2008 Eighth IEEE International Conference on Data Mining

# Prediction of Skin Penetration Using Machine Learning Methods

Yi Sun[1*], Gary P. Moss[2*], Maria Prapopoulou[2*], Rod Adams[1*],
Marc B. Brown[2*], Neil Davey[1*]
1* Science and Technology Research School, University of Hertfordshire,
United Kingdom, AL10 9AB
{comrys, r.g.adams, n.davey}@herts.ac.uk
2* School of Pharmacy, University of Hertfordshire,
United Kingdom, AL10 9AB
maria.2.prapopoulou@kcl.ac.uk, {g.p.j.moss, m.7.brown}@herts.ac.uk

## Abstract

*Improving predictions of the skin permeability coefficient is a difficult problem. It is also an important issue with the increasing use of skin patches as a means of drug delivery. In this work, we apply $K$-nearest-neighbour regression, single layer networks, mixture of experts and Gaussian processes to predict the permeability coefficient. We obtain a considerable improvement over the quantitative structure-activity relationship (QSARs) predictors. We show that using five features, which are molecular weight, solubility parameter, lipophilicity, the number of hydrogen bonding acceptor and donor groups, can produce better predictions than the one using only lipophilicity and the molecular weight. The Gaussian process regression with five compound features gives the best performance in this work.*

## 1. Introduction

In this paper, we address the problem of predicting the rate at which various compounds penetrate human skin. This is an important issue with the increasing use of skin patches as a means of drug delivery. There are two main approaches used to predict and understand the skin penetration procedure. One is quantitative structure-activity relationships (QSARs), and the other is mathematical modelling [4]. Recently more new approaches, for example, artificial neural network and fuzzy modelling, have been applied to this domain [3].

One problem addressed here is how to improve predictions of the *skin permeability coefficient* by applying advanced machine learning techniques, for example, *Gaussian Processes*. To the best of our knowledge this is the first time that Gaussian processes modelling has been applied to skin permeability data. One key feature of this problem domain is that the *target*, skin permeability coefficient, may have a strongly non-linear relationship with the compound descriptors (features). We demonstrate a feasibility of prediction improvement by using computational regression modelling methods.

Another issue discussed in this work is to investigate new compound descriptors to the problem. We show that involving three new descriptors, which are *solubility parameter*, the number of *hydrogen bonding acceptor* and *donor* groups, simply gives a more detailed description of the molecule in relation to the skin.

## 2 Problem Domain

Percutaneous absorption has been a significant challenge for pharmaceutical scientists for the last $50$ years. As recently as the $1960'$s, it was assumed that drugs could not pass through the skin. However, as knowledge of the detailed structure of the skin barrier - the *stratum corneum*, the skin's outermost layer - increased, new technologies gradually became available for the treatment of medical conditions by transdermal therapy. The stratum corneum is the main barrier to percutaneous absorption, and this is due to its structure and properties. It is a very thin layer, commonly $15 - 30\mu$m on the volar forearm, for example. This layer effectively governs the rate of passage of exogenous chemicals across the skin and into the viable tissues from the external environment. It is a densely packed layer consisting of dead, flattened keratin cells enmeshed in a lipid domain.

While qualitative estimates of percutaneous absorption were common until the $1980'$s, it was not until $1990$, and the publication of the *Flynn dataset* [5] that a quantitative approach to skin absorption was proposed. Flynn determined,

1550-4786/08 $25.00 © 2008 IEEE
DOI 10.1109/ICDM.2008.97

in a semi-quantitative manner, that skin absorption was influenced predominately by two compound descriptors - the *lipophilicity* of a molecule and its *molecular weight* ($MW$). The former descriptor defines the ability of a molecule to partition between lipid and aqueous layers, represented by *n*-octanol and water (or buffer), respectively. This is then defined as $\log P$, the logarithm of the ratio of concentrations in, respectively, the lipid and aqueous domains. Molecular weight is effectively a measure of the size of the molecule, and is related to the lipophilicity in a generalised sense that, when $MW$ is increased it is most likely due to the addition of lipophilic groups onto the molecule.

