Comparing Multi-Label Classification Methods for Provisional Biopharmaceutics Class Prediction

Danielle Newby[a], **Alex. A. Freitas**[b], **Taravat Ghafourian**[a,c*]

[a]*Medway School of Pharmacy, Universities of Kent and Greenwich, Chatham, Kent, ME4 4TB, UK*

[b]*School of Computing, University of Kent, Canterbury, Kent, CT2 7NF, UK*

[c]*Drug Applied Research Centre and Faculty of Pharmacy, Tabriz University of Medical Sciences, Tabriz, Iran*

*** Corresponding Author,** Email: T.ghafourian@kent.ac.uk; Tel +44(0)1634 202952; Fax +44 (0)1634 883927

Abstract

The biopharmaceutical classification system (BCS) is now well established and utilised for the development and biowaivers of immediate oral dosage forms. The prediction of BCS class can be carried out using multi-label classification. Unlike single label classification, multi-label classification methods predict more than one class labels at the same time. This paper compares two multi-label methods, binary relevance and classifier chain, for provisional BCS class prediction. Large datasets of permeability and solubility of drug and drug-like compounds were obtained from the literature and were used to build models using decision trees. The separate permeability and solubility models were validated and an BCS validation set of 127 compounds where both permeability and solubility were known was used to compare the two aforementioned multi-label classification methods for provisional BCS class prediction. Overall, the results indicate that the classifier chain method, which takes into account label interactions, performed better compared to the binary relevance method. This work offers a comparison of multi-label methods and shows the potential of the classifier chain multi-label method for improved biological property predictions for use in drug discovery and development.

Keywords:

Multi label, BCS, classification, permeability, solubility, Oral absorption, *in silico*

1. Introduction

Oral absorption is dependent on many physiological, physicochemical and formulation effects. Two of these physicochemical main factors are permeability and solubility, which are considered the main fundamental properties that govern the rate and extent of oral absorption. The importance of these two properties has been emphasised in the Biopharmaceutics Classification System (BCS)[1]. The BCS system was developed to classify drugs into one of four classes based on solubility or dissolution properties and intestinal permeability (**Figure 1**). The BCS has been adopted by many regulatory authorities as a standard for the justification of biowaivers for costly bioequivalence studies. Compounds that are eligible for biowaviers under the BCS are immediate release dosage forms with high permeability and high solubility (BCS class 1) and are experimentally shown to exhibit rapid dissolution. In addition, the EMA (2010)[2] has extended the eligibility of biowaviers to include certain class 3 compounds. Therefore the BCS is shown to be a vital cost effective tool of compound development [1, 3].

	High permeability	Low permeability
High solubility	Class I	Class III
Low solubility	Class II	Class IV

Figure 1. The Biopharmaceutics Classification System (BCS)

In drug discovery the characterization of preliminary BCS classification is of great interest. The use of a provisional BCS class prediction can help guide decision making and formulation of compound development strategies[4-9]. In addition, it has been observed that knowledge of the different BCS classes can give an indication of the rate limiting steps of absorption as well as potential metabolic routes and transporters interactions[8, 10].

There are many classification models in the literature that predict oral absorption, solubility or permeability classes in separate models[11-13]. These classification models predict just one property and assign a compound to one class label out of two or more mutually exclusive class labels, for example high or low absorption. This is single label classification. The problem with this is that in a real life scenario most objects belong to more than one class at the same time. For example a drug molecule can be highly absorbed but can also have high

solubility or low solubility. The prediction of multiple class labels at the same time is termed multi-label classification[14-16]. Due to the relationship between solubility and permeability with oral absorption, a potential multi-label problem exists.

Early research into multi-label modelling has focussed on text categorization[17, 18] and now this type of method has expanded into being utilised in many different fields such as gene function prediction[19], medical diagnosis[20], and drug discovery[21]. There are two main types of multi-label methods; problem transformation and algorithm adaption methods. Problem transformation methods involve transformation of the multi-label data into single label data to then carry out conventional single label classification. Therefore problem transformation methods can also be termed algorithm independent methods and be used with any single label classification method. Algorithm adaption methods involve the adaption of original single-label algorithms to deal with multi-label data directly[14-16].

Problem transformation is a more common route for dealing with multi-label data. There are several different strategies in order to transform multi-label data into single label data for analysis. A common approach is the binary relevance method (**Figure 2**). This is where each class label, or property, is separately predicted. The results are then combined to give the results for the multi-label problem. In relation to the BCS prediction, solubility and permeability are predicted separately then the predicted BCS is assigned based on the combined permeability and solubility prediction based on the two separate labels. This method is simple and any single label classification algorithm can be used. A benefit of this method is that the numbers of compounds in the datasets do not need to be identical as the properties are modelled separately; therefore all available data is used. However one important drawback of this method is that it fails to take into account label interactions[14-16].

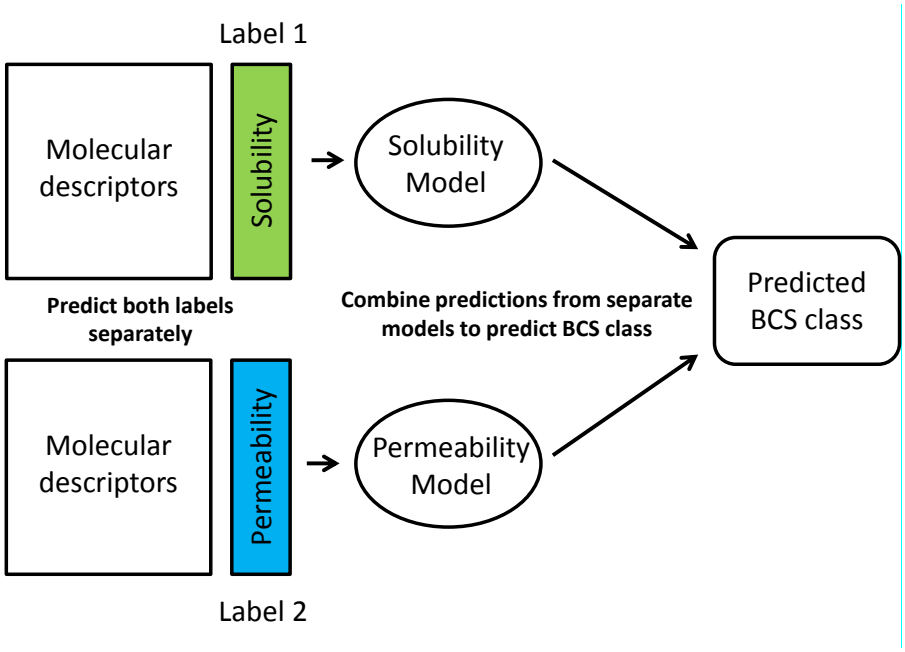

Figure 2 How the binary relevance problem transformation method works for BCS prediction

An example of binary relevance multi-label method utilised in the literature for BCS classification is by Pham-The and co-workers[6]. Although the multi-label method term binary relevance is not mentioned in this study it built separate models for the *in silico* prediction of solubility or Caco-2 cell permeability. The results from the models were then combined to give a provisional BCS prediction[6]. A similar study predicts solubility and rate of metabolism separately to predict biopharmaceutical drug disposition classification class (BDDCS)[10] using the combined predictions[22].

Another typical multi-label method in the problem transformation category is called label power set. This is where the two labels to be predicted are converted into a single label by combining the labels[14]. In the context of BCS, this method is basically the prediction of BCS classes directly. Therefore rather than a prediction of solubility and permeability a BCS class is predicted. The only relevant examples in the literature predict BDDCS class[10, 23], instead of predicting BCS class. In one example the prediction of BDDCS class was carried out using recursive partitioning (building a single decision tree), random forest (building a set of decision trees) and support vector machine[23]. Although this method takes into account interactions between labels, the main problem with this method is the lack of representation of some of the classes. In other words some classes may have fewer examples compared to the rest and leads to a poor prediction accuracy for that underrepresented class[22]. In addition,

models can only be built when both labels are known, therefore not utilising all of the data available. Therefore, for this work this method was not utilised due to the drastic reduction of data available for modelling. Note that it is also possible to predict continuous values of permeability and solubility, or another approach would be to classify compounds into multiple categories (low, medium, high)[24]. However these approaches are out of the scope of this current work since we are engaged in binary classification of chemicals according to the BCS system.

A less well known multi-label method is classifier chain[16]. This method seeks to overcome the drawbacks of binary classifier by taking into account label interactions. The method works by firstly predicting one label, then using the predicted label, along with any other predictors (molecular descriptors) models are built, in order to predict the second label (**Figure 3**). Then the predictions from both labels are combined like binary relevance for the final BCS prediction. A potential issue with this method could be the noisy data created from using the predicted value of the first label as a descriptor to predict the second label.

One of the problems of this method is deciding which label to predict first[25]. In some cases there may be a definite order of the labels from a mechanistic point of view, making this choice obvious. For example, in the case of solubility and permeability prediction, solubility would be the first label and permeability would be the second. This is because solubility is a basic property that can affect permeability of molecules, whereas permeability is a higher level property. Molecules need to be dissolved and solubilised first, before they can permeate the intestinal wall.

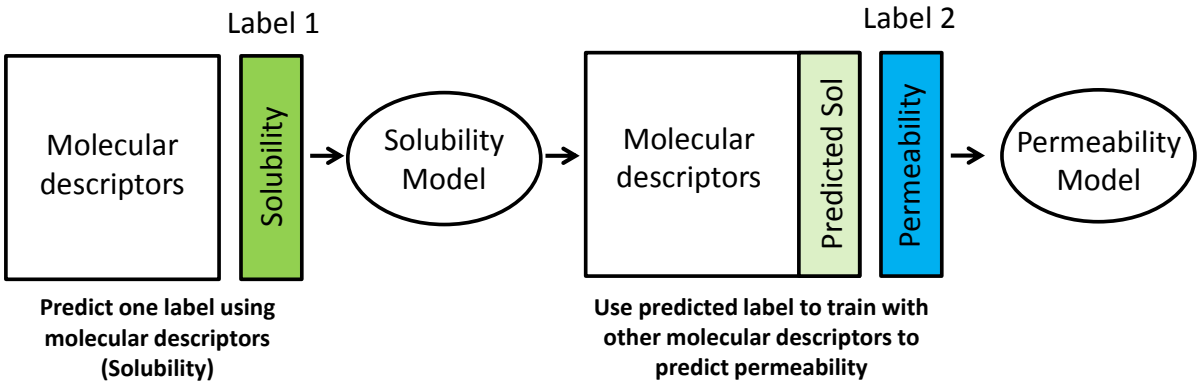

Figure 3 Prediction of BCS using the classifier chain multi-label method

Both binary relevance and classifier chain also require an extra step to convert the single labels into a final label result (BCS class assignment). Both have the benefit of utilising all available data for modelling without being restricted like the power set method.

An overview of the methods mentioned can be found in **Table 1**. Binary relevance and classifier chain were the methods utilised in this work.

Table 1. A comparison of multi-label classification methods.

