



# Machine learning using multi-modal data predicts the production of selective laser sintered 3D printed drug products

Youssef Abdalla, Moe Elbadawi, Mengxuan Ji, Manal Alkahtani, Atheer Awad, Mine Orlu, Simon Gaisford, Abdul W. Basit\*

UCL School of Pharmacy, University College London, 29-39 Brunswick Square, London WC1N 1AX, UK

## ARTICLE INFO

### Keywords:

Additive manufacturing of medicines  
Powder bed fusion 3D printing of pharmaceuticals  
Computational modeling  
Data fusion  
Digital manufacturing and internet of things  
Industry 4.0 and drug delivery

## ABSTRACT

Three-dimensional (3D) printing is drastically redefining medicine production, offering digital precision and personalized design opportunities. One emerging 3D printing technology is selective laser sintering (SLS), which is garnering attention for its high precision, and compatibility with a wide range of pharmaceutical materials, including low-solubility compounds. However, the full potential of SLS for medicines is yet to be realized, requiring expertise and considerable time-consuming and resource-intensive trial-and-error research. Machine learning (ML), a subset of artificial intelligence, is an *in silico* tool that is accomplishing remarkable breakthroughs in several sectors for its ability to make highly accurate predictions. Therefore, the present study harnessed ML to predict the printability of SLS formulations. Using a dataset of 170 formulations from 78 materials, ML models were developed from inputs that included the formulation composition and characterization data retrieved from Fourier-transformed infrared spectroscopy (FT-IR), X-ray powder diffraction (XRPD) and differential scanning calorimetry (DSC). Multiple ML models were explored, including supervised and unsupervised approaches. The results revealed that ML can achieve high accuracies, by using the formulation composition leading to a maximum F1 score of 81.9%. Using the FT-IR, XRPD and DSC data as inputs resulted in an F1 score of 84.2%, 81.3%, and 80.1%, respectively. A subsequent ML pipeline was built to combine the predictions from FT-IR, XRPD and DSC into one consensus model, where the F1 score was found to further increase to 88.9%. Therefore, it was determined for the first time that ML predictions of 3D printability benefit from multi-modal data, combining numeric, spectral, thermogram and diffraction data. The study lays the groundwork for leveraging existing characterization data for developing high-performing computational models to accelerate formulation development.

## 1. Introduction

Three-dimensional (3D) printing is a digitalized fabrication technique that is drastically redefining the drug development pipeline (Awad et al., 2021a). The technology provides unprecedented control in designing drug delivery systems (DDS) compared to conventional fabrication methods (Barakh Ali et al., 2019). In a short period, 3D printing has been successfully demonstrated to print a range of DDS, including microneedles, thin films, microparticles, gastro-retentive electronic devices, cardiac stents, and ocular implants (Awad et al., 2021b; Fenton et al., 2020; Melocchi et al., 2021; Melocchi et al., 2020; Seoane-Viaño et al., 2021; Wang et al., 2021). Moreover, 3D printing can be integrated with other digital technologies, such as a 3D scanner, to achieve personalized medicines and devices with improved patient

compliance (Goyanes et al., 2016). Additionally, the rapid nature of the technology, its digital precision, and the ability to print net-shape products have made 3D printing a desirable technology in addressing research-wide issues, such as reproducibility and sustainability (Ford and Despeisse, 2016). Thus, the benefits of 3D printing are increasingly being realized in both laboratories and clinical settings.

3D printing, also referred to as additive manufacturing, is a collection of seven technologies that share similar traits, such as utilizing computer-aided software to design the print, slicing the model and building the part in a layer-by-layer approach (Awad et al., 2021a; Culmone et al., 2019). Selective laser sintering (SLS) is a newer technology for producing medicines and has attracted considerable attention from industrial organizations distant from pharmaceuticals (Cai et al., 2021). The process utilizes a laser to sinter a free-flowing powder into a

\* Corresponding author.

E-mail address: [a.basit@ucl.ac.uk](mailto:a.basit@ucl.ac.uk) (A.W. Basit).

<https://doi.org/10.1016/j.ijpharm.2023.122628>

Received 29 November 2022; Received in revised form 11 January 2023; Accepted 16 January 2023

Available online 20 January 2023

0378-5173/© 2023 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

monolithic structure. Compared to other 3D printing technologies, SLS can be adapted to operate using FDA-approved materials, has simple pre- and post-processing steps, fine resolutions, and is amenable to large-scale production (Hettesheimer et al., 2018). Moreover, one of the salient features of SLS in medicines is its ability to produce fast-dissolving tablets, with improved solubility and a tunable dissolution profile (Davis et al., 2021; Kulinowski et al., 2022; Trenfield et al., 2023). Other applications of SLS have included gyroid lattices, implants and braille tablets (Fina et al., 2018; Awad et al., 2020; Salmoria et al., 2018). Despite these advantages and the potential of SLS, the technology remains heavily under-explored in pharmaceuticals.

The preparation of SLS feedstock involves mixing the starting materials and then pouring the admixture onto the reservoir platform. The laser will write the 2D cross-sectional design of the desired geometry onto the printing platform, consolidating the powder. After each layer, the roller will apply a fresh layer of powder that is subsequently sintered by the laser. The process proceeds until the desired 3D geometry is obtained (Fina et al., 2017). Fig. 1A summarizes the SLS printing pipeline, despite its simplicity, optimization is needed to ensure that the finished product is defect-free. Examples of anomalies in SLS include both over- and under-sintering, charring, incomplete spreading, misprint and part damage (Scime et al., 2020). Currently, there is no tool to anticipate such defects when using raw pharmaceutical materials, thus resulting in a costly and time-consuming trial-and-error approach to optimize formulations. Powder characterization of the feedstock is performed to assess printability to some degree; however, such analyses require expertise (Fig. 1B). Moreover, while observations have linked the chemical structure and thermal properties of the formulation powder to its sintering behavior, a thorough correlation is yet to be achieved. Indeed, a predictive tool will widen the usability of SLS for the production of medicines, as well as in allied fields.

*In silico* tools offer a potential solution to optimize formulation development, having demonstrated their prowess in other fields. In 3D printing, *in silico* tools such as finite element methods have been used to optimize the printing process (Chen et al., 2017; Ganeriwala and Zohdi, 2016; Ratsimba et al., 2021). However, such methods have limited capacity to comprehend the entire printing process. Alternatively, machine learning (ML), a subset of artificial intelligence (AI), has arisen as a promising candidate for accurately predicting the 3D printability of formulations (Elbadawi et al., 2021b; Ong et al., 2022). ML is a transformative *in silico* model that has been demonstrated to revolutionize numerous sectors, being able to predict future outcomes with an unprecedented degree of accuracy (Trenfield et al., 2022a). In healthcare, ML is used in clinical trials, diagnostics and surgery (Giorgio et al., 2022; Halamka et al., 2022; Myszczyńska et al., 2020; Shah et al., 2019; Zame et al., 2020). In pharmaceuticals, ML models have been applied to model drug-food interactions, drug-microbiome interactions, and formulation development (Gavins et al., 2022; McCoubrey et al., 2021; Wang et al., 2022). For 3D printing medicines, ML has been demonstrated to predict printability, drug release rate, and accelerating quality control (Elbadawi et al., 2020a; Muñoz Castro et al., 2021; O'Reilly et al., 2021). Compared to other *in silico* models, ML can comprehend high-dimensional data, different data formats and fast prediction times.

To that end, ML was applied to predict the printability of SLS formulations to help accelerate the technology's developments. In-house formulations were prepared and characterized using three conventional analytical techniques: Fourier-transformed infrared spectroscopy (FT-IR), X-ray powder diffraction (XRPD) and differential scanning calorimetry (DSC). These characterization techniques were performed to understand the sintering behavior of pharmaceutical-grade powders and were inputted into the ML model to predict SLS printability based on their insight. Unsupervised and supervised machine learning techniques

