Journal Pre-proofs

M3DISEEN: A Novel Machine Learning Approach for Predicting the 3D Printability of Medicines

Moe Elbadawi, Brais Muñiz Castro, Francesca K.H. Gavins, Jun Jie Ong, Simon Gaisford, Gilberto Pérez, Abdul W. Basit, Pedro Cabalar, Álvaro Goyanes

PII: S0378-5173(20)30822-X
DOI: https://doi.org/10.1016/j.ijpharm.2020.119837
Reference: IJP 119837

To appear in: *International Journal of Pharmaceutics*

Received Date: 20 July 2020
Revised Date: 26 August 2020
Accepted Date: 27 August 2020

Please cite this article as: M. Elbadawi, B. Muñiz Castro, F.K.H. Gavins, J. Jie Ong, S. Gaisford, G. Pérez, A.W. Basit, P. Cabalar, A. Goyanes, M3DISEEN: A Novel Machine Learning Approach for Predicting the 3D Printability of Medicines, *International Journal of Pharmaceutics* (2020), doi: https://doi.org/10.1016/j.ijpharm.2020.119837

M3DISEEN: A Novel Machine Learning Approach for Predicting the 3D Printability of Medicines

Moe Elbadawi[1a], Brais Muñiz Castro[2a], Francesca K.H. Gavins[1], Jun Jie Ong[1], Simon Gaisford[1,3], Gilberto Pérez [2], Abdul W. Basit[1, 3, *], Pedro Cabalar[4], Álvaro Goyanes[3, 5, *]

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

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

[3]FabRx Ltd., 3 Romney Road, Ashford, Kent, TN24 0RW, UK.

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

[5]Departamento de Farmacología, Farmacia y Tecnología Farmacéutica, I+D Farma Group (GI-1645), 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

Artificial intelligence (AI) has the potential to reshape pharmaceutical formulation development through its ability to analyze and continuously monitor large datasets. Fused deposition modeling (FDM) 3-dimensional printing (3DP) has made significant advancements in the field of oral drug delivery with personalized drug-loaded formulations being designed, developed and dispensed for the needs of the patient. However, the optimization of the fabrication parameters is a time-consuming, empirical trial approach, requiring expert knowledge. Here, M3DISEEN, a web-based pharmaceutical software, was

developed to accelerate FDM 3D printing, which includes producing filaments by hot melt extrusion (HME), using AI machine learning techniques (MLTs). In total, 614 drug-loaded formulations were designed from a comprehensive list of 145 different pharmaceutical excipients, 3D printed and assessed in-house. To build the predictive tool, a dataset was constructed and models were trained and tested at a ratio of 75:25. Significantly, the AI models predicted key fabrication parameters with accuracies of 76% and 67% for the printability and the filament characteristics, respectively. Furthermore, the AI models predicted the HME and FDM processing temperatures with a mean absolute error of 8.9 ºC and 8.3 ºC, respectively. Strikingly, the AI models achieved high levels of accuracy by solely inputting the pharmaceutical excipient trade names. Therefore, AI provides an effective holistic modeling technology and software to streamline and advance 3DP as a significant technology within drug development. M3DISEEN is available at (http://m3diseen.com/predictions/).

Keywords: additive manufacturing, feature engineering, personalized pharmaceuticals, gastrointestinal drug delivery, 3D printed drug products, material extrusion, fused filament fabrication.

