B. Muñiz Castro, M. Elbadawi, J. J. Ong, T. Pollard, Z. Song, S. Gaisford, G. Pérez, A. W. Basit, P. Cabalar, and A. Goyanes, "Machine learning predicts 3D printing performance of over 900 drug delivery systems", Journal of Controlled Release, Vol. 337, pp. 530-545, 10 Sept. 2021, doi: 10.1016/j.jconrel.2021.07.046

Link to published version: https://doi.org/10.1016/j.jconrel.2021.07.046

Machine Learning Applied to over 900 3D Printed Drug Delivery Systems

Brais Muñiz Castro[1a], Moe Elbadawi[2a], Jun Jie Ong[2], Thomas Pollard[2], Zhe Song[2], Simon Gaisford[2,3], Gilberto Pérez[1], Abdul W. Basit[2, 3, *], Pedro Cabalar[4], Alvaro Goyanes[2,3,5, *]

[1]IRLab, CITIC Research Center, Department of Computer Science, University of A Coruña, Spain

[2]Department of Pharmaceutics, UCL School of Pharmacy, University College London, 29-39 Brunswick Square, London WC1N 1AX, UK.

[3]FabRx Ltd., Henwood House, Henwood, Ashford, Kent, England, TN24 8DH, UK.

[4]IRLab, Department of Computer Science, University of A Coruña, Spain.

[5]Departamento de Farmacología, Farmacia y Tecnología Farmacéutica, I+D Farma (GI-1645), Facultad de Farmacia, Health Research Institute of Santiago de Compostela (IDIS), Universidade de Santiago de Compostela, 15782, Spain.

[a] These authors contributed equally to this work.

[*] Corresponding author at: UCL School of Pharmacy, University College London,

29-39 Brunswick Square, London WC1 N 1AX, UK.

E-mail addresses: a.goyanes@FabRx.co.uk (A. Goyanes), a.basit@ucl.ac.uk (A.W. Basit).

gilberto.pvega@udc.es (G. Pérez)

Abstract

Three-dimensional printing (3DP) is a transformative technology that is advancing pharmaceutical research by producing personalized drug products. However, advances made via 3DP have been slow due to the lengthy trial-and-error approach in optimization. Artificial intelligence (AI) is a technology that could revolutionize pharmaceutical 3DP through analyzing large datasets. Herein, literature-mined data for developing AI machine learning (ML) models was used to predict key aspects of the 3DP formulation pipeline and *in vitro* dissolution properties. A total of 968 formulations were mined and assessed from 114 articles. The ML techniques explored were able to learn and provide accuracies as high as 93% for values in the filament hot melt extrusion process. In addition, ML algorithms were able to use data from the composition of the formulations with additional input features to predict the drug release of 3D printed formulations. The best prediction was obtained by an artificial neural network that was able to predict drug release times of a formulation with a mean error of ±24.29 minutes. In addition, the most important variables were revealed, which could be leveraged in formulation development. Thus, it was concluded that ML proved to be a suitable approach to modelling the 3D printing workflow.

Keywords: additive manufacturing and continuous manufacturing, personalized and precision pharmaceuticals, machine learning and predictive analysis, digital health and digital technologies, fused filament fabrication, drug delivery

1 Introduction

Three-dimensional printing (3DP), or additive manufacturing, is a cutting-edge fabrication technology that involves the layer-by-layer fabrication of a 3D object based on a computer-aided design (CAD) model [1-6]. Since the approval of the first 3D printed medicine, Spritam®, 3DP has been touted as the next disruptor of the pharmaceutical manufacturing industry [7, 8]. Promising bespoke medicines with precise dosing, pharmaceutical 3DP may contribute to the clinical goal of precision medicines, allowing every individual to be able to receive the right dose at the right time [9-14]. The growing interest in this field has led to an ever-expanding number of 3DP technologies deemed suitable for fabricating tailored medicines. These can be grouped based on the technique; (1) Material Extrusion, which includes Fused Filament Fabrication (better known as Fused Deposition Modelling (FDM™)) [15-20], Semi-solid Extrusion (SSE) [21-25], and Direct Powder Extrusion (DPE) [26, 27]; (2) Powder Bed Fusion, which includes Selective Laser Sintering (SLS) [28-32]; (3) VAT Photopolymerization, which includes Stereolithography (SLA) [33-36]; and (4) Material Jetting, which includes Inkjet Printing (IJP) [37-41]. Each of these technologies possess unique features and advantages; for example, IJP is capable of printing unique patterns such as QR codes that can help in the international war against counterfeit medicines [42, 43]. Amongst these, FDM is the most actively explored 3DP technology in pharmaceutics [7, 44-46].

FDM is a thermal material extrusion technology whose popularity is mainly attributed to its affordability, versatility and compact size [7, 17, 47]. It involves processing raw pharmaceutical material through hot melt extrusion (HME) to obtain long strands of filament, which are subsequently fed into an FDM 3D printer [48]. The printer melts the filament and it is deposited layer-by-layer onto a build plate to create a 3D object. The size and shape of the object can be easily modified using software. This technology has been used within the pharmaceutical arena to produce an array of drug products, ranging from printlets (3D printed tablets) [49] and capsules [13], to transdermal microneedles [50], subcutaneous implants [51], and other innovative drug delivery devices [52-55]. Yet, developments in pharmaceutical FDM 3DP has been hampered by the empirical process of formulation development. Numerous parameters within this two-step process can influence the performance of the final product. These include, but are not limited to, pre-HME variables (e.g. proportion of materials, object design), HME variables (e.g. extrusion temperature, torque, extrusion speed), and FDM 3DP variables (e.g. printing speed, printing temperature, platform temperature) [56, 57]. Consequently, in order to produce the desired product, researchers must undergo a process of

trial-and-error, slowly adjusting each parameter one at a time and evaluating the performance of each prototype. Not only is this time-consuming and inefficient, it also necessitates large amounts of material waste and monetary costs.

Therefore, to have a means of predicting the optimal parameters that will produce the 3D printed object with the best performance would be desirable. Machine Learning (ML) may hold the key to optimising this process [58, 59]. ML is an Artificial Intelligence (AI)-based, *state-of-the-art* technology that enables pattern recognition from complex datasets [60-63]. Recent years have seen AI receive immense and well-deserved media coverage, owing to its successes in affording unparalleled insights and enhanced efficiency in numerous disciplines. For instance, Google DeepMind's AI program (AlphaFold) determines the 3D shapes of proteins from its amino-acid sequence, potentially saving computational biologists time and resources compared to existing lab techniques such as X-ray crystallography [64]. Successful applications of AI in other sectors have prompted the pharmaceutical industry to re-evaluate the traditional costly and time-consuming process of bringing drugs into market [65-69]. Indeed, AI is a versatile and revolutionary technology that warrants consideration for accelerating and transforming pharmaceutical 3DP [70].

We have previously reported an AI-based web application, named M3DISEEN (http://m3diseen.com), that employs five ML techniques to enhance the efficiency of FDM formulation development [71]. This software was successful at predicting four key process parameters: extrusion temperature, filament mechanical characteristics, printing temperature and printability. The dataset comprised a total of 614 drug-loaded formulations evaluated by expert HME and FDM operators from University College London – School of Pharmacy and the company FabRx, using 145 excipients and drugs. An advantage of ML is its ability to improve its predictive performance as the sample size increases. Expanding the M3DISEEN dataset could be achieved by conducting further experiments in-house, however, this approach is time-consuming. Alternatively, a potentially more efficient strategy would be to data mine FDM formulations from published studies. This strategy would also present the opportunity to gather data generated by other research groups, thus minimising potential bias. In addition, more information could be extracted from the literature e.g., drug dissolution results from formulations.

As more intricate 3D designs are fabricated via FDM 3D printing, it may become more difficult to gauge the drug release profile *a priori*. Thus, the ideal prediction model should include this feature. Dissolution testing is a fundamental analysis in formulation development, used to conclude the suitability of a drug product and for further development. As a product is

formulated, it is important to ensure that the drug release occurs in an appropriate manner. The dissolution process may be time-consuming, particularly if the experiments are conducted over weeks or months, which cannot be avoided. Due to its necessity, researchers have investigated modelling techniques to predict dissolution behaviour, particularly for controlled release systems [72, 73]. A mathematical description of the release profile is rather difficult, given the numerous factors that will need to be considered. This is particularly true for FDM, since it affords researchers the ability to produce different and intricate designs [48]. ML on the other hand can utilise existing data, which is made possible by the abundance of dissolution data published, to predict dissolution results of new formulations.

The present study reports the ML pipeline developed, using formulations mined from previously published studies, to predict key HME and FDM 3D printing conditions and drug dissolution properties. The key parameters predicted are extrusion temperature, filament mechanical characteristics, printing temperature and printability. The work especially focussed on the prediction of the drug dissolution performance of the 3D printed formulations and the features that affected dissolution. This study will provide a critical analysis of the performance of ML techniques for the prediction of different parameter of 3D printed formulations from data obtained from the literature and the requirements of the collected data.

2 Materials and methods

2.1 Data mining from literature

PubMed, Google Scholar, and Web of Science were used to search for articles published in English using the terms “hot melt extrusion”, or “fused deposition modelling”, or “fused filament fabrication”, and “drug”, or “tablet”, or “capsule”, or “printlet”, or “drug device”, or “printability” between Jan 1, 2013, and November 30, 2020.

2.2 Data collection

The data collection from the literature were arranged as shown in Table 1.

2.2.1 Identification of the Formulation

The formulations extracted from literature were identified by the article’s DOI, author ID, formulation ID in the manuscript and year of publication.

2.2.2 Composition

The components and their respective weight ratio for each formulation was recorded. Any formulations where the accumulative ratio did not sum to 1 (i.e. 100 w/w%) were removed from the analysis.

Table 1. The variables used within this study

Identification of the formulation	**Article DOI**	DOI_1	DOI_2	…	DOI_n
	Author	Author_1	Author_2		Author_n
	Formulation ID	ID_1	ID_2	…	ID_n
Composition	**Material 1**	0.2	0.5	…	…
	Material 2	0.3	0	…	…
	…	…	…	…	…
	Material 410	0.1	0.1	…	…
Hot Melt Extrusion	**Extruder (brand type)**	HAAKE_MiniCTW	Noztek_Pro	…	…
	Extrusion Speed (RPM)	22.5	135	…	…
	Extrusion temperature (ºC)	145	169	…	…
	Extrusion torque (N.cm)	15	15	…	…
	Filament aspect	Good	Good	…	…
3D printing	**Printer (brand type)**	Makerbot_Replicator_2X	Makerbot_Replicator_2X	…	…
	Nozzle diameter (mm)	0.4	0.4	…	…
	Printing Speed (mm/s)	90	10	…	…
	Printing temperature (ºC)	210	200	…	…
	Platform temperature (ºC)	30	80	…	…
	Printability	Yes	Yes	…	…
3D printed formulation	**Object**	Tablet	Film	…	…
	Shape	Cylinder	Square	…	…
	Type of shell	1	1	…	…
	Length (mm)	10	20	…	…
	Width, Diameter (mm)	10	20	…	…
	Depth, Thickness (mm)	3.2	0.2	…	…
	Volume (mm3)	258.97	80	…	…
	Surface area (mm2)	257.61	816	…	…
	Surface area/volume	0.995	10.2	…	…
	Weight (mg)	181.02	112.8	…	…
	Layer thickness (mm)	0.2	0.05	…	…
	Shell (top/bottom) (mm)	0.2	0.4	…	…
	Shell (lateral) (mm)	0.2	0.4	…	…
	Infill (%)	0	60	…	…
	Infill type	Rectilinear	Hexagonal	…	…
	3D printed product aspect	Good	Good		
Dissolution test	**Dissolution T20 (min)**	20	y	…	…
	Dissolution T50 (min)	80	y	…	…
	Dissolution T80 (min)	230	y	…	…
	pH of the dissolution media (pH)	Acid	Mixed	…	…
	Volume of dissolution media (ml)	900	50	…	…
	Dissolution apparatus	USP_II	bottle	…	…
	Dissolution speed (RPM)	50	50	…	…
Drug solubility	**Drug Solubility (mg/L)**	0.1	0.007		

*"y" was used to represent information that could not be found

2.2.3 Hot Melt Extrusion

The HME process parameters recorded were extruder type, extrusion speed, extrusion temperature (as per the temperature reported in the respective manuscripts; this may refer to the nozzle temperature or maximum barrel temperature), extrusion torque, and filament mechanical characteristics (good, brittle or flexible).

2.2.4 3D Printing

The FDM printing process parameters recorded were printer brand and type (e.g. direct drive), nozzle diameter, printing speed, printing temperature, platform temperature, and if the formulation was printable or not.

2.2.5 3D Printed Formulations

This part included the information about the object printed, shape of the object, dimensions of the object (Length x Width x Height), weight, layer thickness, the type of shell, thickness of the shell, and percentage infill. The printed products were classed by a feature called ‘object’ that refers to the type of delivery system, either a tablet, film, device or other. Since 3D printing can produce complex shapes, a feature called ‘shape’ was created to detail the shape of the delivery system. This feature helped to elaborate whether a film was cylindrical or square; or whether a tablet was a cylinder or in the shape of a unique structure, such as a radiator [74]. Examples of objects and shape can be found in Figure 1.

