Journal of Research High School

Vol. 2026 (5) 428 – 437
ISSN 2688-3651 [online]

Evaluation of Four Machine Learning Methods to Estimate Dermal Permeation Coefficient for Organic Chemicals

Timothy Yoo[1] *

[1]Stone Bridge High School, Ashburn, VA, USA
*Corresponding Author: 1074437@lcps.org

Advisor: Myoungwoo Kim, information@bluebellenvirobotics.com

Received July 13, 2025; Revised February 5, 2026; Accepted March 16, 2026

Abstract

Accurately predicting skin permeability (logKp) is essential for evaluating the dermal absorption of chemical compounds in drug discovery and development. Traditional models, for example, the equation by Potts and Guy, rely primarily on logP and molecular weight to predict the skin permeability. However, these approaches often suffer from limited accuracy from oversimplification. In this study, the use of machine learning models including classification and regression trees (CART), extreme gradient boosting (XGBoost), random forest (RF), and artificial neural network (ANN), was investigated to enhance the prediction of the skin permeability. To form training data, logKp values of 245 organic chemicals were collected from a public dataset. Subsequently, 8 to 10 features out of 1,444 molecular descriptors calculated by PaDEL program were selected using sequential feature selection (SFS), a method that typically selects features based on categorical correlation between a dependent variable and multitude of independent variables. Results have found that classification and regression-based models such as CART, XGBoost, and RF displayed higher R^2 values ranging from 0.84 to 0.92, which significantly outperformed other traditional statistical models. However, ANN exhibited relatively lower R^2 of 0.42, showing that neural networks performed poorer than other models. These findings revealed that XGBoost paired with SFS works the best for logKp *in silico* modeling with a large number of training dataset and provided valuable insight into the molecular features most relevant to dermal absorption.

Keywords: Dermal permeability coefficient, Machine learning, In silico, Molecular descriptors

1. Introduction

It is essential to be able to understand and predict skin permeability (K_p) in fields like drug delivery, chemical safety, and regulatory toxicology. Drug delivery refers to the targeting of drugs to lesions in the skin with minimum systemic absorption, and is widely used to treat dermatitis, skin cancer, and microbial infections (Badili et al., 2018). In fields such as regulatory toxicology, dermal exposure has been recognized as the most relevant route of exposure specifically for pesticide applications (Lebailly, 2009). This is because many environmental chemicals are lipophilic and can easily be absorbed by the skin fat. Studies have shown exposure to chemicals had many negative effects on health. Acute health effects include irritation, chronic health effects include neurological and mental damage, reproductive effects, and even cancer (Macfarlane, 2013). Across a multitude of fields, skin permeability comes in as a very important data for experiments and developments of new chemicals. In order to obtain this information on various chemicals, *in vivo* animal studies have been conducted to estimate K_p values. However, these methods using animals have introduced a variety of issues involving ethical matters, cost effectiveness, and time consumption.

To address these issues, several government agencies including United States (U.S.) Environmental Protection Agency (EPA) and U.S. Food and Drug Administration (FDA) have made efforts to reduce animal studies with the usage of New Approach Methods (NAMs). NAMs include any technology, methodology, approach that can provide

Vol. 2026 (5) 428 – 437
ISSN 2688-3651 [online]

data on chemical hazard and risk assessment while avoiding conducting animal testing (U.S. Food and Drug Administration, 2024). With the rapid advancement of technology, this is made possible through the use of in silico modeling, which are computational modeling approaches and, in this case, models accompanied by statistics or machine learning. Another advantage of NAMs is the capability for cross-species extrapolation of data. Cross-species of extrapolation involve using knowledge about one species to predict or estimate effects in another species, essential in drug development. With this method, incredible amounts of data gaps can be filled and can also help in predicting human data (U.S. FDA, 2024).

However, challenges are still present in this approach. To use NAMs, specifically machine learning, the data should be treated and processed accordingly. This is essential for any machine learning studies and includes taking care of outliers, abnormal distribution, and excessive zeros. Without preprocessing the dataset, results would lack precision and accuracy. Additional errors can come from lack of known data or not using the algorithm of best fit. Recent studies demonstrate how these new approaches can be applied using various chemical properties and model architectures to predict the dermal permeability of a chemical.