Potts and Guy [14] used the Flynn dataset to derive a linear equation that quantified percutaneous absorption:

$$\log K_p = 0.71 \log P - 0.0061\text{MW} - 6.3 \ , \qquad (1)$$

where $\log K_p$ is the permeability coefficient, a concentration-corrected measure of drug transport, or flux, across a membrane (such as the skin), $\log P$ is the lipophilicity and $MW$ the molecular weight.

In essence, Flynn used the data that was available to him from the literature. This means that there is an uneven distribution of this data across the whole range of interest, and that the data is generally skewed by certain compound descriptors, such as $\log P$ and $MW$.

Moss and Cronin [8] developed the Potts and Guy model by evaluating the role of steroids in the dataset. Thus, the model is based on a slightly larger and more robust dataset. It is represented by the following equation:

$$\log K_p(cm/s) = 0.74 \log P - 0.0091 MW - 2.39 \ , \quad (2)$$

where $\log K_p$, $\log P$ and MW are as defined earlier.

A number of similar equations have been derived since the publication of Potts and Guy's model. Moss et al [10] compared a series of published models. They showed that there were significant differences between $\log K_p$ values that were measured experimentally and those that were determined using the Potts and Guy (and other) equations. Interestingly, they showed that the greatest difference between experimental and predicted values was found at high $\log P$ values. In effect, the absorption is Gaussian in its distribution, if assessed by $\log P$ in particular. This contradicts the linear nature of the Potts and Guy (and similar) equations, although the concomitant increase in $\log P$ and $MW$ is often offset by the negative sign in front of $MW$.

Therefore, there is compelling qualitative evidence that suggests the non-linear response to skin absorption that is missing from the linear equations, such as Potts and Guy. Non-linear modelling might provide a prediction for percutaneous absorption for molecules with a wide range of $\log P$ values.

In [2], an artificial neural network was developed, where $MW$, $\log P$ and *partial change* of the penetrant molecule were used as inputs, and was applied to skin permeability data for the first time. In [11], an ensemble model using $K$-nearest-neighbour and ridge regression to predict skin permeability coefficients was proposed, where three computational descriptors, which are $MW$, $\log P$ and solvation free energy, were used as inputs. Note that both partial change of the penetrant molecule and solvation free energy are useful descriptors, but they are difficult to be calculated without specialist software.

In previous work [13], Gaussian processes (GP) modelling has been used to predict $\log P$. To the best of our knowledge, the GP methods have not been applied to skin permeability data. In this work, we investigate the non-linear response to skin absorption by applying computational regression modelling methods: $K$-nearest-neighbour regression, single layer networks (SLN), mixture of experts and Gaussian processes.

## 3 A Description of the Data

The dataset employed in this study has been collated with reference to a range of literature sources. It predominately consists of the Flynn dataset, used by Potts and Guy, and others. It contains several additions, whose origins are described in Moss et al. [9]. The whole dataset consists of $149$ compounds. Usually, $\log P$ and $MW$ appear to be the only significant features in QSAR forms. However, in some cases (such as [12]) other features achieve significance; these features are often calculated using expensive specialised software. Since they often provide only marginal improvements in the prediction of $\log K_p$ compared to other QSAR models, there is little application of them in the field. In this work, $5$ molecular features in total are involved. They are *molecular weight* ($MW$), *solubility parameter* ($SP$), $\log P$ (often described, for example by Potts and Guy, as $\log P_{\text{known}}$), counts of the number of hydrogen bonding acceptor ($HA$) and donor groups ($HD$), respectively, that can be found on a molecule.

