Title	Machine learning methods for prediction of food effects on bioavailability: A comparison of Support Vector Machines and Artificial Neural Networks
Author(s)	Bennett-Lenane, Harriet; Griffin, Brendan T.; O'Shea, Joseph P.
Publication date	2021-09-24
Original citation	Bennett-Lenane, H., Griffin, B. T. and O'Shea, J. P. (2021) 'Machine learning methods for prediction of food effects on bioavailability: A comparison of support vector machines and artificial neural networks', European Journal of Pharmaceutical Sciences. doi: 10.1016/j.ejps.2021.106018
Type of publication	Article (peer-reviewed)
Link to publisher's version	http://dx.doi.org/10.1016/j.ejps.2021.106018 Access to the full text of the published version may require a subscription.
Rights	© 2021, the Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (https://creativecommons.org/licenses/by/4.0/) https://creativecommons.org/licenses/by/4.0/
Item downloaded from	http://hdl.handle.net/10468/12059

Journal Pre-proof

Machine Learning Methods for Prediction of Food Effects on Bioavailability: A Comparison of Support Vector Machines and Artificial Neural Networks

Harriet Bennett-Lenane , Brendan T. Griffin , Joseph P. O'Shea

Please cite this article as: Harriet Bennett-Lenane , Brendan T. Griffin , Joseph P. O'Shea , Machine Learning Methods for Prediction of Food Effects on Bioavailability: A Comparison of Support Vector Machines and Artificial Neural Networks, *European Journal of Pharmaceutical Sciences* (2021), doi: https://doi.org/10.1016/j.ejps.2021.106018

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Machine Learning Methods for Prediction of Food Effects on Bioavailability: A Comparison of Support Vector Machines and Artificial Neural Networks

Harriet Bennett-Lenane, Brendan T. Griffin, Joseph P. O'Shea[*]

School of Pharmacy, University College Cork, Cork, Ireland

[*]Correspondence to: Joseph O' Shea; Tel.: +353 (0)21 4901665, Email address: joseph.oshea@ucc.ie

Graphical Abstract
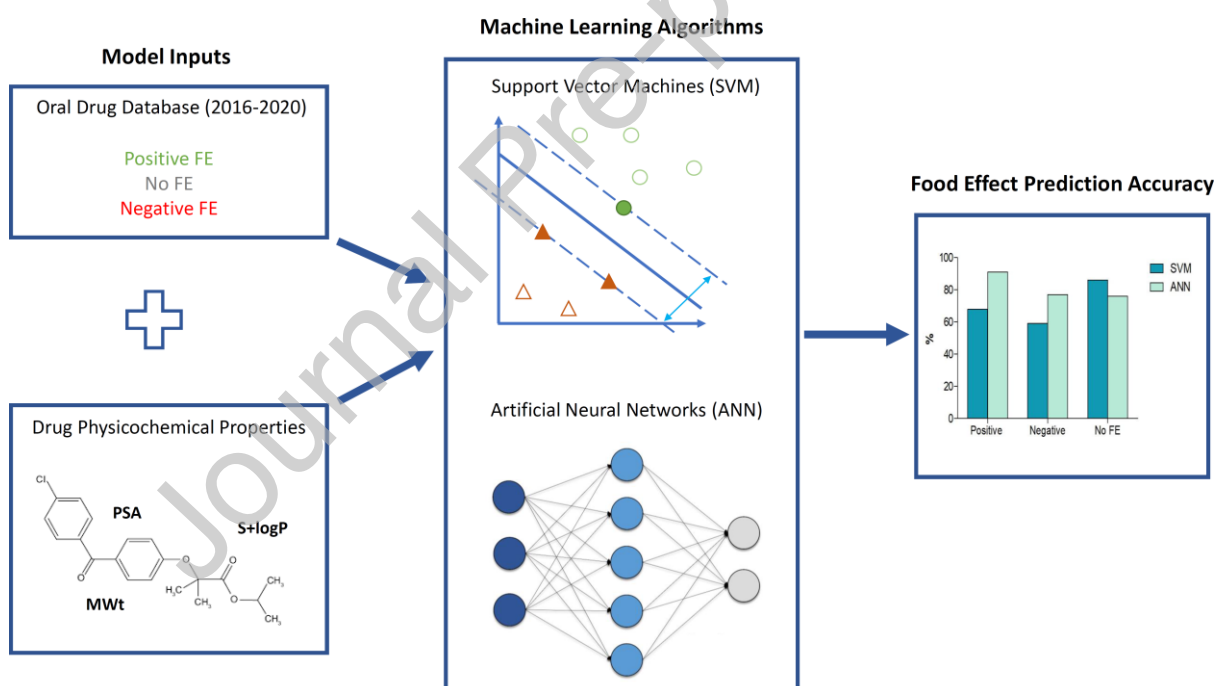
Abstract

Despite countless advances in recent decades across various *in vitro*, *in vivo* and *in silico* tools, anticipation of whether a drug will show a human food effect (FE) remains challenging. One means to predict potential FE involves probing any dependence between FE and drug properties. Accordingly, this study explored the potential for two machine learning (ML) algorithms to predict likely FE. Using a collated database of drugs licensed from 2016-2020, drugs were classified into three groups; positive,

negative or no FE. Greater than 250 drug properties were predicted for each drug which were used to train predictive models using Support Vector Machine (SVM) and Artificial Neural Network (ANN) algorithms. When compared, ANN outperformed SVM for FE classification upon training (82%, 72%) and testing (72%, 69%). Both models demonstrated higher FE prediction accuracy than the Biopharmaceutics Classification System (BCS) (46%). This exploratory work provided new insights into the connection between FE and drug properties as the Octanol Water Partition Coefficient (S+logP), Number of Hydrogen Bond Donors (HBD), Topological Polar Surface Area (T_PSA) and Dose (mg) were all significant for prediction. Overall, this study demonstrated the utility of ML to facilitate early anticipation of likely FE in pre-clinical development using four well-known drug properties.

Keywords

Oral Drug Delivery, Bio-Enabling Formulations, Food Effect, Machine Learning.

Abbreviations

FE, Food Effect; AUC, Area Under the Curve; C_{max}, peak plasma concentration; T_{max}, time to peak plasma concentration, BCS, Biopharmaceutics Classification System; BDDCS, Biopharmaceutical Drug Disposition Classification System; PBPK, Physiologically Based Pharmacokinetic Modelling; M&S, Modelling and Simulation, ML, Machine Learning; SVM, Support Vector Macines; ANN, Artifical Neural Networks, EMA, European Medicines Agency; FDA, Food and Drug Administration; MAD, Maximum Absorbable Dose; S+logP, the Octanol Water Partition Coefficient; HBD, Hydrogen Bond Donors; T_PSA, Topological Polar Surface Area; S+logD, Partition Coefficient pH 7.4; S+Sw, Aqueous Solubility; D/S, Dose/Solubility Ratio; MWt, Molecular Weight; RB, Rotatable Bonds; MCC, Matthews Correlation Coefficient; logHLC, logarithm of the air-water partition coefficient (Henry's Law Constant at 25°C); F_HBP, Population Average Number of Protons Available for Hydrogen Bonding Divided by the Number of Non-Hydrogen Atoms; P_{eff}, Human Jejunal Permeability; LBF, Lipid-Based Formulations.

Introduction

It is widely recognised that concomitant administration of oral dosage forms with food can alter drug pharmacokinetic profiles (1-3). As oral dosage forms are both widely and often chronically administered, understanding of the biological processes triggered by food consumption and its complex and drug-specific impact on oral bioavailability. The numerous underlying mechanisms by which food exerts this effect on drug absorption include physiological changes in pH, gastric emptying times, fluid volumes, bile salt concentrations and intestinal enzyme activity, in addition to specific food effects including binding, metabolism or interference with transporters (1, 3, 4). The clinical consequences of these changes are assessed through comparison of pharmacokinetic parameters describing the rate and extent of bioavailability i.e. peak plasma concentration (C_{max}), time to peak plasma concentration (T_{max}) and area under the curve (AUC) in both the fed and fasted state (5). A food effect (FE) is defined as when the 90% confidence intervals for the ratio of population geometric means, based on log transformed data, for either $AUC_{0\to\infty}$ or C_{max} fall outside the 80–125% bioequivalence limits relative to the same formulation administered in the fasted stated (5). These FE studies are subject to stringent regulatory requirements (5, 6), and the consequences of food-mediated effects on bioavailability have been widely reported (3, 7-12).

Previous research that 40% of drugs licensed between 2010 and 2017 displayed significant FE (1) suggests that within the current drug development paradigm, anticipation of the impact of food on drug absorption is pertinent. Moreover, in addition to guiding the design of improved formulations that are food effect resistant, this information is fundamental to optimise exposure of medicines with narrow therapeutic ranges in a clinical setting, to meet strict fed state bioequivalence study requirements of international regulatory authorities and reduce costs associated with product failures due to variability in exposure (1, 3, 7, 13). Consequently, the ability to predict and anticipate FE on oral drug delivery is of immense value to drug development. To date, extensive mechanistic tools spanning a wide range *in vitro, in vivo* and *in silico* methods to predict FE have been described in literature (14-18). Typically, drug performance under fasted and fed conditions are anticipated *in vitro* with dissolution tests in biorelevant media mimicking the human gastrointestinal environment (19-21) while *in vivo* predictions using various animal models (14, 16, 22), including canine and porcine examples, have also been employed, along with drug solubility testing in aspirated intestinal fluids (15, 23). However, despite the varying levels of success achieved by these methods, the multifaceted factors associated with drug, meal type and physiological conditions mean that, as of yet, no universally comprehensive approach for FE prediction has been found and that all current models exhibit limitations.

Owing to the limitations of *in vitro* methods, and the challenges in performing *in vivo* pharmacokinetic studies, *in silico* methods have emerged as the "go-to" approaches for predictive biopharmaceutics (24). Drug classification tools comprising rules of thumb based on drug biopharmaceutical properties,

and both the Biopharmaceutics and Biopharmaceutical Drug Disposition Classification Systems (BCS and BDDCS), provide simple guides to anticipate FE based upon related drug physicochemical properties (14, 25, 26). While such approaches provide a readily accessible prediction of likely FE, the relatively low accuracy and precision of these predictions has necessitated the development of mechanistic and data driven *in silico* models of FE (9). Physiologically based pharmacokinetic (PBPK) models lie at the centre of this diversifying modelling and simulation (M&S) toolbox, and have achieved increasingly accurate predictions of FE (27-30). Using mathematical equations to model physiological processes and anatomical parameters, these compartment-based absorption models mechanistically simulate a drugs plasma concentration-time profile in the fasted and fed states and can be applied in both pre-clinical studies with further potential application as decision-making aids for regulatory agencies (31). While PBPK models require comprehensive data about the physiological, biochemical, and physicochemical processes that occur, data-driven modelling tools which establish statistical relationships between FE and drug molecular properties, signify complementary additions to this ever-expanding M&S toolbox without the need for such. To date several attempts, spanning the last three decades, have been made to probe dependence between FE and drug physicochemical or molecular properties. These include quantitative linear correlations of AUC with individual drug properties (32) and a tool for computational biopharmaceutical profiling of ligands, based on predicted increases in fed state simulated intestinal fluid (FeSSIF) solubility to signal a positive FE (17). Finally, qualitative computational models predicting if a drug displays a positive, negative, or no FE according to the AUC_{fed}/AUC_{fasted} ratio, have also been published, using logistic regression (33) and more recently decision tree analysis (34).