One study, done by Kuster et al., presents a Random Forest model made to predict dermal absorption of active substances in plant protection products. The study used a dataset of 445 *in vitro* human skin studies covering 160 active substances and 28 different formulation types. All these datasets were obtained from the EFSA dermal absorption database and proprietary experimental data. 13 different physicochemical descriptors, including molecular weight, molecular volume, logD (lipophilicity), solubility, and hydrogen bonding characteristics, were used as input features to train the model. The model was trained using leave-one-out cross-validation (LOO-CV) and externally validated with 17 additional substances, which showed a strong correlation between predicted and experimental values. One contribution of this paper is the use of percentile-based predictions. This new approach was used to take account of biological variability across donors, formulations, and concentrations. Instead of predicting one point value, the model helps safety assessors to account for the upper bounds of dermal absorption. The study also states that existing models only focus on the stratum corneum and ignore formulation effects or variability (Kuster et al., 2022). A challenge in this study was dependence on a single machine learning algorithm. Relying solely on Random Forest models cannot capture all the nonlinearities the dataset has.

Another study titled "Machine Learning for Skin Permeability Prediction: Random Forest and XGBoost Regression" by Ita and Prinze, presents two regression models as said in the title. K_p is predicted using both Random Forest and XGBoost using Abraham solute descriptors of the chemical. These descriptors included excess molar refraction (E), dipolarity/polarizability (S), hydrogen bond acidity (A), hydrogen bond basicity (B), and McGowan's characteristic volume (V), all of which are known to influence a molecule's skin penetration. The test validation splits the training data into 8:2 ratio and evaluated models using R^2, mean absolute error, mean squared error, and root mean squared error. Based on the results, the XGBoost model outperformed Random Forest slightly in predictive accuracy, which was reasonable due to its strength in modeling non-linear relationships. It was found that the most important descriptor among the Abraham solute descriptors was the hydrogen bond basicity (B) for both models. This is consistent with the known mechanisms of skin absorption were hydrogen bonding influences transport through lipid membranes. The importance of choosing interpretable descriptors is emphasized in the article as it helps find the balance between prediction and model transparency, making it easier for applications (Ita & Prinze, 2024). Some challenges of this study included the lack of input features, which only used 5 abrahamic descriptors. These descriptors can be informative but can also be oversimplified, causing errors in the model.

Although recent studies with machine learning have been proven to be strong in predicting dermal permeability, important gaps remain. Many existing models use multiple molecular descriptors within the same category as input features, which introduces redundancy and can obscure true relationships between the descriptors and the permeability. In addition, most studies evaluate only one or two machine learning algorithms, limiting insight into how the performance varies across different models.

This study aims to predict the dermal permeability coefficient through machine learning using molecular descriptors as input features. Molecular descriptors (approximately 3,000 - 10,000 in quantity) are numerical representations of a molecule's structural information, used to predict its properties or activity (Grisoni et al., 2018). Past research findings have shown that different machine learning algorithms will indicate different chemical

Vol. 2026 (5) 428 – 437
ISSN 2688-3651 [online]

properties as the best performing features. This indicates that separate steps should be taken to select molecular descriptors, and in this study, sequential feature selection (SFS) will be used for this process. SFS is an algorithm that continuously adds or removes features in a dataset and tests its accuracy with the machine learning algorithm (GeeksForGeeks, 2025). Before applying SFS, the molecular descriptors will be sorted by its descriptor type, which will reduce variability and data deception by reducing the dimension. Dimensionality reduction is the conversion from a high dimensional and complex space into a lower dimensional and simplified form of information. In essence, the goal of dimensionality reduction in the context of this study is to remove overlapping input features, or those in the same category, so that the result is not affected by any repeated features. This process will allow SFS to select the best descriptor within each descriptor type.

Finally, this study aims to compare multiple machine learning models. The four machine learning algorithms that will be used are Classification and Regression Tree (CART), Random Forest, Extreme Gradient Boost (XGBoost), and Artificial Neural Network (ANN). After the process of training, whichever model that performs the best will be chosen to be used as the final model. This way, the challenges of other studies, such as dependence on a single machine learning, can be solved, and the limited input features can be solved by using SFS. With these structured ideas, the research question was to find out whether SFS had a significantly positive effect compared to traditional machine learning models.