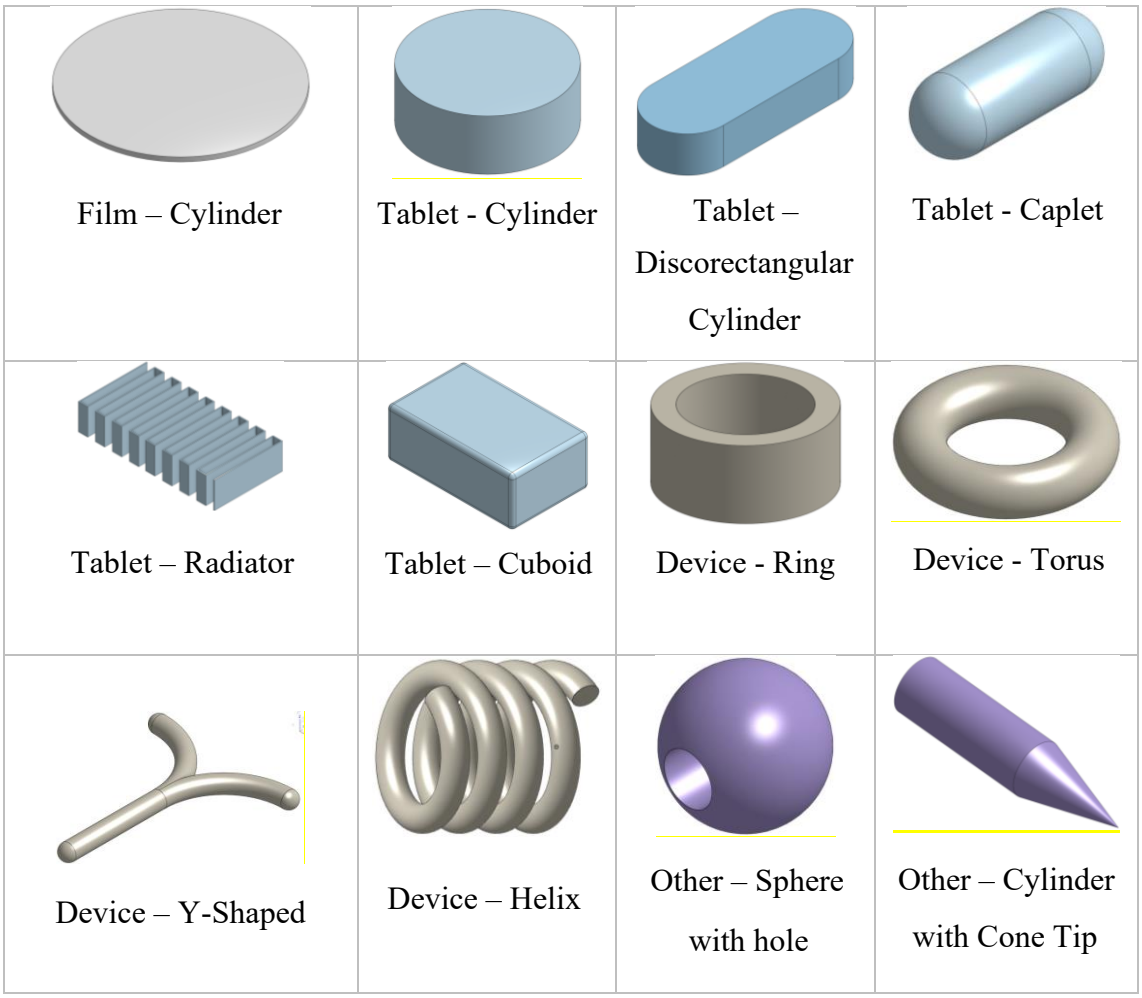

Figure 1. Examples of some 3D designs of objects and shapes found in the literature (object – shape)

Any 3D printed object consists of an external structure called *shell* that provides the shape to the object, and the internal structure called *infill* (Figure 2). The information about the percentage of infill of the 3D printed object was also recorded. The information related to the type of shell were represented through 3 options: "0" - no shell, "1" represented an object with lateral or top/bottom shell, and "2" represented an object with lateral and top/bottom shells. Cylindrical objects that were printed with 100% infill were consistently regarded as having both lateral and top/bottom shells, i.e. shell type 2. The formulations that contain multiple drugs or structures with different composition for the shell and the infill (e.g. 3D printed enteric coating) were not taken into account for the prediction of the dissolution profiles.

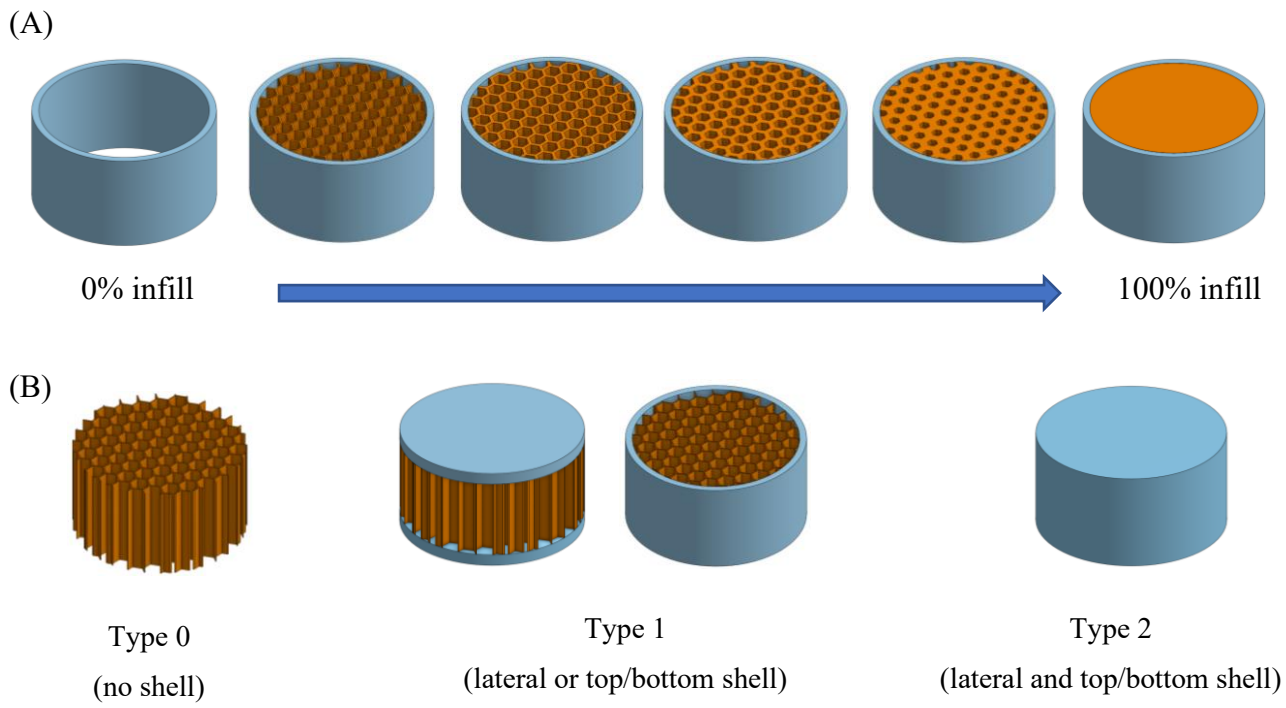

Figure 2. Schematic representation of (A) cylinder with different infill percentage (from 0% left to 100% right) and of (B) different shell type "0" represented "with no shell", "1" represented "with lateral or top/bottom shell", and "2" represented "with lateral and top/bottom shells". The composition of the shell and the infill is the same in all the analysed formulation, the different colour is for visualization purposes.

Shell thickness was extracted from the information from the articles or calculated by multiplying the thickness of the FDM extrudate by the number of shells for the lateral shell thickness; and multiplying the layer height by the number of shells for either the top or bottom shell thickness.

The volume and surface area were calculated using the dimensions of the object, as reported in the respective articles, and basic geometric formulas. However, for objects with complicated structures, image processing techniques in MATLAB (version R2020a, MathWorks, USA) were used to estimate their volume and surface area. Briefly, the images were first binarized according to their colour, which allowed the image of the drug product to be separated from the background. By calculating the area of the segmented image, it was possible to determine the surface area, volume and surface area to volume.

2.2.6 Drug Solubility

Drug solubility values in water were obtained from the relevant supplier datasheets or from reported literature. The parameter called weighted drug solubility was calculated using the drug solubility of the drug multiplied by the percentage of drug in each formulation.

2.2.7 Dissolution Test

The dissolution profiles reported in previous studies varied in scale, whereby different studies measured the drug release to different time points. Instead, the time taken to reach 20% (T20), 50% (T50) and 80% (T80) drug release were recorded to ensure a consistent and complete feature was created. As most articles reported results from drug release studies in the form of graphs, an online software named Digitizer (version 4.3, Ankit Rohatgi, USA) was used to determine the time at the relevant percentage drug release. Each dissolution figure was uploaded to the software, which was able to determine the time points by defining the axes. For sustained release formulations where the dissolution test did not reach a specific percentage the time was omitted from the dataset. Other dissolution features included; volume and pH of the dissolution media, type of dissolution apparatus and its speed. The pH of the dissolution media was recorded in the dataset as "acid" for tests conducted in stomach pH-simulating media (taken as media less than pH 4.5) and "basic" intestinal pH-simulating media (taken as media more than pH 4.5). The rationale for choosing pH 4.5 as the threshold between the two types of media is based on gastric pH typically ranging from 1.5 to 4.5. The dissolution studies performed partially in acid media and then in basic media were recorded as "mixed" pH.

2.2.8 General considerations

Information fields that were relevant but were not reported in the article were represented using "y". Examples of such information include extrusion torque if the filament was extrudable, and dissolution time if the 3D object was printable but not evaluated in dissolution tests. The notation "x" was used to represent information when downstream processes were not applicable, e.g. printing speed and temperature were marked "x" when the filament was not extrudable.

2.3 Predicted target variables

The key parameters that the study aimed to predict were the extrusion temperature, filament mechanical characteristics, printing temperature, printability, and T20, T50 and T80 (Table 2). These are referred to as *targeted variables*.

Table 2. Summary of the predicted targeted variables

Targeted variables	Values	Analysis Type
Extrusion temperature	HME temperature (ºC)	Regression
Filament mechanical characteristics	Unextrudable, Flexible, Good or Brittle	Multi-classification
Printing temperature	Printing temperature (ºC)	Regression
Printability	Yes or No	Binary Classification
Dissolution time (T20, T50 and T80)	Time (min)	Regression

Regression analyses were used to predict the HME temperature, FDM printing temperature and dissolution time, since these target variables were continuous numerical values. Classification analyses were performed to predict the filament mechanical characteristics and printability [71], since these target variables are categorical. The labels used for filament mechanical behaviour were either 'Good', 'Brittle', 'Flexible' or 'Unextrudable' based on the comments found in the reported studies. The definition of 'Good', 'Flexible', 'Brittle' and 'Unextrudable' can be found in a previous publication [71]. Printability was classified as either 'Yes' or 'No' to indicate whether the filament was printable via FDM, given the selected printing parameters. The drug release results reported in the studies varied in scale because different studies measured the drug release at different time points. For dissolution prediction, the time in minutes taken to reach 20% (T20), 50% (T50) and 80% (T80) drug release were recorded to ensure the feature was consistent.

2.4 Feature set selection and creation

Five feature sets used herein were *material, material name, material type, physical properties and physical properties per material type*. The feature sets were created similarly to those previously reported [71]. Briefly, material refers to the individual excipient or drug, respective of supplier, and uses the weight fraction of the material as input. Material name is the same as material, but materials from different suppliers were grouped together (Figure 3). The feature set material type groups materials by their chemical structure, whereas physical properties uses the weighted glass transition temperature, melting temperature and molecular weight as inputs. The final feature set is a combination of physical properties and material type, where the materials are grouped by their chemical structures and the input is the weighted physical properties. Schematics illustrating the creation of the feature sets are presented in Figure 3.

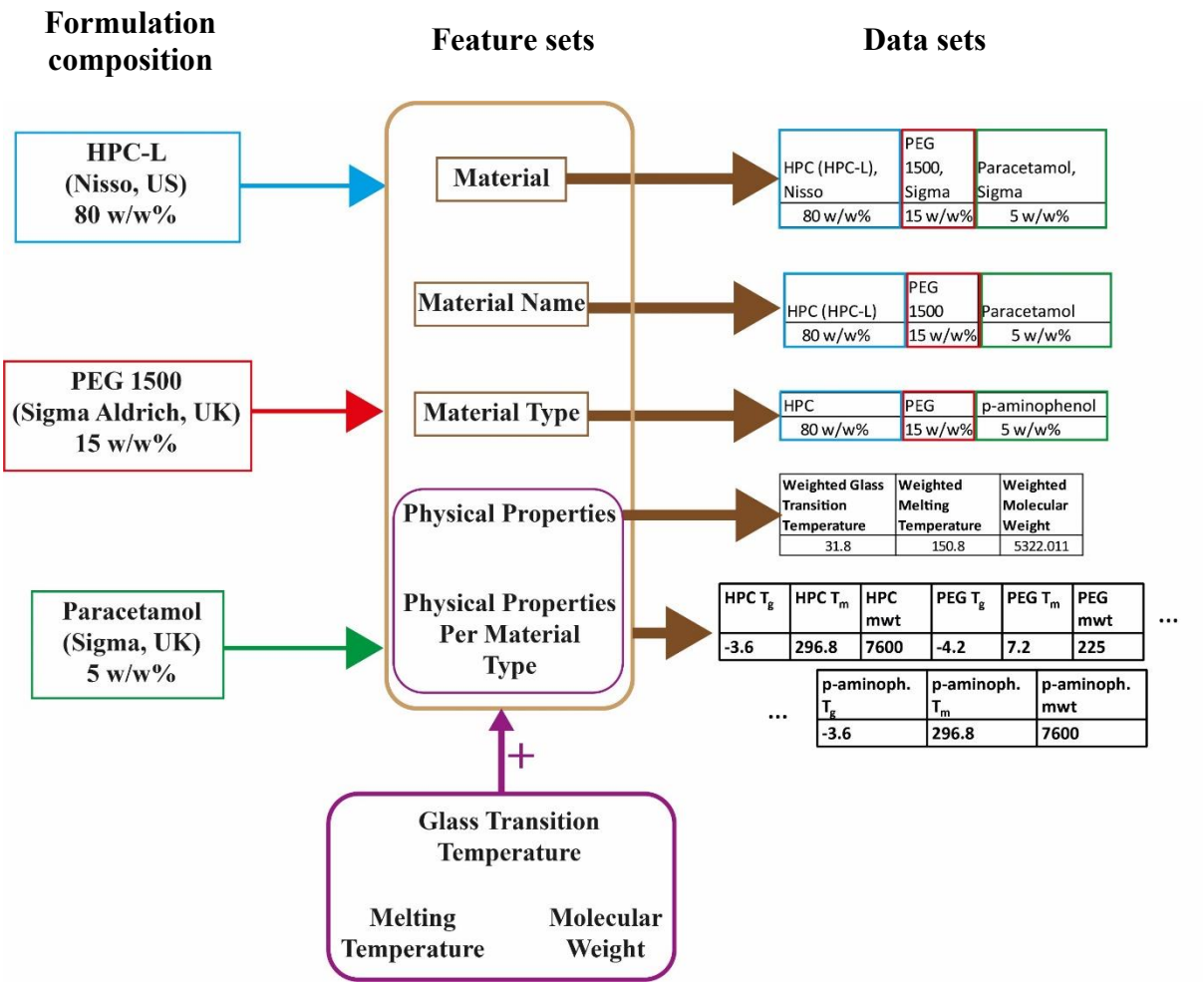

Figure 3. Schematic illustrating how materials from the formulations were classified in the different feature sets: material, material name, material type, physical properties and physical properties per material type.

2.5 Data analysis - Machine learning (ML) techniques

A standard PC (running on Operative system: Debian 5.4.19-1 x86_64) was used for the data analysis and the development of the algorithms described below (Processor: Intel® Xeon® CPU E5620 (2.40 GHz), RAM Memory: 32 GB).