Consequently, while modelling efforts to correlate the effect of food on AUC with drug properties exist, the application of multiple machine learning (ML) classification algorithms to predict FE remains unexplored. Support Vector Machine (SVM) classification and Artificial Neural Network (ANN) algorithms have been gaining interest across various facets of drug design and development, supporting streamlined and decision-based pre-clinical testing (35-41). SVM is a pattern recognition method widely used in data mining which works by finding an optimal separation line (hyperplane) which accurately separates and maximises the margins between two or more classes. Separation of non-linear data is also achievable through kernel functions which map the original data to a higher-dimensional "feature space" facilitating linear separation, as previously described (42, 43). While as a ML algorithm, ANN detects complex non-linear relationships between datasets, by mimicking basic human biological information processing methods. Adopting a general structure consisting of an input layer, hidden layer(s), output layer and using activation functions and connection weights, nodes (artificial neurons) from the input layer send data to the hidden and output layers through weighted connections ("synapses") to predict Y, as described in detail previously (44, 45). Using these ML algorithms, the broad objective of this work sought to develop a *in silico* model which could identify important drug properties for accurate FE prediction. Here, drugs were classified as displaying a positive, negative or no FE according to change in the extent of drug absorption (AUC) in the fed versus fasted states. Using a collated database of newly licensed drugs from 2016-2020, both SVM and ANN were employed to explore any relationship between the three FE classes and drug

properties. The study design facilitated investigation into the prevalence of FE among drugs licensed in the last 5 years, while also assessing if either ML algorithm or the BCS classification tool provided the highest prediction accuracy for this dataset. Accordingly, this study provides the first evidence of the capacity of ANN and SVM to facilitate early anticipation of likely FE.

Methods

Database collation

The drug database used in this study was obtained from drug products licensed in the European Medicines Agency (EMA) and Food and Drug Authority (FDA) between January 2016 and December 2020. The database of original new drug applications (NDA's) on the FDA website was searched by month from January 2016 to September 2020 along with the European Public Assessment Report (EPAR) of the EMA licensed products each year from 2016-2020. Exclusion criteria were any non-oral product, any biological product, any modified release product, any product where no FE information was available, any generic of a previously authorised product or any authorisation submission referring purely to changes of product indication. The products eligible for analysis were new molecular entities (NME), new combination products and reformulations of products which have been previously marketed. General information recorded included year of licensing, generic name, commercial name and any label restriction of the drug administration regarding food. Information regarding FE on absorption was obtained from the product EPAR or FDA label for each product where the ratios of AUC_{fed}/AUC_{fasted} were recorded. In cases where the documentation stated a product showed no change or a non-significant change in AUC_{fed}/AUC_{fasted}, with no values or ratios provided, a value of 1 was assigned. As FE information was obtained from regulatory submissions only, variables related to meal composition that might affect the interpretation of results were minimised.

Food Effect Classification

In this study drugs were predicted to belong to one of 3 FE classes designated according to the change in the extent of drug absorption in the fed versus fasted state (AUC_{fed}/AUC_{fasted}) alone, with "positive" referring to significantly increased and "negative" referring to significantly decreased extent of drug absorption in the fed state. This lone classification parameter was chosen as information regarding C_{max} was more frequently omitted from EPARs and FDA drug labels. Previous comparative studies also used this classification criterion, and the toxicity, efficacy and clinical significance of numerous drugs including those which are chronically dosed, correlates better with total exposure (AUC) than C_{max} (1, 32). The FE ratio (AUC_{fed}/AUC_{fasted}) was obtained for all drugs in the final dataset (141 drugs). A positive or negative FE was considered significant if the ratio fell outside 80–125% in reference to the currently accepted 90% CI for the ratio of population geometric means between fed and fasted treatments for concluding a lack of food-effect (5). Drugs with $AUC_{fed/fasted} > 1.25$ were classified with a "positive FE", $AUC_{fed/fasted}$ between 0.8-1.25 were deemed to have "no FE" and

$AUC_{fed/fasted}$ <0.8 a "negative FE". The final database of 141 drug compounds consisted of 44 Positive FE, 80 No FE and 17 Negative FE drugs.

Compilation of Physicochemical Descriptors

More than 250 descriptors for each drug were obtained from ADMET Predictor 9.5 (Simulations Plus, USA). Molecular structures were acquired as smiles from PubChem and used as inputs for the ADMET Predictor software (Version 9.5, Simulations Plus, California, USA) to calculate the molecular descriptors. A conscious effort was made to ensure repetition of drug properties found to be significantly correlated with food effect in previously published reports, to facilitate comparisons (32-34). The drug dosage strength used in the FE bioequivalence study for each drug was obtained from the EPAR or FDA product label respectively. A maximum absorbable dose (MAD) was calculated using a predicted fasted state simulated intestinal fluid (FaSSIF) solubility, again using ADMET Predictor, using the equation described previously (46). The Dose Solubility ratio was calculated using the dose as described above divided by an aqueous drug solubility predicted from ADMET Predictor.

Statistical analysis

Prior to employment of ML, to analyse any linear univariate correlations between FE classification and selected drug properties, a stepwise statistical analysis approach, as described previously (47), was adopted using SPSS (IBM Corporation, US) on the full drug database. It was hoped that these preliminary results could inform which properties may be significant for ML prediction. In brief, frequency distributions of the variables were graphed for each FE classification (positive, negative, no FE) and normality was checked visually with Q-Q and P-P plots. Ratios of samples sizes between the 3 groups were obtained. Variances of the datasets were analysed and compared to Levene's Test for Equality of Variances. A p-value <0.05 indicated a violation of equal variance. The null hypotheses were that no statistical differences were seen in a drug property between drug classes. Three separate comparisons were made i.e. Negative versus No FE, No FE versus Positive, Positive versus Negative. Comparison between groups were made using the t-test, Welch's test or Bootstrap independent samples test (5000 samples.) A p-value of 0.05 was used as the significance level for all tests. Boxplots were produced to facilitate visual interpretation of the data again using SPSS (IBM Corporation, US) and descriptive statistics including Median, Mean, Standard Deviation of Mean, Q1, Q3, Minimum, Maximum, Variance were obtained for each drug property for the 3 groups. The properties selected were S+logP (Octanol Water Partition Coefficient), HBD (Number of Hydrogen Bond Donors), HBA (Number of Hydrogen Bond Acceptors), T_PSA (Topological Polar Surface Area), Dose (mg), S+logD (Partition Coefficient pH 7.4), S+Sw (Aqueous Solubility), MAD (Maximum Absorbable Dose), D/S (Dose/Solubility Ratio), MWt (Molecular Weight) and RB (Rotatable Bonds).

BCS Classification

The BCS class of the drugs studied were obtained where available from the EPAR, FDA label or from literature. Fleisher et al. (4) previously described the general trend of FE on drug absorption (AUC) based on BCS classification where, BCS Class I compounds are likely to have no FE; BCS Class II compounds are likely to have a positive effect; BCS Class III compounds are likely to exhibit a negative effect while there is insufficient evidence for any clear identifiable trends for BCS Class IV compounds. Further separation of drugs within BCS classes as previously recognised (15), was not conducted to facilitate comparisons to a previously published analysis (33). Accordingly, the database was classified into these 3 FE categories, while BCS class 4 drugs were disregarded for this portion of the analysis.

Machine Learning/Model Development

FE classifications were predicted using two ML algorithms, ANN and SVM. To facilitate direct comparison of the predictive power of both algorithms the same training:test split was used. Principal Component Analysis (PCA) using the Unscrambler XI (Camo Analytics, US) was applied for a randomised assignment of training:test data. Training set criteria was that it covered the chemical space of the test set and ensured an almost equal representation of positive, negative and no FE drugs in the training set to avoid any potential for classification bias. Such imbalance in datasets has proved to be a widely reported obstacle to classification problems in ML in the past (42). In accordance with previous reports of classification prediction using ML algorithms (41, 48), initial variable reduction was conducted. Either a one-way ANOVA analysis with Tukey Multiple comparisons test (parametric) or Kruskal-Wallis analysis with Dunn's multiple comparison test (non-parametric) were applied to the training data. Variables with a p-value less than 0.05 for at least one class pair in the respective post tests were highlighted for further investigation. A correlation analysis of these identified variables was carried out with highly correlated variables clustered into the same group and the most significant variable of the group chosen for inclusion in the model development. Final models as well as BCS predictions were compared in terms of various accuracy statistics including, overall accuracy of prediction, as well as sensitivity, precision, specificity and Matthews Correlation Coefficient (MCC) for each FE class, as previously defined (49), where TP, FP, TN and FN refer to true positive, false positive, true negative and false negative results respectively. MCC was previously suggested as a reliable statistical metric for ML performance quality evaluation (50). A high MCC score (close to 1) is only achieved if the model obtained good results in all four confusion matrix metrics (TP, FP, TN, FN).

$$\text{Overall Accuracy:} \frac{\sum_{i=1}^{M} TP_i}{\sum_{i=1}^{M} TP_i + FN_i} \, x \, 100\%$$

$$\text{Sensitivity:} \frac{TP}{TP + FN} \, x \, 100\%$$

$$\text{Precision: } \frac{TP}{TP + FP} \; x \; 100\%$$

$$\text{Specificity: } \frac{TN}{TN + FP} \; x \; 100\%$$

$$\text{MCC: } \frac{(TP \; x \; TN) - (FP \; x \; FN)}{\sqrt{(TP + FN) \; x \; (TP + FP) \; x \; (TN + FN) \; x \; (TN + FP)}}$$

Support Vector Machine Classification

A SVM algorithm was used to build a classification prediction model using Unscrambler XI (Camo Analytics). Resulting variables from the initial variable reduction protocol were mean centred, de-identified and standardized through scaling by standard deviation. Variables were added one-by-one to assess which combinations produced the highest accuracy in both the training and test sets. C-SVC was used, where as part of the optimisation procedure the optimum values of key parameters of the regularization parameter C and gamma were sought from a grid search, performed across 10 orders of magnitude in logarithmic scale. In this grid search these key parameters were varied systematically to monitor which combination provided the highest classification accuracy in training and cross validation. This grid search was repeated for each variable combination using various kernel types (linear, polynomial, radial basis function and sigmoid) until the combination of kernel, C, gamma and input variables resulting in the best classification performance was obtained. The variables with the highest contribution to SVM classification were interpreted from the loadings plot of the PCA analysis which mapped the multidimensional data into a two-dimensional space.