Method	Advantages	Disadvantages
Binary relevance (BR)	• Any single label classification algorithm can be used • Simple	• Higher computational cost than power set • Ignores potential label interactions
Label Power set (PS)	• Any single label classification algorithm can be used • Takes into account label interdependences	• Often, there are several classes representing combined labels with few compounds, which tends to cause over fitting
Classifier chain (CC)	• Takes into account label interdependences	• Which label to choose first? Order of chain has an effect on accuracy[25] • Noisy data created from using predicted value of the first label

There are a number of methods in the literature that assign BCS for drug compounds[26-28]. However these do not offer a computational prediction of BCS class based on chemical structure alone using Quantitative Structure Activity Relationships (QSAR). Where there are a lot of studies that predict either permeability or solubility in separate studies there are few that utilise multi-label classification. Therefore the aim of this work is to compare two multi-label methods for the prediction of BCS. To our best knowledge there are no other works in the literature which compare multi-label methods for provisional BCS prediction suitable for use in drug discovery. Binary relevance is a simple multi-label methods however one disadvantage is it cannot take into account any interactions between the labels. Based on this, this work introduces the classifier chain multi-label classification method for application in pharmaceutics and drug discovery field – to the best of our knowledge; this is the first work using classifier chains in pharmaceutical sciences. It is anticipated that, by using this method and taking into account the label interactions, more accurate models can be produced for provisional BCS prediction. This work shows the potential of multi-label classification

methods, which can be used for the future prediction of many pharmacokinetic properties in drug discovery and development.

2. Methods and Materials

2.1 Datasets

In vitro Permeability

The permeability dataset to build the initial permeability models was taken from the published dataset of Pham-The et al. 2013[6]. This dataset contained apparent permeability values for 1301 compounds from the Caco-2 cell line, measured in the pH range 6.5-7.4. Apparent permeability (P_{app}) is the rate of permeation across cell monolayers and is usually measured in cm/s^{-1}. Upon the removal of duplicates, erroneous compounds and compounds with molecular weights greater than 3000, a dataset of 1288 compounds remained for permeability modelling. In addition, one compound (HBED) was found to have the incorrect SMILES (Simplified Molecular-Input Line-Entry System) code and was corrected. Based on previous work, the benchmark threshold to define the boundary between high and low permeability for 80% Human Intestinal Absorption (HIA) was set at 7.08×10^{-6} cm/s (logPapp of -5.15)[29]. Therefore, a compound with *in vitro* permeability $< 7.08 \times 10^{-6}$ cm/s would be defined as poorly permeable and a compound with permeability $\geq 7.08 \times 10^{-6}$ cm/s would be defined as highly permeable[29].

In addition, *in vitro* permeability data collected from Caco-2 and MDCK cell lines, measured in pH range 6.5-7.4, for 127 compounds were compiled from our previously published dataset[29]. These 127 compounds were not present in the Pham-The et al.'s published permeability dataset, therefore those compounds were to act as an BCS validation set for provisional BCS prediction. This BCS validation set contained 127 compounds where both *in vitro* permeability and aqueous solubility were known (**Supporting Information S1**).

Solubility

Experimental and qualitative aqueous solubility data were obtained from the previously published dataset[29] and combined to give a final total of 750 solubility values (**See Supporting information S2**). The majority of these solubility values were obtained from the AQUASOL dATAbASE (6^{th} Edition)[30] and Martindale (2009)[31]. For the 250 qualitative solubility values that were obtained, these were converted to numerical values based on the principles of Kasim et al.[32] according to the **Table 2**.

Table 2. Solubility definitions adapted from Kasim et al[32]

Descriptive term (solubility definition)	Solubility assigned (mg/mL)
Very soluble (VS)	1000
Freely soluble (FS)	100
Soluble (S)	33
Sparingly soluble (SPS)	10
Slightly soluble (SS)	1
Very slightly soluble (VSS)	0.1
Practically insoluble (PI)	0.01

From this initial dataset of 750 compounds, 127 compounds whose permeability was known were used for the BCS validation set. This resulted in a smaller dataset of 623 compounds used to build and validate the resulting solubility models.

In the BCS, the definition of the boundary between high and low solubility is determined using the dose number ($D_o = (M_o/V_o)/S$), where M_o is the highest dose strength, V_o is 250ml and S is the aqueous solubility (mg/ml)), compounds with $D_o \leq 1$ are classed as highly soluble and drugs with $D_o > 1$ are assigned as poorly soluble drugs[1, 3]. However, in early drug discovery the clinical dose is usually unknown; therefore a suitable threshold needs to be defined. Additionally, D_o is a property of the drug formulation and not a specific property of the active compound. For this work, a solubility cut off of 0.2 mg/ml was set. Hence, any drug with solubility ≥ 0.2 mg/ml was defined as highly soluble and drugs with solubility < 0.2 mg/ml were classed as poorly soluble. A value of 0.2 mg/ml was used as, according to Lipinski et al.[33], this value is the minimum solubility required to get a projected clinical dose of 1 mg/kg for compounds with low permeability. This cut-off for solubility has also been used in a recent work for BCS using MDCK permeability and solubility[5].

2.2 Molecular descriptors

2D and 3D Molecular descriptors were calculated using TSAR 3D v3.3 (Accelrys Inc.), MDL QSAR (Accelrys Inc.), MOE (Chemical Computing Group Inc.) v2012.10 and Advanced Chemistry Development ACD Laboratories/LogD Suite v12. For molecular descriptors calculated based on their 3D structure, the 3D structures of the molecules were first optimised. This was done after removing all the salts and then assigning atomic partial

charges. Molecules were minimized to their lowest energy conformation using AM1 semiempirical method as implemented in MOE software (version 2012.10). A total of 492 molecular descriptors were generated and made available for a feature selection procedure carried out in a data pre-processing phase, before model building.

2.3 Training and validation sets

The compounds in each permeability and solubility dataset were sorted based on either ascending logPapp or logS (mg/mL) separately (excluding the 127 compounds used for the BCS validation set). For each individual dataset, from each group of five consecutive compounds, four were assigned to the training set, and one compound was allocated to the validation set randomly. By doing this a similar distribution of values in the training and validation sets was achieved for both datasets. The resulting compound numbers in the training and validation sets are shown in **Table 3**.

Table 3. Training and validation set compound numbers used in this work

Type of dataset	Training (n)	Validation (n)	BCS Validation (n)
Permeability	1026	262	127
Solubility	490	133	127

The training sets were used to build separate models to predict permeability and solubility classes. The individual validation sets for the permeability and solubility datasets were used to measure the predictive performance of the individual models for the two types of classes. Lastly, in order to compare the two multi-label methods for the provisional BCS classification, an additional BCS validation set containing 127 compounds with known permeability and solubility values was used (BCS validation set).

2.4 Feature selection

Feature selection reduces the number of molecular descriptors used to describe the property (class) being modelled i.e. solubility or permeability. Feature selection can improve interpretability, model accuracy and reduce over fitting of resulting models[34, 35]. Initially, using the training sets only, molecular descriptors with more than 10 missing values were

removed, so that 14 molecular descriptors were removed from each training set and this resulted in 478 molecular descriptors available for feature selection. However there were still certain molecular descriptors with less than 10 missing values in the dataset.

Based on previous work we used predictor importance ranking in random forest to obtain the top 20 molecular descriptors[35]. Using only the training set, optimisation of the random forest was carried out based on the plot of misclassification rate vs the number of trees. The misclassification rate is the number of misclassified compounds divided by the total number of compounds. Based on this plot the optimum number of trees was selected (106 for the solubility, 109 for the permeability). The maximum number of levels for each tree was set to the default 10. The software default value of nine was used for the number of molecular descriptors used in each tree. From the random forest model, the top 20 molecular descriptors were selected based on a ranking function called predictor importance in STATISTICA v 12. For a more detailed description of the feature selection method, see reference[35]. The top 20 molecular descriptors for each property (solubility and permeability) can be found in the supporting information (**Supporting Information S3**).

2.5 Classification and regression trees (C&RTs)

STATISTICA v12 (StatSoft Ltd.) software was used for building each classification model using C&RT analysis. C&RT analysis is a statistical technique that uses decision trees to solve regression and classification problems developed by Breinman et al.[36].

For the binary relevance method, each class – i.e. solubility or permeability variable – was set as the dependent variable and binary classification was carried out using selected molecular descriptors as the independent variables to create individual models for each class label.

For the classifier chain method, initially individual solubility classification models were built using the top 20 molecular descriptors as chosen by feature selection. These models were then used to predict the solubility class for the whole permeability dataset. The permeability model was then built setting permeability class as the dependent variable, while the predicted solubility and the top 20 molecular descriptors pre-selected for permeability were set as the independent variables. The preliminary results indicated that predicted solubility class (acting as a molecular descriptor) would not be used high up in the tree (if at all); therefore predicted solubility was selected manually as the first molecular descriptor in the C&RT model for permeability. The rest of the C&RT decision tree was allowed to be built automatically.

For this work the stopping factors used when growing the C&RT tree were minimum number of compounds for splitting. These stopping factors were the default values for the software and are based on the number of compounds in the dataset. This enables pruning of the tree and prevents over-fitting of the decision tree. For the permeability and solubility datasets, stopping factors of 25 and 12 respectively were used.

2.5.1 Misclassification costs for classification models

Misclassification costs are a useful method to overcome the dataset bias of imbalanced class distributions (where one class value is much more frequent than another) without reducing dataset size[35, 37]. Even if the dataset has a balanced class distribution, the application of higher misclassification cost for a specific class can increase the predictive accuracy and reduce misclassification errors of that specific class.

The solubility and permeability datasets have roughly balanced class distributions, therefore misclassification costs can remain as equal (FP:FN of 1:1, where FP:FN is the ratio of the number of false positives to the number of false negatives). However usually there is an under representation of BCS classes 3 and 4 due to the low number of poorly permeable compounds and compounds with both poor permeability and poor solubility. Therefore in order to potentially improve the predictive accuracy of these underrepresented classes, higher misclassification costs can be applied to reduce false positives (i.e. the number of compounds in the poor solubility and poor permeability classes which are wrongly predicted as having high solubility or high permeability), in order to take into account the lack of compound representation for these classes when combining the solubility and permeability predictions. A higher misclassification cost of 1.5 was applied to the false positive class (FP:FN of 1.5:1) based on the data distribution of the permeability and solubility datasets.

2.6 Statistical evaluation of classification models

2.6.1 Single label models of permeability and solubility

Specificity (SP), sensitivity (SE), cost normalized misclassification index (CNMI), and SP × SE were used to measure the predictive performance of the classification models. Specificity is defined as SP =TN/(TN + FP), where TN is the number of true negatives and FP is the number of false positives. SP is the fraction of poorly permeable/soluble compounds correctly classified by the models. Sensitivity (SE) is the ratio of highly permeable/soluble compounds correctly classified by models, and is defined as SE = TP/(TP + FN), where TP is the number of true positives and FN is the number of false positives. The overall predictive

performance of a model was measured by multiplying the specificity and sensitivity (SP × SE). This measure is an effective way to assess a model's predictive performance as it takes into account the effect of class distribution. By contrast, conventional accuracy measures usually define the ratio of correct over the total number of predictions and do not consider the class-imbalance of datasets. This SP x SE measure has been used in previous investigations for oral absorption prediction[35, 37]. Finally, to take into account misclassification costs in the models, the cost normalized misclassification index (CNMI) was calculated. CNMI can be calculated by **Equation 1**.

$$CNMI = \frac{(\mathrm{FP\ X\ Cost_{FP}}) + (\mathrm{FN\ X\ Cost_{FN}})}{(\mathrm{Neg\ X\ Cost_{FP}}) + (\mathrm{Pos\ X\ Cost_{FN}})} \qquad \textbf{Eq 1.}$$

$Cost_{FP}$ and $Cost_{FN}$ are the misclassification costs assigned for false positives and false negatives, and Neg and Pos define the total number of negative and positive observations, respectively. The calculated CNMI value will be between zero and one, where zero shows no misclassification errors and as the number of misclassification increases the value increases towards 1 (complete misclassification error). For a more detailed explanation of **Equation 1**, see reference 36.

