



Article

Machine Learning Predicts Drug Metabolism and Bioaccumulation by Intestinal Microbiota

Laura E. McCoubrey, Stavriani Thomaidou, Moe Elbadawi, Simon Gaisford , Mine Orlu and Abdul W. Basit *

Department of Pharmaceutics, UCL School of Pharmacy, University College London, London WC1N 1AX, UK; laura.mccoubrey.18@ucl.ac.uk (L.E.M.); stavriani.thomaidou.20@ucl.ac.uk (S.T.); m.elbadawi@ucl.ac.uk (M.E.); s.gaisford@ucl.ac.uk (S.G.); m.orlu@ucl.ac.uk (M.O.)

* Correspondence: a.basit@ucl.ac.uk; Tel.: +44-2039872769

Abstract: Over 150 drugs are currently recognised as being susceptible to metabolism or bioaccumulation (together described as depletion) by gastrointestinal microorganisms; however, the true number is likely higher. Microbial drug depletion is often variable between and within individuals, depending on their unique composition of gut microbiota. Such variability can lead to significant differences in pharmacokinetics, which may be associated with dosing difficulties and lack of medication response. In this study, literature mining and unsupervised learning were used to curate a dataset of 455 drug–microbiota interactions. From this, 11 supervised learning models were developed that could predict drugs' susceptibility to depletion by gut microbiota. The best model, a tuned extremely randomised trees classifier, achieved performance metrics of AUROC: $75.1\% \pm 6.8$; weighted recall: $79.2\% \pm 3.9$; balanced accuracy: $69.0\% \pm 4.6$; and weighted precision: $80.2\% \pm 3.7$ when validated on 91 drugs. This machine learning model is the first of its kind and provides a rapid, reliable, and resource-friendly tool for researchers and industry professionals to screen drugs for susceptibility to depletion by gut microbiota. The recognition of drug–microbiome interactions can support successful drug development and promote better formulations and dosage regimens for patients.

Keywords: artificial intelligence; classification; semi-supervised learning; gastrointestinal microbiome; drug stability; drug discovery and development; pharmacokinetics; in silico prediction; principal component analysis; feature selection



Citation: McCoubrey, L.E.; Thomaidou, S.; Elbadawi, M.; Gaisford, S.; Orlu, M.; Basit, A.W. Machine Learning Predicts Drug Metabolism and Bioaccumulation by Intestinal Microbiota. *Pharmaceutics* **2021**, *13*, 2001. <https://doi.org/10.3390/pharmaceutics13122001>

Academic Editor: David Barlow

Received: 28 October 2021

Accepted: 23 November 2021

Published: 25 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Over 150 drugs are recognised as being susceptible to metabolism or bioaccumulation by intestinal microbiota [1–15]. These direct effects on drug concentration can lead to significant inter-individual variability in pharmacokinetics, arising due to differences between individuals' gut microbiome compositions [16–18]. Microbial drug metabolism or bioaccumulation (termed henceforth as depletion) is often dependent upon the production of specific enzymes that may be variably expressed between patients [3,19–23]. For example, digoxin is inactivated by strains of *E. lenta* that produce the cardiac glycoside reductase (CGR) enzyme [24]. Research has shown that the abundance of the CGR gene relative to *E. lenta* concentration in patients' faeces is significantly correlated with their ex vivo metabolism of digoxin, providing strong evidence that patients colonised by CGR-encoding bacteria metabolise digoxin to a greater extent in vivo [25]. Another drug whose microbial metabolism has been explored, with consideration for inter-individual variability, is tacrolimus [26]. Lee et al. have highlighted that the abundance of *F. prausnitzii* in patients' stools is positively correlated with their tacrolimus dosing requirements [27]. This relationship is thought to arise because *F. prausnitzii* transforms tacrolimus to a metabolite known as M1, which has 15-fold lower immunosuppressant activity than the original drug [26].

The inter-individual pharmacokinetic variability arising from the microbial depletion of drugs could contribute towards treatment failure or toxicity in some patients, in addition

to difficulties finding an optimum dose [28–30]. Further, pharmacokinetic variability could risk the progression of novel treatments through clinical trials to market approval. Despite these risks, the susceptibility of drugs to microbiota depletion is not routinely tested during preclinical or clinical development [31]. In rare cases when microbial metabolism is explored, it is usually conducted to determine drugs' stability in the colonic environment rather than to study pharmacokinetic variability [7]. Due to its underexplored nature, there is currently no universally accepted method for the confirmation of microbial drug bioaccumulation [3]. Common methods used to quantify microbial drug metabolism involve incubating drugs in either human or animal faecal slurries, microbial cultures, and less commonly in intestinal fluids to measure drug degradation over a defined period (usually ≤ 24 h) [1,2,6]. Experimental determination of drugs' susceptibility to microbial depletion can be a time-consuming and resource-intensive process. High-throughput screening of many investigative active pharmaceutical ingredients (APIs) may not be feasible in industry for several reasons. Firstly, drug development pipelines may not have the necessary time required to measure many drugs' degradation and metabolite formations, as drug-specific high-performance liquid chromatography or mass spectrometry methods are frequently required [2]. Moreover, it is desirable to minimise the use of animals, and sourcing either human faecal or intestinal fluid samples can be difficult and expensive. For this reason, *in silico* methods of predicting drugs' microbiome depletion hold significant potential [32–36].

To date, there are a few examples of the *in silico* prediction of microbial drug depletion [37]. In 2017, Sharma et al. used random forest ML to develop a predictive model known as DrugBug [38]. The group identified 324,697 metabolic enzymes from 491 gut bacterial genomes. Non-drug substrates of these bacterial enzymes ($n = 1609$) were then retrieved from the KEGG database (a resource containing thousands of biochemical interactions). The substrates were used as a training dataset to classify the bacterial enzymes most likely to metabolise drugs. Though the model achieved $>90\%$ accuracy in predicting the non-drug substrates' associated enzymes, the field has evolved since, requiring new input considerations. Firstly, the model was developed at a time when only a handful of drugs had characterised bacterial metabolic reactions, resulting in an outdated model. Further, the outcome of drug stability was not considered (the model assumes all drugs were metabolised by bacterial enzymes); however, it is now suspected that the majority of drugs are resistant to microbial transformation [1,2]. In 2019, as part of their high-throughput screening study, Zimmerman et al. used a hierarchical clustering algorithm to examine the functional groups that increase drugs' risk of being microbially depleted [1]. They observed that drugs containing urea, azo, lactone, and nitro functional groups were more likely to be depleted by at least one of the tested 76 gut bacterial strains. Whilst this explorative analysis is interesting, it should be explained with a quantifiable means to assess untested drugs' risk of microbial depletion to provide further insight into microbiome activity. Elsewhere, Elmassry et al. evaluated drugs' risk of metabolism by bacterial β -glucuronidases using a common substructure algorithm [39]. Drugs inferred to undergo reaction with microbial β -glucuronidases, due to their known reaction with hepatic β -glucuronidases, were clustered based on their chemical structures. Structural similarity of new drugs with those included in the study could facilitate predictions for untested β -glucuronidase metabolism. However, such predictions would be based on the untested assumption that drugs in the training set really do undergo metabolism by microbial β -glucuronidases.

This study aimed to develop a classification algorithm capable of predicting whether small molecule drugs are susceptible to direct depletion (i.e., metabolism or bioaccumulation) by gut microbiota. The model output is binary, i.e., depleted or not depleted, and includes a level of predictive confidence. This technology could become a valuable tool for *in silico* prediction of drug–microbiota interactions, and the compounding effects on bioavailability and pharmacokinetic variability.