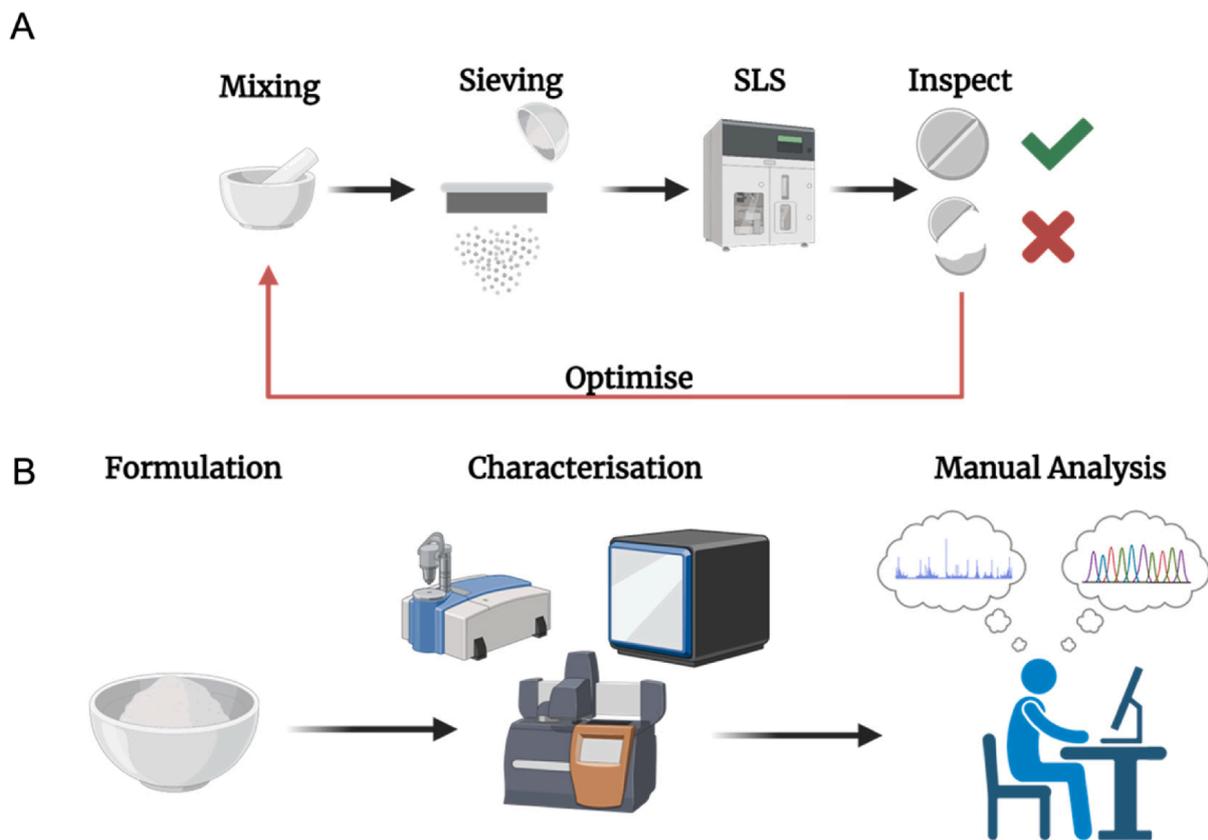


Fig. 1. (A) Schematic of the SLS printing pipeline. Currently, the process requires empirical trials to know if a formulation is printable, which results in both wasted time and resources. Characterization techniques can be performed (B) to provide insight into the formulation, however, the data points generated require expertise to be interpreted.

(MLTs) were explored using four different feature sets. The study set out to determine the optimal ML pipeline for predicting the printability of SLS medicines.

## 2. Materials and methods

### 2.1. Materials

All materials and suppliers are enumerated in Table S1.

### 2.2. Formulation preparation process

Multiple pharmaceutical materials were used to prepare 170 different formulations, based on previously trialed in-house preparations. Formulations contained a variety of medicines and excipients, however, all formulations contained Candurin, a photoabsorbent which is needed for printing using a 2.3 W blue diode laser that operates at a wavelength of 445 nm. Drugs and excipients were weighed separately to make up 15 g of the final product. The materials were first sieved using a 150  $\mu\text{m}$  sieve and then mixed thoroughly using a pestle and mortar for approximately 15 minutes until a uniform colour was obtained. The produced powder was used for product characterization and SLS printing.

### 2.3. Formulation characterization

#### 2.3.1. Fourier-transformation infrared

Attenuated total reflective FT-IR spectra were obtained using a Spectrum 100 spectrometer (PerkinElmer, CT, USA). The formulation powder was added onto the crystal and the force of the arm of the Universal Attenuated Total Reflectance Accessory (UATR) was set to 130. The spectral data was analyzed with the Essential FT-IR software (V3.10.016, Operant LLC, WI, USA). Data was collected over the wavenumber range from 4000 to 650  $\text{cm}^{-1}$ , with a resolution of 2  $\text{cm}^{-1}$  and 4 scans obtained per sample. For data analysis, to reduce the computational demand, only data between 1800 and 650  $\text{cm}^{-1}$ , was used.

#### 2.3.2. X-ray powder diffractometry

The formulation powders were analyzed using XRPD. A Cu K $\alpha$  X-ray source ( $\lambda = 1.5418 \text{ \AA}$ ) was used to gather the XRPD patterns in a Rigaku MiniFlex 600 (Rigaku, TX, USA). The voltage and intensity were 40 kV and 15 mA, respectively. Data acquisition had an angular range of 2–60° 2 $\theta$ , a step size of 0.02°, and a speed of 10°/min.

#### 2.3.3. Differential scanning calorimetry

Powdered samples (5–10 mg) were analyzed using pierced-lid, hermetically sealed aluminum pans (TA instruments, Waters LLC, USA). A Q2000 DSC (V4.5.0.5, TA Instruments, DE, USA) equipped with an autosampler and nitrogen for both cooling and purging (50 mL/min) was used to determine the thermal profile of each formulation in temperatures pertinent to SLS. Following initial acclimatization to 20 °C, the temperature was raised to 200 °C at a heating rate of 10 °C/min.

### 2.4. SLS printing

The formulation powders were transferred to an SLS printer (Sintratec Kit, AG, Brugg, Switzerland) to print the products. Cylindrical discs (10 mm diameter  $\times$  3.6 mm height) were designed on Onshape (Boston, MA, USA). The Standard triangle language (STL) files from the 3D models were exported into the Sintratec central programme for 3D printing. As a starting point, the chamber temperature, the surface temperature, and the laser scanning speed were all set to 100 °C, 80 °C, and 300 mm/s, respectively. SLS printing followed the standard procedure found in the literature (Trenfield et al., 2020). Once printed, the individual discs were left to cool, they were then taken out of the printer

and any extra powder that hadn't been sintered was brushed off. Formulations were considered printable if the produced discs had good structural integrity and shape, with no deformations or charring and maintained integrity during post-printing processing. If a disc was not obtained from the initial printing parameters, then a further three printing attempts were made using expert knowledge to adjust the printing parameters.

### 2.5. Machine learning models

All MLTs were run on a MacBook Pro (Operating System: macOS 12.6; Processor: 2.9 GHz Dual-Core Intel Core i5; RAM Memory: 8 GB). Python (v3.10.4) was used to run the MLTs (Python Software Foundation). All MLTs were deployed using the Scikit-learn (v1.1.1) Python package (Pedregosa et al., 2011), except for extreme gradient boosting which was deployed through its library (xgboost v1.6.1) (Chen and Guestrin, 2016).

Unsupervised learning methods, namely principal component analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE) were used to visualize the high dimensional data in low dimensional space. PCA was also used to reduce the dimensions of the characterization data before applying supervised learning methods to reduce the computational demand of the MLTs.

To determine which MLT is best able to make predictions based on the dataset, nine different supervised MLTs were deployed. These were random forest (RF), logistic regression (LR), support vector machine (SVM), gradient boosting (GB), extreme gradient boosting (XGBoost), decision tree (DT), multilayer perceptron (MLP), k-nearest neighbors (kNN), and extremely randomized trees (EXTr). The MLTs were used for classification - to make a binary decision - whether the formulation is printable using SLS printing. The optimum hyperparameters for each model were found using grid search or MLflow (v1.28.0) (Zaharia et al., 2018).

The accuracy of the predictions made by the models was evaluated using k-fold cross-validation (CV). In this process, data sets are split into pairs of training and test sets. The data is split into  $k$  subsets, or folds, which are approximately the same size. For each fold, the fold is taken as the test data set and the remaining folds as the training data. The model is fitted to the training data and evaluated on the test data, and subsequently, the evaluation score is retained. Once iterated over all the folds, the model is evaluated as the average of the evaluation scores (Ting et al., 2010).

When measuring accuracy, precision and recall should be considered. Precision measures the proportion of positive predictions that are true positives, the equation is:

$$\text{Precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}} \quad (1)$$

Recall is a measure of the proportion of correct predictions that are true positives, the equation is:

$$\text{Recall} = \frac{\text{true positives}}{\text{true positives} + \text{true negatives}} \quad (2)$$

To find a balance between precision and recall the F1 score of accuracy was used for CV (Hossin and Sulaiman, 2015), the equation is:

$$F1 \text{ Score} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (3)$$

Herein, accuracy refers to the F1 score metric.