2.6.2 Multi-label models of provisional BCS class

The evaluation of multi-label classification models requires different measures compared to conventional single label classification models[14, 15]. The statistical evaluation of multi-label work can be difficult as a result can be fully correct, partially correct or fully incorrect. Therefore, it is important to have several different evaluation measures, due to the issue of multiple class labels, to help select the best model, i.e. the one with the best model performance over a set of evaluation measures.

For multi-label classification there are two broad types of evaluation measures. These are label based evaluation measures and label set evaluation measures[14-16]. Label based evaluation measures are those based on the individual single labels, such as Hamming loss[38] and classification/subset accuracy[17, 39]. In this work, the accuracy of the individual four BCS classes was used, which is essentially the converse of the Hamming loss – in the sense that the latter is to be minimized, whilst the individual accuracy per class is to be maximized. The individual class accuracy for each class was calculated by dividing the correct number of

predictions for compounds of that class divided by the total number of compounds of that class, resulting in four accuracy measures for the individual four BCS classes. Additionally; for this work the SP X SE accuracy measure of the individual permeability and solubility labels was calculated.

Label set evaluation measures are based on the prediction of all labels together. Therefore this type of measures can be very harsh, as if there is not a perfect prediction of both labels for a compound, that prediction will be considered completely wrong, even if one of the two labels was correctly predicted. Examples of label set evaluation measures are micro-averaging and macro-averaging[40]. The label set evaluation measures used in this work are based on macro-averaging[40]. Macro-averaging is the average, by compound, of all the accuracies for the different BCS classes. To calculate the overall accuracy, the number of correct predictions (regardless of class) was divided by the total number of compounds. However, this value could be biased and not give an accuracy measure which would show the predictive accuracy of all four classes. Therefore, in addition the geometric mean of all four individual predictive accuracy measures for the BCS classes was calculated. The geometric mean is measured by multiplying all the four BCS class accuracy measures and taking the fourth route of this product. The benefit of this measure is that it will not be biased towards the distribution or predictive accuracy of any individual BCS class. In other words, if a model can predict three out of four classes with high accuracy but is unable to predict accurately for one class, the geometric mean accuracy will be low.

3. Results

3.1 Permeability and solubility C&RT models

In this work we are investigating the use of two multi-label classification methods to predict provisional BCS class using permeability and solubility from the literature and published datasets. Separate models of permeability and solubility were built using training sets of 1026 and 490 compounds respectively, using the top 20 molecular descriptors selected by the random forest-based feature selection method. The predictions from the solubility and permeability models were then combined to give a provisional BCS class for an BCS validation set of 127 compounds. All the C&RT decision trees that produced the results reported in **Tables 4 and 5** can be found in the **Supporting Information S4**. In **Tables 4 and 5**, the best models are those that have the highest SP, SE and SP X SE and the lowest CNMI. These have been highlighted in **bold** for the training and validation sets in these

tables. Firstly, the two solubility models whose results are shown in **Table 4** are models with equal and higher misclassification costs applied to reduce false positives – models 1 and 2, respectively. The compound numbers in training and validation sets for solubility and permeability for **Tables 4 and 5** are lower than the original numbers in Table 3. This is because for certain compounds molecular descriptors were unable to be calculated and therefore were unable to be classified in the terminal nodes. Therefore, the compound numbers in **Tables 4 and 5** represent the compound numbers classified by the models.

Table 4. Results of C&RT Analysis for the Classification of Solubility

Model	Misclassification cost ratio (FP:FN)	Set	n	SP X SE	SE	SP	CNMI
1	1:1	t	485	0.621	**0.784**	0.792	0.212
		v	128	**0.578**	**0.795**	0.727	**0.234**
2	1.5:1	t	485	**0.638**	0.706	**0.903**	**0.178**
		v	128	0.538	0.658	**0.818**	0.243

Note that the numbers of compounds used in the analysis are lower than the available compounds due to missing descriptor values for some chemicals

Both solubility models from **Table 4** can be considered the best depending on the intended use and purpose of the model. Model 1 has the highest sensitivity for the training set and validation set as well as overall accuracy for the validation set. Whereas model 2, as expected has the highest SP for the training and validation set due to the application of higher misclassification costs to reduce false positives. Therefore, if the aim of the model is to predict poorly soluble compounds, model 2 would be the best model; but model 1 would be the best to use if the aim was to predict highly soluble compounds. Model 1 may be considered as the best C&RT model in this work (shown in **Figure 4**), since for the validation set there is more of a balanced prediction for poorly and highly soluble compounds (higher SP X SE). Both solubility models were then used to predict solubility for compounds in the permeability dataset, which was in turn used as an additional descriptor (independent variable or feature) for building permeability model – this process implements the classifier chain approach for multi-label classification, discussed earlier.

The statistical parameters of the permeability models produced in this work are shown in **Table 5**. Initially permeability models were built using only the top 20 molecular descriptors selected by the random forest-based feature selection method (models 1 and 4). Next, permeability models were built using the predicted solubility either from the solubility model

1 or from solubility model 2 in **Table 4** in addition to the top 20 molecular descriptors as the independent variables. Again models were also built with equal (models 1-3) or higher misclassification costs (models 4-6) applied to reduce false positives (FP:FN 1.5:1).

Table 5. Results of C&RT Analysis for the Classification of Permeability (with and without predicted solubility incorporated in the model)

Model	Misclassification cost ratio (FP:FN)	Solubility Model included	Set	n	SP X SE	SE	SP	CNMI
1	1:1	none	t	1016	0.653	**0.847**	0.771	0.192
			v	261	0.503	0.727	0.692	0.291
2		1	t	1016	0.655	0.841	0.778	0.191
			v	261	**0.519**	**0.742**	0.699	0.280
3		2	t	1016	0.638	0.761	0.838	0.200
			v	261	0.482	0.641	0.752	0.303
4	1.5:1	none	t	1016	**0.659**	0.807	0.817	0.188
			v	261	0.484	0.664	0.729	0.298
5		1	t	1016	0.630	0.716	0.880	**0.185**
			v	261	0.489	0.586	**0.835**	**0.265**
6		2	t	1016	0.625	0.706	**0.884**	0.187
			v	261	0.489	0.586	**0.835**	**0.265**

Note that the numbers of compounds used in the analysis are lower than the available compounds due to missing descriptor values for some chemicals

Based on the validation set, the best permeability model to choose would be model 2. This permeability model was built using the predicted solubility from model 1 in **Table 4** and equal misclassification costs applied. This model achieved the highest overall accuracy (SP X SE) and sensitivity for the validation set of 0.519 and 0.742, respectively. In addition, it also had the second highest SP X SE and SE for the training set and the lowest CNMI for the training and validation sets, when comparing the other models with equal misclassification costs applied (models 1-3).

Table 5 shows that when equal misclassification costs are applied (models 1-3), a higher overall accuracy model (based on the validation set) is produced using predicted solubility (from solubility model 1 in Table 4) as a molecular descriptor to predict permeability class. Although model 3 has a lower overall accuracy, its specificity is much higher and this could be due to the influence of the solubility model included in the permeability model (solubility model 2). In other words, improving the prediction of poorly soluble compounds resulted in higher prediction accuracy for poorly permeable compounds according to **Table 5**.

When higher misclassification costs are applied to false positives in the permeability models, models 5 and 6 have better overall accuracy (SE X SP) for the validation set and the lowest CNMI for the training set was obtained by model 5. Overall, the application of higher misclassification costs to reduce false positives resulted in the increased specificity and lower misclassification errors (CNMI), but overall accuracy is lower in models 4-6 in comparison with models 1-3. As expected, model 6, which included predicted solubility from model 2 in Table 4, had a higher specificity due to the higher misclassification costs originally applied to the solubility model – which have been utilised to improve prediction accuracy for poorly permeable compounds.

3.2 Interpretation of selected solubility and permeability models

Solubility classification models were developed using the top 20 molecular descriptors. In addition, permeability models were developed using either the top 20 molecular descriptors (selected using random forest) or the top 20 molecular descriptors plus predicted solubility from solubility models built in this work. It must be noted that although the top 20 molecular descriptors were given as input to the algorithm that builds the C&RT tree, not all the molecular descriptors were used to build the decision trees, since the C&RT also performs an additional 'embedded' feature selection process, adding to the tree only attributes deemed relevant for class prediction by the algorithm[35]. Furthermore, some molecular descriptors can be used more than once in a C&RT tree, as discussed below. **Figure 4** is the selected solubility model 1 based on the classification decision tree.

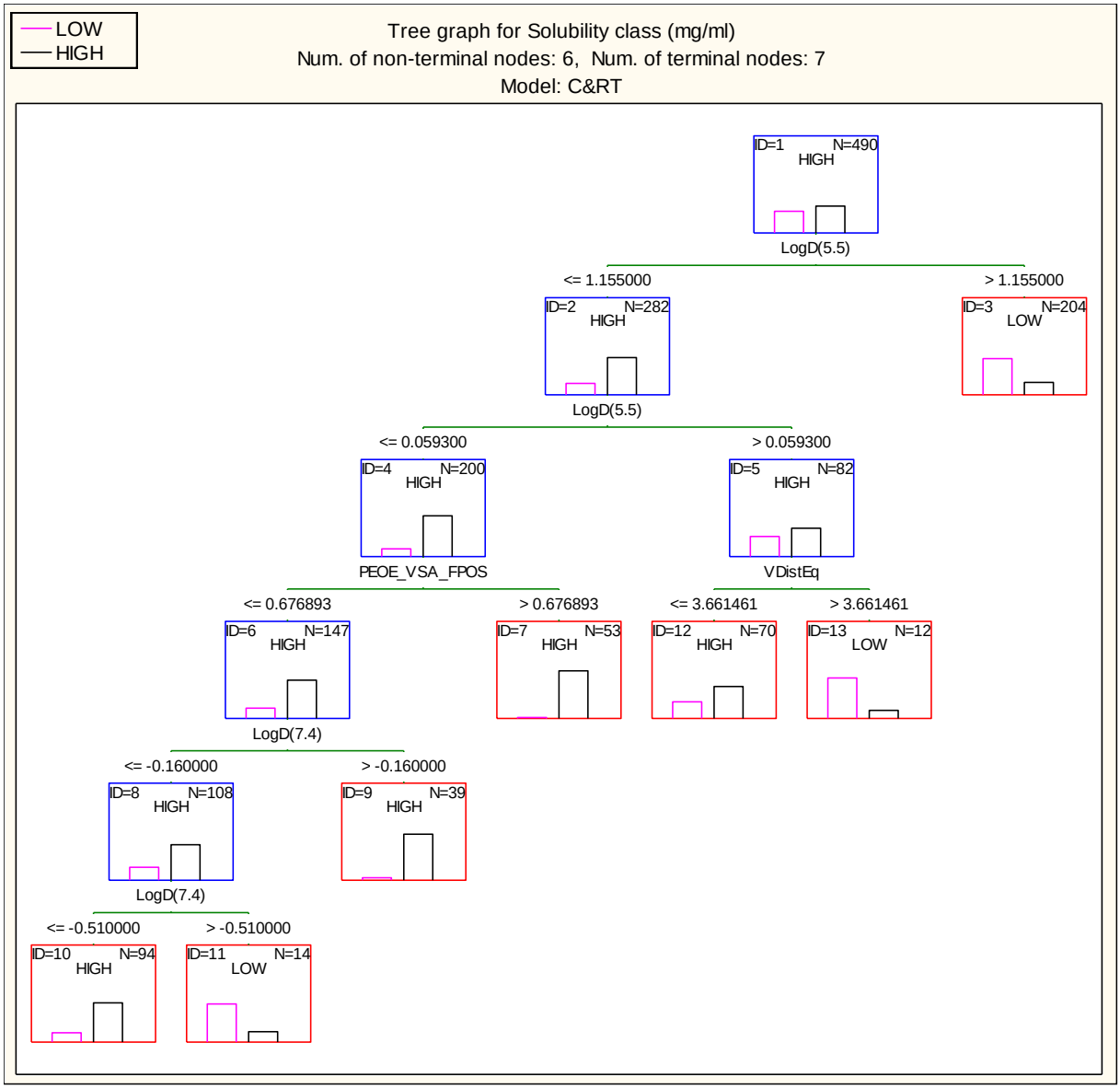

Figure 4. Tree graph for C&RT analysis for the prediction of solubility class with equal misclassification costs (model 1 in Table 4)