2. Materials and Methods

2.1. Dataset Curation and Labelling

Experimental data describing the depletion of drugs by gut microbiota were compiled from several studies, with the majority from work by Zimmerman et al. and Javdan et al. [1–3,6,7,10,12]. Zimmerman et al. incubated 271 drugs independently with 76 gut bacterial isolates anaerobically for 24 h. Drugs were labelled as being significantly ($p < 0.05$) depleted by at least one bacterial strain if $\geq 20\%$ reduction in starting concentration occurred [1]. Javdan et al. anaerobically incubated 438 drugs in the presence of gut microbiota sourced from a single human donor (sex not disclosed) [2]. Drugs were labelled as metabolised if the drug was observed to be entirely consumed and a new metabolite was formed after 24 h, in at least 2 of 3 independent experiments. Drugs from other included studies were incubated under anaerobic conditions for 24 h in faecal slurry sourced from multiple healthy human donors (both sexes) and labelled as metabolised if they were significantly degraded compared to controls [6,7,10,12]. In this manner, drugs were labelled into two categories: depleted or not depleted. Where the same drug was tested in two studies (this included multiple salt forms of the same drug), and results disagreed, another study from the literature was sought to provide a 3rd opinion and enable the assignment of a label. Importantly, only studies based on gut bacterial isolates or human faecal/intestinal fluid were considered; studies examining drug metabolism in animals were not included, as the microbiome composition of animals is significantly different to that of humans [40]. Where a 3rd study could not be found, then drugs were placed into a temporary category known as undefined. These drugs ($n = 86$) were then assigned a label using a K-nearest neighbours algorithm trained on the labelled drugs ($n = 469$). Model inputs were the 86 unlabelled drugs attached to 200 physicochemical parameters from Python's RDKit library (version 2021.03.1) combined with Morgan fingerprints (radius 2, 1024 bits). $K = 11$ was chosen for the K-nearest neighbours algorithm, as this value was found to have the lowest mean error rate during training, from K values of 1–25 (Figure 1). A mean error rate of 0 signifies that no errors were made by the model during classification of the training dataset, whereas mean errors closer to 1 demonstrate higher instances of incorrect prediction. A k value of 11 indicates that a sample was assigned a group based on the classifications of its nearest 11 neighbours in the multidimensional feature space. Following this, a labelled dataset of 555 drugs was formed. Because the dataset was unbalanced (drugs labelled as not depleted = 411; drugs labelled as depleted = 144), which can affect the reliability of machine learning models, 100 drugs from the not depleted category were removed from consideration using a random seed of 10 in Python. This resulted in 455 drugs being considered in a more balanced dataset (311 not depleted; 144 depleted). The predictive performance of the original and more balanced datasets was assessed.

2.2. Feature Generation

Simplified molecular-input line-entry system (SMILES) notations were obtained from PubChem for each drug. Based on these SMILES structures, two types of molecular fingerprints were assigned to each drug: Morgan fingerprints (radius 2, 1024 bits) and 200 physicochemical parameters from Python's RDKit (version 2021.03.1). Morgan fingerprints are one of the most popular chemical descriptors used for small molecules and provide a fingerprint based on multiple substructures around each atom in a molecule [41–43]. In comparison, the descriptors generated from RDKit include more functional, property-based features of drugs such as molecular weight and SLogP. Both Morgan and physicochemical fingerprints were investigated as drugs' inputs for the machine learning models, trialled together and in isolation to determine the best descriptors.

2.3. Data Preprocessing and Visualisation

Drug labels were encoded as 0 (not depleted) or 1 (depleted) using Sklearn's LabelEncoder. To remove noise arising from unit variance, drug features were standardised using the StandardScaler tool in the Python sklearn.preprocessing library. The dataset was also

checked for NA values, and none were present. To visualise the spread of data, a principal component analysis (PCA) algorithm was applied using the standard PCA tool in Python's Sklearn package (random state = 0). The entire labelled dataset was decomposed into principal components, and the percentage explained variances of the top 10 components were elucidated using a Scree plot. Following this, the top 2 principal components were plotted with drugs labelled as depleted/not depleted.

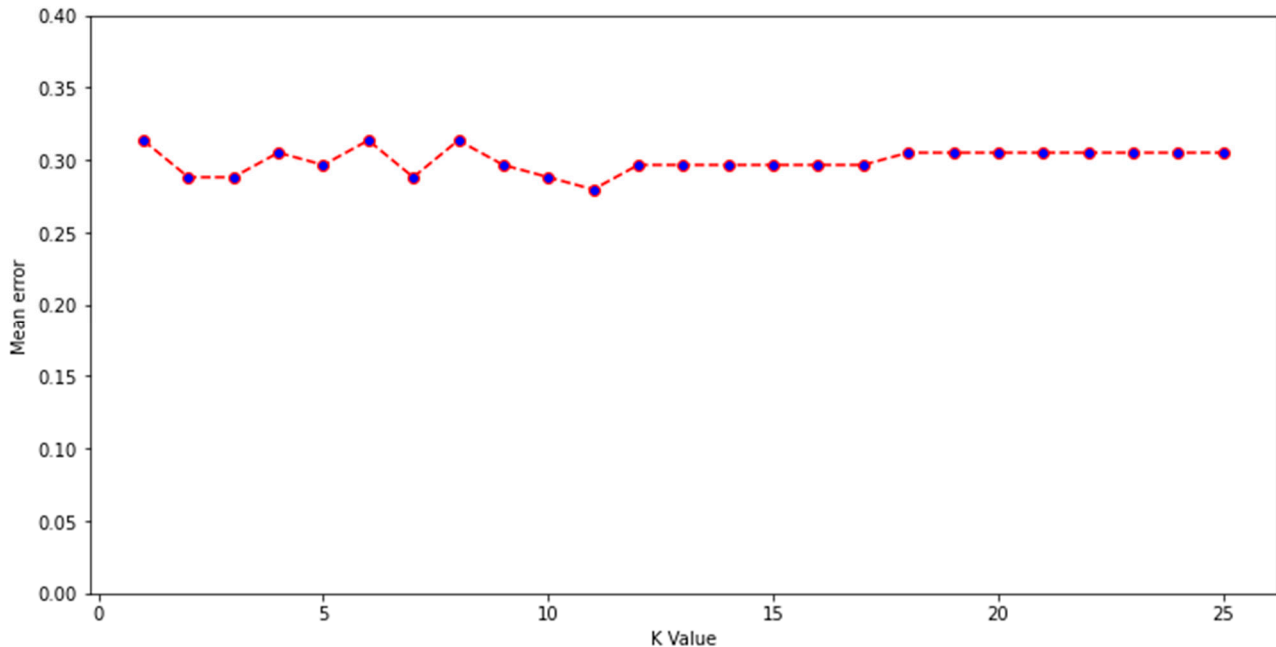


Figure 1. Mean error for K values of 1–25 during construction of a K-nearest neighbour algorithm developed to label unlabelled drugs.

2.4. Development of Machine Learning Models

Eleven ML techniques were investigated in this study to capture performance across a range of learning styles: extra trees, random forest, K-nearest neighbours (kNN), multilayer perceptron (MLP), decision tree, support vector machine (SVM), gradient boosting, logistic regression, stochastic gradient descent (SGD), perceptron, and passive aggressive classification. A background on various ML methodologies has been published by Badillo et al. [44]. During initial development, all models were used in their baseline state, imported from Python's Sklearn library (random state = 0 where applicable). The 11 baseline models were assessed for their ability to predict drugs' microbiome status using 4 performance metrics with Sklearn: mean area under the curve of the receiver operating characteristic (AUROC), weighted precision, weighted recall, and balanced accuracy. Weighted and balanced metrics were used to offset bias arising from the imbalance between the number of depleted vs. not depleted drugs in the dataset. Weighting performance involved calculating metrics for both classes and then finding their average based on the number of instances in that class. Performance metrics were chosen to give a global understanding of models' performances [45]. The AUROC score summarises a model's true positive (TP) prediction rate as a function of false positive (FP) prediction rate. Precision equals the number of TPs divided by the total number of positives (TPs + FPs) generated by a model, and as such it captures the likelihood of positive cases being overestimated (type 2 error). Recall is defined as the number of TPs divided by the number of actual positives (TPs + false negatives (FNs)) within a model, thus measuring the risk of positive cases being missed (type 1 error). Finally, balanced accuracy equals the number of correct predictions, weighted by the number of true and false samples in the dataset, divided by 2 (Equation (1)). Each model's performance metrics were calculated using the ShuffleSplit cross-validation tool in Sklearn (number of splits = 10, test size = 20%, random state = 0).