Five different ML techniques were used in this study for classification tasks, which were support vector machines (SVM), random forests (RF), artificial neural networks (ANN), K-nearest neighbors (KNN) and logistic regression (LR). Different ML techniques were used since each ML technique has its own learning characteristics. Three different ML techniques were used for regression task, which were SVM, RF and ANN. Multi-linear regression and KNN were unable to result in meaningful predictions, and hence the results are not included in this study for regression analyses. Brief explanations of each ML technique can be found in a

previous study [71]. The ML techniques were developed using python 3.7 (Python Software Foundation), using the Scikit-Learn package (scikit-learn package, v0.21.3). A 75:25 split was used for training and testing the ML techniques.

For developing models to predict the dissolution time the original five feature sets (Figure 3) were used, however additional features were taken into account (Table 1, sections 3D printed formulation, Dissolution test, Drug solubility). These features (e.g. surface area, weight, infill, pH of the media) were included since they could affect the drug dissolution results and could be considered dissolution-related data.

Predicting the dissolution profile was more demanding than, for example, predicting printability or printing temperature. This was because not every literature mined 3D printed formulation contained dissolution data, and hence the results had to be discarded prior to performing ML. Additionally some articles may report some features (e.g. weight of the formulation) but not others (e.g. infill or shell thickness), whereas ML techniques need to be fed with complete dataset, without missing values. The more data fed into the ML algorithms the greater their performance would be, but due to the missing values in some features, feeding the algorithms with all the dissolution related features would reduce the number of rows (formulations). For example, if weight, shape, pH and dissolution speed were included and then any row containing any null values were removed, which resulted in a 351 formulations dataset; if infill, weight and dissolution speed were selected, then this resulted in 336 formulations. Generally, it was observed that including more features resulted in a higher percentage of missing data, and hence the smaller the size of the data set and the number of formulations included (Figure 4). To avoid this situation, different combinations of input features were tested and compared in terms of the ML algorithms prediction performance.

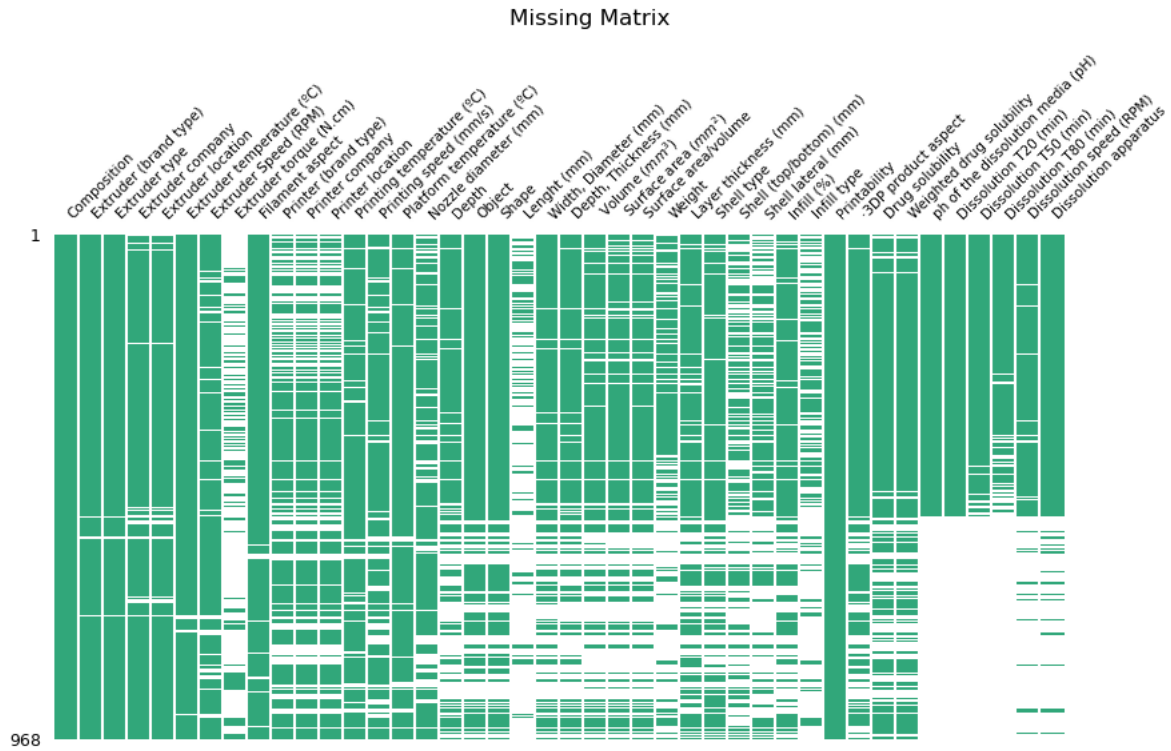

Figure 4. Diagram representing the dataset, used to illustrate the missingness of the data for each of the 968 formulations. Green indicates information was available in the literature, whereas white areas indicates the data was missing.

In this study each possible combination of the 12 features that can affect drug dissolution were computed (shape, type of shell, surface area/volume, weight (mg), infill (%), infill type, pH of the dissolution media (pH), volume of dissolution media (ml), dissolution apparatus, dissolution speed (RPM), drug solubility (mg/L), weighted solubility). This led to a to 2 to the power of 12 ($2^{12} = 4096$) combinations of features that were merged with the 5 feature sets that take in to account the composition of the formulations (Figure 3). We disregarded those datasets that lost more than the 40% of the original formulations and used the rest for training a ML model for each algorithm (RF, SVM and ANN). This led us to consider a total of (2^{12}) × 5 × 3 different ML experiments. Additionally, each experiment was tested in 50-fold random-split cross validation to avoid the negative impact of outliers (Figure S1). The dissolution data is spread on a considerably large scale (e.g. T20 could be either 5 min or 2000 min), where the effect of randomly splitting the data into training and testing had a pronounced effect on the results and an undesirable impact in the metrics. The ML pipeline for predicting the dissolution times is detailed and illustrated in the supplementary document (Figure S1). Categorical values (e.g. print shape) were label encoded, and numerical values (e.g. surface area, dissolution time)

with large ranges were quantile transformed. Label encoding is one means of vectorising categorical data. Using shape features as an example, cylinder, caplets and capsules were represented as 0, 1 and 2, respectively.

2.6 Data evaluation

Different metrics were used for scoring the accuracy of the ML techniques, as no single metric conveys a complete picture of a model's performance. A brief explanation of each metric can be found in our previous study [71]. For classification analyses, five classification metrics were used; *accuracy*, Cohen's *kappa*, *precision*, *recall*, and *F1*. For the processing temperature and dissolution time predictions, two regression metrics were used: the *mean absolute error* (MAE), and the *coefficient of determination* (R^2).

An additional metric that we called RADOC (Real Area Difference Of Curves) was developed for predicting the dissolution times. The metric is used to compare two "curves", in a two-dimensional space, formed by the two series of points (the experimental and the predicted points) respectively connected by straight lines. RADOC computes the area corresponding to the absolute difference between those two curves (Figure S2 (A)). The smaller this difference area, the more similar the shape of the two curves will be, leading to a more fine-grained measure of the dissolution dynamics. That difference area is then relativized against the area under the real curve (Figure S2 (B) and (C)) (leading to a [0%, ∞%] error range), which helped us to also address the scale problem.

3 Results and Discussion

3.1 Exploratory data analysis

A total of 968 formulations were literature mined from 114 articles, and only formulations incorporating drugs were added to the database. Information relating to the starting materials, HME process, 3DP and drug dissolution was obtained, which were identified as having a potential effect on the fabrication workflow and drug release profile. Figure 4 illustrates the distribution of the data collected. During the data collection stage, it was clear that there was a lack of data in some of the selected parameters, which could be a potential problem for the machine learning (ML) algorithms. It is worth mentioning that only 57.02% of FDM articles reported the drug dissolution profile of their printed product.

In total, 411 excipients and drugs were recorded from 121 different suppliers. Grouping similar materials together, irrespective of supplier, resulted in a total of 254 materials, presented as packed bubble diagrams in Figure 5, where it is evident that a large number of excipients had been used. Figure 5 (B) presents the materials when grouped by similar chemical structure. From both analyses, it appears that materials were used evenly, displaying equal distribution. The most widely used excipient type was acrylics, which was used slightly more used than HPMC and PVA. Similarly, the most used drug was theophylline, which was marginally more used than paracetamol.

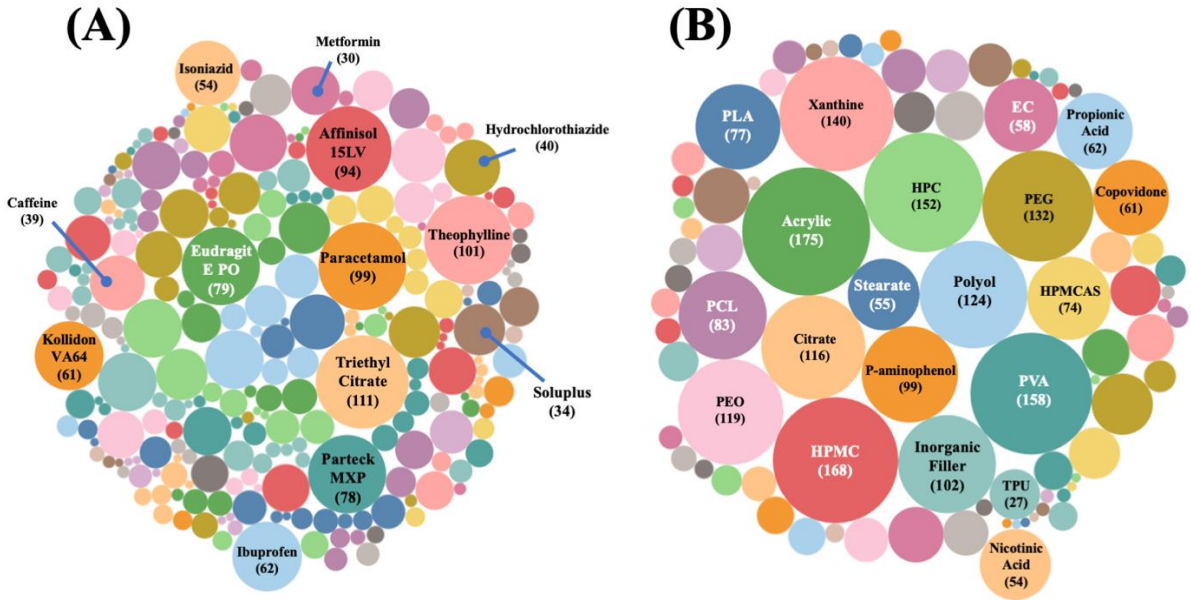

Figure 5. Packed bubble diagrams to illustrate the distribution of (A) individual materials used and (B) material types.

Four different physical properties pertaining to each material were recorded in the present study. The glass transition temperatures (T_g) of the individual materials ranged from -107.65 to 1201.85ºC, with the majority possessing a T_g below 200 ºC (Figure 6 (A)). The melting temperatures (T_m) of the materials ranged from -76 ºC to 1,974 ºC, with the majority of materials possessing T_m values below 400 ºC (Figure 6 (B)). The small number of outliers with high T_m and T_g values correspond to inorganic fillers, such as titanium dioxide and barium sulphate. The molecular weight of materials ranged from 58.4 to 7,000,000 g/mol (Figure 6(C)). Drug solubility is also a determinant of the dissolution behaviour, and the value for each formulation was recorded, ranging from 0.0004 to 2,450 mg/L (Figure 6 (D)).

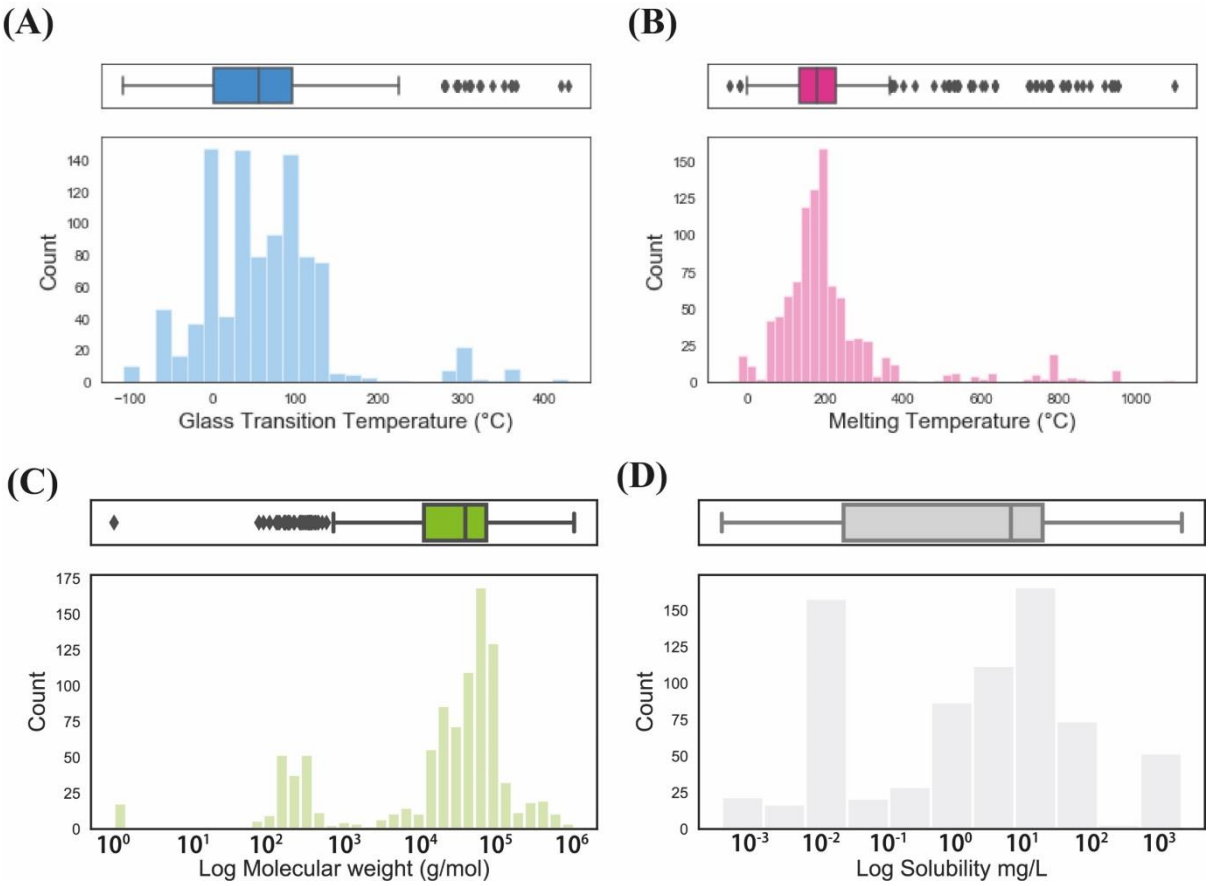

Figure 6. Box plot-histogram depicting the distribution of (A) glass transition temperature, (B) melting temperature, (C) molecular weight and (D) drug solubility of the formulation.