The first split variable in **Figure 4** is ACDLogD(5.5), the logarithm of the apparent distribution coefficient between octanol and water at pH 5.5, a measure of hydrophobicity. This descriptor as well as logP has been used in many publications for modelling of different properties such as oral absorption[11, 37], permeability[12, 41] as well as solubility models[13, 42]. The use of logD at pH 5.5, despite solubility being measured at pH 7.4, is justified based on the fact that this descriptor indicates not only the effect of lipophilicity, but also the effect of acid/base property of the compounds. For example, an acidic and a basic compound of similar logP values will have different logD at this pH depending on their percentage of ionisation. At pH 5.5, the acidic compound will be mainly unionised and hence its logD(5.5)

will be close to its logP value, whereas the basic compound will be highly ionised therefore it will have a lower logD(5.5) than its logP value. In relation to solubility, highly lipophilic compounds can give rise to poor solubility, as indicated by **Figure 4**. In this model compounds are poorly soluble if they have a LogD(5.5) > 1.16 and examples of poorly soluble drugs in this node are diclofenac and ibuprofen, both are BCS class II compounds (poorly soluble but highly permeable)[43, 44]. There is no further splitting of the highly lipophilic, poorly soluble compounds indicating that this molecular descriptor is useful to define poor solubility (< 0.2 mg/mL) for this tree. The relatively less lipophilic compounds (LogD(5.5) ≤ 1.16) are further characterised into high/low solubility using LogD(5.5); this time a lower threshold of 0.06 is used. In this case both nodes 4 and 5 are associated with high solubility; however, compounds that have higher LogD(5.5) (but lower than 1.16) are poorly soluble only if they have a vertex distance equality index (VDistEq) > 3.66. Computed from a distance matrix, VDistEq is mainly related to the size and shape (branching) of a molecule[45]. Compounds with larger VDistEq tend to be larger and in most cases (less branched) linear molecules.

For compounds with lower LogD(5.5) than 0.06, the next molecular descriptor to split the tree is the partial charge descriptor, PEOE_VSA_FPOS. Using PEOE partial charge calculation[46], PEOE_VSA_FPOS is the sum of the van der Waals surface area of positively charged atoms divided by the total surface area of the molecule[45]. According to **Figure 4**, those compounds with a PEOE_VSA_FPOS > 0.67 will be highly soluble indicating that those with more positive partial charges (an indication of higher polarity and ionization) will be highly soluble. This is in agreement with the literature, where more polar molecules tend to be more soluble in water[47].

However, as depicted by this tree, node 6 (containing less polar compounds with PEOE_VSA_FPOS
≤ 0.67) is not pure at all and needs more splitting with other molecular descriptors; in this case, LogD(7.4) is used twice in the tree for these compounds. In **Figure 4** compounds will be classed as poorly water soluble if -0.51 < LogD(7.4) ≤ -0.16. It must be noted here that all these compounds have a LogD(5.5) below 1.155, as a result of division of node 2 and therefore they are hydrophilic enough to be classed as water soluble. Examples of these poorly water soluble compounds in node 14 are rofecoxib [48] and pindolol [49]. Overall, from the solubility model, the main molecular descriptors used to classify solubility are those

related to lipophilicity, ionization, polarity, size and shape, which is in accordance with the literature[47, 50, 51].

The best permeability model selected was model 2 in **Table 5**. Due to the size of the tree, in order to facilitate its interpretation the tree has been split into two trees (**Figures 5 and 6**). **Figure 5** shows the half of the permeability decision tree that is built for those compounds predicted as poorly soluble by the solubility model 1 in **Table 4**. Figure 6 shows half of the C&RT tree for permeability built for those compounds predicted as highly soluble from the same solubility model. It must be noted that the trees in **Figures 5** and **6** were originally one tree and the combined version, as well as all the other C&RT models presented in this work, is in the **Supporting Information S4**.

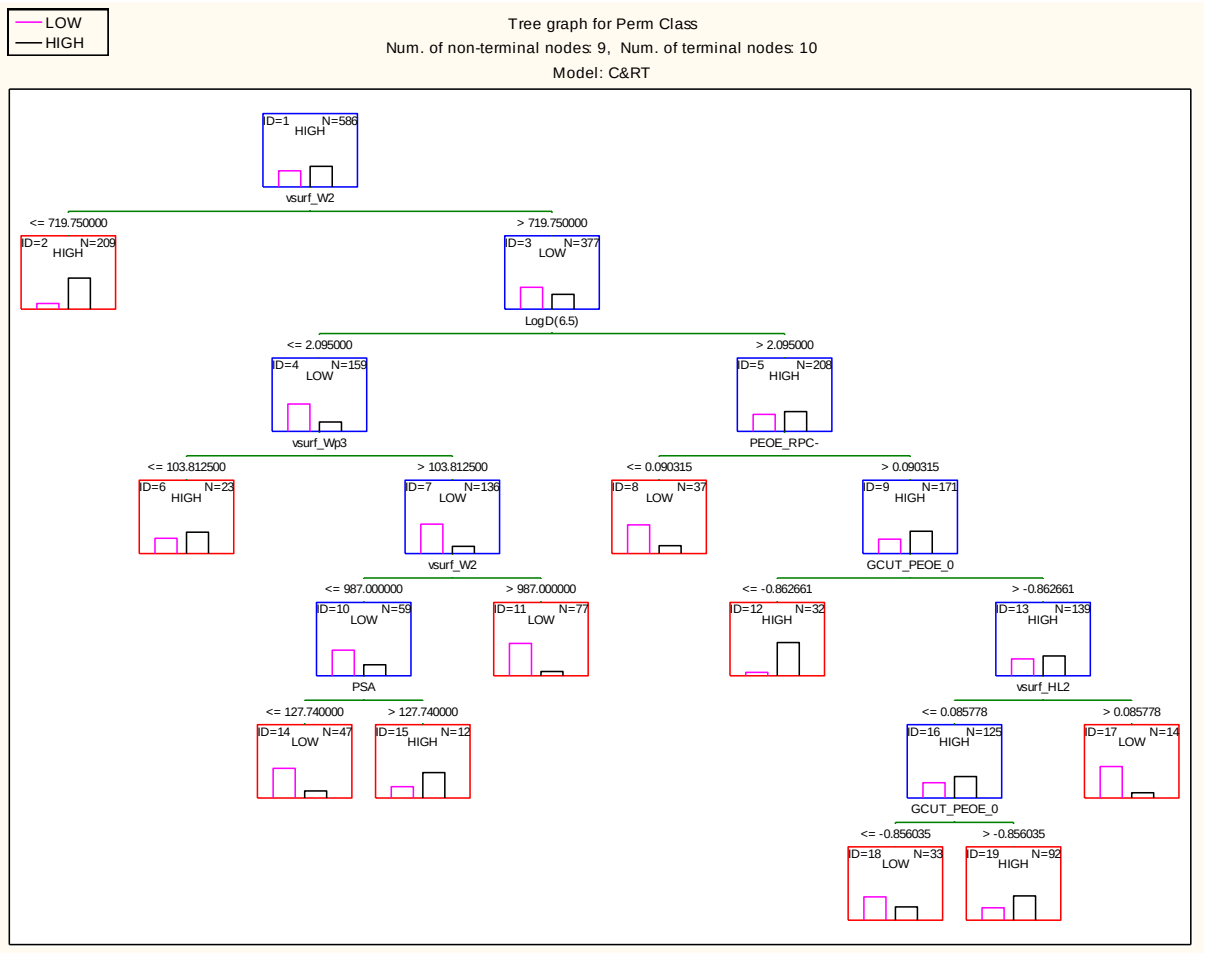

Figure 5. Tree graph for C&RT analysis (part of model 2 in Table 5) for the prediction of permeability class for predicted poorly soluble compounds from solubility model 1 (shown in Figure 4)

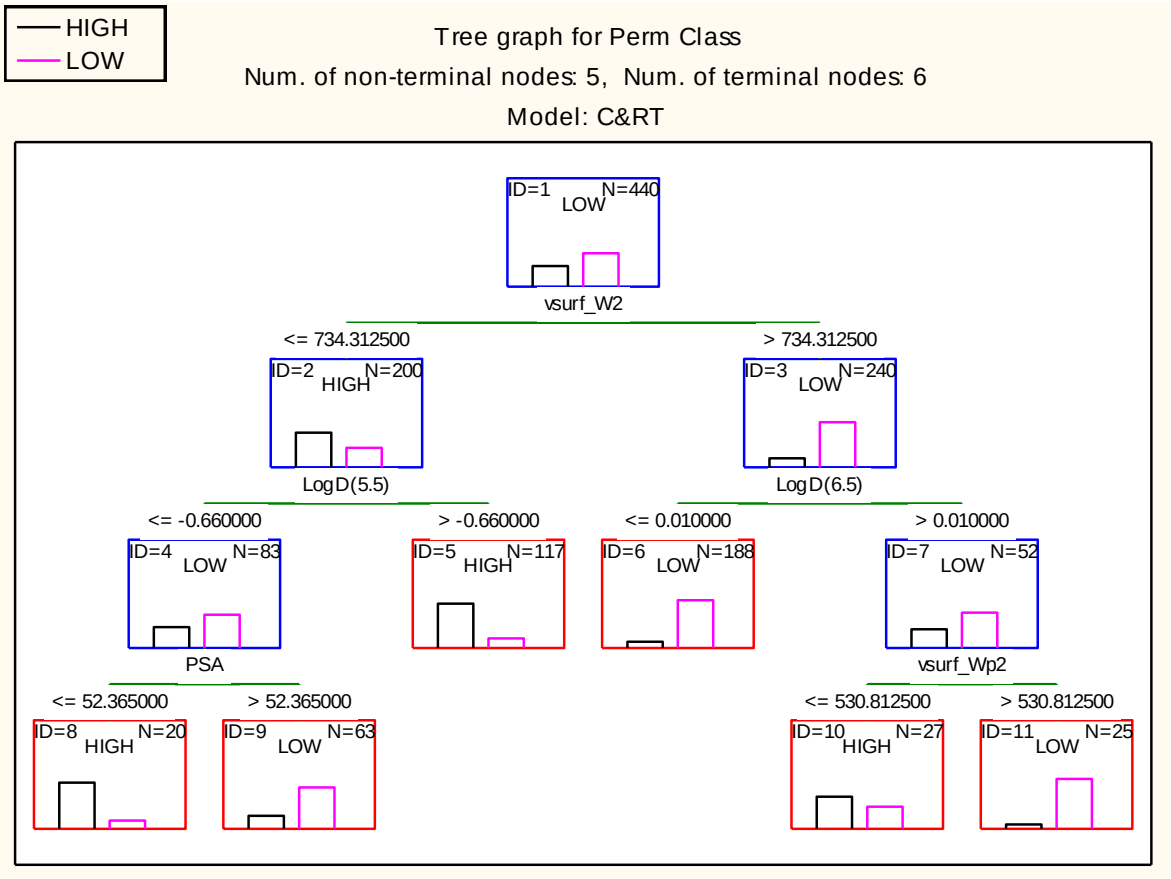

Figure 6. Tree graph for C&RT analysis (part of model 2 in Table 5) for the prediction of permeability class with equal misclassification costs for predicted highly soluble compounds from solubility model 1 (show in Figure 4)

Comparing **Figures 5** and **6**, it is noted that there is a slightly larger number of poorly soluble compounds (**Figure 5**) than highly water soluble compounds (**Figure 6**) in the permeability dataset and those poorly soluble compounds are mainly highly permeable (**Figure 5**) and *vice versa*. The first split of the tree in Figure 5 is using the vsurf_W2 molecular descriptor as calculated by MOE[52]. Vsurf and related molecular descriptors are Volsurf descriptors described by Crucciani et al (2000)[53] that describe the size, shape, polarity, hydrophobicity and the balance between these properties on molecules. More specifically, vsurf_W descriptors describe the volume of hydrophilic regions of a molecule, calculated at certain interaction energy levels. In this case vsurf_W2, calculated at energy level 0.5 kcal/mol, accounts for the polarizability and dispersion forces in the hydrophilic regions of the molecules[52]. According to this tree, poorly soluble compounds in **Figure 5** will be classified as highly permeable so long as they have small hydrophilic volume (node 2). Compounds with larger hydrophilic volumes in nodes 3 have been divided further according to logD6.5.

