.768	3.620	3.556	261	Malathion	4.917	3.658	3.005	3.173
230	Captan	3.615	3.067	3.701	3.098	262	DDT	4.937	3.470	4.595	4.008
231	Benfluralin	3.617	3.525	3.778	2.976	263	Kathon 930	5.019	2.909	3.075	2.821
232	3-Iodo-2-propynyl butylcarbamate	3.623	2.193	2.307	1.718	264	Rotenone	5.317	3.319	4.975	3.336
233	Octhilinone	3.657	2.226	2.206	1.946	265	Dieldrin	5.502	4.989	5.370	4.342
234	TDE	3.66	3.198	3.660	3.684	266	Fenpyroximate	5.583	2.931	5.625	3.117
235	Tetrapropyl thiodiphosphate	3.697	3.849	4.039	3.736	267	Tralomethrin	5.619	6.587	5.652	6.878
236	Pentachlorophenol	3.894	3.175	3.417	3.597	268	Pyridaben	5.705	2.697	3.041	2.890
237	Bromonitrostyrene	3.927	2.933	2.114	2.450	269	Endrin	5.712	4.989	5.370	4.342
238	Ethalfuralin	3.955	3.469	3.358	2.902	270	Deltamethrin	5.744	5.660	6.087	5.977
239	Fluthiacet methyl	3.973	3.905	3.480	3.451	271	Beta cypermethrin	6.028	5.555	5.869	5.877
240	Dienochlor	3.977	4.415	4.319	4.398	272	Cyphenothrin	6.043	5.311	6.270	5.711
241	Chloropicrin	3.998	4.158	3.455	3.928	273	Lambda-Cyhalothrin	6.273	5.727	5.931	6.243
242	Bis(trichloromethyl)sulfone	4.016	4.207	4.358	4.137	274	Bifenthrin	6.45	5.291	6.107	6.130
243	Fluazinam	4.111	3.678	3.769	4.043	275	Tefluthrin	6.844	5.152	5.101	6.166
244	Chlorpyrifos	4.113	3.418	4.673	3.524						

Table 2. Statistical parameters produced by the different models

Methods	No. descriptors	Parameter	Training set (N=275)			Cross-validation (5 fold)			
			R ²	RMSE	MAE	R ² _{cv}	RMSECV ^a	MAECV ^b	Jp ^c
MLR	16	-	0.688	0.871	0.650	0.649	0.923	0.693	0.957
SVM	19	C=100, γ =0.0041, ν =0.5	0.812	0.663	0.486	0.677	0.887	0.674	0.914
ANN	19	Learning rate=0.5, momentum=0.6, no.hidden nodes=6	0.709	0.840	0.612	0.645	0.931	0.708	1.000

^aroot mean square error of cross-validation, ^bmean absolute error of cross-validation, ^cJp statistics.

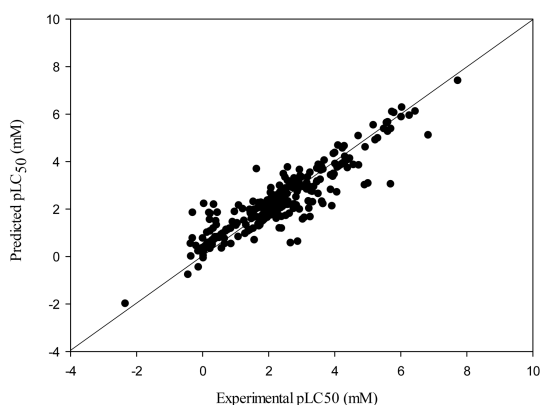


Fig. 3. Plot of experimental vs. predicted pLC₅₀ in the best SVM model.