Artificial Neural Networks

Multilayer Perception Artificial Neural Networks (MLP-ANN) were produced using SPSS Statistics (Version 26, IBM Corporation, US). The output layer consisted of three responses/categories: positive, negative or no FE. After initial variable reduction, as described above, the remaining significant variables were rescaled through standardisation where values were converted to their z-scores. Hyperbolic tangent was chosen as the activation function for the hidden layer, while an identity output function was used as the output layer activation function (51). Supervised learning using the Scaled Conjugate Gradient algorithm was chosen (52). Batch training was selected. Topologies with only one hidden layer were considered. The optimum number of neurons in the hidden layer was identified following a systematic trial-and-error approach were the number of neurons in the hidden layer were manually altered between 2 and 50, with runs being performed in triplicate. The optimal network size was chosen from the solution which resulted in the highest prediction accuracy in the training and test sets. All combinations of the significant variables were tested to ascertain which combination produced the highest prediction accuracy. This procedure was repeated until no improvement in model performance was observed. The relative contributions of each variable in the final network were elucidated from the Normalised Importance Chart.

Results

Analysis of Trends in Physicochemical Descriptors between Food Effect Classifications

A selection of well-known molecular and physicochemical properties of drugs licensed in the last five years were statistically compared with respect to the three FE classes. It was anticipated that as a result of this preliminary analysis of the entire dataset, any significant differences in properties between classes could inform which properties may be significant for subsequent ML prediction. Tabular results of the statistical analysis are shown in Supplementary Materials. A visual representation of properties which demonstrated significant differences is illustrated in Figure 1.

Upon statistical analysis, the properties of S+logP, HBD, T_PSA, Dose, S+Sw, S+logD, MWt, D/S and MAD were all significantly different for at least one pairing i.e. negative versus positive, negative versus no FE or positive versus no FE. Of these, 6 properties, namely S+logP, HBD, Dose, S+Sw, D/S and S+logD were found to be significantly different between drugs classified as having either a positive or negative FE. In particular, the mean S+logP value was observed to be almost 2.5 times greater for positive FE drugs. Conversely, it was observed that drugs with a positive FE could be differentiated from drugs which displayed no FE in terms of the 8 properties of S+logP, T_PSA, Dose, S+Sw, S+logD, MWt, D/S and MAD. Where the median dosage strength for positive FE drugs was four times higher than that of drugs with no FE (200mg vs. 50mg). Finally, drugs classified as either displaying a negative or no FE appeared more difficult to differentiate as only three properties were found to be statistically different for this pairwise comparison, namely S+logP, HBD and T_PSA, where the median number of HBD for negative FE drugs was double that of no FE (4 vs 2). Overall, it was observed that only one property, S+logP, was found to be significantly different between the all three classification groups (Figure 1). Furthermore, the properties of HBA, RB, were not found to be significantly different between any of the three pairwise comparisons.

Food Effect Prediction using BCS Classification

The dataset collated in this study consisted of 61/141 (43%) drugs displaying either a significant positive or negative FE. Previous reports indicate that BCS classifications (BCS Class I, II, III), using dose:solubility ratio and extent of absorption, can aid identification of likely FE, and such an approach was investigated to facilitate comparison of this simple classification rule of thumb versus data-driven ML techniques (4). In this study, using classifications based on the BCS, a poor overall accuracy of prediction (46%) was obtained. While the sensitivity of the method was acceptable for positive (87%) and negative (69%) FE drugs, relating to lower numbers of false negative classifications, the same could not be said for the no FE class where only 22% of drugs found to display no FE were BCS class 1 (Figure 2). While in terms of precision i.e. positive prediction rate, using the BCS tool, poor results were seen for positive (41%) and negative FE (35%) drugs with a poor rate of specificity also seen for

the positive FE group (49%). Finally, in terms of MCC, scores close to zero of 0.3 (positive), 0.2 (negative) and 0.2 (no FE) respectively were calculated (Figure 2). Therefore, these results suggest that use of the BCS tool to predict FE classification results in high relative amounts of either false positive or false negative predictions.

Applying ANN and SVM to Predict Food Effect Classification.

Two ML algorithms, ANN and SVM were employed to qualitatively predict if a drug displayed positive, negative or no FE. Using the SVM algorithm, the optimum model required 6 drug properties for prediction, namely S+logP, HBD, T_PSA, dose, logarithm of the air-water partition coefficient (Henry's Law Constant at 25°C) (logHLC) and population average number of protons available for hydrogen bonding divided by the number of non-hydrogen atoms (F_HBP). Using the radial basis function (RBF) kernel which outperformed the other kernel functions tested, results demonstrated that the SVM model could correctly predict FE classification of the training and set sets with 72% and 69% overall accuracy respectively (Figure 3). No FE drugs were predicted with the highest sensitivity upon both training (86%) and testing (71%) (Supplementary Materials), demonstrating better or comparative performance to ANN in these cases. Conversely, it was observed that the SVM algorithm was less successful in differentiating drugs with a negative FE compared to the ANN model described below (59%). In terms of precision, positive prediction rates of 79%, 91% and 60% were observed for the positive, negative and no FE groups (Figure 3). The SVM model showed the highest specificity in negative FE prediction (98%) where the lowest specificity was calculated for no FE drugs (69%). Intermediate MCC scores of 0.6, 0.7 and 0.5 were observed for the positive, negative and no FE classes respectively. Upon interpretation of the PCA correlations loading plot, HBD was the most influential property for classification prediction using this SVM-based model.

ANN model development resulted in an optimum three-layer feed forward network denoted MLP 4-13-3 (Figure 4). This network consisted of a single input layer with four descriptors, octanol water partition coefficient (S+logP), number of hydrogen bond donors (HBD), Topological Polar Surface Area (T_PSA) and Dose (mg), a single hidden layer with 13 nodes and an output layer with three output variables representing the 3 possible FE classifications (positive, negative, no FE). This network achieved high overall prediction accuracy in both the training (82%) and test sets (72%) (Figure 3). In terms of the individual classes, in contrast to the SVM model, drugs displaying positive FE could be distinguished the easiest displaying the highest sensitivity rates in both the training (91%) and test sets (73%), while the lowest sensitivity was seen for the no FE group (76% training, 71% test). For the positive, negative and no FE groups respectively, the precision of the predictions were 83%, 93% and 73%, specificity was 90%, 98% and 85% and higher MCC scores of 0.8, 0.8 and 0.6 closer to 1 were obtained (Figure 3). From the normalised importance chart, the most important property for prediction was S+logP, followed by T_PSA, HBD and Dose. However, all properties displayed over 67% normalised importance to prediction.

Overall, relatively similar overall prediction accuracy was achieved using both ML algorithms, as ANN marginally outperformed SVM. The proposed ANN network demonstrated equivalent or higher sensitivity, precision and specificity statistics for all but one metric for the three FE classes. In terms MCC, which have been reported as a more reliable overall performance evaluator, comparatively higher scores were observed for the ANN model. Resultantly, considering these superior and more consistent classification results along with the requirement for less input descriptors, it was concluded that the ANN algorithm produced the more robust model for this dataset.

Discussion

The significant effects of concomitant food intake on the pharmacokinetics of many drugs highlights the importance of FE predictions The capability to predict FE early in the drug development process can potentially expedite progression of drug products through pipelines, and streamline the regulatory approval process through identifying where clinical FE studies may or may not be required. As a result, the earliest possible elucidation of important drug properties which indicate a likely FE is vital as it was seen in this study that 43% of drugs licensed in the last 5 years reported either a significant positive or negative FE. While various efforts, dating as far back as 1996 (53), have been made to correlate FE with drug properties, the purpose of this study was to investigate if modern ML capabilities could predict the FE classification of drugs. As the dataset employed embodied drugs brought to market in the last 5 years, which are largely diverse in terms of their structural, physiochemical and pharmacokinetic properties, it was analysed if these ML algorithms could yield predictive tools relevant for the contemporary drug development landscape.

Our study demonstrated the capability of ML algorithms to predict FE classification based on physicochemical drug descriptors. Upon comparison of the two ML algorithms, results suggested that while both approaches demonstrated capacity for prediction, ANN outperformed SVM. The optimum ANN, containing 1 hidden layer of 13 nodes and using 4 input properties (S+logP, T_ PSA, HBD, Dose), yielded strong overall accuracy in classification for both training (82%) and test sets (72%). Comparatively, the SVM model which employed 6 input properties (S+logP, T_PSA, HBD, Dose, F_HBP, logHLC) for prediction, did also demonstrate relatively good accuracy upon training (72%) and testing (69%). However, in terms of the performance metrics analysed i.e. precision, sensitivity, specificity and MCC, the ANN model proved either superior or equivalent to SVM in 11/12 cases, including MCC scores closer to 1 for all three classes. The exact reason for the improvement in performance metrics and requirement for less inputs for ANN, while difficult to pinpoint, may be attributable to the varying mathematical approaches used by both methods to classify this non-linearly separable dataset and map the data to higher dimensional spaces, i.e. kernel tricks (SVM) and non-linear activation functions (ANN) (44, 49, 54, 55). While the comparative accuracy of SVM versus ANN modelling is dependent on the specific dataset involved, the easily interpretable, readily obtainable, and widely recognisable nature of the four properties used in the ANN model to formulation scientists expands its applicability in a preclinical setting, as at this time only limited resources and preliminary information are available regarding a model compound. Accordingly, this study supports the application of ML algorithms, in particular ANN, to provide accessible tools for FE prediction of newly licensed drugs.

Our models identified key physicochemical and molecular properties which contribute to the classification of newly licensed drugs according to the likely effect of food on extent of drug absorption. While the ANN model found four properties of S+logP, Dose, HBD and T_PSA to be noteworthy, using SVM the properties of logHLC and F_HBP were also required. The inclusion of S+logP as the most important property for ANN prediction matches observations from our preliminary statistical analysis. There S+logP, a widely used drug parameter used in PBPK modelling (56), was the only property significantly different between all three pairwise comparisons (Figure 1). While another type of partition coefficient, logHLC was also significant for SVM modelling, S+logP or increasing drug lipophilicity in particular has previously been linked with increased susceptibility to positive FE (17, 32, 57), through the increased dissolution and solubilisation effects of food for lipophilic drugs. This is a partition property, while a hypothesis for its significance may relate to ingestion of fatty foods resulting in greater partitioning of lipophilic drugs between the digestion phases which could be related to the kinetics of partition between octanol and water. Using a cohort of pre-2005 drugs, Singh et al. found a significant positive correlation between AUC ratio and logP, where the positive effect of food was more pronounced for lipophilic drugs (32). In agreement, in this study, we observed that 41% of drugs with a positive FE had S+logP values >4, with 84% >2. Gu et al. and Omachi et al. have also successfully utilised logP or the closely related logD as a critical parameter to predict FE on AUC using both logistic regression and decision tree analysis (33, 34). Nevertheless, while the significance of S+logP in our model suggests the importance of a lipophilicity indicator to predict FE on AUC, caution has previously been advised in terms of using this blanket generalisation of likely positive FE for all PWSD as exceptions exist (32). In addition to its importance as a partition property, the requirement for S+logP may also reflect its significance to predict other biopharmaceutical properties such as membrane permeability, as it is possible that some dietary foods may affect permeability, however, it is unlikely that this would be unique to certain drugs. Lipophilicity has previously been correlated with permeability measurements for compounds which are passively transported (58-62).