Using this form of cross-validation, the entire dataset was shuffled and split into 10 random groups, within which 80% of data were used for training and the remaining 20% for testing. Using cross-validation leads to higher trust in models' performance, as scores are generated using several different test datasets, thus also providing protection against model overfitting. Here, models' performance scores are presented as averages (means) and standard deviations across all cross-validation folds.

$$\text{Balanced accuracy} = \frac{1}{2} \left(\frac{\text{TP}}{\text{TP} + \text{FN}} + \frac{\text{TN}}{\text{TN} + \text{FP}} \right) \quad (1)$$

2.5. Selection and Optimisation of Best Model

The best ML model was selected by considering global performance. Where a model achieved the 1st, 2nd, or 3rd best score for a single metric, it was assigned 3, 2, or 1 point(s), respectively. Models' scores across the 4 metrics were totalled and the models were ranked, with the best attaining the highest overall score. The best model was then optimised by hyperparameter tuning, using the RandomizedSearchCV function within Python's Sklearn package (param_distributions = random_grid, n_iter = 50, cv = 3, verbose = 2, random_state = 0, n_jobs = -1). The parameters included in the randomised search were n_estimators, max_features, max_depth, min_samples_split, min_samples_leaf, bootstrap, and class_weight. The Python code for the final model is available in the Supplementary Materials, with which users can make predictions for untested APIs. With the model prediction (depleted/not depleted), there is a level of predictive confidence supplied by training an ActiveLearner (modAL active learning framework for Python version 0.4.1 (Szeged, Hungary)) based on the final model. This feature outputs classifier uncertainty (from 0.00–1.00) for each prediction, whereby lower uncertainty scores signify that the model has higher confidence in a specific prediction.

2.6. Data Analysis and Statistics

A PC (running on operative system: Windows 10 64-bit, processor: Intel® Core i7 3770 K (Santa Clara, CA, USA) (overclocked 4.5 GHz), RAM: 16 GB DDR3, and graphics card: Asus Phoenix GTX 1660 OC Edition (Taipei, Taiwan)) was used for data analysis and ML model construction. The ML dataset was compiled within Microsoft® Excel® for Microsoft 365 MSO (16.0.13231.20372) 64-bit. Dataset cleaning and preprocessing, and model construction and evaluation, were completed using Python version 3.9.0 (Wilmington, DE, USA) on Jupyter Notebook version 6.0.3 (San Francisco, CA, USA). All ML techniques were developed using Python's scikit-learn package, version 0.23.2. Metrics used to assess models' performance: AUROC, weighted precision, weighted recall, and balanced accuracy. The statistical difference between models was calculated using either a t-test (when comparing two models) or one-way ANOVA (when comparing > 2 models) using the Scipy package (version 1.7.1) in Python, with $p < 0.05$ taken as significant. The time taken to fit models ($n = 3$ for each measurement) was computed using Python's time function. Feature shuffling was performed by randomly relocating features' positions in the training dataset, followed by analysis of the ML model performance, to ensure the validity of the final model's performance scores. The 3 shuffled datasets are available in the Supplementary Materials. Plots were constructed using the Matplotlib package in Python and OriginPro (version 2021b).

3. Results and Discussion

3.1. Unsupervised Learning

To visualise how drugs' physicochemical features relate to their microbial depletion status, the dataset was decomposed using PCA into principal components (PCs), and the first two PCs (accounting for around 6.5% of total variance) were plotted (Figure 2). PCA is an unsupervised learning technique that can find inherent differences in data without the need for a labelled dataset. PCA revealed that depleted drugs have similar physicochemical properties to those not depleted, depicted by many drugs sharing a similar physicochemical

space regardless of label. That said, certain physicochemical fingerprints may be predictive of microbial depletion, as suggested by regions of mainly non-depleted drugs in the PCA plot. Alone, the PCA analysis cannot provide sufficient predictive power because distinct clusters of non-depleted/depleted drugs are not apparent in Figure 2. Therefore, numerous supervised ML techniques were subsequently explored to facilitate such predictions.

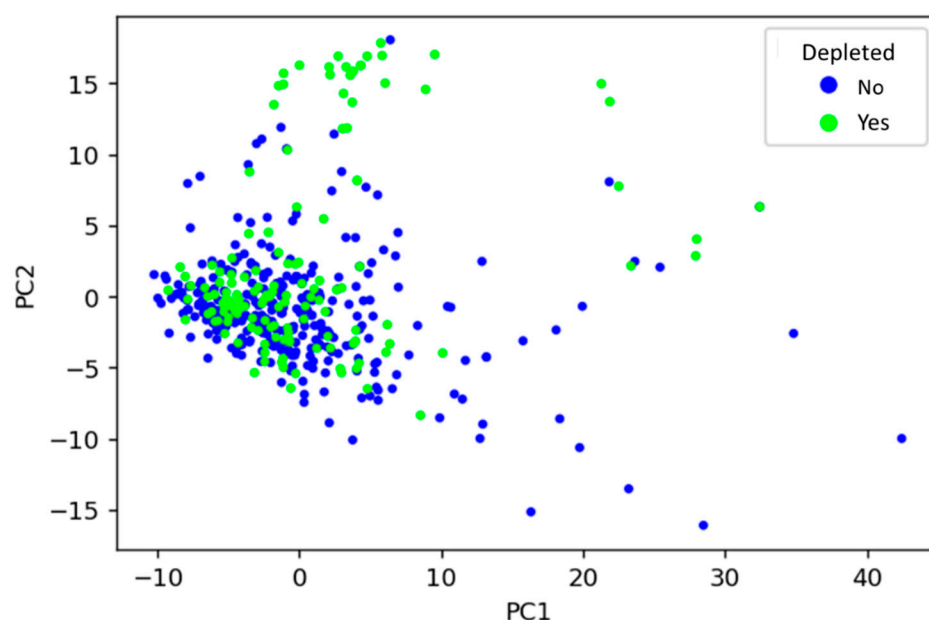


Figure 2. Visualisation of the relationship between drugs' physicochemical features and whether they are depleted by gut microbiota, using principal component analysis. PC1: principal component 1, PC2: principal component 2.

3.2. Dataset Balancing

Increasing the balance between the two classes within the training dataset (depleted vs. not depleted) resulted in performance changes; however, these were not statistically different (AUROC: $p = 0.31$, weighted precision: $p = 0.16$, weighted recall: $p = 0.14$, balanced accuracy: $p = 0.26$) (Figure 3). Though observed changes were insignificant, it is generally recognised in the field that unbalanced datasets need amendment to produce a reliable ML model, otherwise classifications may be biased towards the overrepresented group [44]. Moreover, the more balanced dataset was significantly faster to train than the unbalanced set (2.63 s compared to 3.07 s, $p = 3.44 \times 10^{-7}$) due to its smaller size. Based on this reasoning, the more balanced dataset was selected for use going forward.

3.3. Feature Selection

Three different feature selections were explored in this study, where all were found to yield high metric scores (Figure 4). Performances were determined as statistically indifferent across the sets (AUROC: $p = 0.66$, weighted precision: $p = 0.19$, weighted recall: $p = 0.34$, balanced accuracy: $p = 0.95$). Therefore, all could be feasibly used as drug features in the ML models. However, the physicochemical set has fewer features than the Morgan fingerprints (by 824 features), thus it was hypothesised that ML model training using physicochemical parameters alone would be computationally less intensive than when incorporating Morgan fingerprints. This was reflected in the time taken to train models, whereby the combined feature set and Morgan fingerprints alone required an average of 2.63 s and 2.67 s to train, respectively, compared to just 1.71 s for the physicochemical parameters in isolation. Accordingly, the physicochemical feature set needed significantly less time to train the representative extremely randomised trees (extra trees) ML model than when combined with Morgan fingerprints or the Morgan fingerprints alone ($p = 6.85 \times 10^{-8}$ and 7.53×10^{-8} , respectively). Computational efficiency is an important consideration

when developing ML models, especially in the context of this study, in which numerous models were screened. In practice, more efficient models would be better suited to the high-throughput *in silico* screening of drug–microbiota interactions, as more predictions could be generated in a given timeframe. In addition, utilising the smaller physicochemical feature set ensured that the number of features was comparable to the number of drug–microbiota observations, minimising overfitting [46,47]. In recognition of these factors, all subsequent models were trained on drugs’ physicochemical features alone.