1. Introduction

Three-dimensional printing (3DP) is the *state-of-the-art* fabrication technology, which has achieved disruptive innovations across a number of fields (Sun et al., 2019). 3DP is an additive manufacturing technology, with the unique ability to produce personalized objects with complex designs at reduced costs and with high resolution. These traits are ideal for achieving precision solid dosage forms (Alhnan et al., 2016; Goyanes et al., 2015c; Trenfield et al., 2018). There are several 3DP technologies currently being investigated for pharmaceutical applications, including powder bed inkjet printing that was used to manufacture the anti-epileptic drug delivery system Spritam (Aprecia_Pharmaceuticals,

2015). This was the first 3DP drug to reach the US market and is available in four dose strengths. Other 3DP technologies include *fused deposition modeling* (FDM) (Goyanes et al., 2014; Isreb et al., 2019; Kempin et al., 2018), *selective laser sintering* (Allahham et al., 2020; Awad et al., 2020), *semisolid extrusion* (Goyanes et al., 2019; Vithani et al., 2019), *direct powder extrusion* (Ong et al., 2020) and *stereolithography* (Xu et al., 2020).

FDM 3DP has been extensively explored in the pharmaceutical field (Alhijjaj et al., 2016; Jamróz et al., 2017) because of its small and quick production runs, ability to print multi-material and multi-drug (Gioumouxouzis et al., 2018) delivery systems with tailored shapes (Goyanes et al., 2015b) and release characteristics (Genina et al., 2017; Gioumouxouzis et al., 2017; Goyanes et al., 2015a; Maroni et al., 2017), personalized to the needs of patients (Pereira et al., 2020; Zema et al., 2017; Zhang et al., 2017). Its low material wastage, low capital cost and compact size offers the possibility of deploying the technology in clinical settings, such as hospitals and community pharmacies. Although first, several challenges must be addressed to achieve this long-term goal of personalized 3DP drug delivery systems.

The preparation of personalized drug delivery systems using FDM 3D printing is a two-step process (Figure 1) (Awad et al., 2018). First, the raw pharmaceutical materials (drug and excipients) are thoroughly mixed and poured into a hot melt extruder (HME) that applies both heat and shear stress to homogenize the admixture. The molten materials are then extruded via the nozzle of the HME to obtain a filament (Tiwari et al., 2016). Then, the filament, which is the feedstock for the FDM 3D printer, is fed into the printing nozzle, where it is heated again to a semi-molten state and is extruded to form the designed object. The 3D shape is pre-determined using a computer-assisted design (CAD) software.

Figure 1. Schematic of the HME and FDM 3D printing process. The blue text shows examples of variable that can influence the fabrication process and the red text relates to characteristics of the products at each stage

The fabrication of a printable filament is a challenging and, so far, very empirical process. Many variables affect the process, including the proportion of the starting materials, the extrusion temperature and the printing temperature (Figure 1). The mechanical properties of the filament are paramount, as filaments that are notably too flexible or too brittle, or not within a given dimensional tolerance will not be compatible with the printer (Nasereddin et al., 2018). Since there are many pharmaceutical materials and brands that could potentially be tested, an empirical trial approach is used where different drug and pharmaceutical compositions are manually tested. Therefore, the formulation development stage is time-consuming, costly and resource intensive, even for an experienced FDM scientist. There is a need to optimize the filament development process to improve the existing approach, and thus improve the efficiency of the fabrication process.

In other fields, such as aerospace, standard fabrication techniques have benefited from specialized software to facilitate the development stage. However, these are rarely readily available and not used in formulation development in pharmaceutics (Leuenberger and Leuenberger, 2016). Pharmaceutical research should look to other industries for creative innovations as inspiration to apply such modeling tools within the pharmaceutical technology field (Ekins, 2016). For example, finite element analysis and computational fluid dynamics are professional software that are readily accessible as web services for engineers to guide manufacturing. In pharmaceutical sciences, the most used software is usually based on *design of experiments* (DOE), which aims to identify the key parameters that affect the process or the result. However, DOE is limited to datasets with low dimensions (i.e. number of variables), requires prior knowledge of the fabrication process, needs numerous experiments

to complete the model, and is unable to learn from existing experimental data (Hussain et al., 1991; Paulo and Santos, 2017; Rantanen and Khinast, 2015; Singh et al., 2011). Additionally, conventional experimental design studies are usually limited to an offline mode, and are not suitable for dynamic studies (Acherjee et al., 2011; Dehnad, 2012). The use of specific experimental design tools often requires the preparation of unprintable filaments that experienced FDM researchers would not prepare, but the software requires such preparation and inputs in order to achieve appropriate levels of statistical power and sensitivity. For all these reasons, the use of alternative data processing approaches is required to further our understanding of the fabrication process for the development of new drug-loaded formulations (Landin and Rowe, 2013).