In this study S+logP alone did not prove sufficient for accurate prediction of FE, as other properties including HBD and T_PSA were found in the final models. While these were not previously significant parameters for FE classification using logistic regression (33), using this contemporary cohort of drugs ML identified their predictive abilities. Both parameters, have been formerly associated with membrane permeability (59), as poor permeation is suggested to be likely with drugs of PSA >140Å (63, 64), where T_PSA is the commonly used PSA descriptor (17), and an excessive number of HBD groups impairs permeability across a membrane bilayer (65-67), as specified by a cut-off of 5 in Lipinski's rule of 5 developability criteria (59). Both parameters were used previously along with logP, to predict human jejunal permeability (P_{eff}) using partial least squares regression (60). It could be suggested that the role of both T_PSA and HBD in the ML models was to aid identification of drugs with negative FE as HBD and T_PSA demonstrated the highest mean values for the negative FE class upon preliminarily statistical analysis (Figure 1). This finding correlates well to the Fleisher et al. summary of FE prediction based on BCS class where Class 3 drugs, of poor permeability and good

solubility, are likely to exhibit negative FE (4). Such negative food effects were previously associated with highly hydrophilic drugs displaying a narrow window of absorption (18, 33).

Dose was also a significant parameter in classify drugs by food effects on AUC. Upon early statistical analysis, it was observed that over 18% of drugs displaying a positive FE used a dose >500 mg in their respective pharmacokinetic studies, compared to 0% and 6.25% of negative and no FE drugs respectively, with 59% of this positive class using a dose >200 mg. While the exact reason for this substantial difference in dosage strength between classes is unknown, previously a logistic regression model found dose number and MAD to be significant for FE classification (33). However, when tested neither were found to be significant in our ML models. In addition, the Dose/Solubility ratio of a drug was previously significantly correlated with AUC however, in our analysis neither this ratio nor S+Sw, despite displaying significant differences between classes in our preliminary statistics, were important for ML prediction. It could be hypothesised that this dose parameter aided the prediction of positive FE drugs due to its ability to differentiate positive FE drugs from both negative and no FE classes in our initial statistical analysis (Figure 1). Overall, the properties most significant for FE classification demonstrate the importance of both solubility and permeability for FE prediction. These properties reflect widely known drug-likeness filters, drug classification systems and properties used in previous models for FE prediction. It is likely that any differences in significant properties compared with previous publications reflect the different datasets of drugs used to build the respective models, as this study reflects the most contemporary drug products licensed in the last 5 years.

As previously stated, in terms of accuracy the ANN model outperformed SVM for FE classification. However, when compared to predictions using BCS class, both ML models performed strongly as the overall BCS accuracy (46%) was substantially weaker than that of the ML algorithms. BCS classifications appeared inadequate compared to ML in terms of consistency in precision, sensitivity, specificity and MCC scores. Owing in part to a large number of false positive predictions (58 drugs), this BCS accuracy of 46% is substantially lower than its previous 67% accuracy in classifying FE for a database of pre-2007 drugs (33). Possible reasons for this inaccuracy compared to pre-2007 drugs are unclear but may reflect that the drugs licensed between 2016-2020 represent a different chemical design space. While direct accuracy comparisons with the ML models are not achievable as no trends in FE have been suggested for BCS class IV drugs, perhaps improved predictions with the BCS tool could be obtained going forward with further subdivision of categories into weak acids, weak bases, and lipophilic compounds (15), however this was not within the scope of this study. Additionally, the accuracy of the ANN model compared favourably to accuracies of 80% and 66% achieved in previous computational approaches to predict FE classification (33, 34). Ultimately, direct comparisons of prediction accuracies are unachievable due to the absence of external test sets in these previous publications and differences in the datasets used to build the respective models. However, this work identifies ANN as an accurate and efficient solution in detecting correlations between FE and drug properties from a cohort of newly licensed drugs using only 4 well-known drug properties. In the grand paradigm of drug development and computational pharmaceutics this model would provide early indication if significant FE are relevant, allowing informed decisions in drug development. Including

whether redesign of the drug candidate may be relevant based on drug properties or whether a bio-enabling strategy should be applied, prompting subsequent application of other computational models such a previous example from our own group (68-71), which would give an indication as to whether a lipid-based formulation (LBF) or other formulation strategies which are reported to overcome FE, may offer a high likelihood of success.

The effects of food on bioavailability are multifaceted, involving physiological, physicochemical and biochemical mechanisms, and as a result, similar to all current methods to anticipate FE, limitations of our models are acknowledged. By design, data-driven modelling approaches identify correlations between individual drug properties of the dataset and classifications, thereby facilitating prediction of the general physiological changes exert by food on drug absorption. However, such an approach, limited by the physiochemical parameters employed, cannot capture all contributing factors drug specific FE such as effect on metabolism, bile-micelle binding, specific chelation between food and a drug or activity of a specific transporter or enzyme (1, 14, 15). As understanding of such effects continues to grow across future studies, in line with continued improvements and understanding of other approaches of FE prediction, opportunities will exist to broaden the utility of these predictions to incorporate such factors. This aside, these current models successfully predict FE category, providing evidence of a computational tool suited for easy integration within current pre-clinical drug development.

Conclusion

In this study, innovative predictive models using two ML algorithms (SVM and ANN) were developed which accurately predicted the FE category of drugs licensed between 2016-2020, of which 43% demonstrated significant FE. These models were found to possess greater prediction accuracy than FE predictions using the BCS criteria and performed strongly upon comparison to previously published tools using older drug datasets. This predictive modelling enabled key physiochemical parameters that contribute to the effect of food on the extent of drug absorption to be identified, namely S+logP, T_PSA, HBD and Dose. Therefore, this exploratory work provides a further mechanistic basis to understand a drugs behaviour in fed and fasted conditions using a contemporary cohort of licensed drugs. Of course, the rationale and requirements for FE determination will differ depending on the stage of drug development, be that preliminarily formulation testing or the investigation of specific pharmacokinetic parameters. Regardless, the ML tools, particularly the ANN produced in this study can facilitate screening of drug candidates for potential FE, with little cost and effort at the early stages of drug development, utilising only easily recognisable drug properties.

mmc1.pdf

Credit Author Statement:

Harriet Bennett-Lenane: Writing - Original draft, Methodology, Computational Modelling, Conceptualization.

Brendan T. Griffin, Joseph P. O'Shea: Supervision and Editing.

Acknowledgements

References

1.   O'Shea JP, Holm R, O'Driscoll CM, Griffin BT. Food for thought: formulating away the food effect - a PEARRL review. J Pharm Pharmacol. 2019;71(4):510-35.
2.   Welling PG. Effects of food on drug absorption. Annu Rev Nutr. 1996;16:383-415.
3.   Koziolek M, Alcaro S, Augustijns P, Basit AW, Grimm M, Hens B, et al. The mechanisms of pharmacokinetic food-drug interactions – A perspective from the UNGAP group. European Journal of Pharmaceutical Sciences. 2019;134:31-59.
4.   Fleisher D, Li C, Zhou Y, Pao LH, Karim A. Drug, meal and formulation interactions influencing drug absorption after oral administration. Clinical implications. Clin Pharmacokinet. 1999;36(3):233-54.
5.   FDA. Food-effect bioavailability and fed

bioequivalence studies: guidance for industry. 2002 [Available from: Guidances/UCM126833.pdf.
6.   Farha M, Masson E, Tomkinson H, Mugundu G. Food Effect Study Design With Oral Drugs: Lessons Learned From Recently Approved Drugs in Oncology. J Clin Pharmacol. 2019;59(4):463-71.
7.   Varum FJ, Hatton GB, Basit AW. Food, physiology and drug delivery. Int J Pharm. 2013;457(2):446-60.
8.   Yasuji T, Kondo H, Sako K. The effect of food on the oral bioavailability of drugs: a review of current developments and pharmaceutical technologies for pharmacokinetic control. Ther Deliv. 2012;3(1):81-90.
9.   Yan J-H. Food Effect on Oral Bioavailability: Old and New Questions. Clinical Pharmacology in Drug Development. 2017;6(4):323-30.
10.   Zhuang X, Lu C. PBPK modeling and simulation in drug research and development. Acta Pharmaceutica Sinica B. 2016;6(5):430-40.

11.   Toothaker RD, Welling PG. The Effect of Food on Drug Bioavailability. Annual Review of Pharmacology and Toxicology. 1980;20(1):173-99.

12.   Koziolek M, Schneider F, Grimm M, Modeβ C, Seekamp A, Roustom T, et al. Intragastric pH and pressure profiles after intake of the high-caloric, high-fat meal as used for food effect studies. J Control Release. 2015;220(Pt A):71-8.

13.   Savla R, Browne J, Plassat V, Wasan KM, Wasan EK. Review and analysis of FDA approved drugs using lipid-based formulations. Drug Dev Ind Pharm. 2017;43(11):1743-58.

14.   Zhang T, Wells E. A Review of Current Methods for Food Effect Prediction During Drug Development. Current Pharmacology Reports. 2020;6(5):267-79.

15.   Lentz KA. Current methods for predicting human food effect. Aaps j. 2008;10(2):282-8.

16.   Christiansen ML, Müllertz A, Garmer M, Kristensen J, Jacobsen J, Abrahamsson B, et al. Evaluation of the Use of Göttingen Minipigs to Predict Food Effects on the Oral Absorption of Drugs in Humans. Journal of Pharmaceutical Sciences. 2015;104(1):135-43.

17.   Bergstrom CAS, Charman WN, Porter CJH. Computational prediction of formulation strategies for beyond-rule-of-5 compounds. Adv Drug Deliv Rev. 2016;101:6-21.

18.   Kawai Y, Fujii Y, Tabata F, Ito J, Metsugi Y, Kameda A, et al. Profiling and trend analysis of food effects on oral drug absorption considering micelle interaction and solubilization by bile micelles. Drug Metab Pharmacokinet. 2011;26(2):180-91.

19.   Borbás E, Kádár S, Tsinman K, Tsinman O, Csicsák D, Takács-Novák K, et al. Prediction of Bioequivalence and Food Effect Using Flux- and Solubility-Based Methods. Mol Pharm. 2019;16(10):4121-30.