The contribution of this work lies in the systematic integration of SFS with multiple machine learning models applied to the dermal permeability dataset. While prior studies primarily focused on optimizing predictive accuracy using fixed or limited descriptor sets, this study advances QSAR-based dermal modeling by explicitly addressing descriptor redundancy and overlap in high-dimensional molecular descriptor dataset. By applying SFS across four models, this work shifts the focus from identifying a single best-performing model to understanding how feature selection interacts with model structure to influence interpretability and accuracy.

2. Materials and Methods

2.1 Data Mining

Skin permeability is the measure of how well substances can pass through a skin. These values are often expressed through cm/h, cm/s, or log K_P (Akhtar et. al., 2016). For this study, a set of data from another study was used. This set of data included empirically- based values of K_P for 245 organic compounds including industrial, cosmetic, agricultural, plastic polymer compounds. The compounds that didn't have CAS numbers were removed from the dataset as they were experimental and were not publicly recognized. After excluding 10 compounds that are not registered as CAS number, the used data, which was 235 compounds in total, consisted of the chemical name, chemical abstracts service (CAS) number, often used to identify the chemical, the log K_P values, original values of K_P, molecular weight, and log K_{ow} (Brown et al., 2016). With this information, the SMILES code was determined using the database from the National Institutes of Health (NIH). With the SMILES code to the corresponding organic compound, 1D and 2D molecular descriptors were calculated using the PaDEL program.

2.2 Calculation of Molecular Descriptors Using PaDEL

Once the SMILES strings for all 245 compounds were collected, the calculation of molecular descriptors was conducted using PaDEL-Descriptor, an open-source software tool developed for cheminformatics analysis. PaDEL computes a broad array of molecular descriptors that numerically encode the chemical structure and properties of a compound. In PaDEL, these descriptors are categorized into different classes based on what they measure. Examples include topological indices, atom-type counts, charge distribution, autocorrelation metrics, and indicators of molecular size, symmetry, and polarity. In this study, both 1D and 2D descriptors were computed, resulting in 1,444 descriptors per compound. These descriptors serve as independent variables in various machine learning models within this field. The PaDEL program also had divisions of molecular descriptors based on its properties, and this study used 1 descriptor per category, ensuring no redundancy between the input features. The selection of the single desccriptor was done through SFS (see more in 2.3) because having multiple correlated variables gives no additional predictive information but rather just inflates variance in coefficient estimates.

Vol. 2026 (5) 428 – 437
ISSN 2688-3651 [online]

2.3 Sequential Feature Selection (SFS)

Sequential Feature Selection (SFS) is a wrapper-based feature selection method that aims to identify a subset of the most relevant features from a larger set of input features. It works through a greedy algorithm, starting either with an empty set, namely the forward selection, or the full feature set, called the backward selection. With this, the algorithm iteratively adds or removes features based on their importance to the model's performance. SFS provides several benefits, especially in a data set with high dimensions. First, by selecting only the most informative features, the risk of overfitting is greatly diminished (GeeksforGeeks, 2025). This would be even more effective since the number of input features is far greater than the sample size. In addition, using SFS increases model transparency, because optimized subset of features allows for much easier interpretation of the model, especially important in fields such as bioinformatics or cheminformatics. SFS also enhances computation efficiency, as training models with fewer features greatly reduce computational load and training time. In this study, SFS was used within each PaDEL descriptor type to ensure no redundancy between the input features and within each descriptor type; the most informative one was selected.

2.4 Machine Learning Models

Machine learning refers to a set of algorithms that learn patterns from data to make predictions or decisions. Its significance lies in its ability to navigate through complex, nonlinear relationships in data that traditional statistical models may omit. In the context of predicting skin permeability (logKp), traditional empirical models use octanol water partition coefficient and molecular weight as the main variable (Potts & Guy, 1992), which introduce considerable uncertainty, as there are multitudes of other factors that contribute to skin permeability To address this issue, machine learning is used to identify key descriptors and generate a model with greater predictive power.