#### 2.5.1. Principal component analysis

PCA (Jolliffe and Cadima, 2016) is a linear dimensionality reduction technique. It transforms the features into vectors known as principal components which are computed to maximize the variance of the original features, for example, an entire FTIR spectrum can be decomposed as a single 2D point using PCA (Fig. 2). Kernel PCA, an alternative type

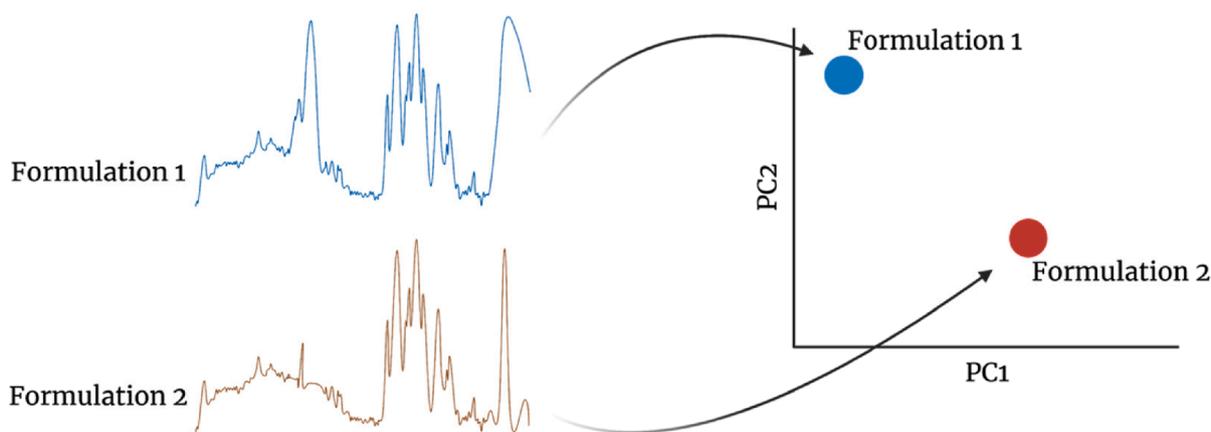


Fig. 2. An illustrative example of the PCA process. In this study, for example, the entire FT-IR data is decomposed into single scatter points in a 2D space. This allows for the intuitive visualization of multiple spectra in one plot.

of PCA which is non-linear was also used. Three different curve functions were trialed - polynomial, radial basis function and cosine.

#### 2.5.2. *t*-distributed Stochastic neighbor embedding

t-SNE (van der Maaten and Hinton, 2008) is another non-linear dimension reduction technique, which is used for exploratory data analysis. t-SNE maps the data around individual points in high-dimensional space using a Gaussian distribution and maps the same points in low-dimensional space using a Student's *t* distribution. A cost function is then used to optimize the two similarity measures. The hyperparameter perplexity - which can be interpreted as the effective number of neighbors the algorithm considers - was tuned between 5 and 50, as recommended by the original paper (van der Maaten and Hinton, 2008).

#### 2.5.3. Logistic regression

LR (Cox, 1958) combines input values linearly and then a logistic function is used to transform predictions. The model outputs a binary value (1 or 0), allowing printability predictions. The hyperparameters tuned were the penalty (l1, l2, elasticnet, none), C parameter (between 0.0001 and 10,000, logarithmic scale), the solver (newton-conjugate gradient, low memory Broyden-Fletcher-Goldfarb-Shanno algorithm (lbfgs), stochastic average gradient (sag), 'saga', 'liblinear') and maximum iterations (between 100 and 5000).

#### 2.5.4. *K*-nearest neighbors

kNN (Mucherino et al., 2009) makes classification based on the label of *k* neighbors in closest proximity to the input. The hyperparameters tuned were the number of neighbors (between 1 and 30) and the power parameter (between 1 and 10).

#### 2.5.5. Support vector machine

SVM (Zhang, 2012) maps the data and creates a hyperplane which best separates the input variable by a class, in this case, whether printing is successful. The hyperparameters tuned were the C parameter (between 0.1 and 100, logarithmic scale),  $\gamma$  (between 0.001 and 1, logarithmic scale) and the kernel (radial basis function, polynomial or sigmoid).

#### 2.5.6. Decision tree

DT (Breiman et al., 2017) utilizes a flowchart-like structure where decisions are made through internal nodes, and the final or 'leaf' node represents a class label - printable or not printable. The hyperparameters tuned were the maximum depth of the tree (between 1 and 30), the minimum number of samples required to split an internal node (between 2 and 20), the minimum number of samples required to be at a

leaf node (between 1 and 20), and the number of features to consider when choosing the best split (between 1 and 12).

#### 2.5.7. Extremely randomized trees

EXTr (Geurts et al., 2006) utilizes multiple decision trees to make decisions, and the final prediction is based on majority voting from the individual decision trees. The hyperparameters tuned were the number of estimators (between 10 and 100), the minimum number of samples required to split an internal node (between 2 and 10), and the number of features to consider when choosing the best split (between 1 and 20).

#### 2.5.8. Random forest

RF (Breiman, 2001) is an ensemble of decision trees, each created from a different bootstrap sample from the data, the final prediction is based on majority voting from the individual decision trees. The hyperparameters tuned were the maximum depth of the tree (between 1 and 40), the minimum number of samples required to split an internal node (between 2 and 20), the minimum number of samples required to be at a leaf node (between 1 and 20), and the number of features to consider when choosing the best split (between 1 and 14, the square root of the number of features or  $\log_2$  of the number of features).

#### 2.5.9. Gradient boosting

GB (Friedman, 2001) utilizes an ensemble of weak learners - decision trees, which are added using a gradient descent-like procedure. The hyperparameters tuned were the number of estimators (between 10 and 100), the learning rate (between 0.0001 and 1), the maximum depth of the tree (between 1 and 20), the minimum number of samples required to split an internal node (between 2 and 20), the minimum number of samples required to be at a leaf node (between 1 and 20), and the subsample used to fit the individual decision trees (0.1 to 1).

#### 2.5.10. Extreme gradient boosting

XGBoost (Chen and Guestrin, 2016) is an optimized GB library which introduces multiple techniques to accelerate model training. The hyperparameters tuned were the number of estimators (between 5 and 100), the learning rate (between 0.0001 and 1), the maximum depth of the trees (between 1 and 10), and the subsample used to fit the individual decision trees (0.1 to 1).

#### 2.5.11. Multilayer perceptron

MLP (Murtagh, 1991) is the most fundamental type of artificial neural network (ANN). The input layer is connected to hidden layers, with each layer feeding forwards into the next, and final classifications are made at the output layer. The hyperparameters tuned were the activation function (identity, logistic, rectified linear tan (ReLU),

hyperbolic tan), the solver for weight optimization (lbfgs, stochastic gradient descent, adam), initial learning rate (between 0.0001 and 1), learning rate (constant, invscaling, adaptive), hidden layer sizes (a random array with length 1–3 was generated with random numbers between 8 and 64) and the maximum number of iterations (between 100 and 800).

### 2.5.12. Ensemble voting classifier

The predictions from multiple models can be combined to make a final prediction, using ‘hard’ or ‘soft’ voting. Hard voting chooses the prediction with the greatest number of votes across the models. On the other hand, soft voting examines the probabilities of each prediction for each model and the final prediction is the one with the highest probability (Kumari et al., 2021).

## 3. Results

### 3.1. Exploratory data analysis

Herein, the aim was to prepare formulations based on previously trialed in-house preparations and identify the ability of different ML algorithms to predict the printability of SLS formulations. A total of 170 in-house formulations were developed and their printability via SLS was evaluated. Before the application of MLTs, exploratory data analysis was performed to provide an overview of the dataset and identify any anomalies that may skew results or lead to over-fitting data. Formulations were prepared from 78 pharmaceutical materials, encompassing drugs, polymers, and other excipients. Table S2 enumerates the frequency of use of each material. Fig. 3A summarizes the distribution of the number of uses of each material in the formulations. 45% of the materials were only used once, Candurin gold sheen or Candurin red sparkle was used in all the formulations as a sintering agent; it is a photo-absorbing excipient which enhances radiation absorption and improves printability (Trenfield et al., 2022b). The most used drug was Paracetamol, which was in 21% of all the formulations. Other materials such as magnesium stearate, mannitol, methylparaben and Kollicoat instant release (IR) were in approximately 20% of the formulations; these were used as lubricants, plasticizers, or binders. Overall, 50.29% of the formulations were successfully printed (Fig. 3B), meaning the target predictions were balanced.

The developed formulations were characterized using FT-IR, XRPD and DSC. Before the application of any MLTs, this data was normalized. Data normalization transforms data within the range (0–1) (Milligan and Cooper, 1985). Normalized data improves the recovery of the class structure, by ensuring that equal weights are attributed to the different inputs (Doherty et al., 2007).

Unsupervised learning using PCA was initially investigated to ascertain whether any inherent clustering is exhibited for printable/non-printable formulations. Linear PCA showed no complete clustering for

FT-IR, XRPD, and DSC, both in 2- and 3-dimensional space (Fig. 4).

However, the data may have a non-linear pattern which cannot be visualized with linear models (McClurkin et al., 1991; Murase and Nayar, 1995). Therefore, a non-linear PCA method was trialed, kernel-PCA. Although exhibiting minor clustering, none of the kernels displayed a clear segregated clustering between printable and unprintable formulations for FT-IR (Fig. 5A), XRPD (Fig. 5B) and DSC (Fig. 5C). The small clusters observed could be due to similar formulations with minor compositional changes.

Since no clustering was observed with kernel-PCA, another non-linear dimensionality reduction model was used – t-SNE. t-SNE has been shown to be significantly better for the visualization of high-dimensional data than traditional methods (van der Maaten and Hinton, 2008). Again, no distinct pattern was observed with the data for FT-IR (Fig. 6A), XRPD (Fig. 6B) and DSC (Fig. 6C).