20.   Shono Y, Jantratid E, Janssen N, Kesisoglou F, Mao Y, Vertzoni M, et al. Prediction of food effects on the absorption of celecoxib based on biorelevant dissolution testing coupled with physiologically based pharmacokinetic modeling. European Journal of Pharmaceutics and Biopharmaceutics. 2009;73(1):107-14.

21.   Gamsiz ED, Ashtikar M, Crison J, Woltosz W, Bolger MB, Carrier RL. Predicting the effect of fed-state intestinal contents on drug dissolution. Pharm Res. 2010;27(12):2646-56.

22.   Henze LJ, Koehl NJ, O'Shea JP, Holm R, Vertzoni M, Griffin BT. Toward the establishment of a standardized pre-clinical porcine model to predict food effects – Case studies on fenofibrate and paracetamol. International Journal of Pharmaceutics: X. 2019;1:100017.

23.   Bennett-Lenane H, Jørgensen JR, Koehl NJ, Henze LJ, O'Shea JP, Müllertz A, et al. Exploring porcine gastric and intestinal fluids using microscopic and solubility estimates: Impact of placebo self-emulsifying drug delivery system administration to inform bio-predictive in vitro tools. Eur J Pharm Sci. 2021;161:105778.

24.   Lennernäs H, Aarons L, Augustijns P, Beato S, Bolger M, Box K, et al. Oral biopharmaceutics tools - time for a new initiative - an introduction to the IMI project OrBiTo. Eur J Pharm Sci. 2014;57:292-9.

25.   Heimbach T, Xia B, Lin TH, He H. Case studies for practical food effect assessments across BCS/BDDCS class compounds using in silico, in vitro, and preclinical in vivo data. Aaps j. 2013;15(1):143-58.

26.   Custodio JM, Wu CY, Benet LZ. Predicting drug disposition, absorption/elimination/transporter interplay and the role of food on drug absorption. Adv Drug Deliv Rev. 2008;60(6):717-33.

27.   Cheng L, Wong H. Food Effects on Oral Drug Absorption: Application of Physiologically-Based Pharmacokinetic Modeling as a Predictive Tool. Pharmaceutics. 2020;12(7).

28.   O'Shea JP, Faisal W, Ruane-O'Hora T, Devine KJ, Kostewicz ES, O'Driscoll CM, et al. Lipidic dispersion to reduce food dependent oral bioavailability of fenofibrate: In vitro, in vivo and in silico assessments. Eur J Pharm Biopharm. 2015;96:207-16.

29.   Gao H, Wang W, Dong J, Ye Z, Ouyang D. An integrated computational methodology with data-driven machine learning, molecular modeling and PBPK modeling to accelerate solid dispersion formulation design. Eur J Pharm Biopharm. 2021;158:336-46.

30.   Riedmaier AE, DeMent K, Huckle J, Bransford P, Stillhart C, Lloyd R, et al. Use of Physiologically Based Pharmacokinetic (PBPK) Modeling for Predicting Drug-Food Interactions: an Industry Perspective. The AAPS Journal. 2020;22(6):123.

31.   Kesisoglou F. Can PBPK Modeling Streamline Food Effect Assessments? J Clin Pharmacol. 2020;60 Suppl 1:S98-s104.

32.   Singh BN. A quantitative approach to probe the dependence and correlation of food-effect with aqueous solubility, dose/solubility ratio, and partition coefficient (Log P) for orally active drugs administered as immediate-release formulations. Drug Development Research. 2005;65(2):55-75.

33.   Gu CH, Li H, Levons J, Lentz K, Gandhi RB, Raghavan K, et al. Predicting effect of food on extent of drug absorption based on physicochemical properties. Pharm Res. 2007;24(6):1118-30.

34.   Omachi F, Kaneko M, Iijima R, Watanabe M, Itagaki F. Relationship between the effects of food on the pharmacokinetics of oral antineoplastic drugs and their physicochemical properties. J Pharm Health Care Sci. 2019;5:26.

35.   Aksu B, Paradkar A, de Matas M, Ozer O, Güneri T, York P. Quality by design approach: application of artificial intelligence techniques of tablets manufactured by direct compression. AAPS PharmSciTech. 2012;13(4):1138-46.

36.   Damiati SA, Martini LG, Smith NW, Lawrence MJ, Barlow DJ. Application of machine learning in prediction of hydrotrope-enhanced solubilisation of indomethacin. International Journal of Pharmaceutics. 2017;530(1):99-106.

37.   Manda A, Walker RB, Khamanga SMM. An Artificial Neural Network Approach to Predict the Effects of Formulation and Process Variables on Prednisone Release from a Multipartite System. Pharmaceutics. 2019;11(3).

38.   Zhang Y, Evans JRG, Yang S. Exploring Correlations Between Properties Using Artificial Neural Networks. Metallurgical and Materials Transactions A. 2020;51(1):58-75.

39.   DeBoyace K, Wildfong PLD. The Application of Modeling and Prediction to the Formation and Stability of Amorphous Solid Dispersions. J Pharm Sci. 2018;107(1):57-74.

40.   Alhalaweh A, Alzghoul A, Mahlin D, Bergström CAS. Physical stability of drugs after storage above and below the glass transition temperature: Relationship to glass-forming ability. Int J Pharm. 2015;495(1):312-7.

41.   Alhalaweh A, Alzghoul A, Kaialy W, Mahlin D, Bergström CAS. Computational Predictions of Glass-Forming Ability and Crystallization Tendency of Drug Molecules. Molecular Pharmaceutics. 2014;11(9):3123-32.

42.   Cervantes J, Garcia-Lamont F, Rodríguez-Mazahua L, Lopez A. A comprehensive survey on support vector machine classification: Applications, challenges and trends. Neurocomputing. 2020;408:189-215.

43.   Kausar N, Belhaouari Samir B, Abdullah A, Ahmad I, Hussain M, editors. A Review of Classification Approaches Using Support Vector Machine in Intrusion Detection. Informatics Engineering and Information Science; 2011 2011//; Berlin, Heidelberg: Springer Berlin Heidelberg.

44.   Bourquin J, Schmidli H, van Hoogevest P, Leuenberger H. Basic concepts of artificial neural networks (ANN) modeling in the application to pharmaceutical development. Pharm Dev Technol. 1997;2(2):95-109.

45.   Tu JV. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. Journal of Clinical Epidemiology. 1996;49(11):1225-31.

46.   Ding X, Rose JP, Van Gelder J. Developability assessment of clinical drug products with maximum absorbable doses. Int J Pharm. 2012;427(2):260-9.

47.   Bennett-Lenane H, O'Shea J. P., O'Driscoll, C. M., Griffin, B. T. A Retrospective Biopharmacetical Analysis of >800 Approved Oral Products: Are Drug Properties of Solid Dispersions and Lipid-Based Formulations Distinctive? J Pharm Sci (In Review)2021.

48.   Alhalaweh A, Alzghoul A, Mahlin D, Bergström CAS. Physical stability of drugs after storage above and below the glass transition temperature: Relationship to glass-forming ability. International Journal of Pharmaceutics. 2015;495(1):312-7.

49.   Bisgin H, Bera T, Ding H, Semey HG, Wu L, Liu Z, et al. Comparing SVM and ANN based Machine Learning Methods for Species Identification of Food Contaminating Beetles. Scientific Reports. 2018;8(1):6532.

50.   Chicco D, Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genomics. 2020;21(1):6.

51.   Alshalif SA, Ibrahim N, Herawan T, editors. Artificial Neural Network with Hyperbolic Tangent Activation Function to Improve the Accuracy of COCOMO II Model. Recent Advances on Soft Computing and Data Mining; 2017 2017//; Cham: Springer International Publishing.

52.   Møller MF. A scaled conjugate gradient algorithm for fast supervised learning. Neural Networks. 1993;6(4):525-33.

53.   Milton KA, DJ N. The predictability of food effects on bioavailability (BA). Eur J Pharm Sci. 1996;4(Suppl. 1):S94.

54.   Ren J. ANN vs. SVM: Which one performs better in classification of MCCs in mammogram imaging. Knowledge-Based Systems. 2012;26:144-53.

55.   Kavzoglu T, Colkesen I. A kernel functions analysis for support vector machines for land cover classification. International Journal of Applied Earth Observation and Geoinformation. 2009;11(5):352-9.

56.   Jones HM, Gardner IB, Watson KJ. Modelling and PBPK simulation in drug discovery. Aaps j. 2009;11(1):155-66.

57.   Marasanapalle VP, Boinpally RR, Zhu H, Grill A, Tang F. Correlation between the systemic clearance of drugs and their food effects in humans. Drug Dev Ind Pharm. 2011;37(11):1311-7.

58.   Kristl A, Tukker JJ. Negative correlation of n-octanol/water partition coefficient and transport of some guanine derivatives through rat jejunum in vitro. Pharm Res. 1998;15(3):499-501.

59.   Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Advanced Drug Delivery Reviews. 1997;23(1):3-25.

60.   Winiwarter S, Bonham NM, Ax F, Hallberg A, Lennernäs H, Karlén A. Correlation of Human Jejunal Permeability (in Vivo) of Drugs with Experimentally and Theoretically Derived Parameters. A Multivariate Data Analysis Approach. Journal of Medicinal Chemistry. 1998;41(25):4939-49.

61.   Artursson P, Karlsson J. Correlation between oral drug absorption in humans and apparent drug permeability coefficients in human intestinal epithelial (Caco-2) cells. Biochem Biophys Res Commun. 1991;175(3):880-5.

62.   Takagi T, Ramachandran C, Bermejo M, Yamashita S, Yu LX, Amidon GL. 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):631-43.

63.   Pajouhesh H, Lenz GR. Medicinal chemical properties of successful central nervous system drugs. NeuroRx. 2005;2(4):541-53.

64.   Palm K, Stenberg P, Luthman K, Artursson P. Polar Molecular Surface Properties Predict the Intestinal Absorption of Drugs in Humans. Pharmaceutical Research. 1997;14(5):568-71.

65.   Abraham MH, Chadha HS, Whiting GS, Mitchell RC. Hydrogen bonding. 32. An analysis of water-octanol and water-alkane partitioning and the delta log P parameter of seiler. J Pharm Sci. 1994;83(8):1085-100.

66.   Paterson DA, Conradi RA, Hilgers AR, Vidmar TJ, Burton PS. A Non-aqueous Partitioning System for Predicting the Oral Absorption Potential of Peptides. Quantitative Structure-Activity Relationships. 1994;13(1):4-10.

67.   Norinder U, Bergström CA. Prediction of ADMET Properties. ChemMedChem. 2006;1(9):920-37.

68.   Bennett-Lenane H, Koehl NJ, O'Dwyer PJ, Box KJ, O'Shea JP, Griffin BT. Applying Computational Predictions of Biorelevant Solubility Ratio Upon Self-Emulsifying Lipid-Based Formulations Dispersion to Predict Dose Number. J Pharm Sci. 2020.