Machine Learning is an Artificial Intelligence (AI)-based emerging technology that allows pattern recognition from complex datasets. AI is sparking global interest having already created and executed breakthrough developments across a number of disciplines (Baker, 2018; Hatamlou, 2013; Nikolaev et al., 2016; Popova et al., 2018; Xianyu et al., 2018). In the medical field, the success of AI is well-publicized, most notably in outperforming clinicians in diagnostic tests (Hosny et al., 2018). In the pharmaceutical industry, AI is affording researchers a renewed perspective towards minimizing the time-consuming and costly process of bringing drugs to the market (Han et al., 2019; Harrer et al., 2019; Schneider et al., 2019) and making progress towards precision and personalized drug delivery systems (Xu et al., 2019). However, the current use of AI in both pharmaceutical sciences and 3D printing remains underdeveloped. A few studies have focused on either quality control (Nam et al., 2020; Shen et al., 2020), or establishing a correlation between structure and property (Li et al., 2019; Nasereddin et al., 2018). Integrating AI with formulation development requires substantial datasets for training the AI models. Given the large number of drug delivery systems that could be developed from the many available

materials and the parameters that require prediction, AI machine learning techniques constitute the perfect tools for formulation development in FDM 3D printing. AI models can comprehend both structured and unstructured data (Balducci and Marinova, 2018). Furthermore, AI is capable of continuous learning (D'Souza et al., 2020), and hence will not require an experienced user to continually train the models.

The aim of this study was to explore AI machine learning techniques to increase the efficiency of the FDM formulation development process by developing a web-based software. Here, several AI machine learning methods were trained, tested, compared and evaluated to predict key fabrication parameters. Subsequently, M3DISEEN was developed to allow users to design 3D printed drug-loaded formulations and predict four key process parameters, namely extrusion temperature, filament mechanical characteristics, printing temperature and printability in an off-site setting, expediting the fabrication of 3D printed drug-loaded products.

2. Experimental Section

2.1 Pharmaceutical materials

Drugs and excipients were purchased from different suppliers. A list of the materials used in this study can be found in the Table S1 (Supplementary Material).

2.2 Preparation of drug-loaded filaments by hot melt extrusion (HME)

Pharmaceutical materials (drug and excipients) were selected for the preparation of Printlets (3D printed tablets) with diverse drug release rate characteristics. For each proposed drug-loaded formulation, a mixture of materials of 40 g was prepared with a theoretical drug content ranging from 4.5 to 40 % w/w. The selected materials were mixed for at least 5 minutes using a mortar and pestle until a homogenous blend was obtained; and subsequently

extruded using a single-screw filament extruder (Noztek Pro hot melt extruder, Noztek, UK) equipped with 1.75 mm nozzle, and a screw speed 15 rpm with extrusion temperature varied from 50-200 °C. The extruded filaments obtained were protected from light and kept in a vacuum desiccator until printing.

2.3 3D Printing of drug-loaded formulations by fused-deposition modeling (FDM)

Printlets were prepared from the drug-loaded filaments using a commercial fused-deposition modeling 3D printer (MakerBot Replicator 2X, MakerBot Inc, USA). AutoCAD 2014 (Autodesk Inc., USA), which was used to design the templates of the Printlets, exported as a stereolithography (`.stl`) file into 3D printer software (MakerWare v. 3.7.0, MakerBot Inc., USA). The selected 3D geometry was a cylinder (10 mm diameter × 3.6 mm height). The `.stl` format contains only the object surface data, and all the other parameters need to be defined by the MakerBot software in order to print the desired object. The printer settings were as follows: standard resolution without raft, extrusion temperature of 70-220 °C, extrusion speed 90 mm/s, travel speed 150 mm/s, number of shells 2, infill percentage 100% and layer height 0.10 mm.