Random Forest

Random Forest is an ensemble learning algorithm that builds multiple decision trees during training and outputs the average prediction (regression) or the majority vote (classification) of the individual trees. It helps reduce overfitting and improve generalization by using bootstrapped samples and random feature selection at each split.
The equation for random forest is shown by the following (Breiman 2001):

$$y = \frac{1}{T}\sum f_t(x) \tag{1}$$

Where T represents the number of trees

CART

CART is a decision tree algorithm that splits data recursively based on input features to predict either categorical outcomes (classification) or continuous values (regression). It uses metrics such as Gini impurity for classification and mean squared error for regression to determine the best splits. While CART model doesn't have a single compact equation like other models, its core idea is to recursively partition the feature space to minimize impurity. (Breiman, 2017)

XGBoost

XGBoost is a scalable and regularized gradient boosting framework. It builds an ensemble of decision trees sequentially, where each new tree corrects the residual errors of the previous trees. It uses second-order gradients and includes regularization terms to reduce overfitting and enhance performance. Regularization is a machine learning technique that penalizes the model for having complexity, which encourages simpler architecture and therefore prevents model from memorizing the training data. The equation for XGBoost is shown by the following:

$$L(\varphi) = \sum l(\hat{y}_i, y_i) + \sum \Omega(f_k) \tag{2}$$

$$Where\ \Omega(f) = \gamma T + \frac{1}{2}\lambda||w||^2$$

Vol. 2026 (5) 428 – 437
ISSN 2688-3651 [online]

In this equation, l is the differentiable convex loss function, and Ω represents penalization of the complexity of the model (Chen & Guestrin, 2016). The convex loss function in this equation creates a bowl-shaped landscape with a single minimum, which allows for efficient optimization and gradient descent techniques.

ANN

Artificial Neural Networks are inspired by biological neural networks and consist of interconnected layers of nodes (neurons). Each node applies an activation function to a weighted sum of its inputs. ANNs are capable of capturing complex, non-linear relationships between features and targets. The equation for a single neuron in ANN is shown by the following (Goodfellow et al., 2016).

$$\hat{y} = \sigma\left(\sum_{i=1}^{n} w_i x_i + b\right) \quad (3)$$

Where w_i represents the weight of the input feature, b represents the bias, and σ is the activation function

2.5 Model Training and Validation

For four of these machine learning models, sequential feature selection was performed using 5-fold cross validation with R^2 as optimization metric, selecting 10 features for CART, 8 for Random Forest, 7-10 for XGBoost, and 8 for ANN using scikit-learn's SFS.

Model configurations were optimized through grid search approaches across multiple train-test splits. Test set sizes of 10%, 20%, and 30% were evaluated across 50 random states for CART, RF, and XGBoost. The ANN model was evaluated using test sizes of 10%, 15%, 20%, and 30% with a fixed random state of 42 and varying training epochs (50, 100, 150). The ANN architecture consisted of two hidden layers (64 and 32 neurons) with ReLU activation, paired with the Adam optimizer. The model yielding the highest test set R^2 value was chosen for each algorithm.

Model performancce was evaluated using R^2 value and Average Absolute Fold Error (AAFE), calculated as mean. All models were implemented in Python 3 using scikit-learn for CART, RF, and feature selection; XGBoost for gradient boosting; and TensorFlow/Keras for ANN.

3. Results

3.1 Sequential Feature Selection (SFS) and Machine Learning

Table 1 presents the statistical performance indicators for each trained model, along with the molecular descriptors used in their respective equations. The Average Absolute Fold Error (AAFE) was calculated by computing the fold error for each prediction, subtracting 1, taking the absolute value, and then averaging the results. In this formulation, lower AAFE values indicate smaller deviations between predicted and experimental values.

Table 1. Selected Descriptors along with R^2 and AAFE by SFS.

Machine Learning	No. of Features Selected	Selected Descriptors	R^2	AAFE*
Random Forest	8	'ATS0m', 'nBase', 'nHBAcc', 'nHBDon', 'nAtomP', 'MLogP', 'MDEC-11', 'MPC2'	0.8972	0.1862
CART	10	'nAcid', 'nBase', 'C1SP1', 'SCH-3', 'nHBd', 'nHBDon', 'nAtomLC', 'nAtomLAC', 'nRing', 'LipinskiFailures'	0.8459	0.2502
XGBoost	8	'nBase', 'nHBAcc', 'nHBDon', 'nAtomP', 'MLogP', 'MDEC-11', 'MPC2', 'nRing'	0.9176	0.1343
ANN	8	'nAcid', 'nBase', 'CrippenLogP', 'nHBDon','MLFER_A', 'nRing', 'LipinskiFailures', 'TopoPSA'	0.4177	0.2182
*AAFE refers to Average Absolute Fold Error				