69.	Persson LC, Porter CJ, Charman WN, Bergstrom CA. Computational prediction of drug solubility in lipid based formulation excipients. Pharm Res. 2013;30(12):3225-37.
70.	Alsenz J, Kuentz M. From Quantum Chemistry to Prediction of Drug Solubility in Glycerides. Molecular Pharmaceutics. 2019;16(11):4661-9.
71.	Kuentz M, Imanidis G. In silico prediction of the solubility advantage for amorphous drugs - Are there property-based rules for drug discovery and early pharmaceutical development? Eur J Pharm Sci. 2013;48(3):554-62.

Figure Legends
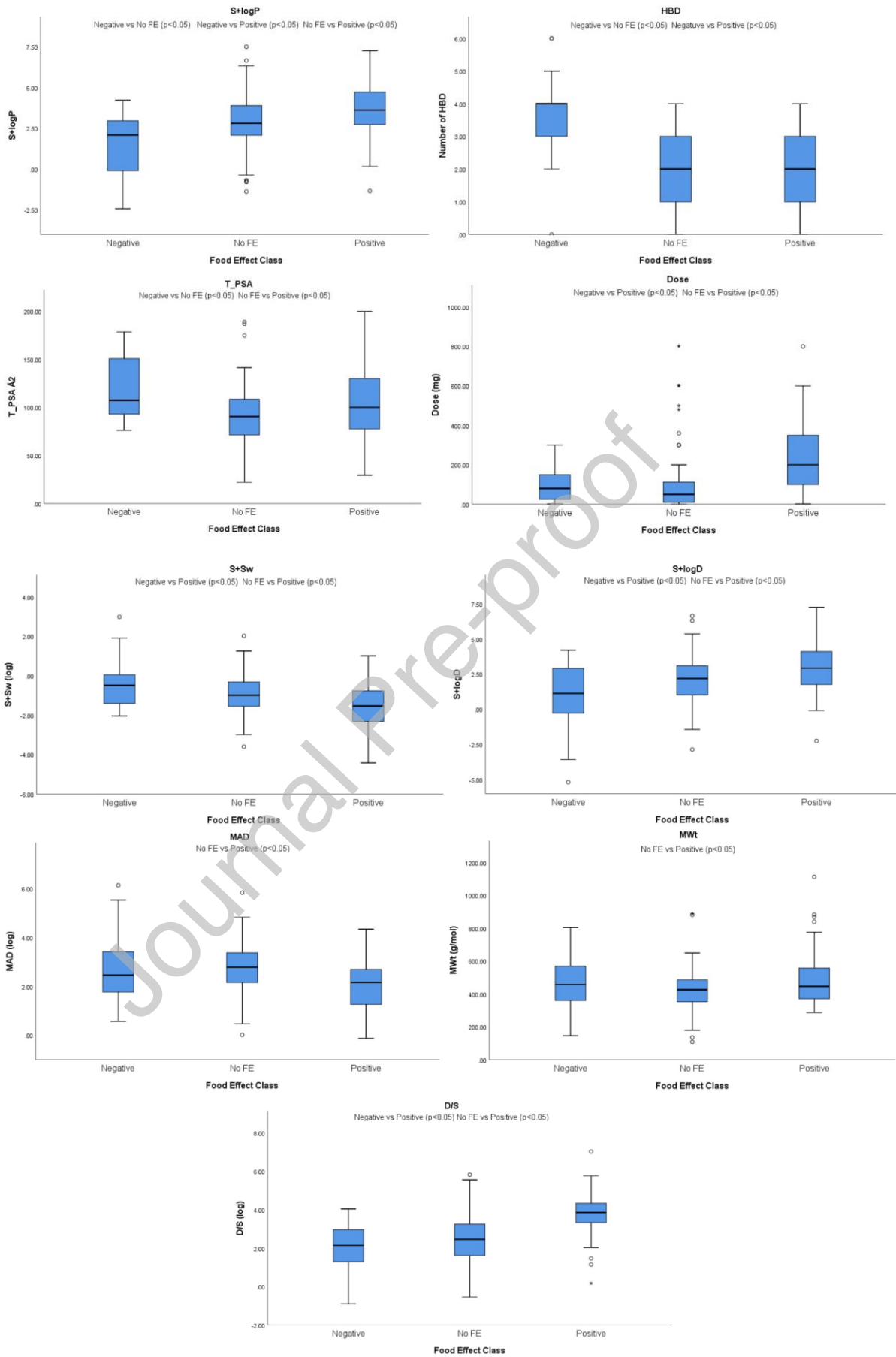
Figure 1 A+B*:* Visual representation of the statistically significant differences found between Positive, Negative and No Food Effect Classes as part of the preliminary statistical analysis. p-values for the statistically significant pairwise comparisons are shown. The dark line in the middle of the boxes is the median. The bottom and top of the box indicates the 25th (Q1) and 75th percentile (Q3). The T-bars are inner fences/whiskers which extend to 1.5 times the box height. The points are outliers that do not fall in the inner fences. The asterisks are extreme outliers which have values greater than three times the height of the boxes.
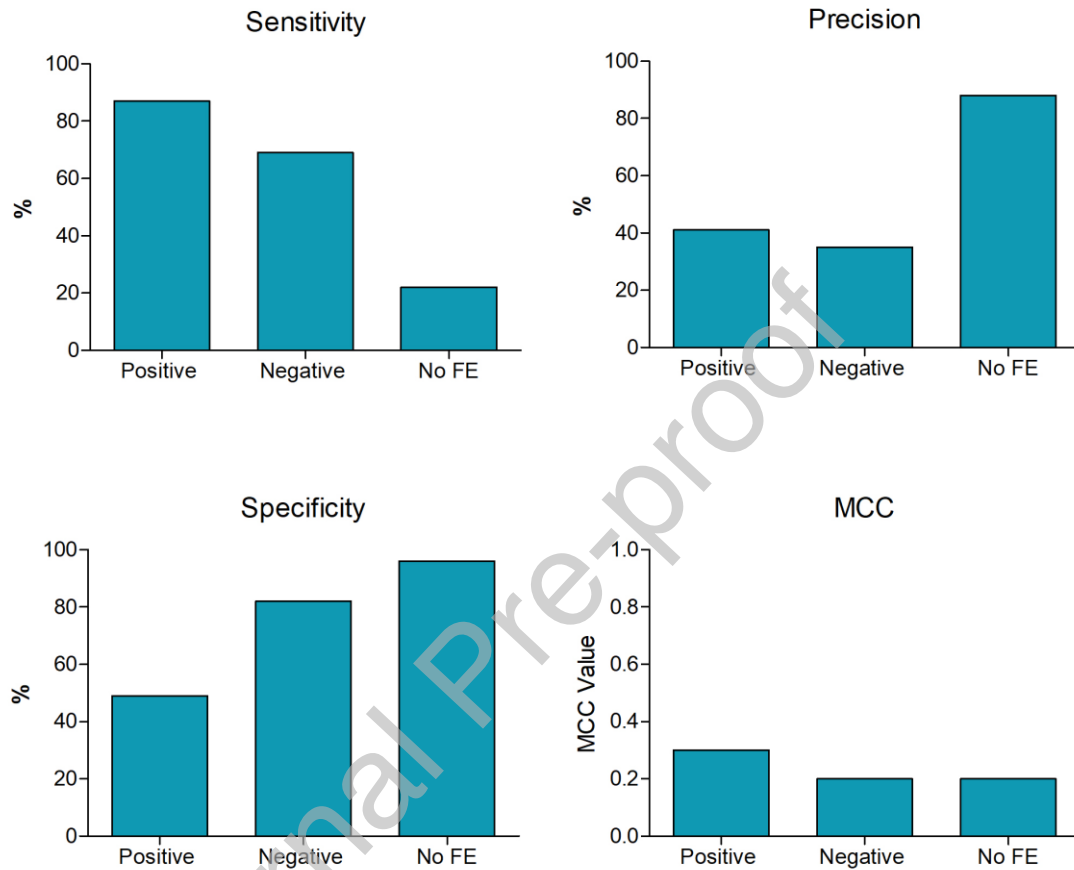
Figure 2: Visual representation of the sensitivity, precision, specificity and Matthews Correlation Coefficient (MCC) scores of the classification predictions (positive, negative and no Food Effect (FE)) using the biopharmaceutics classification system (BCS) criteria where an overall accuracy of 46% was achieved for the dataset.
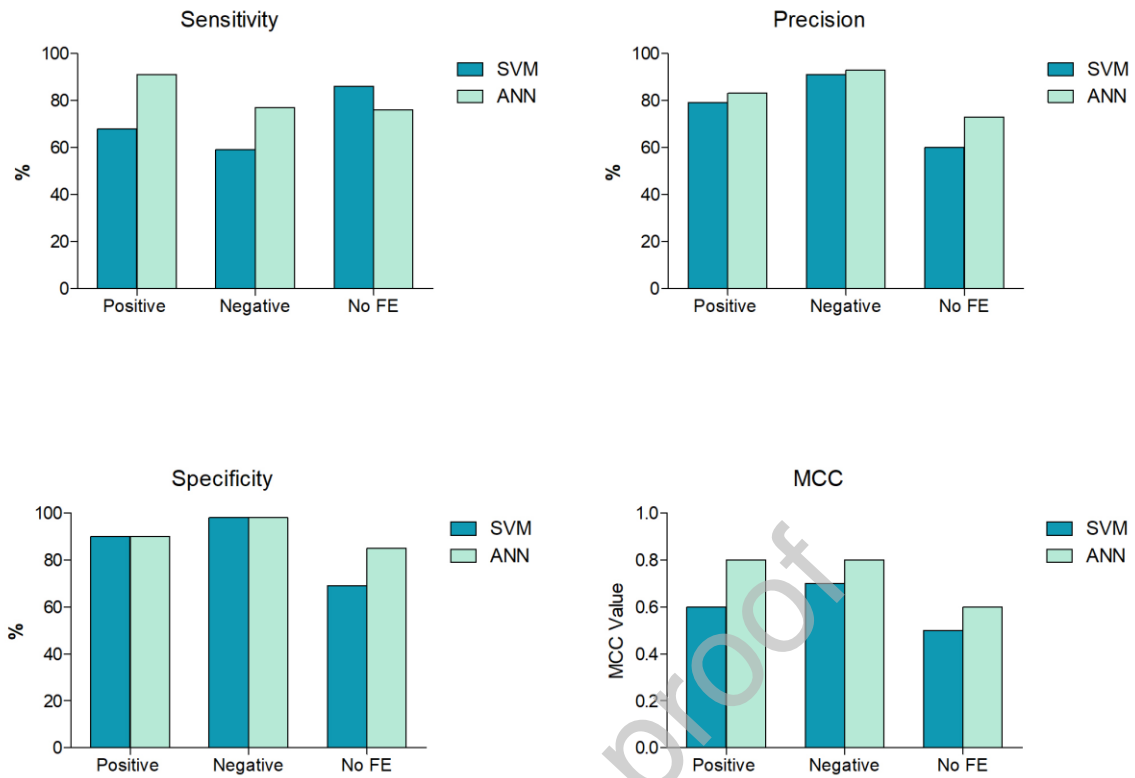
Figure 3: Visual comparison of the sensitivity, precision, specificity and Matthews Correlation Coefficient (MCC) performance metrics calculated for the optimum Support Vector Machines (SVM) and Artificial Neural Networks (ANN) models produced in this study to predict food effect (FE) classification.
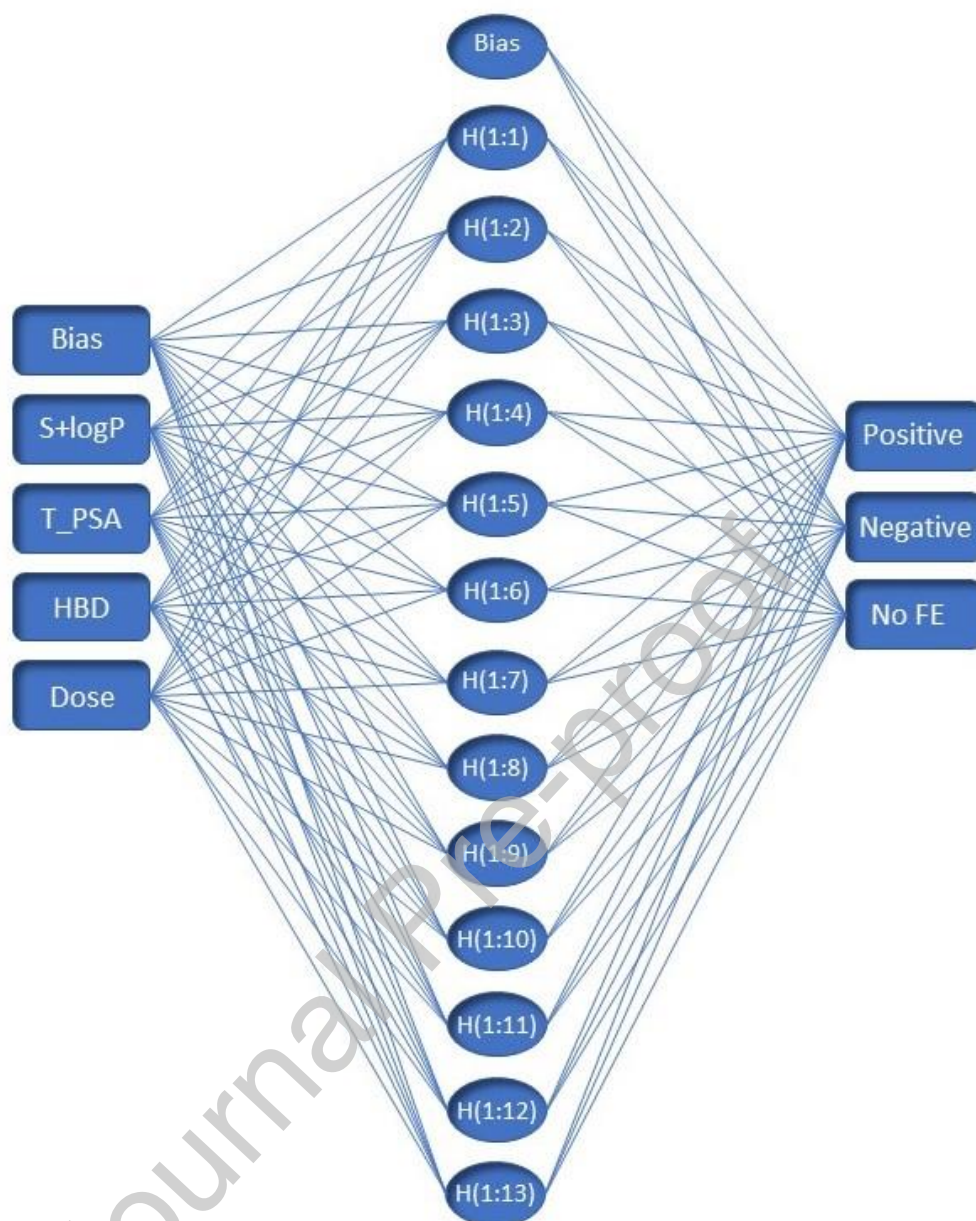
Figure 4: Schematic representation of the optimum multilayer perceptron (MLP) 4-13-3 Artificial Neural Network (ANN) produced which outperformed Support Vector Machines (SVM) for the food effect (FE) classification of drugs licensed from 2016-2020. H refers to hidden layer node.