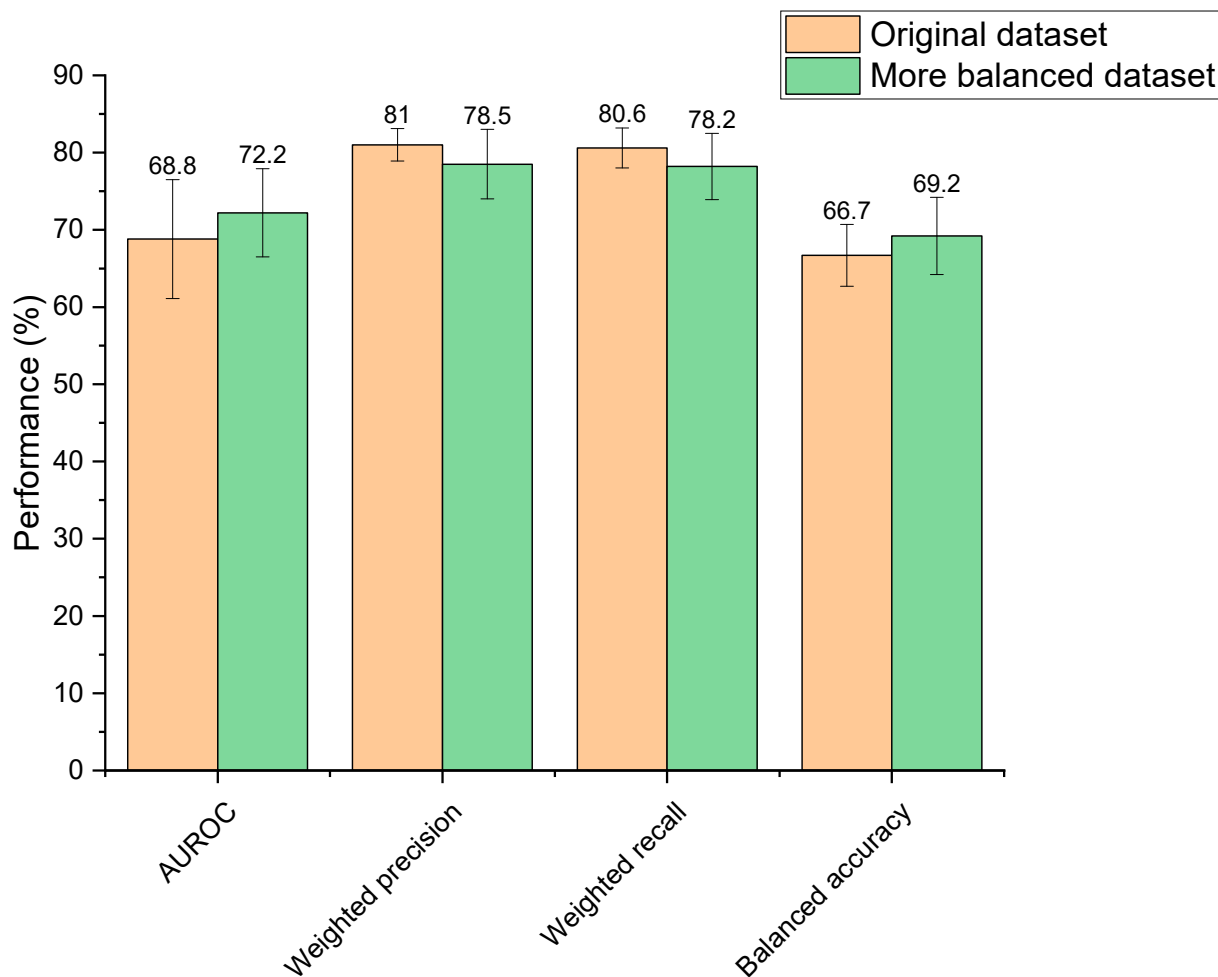


Figure 3. Performance of two extra trees machine learning models in predicting drugs’ microbial depletion status. Orange bars indicate scores for the model trained on the original unbalanced dataset. Green bars indicate scores for the model trained on a more balanced dataset.

3.4. Supervised Machine Learning

3.4.1. Baseline Models

Figure 5 shows the performances of the 11 supervised ML models developed to classify whether drugs are microbially depleted or not. Based on mean scores alone, the extra trees model received the highest ranking (with nine ranking points) followed MLP, SVM, and random forest (holding four ranking points each). However, when examined for significant differences, the models were found to be statistically similar based on AUROC ($p = 0.98$) and balanced accuracy ($p = 0.28$). For weighted precision and recall, the MLP model had significantly lower performance than the extra trees, SVM, and random forest models ($p < 0.05$), though extra trees, SVM, and random forest were all statistically indistinguishable. The extra trees model was selected for further optimisation, as it is a reliable tree-based method that can learn both linear and non-linear relationships, and has been successfully used to model drug–microbiome interactions in the past [48]. Compared to a

simple decision tree, extra trees has the benefit of utilising multiple randomised trees to generate predictions on subsections of data, where final predictions are averages across all trees [49]. This feature provides added control against overfitting and can improve predictive accuracy [50]. The extra trees model in this study had significantly higher weighted precision ($p = 0.002$), weighted recall ($p = 0.0001$), and AUROC ($p = 0.01$) scores than the decision tree. Extra trees also has the added benefit of being more computationally efficient and reducing the variance of the trees compared to random forest, making it more generalisable to new data [51,52]. Here, the extra trees model was found to be more computationally efficient than the random forest model, requiring on average 0.49 s less to train ($p = 5.84 \times 10^{-6}$).

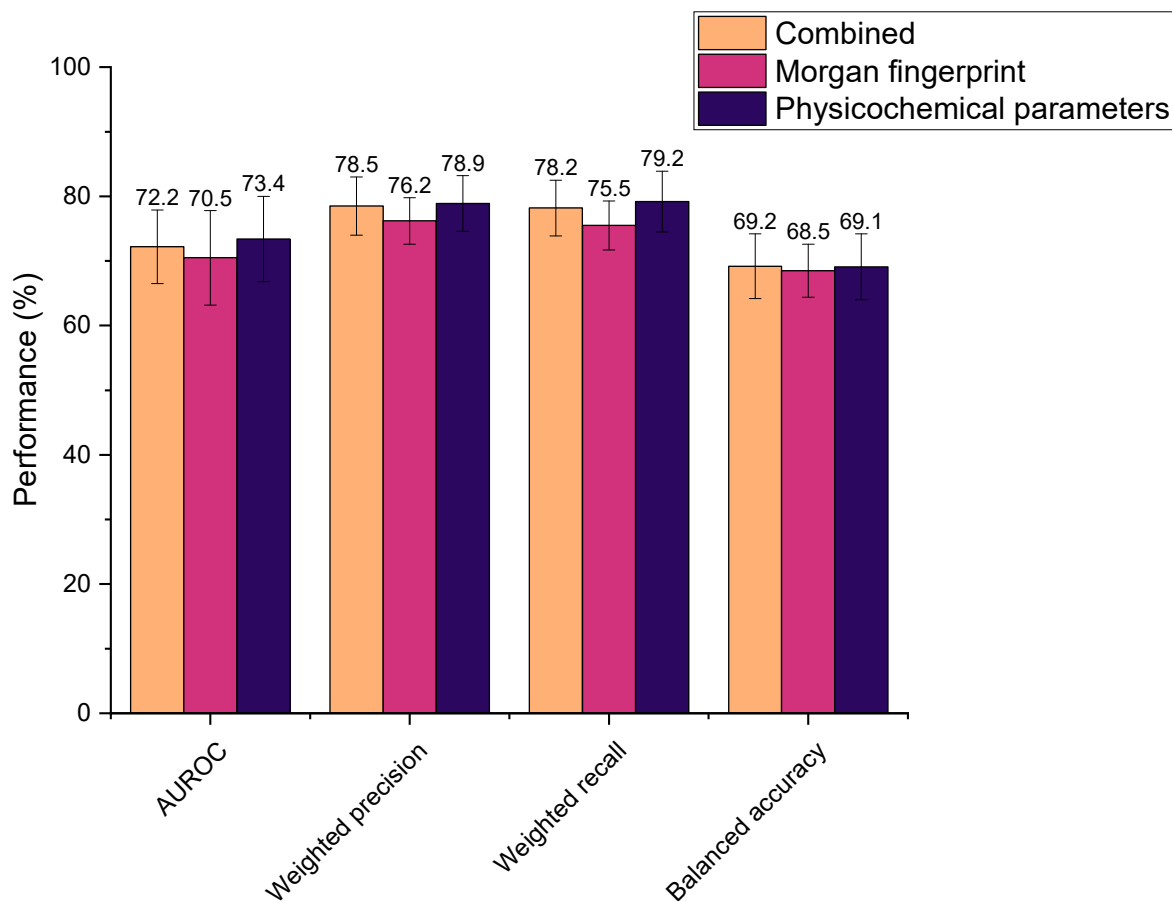


Figure 4. Performance of extra trees machine learning models in predicting drugs' microbial depletion status. Bar colours indicate the features used to describe drugs in the training dataset. Combined: both Morgan fingerprints (radius 2, 1024 bits) and 200 physicochemical parameters (from Python's RDkit).