Table 2 shows that several descriptors recur across multiple machine learning models, highlighting their strong

Vol. 2026 (5) 428 – 437
ISSN 2688-3651 [online]

relevance to predicting skin permeability (log Kp). One of the most consistently selected descriptors is nHBDon, which appears in all four models. This reflects the well-established role of hydrogen bond donors in reducing skin permeability, as stronger intermolecular interactions with the skin barrier hinder passive diffusion. Similarly, nBase and nAcid, found in three models, indicate that the presence of ionizable groups influences molecular polarity and solubility, all factors that directly affect absorption.

Topological and shape-based descriptors like nRing also recur across models. nRing, which represents the number of rings in a molecule, helps capture molecular rigidity and compactness, both of which influence how easily a compound can navigate through the stratum corneum.

Additionally, the inclusion of lipophilicity descriptors such as CrippenLogP and MLogP underscores the importance of balancing hydrophobicity and hydrophilicity. Since the skin barrier is primarily lipid-based, compounds with moderate lipophilicity are more likely to partition into and diffuse through it. The repeated selection of these descriptors across diverse models suggests that hydrogen bonding, ionization, molecular shape, and lipophilicity are fundamental drivers of skin permeability. Their consistent appearance strengthens confidence in their predictive power and physical relevance.

Table 2. Explanation of molecular descriptors selected for each model.

Random Forest		**CART**	
ATS0m	Broto-Moreau autocorrelation - lag 0 / weighted by mass	nAcid	Number of acidic groups.
nBase	Number of basic groups.	nBase	Number of basic groups.
nHBAcc	Number of hydrogen bond acceptors	C1SP1	Triply bound carbon bound to one other carbon
nHBDon	Number of hydrogen bond donors	SCH-3	Simple chain, order 3
nAtomP	Number of atoms in the largest pi system	nHBd	Count of E-States for (strong) Hydrogen Bond donors
MLogP	Mannhold LogP	nHBDon	Number of hydrogen bond donors
MDEC-11	Molecular distance edge between all primary carbons	nAtomLC	Number of atoms in the largest chain
MPC2	Molecular path count of order 2	nAtomLAC	Number of atoms in the longest aliphatic chain
		nRing	Number of rings
		LipinskiFailures	Number failures of the Lipinski's Rule Of 5
XGBoost		**ANN**	
nBase	Number of basic groups.	nAcid	Number of acidic groups.
nHBAcc	Number of hydrogen bond acceptors	nBase	Number of basic groups.
nHBDon	Number of hydrogen bond donors	CrippenLogP	Crippen's LogP
nAtomP	Number of atoms in the largest pi system	nHBDon	Number of hydrogen bond donors
MLogP	Mannhold LogP	MLFER_A	Overall or summation solute hydrogen bond acidity
MDEC-11	Molecular distance edge between all primary carbons	nRing	Number of rings
MPC2	Molecular path count of order 2	LipinskiFailures	Number failures of the Lipinski's Rule Of 5
nRing	Number of rings	TopoPSA	Topological polar surface area

3.2 Comparison of the Different Machine Learning Models

The bar graph shown in Figure 1 exhibits the relative importance of each feature used by the model. Feature importance was calculated based on each variable's contribution to reducing the prediction error. Higher bars indicate greater influence on the model's output. nHBa, the quantity of hydrogen bond acceptor showed the most important contribution to the model for both XGBoost and random forest, while CART had nHBd as its primary descriptor. Both nHBa and nHBd describes the compound's ability for hydrogen bonding, which is closely related to polarity and hydrophilicity. This makes the two molecular descriptors very important, as the stratum corneum is mostly lipid,

Vol. 2026 (5) 428 – 437
ISSN 2688-3651 [online]

meaning it is highly influenced by octanol-water partition coefficient. Some of the other key features such as MLP, and nHBDn had also overlapped between models, while other features remained exclusive to the respective models.