$SP$ relates the solubility of a penetrant to the solubility of the stratum corneum, the skin barrier. If a molecule is too soluble in the stratum corneum it might not be soluble in the rest of the skin, and it might therefore have poor skin penetration overall. So, the use of $SP$ will provide an indication of the mostly lipophilic molecules likely to exhibit this behaviour. $HA$ and $HD$ are the hydrogen-bonding terms. Hydrogen bonds (H-bonds) are specific covalent bonds between a hydrogen on one molecule and a highly electronegative atom on a different molecule. $HA$ is a H-bond acceptor, such as oxygen, nitrogen or fluorine. $HD$ is a H-bond donor, normally a hydrogen which is itself attached to a highly electronegative atom. $HA$ and $HD$ are associated with electronegativity and H-bonding. This is a polar phenomenon and as such might be associated with retarding

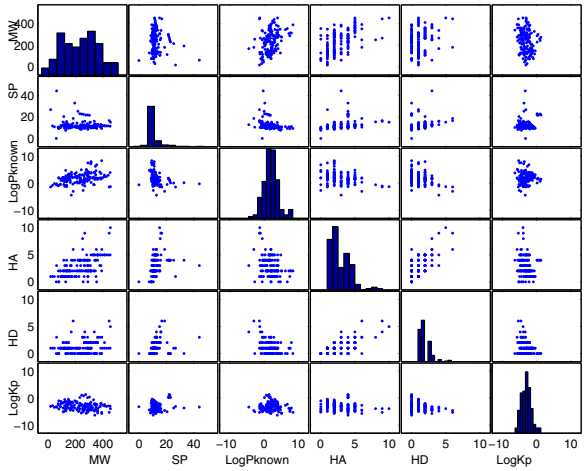


**Figure 1. A scatter plot matrix of the skin dataset. The diagonal shows the shape of the distribution of each feature. The graphs in the lower triangle are the transpose of the graphs in the upper triangle.**

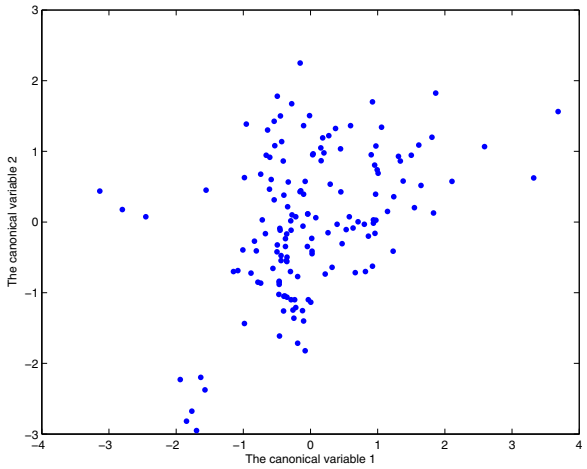


**Figure 2. The canonical correlation between 5 compound descriptors and $\log K_p$.**

percutaneous absorption.

There is very little work which uses the five descriptors we employ in this work. We are interested in looking at new compound descriptors. The descriptors above are chosen not only for their applicability but for the ease with which they can be determined.

### 3.1 Visualisation of the skin data

The scatter plot matrix in Figure 1 shows data for all 149 compounds with 5 features against each other, and the diagonal shows the shape of the distribution of each feature. The subplot appearing in the first row and last column shows $MW$ against $\log K_p$. It suggests that very similar $\log K_p$ values can correspond to many different $MW$ values. This is also true of $\log P_{known}$ and $\log K_p$. It can also be seen that there is no simple linear replationship between any pair of descriptors.

### 3.2 Canonical correlation analysis

Canonical correlation analysis (CCA) [6] can be used to find a projection that maximises the correlation between two sets of variables.

In this work, we group $MW$, $SP$, $\log P$, $HA$ and $HD$ into one set, and $\log K_p$ into another set. Our aim is to investigate the correlating linear relationship between $\log K_p$ and the 5 compound descriptors. The canonical variable 1 (*CV1*) in Figure 2 is a combination of 5 descriptors used in this work: $CV1 = 0.002MW - 0.116SP + 0.033 \log P + 0.107HA + 0.6655HD$, while the canonical variable 2 (*CV2*) in Figure 2 is given by $CV2 = -0.686 \log K_p$. As one can see that there is no linear relationship between the two sets of variables. It is interesting to note that in *CV1* the least important features (those with lowest coeffecients) are $MW$ and $\log P$. Actually, the *canonical correlation coefficient* is approximately 0.415, while the canonical correlation coefficient between $\log K_p$ and a group of two variables, $MW$ and $\log P$ is about 0.238.