In this case, the general trend is that less lipophilic compounds (logD6.5 ≤ 2.10) will be mostly poorly permeable (node 4), which matches previous observations in caco-2 and other *in vitro* permeability cell lines[29, 54]. For those less lipophilic compounds (logD6.5 ≤ 2.10), the descriptor vsurf_Wp3 is used to discriminate between compounds with small polar volume (Vsurf_Wp3 ≤ 103.8) which are highly permeable, and compounds with large polar volume of the molecule (node 7). Compounds will be classified as poorly permeable due to their large polar volume unless they have smaller volume (Vsurf_W2 ≤ 987), but a polar surface area (PSA) greater than 127.7 (node 13). Polar surface area (PSA) is a common molecular descriptor used in oral absorption models as well as permeability models[11]. PSA is the area of the van der Waals surface that arises from oxygen and nitrogen atoms or hydrogen atoms bound to these atoms[55]. PSA has been cited to have a negative effect on oral absorption and hence permeability; this was also observed in previous works using oral absorption dataset [11,29,35,37]. However this is not what is presented in **Figure 5** for the permeability dataset. The maximum PSA in this list of compounds (159 Å) is still moderate in comparison with the rest of the dataset. On closer inspection, the vast majority of these highly permeable compounds contain a sulphonamide or thaizole group. The polarity measure of these sulphur-containing functional groups using PSA seems to not correlate with the expected reduced absorption of polar compounds. Examples of these highly permeable compounds with large PSA values are glipizide and two oxazolidinones, antimicrobial agents PNU-182945 and PNU-183981.

For highly lipophilic compounds (logD 6.5 > 2.1) the next descriptor used to discriminate between high and low permeability is the relative negative partial charge descriptor calculated by PEOE (RPC-). This molecular descriptor is calculated by dividing the smallest negative charge by the sum of (most negative) charges on the whole molecule. Therefore, a higher number of hydrogen bond acceptors such as oxygen atoms in the molecules leads to lower values of RPC-. In this instance, compounds with a lower relative negative partial charge (≤ 0.09) are poorly permeable. Compounds with a higher RPC- are mainly highly permeable but can be split further by the molecular descriptor, GCUT_ PEOE_0. GCUT descriptors are calculated from the eigenvalues of a modified graph distance matrix with the diagonal using in this case charges calculated from PEOE partial charges. A minority of compounds with a lower GCUT-PEOE_0 than -0.86 have been classed as highly absorbed. These are structurally large and complex molecules with many rings and branches, mostly belonging to nucleotide based antivirals. Due to similarity of these compounds to natural

metabolites, it is likely that they may have the possibility of being transported by carrier proteins.

Compounds with a higher GCUT_ PEOE_0 are also classified as highly permeable unless if they have a vsurf_HL2 > 0.086 or despite a smaller Vsurf_HL2, have GCUT_PEOE_0 ≤ - 0.856. Vsurf_HL2 describes the hydrophilic-lipophilic balance, which is the calculated ratio between the hydrophilic regions measured at 4 kcal/mol and the hydrophobic regions measured at 0.8 kcal/mol[52]. According to the tree in **Figure 5**, compounds are predicted as poorly permeable if they have a higher ratio of hydrophilic to lipophilic effect, and examples include bromocriptine and lansoprazole.

Figure 6 is the permeability model for compounds predicted as highly soluble according to solubility model 1. In this figure the same top molecular descriptor as in **Figure 5** is selected to split the compounds into high/low permeability in node 1. Compounds with vsurf_W2 values greater than 734.2, i.e. larger hydrophilic volume, are more likely to be poorly permeable according to this tree. This is unless they have a higher lipophilicity (logD6.5 > 0.01) and lower polar volume, according to vsurf_Wp2 ≤ 530.8. On the other side of the tree, majority of compounds with relatively small hydrophilic volume are highly permeable unless they are relative hydrophilic at pH5.5 (LogD(5.5) ≤ -0.66) and have a PSA higher than 52.4. In this instance, this PSA threshold is similar to the threshold of 60 Å used for recent permeability modelling of Caco-2 permeability[41]. Based on **Figures 5** and **6**, it is interesting to note that hydrophilic volume of a molecule is a better measure of permeability than the most widely known parameter, partition coefficient. For instance, in **Figure 6**, node 2, it can be seen that a good fraction of compounds with lower LogD(5.5) than -0.66 are highly permeable given the polar surface area is not too large (≤ 52.3).

3.3 Provisional BCS class prediction in an BCS validation set using multi-label methods

The permeability and solubility models created previously were used to predict the BCS of an BCS validation set of 127 compounds with known values for both properties collected from the literature (BCS validation set). Different combinations of permeability and solubility models were tried in order to see which effect this would have on the overall results. **Table 6** shows the results from the different combinations of the permeability and solubility models presented in **Tables 4 and 5**. For example in **Table 6**, model 1 is the combination of the solubility model 1 (**Table 4**) and permeability model 1 (**Table 5**).

Recall, the multi-label method binary relevance (BR), involves the prediction of permeability and solubility separately (models 1-2, 7-8 in **Table 6**), however it fails to take into account the relationship between these interrelated properties. Whereas the classifier chain (CC) method, which uses a predicted solubility alongside structural molecular descriptors to help predict permeability, takes into account the label interactions (Models 3-6, 9-12 in **Table 6**). In **Table 6**, the overall accuracy (SP X SE) of the permeability and solubility models for the BCS validation set has also been included. In addition, the overall accuracy and geometric mean have been calculated alongside the individual class accuracies in order to help with interpretation.

Table 6. Results of the provisional BCS classification of an BCS validation set (n=127) to compare the binary relevance and classifier chain multi-label methods

Model	Multi-label method	Permeability Model Used (Table 5)	Solubility model Used (Table 4)	Permeability Accuracy (SP X SE)	Solubility Accuracy (SP X SE)	Overall Accuracy[a]	Geometric mean[b]	Class 1 accuracy (n=53)[c]	Class 2 accuracy (n=40)[c]	Class 3 accuracy (n=26)[c]	Class 4 accuracy (n=8)[c]
1	BR[d]	1	1	0.525	0.565	0.606	0.000	0.566	**0.725**	0.692	0.000
2			2		0.551	0.591	0.496	0.509	**0.725**	0.653	0.250
3	CC[e]	2	1	0.641	0.565	**0.630**	0.523	0.585	0.700	0.731	0.250
4			2		0.551	0.606	**0.590**	0.528	0.700	0.654	0.500
5	CC[e]	3	1	0.642	0.565	0.598	0.508	0.528	0.625	**0.806**	0.250
6			2		0.551	0.575	0.574	0.453	0.625	0.769	0.500
7	BR[d]	4	1	0.480	0.565	0.543	0.000	0.453	0.675	0.692	0.000
8			2		0.551	0.528	0.456	0.415	0.675	0.615	0.250
9	CC[e]	5	1	0.581	0.565	0.559	0.472	**0.604**	0.450	0.731	0.250
10			2		0.551	0.543	0.563	0.547	0.450	0.654	**0.625**
11	CC[e]	6	1	0.587	0.565	0.559	0.481	0.528	0.500	**0.808**	0.250
12			2		0.551	0.528	0.537	0.434	0.500	0.500	0.500

[a]Overall accuracy, calculated as correct number of predictions divided by total number of predictions; [b]Geometric mean, multiplication of all accuracy predictions of classes 1-4 and taking the fourth root of this product; [c]Class average ,number of correct class prediction divided by total number of the specific class;[d]BR, Binary relevance; [e]CC, Classifier chain

From **Table 6**, based on the overall accuracy i.e. the highest percentage of correct predictions, the best model to choose would be model 3. This model had an overall accuracy 0.630 (80/127) and was created combining the solubility model 1 and permeability model 2 (with incorporated predicted solubility). Although this model has the highest number of correct predictions, it has a poorer predictive accuracy for class 4. Therefore, using the geometric mean, which gives an average overall accuracy of all four classes, the best model would be model 4. This model was created combining the solubility model 2 and permeability model 2 (with incorporated predicted solubility). The difference between models 3 and 4 in **Table 6** is the solubility model used with permeability model 2 to put compounds into BCS classes. Solubility model 1 from **Table 4** is with equal misclassification costs and solubility model 2 is with higher misclassification costs to reduce false positives. Different combinations of the permeability and solubility models result in the different models having the best accuracy for all four classes. It is difficult to pick the best model based on the individual accuracies of the four classes. However, for overall accuracy the best model to choose would be either model 3 or model 4.

Models 1-6 were all derived from permeability models using equal misclassification costs applied, whereas Models 7-12 were derived from permeability models with higher misclassification costs applied to reduce false positives. Overall the application of higher misclassification costs to false positives in the permeability models (models 7-12) has led to lower overall accuracy and geometric mean accuracy; however, it has also led to the highest class accuracy for class 3 (model 11) and class 4 (model 10), due to better prediction of the low permeability compounds as expected.

In order to compare the models built by the two multi-label methods, firstly models 1 and 2 in **Table 6** can be compared with models 3-6. Models 1 and 2 were built by the binary relevance method, whereas models 3-6 were built by the classifier chain method multi-label method. Overall, based on the geometric mean the classifier chain method obtained higher predictive ability across all classes. The only exception is that although models 5 and 6 have a higher geometric mean, they have a slightly lower overall accuracy compared with the binary relevance models 1 and 2. The superiority of classifier chain method can also be seen from the permeability accuracy which was higher for the models built by the classifier chain method, indicating that incorporating predicted solubility into models results in higher predictive accuracy for permeability. These patterns are also seen when comparing models 7-

12, where higher misclassification costs have been applied to reduce false positives for the permeability models.

4. Discussion

This work has explored attempts to build permeability and solubility models to computationally predict a provisional BCS for chemicals in drug discovery by comparing two multi-label classification methods. The predictions can be very useful in early drug development and can streamline formulation and chemical optimization strategies. In addition, the BCS predictions can give insight into the mechanistic absorption properties of drugs, such as rate limiting steps like transporter effects or dissolution limiting solubility.