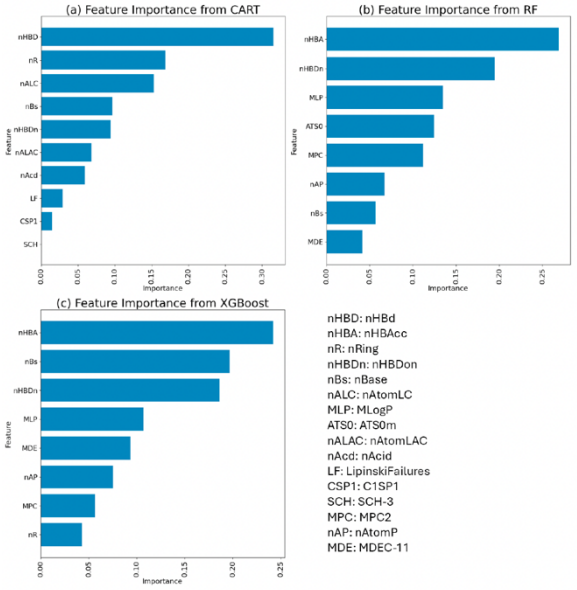

Figure 1. Feature importance from each ML model. ANN does not provide direct feature importance.

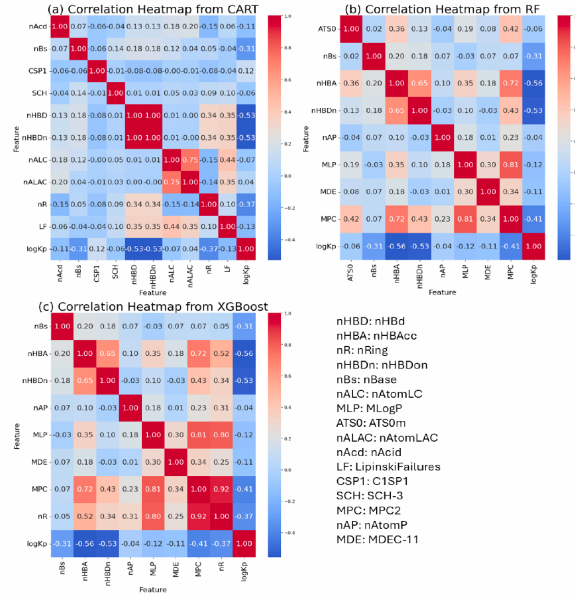

Figure 2. Correlation heatmap for each ML model. ANN does not provide correlation heatmap.

Figure 2 displays correlation heatmaps for descriptors selected by CART, Random Forest, and XGBoost, showing how each variable relates linearly to log Kp. Unlike feature importance plots, which rank variables by their overall contribution to predictive accuracy, these heatmaps show raw Pearson correlation coefficients. Descriptors such as nHBDon, nHBd, and nHBAcc consistently exhibit moderate negative correlations with log Kp, ranging from -0.53 to 0.56, reflecting their inhibitory effect on skin permeability due to increased polarity and hydrogen bonding. Other descriptors, including nBase, nRing, MPC2, and chain-length features like nAtomP or SCH-3, show weaker or mixed correlations, yet they still appear in feature importance rankings due to their complex interaction. This highlights that low linear correlation does not necessarily mean low predictive value, especially in complex models like Random Forest and XGBoost.

As shown in Figure 3, a density heatmap was also used to examine the distribution and relationship of predicted versus actual values across all models. In each case, the heatmap revealed a thin, diagonal structure with a slope close to 1, indicating that the majority of predictions closely matched the true values. This pattern reflects strong model performance, as the densest regions of the heatmap align with the ideal prediction line (where predicted = actual). The narrowness of the band suggests low variance and high consistency in predictions across the dataset. Additionally, the lack of significant spread or clustering away from the diagonal implies that the models generalized well and did not suffer from systematic over- or under-prediction. Thus, the density heatmap provides a visual confirmation of the accuracy and reliability of the machine learning models used in predicting skin permeability.