Exploratory data analysis of the outcome of HME revealed that 84.6% of the filaments reported in the literature were identified as 'Good' with respect to filament characteristics (Figure 7). These values are likely to be positively skewed, due to bias reporting wherein researchers are incentivised to only publish positive results. As illustrated by the Sankey diagram in Figure 7, the majority of 'Good' filaments were printable. Conversely, filaments exhibiting either 'Flexible' or 'Brittle' characteristics were found to mainly yield unprintable formulations. Nevertheless, the majority of the 968 formulations reported in the literature were printable (85.74%), which highlight again that most of the articles only report positive results.

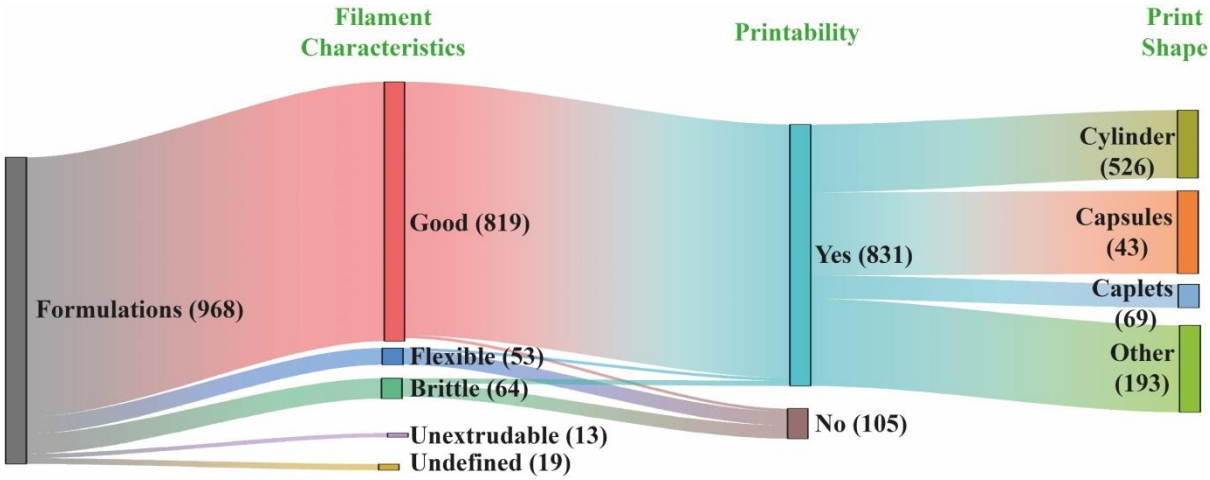

Figure 7. Sankey diagram depicting the flow of literature-mined formulations across three different features.

The extrusion temperatures used in HME ranged from 22 to 210 ºC, with a mean of 132 ºC (Figure 8 (A)). Twenty-four extruder brands were used to prepare filaments, with the Thermo Scientific Process 11 filament extruder and the HAAKE MiniCTW found to be the most used. Extrusion speeds ranged from 5 to 200 rpm. Values of torque during extrusion were reported in some articles but, due to low levels of reporting, this feature was not further analysed. The printing temperatures used in FDM 3DP ranged from 53 to 240 ºC, with a mean of 174 ºC (Figure 8 (B)). As evidenced by the box-plot, there are a notably larger number of outliers in the printing temperature compared to the HME temperatures. Outliers due to incorrect information can negatively impact modelling performance since the ML techniques will be making predictions based on incorrect relationships. However, these outliers, although statistically determined as outliers by the box-plot, were in fact correct values. These outliers reflect that, despite being a relatively high-temperature fabrication process (> 100 °C), a small number of studies have investigated whether certain formulations can be printed at lower temperature. Keeping the outliers in the dataset provides the potential to develop a modelling technique for low-temperature FDM processing, which will benefit researchers investigating thermally labile drugs.

The platform temperature is also an important feature because it can affect the adherence of the formulations to the build plate while printing. These values ranged from 16 to 115 ºC, with a mean of 41 ºC, although in 47% studies the temperature was not controlled, and hence the value was room temperature. A total of thirty different types of printer brands were used in the studies, with Makerbot Replicator 2X and Prusa i3 3D desktop printer being

the most commonly used, and with nozzle diameters ranging from 0.2 to 0.5 mm (mode 0.4 mm). Values of Printing Speed ranged from 0.5 to 500 mm/s, with a mode of 90 mm/s.

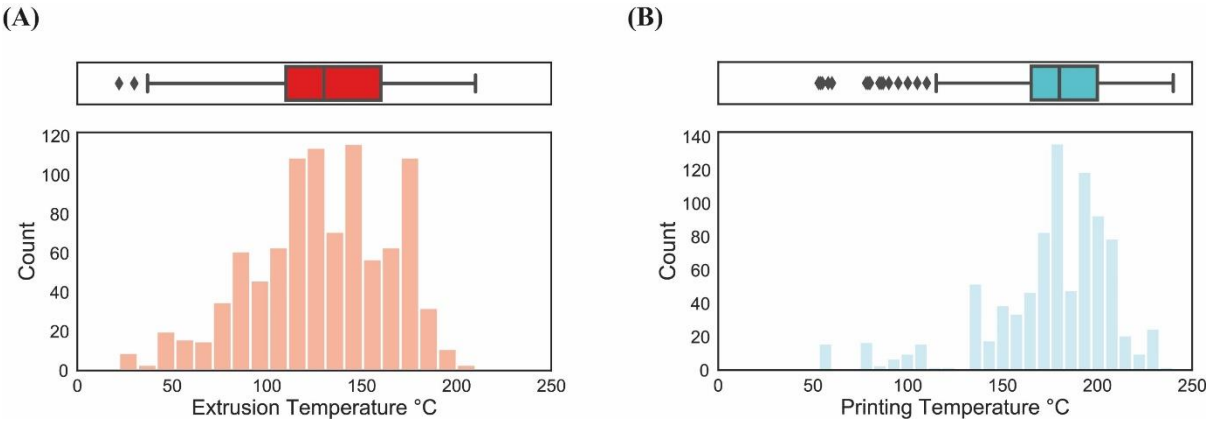

Figure 8. Box plot-histogram plots depicting the distribution of (A) extrusion and (B) printing temperatures recorded in the dataset.

Regarding the 3D printed objects, FDM 3DP can be used to fabricate a range of items, however the majority of objects printed were oral formulations that were encoded as “tablets”, with a comparatively smaller proportion of “films” and “devices” printed (Figure 9 (A)). Although 3DP can print complex geometries, most of the literature has focused on developing cylinders, capsules and caplets (Figure 9 (B)). Overall, a total of 38 different shapes were recorded, with the most common shape printed being a cylinder (48.03%), followed by caplets (6.98%) and elliptical cylinder (4.65%).

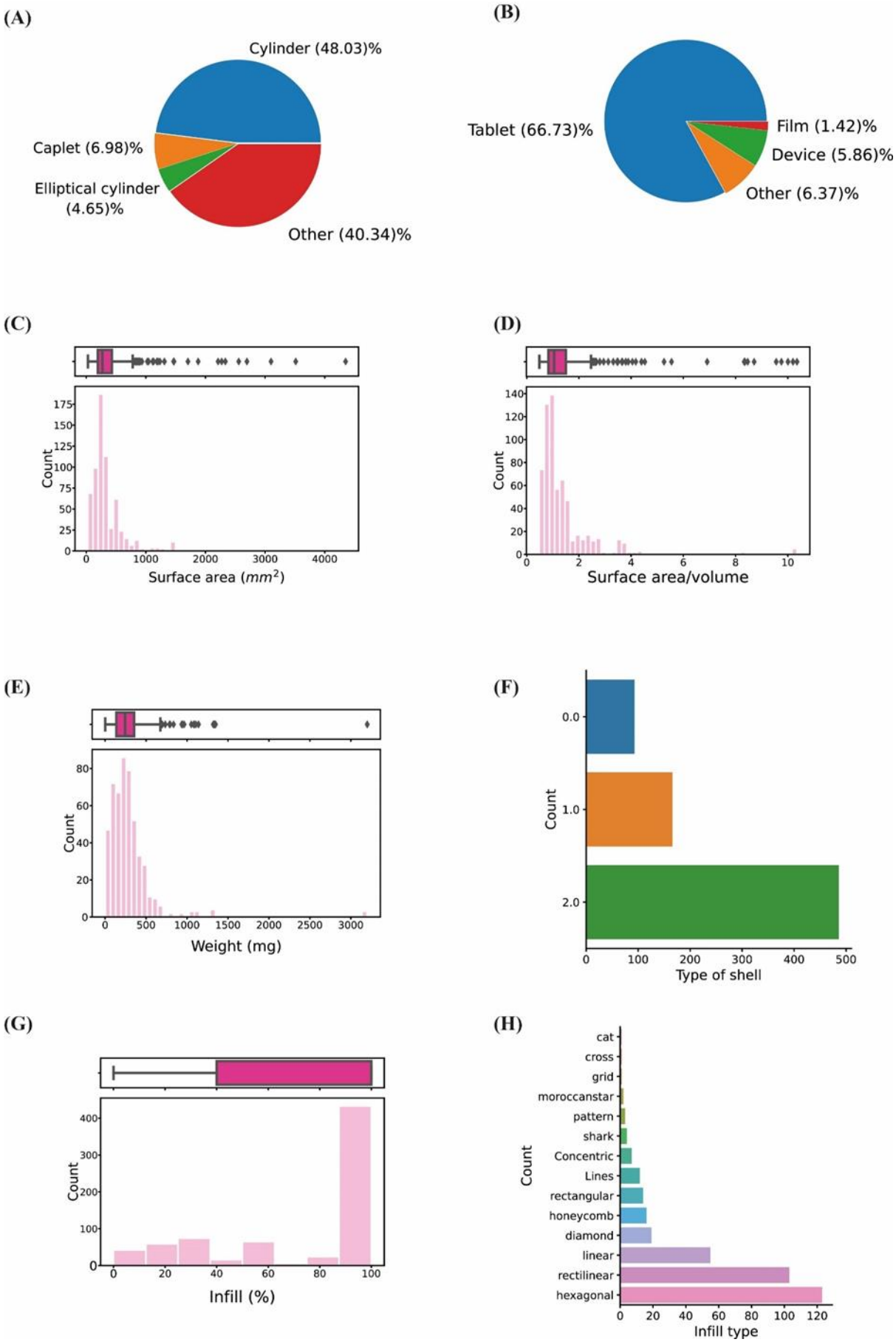
(A)
Cylinder (48.03)%
Caplet (6.98)%
Elliptical cylinder (4.65)%
Other (40.34)%
(B)
Tablet (66.73)%
Film (1.42)%
Device (5.86)%
Other (6.37)%
(C)
Count
Surface area (mm^2)
(D)
Count
Surface area/volume
(E)
Count
Weight (mg)
(F)
Count
0.0
1.0
2.0
Type of shell
(G)
Count
Infill (%)
(H)
Count
cat
cross
grid
moroccanstar
pattern
shark
Concentric
Lines
rectangular
honeycomb
diamond
linear
rectilinear
hexagonal
Infill type

Figure 9. Pie charts, box plot-histograms and bar charts illustrating the proportion of (A) objects and (B) shapes printed, (C) surface area, (D) surface area to volume ratio, (E) weight, (F) type of shell, (G) infill and (H) infill type.

Other physical characteristics of the 3D printed objects that could be relevant due to their potential effect on the drug release from the formulation were collected and analysed (Figure 9). The dimension of the objects (length, width, diameter, depth) were collected and were used to derive features like volume (ranged from 10.6 mm^3 to 1658.8 mm^3, with a mean of 332.8 mm^3), surface area (ranged from 26.6 to 4350.4 mm^2, with a mean of 384.8 mm^2), and surface area to volume ratio (ranged from 0.5 to 10.4, with a mean of 1.5) (Figure 9).

The weight of the printed object ranged from 30 to 3200 mg, with a mean of 308.5 mg and the layer thickness ranged from 0.05 to 0.5 mm, with a mean of 0.18 mm. Most of these objects (65.2 %) were printed with including lateral and top/bottom shells (Figure 9). Only 12.5 % of the objects did not include any external shell. The thickness of top/bottom shells ranged from 0.05 to 2.4 mm with a mean of 0.4 mm, and thickness of the lateral shells ranged from 0.1 to 2.4 mm, with a mean of 0.7 mm. A wide range of infill percentages were used (from 0 to 100 %) with a mode of 100 %. Fourteen types of infills were used in the mined studies, with rectilinear and hexagonal infills being the most used. Due to the missing data, the feature infill type was not used for further analysis.

Data mining the literature allowed the extraction of the dissolution behaviour of 3D printed formulations. The results revealed that 48.04% of the printable formulations were analysed for their drug releasing characteristics. The distribution of times taken for the formulation to reach 20%, 50% and 80% drug release are presented in Figure 10. The times spanned several orders of magnitude, ranging from 0.4 min to 46,123 min (32 days). This reflects the ability of FDM to be applied in a range of drug delivery systems capable of both immediate and extended-drug release. However, the data is positively skewed, highlighting that the majority of studies focused on release in the order of hours. Skewed data is known to negatively impact ML techniques, and hence the data will need to be transformed prior to modelling. Skewed data will result in ML techniques being trained on a disproportionately higher number of shorter dissolution times, and will be less likely to accurately predict times for larger dissolution times. Addressing this issue usually involves collecting more data to balance the distribution, which is not feasible since all the published results have already been collected. Alternatively, the majority class can be minimised to balance the distribution, but

this will come at the expense of a smaller dataset. Hence, in this instance, it is better to transform the data. The log transformed data highlights that when the data is transformed it results in a near-normally distributed data across several orders of magnitude (Figure 10 (B)).