## 4 Modelling Methods

### 4.1 Single layer network

First of all, we consider a single layer network (SLN) on the regression problem, which is a simple linear regression. The output $y$ is the weighted sum of the components of an input $\mathbf{x}$, The weights are set so that the sum squared error function is minimised on a training set.

### 4.2 K-nearest-neighbour regression

Given a test input $\mathbf{x}$, the algorithm finds $K$ closest points to $\mathbf{x}$ among all the training inputs. The prediction of the model is therefore the average of those $K$ target values. In our experiments, we apply Euclidean distance in the $K$-nearest-neighbour (KNN) regression to do the distance measurement.

### 4.3 Mixture of experts

The mixture of experts [7] divides the input space into a nested set of regions. In each region a simple surface is

fitted to the data. It consists of a gating network and experts. The function of the gating network is to partition the input space so that each expert only needs to model a small region. The gating network receives the input $\mathbf{x}$, and outputs a scalar value $p_i$ with the property that $p_i \geq 0$ and $\sum_i p_i = 1$. The final prediction of the model is a sum of the expert predictions weighted by $p_i$. In this work, all local experts are linear regression models.

### 4.4 Gaussian process regression - GPR

*Gaussian process* (GP) modelling is a non-parametric method, which does not produce an explicit functional representation of the data. Here it is assumed that the underlying function, $f(\mathbf{x})$, that produces the data will remain unknown, but that the data is produced from a (infinite) set of functions, with a Gaussian distribution in the function space.

A Gaussian process is completely characterised by its mean and covariance function. For simplicity, we usually consider the mean function to be the zero everywhere function. The covariance function, $k(\mathbf{x}_i, \mathbf{x}_j)$, is crucial to GP modelling. It expresses the expected correlation between values $f(\mathbf{x})$ at the two points $\mathbf{x}_i, \mathbf{x}_j$. In other words, it defines nearness or similarity between data points.

In this work, we apply the squared exponential covariance function, which incorporates noise into the model, as follows:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \sigma_f^2 \exp\left(-\frac{1}{2}(\mathbf{x}_i - \mathbf{x}_j)^T M (\mathbf{x}_i - \mathbf{x}_j)\right) + \sigma_n^2 \delta_{ij} \ , \tag{3}$$

where $M = l^{-2} I$, $l$ is *characteristic length-scale*, $\sigma_f$ is *signal variance*, $\sigma_n$ is *noise variance*, and $\delta_{ij}$ is a Kronecker delta which is one iff $i = j$ and zero otherwise.

To make a prediction $f(\mathbf{x}_*)$ at a new input $\mathbf{x}_*$, we need to compute the conditional distribution $p(f(\mathbf{x}_*)|y_1, \ldots, y_{N_{trn}})$ on the observed vector $[y_1, \ldots, y_{N_{trn}}]$. Since our model is a Gaussian process, this distribution is also a Gaussian and is completely defined by its mean and variance. The mean at $\mathbf{x}_*$ is given by

$$E[f(\mathbf{x}_*)] = \mathbf{k}_*^T (\mathbf{K} + \sigma_n^2 \mathbf{I})^{-1} \mathbf{y} \ . \tag{4}$$

In eq(4), $\mathbf{k}_*$ denotes the vector of covariances between the test point and the $N_{trn}$ training data; $\mathbf{K}$ denotes the covariance matrix of the training data; $\sigma_n^2$ denotes the variance of an independent identically distributed Gaussian noise, which means observations are noisy; and $\mathbf{y}$ denotes the vector of training targets.

The variance at $\mathbf{x}_*$ is given by

$$var[f(\mathbf{x}_*)] = k(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{k}_*^T (\mathbf{K} + \sigma_n^2 \mathbf{I})^{-1} \mathbf{k}_*, \tag{5}$$

where $k(\mathbf{x}_*, \mathbf{x}_*)$ denotes the variance of $y_*$.

We use the mean as our prediction and the variance as error bars on the prediction.

## 5 Performance Measures

We apply three common measurements for assessing regression performance, which are *normalised mean squared error* (NMSE), *percent improvement over a naive model* (ION), and *negative log loss* (NLL).