### 3.2. Predicting printability based on formulation composition

Since no distinct clustering was observed that would lead to accurate ML model development, supervised learning approaches were explored. Initially, MLTs were fed the composition and printability of formulations (Fig. 7A). The MLTs trialed were RF, LR, SVM, XGBoost, DT, MLP kNN and EXTr. CV with 5, 10 and 25 folds were compared (Figure S1). As the number of folds increased, the F1 score increased, and variance reduced (i.e. error bars). However, accuracy was greatest when leave one out cross-validation (LOOCV) was utilized. In LOOCV, the number of folds is equal to the number of instances in the data, which reduces the bias of the predictions (Efron, 1982). Fig. 7B summarizes the F1 score of all the models used. SVM (81.9%), XGBoost (80.7%) and RF (79.5%) had the greatest F1 scores. These models were subsequently ensemble to make a final prediction using ‘soft’ or ‘hard’ voting (Fig. 7C). The F1 score of the hard voting model (83.6%) was greater than that of the soft voting model (83.0%).

### 3.3. Predicting printability using FT-IR, XRPD and DSC

While the MLTs were able to predict printability with good accuracy based on formulation composition, this has limited applications to new formulations containing other materials, or even the same materials from different manufacturers. Therefore, before printing the formulations, powders were characterized with FT-IR, XRPD and DSC and the results were used as inputs for the ML models to predict formulation printability. Before the application of supervised ML models, PCA was used to reduce the dimensions of the data to 25 to decrease the computational demand of the MLTs (Fig. 8A). The MLTs trialed were RF, LR, SVM, XGBoost, DT, MLP kNN and EXTr. The highest-performing models for FT-IR were RF (84.2%), GB (83.0%) and MLP (82.5%) (Fig. 8B), for XRPD they were MLP (81.3%), EXTr (81.3%) and kNN (80.7) (Fig. 8C), and for DSC they were EXTr (80.1%), GB (77.8%) and

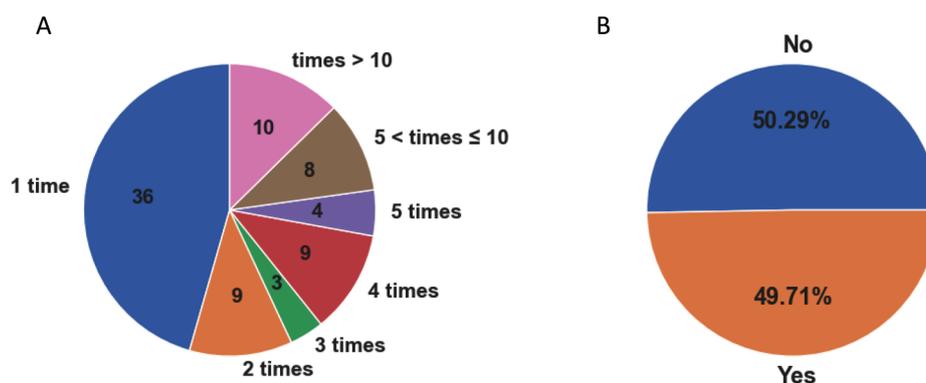
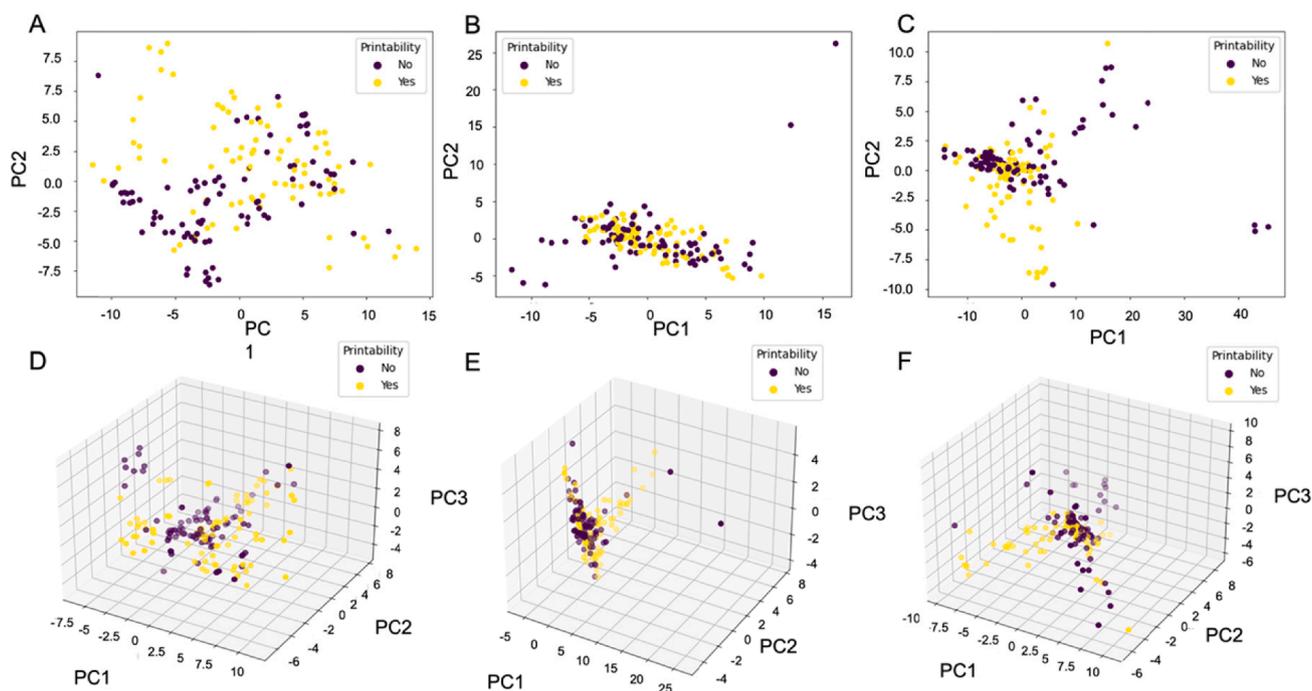


Fig. 3. Exploratory data analysis. A) The distribution of use of different materials in the formulations. B) Printability of different formulations.



**Fig. 4.** Linear PCA for A) FT-IR, B) XRPD and C) DSC and 3D linear PCA for D) FT-IR, E) XRPD and F) DSC. Each scatter point represents a single formulation.

XGBoost (77.2%) (Fig. 8D). Overall, the MLTs showed similar performance with different feature inputs – formulation composition, FT-IR, XRPD and DSC.

To see whether using all three characterization feature sets increases the F1 score, the three data sets were combined, and the MLTs were applied to the combined data (Fig. 9A). Model performance increased with the greatest-performing models being RF (86.5%), LR (83.6%) and SVM (83.0%) (Fig. 9B). As an alternative method to combine feature sets, the best-performing models for each characterization method – RF for FT-IR and EXTr for XRPD and DSC – were chosen and ensembled (Fig. 9C). Prediction accuracy increased further to 87.7% for soft voting and 88.9% for hard voting. Therefore, model performance is greatest when all three characterization methods are used to make separate predictions, which are then combined to make a final prediction on formulation printability.

#### 4. Discussion

Herein, ML was demonstrated to predict the printability of SLS formulations with high accuracies. The successful predictions are expected to expedite formulation development in SLS printing of medicines where knowledge thereof currently lags behind other 3D printing techniques. It was found that different feature sets were able to achieve high predictions, including that of FT-IR which is amenable to real-time measurements. Thus, the present study further expands the type of data that can be used for ML model development in predicting 3D printing outcomes, thereby highlighting the utility of ML to accommodate various data formats.

FT-IR, XRPD and DSC are common characterization techniques in pharmaceutics that provide insight into the formulation to assist in subsequent decision-making. Their high demand has resulted in high-throughput instruments for both rapid and small-sample requirements. In all cases, milligrams of powder were needed, and except for DSC, the analyses were non-destructive. Remarkably, despite their simplicity and low ‘sample cost’, all three characterization methods can provide information that is beyond current human interpretation. Limited studies have found patterns in chemical structures that provide a correlation to the sinterability of polymers and are also limited to a few model

polymers (Bourell et al., 2014; Goodridge et al., 2012). Such data suggests there is potential for the insight generated by the three techniques for model development. Individually, accuracies comparable to ‘material name’ were achieved. FT-IR showed marginally greater accuracy than XRPD and DSC. This may be because the formulations contained a variety of polymers of different crystalline states, including amorphous polymers which are more difficult to characterize using DSC or XRPD. Given that each characterization technique provides different information, the effect of combining all three characterization methods to predict printability was explored. Further integration of all three datasets yielded an increase in the mean F1 score. Thus, highlighting the benefit of using multi-modal data to make predictions. ‘Consulting’ multiple MLTs has been reported to improve model performance, which is analogous to consulting multiple human experts (Alizadeh et al., 2019; Phung and Rhee, 2019). Overall, the present study provides a compelling impetus for the use of multi-modal data for ML development to help improve the F1 score.