3.4.2. Hyperparameter Optimisation

Optimisation of the extra trees model's hyperparameters led to altered performance scores that were statistically indifferent to the baseline model (Figure 6). Whilst mean AUROC, weighted recall, and weighted precision increased post tuning, these differences were not significant ($p > 0.59$ for all metrics). This reflects that the baseline extra trees model was already well suited to modelling relationships between drug features and microbial depletion. That said, hyperparameter tuning reduced standard deviations for weighted recall, weighted precision, and balanced accuracy scores, demonstrating that the tuned model had slightly lower variability than the baseline model. In the case of AUROC, the standard deviation of the tuned model was only 0.2% lower than the baseline model. Based on this, the tuned extra trees algorithm was selected as the final model for the prediction of microbial drug depletion. The final model parameters were: `n_estimators: 1400`, `min_samples_leaf:`

1, max_features: 'sqrt', max_depth: 90, class_weight: 'balanced', bootstrap: False. Final metrics were calculated as: AUROC: 75.1% \pm 6.8; weighted recall: 79.2% \pm 3.9; weighted precision: 80.2% \pm 3.7; and balanced accuracy: 69.0% \pm 4.6. These metrics represent a model with high predictive performance, capable of predicting microbial drug depletion far better than human guess (50%) in a task that has never been accomplished before. The slightly lower balanced accuracy score may demonstrate that negative cases of drug depletion were more difficult to predict than positive cases, as the other three metrics focus mainly on positive prediction rates (see Section 2.4). The especially high precision score demonstrates that the model is very unlikely to make a type 2 error, in which a positive case of drug depletion is mistaken for a negative case. In practice, a low type 2 error rate is very important, as researchers can trust predictions that identify drugs as having a high risk of microbial depletion. This ensures that resources will not be wasted if only drugs predicted to be depleted are to be experimentally tested.

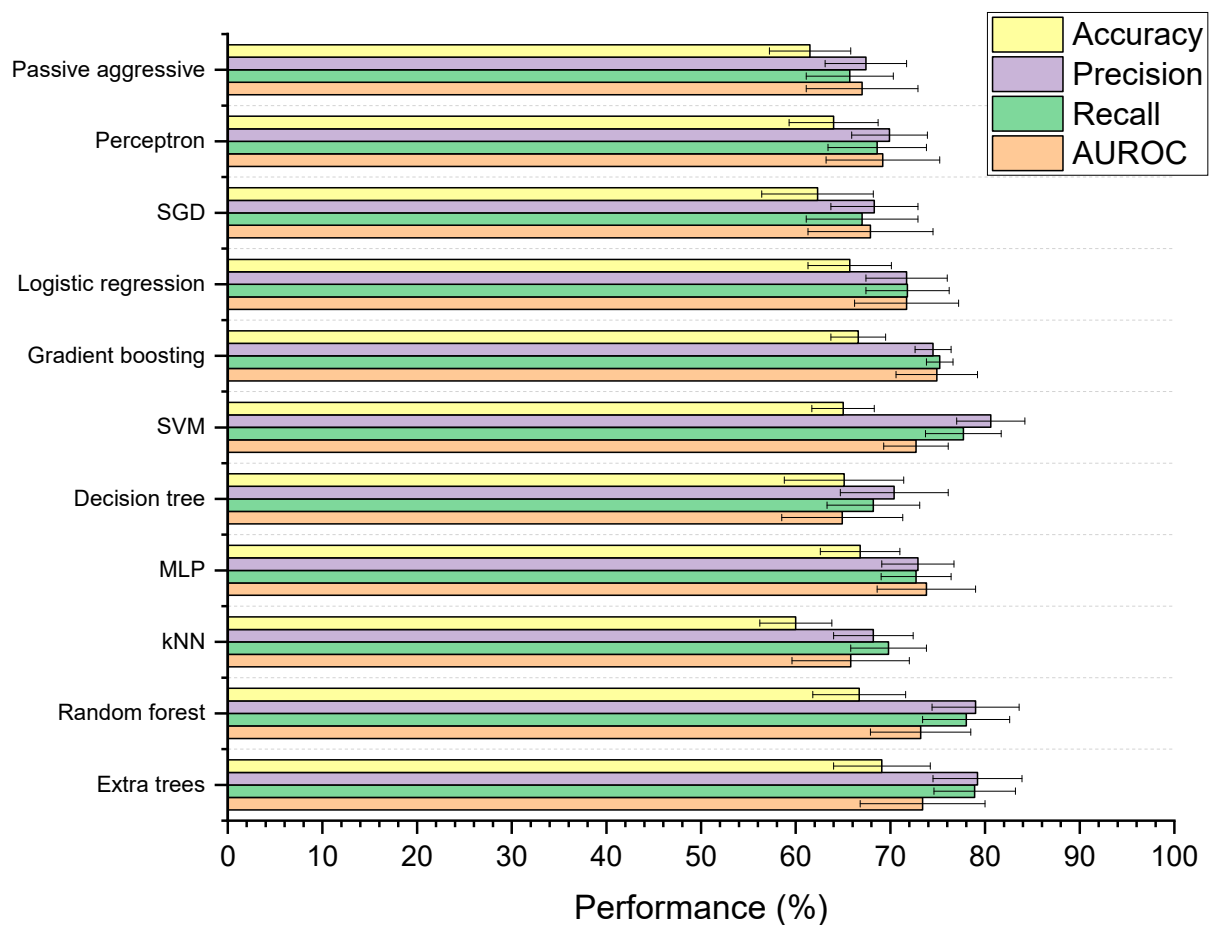


Figure 5. Performance of machine learning (ML) models designed to classify drugs as either susceptible or not susceptible to depletion by gut microbiota. Accuracy is balanced, and precision and recall are weighted. The 11 ML techniques were: extra trees, random forest, K-nearest neighbours (kNN), multilayer perceptron (MLP), decision tree, support vector machine (SVM), gradient boosting, logistic regression, stochastic gradient descent (SGD), perceptron, and passive aggressive classification.

3.4.3. Feature Shuffling

Figure 7 demonstrates the validity of the final model, as the performance metrics were not significantly altered following shuffling of the features within the dataset ($p > 0.95$ for all metrics). These results show that the model's high performance can be trusted, as slight changes to its inputs did not alter predictive performance.

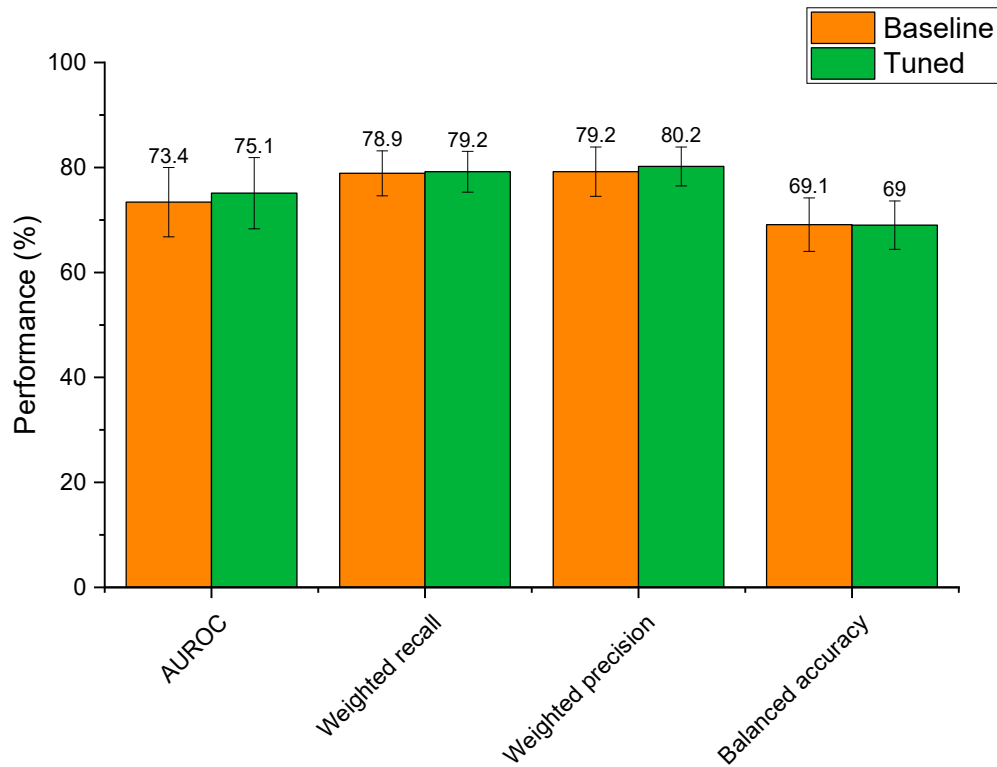


Figure 6. Performance of the baseline extra trees model compared to the extra trees model with hyperparameter tuning.