The *mean squared error* (MSE) measures the average squared difference between model predictions and the corresponding targets . Here we report the NMSE which is to normalise MSE by the variance of target values.

In the naive model for any input the prediction is always the same value, namely the mean of $\log K_p$ in the training set. Denote mean squared error of a naive model as $MSE_{naive}$. The degree of improvement of the model over the *Naive* predictor can be quantified by the ION measure [13], that is $ION = \frac{MSE_{naive} - MSE}{MSE_{naive}} \times 100\%$.

Finally, when we investigate GP's results, we also consider the average *negative log estimated predictive density NLL*, given by $NLL = \frac{1}{N} \sum_n -\log p(y_n|\mathbf{x}_n)$, where $N$ is the number of test cases, and $-\log p(y_*|\mathbf{x}_*) = \frac{1}{2}\log(2\pi\sigma_*^2) + \frac{(y_* - E[f(\mathbf{x}_*)])^2}{2\sigma_*^2}$, in which case $\sigma_*^2$ is the predictive variance, and $y_*$ is the target value at $\mathbf{x}_*$. A small value shows good performance.

When we analyse the results, we want a model which on the test set provides low values of both NMSE and NLL and high values of both ION and the correlation coefficient (CORR).

## 6 Experiments

### 6.1 Simulation setup

In this work, we randomly divided the whole dataset into a training set including 130 compounds, and an independent test set consisting of the remaining 19 compounds. We then applied those modelling methods described in Section 4 on the training set and do prediction on the independent test set using the trained models. We repeated this whole procedure 10 times, each time for a different randomly assigned training and test sets. To investigate whether predictions can be improved by involving all 5 features rather than the original 2 features used in the QSAR forms, ($MW$ and $\log P$), we employed regression modelling methods with both 2 and 5 compound features.

In $K$-nearest-neighbour modelling, we varied the number of neighbours, $K$, between 1 and 10; in the mixture of experts, we set the number of experts between 2 and 5. In Gaussian process modelling, we initialised the logarithm of *characteristic length-scale*, the logarithm of *signal variance*, and *noise variance* from the set $[\log(1.0)\ \log(1.0)\ \log(0.1); \log(2.0)\ \log(1.0)\ \log(0.1);$ $\log(1.0)\ \log(2.5)\ \log(0.3);\ \log(0.7)\ \log(2.5)\ \log(0.3);$

$\log(0.3)$ $\log(1.0)$ $\log(0.1)$; $\log(1.0)$ $\log(1.2)$ $\log(0.9)$; $\log(2.0)$ $\log(1.2)$ $\log(0.9)$; $\log(0.3)$ $\log(1.2)$ $\log(0.9)$; $\log(1.0)$ $\log(0.6)$ $\log(0.01)$; $\log(0.7)$ $\log(0.6)$ $\log(0.01)$].

We used a 5-fold cross-validation procedure to select optimal parameters for each of $K$-nearest-neighbour, the mixture of experts, and Gaussian process. In these cases, each training set is further divided into training and validation sets.

We applied Rasmussen and Williams's [15] GP toolbox to do Gaussian process modelling; and employed the Bayes Net Toolbox, which is publicly available at http://www.cs.ubc.ca/∼murphyk/Software/BNT/bnt.html#ack, to do the mixture of experts modelling.

### 6.2 Experimental results

Before we do experiments using the trainable regression models, we first apply the QSAR forms (eqs. (1) & (2)) introduced in section 2 to the whole dataset. Table 1 shows the results. The results are the averages on the 10 independent test sets. For comparison, Table 1 also shows results from the Naive model. In general, all QSAR predictions are worse than naive predictions, especially with Potts' QSAR form.

Table 2 shows results obtained using computational modelling methods from the machine learning field with $MW$ and $\log P$ as descriptors. One can see that all 4 methods have improved the naive predictions, with $K$-nearest-neighbour giving the best results. The average of the optimal number of neighbours, $K$, was equal to 8.4. Not surprisingly, the single layer network, which is a simple linear regression model, performed worst.

Results obtained with 5 compound descriptors are shown in Table 3. Comparing with Table 2, one can see that all 4 regression modelling methods have improved their performance when using 5 features rather than 2. This big improvement in performance confirms that the 3 additional features are very important in predicting skin permeability.