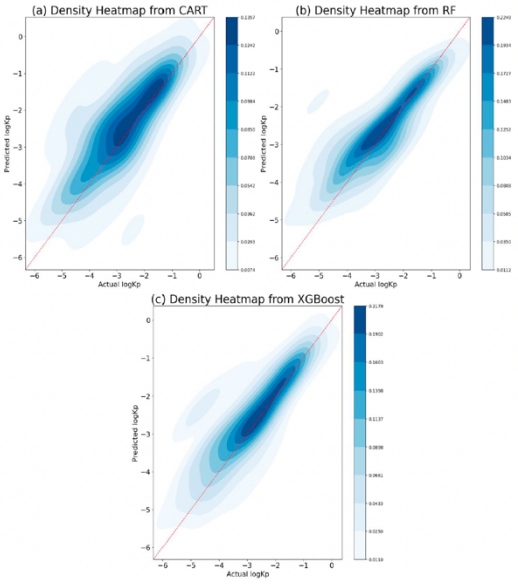

Figure 3. Density Heatmap from each ML model. ANN does not provide a density heatmap.

Vol. 2026 (5) 428 – 437
ISSN 2688-3651 [online]

4. Discussion

The machine learning models developed in this study demonstrated significant improvements in predictive accuracy for dermal permeability coefficient when compared with prior research. The final model achieved an R^2 of 0.9176 and an AAFE of 0.1343, indicating both strong correlation with experimental values and low predictive error as shown in Table 1.

The superior performance of this study's model can be attributed to several methodologies. First, the use of sequential feature selection allowed for the identification of the most relevant molecular descriptors from a large set. Descriptors such as hydrogen bond donors (nHBDon), acidic/basic groups (nAcid, nBase), ring structures (nRing), and polarity measures (TopoPSA, CrippenLogP) were frequently selected across different models, reinforcing their role in dermal permeation. These descriptors are chemically interpretable, showing model transparency, and explaining the high predictive accuracy. First, nHBDon influences a molecule's ability to form hydrogen bonds with the lipid layers of the stratum corneum. Molecules with strong hydrogen bonding may permeate the skin less effectively due to increased interaction with the aqueous environment of the skin. Other descriptors such as CrippenLogP and MLogP describe the lipophilicity of a molecule, a well-recognized predictor of dermal absorption. Higher logP values generally correlate with higher skin permeability up to a certain point. Excessively lipophilic compounds tend to remain in the lipid layers and not penetrate deeper. Out of all these descriptors, the most important ones were nHBAcc or nHBd, with its importance ratio close to 0.25-0.30. It is also shown that many of these descriptors have low correlation to each other, which would improve the model's accuracy by removing excess bias that the overlapping descriptors create.

Additionally, tree-based ensemble models such as XGBoost and Random Forest outperformed artificial neural networks (ANN), which is expected given the tabular structure and relatively small sample size of the dataset. These models are well suited for QSAR applications because they capture nonlinear, threshold-driven relationships between molecular descriptors and permeability such as changes associated with molecular weight, lipophilicity, and hydrogen bonding. Random Forest aggregates multiple decision trees to model general structure-permeability trends and reduce overfitting, whereas XGBoost further enhances performance by sequentially learning residual errors. In contrast, ANN performance was limited by the data-intensive nature of neural networks and their lack of inductive bias for small, structured tabular datasets, causing a higher chance of overfitting.

The results show that CART, XGBoost, and Random Forest achieved high predictive performance, with R2 values ranging from 0.84 to 0.92 and AAFE values between 0.13 and 0.25. These results demonstrate significantly greater accuracy compared to traditional statistical models, which reported an R2 of 0.62 on the same dataset (Brown et al., 2016). In contrast, the Artificial Neural Network (ANN) model performed notably worse, yielding a lower R^2 of 0.42 and a higher AAFE of 0.22, suggesting that neural networks were less effective than regression- and classification-based approaches in this context. Out of the treebased models, XGBoost alone achieved the highest R^2 value of 0.9176, proving its strength in capturing non-linear relationships without the problems of overfitting.

Compared to another study done with the same set of data, this is a significant improvement, as they had a R^2 value of 0.62 (Brown et al., 2016). The equation for the final model is shown by the following:

$$logKp = (0.1967 * nBase) + (0.2418 * nHBAcc) + (0.1862 * nHBDon) + (0.0755 * nAtomP) + (0.1070 * MLogP) + (0.0933 * MDEC - 11) + (0.0565 * MPC2) + (0.0429 * nRing) + Bias \quad (4)$$

Despite these promising results, several limitations in this study should be acknowledged. First, the dataset consisted of only 245 compounds, which is relatively small for machine learning applications and may limit the model's generalizability to other organic compounds that were not included in the training set. Second, the study relied exclusively on a single published dataset without external validation on independent test sets. However, the study that had published the dataset was done with the Environmental Protection Agency (EPA), which gives accountability and validity, which forgives some of the limitations. Finally, the compounds examined were restricted to organic molecules, potentially limiting the model's utility for inorganic or organometallic substances that may be relevant in the context of dermal exposure. Future research should address these limitations by expanding the dataset size and chemical diversity and exploring more advanced machine learning architectures.