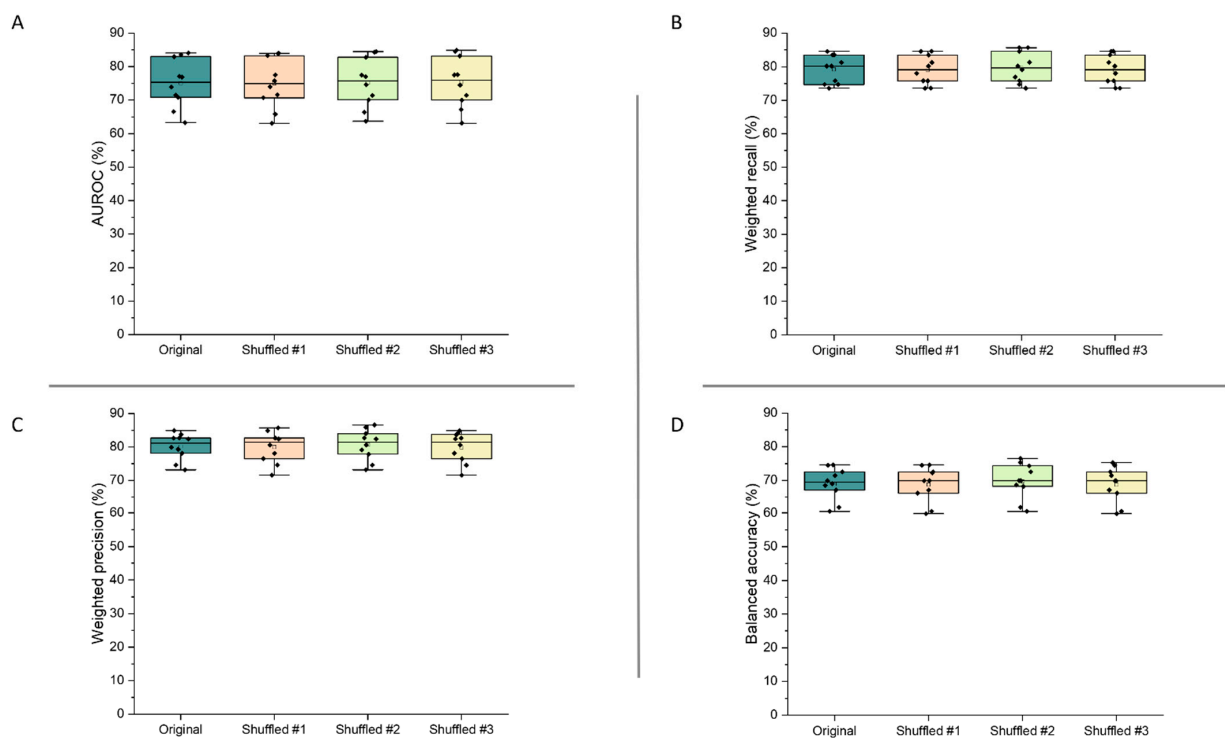


Figure 7. AUROC (A), weighted recall (B), weighted precision (C), and balanced accuracy (D) results from assessment of the final extra trees model by cross-validation, when feature order was shuffled in the training dataset (number of shuffles = 3).

In practice, a positive prediction for drug depletion can be taken to infer a $\geq 60\%$ reduction in drug concentration during exposure to intestinal microbiota for 24 h, as this percentage sits at the intercept of the cut-off points used by Zimmerman et al. ($\geq 20\%$ reduction) and Javdan et al. (total reduction), whose data formed the majority of the ML

model training dataset [1,2]. This information could be a useful tool in the pharmaceutical industry for the rapid screening of investigative molecules or the identification of unknown drug–microbiome interactions in research [48]. Work by Astra Zeneca has highlighted a strong correlation ($R^2 = 0.90$) between drug stability in human faecal slurry and the fraction of drug absorbable in vivo in the colon [53]; therefore, predictions could also provide important information for the development of drugs with targeted release in the lower gastrointestinal tract. The pharmaceutical industry is rapidly embracing ML technology within their processes, and as the first of its kind, this model could offer significant value within the increasingly digitalised drug development pipeline [54–59]. In this model, microbial metabolism and bioaccumulation were combined within the same class because the data from Zimmerman et al. did not differentiate between the two mechanisms [1]. Drug accumulation by intestinal bacteria is a newly recognised concept, and as such sufficient data with which to predict metabolism and bioaccumulation as separate outcomes do not yet exist [3]. In the future, many more instances of bioaccumulation will likely be mapped, allowing distinction between metabolism and bioaccumulation.

4. Conclusions

In this study, a dataset describing 455 drugs' depletion by intestinal microbiota was compiled by extracting data from the literature. Drugs were assigned into two classes (depleted/not depleted), with support from unsupervised learning, and after dataset balancing it was observed that providing physicochemical features as drug descriptors generated ML classifications with high average performance scores and good computational efficiency. The performances of 11 ML models with different learning techniques were compared and an extra trees model was selected for further optimisation, based on its high performance and efficient learning style. After hyperparameter tuning, the model achieved good performances (AUROC: $75.1\% \pm 6.8$; weighted recall: $79.2\% \pm 3.9$; weighted precision: $80.2\% \pm 3.7$; and balanced accuracy: $69.0\% \pm 4.6$) via cross-validation. The model maintained its performance in response to feature shuffling. This ML model is the first to accurately predict whether drugs will or will not be substantially metabolised or accumulated by intestinal microbiota. The algorithm could be a useful tool during the development of new drugs by allowing rapid screening of compounds' susceptibility to microbial interactions that could significantly affect their pharmacokinetics. The code and dataset required to use the ML model can be found in the Supplementary Materials.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/pharmaceutics13122001/s1>, Excel file S1 (training.xlsx): Training dataset for final extra trees model; Jupyter Notebook file S2 (Final code ML depletion (1).ipynb): Python code to use final extra trees model; Excel file S3 (training1.xlsx): Shuffled training dataset no. 1; Excel file S4 (training2.xlsx): Shuffled training dataset no. 2; Excel file S5 (training3.xlsx): Shuffled training dataset no. 3.