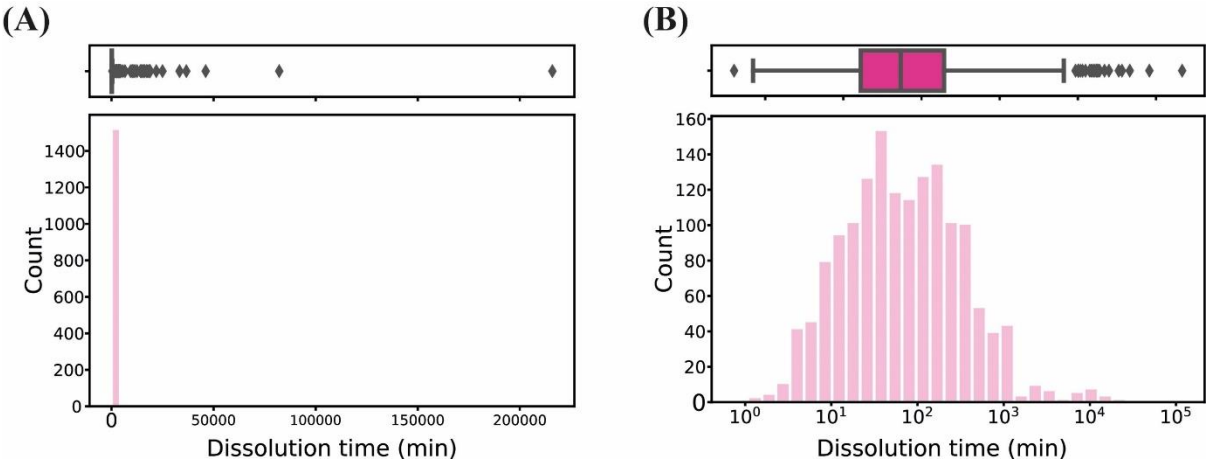

Figure 10. Histogram and boxplot depicting (A) the distribution of time taken to reach 20%, 50% and 80% drug release and (B) the log transformed data. The log transformation clearly illustrates the distribution of dissolution times were recorded across several orders of magnitude.

The values of other dissolution test parameters that could affect the drug dissolution rate were also collected and analysed. 45.2% of the formulations were tested in simulating intestinal pH condition using a "basic" dissolution media (pH media higher than pH 4.5), 36.5% of tests were conducted in stomach pH-simulating conditions (pH media lower than pH 4.5) and some studies (14.3%) evaluated the formulations first in acid and then in basic pH media, simulating the transit through the GI tract (Figure S3). Some studies (3.9%), especially for formulations made with materials that are pH dependent, e.g. enteric polymers, evaluated the drug release of the same formulations using acid and basic pH media. The volume of dissolution media ranged from 1 to 1000 mL, with a mode of 900 mL. The main type of dissolution apparatus used in those studies was USP type II, and the dissolution speeds ranged from 10 to 200 rpm, with a mode of 50 rpm (Figure S3).

3.2 Predictability evaluation

3.2.1 Predicting Filament Mechanical Characteristics

ML techniques were used to predict the filament characteristics using the literature dataset. ANN obtained the highest accuracy of 91%, with the feature set Material Name (Figure 11 (A)). Similarly, this feature obtained the highest *kappa* value of 0.49.

For imbalanced datasets, using the accuracy as a metric to compare different datasets can be misleading, particularly if one dataset has a greater imbalance. For example, the literature-mined dataset contained 84.6% labelled as 'Good' for printability. If as prediction criterion, one blindly assigned all formulations as 'Good', then one would trivially obtain an accuracy of 84.6%. This high accuracy value may incorrectly seem a good result while, in reality, the trivial ML "algorithm" would not be learning any patterns as it would just be predicting the majority class for all formulations. Thus, despite the simplicity for calculating the accuracy, it is more informative to use a metric that factors in a baseline value, such as the *kappa* value. The *kappa* value factors in the probability of a chance agreement (i.e. random guessing), and measures the predictive performance of an ML technique compared to random guessing. *Kappa* values can be negative, indicating the ML technique performed worse than random guessing; 0, indicating a performance comparable to random guessing; or a positive value, indicating the performance was better than random guessing. From the results presented in Figure 11, it can be concluded that ML techniques are able to perform better than random guessing. There were some exception, primarily with using the Physical Properties feature set as input, where the *kappa* value was 0 for ANN, SVM and LR. Nevertheless, from a practical sense, and using the Material name feature set, ML will provide researchers with an enhanced accuracy in predicting the filament characteristics compared to random guessing. The precision and recall metrics are equally informative for 3DP researchers from a practical perspective. These metrics reveal how well a model is able to predict the positive class ('Good', in the current study).

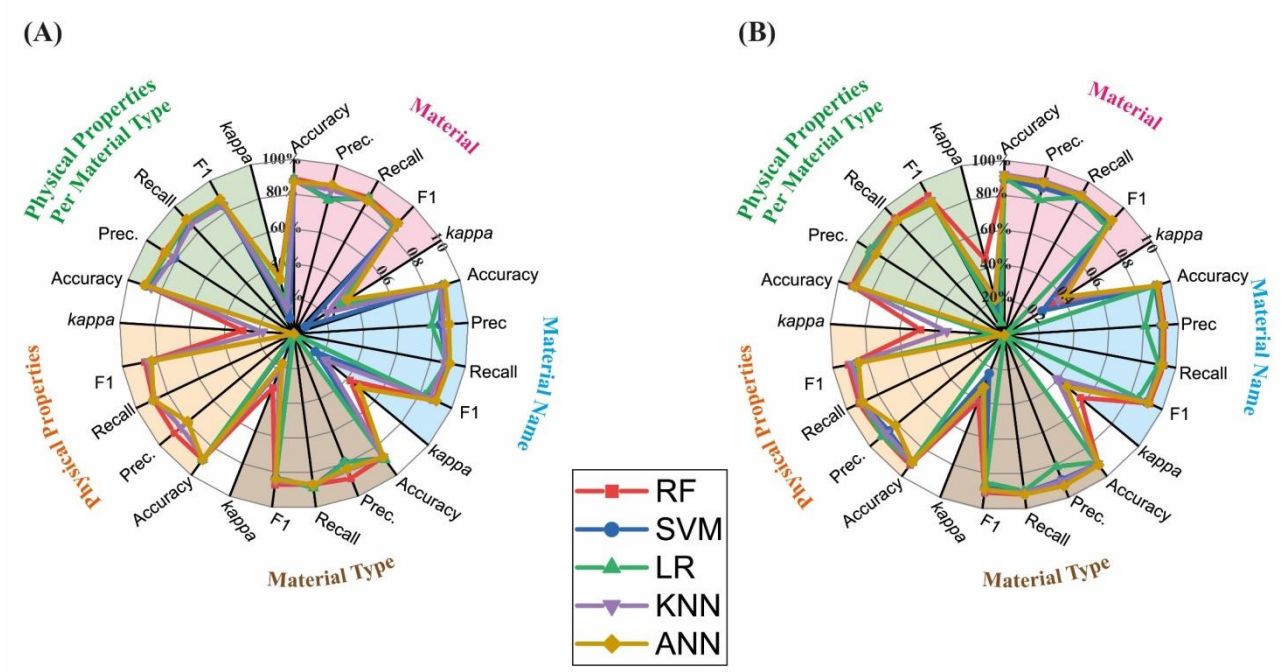

Figure 11. Radar plot with the metrics result for the (A) filament mechanical characteristics and (B) printability. RF - random forests, SVM - support vector machines, LR - logistic regression, KNN - K-nearest neighbors, ANN - artificial neural networks. Please see Table S1 & S2 for the specific values.

3.2.2 Predicting printability

The printability metrics for the literature are presented in Figure 11 (B). The feature set Material was found to produce the highest metrics, which were obtained using RF. The accuracy and kappa values were 93% and 0.56, respectively. The positive label was set to 'Yes' for precision and recall, since there is more interest in knowing if a filament will be printable. The precision and recall values were 82% and 83%, respectively. In a practical sense, the recall value suggests that for every ten formulations, there will be 1.7 formulations that are printable but incorrectly predicted as unprintable by RF.

As previously mentioned, overall, the classification analyses revealed that the Material features set produced the highest metrics. This feature set possessed the largest number of features, a total of 411, and hence provided comparatively the most comprehensive information pertaining to the materials. Equally, the Physical Properties feature set comprised of only three features, which could explain why the lowest predictive accuracies were obtained with it. It should also be noted that more effective models could be developed if the dataset was more

balanced. However, the imbalance reflects the current state of academic publishing, which is to publish mainly the positive results.

3.2.3 Predicting extrusion temperature

The extrusion temperature is a parameter difficult to anticipate, especially without prior knowledge. The values are continuous, ranging from 20 to 220 °C, and thus a regression task was performed to predict the individual temperature values for each formulation. The metrics used were the coefficient of determination (R^2) and the mean absolute error (MAE). R^2 measures the variance in the data between the actual temperature and the predicted temperature, with a perfect prediction resulted in an R^2 of 1.00. For more practical usage, the MAE measures the absolute errors between the actual and predicted temperatures. The lower the error the more accurate the prediction, with a perfect prediction producing an MAE of 0 °C. MAE is more practical because a value, e.g. of 5 °C indicates that on average, the predicted temperature will deviate by ± 5 °C.

The optimal MAE and R^2 were achieved with ANN; 5.18 °C and 0.90, respectively, again using the Material feature set (Figure 12 (A)). These results were an improvement over previous work, that used a smaller dataset [71], wherein the MAE and R^2 were 10.8 °C and 0.56, respectively. This was despite the present work possessing a wider temperature range, where a larger error would have been expected to account for the wider range. The increase in R^2 clearly highlights the significant improvement in the predictive performance of the present study, suggesting that collecting data from the literature could be a suitable approach for predictions, and is even better than generating the data in house.

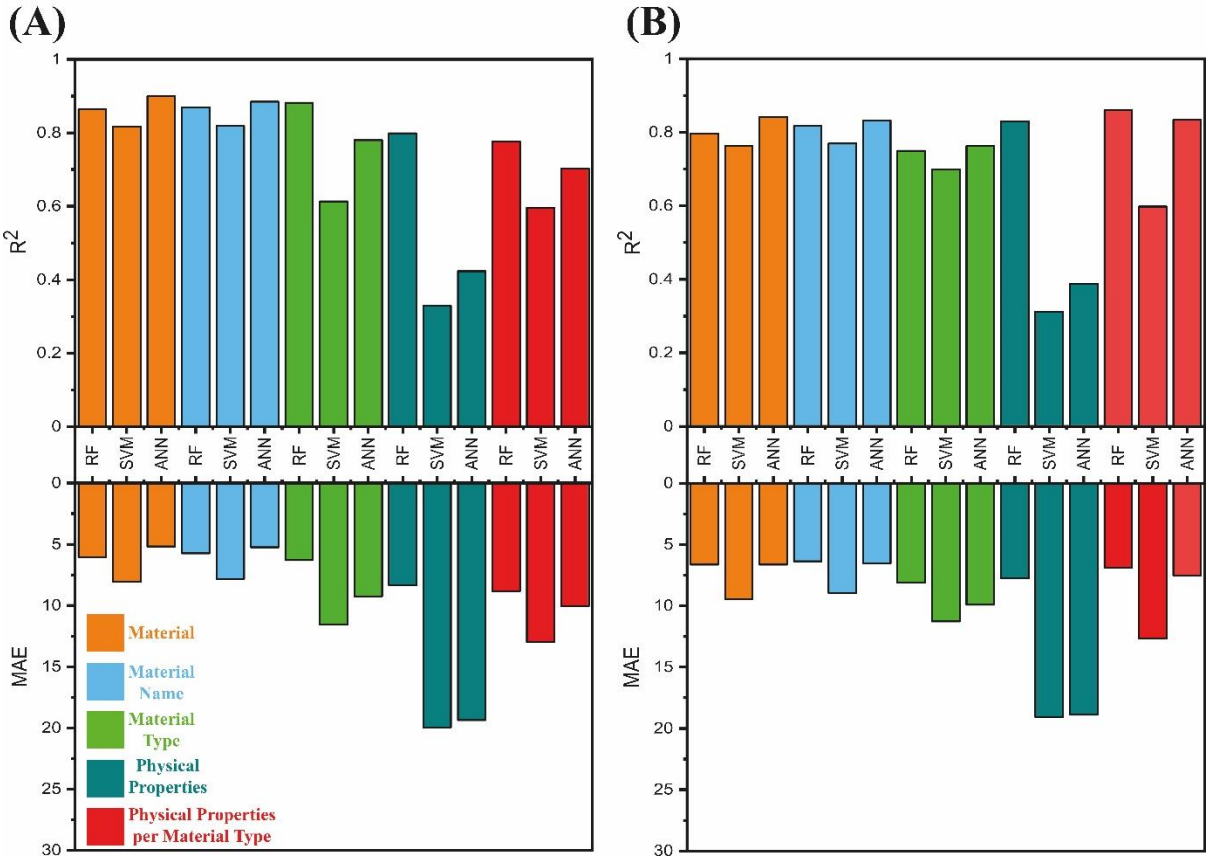

Figure 12. The R^2 and MAE for the (A) extrusion and (B) printing temperatures for the different ML algorithms. RF - random forests, SVM - support vector machines, ANN - artificial neural networks

3.2.4 Predicting printing temperature

The printing temperature is an important variable that affects the printability of a formulation but predicting its value is a time-consuming approach without prior knowledge. Similar to HME, the incorrect temperature can result in nozzle blockage if the temperature is too low, or blockage caused by degradation of the polymer and the drug if the temperature is too high. To date, there is no *rule-of-thumb* or an established model for pre-determining the printing temperature, other than the assumption that the printing temperature should be higher than the extrusion temperature in the HME. The optimal MAE and R^2 were obtained by RF, which were 6.87 °C and 0.86, respectively, using the Physical Properties per Material Type feature set (Figure 12 (B)). The MAE and the R^2 values were better than the values in the previous study (8.3 ºC and 0.83, respectively) [71], where all the data was obtained using the same FDM 3D printer brand and generated in-house. These new results were remarkable, indicating that

printing temperature data obtained from the literature, published by many different research groups using many different FDM printer models, were comparable or even better at predicting printing temperature. Nevertheless, the MAE infers that using the literature-mined data can yield an accuracy of ± 6.87 °C, which is a narrow range considering that the printing temperatures attempted to date vary from 40 to 260 ºC.

3.2.5 Predicting Dissolution Behaviour

The drug dissolution behaviour of the formulations is affected by more than just the material components of the delivery system. The drug dissolution is influenced by design parameters of the formulation, such as weight and surface area-to-volume ratio [8, 48], drug solubility [75]; and the dissolution conditions, such as media pH and volume. The physical characteristics of the 3D printed object, the conditions of the dissolution test and the solubility of the drug were therefore used as inputs for each one of the feature configurations. Hence, developing a predictive model requires additional inputs to those used for modelling printability. The complete list of input variables that could affect drug dissolution profiles are detailed in Table 1.