지 도달하지 못하고, 너무 낮은 AlogP98값은 물에 많이 분포하나 독성 물질이 큰 극성을 띠기 때문에 비극

성인 지질층을 통과하지 못한다.¹⁹ 결론적으로 물에 어느 정도 녹으면서도 비극성 성질을 지닌 물질이 독성으로 크게 작용 할 것이다. HRNCS (The relative negative charged surface area to H-bond acceptors atoms) 표현자는 수소결합받게 원자당 (-)전하를 지닌 표면적으로서, 수소결합받게의 수에 비해 상대적으로 (-)전하를 지닌 표면적이 크면 어류의 표피와 반응성이 증가되어 독성이 증가한다고 볼 수 있다. FraVSA_{don}²⁰는 전체 분자 표면적에 대한 수소 결합 주계 표면적의 비를 나타내며, FraVSA_{don}값이 커질수록 어류의 지질층을 통과하기 어려워져 독성이 낮은 것으로 판단된다. 이러한 경향은 CATS_{AccNeg2} (수소 결합 받게와 (-) 전하를 띤 원자간의 위상 거리가 2정도 떨어져 있음을 가리킴), (+)PSA_{MPEOP} (MPEOE방법^{21,22}으로 계산한 전하를 이용하여 (+) 전하를 띤 극성 표면적),

Table 3. Molecular descriptors involved in the best MLR model derived for pLC₅₀

No	Molecular descriptors	Description	Coefficient	t-ratio
1	No _{ring3}	The number of 3 membered rings	1.478	5.93
2	AlogP98	Calculated logP by Ghose's atom additive method	0.254	5.88
3	HRNCS	The relative negative charged surface area to H-bond acceptors atoms	0.015	4.25
4	CATS _{AccNeg2}	Indicator for atom pair in 2 distance between H-bond acceptor atom and (-) chargable atom	-2.212	-3.96
5	FraVSA _{don}	Fraction of 2D Van der Waals H-bond donor surface area	-4.972	-3.85
6	FPSA1	The partial positive VDW surface area divided by the total VDW surface area	-0.906	-3.63
7	NoSdO	The number of double bonds between S atoms and O atoms	-0.484	-3.57
8	Vchi 3	Kier & Hall valence connectivity index of order 3 (path)	0.102	3.44
9	Fra _{hetero}	Fraction of hetero atoms	1.746	3.10
10	CATS _{DonAcc10}	Indicator for atom pair in 10 distance between H-bond donor atom and H-bond acceptor atom	1.920	2.67
11	E _{stateS_{SS}}	Sum of E-state for -S- type	-0.260	-2.46
12	No _{esterr}	The number of ester groups	0.324	2.46
13	CATS _{DonDon5}	Indicator for atom pair in 1 distance between H-bond donor atom and H-bond donor atom	1.514	2.38
14	CATS _{DonAcc5}	Indicator for atom pair in 1 distance between H-bond donor atom and H-bond acceptor atom	-0.400	-1.86
15	No _{(-) groups}	The number of (-) chargable groups	0.371	1.45
16	(+)PSA _{MPEOE}	The sum of VDW surface area with values of MPEOE ^b charges greater than 0.2	-0.004	-1.37
17	Intercept	Intercept	1.358	-

^a) chemical Advanced Template Search, ^b) the Modified Partial Equalization of Orbital Electronegativities (ref. 21-22)

FPSA1 (전체 표면적에서 (+)전하를 띤 표면적의 비율), NoSdO (O=S<기의 수) 등의 표현자에서도 나타나며, 마찬가지로 전하를 지닌 표면적이 증가할수록 어류의 지질층을 통과하기 어려운 것으로 판단된다. 그 이외의 표현자들은 모델 내에서의 기여도가 낮아서 상세하게 설명하지 않고 생략하도록 한다. 표현자를 이용한 무지개 송어의 MLR모델 분석을 통해서 특정 화학 구조의 반응성(No_{ing3} , HRNCS)에 의한 어독성을 확인할 수 있었고, 분자의 친지성이 높고(AlogP98), 전하를 띤 표면적이 감소할수록(FraVSA_{don}, CATS_{AccNeg2}, (+)PSA_{MPEOE}, FPSA1, NoSdO) 어류 지질층을 통과하기가 용이하여 어독성을 나타낸다는 사실을 확인할 수 있었다.