This work has involved multi-label classification of *in vitro* permeability and aqueous solubility to provisionally predict BCS classes for new chemical entities (NCEs) for early stage drug discovery. In order to compare the two multi-label methods, individual permeability and solubility models were built and validated. Initially, permeability and solubility models were built using the top 20 molecular descriptors as selected via random forest-based feature selection. Our previous study shows improved prediction accuracy when a pre-processing feature selection is performed prior to C&RT analysis[35]. In addition, permeability models were also built utilising the predicted solubility alongside the selected molecular descriptors to predict permeability class. The use of higher misclassification costs for false positives was also investigated to help improve class prediction of the poorly permeable and poorly soluble classes. Using an BCS validation set with known solubility and *in vitro* permeability, the predictions of the permeability and solubility models were combined and compared to the observed experimental BCS class. In this way, we compared two multi-label methods using the BCS validation set. Binary relevance involves the combination of separate, independently-built solubility and permeability models; however this does not take into account the interactions between these two labels. In order to overcome this, we compared this method to the multi-label method classifier chain. This method, in relation to this work, involved the incorporation of predicted solubility to build and predict permeability class and in doing so this method takes into account the relationship between these properties. Therefore, we are exploring the idea that the classifier chain method can help improve permeability class prediction and in turn provisional BCS class prediction.

4.1 Individual Permeability and solubility models

Both permeability and solubility are important properties in drug discovery. However, both these properties individually are complex and can be difficult to model. Lack of high quality datasets for drug-like compounds can contribute to the difficulty in predictions. BCS class prediction can overcome variable permeability and solubility data by predicting compounds' classes rather than specific values as a first initial drug screen. However, suitable thresholds for discriminating between high and low permeability/solubility must be selected.

Permeability is the rate of drug absorption through Caco-2 cell line and is highly correlated with intestinal absorption[29]. Similar to intestinal absorption, there are many factors affecting and influencing permeability. According to the results of this study using the top 20 molecular descriptors from feature selection, permeability classes can be predicted with good accuracy. On the whole it is easier to predict the high permeability class than it is to predict the poor permeability class when equal misclassification costs were applied on a dataset with balanced class distribution (higher sensitivity than specificity values in **Table 5**). The same pattern emerges in relation to solubility, where according to this work better predictive accuracy is obtained for highly soluble compounds when using equal misclassification costs (**Table 4**). Solubility is also another complex parameter to predict with many complex interlinking factors[56 47].

When equal misclassification costs have been applied, using predicted solubility as a molecular descriptor alongside the other molecular descriptors to build permeability models caused two things: models had better overall accuracy and better accuracy for poorly permeable compounds in comparison with the model not incorporating predicted solubility (see **Table 5**). Therefore, the inclusion of predicted solubility in this way increased the predictive accuracy of the poor permeability class. When higher misclassification costs were applied to improve the prediction of poorly permeable compounds, the specificity of permeability models also increased upon incorporating predicted solubility. Therefore, inclusion of predicted solubility into permeability models has resulted in better models or those that can predict poor permeability class better. This follows on from previous research where by incorporating experimental permeability and experimental and predicted solubility into oral absorption models results in higher predictive accuracy[29]. When higher misclassification costs were applied to reduce false positives for the permeability models, overall lower predictive accuracy was observed. This could be due to the balanced nature of the dataset, containing roughly 50:50 high:low permeability compounds.

4.2 Comparison of molecular descriptors

It is difficult to directly compare different permeability and solubility models used in the literature; however the molecular descriptor subsets used in the models can be compared. The top 20 molecular descriptors selected by random forest using predictor importance can be found in the **Supporting Information S3**. In addition, the top descriptors chosen from the pool of 20, by the C&RT analysis for the two properties can also be compared to see if there are similarities and/or differences, and this can be related back to the property in question. The top molecular descriptors selected by the solubility and permeability (C&RT) models are shown in **Tables 7** and **8** respectively. The top molecular descriptors are counted by how many models they appear in, also noting in **Table 7** is if the molecular descriptor occurs more than once in the same decision tree. For **Table 7**, the molecular descriptors from solubility models 1 and 2 (**Table 4**) were used to show the top solubility molecular descriptors. For **Table 8**, permeability models 1 and 4 and models 2, 3, 5 and 6 (**Table 5**) were used to show the top molecular descriptors for the binary relevance and classifier chain methods respectively.

Table 7. The top molecular descriptors selected by C&RT for the prediction of solubility class (models 1 and 2 in Table 4)

Type of descriptor	Descriptor	Number of C&RT models	Model (From Table 4)
Lipophilicity	LogD(5.5)	4[a]	1,2
	LogD(7.4)	3[a]	1,2
Size/shape	VDistEq	3[a]	1,2
	BCUT_PEOE_0	1	2
	BCUT_SLOGP_2	1	2
Polarity/ Polarization	PEOE_VSA_FPOS	1	1
	PEOE_VSA_POL	1	2
Hydrogen bonding	MaxHp	1	2

[a]Occurred more than once in a single tree model.

For the solubility models 1 and 2 the top molecular descriptors (**Table 7**) picked by C&RT analysis was LogD(5.5). Other studies have identified lipophilicity descriptors relating to LogD(5.5) and LogD(7.4), such as logP, as important for the prediction of solubility [42, 57]. The next most frequently picked molecular descriptor is VDistEq, relating to the size and shape of the molecule. Larger molecules in drugs and drug like molecules tend to have higher lipophilicity[47] and additionally require higher energy to create a cavity in the solvent and solvate (solvation limiting solubility)[58]. Additionally the size and shape of a molecule can

result in a rigidity that can cause high crystal lattice energy resulting in poor solubility (solid-state limiting solubility)[47, 58] Finally those descriptors relating to polarity and hydrogen bonding are also important for solubility prediction[47, 59]. Overall, molecular descriptors relating to lipophilicity, size, shape, polarity and hydrogen bonding are all important for solubility of drug compounds as they relate to the crystal lattice energy, solvent cavity formation energy and solvation energy – all important factors for solubility of drug compounds[47, 59, 60].

Table 8. The top molecular descriptors selected by C&RT for the prediction of permeability class for the binary relevance (models 1 and 4, Table 5) and classifier chain permeability models (models 2, 3, 5 and 6, Table 5)

Type of descriptor	Descriptor	BR permeability models		CC permeability models	
		Number of C&RT models	Model (From Table 5)	Number of C&RT models	Model (From Table 5)
Lipophilicity/ Hydrophobicity	LogD(6.5)	3[a]	1,4	6[a]	2,3,5,6
	LogD(5.5)	1	1	3	2,3,6
	LogD(10)			3	3,5,6
	LogD(7.4)	2	1,4		
	vsurf_HL1	2	1,4		
	vsurf_HL2			4	2,3,5,6
	vsurf_CW4	1	1		
Size of hydrophilic/polar regions	vsurf_Wp3	2	1,4	7[a]	2,3,6
	vsurf_W2	1	1	7[a]	2,3,5,6
	vsurf_W3	1	4	2	5,6
	vsurf_Wp2	1	4	2	2,5
	PEOE_RPC-	1	4	4	2,3,5,6
	PSA			2[a]	2
Size/Shape	xv2	2[a]	4		
	GCUT_PEOE_0	3[a]	1,4	8[a]	2,3,5,6
	chi1_C	2	1,4		
Bascity	FIBpH6.5	2[a]	4		
Hydrogen bonding	vsurf_HB1	5[a]	1,4	3[a]	5

[a]Occurred more than once in a single tree model. BR: Binary Relevance; CC: Classifier Chain

The top molecular descriptors for the permeability models in this work picked by the resulting C&RT analysis can be roughly grouped into five groups: lipophilicity/hydrophobicity parameters, those describing the size of the hydrophilic or polar molecular regions, basicity, hydrogen bonding, and finally size/shape parameters (**Table 8**). Overall, there are 25 cases of lipophilicity/hydrophobicity parameters used in the permeability models and 30 cases of parameters describing the size of the hydrophilic or polar regions of the molecule. These two make up 69% of permeability related features. There are only two instances of the basicity parameters, eight cases of hydrogen bond donor effect and 15 cases of molecular descriptors relating to size and/or shape utilized in the permeability models. The importance of hydrophilic or polar size of the molecular has been seen in previous literature. In particular, polar surface area has been cited to be important for permeability classification between low, medium and high permeability, and is a popular molecular descriptor used in our models[41]. Molecular descriptors relating to hydrogen bonding are also popular in relation to permeability[61] as well as oral absorption. More specifically hydrogen bonding is one of the descriptors used in the widely accepted filter for identifying poorly absorbed compounds, Lipinski's rule of five[62]. Molecular descriptors important for permeability such as those relating to lipophilicity, size/shape, polarity and hydrogen bonding are also important for the prediction of oral absorption[11, 35, 37].

4.3 Comparison with related literature

There are few studies to our knowledge which use QSAR models to predict BCS class. However, there are many individual studies that predict either permeability or solubility. A related work has been published recently by Pham-The et al (2013)[6], which is different from this study in terms of the methods, parameters used and property thresholds.

As a solubility measure, Pham-The et al. used dose number (D_o) defined as the ratio of drug concentration following a given dose in the stomach of 250ml volume, to the saturated solubility. One of the problems with using D_o for a provisional prediction is that D_o is a property of the drug formulation and not a specific property of the active compound. Therefore the maximum dose can depend on many things such as formulation type, toxicity and drug target affinity or even different doses of drug may be used to treat different disease severity or even different disease states[22]. In terms of future predictions, maximum dose will be needed from literature in order to calculate D_o. The advantage of our models described here is that they do not need any experimental values such as the drug dose for future predictions.

They also used a permeability threshold of 16 x 10^{-6} cm/s, based on the permeability of metoprolol, a highly absorbed drug. This threshold is over double the threshold that was objectively selected and statistically validated using the correlation between oral absorption and *in vitro* permeability in previous studies[29]. The individual permeability and solubility models developed by Pham-The et al using a dataset of 322 compounds achieved good overall accuracy for the training and validation sets (>75%). Due to the different datasets and validation and training sets, the accuracy of the models cannot be directly compared. We have used larger datasets for model development that cover a large chemical space. In addition, the different thresholds used lead to different classification problems, each resulting in different levels of difficulty for classification of each property.

Pham-The et al. (2013) validated the models by using firstly an external validation set containing 57 compounds from the WHO (World Health Organization) list of essential medicines. Unfortunately, in this validation set there was no experimental Caco-2 permeability data to validate the permeability prediction, furthermore over half of these compounds are assigned into more than one class, which is potentially inconclusive. Our work involved validation sets to validate permeability and solubility models and in addition an BCS validation set where both permeability and solubility were known, in order to validate BCS prediction.

There are studies in the literature that predict BDDCS class (Biopharmaceutics Drug Disposition Classification System)[10] instead of BCS class. The BDDCS system classifies compounds into one of the four BDDCS classes based on the rate of metabolism, instead of permeability used in the BCS system, and solubility (using dose number). There appears to be a correlation between BCS and BDDCS classes, but only for passively absorbed compounds[22]. With the growing number of compounds being identified as undergoing carrier mediated absorption, the comparison of BCS and BDDCS models could be complicated.