The Gaussian process regression gives the best performance, Figure 3 displays a box plot of normalised mean squared errors from 10 independent test sets on Naive model, the Moss QSAR form, and those 4 computational modelling methods with 5 features. It shows that the Gaussian process regression with 5 features (GPRf5) gives the lowest upper quartile, median and lower quartile values on NMSE. Although the mixture of experts with 5 features (MIXEXPf5) has comparable low median and lower quartile values, its upper quartile value and the largest NMSE value are much bigger than those obtained from GPRf5. It suggests that GPRf5 has a relatively stable and robust performance. On the other hand, one can see the QSAR form (Moss) has the highest lower quartile, median and upper quartile values. Both $K$-nearest-neighbour with 5 features

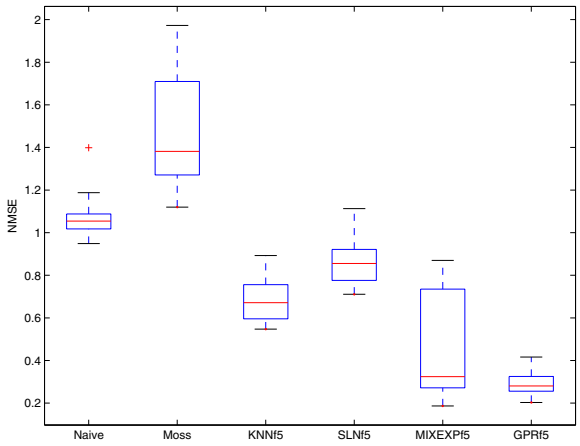


**Figure 3. Box plot of normalised mean squared errors from 10 independent test sets on 6 different models.**

(KNNf5) and single layer network with 5 features (SLNf5) were relatively stable, but in general not as good as GPRf5 and MIXEXPf5.

## 7 Conclusions

There are two main quantitative conclusions from this study. First, we get a considerable improvement over the linear QSAR predictor by using non-linear regression methods. In fact, the QSAR prediction even gives worse prediction than the naive model. Secondly, the trainable regression methods with five features have much better results than those obtained using only two features. We conclude that the Gaussian processes model with five compound descriptors has the best performance on the NMSE, ION and CORR measures.

Involving the three new descriptors, $SP$, $HA$ and $HD$, simply gives a more detailed description of the molecule in relation to the skin. Quantifying a key molecular descriptor might provide more information for the model, and it might allow a more detailed estimation of the key properties of a penetrant. There are lots of other descriptors we could have looked at, and which we should look at in future work. However, the descriptors used in this work are chosen not only for their applicability but for the ease with which they can be determined. This means that any method which uses these descriptors is not only significant but can be easily used by other researchers.

## References

[1] C. M. Bishop. (1995) *Neural Networks for Pattern Recognition*. Oxford University Press, New York.

**Table 1. The results on test sets using different QSAR models.**

| Models | NMSE | ION (%) | CORR |
|---|---|---|---|
| Naive | $\mathbf{1}.08 \pm \mathbf{0}.13$ | $\mathbf{0}$ | - |
| Moss | $1.46 \pm 0.28$ | $-34.71 \pm 18.45$ | $0.21 \pm 0.21$ |
| Potts | $5.75 \pm 1.14$ | $-430.33 \pm 74.38$ | $0.18 \pm 0.22$ |

**Table 2. The results on test sets using different machine learning methods with $2$ features.**

| Models | NMSE | ION (%) | CORR | NLL |
|---|---|---|---|---|
| Naive | $1.08 \pm 0.13$ | 0 | - | - |
| KNN | $\mathbf{0}.89 \pm \mathbf{0}.14$ | $\mathbf{1}7.81 \pm \mathbf{5}.89$ | $\mathbf{0}.44 \pm \mathbf{0}.15$ | - |
| SLN | $1.07 \pm 0.17$ | $1.54 \pm 4.11$ | $0.21 \pm 0.16$ | - |
| MIXEXP | $1.03 \pm 0.14$ | $4.76 \pm 6.76$ | $0.28 \pm 0.12$ | - |
| GPR | $0.98 \pm 0.11$ | $9.85 \pm 5.92$ | $0.32 \pm 0.13$ | $3.06 \pm 0.48$ |