The analysis began by incorporating the new added features and finding the best configuration of features to obtain the highest predictive performance. The best configurations were selected based on a new metric used herein, which is referred to as RADOC, due to the shortcomings of the other metrics. The pragmatism of MAE is useful since the units for this metric are the same as the data under analysis. The MAE is a scale-dependent metric that requires the data, including during the training-test partition, to be on the same scale. However, this was not the case for predicting the dissolution time, where some partitioning exhibited longer dissolution times. Due to the scale difference between T20, T50 and T80, relative metrics such as R^2 or the mean absolute percentage error (MAPE) are more suitable for this task. However, although a high score in those metrics would normally mean the evolution of both profiles is also similar, this is not the case when having only three points (T20, T50 and T80). To address this problem, when selecting the best model, the RADOC metric was used. RADOC is both scale-free and capable of capturing the evolution of the graphs, and hence is suitable for predicting the dissolution times (Figure S2). RADOC compares the relative difference between the area under the curve for both the actual and predicted curves, where the smaller the value the smaller the deviation between the two curves. This helped to determine which configuration provided the best predictive performance. The training-test split

partitioning was performed 50 times using different random splits. This was due to the incompleteness of data, whereby certain formulations would be missing values for particular features (Figure 4). As a result, the same random split could not be achieved for each configuration, which made it difficult to determine the true optimal configuration. Performing the analysis 50 times with varying random splits provided a more holistic determination of the optimal configuration. Again, the RADOC metric proved to be useful when comparing the optimal configuration due to the variability in random splitting.

The features that were the most occurring in the best 100 analyses, in terms of producing the lowest RADOC value, are presented in Figure 13. The main features used in the best analyses were, in descending order, Surface area-to-volume ratio, pH, infill, shape, weighted drug solubility, shell type, drug solubility and weight. The mean RADOC for the best 100 analyses was 48.01 and a standard deviation of 12.37.

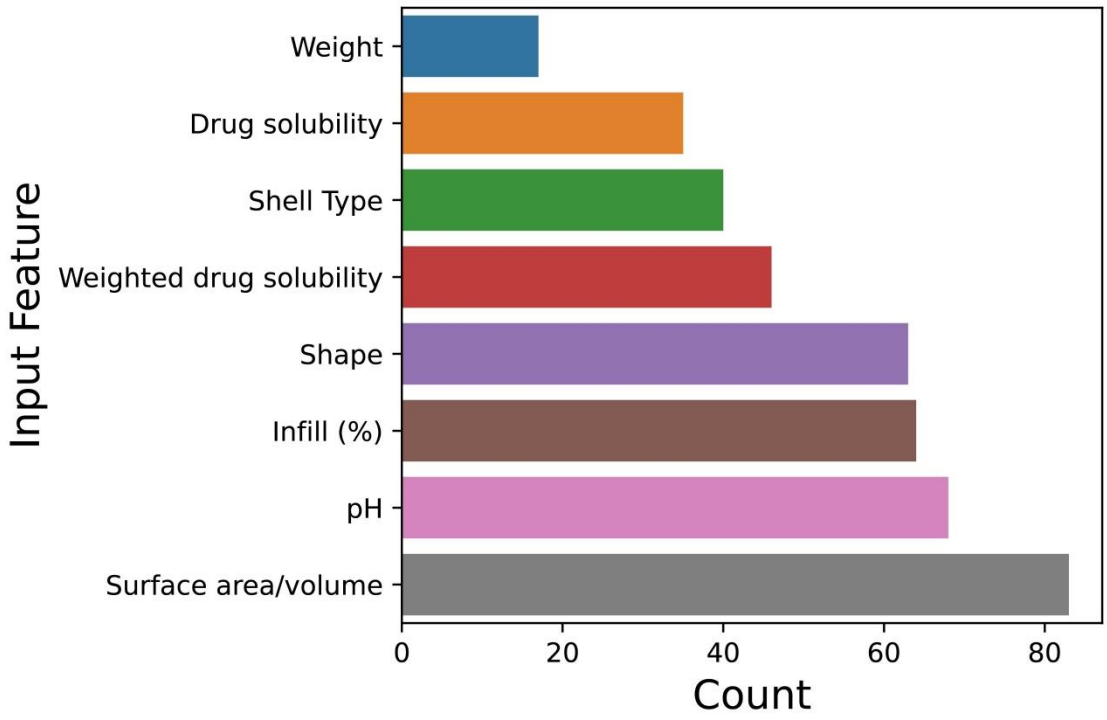

Figure 13. Histogram depicting the feature importance. The count number indicates the number of times a feature was used in the best 100 analysis.

The feature surface area-to-volume ratio was identified as the most important feature and was used in more than 80 of the best predictions. The feature was already identified as a relevant parameter to control dissolution of 3D printed formulations in one of the first studies

in 2015 [48]. This feature is also related to the shape of the 3D printed object that was also identified as a relevant feature, used in more than 60 on the best 100 predictions.

The pH of the media is the second most relevant parameter that needs to be controlled when performing the dissolution test. The pH is not a characteristic of the 3DP formulation but the dissolution media. The pH is included in more than 65 of the best 100 predictions. It is important because some materials used to prepare 3D printer medicines show different properties or solubility in different pH. The best example of this is the enteric polymers that do not dissolve at pH acid (lower than 4.5) but disintegrate/dissolve when the pH is close to 5. Dissolution studies performed in acidic media are typically for immediate release formulations, so the selection of the pH of the media is partially linked to the type of formulations that are evaluated in the dissolution test too.

The infill percentage of the formulations is the third most important feature and was also identified as a relevant in previous studies [76, 77]. Higher infill percentage is associated with longer dissolution times. Other important features are solubility and weighted solubility of the drug used in 45 and 35 of the 100 best predictions, respectively. Higher solubility of the drug leads to faster dissolution. The shell type is a feature that affect the dissolution and it is related to the surface area-to-volume ratio feature; formulations without external shells tend to release the drug faster due to easier penetration of dissolution media to the inner part of the formulations. Moreover, the weight of the formulations also affects the dissolution process, and in some cases higher weight leads to longer dissolution times.

The incorporation of the additional feature inputs resulted in a good predictive performance. The results from the 50-fold random split, for each feature set and algorithms are presented in Figure 14. It was evident that the selected random split and configuration can affect the predictive performance of the MLTs. For example, if the test split contained higher dissolution times, then this was found to increase the error rate. The best prediction was obtained by an ANN algorithm that used the material feature set combined with the surface area-to-volume ratio, volume dissolution media, weighted solubility shape and pH of the media as additional input features. Although each of the inputs gathered in (Table 1) were considered important variables by the authors of this study prior to the ML analyses, they were not all used by the ML algorithms. The ANN algorithm achieved an MAE of 24.29 minutes and a R^2 of 0.86 in the test set, which means that on average it is able to predict the dissolution times (T20, T50 and T80) of a formulation with an error of ± 24.29 minutes. This is remarkable considering that some of the dissolution tests run for days.

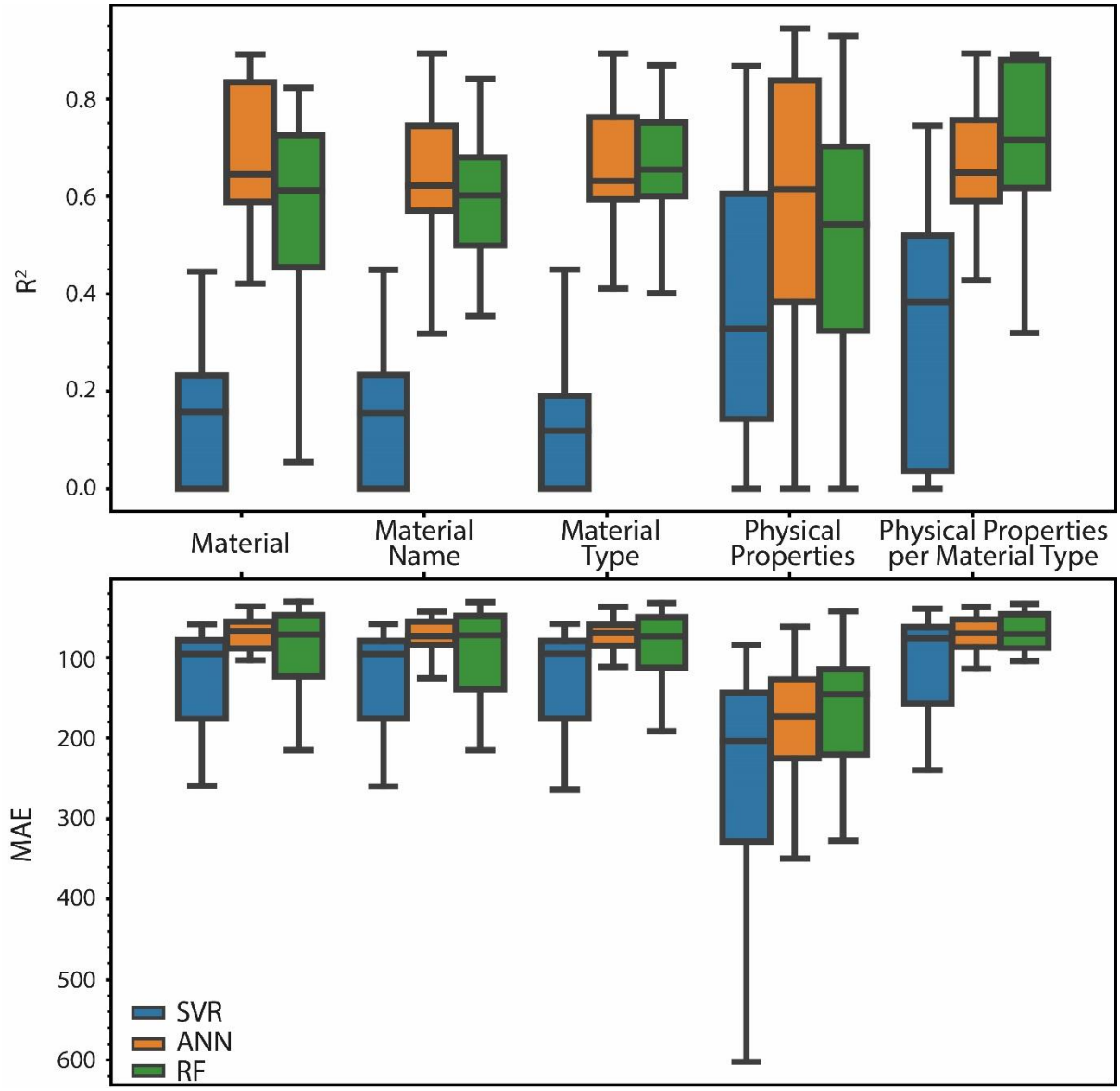

Figure 14. R^2 and mean absolute error results of the 50-fold random split for each of the MLTs, and across the different feature sets for predicting drug dissolution profiles. The results demonstrate that the random split can affect the results of the MLTs, due to the wide range in dissolution times. RF - random forests, SVM - support vector machines, ANN - artificial neural networks.

Figure 15 illustrate the prediction vs actual results from the best performing model. The MAE is an average of the absolute errors and thus influenced by large errors which, as expected, were obtained from sustained release data. This was evidenced when examining both the scatter plot and residual plot (Figure 15(A & B)). The residual plot (Figure 15 (B)) revealed a common trend, whereby an increase in residuals is observed as the actual dissolution time increases, with the exception of a few anomalies. Figure 15 (C-E) presents examples of three

different release studies, illustrating that ML techniques were able to produce accurate simulations of the released drug, thereby confirming the models suitability for both immediate and sustained release. Figure 15 (C-E) also demonstrated that the ML techniques were able to learn the trajectory of the dissolution profile insofar as learning that the concentration of drug release increases over time. A benefit of ML is that multiple predictions can be made from the same data point (i.e. formulation). This was leveraged in the present study by investigating whether the three time points could be predicted simultaneously, rather than developing separate models for each time point, which is a faster approach to model development. This feature was not coded into the ML techniques, and hence all three ML techniques were able to independently learn the graph trajectory.

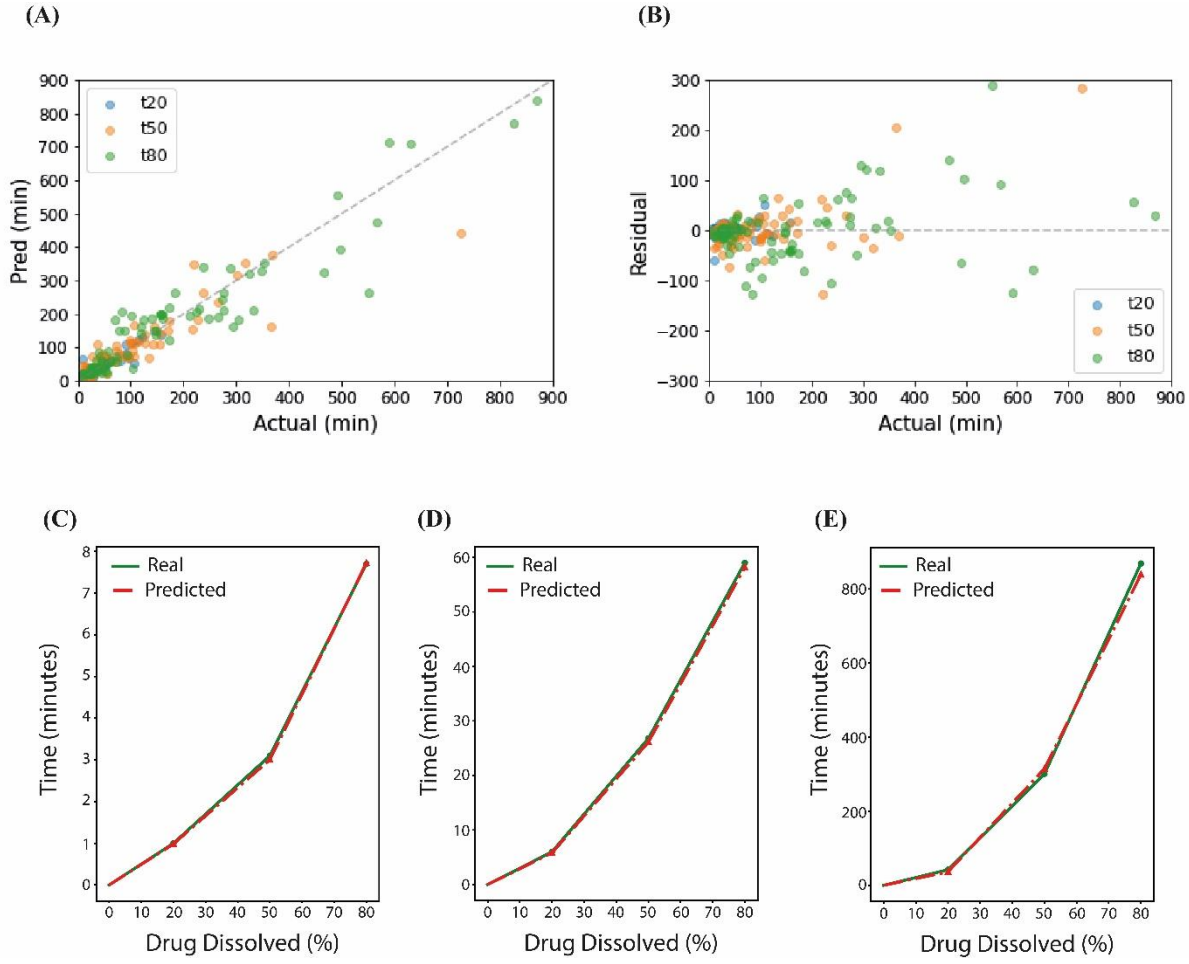

Figure 15. (A) Scatter plot illustrating the actual vs. predicted scatter plots, and (B) the corresponding residual plot of the best performing ML technique. (C-E) Are three representative actual vs predicted dissolution profiles, across three different time scales (8, 60 and 850 min).