Author Contributions: Conceptualization, L.E.M., M.E., S.G., M.O. and A.W.B.; methodology, L.E.M., S.T. and M.E.; software, L.E.M. and M.E.; validation, L.E.M. and M.E.; formal analysis, L.E.M., M.E. and A.W.B.; resources, S.G., M.O. and A.W.B.; data curation, L.E.M. and S.T.; writing—original draft preparation, L.E.M.; writing—review and editing, M.E., S.T. and A.W.B.; visualization, L.E.M. and M.E.; supervision, S.G., M.O. and A.W.B.; funding acquisition, S.G., M.O. and A.W.B. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by The Engineering and Physical Sciences Research Council (grant code EP/S023054/1) to UCL School of Pharmacy.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The training dataset and Python code required to use the final machine learning model can be found in the Supplementary Materials.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zimmermann, M.; Zimmermann-Kogadeeva, M.; Wegmann, R.; Goodman, A.L. Mapping human microbiome drug metabolism by gut bacteria and their genes. *Nature* **2019**, *570*, 462–467. [[CrossRef](#)] [[PubMed](#)]
2. Javdan, B.; Lopez, J.G.; Chankhamjon, P.; Lee, Y.J.; Hull, R.; Wu, Q.; Wang, X.; Chatterjee, S.; Donia, M.S. Personalized Mapping of Drug Metabolism by the Human Gut Microbiome. *Cell* **2020**, *181*, 1661–1679. [[CrossRef](#)]
3. Klünemann, M.; Andrejev, S.; Blasche, S.; Mateus, A.; Phapale, P.; Devendran, S.; Vappiani, J.; Simon, B.; Scott, T.A.; Kafkia, E.; et al. Bioaccumulation of therapeutic drugs by human gut bacteria. *Nature* **2021**, *597*, 533–538. [[CrossRef](#)] [[PubMed](#)]
4. Zimmermann, M.; Zimmermann-Kogadeeva, M.; Wegmann, R.; Goodman, A.L. Separating host and microbiome contributions to drug pharmacokinetics and toxicity. *Science* **2019**, *363*, eaat9931. [[CrossRef](#)]
5. Zimmermann-Kogadeeva, M.; Zimmermann, M.; Goodman, A.L. Insights from pharmacokinetic models of host-microbiome drug metabolism. *Gut Microbes* **2019**, *11*, 587–596. [[CrossRef](#)]
6. Coombes, Z.; Yadav, V.; McCoubrey, L.E.; Freire, C.; Basit, A.W.; Conlan, R.S.; Gonzalez, D. Progestogens Are Metabolized by the Gut Microbiota: Implications for Colonic Drug Delivery. *Pharmaceutics* **2020**, *12*, 760. [[CrossRef](#)]
7. Yadav, V.; Mai, Y.; McCoubrey, L.E.; Wada, Y.; Tomioka, M.; Kawata, S.; Charde, S.; Basit, A.W. 5-Aminolevulinic Acid as a Novel Therapeutic for Inflammatory Bowel Disease. *Biomedicines* **2021**, *9*, 578. [[CrossRef](#)]
8. Chankhamjon, P.; Javdan, B.; Lopez, J.; Hull, R.; Chatterjee, S.; Donia, M.S. Systematic mapping of drug metabolism by the human gut microbiome. *bioRxiv* **2019**, 538215. [[CrossRef](#)]
9. Yadav, V.; Varum, F.; Bravo, R.; Furrer, E.; Basit, A.W. Gastrointestinal stability of therapeutic anti-TNF alpha IgG1 monoclonal antibodies. *Int. J. Pharm.* **2016**, *502*, 181–187. [[CrossRef](#)]
10. Yadav, V.; Gaisford, S.; Merchant, H.A.; Basit, A.W. Colonic bacterial metabolism of corticosteroids. *Int. J. Pharm.* **2013**, *457*, 268–274. [[CrossRef](#)]
11. Wang, J.; Yadav, V.; Smart, A.L.; Tajiri, S.; Basit, A.W. Toward Oral Delivery of Biopharmaceuticals: An Assessment of the Gastrointestinal Stability of 17 Peptide Drugs. *Mol. Pharm.* **2015**, *12*, 966–973. [[CrossRef](#)] [[PubMed](#)]
12. Sousa, T.; Yadav, V.; Zann, V.; Borde, A.; Abrahamsson, B.; Basit, A.W. On the Colonic Bacterial Metabolism of Azo-Bonded Prodrugs of 5-Aminosalicylic Acid. *J. Pharm. Sci.* **2014**, *103*, 3171–3175. [[CrossRef](#)]
13. Sousa, T.; Paterson, R.; Moore, V.; Carlsson, A.; Abrahamsson, B.; Basit, A.W. The gastrointestinal microbiota as a site for the biotransformation of drugs. *Int. J. Pharm.* **2008**, *363*, 1–25. [[CrossRef](#)] [[PubMed](#)]
14. Clarke, G.; Sandhu, K.V.; Griffin, B.T.; Dinan, T.G.; Cryan, J.F.; Hyland, N.P. Gut Reactions: Breaking Down Xenobiotic-Microbiome Interactions. *Pharm. Rev.* **2019**, *71*, 198–224. [[CrossRef](#)]
15. Walsh, J.; Griffin, B.T.; Clarke, G.; Hyland, N.P. Drug-gut microbiota interactions: Implications for neuropharmacology. *Br. J. Pharm.* **2018**, *175*, 4415–4429. [[CrossRef](#)]
16. Vinarov, Z.; Abdallah, M.; Agundez, J.A.G.; Allegaert, K.; Basit, A.W.; Braeckmans, M.; Ceulemans, J.; Corsetti, M.; Griffin, B.T.; Grimm, M.; et al. Impact of gastrointestinal tract variability on oral drug absorption and pharmacokinetics: An UNGAP review. *Eur. J. Pharm. Sci.* **2021**, *162*, 105812. [[CrossRef](#)]
17. Rekdal, V.M.; Bess, E.N.; Bisanz, J.E.; Turnbaugh, P.J.; Balskus, E.P. Discovery and inhibition of an interspecies gut bacterial pathway for Levodopa metabolism. *Science* **2019**, *364*, eaau6323. [[CrossRef](#)]
18. Wilson, I.D.; Nicholson, J.K. Gut microbiome interactions with drug metabolism, efficacy, and toxicity. *Transl. Res.* **2017**, *179*, 204–222. [[CrossRef](#)]
19. Curro, D. The role of gut microbiota in the modulation of drug action: A focus on some clinically significant issues. *Expert Rev. Clin. Pharm.* **2018**, *11*, 171–183. [[CrossRef](#)]
20. Flowers, S.A.; Bhat, S.; Lee, J.C. Potential Implications of Gut Microbiota in Drug Pharmacokinetics and Bioavailability. *Pharmaco-therapy* **2020**, *40*, 704–712. [[CrossRef](#)]
21. Taggart, A.J.; McDermott, B.J.; Roberts, S.D. The Effect of Age and Acetylator Phenotype on the Pharmacokinetics of Sulfasalazine in Patients with Rheumatoid Arthritis. *Clin. Pharmacokinet.* **1992**, *23*, 311–320. [[CrossRef](#)] [[PubMed](#)]
22. Tian, L.; Wang, X.W.; Wu, A.K.; Fan, Y.; Friedman, J.; Dahlin, A.; Waldor, M.K.; Weinstock, G.M.; Weiss, S.T.; Liu, Y.Y. Deciphering functional redundancy in the human microbiome. *Nat. Commun.* **2020**, *11*, 6217. [[CrossRef](#)]
23. Walsh, J.; Olavarria-Ramirez, L.; Lach, G.; Boehme, M.; Dinan, T.G.; Cryan, J.F.; Griffin, B.T.; Hyland, N.P.; Clarke, G. Impact of host and environmental factors on β -glucuronidase enzymatic activity: Implications for gastrointestinal serotonin. *Am. J. Physiol. -Gastrointest. Liver Physiol.* **2020**, *318*, G816–G826. [[CrossRef](#)] [[PubMed](#)]
24. Haiser, H.J.; Seim, K.L.; Balskus, E.P.; Turnbaugh, P.J. Mechanistic insight into digoxin inactivation by *Eggerthella lenta* augments our understanding of its pharmacokinetics. *Gut Microbes* **2014**, *5*, 233–238. [[CrossRef](#)]
25. Haiser, H.J.; Gootenberg, D.B.; Chatman, K.; Sirasani, G.; Balskus, E.P.; Turnbaugh, P.J. Predicting and manipulating cardiac drug inactivation by the human gut bacterium *Eggerthella lenta*. *Science* **2013**, *341*, 295–298. [[CrossRef](#)]
26. Guo, Y.; Crnkovic, C.M.; Won, K.-J.; Yang, X.; Lee, J.R.; Orjala, J.; Lee, H.; Jeong, H. Commensal Gut Bacteria Convert the Immunosuppressant Tacrolimus to Less Potent Metabolites. *Drug Metab. Dispos.* **2019**, *47*, 194. [[CrossRef](#)]
27. Lee, J.R.; Muthukumar, T.; Dadhania, D.; Taur, Y.; Jenq, R.R.; Toussaint, N.C.; Ling, L.; Pamer, E.; Suthanthiran, M. Gut microbiota and tacrolimus dosing in kidney transplantation. *PLoS ONE* **2015**, *10*, e0122399. [[CrossRef](#)] [[PubMed](#)]
28. Hitchings, R.; Kelly, L. Predicting and Understanding the Human Microbiome's Impact on Pharmacology. *Trends Pharm. Sci.* **2019**, *40*, 495–505. [[CrossRef](#)]