3.4. 규제 기준에 의한 모델 평가 및 검증

실제 EU에서는 화합물의 어류 독성에 따라 Table 4와 같이 4가지 분류에 따라 화합물을 규제하고 있다. 본 연구에서는 구현된 각 모델들의 예측값을 EU에서 정한 기준으로 분류하였을 때의 정확도(accuracy, %)를 계산하여 Table 5에 나타내었다. 각 분류 별로 정

Table 4. EC Classification for Fish (Directive 92/32EEC Annex VI Point 5.1)

Class	LC ₅₀	Dangerous for the environment
I	<1 mg/L	Very toxic to aquatic organisms
II	1-10 mg/L	Toxic to aquatic organisms
III	10-100 mg/L	Harmful to aquatic organisms
III	>100 mg/L	May cause long-term adverse effects in the aquatic environment

Table 5. Prediction accuracy of the different models for criteria of EC Classification

Method	MLR				SVM				ANN			
	Class I	Class II	Class III	Class IV	Class I	Class II	Class III	Class IV	Class I	Class II	Class III	Class IV
Class accuracy ^{a)}	82.54	52.87	20.00	48.15	90.48	88.51	80.00	77.78	80.16	73.56	65.71	55.56
Total accuracy ^{b)}	61.82				87.27				73.82			

^{a)} percentage of correctly predicted results for each class, ^{b)} percentage of correctly predicted results for all class

Table 6. Cross-validated results after 50 Y-scrambling test

Methods	Y-scrambling					
	R ² _{cv.ave.}	R ² _{cv.range}	RMSECV _{ave.}	RMSECV _{range}	MAECV _{ave.}	MAECV _{range}
MLR	0.160	0.052-0.280	1.427	1.322-1.516	1.099	1.009-1.163
SVM	0.157	0.038-0.283	1.432	1.324-1.534	1.108	1.030-1.168
ANN	0.097	0.023-0.193	1.490	1.412-1.549	1.151	1.083-1.198

확도를 확인해 보면, MLR과 ANN방법의 경우에는 class I과 II의 정확도는 우수한 반면에 class III과 IV의 정확도가 낮음을 알 수 있다. 반면에 SVM모델의 경우는 가장 낮은 정확도를 나타낸 것이 class IV로 약 78%를 나타내었고, 이 외의 class의 정확도는 모두 80%이상이었다. 전체 화합물에 대한 정확도에서도 SVM방법이 87%로 가장 우수한 것을 확인할 수 있었다.

많은 표현자로부터 다양한 학습 방법에 따라 예측 모델을 구현하였기 때문에, 이 과정에서 우연히 어독성과 상관성을 나타냈을 가능성을 확인하기 위하여 Y-scrambling방법을 통해 우연 상관성(chance correlation)을 확인하였다. 이 과정은 화합물간의 종속변수(Y)만을 뒤섞고, 그 외의 과정(데이터 전처리부터 학습 방법 적용)은 앞에서 최적 모델을 결정할 때와 동일한 과정으로 수행하였고, 그 결과는 Table 6에 나타내었다. 이 과정은 총 50번을 반복하여 수행하였고, 마찬가지로 교차 검증방법에 의해 결정된 적합도(R²_{cv}, RMSECV, MAECV)를 계산하였다. 결과적으로 Y-scrambling test에서 얻은 적합도는 최적 모델의 적합도와 비교하여 큰 차이를 나타내므로, 이 모델은 우연 상관성이 없는 것으로 확인할 수 있었다.

4. 결 론

본 연구에서는 환경에 영향을 미치는 농약류에 대하여 무지개 송어의 어독성을 예측하는 모델을 구현하고 분석하였다. 학습방법으로 선형 모델(MLR), 비선형 모델(SVM, ANN)을 이용하였고, 2차원 화학 구

조에서 쉽게 계산할 수 있는 표현자만을 이용하여 QSAR예측모델을 구현하였다. 본 연구에 사용된 모든 학습 방법은 5묶음 교차 검증 법과 Y-scrambling test에 의해서 얻은 적합도(R^2_{cv} , RMSECV, MAECV)를 통해서 모델의 예측 능력과 견고함을 확인 할 수 있었다. 학습 방법 중에서 SVM방법은 가장 좋은 예측 결과를 나타내었고, MLR방법을 통해서는 어독성을 유발하는 반응성에 대한 구조적 특징과 지질 투과성 및 정전기적인 상호 작용의 영향 등을 모델에 사용된 표현자를 통해서 설명할 수 있었다. 특히 EU에서 규정한 규제 기준에 따라 농약류를 분류하여 87%의 정확도를 나타내므로, 이 모델은 규제기관의 관리용 방법으로 활용할 수 있는 가능성을 제기 할 수 있었다. 특히 규제기관에서 현재 어독성이 측정되지 않은 농약류에 대하여 어독성 실험의 우선순위를 결정하는 분야에 기여할 수 있을 것으로 예상된다.