Table 1: Compilation of licensed oral medicines from 2016-2020 and their AUC_{fed}/AUC_{fasted} ratio, clinical recommendation regarding food take, BCS and Food Effect (FE) Classification. LF, MF and HF refer to Low Fat, Medium Fat and High Fat Meals respectively, [*]refers to the 90% CI limits and [**]refers to the range of values quoted.

Year Licensed	Generic Name	Commercial Name	Clinical Recommendation	Food Effect Classification	AUC_{fed}/AUC_{fasted}	BCS Class
2020	Avapritinib	Ayvakit /Ayvakyt	Taken on an empty stomach, at least one hour before and two hours after a meal	Positive	1.27/1.29	2
2020	Glasdegib	Daurismo	Taken with or without food	No FE	0.84	4
2020	Lefamulin	Xenleta	Taken on an empty stomach, at least 1 hour before or 2 hours after a meal	No FE	0.82	3
2020	Pralsetinib	Gavreto	Taken on an empty stomach (no food intake for at least 2 hours before and at least 1 hour after taking)	Positive	2.22	2
2020	Osilodrostat	Isturisa	Taken with or without food	No FE	0.89	1
2020	Filgotinib	Jyseleca	Taken with or without food	No FE	1	2
2020	Ivacaftor	Kaftrio	Taken with fat-containing food	Positive	1.9-2.5[**]	2
	Tezacaftor			Positive	2.5-4[**]	2
	Elexacaftor			No FE	1	4
2020	Selumetinib	Koselugo	Take on an empty stomach. Do not consume food 2 hours before each dose or 1 hour after each dose	No FE	0. 62	4
2020	Nifurtimox	Lampit	Taken with food	Positive	1.71	2
2020	Siponimod	Mayzent	Taken with or without food	No FE	1	2
2020	Bempedoic Acid	Nilemdo	Taken with or without food	No FE	1	2
2020	Darolutamide	Nubeqa	Taken with food	Positive	2-2.5	2
2020	Bempedoic Acid	Nustendi/Nexlizet	Taken with or without food	No FE	1	2
	Ezetimibe			No FE	1	2
2020	Azacitidine	Onureg	Take with or without food	Negative	0.79	1
2020	Elagolix Sodium	Oriahnn	No instructions with regard to food intake	Negative	0.75	3
	Estradiol			No FE	1	1
	Norethindrone Acetate			No FE	1.23	2
2020	Alpelisib	Piqray	Taken immediately after food, at approximately the same time each day	Positive	LF 1.77 HF 1.73	2
2020	Pretomanid	Pretomanid FGK	Taken with food	Positive	1.88	2
2020	Selpercatinib	Retevmo	Taken with or without food	No FE	1	
2020	Solriamfetol	Sunosi	Taken with or without food	No FE	1	1
2020	Capmatinib Hydrochloride	Tabrecta	Taken with or without food	Positive	LF 1 HF 1.46	2
2020	Fostamatinib	Tavlesse	Taken with or without food	No FE	1.23	4
2020	Dolutegravir Sodium	Tivicay pd	Taken with or without food	Positive	1.66	4

Year Licensed	Generic Name	Commercial Name	Clinical Recommendation	Food Effect Classification	AUC_{fed}/AUC_{fasted}	BCS Class
2020	Tucatinib	Tukysa	Taken with or without food	Positive	1.5	2
2020	Solifenacin Succinate	Vesicare LS	Avoid taking with food due to bitter taste	No FE	1	1
2020	Enzalutamide	Xtandi	Taken with or without food	No FE	1	2
2020	Ozanimod	Zeposia	Taken with or without food	No FE	1	2
2019	Isotretinoin	Absorica Ld	Taken with or without food	No FE	1.2	2
2019	Lumateperone Tosylate	Caplyta	Taken with food	No FE	1.09	1
2019	Ivabradine	Corlanor	Taken with food	Positive	1.2-1.4[**]	1
2019	Trientine Dihydrochloride	Cufence	Take this medicine with water only. Avoid eating or drinking (except water) for 1 hour before, or 2 hours after taking.	Negative	0.55	3
2019	Lemborexant	Dayvigo	Taken immediately before going to bed	No FE	1.18	2
2019	Avatrombopag Maleate	Doptelet	Taken with food	No FE	LF 1.0 HF 1.0	4
2019	Dolutegravir Sodium	Dovato	Taken with or without food	Positive	1.33	4
	Lamivudine			No FE	1	3
2019	Triclabendazole	Egaten	Taken with food	Positive	2	2
2019	Gilteritinib Fumarate	Xospata	Taken with or without food	No FE	0.9	4
2019	Riluzole	Xservan	Taken at least 1 hour before or 2 hours after a meal	No FE	0.85	2
2019	Apalutamide	Erleada	Taken with or without food	No FE	1	2
2019	Colchicine	Gloperba	Taken with or without food	No FE	0.93	3
2019	Ledipasvir	Harvoni	Taken with or without food	No FE	1	2
	Sofosbuvir			Positive	~2	3
2019	Fedratinib Hydrochloride	Inrebic	Taken with or without food	No FE	LF 1.24 HF 1.24	2
2019	Lorlatinib	Lorviqua	Taken with or without food	No FE	1.05	4
2019	Lusutrombopag	Mulpleo	Taken with or without food	No FE	1	4
2019	Istradefylline	Nourianz	Taken with or without food	Positive	1.64	2
2019	Voxelotor	Oxbryta	Taken with or without food	Positive	1.42	2
2019	Naldemedine	Rizmoic/Symproic	Taken with or without food	No FE	1	4
2019	Amifampridine	Ruzurgi	Taken with or without food	No FE	1	3
2019	Talazoparib	Talzenna	Taken with or without food	No FE	1	2
2019	Elexacaftor	Trikafta	Taken with fat containing food	Positive	1.9-2.5[**]	4
	Ivacaftor	(Copackaged)		Positive	2.5-4[**]	2
	Tezacaftor			No FE	1	2
2019	Pexidartinib Hydrochloride	Turalio	Taken on an empty stomach, at least 1 hour before or 2 hours after a meal or snack	Positive	2	2