29. Ke, W.; Saba, J.A.; Yao, C.H.; Hilzendeger, M.A.; Drangowska-Way, A.; Joshi, C.; Mony, V.K.; Benjamin, S.B.; Zhang, S.; Locasale, J.; et al. Dietary serine-microbiota interaction enhances chemotherapeutic toxicity without altering drug conversion. *Nat. Commun.* **2020**, *11*, 2587. [[CrossRef](#)]
30. Mager, L.F.; Burkhard, R.; Pett, N.; Cooke, N.C.A.; Brown, K.; Ramay, H.; Paik, S.; Stagg, J.; Groves, R.A.; Gallo, M.; et al. Microbiome-derived inosine modulates response to checkpoint inhibitor immunotherapy. *Science* **2020**, *369*, 1481. [[CrossRef](#)] [[PubMed](#)]
31. European Medicines Agency. *ICH Guideline M3(R2) on Non-Clinical Safety Studies for the Conduct of Human Clinical Trials and Marketing Authorisation for Pharmaceuticals*; European Medicines Agency: London, UK, 2009.
32. McCoubrey, L.E.; Gaisford, S.; Orlu, M.; Basit, A.W. Predicting drug-microbiome interactions with machine learning. *Biotechnol. Adv.* **2021**, 107797. [[CrossRef](#)] [[PubMed](#)]
33. McCoubrey, L.E.; Elbadawi, M.; Orlu, M.; Gaisford, S.; Basit, A.W. Harnessing machine learning for development of microbiome therapeutics. *Gut Microbes* **2021**, *13*, 1–20. [[CrossRef](#)]
34. Khan, S.; Hauptman, R.; Kelly, L. Engineering the Microbiome to Prevent Adverse Events: Challenges and Opportunities. *Annu. Rev. Pharmacol. Toxicol.* **2021**, *61*, 159–179. [[CrossRef](#)] [[PubMed](#)]
35. Koppel, N.; Maini Rekdal, V.; Balskus, E.P. Chemical transformation of xenobiotics by the human gut microbiota. *Science* **2017**, *356*, eaag2770. [[CrossRef](#)]
36. Wang, W.; Ye, Z.; Gao, H.; Ouyang, D. Computational pharmaceutics—A new paradigm of drug delivery. *J. Control. Release* **2021**, *338*, 119–136. [[CrossRef](#)] [[PubMed](#)]
37. Lam, K.N.; Alexander, M.; Turnbaugh, P.J. Precision Medicine Goes Microscopic: Engineering the Microbiome to Improve Drug Outcomes. *Cell Host Microbe* **2019**, *26*, 22–34. [[CrossRef](#)]
38. Sharma, A.K.; Jaiswal, S.K.; Chaudhary, N.; Sharma, V.K. A novel approach for the prediction of species-specific biotransformation of xenobiotic/drug molecules by the human gut microbiota. *Sci. Rep.* **2017**, *7*, 9751. [[CrossRef](#)]
39. Elmassry, M.M.; Kim, S.; Busby, B. Predicting drug-metagenome interactions: Variation in the microbial beta-glucuronidase level in the human gut metagenomes. *PLoS ONE* **2021**, *16*, e0244876. [[CrossRef](#)]
40. Hatton, G.B.; Yadav, V.; Basit, A.W.; Merchant, H.A. Animal Farm: Considerations in Animal Gastrointestinal Physiology and Relevance to Drug Delivery in Humans. *J. Pharm. Sci.* **2015**, *104*, 2747–2776. [[CrossRef](#)]
41. Capecchi, A.; Probst, D.; Reymond, J.-L. One molecular fingerprint to rule them all: Drugs, biomolecules, and the metabolome. *J. Cheminformatics* **2020**, *12*, 43. [[CrossRef](#)]
42. Morgan, H.L. The Generation of a Unique Machine Description for Chemical Structures—A Technique Developed at Chemical Abstracts Service. *J. Chem. Doc.* **1965**, *5*, 107–113. [[CrossRef](#)]
43. Reker, D.; Shi, Y.; Kirtane, A.R.; Hess, K.; Zhong, G.J.; Crane, E.; Lin, C.H.; Langer, R.; Traverso, G. Machine Learning Uncovers Food- and Excipient-Drug Interactions. *Cell Rep.* **2020**, *30*, 3710–3716. [[CrossRef](#)]
44. Badillo, S.; Banfai, B.; Birzele, F.; Davydov, I.I.; Hutchinson, L.; Kam-Thong, T.; Siebourg-Polster, J.; Steiert, B.; Zhang, J.D. An Introduction to Machine Learning. *Clin. Pharmacol. Ther.* **2020**, *107*, 871–885. [[CrossRef](#)] [[PubMed](#)]
45. Schrynemackers, M.; Kuffner, R.; Geurts, P. On protocols and measures for the validation of supervised methods for the inference of biological networks. *Front. Genet.* **2013**, *4*, 262. [[CrossRef](#)]
46. Chen, W.; Liu, B.; Peng, S.; Sun, J.; Qiao, X. Computer-Aided Grading of Gliomas Combining Automatic Segmentation and Radiomics. *Int. J. Biomed. Imaging* **2018**, *2018*, 2512037. [[CrossRef](#)]
47. Hsu, T.-Y. Machine learning applied to stock index performance enhancement. *J. Bank. Financ. Technol.* **2021**, *5*, 21–33. [[CrossRef](#)]
48. McCoubrey, L.E.; Elbadawi, M.; Orlu, M.; Gaisford, S.; Basit, A.W. Machine Learning Uncovers Adverse Drug Effects on Intestinal Bacteria. *Pharmaceutics* **2021**, *13*, 1026. [[CrossRef](#)] [[PubMed](#)]
49. Geurts, P.; Ernst, D.; Wehenkel, L. Extremely randomized trees. *Mach. Learn.* **2006**, *63*, 3–42. [[CrossRef](#)]
50. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
51. Liang, Y.; Zhang, S.; Qiao, H.; Yao, Y. iPromoter-ET: Identifying promoters and their strength by extremely randomized trees-based feature selection. *Anal. Biochem.* **2021**, *630*, 114335. [[CrossRef](#)]
52. Andersson, S.; Bathula, D.R.; Iliadis, S.I.; Walter, M.; Skalkidou, A. Predicting women with depressive symptoms postpartum with machine learning methods. *Sci. Rep.* **2021**, *11*, 7877. [[CrossRef](#)]
53. Tannergren, C.; Borde, A.; Borestrom, C.; Abrahamsson, B.; Lindahl, A. Evaluation of an in vitro faecal degradation method for early assessment of the impact of colonic degradation on colonic absorption in humans. *Eur. J. Pharm. Sci.* **2014**, *57*, 200–206. [[CrossRef](#)]
54. Dong, J.; Gao, H.; Ouyang, D. PharmSD: A novel AI-based computational platform for solid dispersion formulation design. *Int. J. Pharm.* **2021**, *604*, 120705. [[CrossRef](#)] [[PubMed](#)]
55. Reker, D. Practical considerations for active machine learning in drug discovery. *Drug Discov. Today Technol.* **2019**, *32–33*, 73–79. [[CrossRef](#)] [[PubMed](#)]
56. Awad, A.; Trenfield, S.J.; Pollard, T.D.; Jie Ong, J.; Elbadawi, M.; McCoubrey, L.E.; Goyanes, A.; Gaisford, S.; Basit, A.W. Connected Healthcare: Improving Patient Care using Digital Health Technologies. *Adv. Drug Deliv. Rev.* **2021**, *178*, 113958. [[CrossRef](#)]
57. Muniz Castro, B.; Elbadawi, M.; Ong, J.J.; Pollard, T.; Song, Z.; Gaisford, S.; Perez, G.; Basit, A.W.; Cabalar, P.; Goyanes, A. Machine learning applied to over 900 3D printed drug delivery systems. *J. Control. Release* **2021**, *590*, 119837. [[CrossRef](#)]

-
58. Elbadawi, M.; McCoubrey, L.E.; Gavins, F.K.H.; Ong, J.J.; Goyanes, A.; Gaisford, S.; Basit, A.W. Disrupting 3D printing of medicines with machine learning. *Trends Pharmacol. Sci.* **2021**, *42*, 745–757. [[CrossRef](#)]
 59. Schuhmacher, A.; Gatto, A.; Hinder, M.; Kuss, M.; Gassmann, O. The upside of being a digital pharma player. *Drug Discov. Today* **2020**, *25*, 1569–1574. [[CrossRef](#)]