More sophisticated models were built to address the small data available for SLS in comparison to the data available for fused deposition modelling (Elbadawi et al., 2020b), and the accuracy herein was notably greater. Understanding the discrepancy in accuracy between the two studies will require further analysis. Suffice it to say, the dataset is smaller in terms of the number of formulations and materials, and additionally, different feature sets were used. A concerted effort has been undertaken by scientists across different disciplines to address the need for ML for small datasets (Elbadawi et al., 2021a). There is undoubtedly a fallacy surrounding the idea that large data is needed for feasible predictions (Fujinuma et al., 2022). An effective model can indeed be built with only a few data instances (Baskin, 2019). However, whether this is possible in pharmaceutics, given the complexity of tasks, remains to be seen. There is likely a balance between the size and quality of data, which is yet to be explored in pharmaceutics. Inputting data from characterization techniques, especially as specific as those that provide fingerprinting information, like FT-IR, can minimize model variability and thus reduce the demand for large datasets. Consensus or ensemble models are also used to improve performance in small datasets, which was for the first time, explored herein for 3D printing medicines (Vanpoucke et al., 2020; Wu et al., 2020). As new

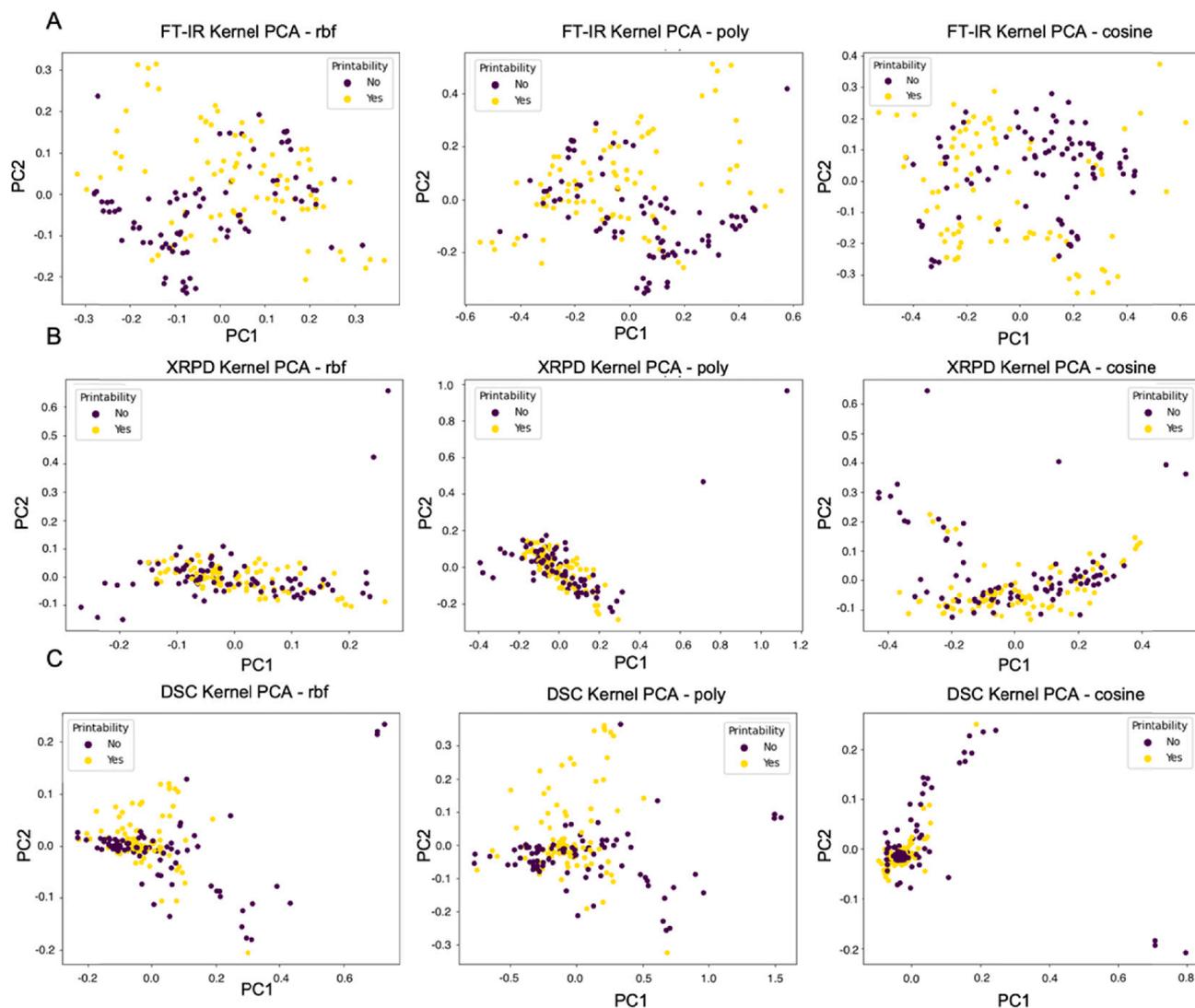


Fig. 5. Kernel PCA for A) FT-IR, B) XRPD and C) DSC. Each scatter point represents a single formulation.

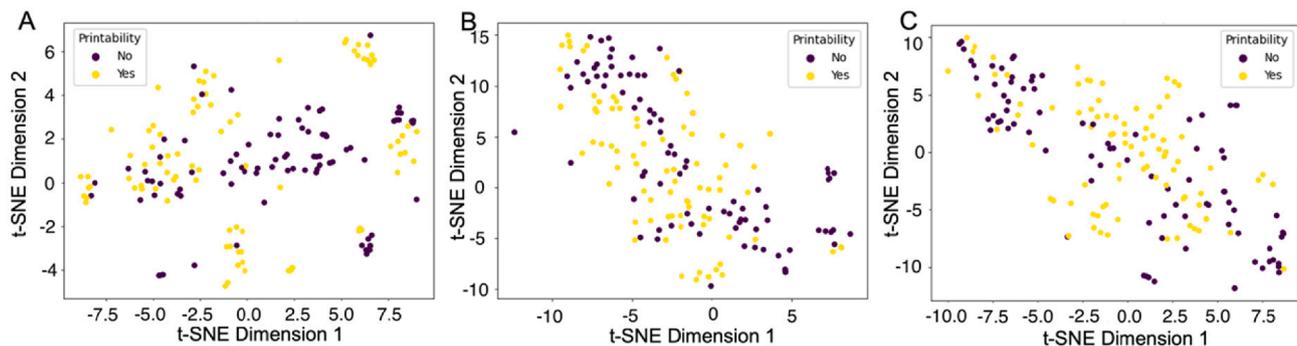


Fig. 6. t-SNE for A) FT-IR, B) XRPD and C) DSC. Each scatter point represents a single formulation.

technologies emerge in 3D printing, ML models for small data will be needed to accelerate developments in the early stages. Future work will seek to elucidate model performance as a function of new formulations, essentially testing its generalizability.

Both supervised and unsupervised MLTs were explored, and the latter was found to achieve accurate predictions. Unsupervised modeling has the salient advantage of overcoming the need to label data (i.e. as printable or not), which itself can be a time-consuming process in

large datasets. However, in agreement with previous work, unsupervised mapping of 3D printing medicines remains ineffective (O'Reilly et al., 2021). Generally, unsupervised models perform weaker than supervised models across many applications but given that unsupervised can reduce the time and cost of the ML pipeline in an already time-consuming and costly research sector, there is a desire to pursue their use.

There is a demand to identify other feature sets to capture as much

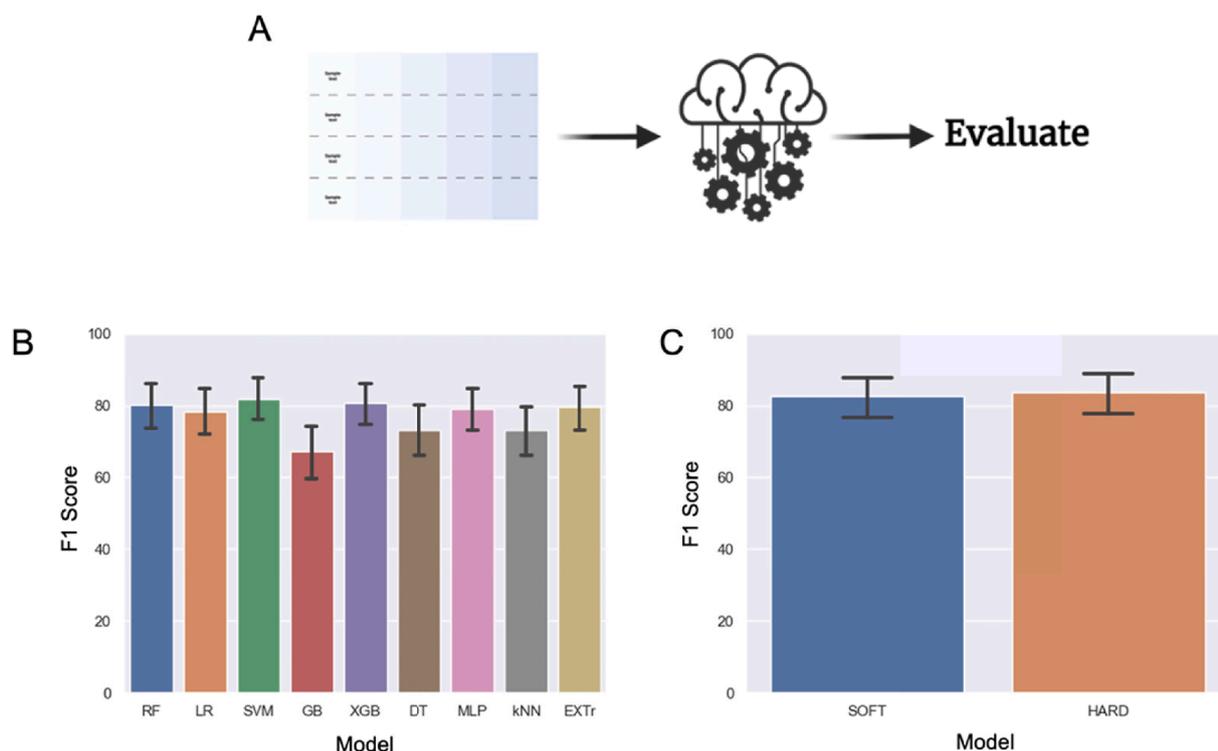


Fig. 7. A) Formulation composition data was fed into the MLTs to predict printability. B) Prediction F1 score measured using leave one out cross-validation. C) Ensemble of the three models that predicted printability with the highest F1 score.