Year Licensed	Generic Name	Commercial Name	Clinical Recommendation	Food Effect Classification	AUC$_{fed}$/ AUC$_{fasted}$	BCS Class
2019	Ubrogepant	Ubrelvy	Taken with or without food	No FE	1	4
2019	Larotrectinib	Vitrakvi	Taken with or without food	No FE	1	1
2019	Dacomitinib	Vizimpro	No instructions with regard to food intake	No FE	1	2
2019	Ceritinib	Zykadia	Taken with food	Positive	LF 1.64 HF 1.39	4
2019	Sotagliflozin	Zynquista	Taken once daily before the first meal of the day	Positive	1.5	2
2018	Brigatinib	Alunbrig	Taken with or without food	No FE	1	1
2018	Tafenoquine Succinate	Arakoda	Taken with food	Positive	1.41	
2018	Bictegravir	Biktarvy	Taken with or without food	No FE	1.24	2
	Emtricitabine			No FE	1	1
	Tenofovir Alafenamide			Positive	1.64	3
2018	Estradiol	Bijuva	Taken with food	No FE	1	1
	Progesterone			Positive	1.79	2
2018	Encorafenib	Braftovi	Taken with or without food	No FE	1	2
2018	Duvelisib	Copiktra	Taken with or without food	Negative	0.63	4
2018	Doravirine	Delstrigo	Taken with or without food	No FE	1.1	2
	Lamivudine			No FE	0.93	3
	Tenofovir Disoproxil Fumarate			Positive	1.27	3
2018	Baloxavir Marboxil	Xofluza	Taken with or without food	No FE	0.64	2
2018	Ibrutinib	Imbruvica	No instructions with regard to food intake	Positive	2	2
2018	Dolutegravir Sodium	Juluca	Taken with a meal	Positive	1.87	4
	Rilpivirine Hydrochloride			Positive	1.72	2
2018	Tolvaptan	Jynarque	Taken with or without food	No FE	1	4
2018	Tafenoquine Succinate	Krintafel	Taken with food	Positive	1.41	
2018	Binimetinib	Mektovi	Taken with or without food	No FE	1	2
2018	Mexiletine Hydrochloride	Namuscla	Should be swallowed with water. In case of digestive intolerance, capsules should be taken during a meal.	No FE	1	1
2018	Neratinib	Nerlynx	Taken with food, preferably in the morning	Positive	2.2	4
2018	Omadacycline Tosylate	Nuzyra	Fast for at least 4 hours and then take	No FE	0.39	3
2018	Elagolix Sodium	Orilissa	Taken with or without food	No FE	0.76	3
2018	Ivacaftor	Orkambi	Mixed with one teaspoon (5 mL) of age-	Positive	3	2
	Lumacaftor		appropriate soft food or liquid and the mixture completely consumed.		2	2

Year Licensed	Generic Name	Commercial Name	Clinical Recommendation	Food Effect Classification	AUC_{fed}/AUC_{fasted}	BCS Class
2018	Doravirine	Pifeltro	Taken with or without food	No FE	1.16	2
2018	Letermovir	Prevymis	Taken with or without food	No FE	0.99	2
2018	Tacrolimus	Prograf	Taken consistently with or without food.	No FE	0.63	2
2018	Rucaparib Camsylate	Rubraca	Taken with or without food	Positive	1.38	2
2018	Brexpiprazole	Rxulti	Taken with or without food	No FE	1	2
2018	Sarecycline Hydrochloride	Seysara	Taken with or without food	Negative	0.73	3
2018	Ertugliflozin	Steglatro	Taken with or without food	No FE	1	1
2018	Ertugliflozin	Steglujan	Taken with or without food	No FE	1	1
	Sitagliptin			No FE	1	2
2018	Ivacaftor	Symdeko/Symkevi	Taken with fat-containing food	Positive	3	2
	Tezacaftor			No FE	1	2
2018	Ivosidenib	Tibsovo	Taken with or without food. Do not administer with a high fat meal due to increase in concentration	Positive	1.98	2
2018	Riluzole	Tiglutik Kit	Taken at least 1 hour before or 2 hours after a meal	No FE	0.91	2
2018	Tecovirimat	Tpoxx	Taken within 30 minutes after a full meal of moderate or high fat	Positive	1.39	2
2018	Abemaciclib	Verzenios	Taken with or without food	No FE	1.09	3
2018	Abiraterone Acetate	Yonsa	Taken with or without food	Positive	4.4	4
2017	Alectinib	Alecensa	Taken with food	Positive	3	4
2017	Deutetrabenazine	Austedo	Taken with food	No FE	1	2
2017	Betrixaban	Bevyxxa	Taken with food	Negative	LF 0.39 HF 0.52	3
2017	Acalabrutinib	Calquence	Taken with or without food	No FE	1	2
2017	Spironolactone	Carospir	Taken with or without food, but should be taken consistently with respect to food	Positive	1.9	2
2017	Allopurinol	Duzallo	Taken with food	No FE	1	4
	Lesinurad			No FE	1	2
2017	Deflazacort	Emflaza	Taken with or without food. Tablet and Suspension	No FE	1	
2017	Tofacitinib	Xeljanz	Taken with or without food	No FE	1	3
2017	Telotristat Etiprate	Xermelo	Taken with food	No FE	3.64	
2017	Pirfenidone	Esbriet	Taken with food	No FE	0.84	1
2017	Tivozanib	Fotivda	Taken with or without food	No FE	1	2
2017	Valbenazine Tosylate	Ingrezza	Taken with or without food	No FE	0.87	1
2017	Deferasirox	Jadenu Sprinkle	Taken on an empty stomach or with a light meal	No FE	LF 1 HF 1.18	2
2017	Ribociclib Succinate	Kisqali	Taken with or without food	No FE	1	4
2017	Macimorelin Acetate	Macrilen	Taken after fasting for at least 8 hours	Negative	0.51	

Year Licensed	Generic Name	Commercial Name	Clinical Recommendation	Food Effect Classification	AUC$_{fed}$/AUC$_{fasted}$	BCS Class
2017	Cladribine	Mavenclad	Taken with or without food	Negative	1	3
2017	Glecaprevir	Maviret	Taken at the same time with food	Positive	1.83-2.63[**]	4
	Pibrentasvir			Positive	1.4-1.53[**]	4
2017	Pitavastatin Sodium	Nikita	Taken with or without food	No FE	1	2
2017	Ritonavir	Norvir	Should be mixed with soft food	Negative	0.51	4
2017	Baricitinib	Olumiant	Taken with or without food	No FE	0.86	3
2017	Valsartan	Prexxartan	No instructions with regard to food intake	No FE	0.92	2
2017	Cariprazine	Reagila	Taken with or without food	No FE	1.12	2
2017	Edoxaban	Roteas	Taken with or without food	No FE	1	4
2017	Oxycodone Hydrochloride	Roxybond	No instructions with regard to food intake	No FE	1.23	3
2017	Midostaurin	Rydapt	Taken with food	No FE	1.6	2
2017	Tenofovir Alafenamide	Symtuza	Taken with food	No FE	1.20	3
	Darunavir			Positive	1.52	2
	Cobicistat			Positive	1.4	2
	Emtricitabine			No FE	1	1
2017	Tenofovir Alafenamide	Vemlidy	Taken with food	Positive	1.51-1.81[**]	3
2017	Niraparib	Zejula	Taken with or without food	No FE	1	1
2017	Pitavastatin Magnesium	Zypitamag	Taken with or without food	No FE	1	2
2016	Brivaracetam	Briviact	Taken with or without food	No FE	0.95	1
2016	Emtricitabine	Descovy	Taken with or without food	No FE	1	1
	Tenofovir Alafenamide			Positive	1.17-1.77[**]	3
2016	Sofosbuvir	Epclusa	Taken with or without food	Positive	1.78	3
	Velpatasvir			No FE	1.21	4
2016	Migalastat Hydrochloride	Galafold	Food should not be consumed at least 2 hours before and 2 hours after taking to give a minimum 4 hours fast	Negative	0.63-0.58[**]	3
2016	Empagliflozin	Glyxambi	Taken with or without food	No FE	0.84	3
	Lignagliptin			No FE	1	3
2016	Palbociclib	Ibrance	Taken with food	No FE	1.2	2
2016	Lenvatinib Mesilate	Kisplyx	Taken at about the same time each day, with or without food	No FE	1	2
2016	Trifluridine	Lonsurf	No instructions with regard to food intake	No FE	1	3
	Tipiracil Hydrochloride			Negative	0.6	3
2016	Sacubitril	Neparvis	Taken with or without food	No FE	1	4
	Valsartan			No FE	1	2
2016	Ixazomib	Ninlaro	Taken at least 1 hour before or at least 2 hours after food	Negative	0.72	3
2016	Pimavanserin	Nuplazi	Taken with or without food	No FE	1.08	

Year Licensed	Generic Name	Commercial Name	Clinical Recommendation	Food Effect Classification	AUC$_{fed}$/ AUC$_{fasted}$	BCS Class
		d				
2016	Obeticholic Acid	Ocaliva	Taken with or without food	No FE	1	2
2016	Emtricitabine	Odefsey	Taken with food	No FE	0.88 (0.85-0.9)[*]	1
	Rilpivirine Hydrochloride			Positive	1.72 (1.49-1.99)[*]	2
	Tenofovir Alafenamide			Positive	1.53 (1.39-1.69)[*]	3
2016	Opicapone	Ongentys	Should not eat food for 1 hour before and for at least 1 hour after intake.	Negative	0.69	2
2016	Saxagliptin	Qtern	Taken with or without food	Positive	1.27	3
	Dapagliflozin Propanediol Monohydrate			No FE	1	3
2016	Dronabinol	Syndros	Administer the first dose on an empty stomach at least 30 minutes before eating. Subsequent doses can be taken without regard to meals.	Positive	2.5	2
2016	Osimertinib Mesylate	Tagrisso	Taken with or without food	No FE	1.06	3
2016	Eluxadoline	Truberzi	Taken with food	Negative	0.4	3
2016	Selexipag	Uptravi	Taken with food	No FE	1.1	2
2016	Venetoclax	Venclyxto	Taken with a meal	Positive	MF 3.4 HF 5.1-5.3[**]	4
2016	Elbasvir	Zepatier	Taken with or without food	No FE	0.89	2
	Grazoprevir			Positive	1.5	2

Table 2: Overview of the Support Vector Machine (SVM) and Artificial Neural Network (ANN) machine learning models produced in this study, detailing the inputs, model architecture and the comparative overall accuracies upon training and testing.

Model Type	Input Properties Used	Architecture	Overall Accuracy Training Set	Overall Accuracy Test Set
SVM	Dose, HBD, F_HBP, S+logP, T_PSA, logHLC	Kernel: RBF Gamma: 0.01 C value: 16.68	72%	69%
ANN	Dose, HBD, S+logP, T_PSA	1 hidden layer 13 hidden nodes	82%	72%