4.4 Comparison of BCS class assignments with the literature

The BCS validation set of 127 compounds contained both *in vitro* permeability and aqueous solubility collected from the literature. Based on the literature data, an observed BCS class was assigned to these compounds using our thresholds for permeability and solubility. Searching the literature, we found reported BCS classes for 71 of the 127 compounds in the validation set. From these 71, 10 compounds were cited in the literature to belong to more than one class and 16 were cited to belong to a different class from what we had assigned

them based on our solubility and permeability thresholds. Different assignments of BCS class to compounds in the literature have also been shown in other studies[63]. On closer inspection of these 16 compounds, the main differences between our assigned BCS class and the literature-assigned BCS class are the effect of maximum dose and pH which have not been considered in our work. In addition there are *in vitro – in vivo* differences due to varying levels of transporter expression in cell lines and gastrointestinal tract. As a result, some compounds that are poorly soluble and poorly permeable or highly permeable but poorly soluble *in vitro* may not necessarily be poorly absorbed *in vivo*. Examples include cinacalcet (Class IV), which is poorly soluble and poorly permeable but is absorbed >80% and dapsone (Class II) which is poorly soluble but it has a %HIA of 90%. The BCS validation set with the experimentally (*in vitro*) assigned and literature assigned compounds can be found in the **Supporting Information S1**. Concerning the 10 compounds cited as belonging to more than one class, it is interesting to see how the best models (those with the best overall accuracy and geometric accuracy, i.e. models 3 and 4 in **Table 6**) predicted these compounds, as their prediction may give more evidence to the assignment of these compounds to that class. For example based on our experimental data, ethosuximide is classified as belonging to class I, however the WHO guidelines state that the classification of this compound could be either class I or class III due to insufficient data on permeability. The models 3 and 4 from **Table 6** both predict that this compound is class I and this is supported by a %HIA of 93%. For the rest of the compounds, the majority are predicted into either one of the cited classes by models 3 and 4.

Using model 4 from **Table 6**, it is interesting to see which class was assigned to the compounds in BCS validation set. This can help understand the error rates associated with the model and the tendency of the model in relation to BCS class prediction. This confusion matrix comparing predicted versus observed BCS classes is shown in **Table 9**.

Table 9. Confusion matrix of model 4 from Table 6 for the prediction of BCS classes for the validation set

	Predicted Class 1	Predicted Class 2	Predicted Class 3	Predicted Class 4	Total	Accuracy (%)
Observed class 1	**28**	15**	6**	4**	54	52.8
Observed class 2	7*	**28**	1	4**	39	70.0
Observed class 3	4*	1	**17**	4**	26	65.4
Observed class 4	1*	2*	1*	**4**	8	50.0
Total Compounds	40	46	25	16		
Precision (%)	70.0	60.9	68.0	25.0		

*Type I errors
**Type II errors
Precision (%) is calculated for each class by adding the number of compounds in the column for that class and dividing by the total number of compounds (column total) for that class. Accuracy (%) is calculated by adding the number of compounds for each class in the row for that class and divided by the total number of compounds (row total) for that class.

Type I and type II errors were calculated for the values reported in **Table 9**. According to Khandelwal et al.[23], Type I errors (false positive error) represent those compounds that are either predicted class I when in fact they are observed to be BCS classes II-IV or predicted class II or III but are actually class IV compounds. Therefore the predicted class is biopharmaceutically more favourable than the observed actual class. Type II errors (false negative error) represent those compounds that are either predicted as class IV but were observed to be BCS classes I-III, or are predicted as class II or III but were observed to be class I. In other words, the predicted class is biopharmaceutically less favourable than the true class. The % of type I errors was 11.8 % and the % of type II errors was 25.9%. The results from a similar study by Pharm-The et al. (2013)[6] calculated type I and type errors II of 10.6% and 14.6% respectively, for their entire dataset (training and validation set) of 322 compounds.

It has been proposed that for BCS class prediction type II errors should be kept as low as possible[6]. This is quite obvious given that BCS class is used for the decision making regarding biopharmaceuticals experimentations required for oral dosage forms. Additionally, it might be more desirable to have good precision of class I compounds, rather than good accuracy, as these compounds are prioritised for biowaivers[3]. This principle of focussing on

precision rather than accuracy may be appropriate for class III compounds too, due to the increasing evidence for the suitability of class III compounds for biowaivers[64]. As seen in Table 9, both of the precision measures for class I and III were higher than the respective accuracy measures. Based on this, it is interesting to see that although class III is not the most popular represented class in the BCS validation set compared with classes I and II, it still has high class accuracy and precision.

It is important to state that the main difficulty for the models in this work was encountered in predicting class IV compounds. This was not entirely unexpected, since although the permeability and solubility datasets had balanced class distributions, the combination of these resulted in an under-representation of class IV. This may not be a major concern for industry; however, from a prediction point of view, by not considering the predictive accuracy of all classes can result in a higher number of misclassifications which could prove costly for industry[23]. This could be resolved by balancing all four BCS classes; however this can drastically reduce the number of compounds and potentially the models' ability to predict new compounds. Our work has utilised all data available and applied misclassification costs to attempt to overcome the BCS class imbalance. However, the poor prediction may not be down to the poor representation of classes and could be also a result of self-association in water, as cited in other research[22, 65].

5. Conclusion

The *in silico* prediction of a provisional BCS class is a challenging task. One of the challenging aspects of BCS class predictions is the potential effect of solubility on permeability prediction. Separate models of permeability and solubility fail to take into account the interactions between the class labels, and modelling each label separately reduces the generalisation for new compounds. It is well known in the literature that poor solubility can give rise to poor and variable absorption. Therefore, permeability prediction should include and so take into account the effects of solubility. Hence, using predicted solubility into permeability models alongside structural molecular descriptors, as performed in this work using the classifier chain multi-label classification method, avoids the disadvantage of other modelling methods for BCS prediction, like binary relevance multi-label classification.

This work has shown that the classifier chain multi-label method can greatly influence permeability models and hence provisional BCS using C&RT analysis. The use of predicted solubility as a descriptor to build and predict permeability, using the classifier chain method,

has been shown to improve a permeability model's predictive accuracy and in turn final provisional BCS prediction. The molecular descriptors used by both solubility and permeability models relate to lipophilicity, hydrogen bonding, polarity, size and shape; however their relationship with these properties is usually inversely related.

The benefit of the binary relevance and classifier chain methods over algorithm adaption methods is the utilisation of large datasets for permeability and solubility. There was no restriction to the dataset just because of missing values, as separate models for permeability and solubility were built based on the available data for each property. One limitation with this type of protocol is the lack of generalisation for the poorly represented class IV compounds. However, this can be improved slightly with the application of higher misclassification costs. The literature reveals a lack of multi-label classification methods for provisional prediction of BCS class suitable for a drug discovery scenario. Therefore, according to our results, the classifier chain method can be used successfully to improve the prediction of permeability class using predicted solubility.

Future extensions to this work would be to utilise more types of multi-label classification methods to perform consensus prediction similar to those in the literature[6], however the method must be able to include and use predicted solubility with the highest weighting in the permeability model.

In conclusion, this work has highlighted the potential benefit of using the classifier chain multi-label method, to predict provisional BCS class prediction for drug discovery.

Supporting Information

A list of 127 compounds in the BCS validation set and their solubility, permeability, and experimental and literature BCS assignments (S1), a list of 750 compounds, with collected solubility used is this work (S2), a list of molecular descriptors picked by the feature selection methods for solubility and permeability (S3) and finally all the C&RT decision trees produced from this work (S4). This information is available free of charge via the Internet at http://pubs.acs.org/.

Notes

The authors declare no competing financial interest.

References

1. Amidon, G. L.; Lennernas, H.; Shah, V. P.; Crison, J. R. A theoretical basis for a biopharmaceutic drug classification - The correlation of in-vitro drug product dissolution and in-vivo bioavalability. *Pharm. Res.* **1995,** *12*, (3), 413-420.
2. EMA *Guideline on the Investigation of Bioequivalence, Committee for Medicinal Products for Human Use (CHMP), http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2010/01/WC500070039.pdf [Accessed 11 September 2014]*; 2010.
3. CDER/FDA *Waiver of In Vivo Bioavailability and Bioequivalence Studies for Immediate-Release Solid Oral Dosage Forms Based on a Biopharmaceutics Classification System, U.S. Department of Health and Human Services - Center for Drug Evaluation and Research*; 2000.
4. Ku, M. S. Use of the biopharmaceutical classification system in early drug development. *AAPS Journal* **2008,** *10*, (1), 208-212.
5. Varma, M. V.; Gardner, I.; Steyn, S. J.; Nkansah, P.; Rotter, C. J.; Whitney-Pickett, C.; Zhang, H.; Di, L.; Cram, M.; Fenner, K. S.; El-Kattan, A. F. pH-Dependent Solubility and Permeability Criteria for Provisional Biopharmaceutics Classification (BCS and BDDCS) in Early Drug Discovery. *Mol. Pharm.* **2012,** *9*, (5), 1199-1212.
6. Pham-The, H.; Garrigues, T.; Bermejo, M.; González-Álvarez, I.; Monteagudo, M. C.; Cabrera-Pérez, M. Á. Provisional Classification and in Silico Study of Biopharmaceutical System Based on Caco-2 Cell Permeability and Dose Number. *Mol. Pharm.* **2013,** *10*, (6), 2445-2461.
7. Bergstrom, C. A. S.; Strafford, M.; Lazorova, L.; Avdeef, A.; Luthman, K.; Artursson, P. Absorption classification of oral drugs based on molecular surface properties. *J. Med. Chem.* **2003,** *46*, (4), 558-570.
8. Lennernas, H.; Abrahamsson, B. The use of biopharmaceutic classification of drugs in drug discovery and development: current status and future extension. *J. Pharm. Pharmacol.* **2005,** *57*, (3), 273-285.
9. Butler, J. M.; Dressman, J. B. The developability classification system: Application of biopharmaceutics concepts to formulation development. *J. Pharm. Sci* **2010,** *99*, (12), 4940-4954.
10. Wu, C. Y.; Benet, L. Z. Predicting drug disposition via application of BCS: Transport/absorption/elimination interplay and development of a biopharmaceutics drug disposition classification system. *Pharm. Res.* **2005,** *22*, (1), 11-23.
11. Ghafourian, T.; Newby, D.; Freitas, A. A. The impact of training set data distributions for modelling of passive intestinal absorption. *Int. J. Pharm.* **2012,** *436*, (1-2), 711-720.
12. Gozalbes, R.; Jacewicz, M.; Annand, R.; Tsaioun, K.; Pineda-Lucena, A. QSAR-based permeability model for drug-like compounds. *Bioorg. Med. Chem.* **2011,** *19*, (8), 2615-2624.
13. Gozalbes, R.; Pineda-Lucena, A. QSAR-based solubility model for drug-like compounds. *Bioorg. Med. Chem.* **2010,** *18*, (19), 7078-7084.
14. Carvalho, A. P. L. F.; Freitas, A., A Tutorial on Multi-label Classification Techniques. In *Foundations of Computational Intelligence Volume 5*, Abraham, A.; Hassanien, A.-E.; Snášel, V., Eds. p 177-195, Berlin, Springer: 2009; Vol. 205, pp 177-195.
15. Tsoumakas, G.; Katakis, I. Multi-label classification: An overview. *Int. J. Data Warehouse. Min.* **2007,** *3*, (3), 1-13.
16. Read, J.; Pfahringer, B.; Holmes, G.; Frank, E. Classifier chains for multi-label classification. *Mach. Learn.* **2011,** *85*, (3), 333-359.
17. McCallum, A. In *Multi-label text classification with a mixture model trained by EM*, AAAI'99 Workshop on Text Learning, 1999; pp 1-7.
18. Schapire, R. E.; Singer, Y. BoosTexter: A boosting-based system for text categorization. *Mach. Learn.* **2000,** *39*, (2-3), 135-168.
19. Schietgat, L.; Vens, C.; Struyf, J.; Blockeel, H.; Kocev, D.; Dzeroski, S. Predicting gene function using hierarchical multi-label decision tree ensembles. *BMC Bioinformatics* **2010,** *11*.
20. Shao, H.; Li, G.; Liu, G.; Wang, Y. Symptom selection for multi-label data of inquiry diagnosis in traditional Chinese medicine. *Sci. China Inf. Sci.* **2013,** *56*, (5), 1-13.