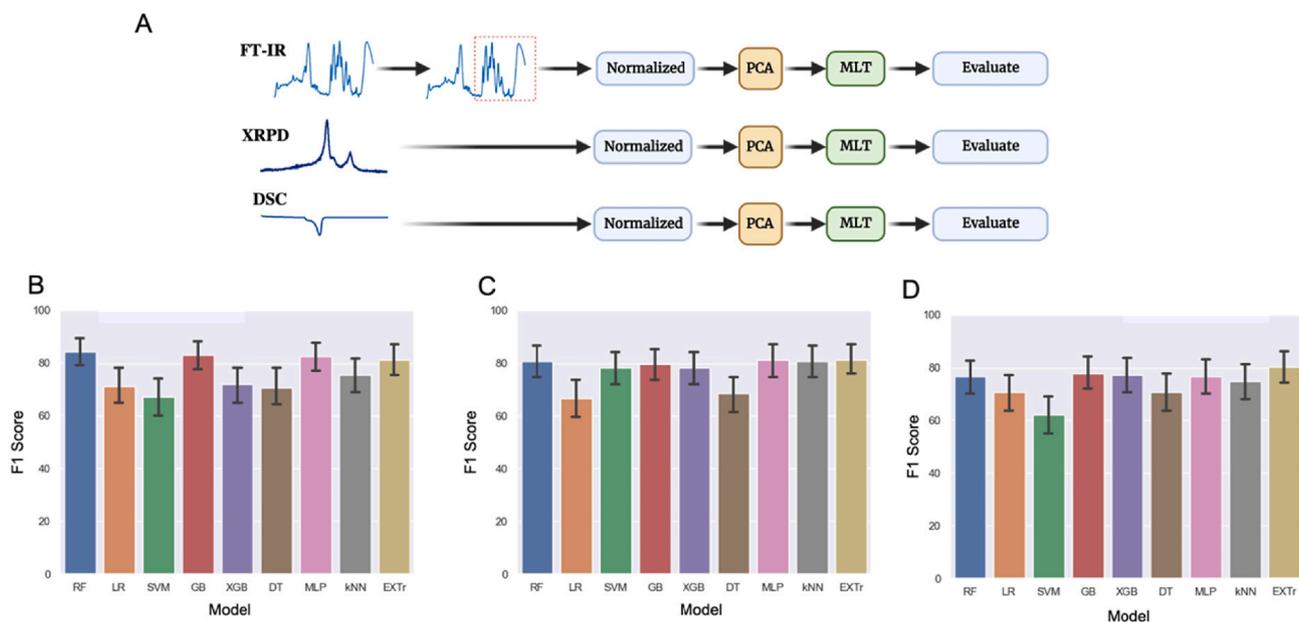
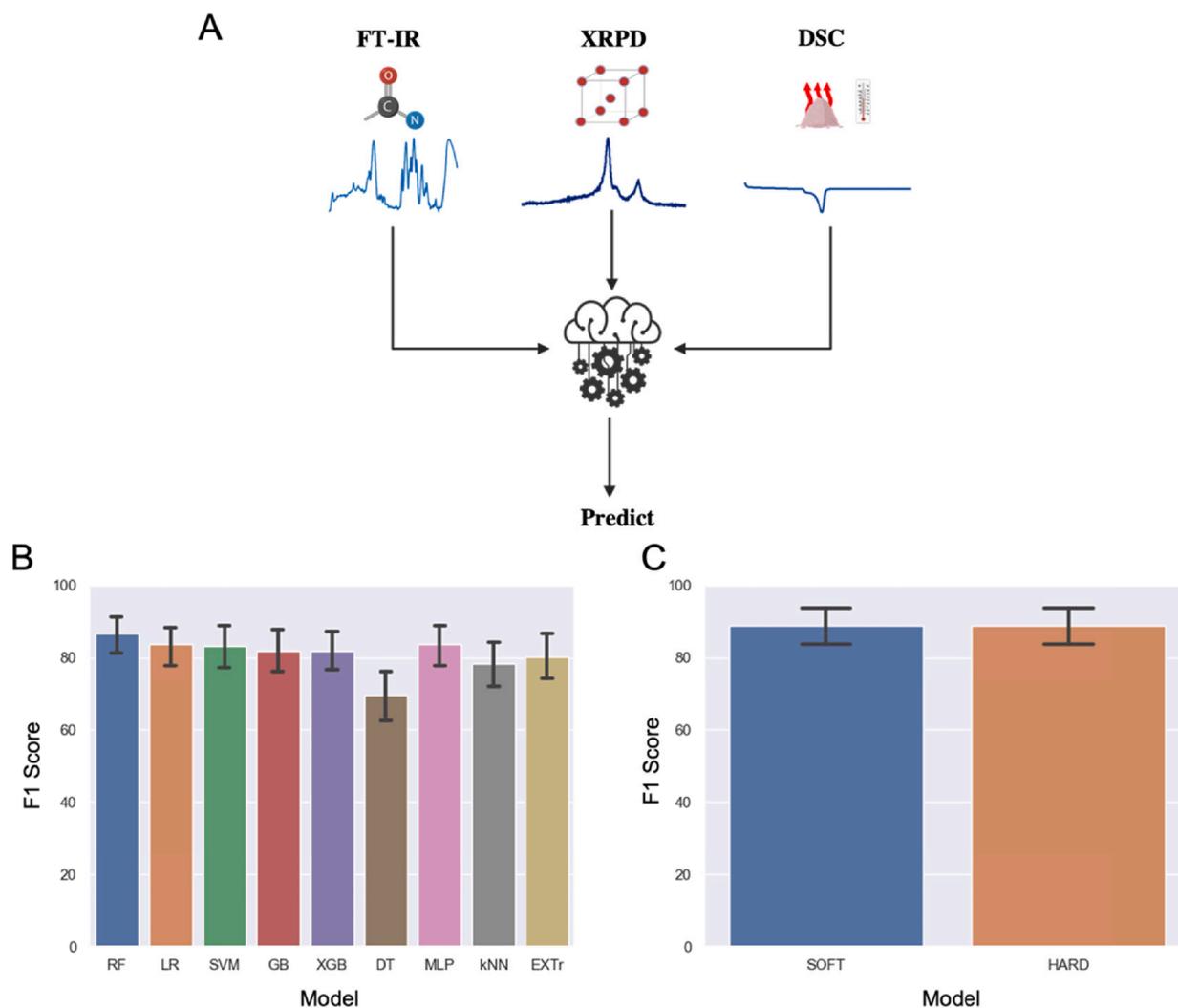


Fig. 8. A) ML Pipeline for making predictions on the printability of formulations using FT-IR, XRPD and DSC data. B) Prediction F1 score measured using LOOCV for FT-IR, C) XRPD and D) DSC measurement.

info as possible and achieve high-performing models. This study is the first to use ML to predict the printability of SLS formulations using multi-modal data, combining numeric, spectral, thermogram and diffraction data, and it does so at a greater accuracy than has previously been shown for other printing methods. The models can be further supplemented with powder characterization data, such as powder flow, to investigate whether this will enhance predictions. Additionally, as an alternative feature set, other characterization methods, which are readily available, such as the Morgan fingerprints of the constituents can also be used to

make predictions, as this removes the need for any lab data and further accelerates drug production.

The present study builds on previous work by the authors in merging two powerful digital technologies: AI and 3D printing; where the ultimate goal is to achieve on-demand, precision medicine. Collectively, they have the potential to transform pharmaceutical research by offering digital precision, personalization, reproducibility, and fast decision-making. Moreover, knowing the printability of a formulation from the outset leads to less waste, and thus a more sustainable research



**Fig. 9.** A) Schematic for FT-IR, XRPD and DSC data fusion process to make predictions. B) Prediction accuracy for all three data sets concatenated and C) an ensemble of the top performing models - RF for FT-IR, EXTr for XRPD and EXTr for DSC.

environment. Considering that SLS can produce intricate formulations, such as multi-layered medicines, there are undoubtedly more exciting opportunities for ML modelling of SLS printing.