Vol. 2026 (5) 428 – 437
ISSN 2688-3651 [online]

In summary, the combination of rigorous feature selection and multiple modeling has resulted in a high-performing and interpretable model for predicting dermal permeability coefficient. Future work could expand on this by increasing the size of the dataset, incorporating more molecular descriptors, such as the 3D descriptors, or exploring deep learning methods like graph neural networks to capture molecular structure more comprehensively. Nonetheless, the current model provides a promising, non-animal, *in silico* approach to estimate skin permeability, contributing to safer chemical design and regulatory risk assessment.

References

Akhtar, N., & Khan, R. A. (2016). Liposomal systems as viable drug delivery technology for skin cancer sites with an outlook on lipid-based delivery vehicles and diagnostic imaging inputs for skin conditions'. *Progress in Lipid Research*, 64, 192–230. https://doi.org/10.1016/j.plipres.2016.08.005

Badilli, U., et al. (2018). Lipid-based nanoparticles for dermal drug delivery. In A. M. Grumezescu (Ed.), Organic materials as smart nanocarriers for drug delivery (pp. 369–413). William Andrew Publishing. https://doi.org/10.1016/B978-0-12-813663-8.00009-9

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. https://doi.org/10.1145/2939672.2939785

Brown, T. N., et al. (2016). Dermal permeation data and models for the prioritization and screening-level exposure assessment of organic chemicals. *Environment International*, 94, 424–435. https://doi.org/10.1016/j.envint.2016.05.025

GeeksforGeeks. (2025). Sequential Feature Selection. https://www.geeksforgeeks.org/sequential-feature-selection/

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

Grisoni, F., Consonni, V., & Todeschini, R. (2018). Impact of Molecular Descriptors on Computational Models. In: Brown, J. (eds) Computational Chemogenomics. Methods in Molecular Biology, vol 1825. Humana Press, New York, NY. https://doi.org/10.1007/978-1-4939-8639-2_5

Ita, K. & Prinze, J. (2024). Machine learning for skin permeability prediction: random forest and XG boost regression. *Journal of Drug Targeting*, 32(1), 57–65. https://doi.org/10.1080/1061186X.2023.2284096

Ita, K., & Roshanaei, S. (2024). Artificial intelligence for skin permeability prediction: deep learning. *Journal of Drug Targeting*, 32(3), 334–346. https://doi.org/10.1080/1061186X.2024.2309574

Kavlakoglu, E. & Murel, J. (2024). What is Ensemble Learning?. International Business Machines Corporation. https://www.ibm.com/think/topics/ensemble-learning#:~:text=Ensemble%20learning%20is%20a%20machine,than%20a%20single%20model%20alone.

Kuster, C. J. et al. (2022). In silico prediction of dermal absorption from non-dietary exposure to plant protection products, *Computational Toxicology*, 24, https://doi.org/10.1016/j.comtox.2022.100242

Lebailly P., et al. (2009). Exposure to pesticides in open-field farming in France. *Ann Occup Hyg*. 2009(53), 69–81. doi: 10.1093/annhyg/men072.

Leo Breiman. (2001). Random Forests. Mach. Learn, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324 Leo Breiman. (2017). Classification and Regression Trees. Chapman and Hall/CRC.

Macfarlane, E., et al. (2013). Dermal exposure associated with occupational end use of pesticides and the role of protective measures. *Safety and health at work*, 4(3), 136–141. https://doi.org/10.1016/j.shaw.2013.07.004

Vol. 2026 (5) 428 – 437
ISSN 2688-3651 [online]

Potts, R. O., & Guy, R. H. (1992). Predicting skin permeability. *Pharmaceutical research*, 9(5), 663–669. https://doi.org/10.1023/a:1015810312465

U.S. Food and Drug Administration. (2024). New Approach Methods (NAMs). FDA