21. Michielan, L.; Terfloth, L.; Gasteiger, J.; Moro, S. Comparison of Multilabel and Single-Label Classification Applied to the Prediction of the Isoform Specificity of Cytochrome P450 Substrates. *J. Chem. Inf. Model.* **2009,** *49*, (11), 2588-2605.
22. Broccatelli, F.; Cruciani, G.; Benet, L. Z.; Oprea, T. I. BDDCS Class Prediction for New Molecular Entities. *Mol. Pharm.* **2012,** *9*, (3), 570-580.
23. Khandelwal, A.; Bahadduri, P. M.; Chang, C.; Polli, J. E.; Swaan, P. W.; Ekins, S. Computational models to assign biopharmaceutics drug disposition classification from molecular structure. *Pharm. Res.* **2007,** *24*, (12), 2249-2262.
24. Macheras, P.; Karalis, V. A non-binary biopharmaceutical classification of drugs: the ABGamma system. *Int J Pharm* **2014,** *464*, (1-2), 85-90.
25. Gonçalves, E. C.; Plastino, A.; Freitas, A. A. In *A Genetic Algorithm for Optimizing the Label Ordering in Multi-Label Classifier Chains*, Proceedings of the 2013 25th IEEE International Conference on Tools with Artificial Intelligence (ICTAI), 2013; IEEE Computer Society Conference Publishing Services (CPS): pp 469-476.
26. Lindenberg, M.; Kopp, S.; Dressman, J. B. Classification of orally administered drugs on the World Health Organization Model list of Essential Medicines according to the biopharmaceutics classification system. *Eur. J. Pharm. Biopharm.* **2004,** *58*, (2), 265-278.
27. Takagi, T.; Ramachandran, C.; Bermejo, M.; Yamashita, S.; Yu, L. X.; Amidon, G. L. A Provisional Biopharmaceutical Classification of the Top 200 Oral Drug Products in the United States, Great Britain, Spain, and Japan. *Mol. Pharm.* **2006,** *3*, (6), 631-643.
28. Dahan, A.; Miller, J.; Amidon, G. Prediction of Solubility and Permeability Class Membership: Provisional BCS Classification of the World's Top Oral Drugs. *AAPS J* **2009,** *11*, (4), 740-746.
29. Newby, D.; Freitas, A. A.; Ghafourian, T. Decision trees to characterise the roles of permeability and solubility on the prediction of oral absorption. *Eur. J. Med. Chem.* **2014,** *(Submitted)*.
30. Yalkowsky, S. H.; Dannenfelser, R. M., AQUASOL Database of Aqueous Solubility. University of Arizona,Tucson, Accessed Jan 2013, 1999.
31. Martindale, *Martindale The Complete Drug Reference*. 36 ed.; Pharmaceutical Press: London, 2009.
32. Kasim, N. A.; Whitehouse, M.; Ramachandran, C.; Bermejo, M.; Lennernäs, H.; Hussain, A. S.; Junginger, H. E.; Stavchansky, S. A.; Midha, K. K.; Shah, V. P. Molecular properties of WHO essential drugs and provisional biopharmaceutical classification. *Mol. Pharm.* **2004,** *1*, (1), 85-96.
33. Lipinski, C. A. Drug-like properties and the causes of poor solubility and poor permeability. *J. Pharmacol. Toxicol. Methods* **2000,** *44*, (1), 235-249.
34. Ghafourian, T.; Cronin, M. T. D. The impact of variable selection on the modelling of oestrogenicity. *SAR QSAR Environ. Res.* **2005,** *16*, (1-2), 171-190.
35. Newby, D.; Freitas, A. A.; Ghafourian, T. Pre-processing Feature Selection for Improved C&RT Models for Oral Absorption. *J. Chem. Inf. Model.* **2013,** *53*, (10), 2730-2742.
36. Breiman, L.; Friedman, J.; Stone, C. J.; Olshen, R. A., *Classification and Regression Trees*. 1 ed.; Chapman and Hall/CRC: Boca Raton, 1984.
37. Newby, D.; Freitas, A. A.; Ghafourian, T. Coping with Unbalanced Class Data Sets in Oral Absorption Models. *J. Chem. Inf. Model.* **2013,** *53*, (2), 461-474.
38. Schapire, R.; Singer, Y. BoosTexter: A Boosting-based System for Text Categorization. *Mach. Learn.* **2000,** *39*, (2-3), 135-168.
39. Zhu, S.; Ji, X.; Xu, W.; Gong, Y. In *Multi-labelled classification using maximum entropy method*, Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval, 2005; ACM: pp 274-281.
40. Yang, Y. An evaluation of statistical approaches to text categorization. *Inf. Retrieval.* **1999,** *1*, (1-2), 69-90.
41. Pham-The, H.; González-Álvarez, I.; Bermejo, M.; Garrigues, T.; Le-Thi-Thu, H.; Cabrera-Pérez, M. Á. The Use of Rule-Based and QSPR Approaches in ADME Profiling: A Case Study on Caco-2 Permeability. *Mol. Inform.* **2013,** *32*, (5-6), 459-479.

42. Duchowicz, P. R.; Talevi, A.; Bruno-Blanch, L. E.; Castro, E. A. New QSPR study for the prediction of aqueous solubility of drug-like compounds. *Bioorg. Med. Chem.* **2008,** *16*, (17), 7944-7955.
43. Chuasuwan, B.; Binjesoh, V.; Polli, J.; Zhang, H.; Amidon, G.; Junginger, H.; Midha, K.; Shah, V.; Stavchansky, S.; Dressman, J. Biowaiver monographs for immediate release solid oral dosage forms: Diclofenac sodium and diclofenac potassium. *J. Pharm. Sci.* **2009,** *98*, (4), 1206-1219.
44. Potthast, H.; Dressman, J.; Junginger, H.; Midha, K.; Oeser, H.; Shah, V.; Vogelpoel, H.; Barends, D. Biowaiver monographs for immediate release solid oral dosage forms: Ibuprofen. *J. Pharm. Sci.* **2005,** *94*, (10), 2121-2131.
45. MOE. QuaSAR-Descriptor help file [Online] Available: http://www.chemcomp.com/journal/descr.htm [Accessed 14 Jan 2014]. **2014**.
46. Gasteiger, J.; Marsili, M. Iterative partial equalization of orbital electronegativity - a rapid access to atomic charges. *Tetrahedron* **1980,** *36*, (22), 3219-3228.
47. Ghafourian, T.; Bozorgi, A. H. A. Estimation of drug solubility in water, PEG 400 and their binary mixtures using the molecular structures of solutes. *Eur. J. Pharm. Sci.***2010,** *40*, (5), 430-440.
48. Davies, N. M.; Teng, X. W.; Skjodt, N. M. Pharmacokinetics of rofecoxib. *Clin. Pharmacokinet.* **2003,** *42*, (6), 545-556.
49. Gazpio, C.; Sánchez, M.; García-Zubiri, I. X.; Vélaz, I.; Martínez-Ohárriz, C.; Martín, C.; Zornoza, A. HPLC and solubility study of the interaction between pindolol and cyclodextrins. *J. Pharm. Biomed. Anal.* **2005,** *37*, (3), 487-492.
50. Bergstrom, C. A. S.; Wassvik, C. M.; Norinder, U.; Luthman, K.; Artursson, P. Global and Local Computational Models for Aqueous Solubility Prediction of Drug-Like Molecules. *J. Chem Inform. Comput. Sci.* **2004,** *44*, (4), 1477-1488.
51. Bergstrom, C. A. S.; Norinder, U.; Luthman, K.; Artursson, P. Experimental and computational screening models for prediction of aqueous drug solubility. *Pharm. Res.* **2002,** *19*, (2), 182-188.
52. MOE *Molecular Operating Environment (MOE), v2012.10*, Chemical Computing Group Inc: Montreal, QC, 2012.
53. Cruciani, G.; Pastor, M.; Guba, W. VolSurf: a new tool for the pharmacokinetic optimization of lead compounds. *Eur. J. Pharm. Sci.* **2000,** *11*, S29-S39.
54. Sherer, E. C.; Verras, A.; Madeira, M.; Hagmann, W. K.; Sheridan, R. P.; Roberts, D.; Bleasby, K.; Cornell, W. D. QSAR Prediction of Passive Permeability in the LLC-PK1 Cell Line: Trends in Molecular Properties and Cross-Prediction of Caco-2 Permeabilities. *Mol. Inform.* **2012,** *31*, (3-4), 231-245.
55. Clark, D. E. Rapid calculation of polar molecular surface area and its application to the prediction of transport phenomena. 1. Prediction of intestinal absorption. *J. Pharm. Sci.* **1999,** *88*, (8), 807-814.
56. Salahinejad, M.; Le, T. C.; Winkler, D. A. Aqueous Solubility Prediction: Do Crystal Lattice Interactions Help? *Mol. Pharm.* **2013,** *10*, (7), 2757-2766.
57. Jain, N.; Yalkowsky, S. H. Estimation of the aqueous solubility I: Application to organic nonelectrolytes. *J. Pharm. Sci.* **2001,** *90*, (2), 234-252.
58. Wassvik, C. M.; Holmen, A. G.; Draheim, R.; Artursson, P.; Bergstrom, C. A. S. Molecular characteristics for solid-state limited solubility. *J. Med. Chem.* **2008,** *51*, (10), 3035-3039.
59. Nelson, T. M.; Jurs, P. C. Prediction of Aqueous Solubility of Organic Compounds. *J. Chem Inform. Comput. Sci.* **1994,** *34*, (3), 601-609.
60. Hewitt, M.; Cronin, M. T. D.; Enoch, S. J.; Madden, J. C.; Roberts, D. W.; Dearden, J. C. In Silico Prediction of Aqueous Solubility: The Solubility Challenge. *J. Chem. Inf. Model.* **2009,** *49*, (11), 2572-2587.
61. Nordqvist, A.; Nilsson, J.; Lindmark, T.; Eriksson, A.; Garberg, P.; Kihlén, M. A General Model for Prediction of Caco-2 Cell Permeability. *QSAR. Comb. Sci.* **2004,** *23*, (5), 303-310.
62. Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **1997,** *23*, (1-3), 3-25.
63. Bergstrom, C. A.; Andersson, S. B.; Fagerberg, J. H.; Ragnarsson, G.; Lindahl, A. Is the full potential of the biopharmaceutics classification system reached? *Eur J Pharm Sci* **2014,** *57*, 224-31.

64. Crison, J. R.; Timmins, P.; Keung, A.; Upreti, V. V.; Boulton, D. W.; Scheer, B. J. Biowaiver approach for biopharmaceutics classification system class 3 compound metformin hydrochloride using in silico modeling. *J. Pharm. Sci.* **2012,** *101*, (5), 1773-82.
65. Ross, D. L.; Riley, C. M. Aqueous solubilities of some variously substituted quinolone antimicrobials. *Int. J. Pharm.* **1990,** *63*, (3), 237-250.