## 5. Conclusion

The present study evaluated the ability of different ML models to predict the printability of SLS formulations using multi-modal data to capture more meaningful information about the formulations. Several ML models were employed, including both supervised and unsupervised learning techniques, where the former was revealed to be of greater use. ML successfully predicted the printability of SLS formulations, based on formulation composition, FT-IR, XRPD and DSC data, with a similar accuracy of around 81%. Subsequently, an ML pipeline was built combining predictions from the top-performing FT-IR, XRPD and DSC models in one consensus model. This model outperformed all other models, where prediction accuracy increased to 88.9%. Thus, laying the groundwork as the first study demonstrating that multi-modal data, combining spectral, numeric, thermogram and diffraction data, improves prediction accuracy. Future work will seek to incorporate more characterization data into the ML pipeline to improve model prediction and accelerate development in the 3D printing of medicines.

## CRediT authorship contribution statement

**Youssef Abdalla:** Data curation, Formal analysis, Software, Methodology, Visualization, Writing – original draft, Writing – review & editing. **Moe Elbadawi:** Conceptualization, Data curation, Software, Formal analysis, Funding acquisition, Project administration, Validation, Visualization, Writing – review & editing. **Mengxuan Ji:** Data curation, Formal analysis, Investigation, Methodology. **Manal Alkhatani:** Data curation, Formal analysis. **Atheer Awad:** Conceptualization, Data curation, Writing - review & editing. **Mine Orlu:** Supervision, Validation, Writing – review & editing. **Simon Gaisford:** Conceptualization, Supervision, Validation, Project administration, Funding acquisition. **Abdul W. Basit:** Conceptualization, Supervision, Writing – review & editing.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

The data that has been used is confidential.

## Acknowledgements

Drs Alvaro Goyanes, Fabrizio Fina and Sarah J. Trenfield and Ms Stephanie Chung for developing the in-house formulations used. Dr Moe Elbadawi would also like to thank the Engineering and Physical Sciences Research Council (EPSRC) UK for its financial support (EP/S009000/1).

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ijpharm.2023.122628>.

## References

- Alizadeh, R., Jia, L., Nellippallil, A.B., Wang, G., Hao, J., Allen, J.K., Mistree, F., 2019. Ensemble of surrogates and cross-validation for rapid and accurate predictions using small data sets. *AI. EDAM* 33, 484–501.
- Awad, A., Trenfield, S.J., Pollard, T.D., Ong, J.J., Elbadawi, M., McCoubrey, L.E., Goyanes, A., Gaisford, S., Basit, A.W., 2021a. Connected healthcare: Improving patient care using digital health technologies. *Adv. Drug. Deliv. Rev.* 178, 113958.
- Awad, A., Fina, F., Goyanes, A., Gaisford, S., Basit, A.W., 2021b. Advances in powder bed fusion 3D printing in drug delivery and healthcare. *Adv. Drug. Deliv. Rev.* 174, 406–424.
- Awad, A., Yao, A., Trenfield, S.J., Goyanes, A., Gaisford, S., Basit, A.W., 2020. 3D Printed Tablets (Printlets) with Braille and Moon Patterns for Visually Impaired Patients. *Pharmaceutics* 12, 172.
- Barakh Ali, S.F., Mohamed, E.M., Ozkan, T., Kuttolamadom, M.A., Khan, M.A., Asadi, A., Rahman, Z., 2019. Understanding the effects of formulation and process variables on the printlets quality manufactured by selective laser sintering 3D printing. *Int. J. Pharm.* 570, 118651.
- Baskin, I.I., 2019. Is one-shot learning a viable option in drug discovery? *Expert. Opin. Drug. Discov.* 14, 601–603.
- Bourell, D.L., Watt, T.J., Leigh, D.K., Fulcher, B., 2014. Performance Limitations in Polymer Laser Sintering. *Phys. Procedia* 56, 147–156.
- Breiman, L., 2001. Random Forests. *Mach. Learn.* 45, 5–32.
- Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 2017. Classification and regression trees. Routledge.
- Cai, C., Tey, W.S., Chen, J., Zhu, Y., Liu, X., Liu, T., Zhao, L., Zhou, K., 2021. Comparative study on 3D printing of polyamide 12 by selective laser sintering and multi jet fusion. *J. Mater. Process. Technol.* 288, 116882.
- Chen, H., Wei, Q., Wen, S., Li, Z., Shi, Y., 2017. Flow behavior of powder particles in layering process of selective laser melting: Numerical modeling and experimental verification based on discrete element method. *Int. J. Mach. Tool. Manu* 123, 146–159.
- Chen, T., Guestrin, C., 2016. Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, pp. 785–794.
- Cox, D.R., 1958. The Regression Analysis of Binary Sequences. *Journal of the Royal Statistical Society. Series. B. (Methodological)* 20, 215–242.
- Culmone, C., Smit, G., Breedveld, P., 2019. Additive manufacturing of medical instruments: A state-of-the-art review. *Addit. Manuf.* 27, 461–473.
- Davis Jr., D.A., Thakkar, R., Su, Y., Williams 3rd, R.O., Maniruzzaman, M., 2021. Selective Laser Sintering 3-Dimensional Printing as a Single Step Process to Prepare Amorphous Solid Dispersion Dosage Forms for Improved Solubility and Dissolution Rate. *J. Pharm. Sci* 110, 1432–1443.
- Doherty, K.A.J., Adams, R.G., Davey, N., 2007. Unsupervised learning with normalised data and non-Euclidean norms. *Appl. Soft. Comput.* 7, 203–210.
- The Jackknife, the Bootstrap and other resampling plans, CBMS-NSF regional conference series in applied mathematics 1982. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.
- Elbadawi, M., Gaisford, S., Basit, A.W., 2021a. Advanced machine-learning techniques in drug discovery. *Drug. Discov. Today* 26, 769–777.
- Elbadawi, M., Gustafsson, T., Gaisford, S., Basit, A.W., 2020a. 3D printing tablets: Predicting printability and drug dissolution from rheological data. *Int. J. Pharm.* 590, 119868.
- Elbadawi, M., McCoubrey, L.E., Gavins, F.K.H., Ong, J.J., Goyanes, A., Gaisford, S., Basit, A.W., 2021b. Harnessing artificial intelligence for the next generation of 3D printed medicines. *Adv. Drug. Deliv. Rev.* 175, 113805.
- Elbadawi, M., Muñiz Castro, B., Gavins, F.K.H., Ong, J.J., Gaisford, S., Pérez, G., Basit, A.W., Cabalar, P., Goyanes, A., 2020b. M3DISEN: A novel machine learning approach for predicting the 3D printability of medicines. *Int. J. Pharm.* 590, 119837.
- Fenton, O.S., Paolini, M., Andresen, J.L., Müller, F.J., Langer, R., 2020. Outlooks on three-dimensional printing for ocular biomaterials research. *J. Ocul. Pharmacol. Ther.* 36, 7–17.
- Fina, F., Goyanes, A., Gaisford, S., Basit, A.W., 2017. Selective laser sintering (SLS) 3D printing of medicines. *Int. J. Pharm.* 529, 285–293.
- Fina, F., Goyanes, A., Madla, C.M., Awad, A., Trenfield, S.J., Kuek, J.M., Patel, P., Gaisford, S., Basit, A.W., 2018. 3D printing of drug-loaded gyroid lattices using selective laser sintering. *Int. J. Pharm.* 547, 44–52.
- Ford, S., Despeisse, M., 2016. Additive manufacturing and sustainability: an exploratory study of the advantages and challenges. *J. Clean. Prod.* 137, 1573–1587.
- Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. *Ann. Stat.* 1189–1232.
- Fujinuma, N., DeCost, B., Hatrick-Simpers, J., Lofland, S.E., 2022. Why big data and compute are not necessarily the path to big materials science. *Commun. Mater.* 3, 59.
- Ganeriwala, R., Zohdi, T.L., 2016. A coupled discrete element-finite difference model of selective laser sintering. *Granul. Matter* 18, 21.
- Gavins, F.K.H., Fu, Z., Elbadawi, M., Basit, A.W., Rodrigues, M.R.D., Orlu, M., 2022. Machine learning predicts the effect of food on orally administered medicines. *Int. J. Pharm.* 611, 121329.
- Geurts, P., Ernst, D., Wehenkel, L., 2006. Extremely randomized trees. *Mach. Learn.* 63, 3–42.
- Giorgio, J., Jagust, W.J., Baker, S., Landau, S.M., Tino, P., Kourtzi, Z., Alzheimer's Disease Neuroimaging, I., 2022. A robust and interpretable machine learning approach using multimodal biological data to predict future pathological tau accumulation. *Nat. Commun.* 13, 1887.
- Goodridge, R.D., Tuck, C.J., Hague, R.J.M., 2012. Laser sintering of polyamides and other polymers. *Prog. Mater. Sci.* 57, 229–267.
- Goyanes, A., Det-Amornrat, U., Wang, J., Basit, A.W., Gaisford, S., 2016. 3D scanning and 3D printing as innovative technologies for fabricating personalized topical drug delivery systems. *J. Control. Release* 234, 41–48.
- Halamka, J., Bydon, M., Cerrato, P., Bhagra, A., 2022. Addressing racial disparities in surgical care with machine learning. *npj. Digital. Med.* 5, 152.
- Hettesheimer, T., Hirzel, S., Roß, H.B., 2018. Energy savings through additive manufacturing: an analysis of selective laser sintering for automotive and aircraft components. *Energ. Eff.* 11, 1227–1245.
- Hossin, M., Sulaiman, M.N., 2015. A Review on Evaluation Metrics for Data Classification Evaluations. *Int. J. Data. Min. Knowl. Manage. Process* 5, 01–11.
- Jolliffe, I.T., Cadima, J., 2016. Principal component analysis: a review and recent developments. *Philos. Trans. A. Math. Phys. Eng. Sci* 374, 20150202.
- Kulinowski, P., Malczewski, P., Laszcz, M., Baran, E., Milanowski, B., Kuprianowicz, M., Dorozynski, P., 2022. Development of Composite, Reinforced, Highly Drug-Loaded Pharmaceutical Printlets Manufactured by Selective Laser Sintering-In Search of Relevant Excipients for Pharmaceutical 3D Printing. *Materials (Basel)* 15.
- Kumari, S., Kumar, D., Mittal, M., 2021. An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier. *Int. J. Cogn. Comput. Eng.* 2, 40–46.
- McClurkin, J.W., Optican, L.M., Richmond, B.J., Gawne, T.J., 1991. Concurrent processing and complexity of temporally encoded neuronal messages in visual perception. *Science* 253, 675–677.
- McCoubrey, L.E., Thomaidou, S., Elbadawi, M., Gaisford, S., Orlu, M., Basit, A.W., 2021. Machine Learning Predicts Drug Metabolism and Bioaccumulation by Intestinal Microbiota. *Pharmaceutics* 13, 2001.
- Melocchi, A., Briatico-Vangosa, F., Uboldi, M., Parietti, F., Turchi, M., von Zeppelin, D., Maroni, A., Zema, L., Gazzaniga, A., Zidan, A., 2021. Quality considerations on the pharmaceutical applications of fused deposition modeling 3D printing. *Int. J. Pharm.* 592, 119901.
- Melocchi, A., Uboldi, M., Cerea, M., Poppoli, A., Maroni, A., Moutaharrik, S., Palugan, L., Zema, L., Gazzaniga, A., 2020. A Graphical Review on the Escalation of Fused Deposition Modeling (FDM) 3D Printing in the Pharmaceutical Field. *J. Pharm. Sci.* 109, 2943–2957.
- Milligan, G.W., Cooper, M.C., 1985. An examination of procedures for determining the number of clusters in a data set. *Psychometrika* 50, 159–179.
- Mucherino, A., Papajorgji, P.J., Pardalos, P.M., 2009. k-Nearest Neighbor Classification. In: Mucherino, A., Papajorgji, P.J., Pardalos, P.M. (Eds.), *Data Mining in Agriculture*. Springer, New York, New York, NY, pp. 83–106.
- Muniz Castro, B., Elbadawi, M., Ong, J.J., Pollard, T., Song, Z., Gaisford, S., Pérez, G., Basit, A.W., Cabalar, P., Goyanes, A., 2021. Machine learning predicts 3D printing performance of over 900 drug delivery systems. *J. Control. Release* 337, 530–545.
- Murase, H., Nayar, S.K., 1995. Visual learning and recognition of 3-D objects from appearance. *Int. J. Comput. Vis.* 14, 5–24.
- Murtagh, F., 1991. Multilayer perceptrons for classification and regression. *Neurocomputing* 2, 183–197.
- Myszczyńska, M.A., Ojames, P.N., Lacoste, A.M.B., Neil, D., Saffari, A., Mead, R., Hautbergue, G.M., Holbrook, J.D., Ferraiuolo, L., 2020. Applications of machine learning to diagnosis and treatment of neurodegenerative diseases. *Nat. Rev. Neurol.* 16, 440–456.
- O'Reilly, C.S., Elbadawi, M., Desai, N., Gaisford, S., Basit, A.W., Orlu, M., 2021. Machine Learning and Machine Vision Accelerate 3D Printed Orodispersible Film Development, *Pharmaceutics*.
- Ong, J.J., Castro, B.M., Gaisford, S., Cabalar, P., Basit, A.W., Pérez, G., Goyanes, A., 2022. Accelerating 3D printing of pharmaceutical products using machine learning. *Int. J. Pharm.* X 4, 100120.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., 2011. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.
- Phung, V.H., Rhee, E.J., 2019. A High-Accuracy Model Average Ensemble of Convolutional Neural Networks for Classification of Cloud Image Patches on Small Datasets. *Appl. Sci.* 9, 4500.
- Ratsimba, A., Zerrouki, A., Tessier-Doyen, N., Nait-Ali, B., André, D., Dupont, P., Neveu, A., Tripathi, N., Franconi, F., Delaizir, G., 2021. Densification behaviour and three-dimensional printing of Y2O3 ceramic powder by selective laser sintering. *Ceram. Int.* 47, 7465–7474.
- Salmoria, G.V., Vieira, F.E., Muenz, E.A., Gindri, I.M., Marques, M.S., Kanis, L.A., 2018. Additive Manufacturing of PE/fluorouracil/progesterone intrauterine device for endometrial and ovarian cancer treatments. *Polym. Test.* 71, 312–317.