**Table 3. The results on test sets using different machine learning methods with $5$ features.**

| Models | NMSE | ION (%) | CORR | NLL |
|---|---|---|---|---|
| KNN | $0.69 \pm 0.11$ | $36.18 \pm 6.15$ | $0.67 \pm 0.08$ | - |
| SLN | $0.87 \pm 0.12$ | $20.12 \pm 5.37$ | $0.50 \pm 0.12$ | - |
| MIXEXP | $0.44 \pm 0.24$ | $59.28 \pm 22.03$ | $0.81 \pm 0.10$ | - |
| GPR | $\mathbf{0}.30 \pm \mathbf{0}.07$ | $\mathbf{7}2.62 \pm \mathbf{5}.03$ | $\mathbf{0}.86 \pm \mathbf{0}.05$ | $\mathbf{1}.48 \pm \mathbf{0}.13$ |


[2] I. Tuncer Degim, J. Hadgraft, S. Ilbasmis, Y. Ozkan. (2003) Prediction of skin penetration using artificial neural network (ANN) modeling, *Journal of Pharmaceutical Sciences*. Vol 92, NO. 3.

[3] I. Tuncer Degim. (2006) New tools and approaches for predicting skin permeability, *Drug Discovery Today*. Vol 11, pp.517-523.

[4] D. Fitzpatrick, J. Corish and B. Hayes. (2004) Modelling skin permeability in risk assessment - the future, *Chemosphere*. Vol 55, pp.1309-1314.

[5] G. L. Flynn. (1990) Physicochemical determinants of skin absorption. In *Principles of Route-to-Route Extrapolation for Risk Assessment*, T. R. Gerrity and C. J. Henry (eds.), Elseviet, New Kork, 1990, pp.93-127.

[6] H. Hotelling (1936) Relations between two sets of variates. *Biometrika*, Vol 28 pp.312-377.

[7] R. Jacobs, M. I. Jordan, S. J. Nowlan, and G. E. Hinton (1991) Adaptive mixtures of local experts. Neural Computation, 3, 79-87.

[8] G. P. Moss and M. T. D. Cronin. (2002) Quantitative structure-permeability relationships (QSPRs) for percutaneous absorption: re-analysis of steroid data. *International Journal of Pharmaceutics*, Vol. 238, pp.105-109.

[9] G. P. Moss, J. C. Dearden, H. Patel, and M. T. D. Cronin. (2002) Quantitative structure-permeability relationships (QSPRs) for percutaneous absorption. *Toxicology in Vitro*, Vol. 16, pp299-317.

[10] G. P. Moss, D. R. Gullick, P. A. Cox, C. Alexander, M. J. Ingram, J. D. Smart and W. J. Pugh. (2006) Design, synthesis and characterization of captopril produgs for enhanced percutaneous absorption. *Journal of Pharmacy and Pharmacology*, Vol 58, pp.167-177.

[11] D. Neumann, O. Kohlbacher, C. Merkwirth, and T. Lengauer (2006) A fully computational model for predicting percutaneous drug absorption, *J. Chem. Inf. Model.* 46, pp.424-429.

[12] H. Patel, W. ten Berge, M. T. D. Cronin. (2002) Quantitative structure-activity relationships (QSARs) for the prediction of skin permeation of exogenous chemicals. *Chemosphere*, Vol 48, pp.603-613.

[13] P. Tino, I. Nabney, B. S. Williams, J. Losel, Y. Sun (2004) Non-linear Prediction of Quantitative Structure-Activity Relationships. *Journal of Chemical Information and Computer Sciences*, 44(5), pp. 1647-1653. (c) ACM

[14] R. O. Potts and R. H. Guy (1992) Predicting skin permeability, *Pharm. Res.* vol(12), 663-669.

[15] C. E. Rasmussen and C. K. I. Williams (2006) *Gaussian Processes for Machine Learning*. The MIT Press.