The predictive performance of the ML strategy applied herein were considered satisfactory. Considering that dissolution studies are performed from days to weeks, an MAE in the order of minutes will indeed prove to be an asset to researchers. Previous work using ML to predict the dissolution profile of 3D printed products has demonstrated that high accuracies can be attained using ML [78, 79]. However, a current limitation of the previous work for predicting dissolution profiles was that the formulations were developed in-house and limited to one drug. In contrast, the model developed herein offers prediction to a larger material pool. Moreover, the information was gathered from different researchers, making it less susceptible to bias and thus providing greater generalisability for making new predictions.

3.2.6 General consideration

This study integrated data from articles published by researchers all over the world, with different materials, methodologies and objectives, which produced ML models that were successfully able to generalize for predicting the targeted variables (extrusion temperature, filament mechanical characteristics, printing temperature, printability and drug dissolution performance). Even though the same MLTs were used as in the previous study, higher predictive performances were obtained in this study, particularly with the HME and FDM temperatures [71]. This was expected as the current study consisted of more formulations. It is also worth acknowledging that in the previous study it took six years to achieve an in-house dataset of 614 formulations, whereas in the same time period 968 formulations were published – an increase of 58% in data – highlighting the fast data generation nature of literature mining. While the data used in the previous study was very straightforward to use, it was somewhat limited, since the data was obtained from the same laboratory and using the same equipment, work methodology and objectives.

Although the findings of the present study provided additional benefits to the previous study in modelling key aspects of the 3DP workflow [71], the integration of the literature-mined data presented several challenges. One salient disadvantage is that the data is not structured and hence it is not machine-learning compatible, requiring an exhaustive and time-consuming pre-processing step to collect and structure the data. For example, for unifying dissolution time in different scales (immediate release, long-term release, etc), the authors had to collect the data as “time to reach a certain percentage of release” rather than “percentage of drug released after a certain time”.

The literature data is biased towards positive results which may have reduced the learning performance of the ML techniques in predicting printability. Most researchers only publish the good results in their studies. Even though there are some unsuccessful formulations in the articles, the information is limited. As a result, most information about the filament aspect and printability is positive, which causes a deficiency of negative examples and this is not ideal for training ML algorithms, as they tend to learn the majority class. In addition, part of the data in this study was estimated by using relevant software. Although estimation is a common data generation technique, it may have contributed and additional error in some of the data, and consequently may have reduced the accuracy of the prediction.

Finally, different articles missed different features when presenting data. For the ML algorithms to work, rows containing null values (i.e. missing) must be removed from both training and test sets, which is known to negatively impact the accuracy of ML algorithms due to fewer learning instances. In addition, removing these null values forced additional pre-processing workload to the ML pipeline. If the literature data was more complete then a simpler pre-processing methodology could have been used, and potentially better results could be achieved for drug dissolution prediction. To assist in developing more effective ML models, the authors of this study encourage other authors in the field to publish complete data including both positive and negative results. All the articles should provide the sufficient information even if the data may not be relevant for the specific aim of the study. Ideally, standards on which data and how it should be reported would avoid some of the problems encountered in this study regarding missing information. The minimum parameters that we consider should be published are included in Table 1, although additional data could be useful for future studies. The features selected herein are known determinants of the target variables. The research in 3DP of pharmaceuticals remains nascent, and as the research develops more information will come to light. This could potentially lead to an improved feature selection, enabling ML techniques to attain a higher accuracy.

Current ML algorithms have the potential to overcome some of the challenges that the field of 3DP of pharmaceuticals faces, including the optimization of the fabrication parameters, reducing the inefficient empirical trial approach, and the requirements of expert knowledge. The performance of the AI tools is expected to drastically improve in the following years, however, one of the main needs of these algorithms to exploit its full potential is Big Data, which means having data with several orders of magnitude of cardinality bigger than the data set used for this study. While in other fields ML is applied to massive amounts of automatically

generated historical data, the application of ML to 3DP of medicines is based on experimental data. This data requires big investment in time and resources as well as human intervention to be generated and reviewed. The optimal amount of data will only be achieved via an open sharing and collaboration-based program. Even if one institution or company were capable of reaching a good amount of data alone, data from different sources would be preferable since it would produce less biased or unbalanced datasets, which subsequently will be more appropriate for training ML models.

Considering the future trajectory of 3DP medicines, the ultimate goal will be to digitally simulate the entire 3DP workflow in an effort to move towards sustainable research, where both costs and material waste are minimised, as well as the time needed to realise the research hypothesis. In essence, the ML models developed could expedite developments in the field of 3DP pharmaceuticals. In addition, digital simulations can offer insight that otherwise would be difficult to experimentally determine. The present study demonstrates that ML could be an effective component of such digital simulation by offering high predictive performance and in rapid time. Moreover, the low computational demands of ML mean that it can be deployed as a web-based software, or seamlessly integrated into other modelling tools similar to the M3DISEEN web-based service. The aim with ML will be to produce an end-to-end model that can simulate the entire 3DP workflow. 3DP and ML (and other AI tools) offer a unique opportunity to move the pharmaceutical development to the next level, and this will indeed depend on the availability of data and the quality thereof.

4 Conclusion

The study investigated the use of literature-mined data for developing artificial intelligence (AI) machine learning (ML) techniques models to predict key aspects of the 3D printing formulation pipeline. The analysis of the literature mined data revealed that positive results are overwhelmingly published, which consequently resulted in an imbalanced dataset for filament aspect and printability. Nevertheless, the ML techniques explored herein were able to learn and provide high predictive accuracies for the values of the filament hot melt extrusion processing temperature, filament aspect, printing temperature and printability. ML algorithms using data based on the composition of the formulations and additional input features that could influence drug release (e.g. surface area/volume, weight, infill percentage, pH and volume of dissolution media, drug solubility) were used to predict the drug release profile of FDM printed formulations. The best prediction was obtained by an ANN algorithm, which was able to

predict the dissolution times (T20, T50 and T80) of a formulation with an error of ±24.29 minutes. Thus, it was concluded that data mined from the literature was an efficient approach to modelling 3D printing workflow. It was also concluded that a structured repository for 3DP data will greatly facilitate the creation of new knowledge via ML.

References:

[1] I. Seoane-Viaño, S.J. Trenfield, A.W. Basit, A. Goyanes, Translating 3D printed pharmaceuticals: From hype to real-world clinical applications, Advanced Drug Delivery Reviews, 174 (2021) 553-575.

[2] S.H. Lim, H. Kathuria, M.H.B. Amir, X. Zhang, H.T.T. Duong, P.C.-L. Ho, L. Kang, High resolution photopolymer for 3D printing of personalised microneedle for transdermal delivery of anti-wrinkle small peptide, Journal of Controlled Release, 329 (2021) 907-918.

[3] A.J. Capel, R.P. Rimington, M.P. Lewis, S.D.R. Christie, 3D printing for chemical, pharmaceutical and biological applications, Nature Reviews Chemistry, 2 (2018) 422-436.

[4] A. Awad, F. Fina, A. Goyanes, S. Gaisford, A.W. Basit, Advances in powder bed fusion 3D printing in drug delivery and healthcare, Advanced Drug Delivery Reviews, 174 (2021) 406-424.

[5] D. Zhi, T. Yang, T. Zhang, M. Yang, S. Zhang, R.F. Donnelly, Microneedles for gene and drug delivery in skin cancer therapy, Journal of Controlled Release, 335 (2021) 158-177.

[6] X. Xu, A. Awad, P. Robles-Martinez, S. Gaisford, A. Goyanes, A.W. Basit, Vat photopolymerization 3D printing for advanced drug delivery and medical device applications, Journal of Controlled Release, 329 (2021) 743-757.

[7] J. Aho, J.P. Bøtker, N. Genina, M. Edinger, L. Arnfast, J. Rantanen, Roadmap to 3D-Printed Oral Pharmaceutical Dosage Forms: Feedstock Filament Properties and Characterization for Fused Deposition Modeling, Journal of Pharmaceutical Sciences, 108 (2019) 26-35.

[8] S.K. Patel, M. Khoder, M. Peak, M.A. Alhnan, Controlling drug release with additive manufacturing-based solutions, Advanced Drug Delivery Reviews, 174 (2021) 369-386.

[9] S.J. Trenfield, A. Awad, A. Goyanes, S. Gaisford, A.W. Basit, 3D Printing Pharmaceuticals: Drug Development to Frontline Care, Trends Pharmacol. Sci., 39 (2018) 440-451.

[10] A. Melocchi, M. Uboldi, A. Maroni, A. Foppoli, L. Palugan, L. Zema, A. Gazzaniga, 3D printing by fused deposition modeling of single- and multi-compartment hollow systems for oral delivery – A review, International Journal of Pharmaceutics, 579 (2020) 119155.

[11] B.C. Pereira, A. Isreb, M. Isreb, R.T. Forbes, E.F. Oga, M.A. Alhnan, Additive Manufacturing of a Point-of-Care "Polypill:" Fabrication of Concept Capsules of Complex Geometry with Bespoke Release against Cardiovascular Disease, Advanced Healthcare Materials, 9 (2020) 2000236.

[12] A. Melocchi, F. Parietti, G. Loreti, A. Maroni, A. Gazzaniga, L. Zema, 3D printing by fused deposition modeling (FDM) of a swellable/erodible capsular device for oral pulsatile release of drugs, Journal of Drug Delivery Science and Technology, 30 (2015) 360-367.

[13] A. Maroni, A. Melocchi, F. Parietti, A. Foppoli, L. Zema, A. Gazzaniga, 3D printed multi-compartment capsular devices for two-pulse oral drug delivery, J. Control. Release., 268 (2017) 10-18.

[14] C.J. Bloomquist, M.B. Mecham, M.D. Paradzinsky, R. Janusziewicz, S.B. Warner, J.C. Luft, S.J. Mecham, A.Z. Wang, J.M. DeSimone, Controlling release from 3D printed medical devices using CLIP and drug-loaded liquid resins, Journal of Controlled Release, 278 (2018) 9-23.

[15] D.K. Tan, M. Maniruzzaman, A. Nokhodchi, Advanced pharmaceutical applications of hot-melt extrusion coupled with fused deposition modelling (FDM) 3D printing for personalised drug delivery, Pharmaceutics, 10 (2018) 203.

[16] I. Xenikakis, M. Tzimtzimis, K. Tsongas, D. Andreadis, E. Demiri, D. Tzetzis, D.G. Fatouros, Fabrication and finite element analysis of stereolithographic 3D printed microneedles for transdermal delivery of model dyes across human skin in vitro, European Journal of Pharmaceutical Sciences, 137 (2019) 104976.

[17] J. Skowyra, K. Pietrzak, M.A. Alhnan, Fabrication of extended-release patient-tailored prednisolone tablets via fused deposition modelling (FDM) 3D printing, European Journal of Pharmaceutical Sciences, 68 (2015) 11-17.

[18] M. Sadia, B. Arafat, W. Ahmed, R.T. Forbes, M.A. Alhnan, Channelled tablets: An innovative approach to accelerating drug release from 3D printed tablets, Journal of Controlled Release, 269 (2018) 355-363.

[19] F. Fina, A. Goyanes, M. Rowland, S. Gaisford, A.W. Basit, 3D printing of tunable zero-order release printlets, Polymers, 12 (2020) 1769.

[20] A. Goyanes, A. Fernández-Ferreiro, A. Majeed, N. Gomez-Lado, A. Awad, A. Luaces-Rodríguez, S. Gaisford, P. Aguiar, A.W. Basit, PET/CT imaging of 3D printed devices in the gastrointestinal tract of rodents, International Journal of Pharmaceutics, 536 (2018) 158-164.

[21] A. Goyanes, C.M. Madla, A. Umerji, G. Duran Piñeiro, J.M. Giraldez Montero, M.J. Lamas Diaz, M. Gonzalez Barcia, F. Taherali, P. Sánchez-Pintos, M.-L. Couce, S. Gaisford, A.W. Basit, Automated therapy preparation of isoleucine formulations using 3D printing for the treatment of MSUD: First single-centre, prospective, crossover study in patients, Int. J. Pharm., 567 (2019) 118497.

[22] S.A. Khaled, J.C. Burley, M.R. Alexander, J. Yang, C.J. Roberts, 3D printing of five-in-one dose combination polypill with defined immediate and sustained release profiles, Journal of Controlled Release, 217 (2015) 308-314.

[23] A. Awad, A. Yao, S.J. Trenfield, A. Goyanes, S. Gaisford, A.W. Basit, 3D printed tablets (printlets) with braille and moon patterns for visually impaired patients, Pharmaceutics, 12 (2020) 172.

[24] K. Vithani, A. Goyanes, V. Jannin, A.W. Basit, S. Gaisford, B.J. Boyd, A Proof of Concept for 3D Printing of Solid Lipid-Based Formulations of Poorly Water-Soluble Drugs to Control Formulation Dispersion Kinetics, Pharmaceutical Research, 36 (2019) 102.

[25] I. Seoane-Viaño, N. Gómez-Lado, H. Lázare-Iglesias, X. García-Otero, J.R. Antúnez-López, Á. Ruibal, J.J. Varela-Correa, P. Aguiar, A.W. Basit, F.J. Otero-Espinar, 3D Printed Tacrolimus Rectal Formulations Ameliorate Colitis in an Experimental Animal Model of Inflammatory Bowel Disease, Biomedicines, 8 (2020) 563.