감사의 글

이 논문은 2010년도 교육과학기술부의 재원으로 한국연구재단의 기초연구사업 (신진연구지원사업) 지원을 받아 수행되었습니다(2010-0006124).

참고문헌

1. J. Ahlers, F. Stock and B. Werschkun, *Environ. Sci. Pollut. Res.*, **15**(7), 565-572 (2008).
2. G. T. Miller, In 'Sustaining the Earth', 6th Ed., pp 211-216, Thompson Learning, Pacific Grove, CA, 2004.
3. OECD, 'OECD Guidelines for Testing Chemicals, METHOD 203, Fish, Acute Toxicity Test', Paris, 1984.
4. 'Technical Guidance Document in support of Commission directive 93/67/EEC for new notified substances and Commission regulation (EC) No 1488/94 on risk assessment for existing substances', Brussels, 1996.
5. J. S. Jaworska, M. Comber, C. Auer and C. J. Van Leeuwen, *Environ. Health Perspect.*, **111**(10), 1358-1360 (2003).
6. 'Development of Environmental Modules for Evaluation of Toxicity of pesticide Residues in Agriculture (DEM-ETRA)', project supported by the European Commission, Contract No. QLK5-CT-2002-00691.
7. A. Roncaglioni, E. Benfenati, E. Boriani and M. Clook, *J. Environ. Sci. Health, Pt. B:Pestic., Food Contam., Agric. Wastes*, **39**(4), 641-652 (2004).
8. M. Casalegno, G. Sello and E. Benfenati, *Chem. Res. Toxicol.*, **19**(11), 1533-1539 (2006).
9. S. K. Lee, S. H. Park, I. H. Lee and K. T. No, PreAD-MET Ver.v2.0, BMDRC: Seoul, Korea, 2007.
10. G. Schneider, W. Neidhart, T. Giller and G. Schmid, *Angew. Chem. Int. Ed. Engl.*, **38**(19), 2894-2896 (1999).
11. N. R. Draper and H. Smith, In 'Applied Regression Analysis', 2nd Ed., pp 294-379, John Wiley & Sons Inc., New York, 1981.
12. C. Cortes and V. Vapnik, *Mach. Learn.*, **20**(3), 273-297 (1995).
13. B. Schölkopf, A. J. Smola, R. C. Williamson and P. L. Bartlett, *Neural Comput.*, **12**(5), 1207-1245 (2000).
14. D. E. Rumelhart, G. E. Hinton and R. J. Williams, *Nature*, **323**(6088), 533-536 (1986).
15. Rapidminer Ver.5.0, Rapid Miner is unquestionable the world-leading open-source system for data mining, Rapid-I: Dortmund, Germany, 2010.
16. B. L. Podlogar, I. Muegge and L. J. Brice, *Curr. Opin. Drug Dis. Dev.*, **4**(1), 102-109 (2001).
17. A. K. Ghose, V. N. Viswanadhan and J. J. Wendoloski, *J. Phys. Chem. A*, **102**(21), 3762-3772 (1998).
18. J. Huang and X. Fan, *Mol. Pharm.*, **8**(2), 600-608 (2011).
19. T. Jager and S. A. L. M. Kooijman, *Ecotoxicology*, **18**(2), 187-196 (2009).
20. P. Ertl, B. Rohde and P. Selzer, *J. Med. Chem.*, **43**(20), 3714-3717 (2000).
21. K. T. No, J. A. Grant, M. S. Jhon and H. A. Scheraga, *J. Phys. Chem.*, **94**(11), 4740-4746 (1990).
22. K. T. No, J. A. Grant and H. A. Scheraga, *J. Phys. Chem.*, **94**(11), 4732-4739 (1990).