- Scime, L., Siddel, D., Baird, S., Paquit, V., 2020. Layer-wise anomaly detection and classification for powder bed additive manufacturing processes: A machine-agnostic algorithm for real-time pixel-wise semantic segmentation. *Addit. Manuf.* 36, 101453.
- Seoane-Viaño, I., Trenfield, S.J., Basit, A.W., Goyanes, A., 2021. Translating 3D printed pharmaceuticals: From hype to real-world clinical applications. *Adv. Drug. Deliv. Rev.* 174, 553–575.
- Shah, P., Kendall, F., Khozin, S., Goosen, R., Hu, J., Laramie, J., Ringel, M., Schork, N., 2019. Artificial intelligence and machine learning in clinical development: a translational perspective. *npj. Digital. Med.* 2, 69.
- Ting, J.-A., Vijayakumar, S., Schaal, S., 2010. Cross-Validation, in: Sammut, C., Webb, G. I. (Eds.), *Encyclopedia of Machine Learning*. Springer US, Boston, MA, pp. 249-249.
- Trenfield, S.J., Awad, A., McCoubrey, L.E., Elbadawi, M., Goyanes, A., Gaisford, S., Basit, A.W., 2022a. Advancing pharmacy and healthcare with virtual digital technologies. *Adv. Drug. Deliv. Rev.* 182, 114098.
- Trenfield, S.J., Januskaite, P., Goyanes, A., Wilsdon, D., Rowland, M., Gaisford, S., Basit, A.W., 2022b. Prediction of Solid-State Form of SLS 3D Printed Medicines Using NIR and Raman Spectroscopy. *Pharmaceutics* 14.
- Trenfield, S.J., Tan, H.X., Goyanes, A., Wilsdon, D., Rowland, M., Gaisford, S., Basit, A. W., 2020. Non-destructive dose verification of two drugs within 3D printed polyprintlets. *Int. J. Pharm.* 577, 119066.
- Trenfield, S.J., Xu, X., Goyanes, A., Rowland, M., Wilsdon, D., Gaisford, S., Basit, A.W., 2023. Releasing fast and slow: Non-destructive prediction of density and drug release from SLS 3D printed tablets using NIR spectroscopy. *Int. J. Pharm.* X 5, 100148.
- van der Maaten, L., Hinton, G., 2008. Visualizing data using t-SNE. *J. Mach. Learn. Res.* 9, 2579–2605.
- Vanpoucke, D.E.P., Knippenberg, O.S.J.v., Hermans, K., Bernaerts, K.V., Mehrkanoon, S., 2020. Small data materials design with machine learning: When the average model knows best. *J. Appl. Phys.* 128, 054901.
- Wang, C., Zhang, L., Fang, Y., Sun, W., 2021. Design, Characterization, and 3D Printing of Cardiovascular Stents with Zero Poisson's Ratio in Longitudinal Deformation. *Engineering* 7, 979–990.
- Wang, F., Elbadawi, M., Tsilova, S.L., Gaisford, S., Basit, A.W., Parhizkar, M., 2022. Machine learning predicts electrospray particle size. *Mater. Des.* 219, 110735.
- Wu, Z., Zhu, M., Kang, Y., Leung, E.-L.-H., Lei, T., Shen, C., Jiang, D., Wang, Z., Cao, D., Hou, T., 2020. Do we need different machine learning algorithms for QSAR modeling? A comprehensive assessment of 16 machine learning algorithms on 14 QSAR data sets. *Brief. Bioinform.* 22.
- Zaharia, M., Chen, A., Davidson, A., Ghodsi, A., Hong, S.A., Konwinski, A., Murching, S., Nykodym, T., Ogilvie, P., Parkhe, M., 2018. Accelerating the machine learning lifecycle with MLflow. *IEEE. Data. Eng. Bull.* 41, 39–45.
- Zame, W.R., Bica, I., Shen, C., Curth, A., Lee, H.-S., Bailey, S., Weatherall, J., Wright, D., Bretz, F., van der Schaar, M., 2020. Machine learning for clinical trials in the era of COVID-19. *Statistics. Biopharm. Res.* 12, 506–517.
- Zhang, Y., 2012. Support Vector Machine Classification Algorithm and Its Application. In: Liu, C., Wang, L., Yang, A. (Eds.), *Information Computing and Applications*. Springer, Berlin Heidelberg, Berlin, Heidelberg, pp. 179–186.