[26] J.J. Ong, A. Awad, A. Martorana, S. Gaisford, E. Stoyanov, A.W. Basit, A. Goyanes, 3D printed opioid medicines with alcohol-resistant and abuse-deterrent properties, International Journal of Pharmaceutics, 579 (2020) 119169.

[27] M. Fanous, S. Gold, S. Muller, S. Hirsch, J. Ogorka, G. Imanidis, Simplification of fused deposition modeling 3D-printing paradigm: Feasibility of 1-step direct powder printing for immediate release dosage form production, International Journal of Pharmaceutics, 578 (2020) 119124.

[28] A. Awad, F. Fina, A. Goyanes, S. Gaisford, A.W. Basit, 3D printing: Principles and pharmaceutical applications of selective laser sintering, Int. J. Pharm., 586 (2020) 119594.

[29] Y. Yang, Y. Xu, S. Wei, W. Shan, Oral preparations with tunable dissolution behavior based on selective laser sintering technique, International Journal of Pharmaceutics, 593 (2021) 120127.

[30] R. Hamed, E.M. Mohamed, Z. Rahman, M.A. Khan, 3D-printing of lopinavir printlets by selective laser sintering and quantification of crystalline fraction by XRPD-chemometric models, International Journal of Pharmaceutics, 592 (2021) 120059.

[31] A. Awad, F. Fina, S.J. Trenfield, P. Patel, A. Goyanes, S. Gaisford, A.W. Basit, 3D printed pellets (miniprintlets): A novel, multi-drug, controlled release platform technology, Pharmaceutics, 11 (2019) 148.

[32] S.J. Trenfield, H.X. Tan, A. Goyanes, D. Wilsdon, M. Rowland, S. Gaisford, A.W. Basit, Non-destructive dose verification of two drugs within 3D printed polyprintlets, International Journal of Pharmaceutics, 577 (2020) 119066.

[33] X. Xu, A. Awad, P. Robles-Martinez, S. Gaisford, A. Goyanes, A.W. Basit, Vat photopolymerization 3D printing for advanced drug delivery and medical device applications, Journal of Controlled Release, (2020).

[34] M.J. Uddin, N. Scoutaris, S.N. Economidou, C. Giraud, B.Z. Chowdhry, R.F. Donnelly, D. Douroumis, 3D printed microneedles for anticancer therapy of skin tumours, Materials Science and Engineering: C, 107 (2020) 110248.

[35] S.N. Economidou, C.P.P. Pere, A. Reid, M.J. Uddin, J.F.C. Windmill, D.A. Lamprou, D. Douroumis, 3D printed microneedle patches using stereolithography (SLA) for intradermal insulin delivery, Materials Science and Engineering: C, 102 (2019) 743-755.

[36] I. Karakurt, A. Aydoğdu, S. Çıkrıkcı, J. Orozco, L. Lin, Stereolithography (SLA) 3D printing of ascorbic acid loaded hydrogels: A controlled release study, International Journal of Pharmaceutics, 584 (2020) 119428.

[37] H.K. Cader, G.A. Rance, M.R. Alexander, A.D. Gonçalves, C.J. Roberts, C.J. Tuck, R.D. Wildman, Water-based 3D inkjet printing of an oral pharmaceutical dosage form, International Journal of Pharmaceutics, 564 (2019) 359-368.

[38] M. Kyobula, A. Adedeji, M.R. Alexander, E. Saleh, R. Wildman, I. Ashcroft, P.R. Gellert, C.J. Roberts, 3D inkjet printing of tablets exploiting bespoke complex geometries for controlled and tuneable drug release, Journal of Controlled Release, 261 (2017) 207-215.

[39] H. Vakili, R. Kolakovic, N. Genina, M. Marmion, H. Salo, P. Ihalainen, J. Peltonen, N. Sandler, Hyperspectral imaging in quality control of inkjet printed personalised dosage forms, International Journal of Pharmaceutics, 483 (2015) 244-249.

[40] H. Öblom, C. Cornett, J. Bøtker, S. Frokjaer, H. Hansen, T. Rades, J. Rantanen, N. Genina, Data-enriched edible pharmaceuticals (DEEP) of medical cannabis by inkjet printing, International Journal of Pharmaceutics, 589 (2020) 119866.

[41] C.C. Dodoo, P. Stapleton, A.W. Basit, S. Gaisford, The potential of Streptococcus salivarius oral films in the management of dental caries: An inkjet printing approach, International Journal of Pharmaceutics, 591 (2020) 119962.

[42] M. Edinger, D. Bar-Shalom, N. Sandler, J. Rantanen, N. Genina, QR encoded smart oral dosage forms by inkjet printing, International Journal of Pharmaceutics, 536 (2018) 138-145.

[43] S.J. Trenfield, H. Xian Tan, A. Awad, A. Buanz, S. Gaisford, A.W. Basit, A. Goyanes, Track-and-trace: Novel anti-counterfeit measures for 3D printed personalized drug products using smart material inks, International Journal of Pharmaceutics, 567 (2019) 118443.

[44] K. Ilyés, N.K. Kovács, A. Balogh, E. Borbás, B. Farkas, T. Casian, G. Marosi, I. Tomuță, Z.K. Nagy, The applicability of pharmaceutical polymeric blends for the fused deposition modelling (FDM) 3D technique: Material considerations–printability–process modulation, with consecutive effects on in vitro release, stability and degradation, European Journal of Pharmaceutical Sciences, 129 (2019) 110-123.

[45] G. Kollamaram, D.M. Croker, G.M. Walker, A. Goyanes, A.W. Basit, S. Gaisford, Low temperature fused deposition modeling (FDM) 3D printing of thermolabile drugs, International Journal of Pharmaceutics, 545 (2018) 144-152.

[46] J. Boetker, J.J. Water, J. Aho, L. Arnfast, A. Bohr, J. Rantanen, Modifying release characteristics from 3D printed drug-eluting products, European Journal of Pharmaceutical Sciences, 90 (2016) 47-52.

[47] S. Cailleaux, N.M. Sanchez-Ballester, Y.A. Gueche, B. Bataille, I. Soulairol, Fused Deposition Modeling (FDM), the new asset for the production of tailored medicines, Journal of Controlled Release, 330 (2021) 821-841.

[48] A. Goyanes, P. Robles Martinez, A. Buanz, A.W. Basit, S. Gaisford, Effect of geometry on drug release from 3D printed tablets, International Journal of Pharmaceutics, 494 (2015) 657-663.

[49] A. Goyanes, F. Fina, A. Martorana, D. Sedough, S. Gaisford, A.W. Basit, Development of modified release 3D printed tablets (printlets) with pharmaceutical excipients using additive manufacturing, Int J Pharm, 527 (2017) 21-30.

[50] M.A. Luzuriaga, D.R. Berry, J.C. Reagan, R.A. Smaldone, J.J. Gassensmith, Biodegradable 3D printed polymer microneedles for transdermal drug delivery, Lab Chip., 18 (2018) 1223-1230.

[51] S.A. Stewart, J. Dominguez-Robles, V.J. McIlorum, E. Mancuso, D.A. Lamprou, R.F. Donnelly, E. Larraneta, Development of a Biodegradable Subcutaneous Implant for Prolonged Drug Delivery Using 3D Printing, Pharmaceutics, 12 (2020).

[52] N.K. Martin, J. Domínguez-Robles, S.A. Stewart, V.A. Cornelius, Q.K. Anjani, E. Utomo, I. García-Romero, R.F. Donnelly, A. Margariti, D.A. Lamprou, E. Larrañeta, Fused deposition modelling for the development of drug loaded cardiovascular prosthesis, International Journal of Pharmaceutics, 595 (2021) 120243.

[53] Z.-L. Farmer, E. Utomo, J. Domínguez-Robles, C. Mancinelli, E. Mathew, E. Larrañeta, D.A. Lamprou, 3D printed estradiol-eluting urogynecological mesh implants: Influence of material and mesh geometry on their mechanical properties, International Journal of Pharmaceutics, 593 (2021) 120145.

[54] G.K. Eleftheriadis, D.G. Fatouros, Haptic Evaluation of 3D-printed Braille-encoded Intraoral Films, European Journal of Pharmaceutical Sciences, 157 (2021) 105605.

[55] M. Vivero-Lopez, X. Xu, A. Muras, A. Otero, A. Concheiro, S. Gaisford, A.W. Basit, C. Alvarez-Lorenzo, A. Goyanes, Anti-biofilm multi drug-loaded 3D printed hearing aids, Materials Science and Engineering: C, 119 (2021) 111606.

[56] A. Awad, S.J. Trenfield, A. Goyanes, S. Gaisford, A.W. Basit, Reshaping drug development using 3D printing, Drug Discovery Today, 23 (2018) 1547-1555.

[57] A. Melocchi, F. Briatico-Vangosa, M. Uboldi, F. Parietti, M. Turchi, D. von Zeppelin, A. Maroni, L. Zema, A. Gazzaniga, A. Zidan, Quality considerations on the pharmaceutical applications of fused deposition modeling 3D printing, International Journal of Pharmaceutics, 592 (2021) 119901.

[58] M. Elbadawi, L.E. McCoubrey, F.K.H. Gavins, J.J. Ong, A. Goyanes, S. Gaisford, A.W. Basit, Disrupting 3D printing of medicines with machine learning, Trends in Pharmacological Sciences, (2021).

[59] D. Reker, Y. Rybakova, A.R. Kirtane, R. Cao, J.W. Yang, N. Navamajiti, A. Gardner, R.M. Zhang, T. Esfandiary, J. L'Heureux, T. von Erlach, E.M. Smekalova, D. Leboeuf, K. Hess, A. Lopes, J. Rogner, J. Collins, S.M. Tamang, K. Ishida, P. Chamberlain, D. Yun, A. Lytton-Jean, C.K. Soule, J.H. Cheah, A.M. Hayward, R. Langer, G. Traverso, Computationally guided high-throughput design of self-assembling drug nanoparticles, Nature Nanotechnology, 16 (2021) 725-733.

[60] R. Han, H. Xiong, Z. Ye, Y. Yang, T. Huang, Q. Jing, J. Lu, H. Pan, F. Ren, D. Ouyang, Predicting physical stability of solid dispersions by machine learning techniques, Journal of Controlled Release, 311-312 (2019) 16-25.

[61] H. Gao, W. Wang, J. Dong, Z. Ye, D. Ouyang, An integrated computational methodology with data-driven machine learning, molecular modeling and PBPK modeling to accelerate solid dispersion formulation design, European Journal of Pharmaceutics and Biopharmaceutics, 158 (2021) 336-346.

[62] D. Reker, Y. Shi, A.R. Kirtane, K. Hess, G.J. Zhong, E. Crane, C.-H. Lin, R. Langer, G. Traverso, Machine Learning Uncovers Food- and Excipient-Drug Interactions, Cell Reports, 30 (2020) 3710-3716.e3714.

[63] P. Bannigan, M. Aldeghi, Z. Bao, F. Häse, A. Aspuru-Guzik, C. Allen, Machine learning directed drug formulation development, Advanced Drug Delivery Reviews, 175 (2021) 113806.

[64] E. Callaway, 'It will change everything': DeepMind's AI makes gigantic leap in solving protein structures, Nature, (2020).

[65] F. Schneider, M. Koziolek, W. Weitschies, In Vitro and In Vivo Test Methods for the Evaluation of Gastroretentive Dosage Forms, Pharmaceutics, 11 (2019) 416.

[66] S. Harrer, P. Shah, B. Antony, J. Hu, Artificial Intelligence for Clinical Trial Design, Trends in Pharmacological Sciences, 40 (2019) 577-591.

[67] M. Elbadawi, S. Gaisford, A.W. Basit, Advanced machine-learning techniques in drug discovery, Drug Discovery Today, 26 (2021) 769-777.

[68] L.E. McCoubrey, S. Gaisford, M. Orlu, A.W. Basit, Predicting drug-microbiome interactions with machine learning, Biotechnology Advances, (2021) 107797.

[69] L.E. McCoubrey, M. Elbadawi, M. Orlu, S. Gaisford, A.W. Basit, Machine Learning Uncovers Adverse Drug Effects on Intestinal Bacteria, Pharmaceutics, 13 (2021) 1026.

[70] M. Elbadawi, L.E. McCoubrey, F.K.H. Gavins, J.J. Ong, A. Goyanes, S. Gaisford, A.W. Basit, Harnessing artificial intelligence for the next generation of 3D printed medicines, Advanced Drug Delivery Reviews, 175 (2021) 113805.

[71] M. Elbadawi, B. Muniz Castro, F.K.H. Gavins, J.J. Ong, S. Gaisford, G. Perez, A.W. Basit, P. Cabalar, A. Goyanes, M3DISEEN: A novel machine learning approach for predicting the 3D printability of medicines, Int J Pharm, 590 (2020) 119837.

[72] L.L. Lao, S.S. Venkatraman, N.A. Peppas, Modeling of drug release from biodegradable polymer blends, European Journal of Pharmaceutics and Biopharmaceutics, 70 (2008) 796-803.

[73] J. Siepmann, H. Kranz, N.A. Peppas, R. Bodmeier, Calculation of the required size and shape of hydroxypropyl methylcellulose matrices to achieve desired drug release profiles, International Journal of Pharmaceutics, 201 (2000) 151-164.

[74] A. Isreb, K. Baj, M. Wojsz, M. Isreb, M. Peak, M.A. Alhnan, 3D printed oral theophylline doses with innovative 'radiator-like' design: Impact of polyethylene oxide (PEO) molecular weight, Int. J. Pharm., 564 (2019) 98-105.

[75] A. Goyanes, M. Kobayashi, R. Martinez-Pacheco, S. Gaisford, A.W. Basit, Fused-filament 3D printing of drug products: Microstructure analysis and drug release characteristics of PVA-based caplets, Int. J. Pharm., 514 (2016) 290-295.

[76] A. Goyanes, A.B. Buanz, A.W. Basit, S. Gaisford, Fused-filament 3D printing (3DP) for fabrication of tablets, Int. J. Pharm., 476 (2014) 88-92.

[77] A. Goyanes, A.B. Buanz, G.B. Hatton, S. Gaisford, A.W. Basit, 3D printing of modified-release aminosalicylate (4-ASA and 5-ASA) tablets, Eur. J. Pharm. Biopharm., 89 (2015) 157-162.

[78] M. Elbadawi, T. Gustaffson, S. Gaisford, A.W. Basit, 3D printing tablets: Predicting printability and drug dissolution from rheological data, International Journal of Pharmaceutics, 590 (2020) 119868.

[79] M. Madzarevic, D. Medarevic, A. Vulovic, T. Sustersic, J. Djuris, N. Filipovic, S. Ibric, Optimization and prediction of ibuprofen release from 3D DLP printlets using artificial neural networks, Pharmaceutics, 11 (2019) 544.