2.4 Data Collection

Drug-loaded formulations were prepared using extrusion (HME) and printing (FDM) from a total of 145 different materials (including 7 drugs) over a period of six years. From the large dataset of prepared drug-loaded formulations, 614 were included in this study, which fulfilled the extrusion and printing conditions detailed in section 2.2 and 2.3. The composition used in each drug-loaded formulation was chosen with the aim of fabricating Printlets with suitable mechanical properties, based on the experience of expert HME and FDM operators from University College London – School of Pharmacy (London, UK) or the company FabRx

(London, UK). The collected data were arranged as shown in Table 1. For each drug-loaded formulation, both the operator's ID and formulation code were recorded. The composition of the excipients (ratio of weight of each component to total weight) was also recorded and care was taken to ensure that the accumulative total ratio was summed to 1. The extrusion temperature, filament mechanical characteristics, printing temperature and whether the drug-loaded formulation was printable or not were also recorded.

Table 1. Example of the collected data matrix

Formulation code	Operator ID	Material 1	Material 2	…	Material 145	Extrusion temperature (°C)	Filament mechanical characteristics	Printing temperature (°C)	Printability
ID1	User X	0.3	0.3	…	0.05	120	Flexible	180	Yes
ID2	User X	0	0.6	..	0.05	100	Good	160	Yes
…	…	…	…	..	…	..	…	…	…
IDn	User X			…					

Materials and their ratio in the formulation

Targeted variables (predicted parameter)

2.5 Predicted Target Variables

The key parameters that the study aimed to predict were the extrusion temperature, filament mechanical characteristics, printing temperature and printability (Table 2), referred to as *targeted variables*. The extrusion temperature and printing temperature would usually be classified as independent variables as the operator of the equipment can select the value of parameters. However, to achieve printable filaments, the optimal values were dependent on the selected materials and their percentage composition in the drug-loaded formulations. For this reason, extrusion temperature and printing temperature are considered herein dependent variables. Other parameters shown in Figure 1, such as HME screw speed, and FDM extrusion speed, travel speed, build-plate temperature, etc., were kept constant for this study.

Table 2. Summary of the predicted targeted variables

Targeted variables	Values	Analysis Type
Extrusion temperature	HME temperature (°C)	Regression
Filament mechanical characteristics	Unextrudable, Flexible, Good or Brittle	Multi-classification
Printing temperature	Printing temperature (°C)	Regression
Printability	Yes or No	Binary Classification

Regression analyses were performed to predict both the HME and FDM processing temperatures, since the targeted variables were continuous numerical values. Categorical analysis was used for predicting filament mechanical characteristics and printability. Filament mechanical behavior varied depending on the composition and was qualitatively classified as either 'good', notably 'brittle', notably 'flexible' or 'unextrudable'. A *good* filament refers to a filament that exhibits similar mechanical behavior to commercial filaments. A *brittle* filament was defined as one that was susceptible to fracturing when it was bent from 180º to 90º, whereas a *flexible* filament was one that would easily bend when held only from one side due to a lack of structural integrity. Pharmaceutical materials, where a filament could not be obtained by HME, even when tested over a wide range of HME temperatures, were classed as *unextrudable*. Printability was qualitatively classified as either 'Yes' or 'No' depending on whether the filament was able to be extruded through the nozzle of the FDM printer given the selected printing parameters.

2.6 Feature set selection and creation

In this study, the starting input variables or *features* used for predictions were the proportions of each material in the drug-loaded formulations. Different learning results can be obtained by changing the *set of features* used in the dataset, either by choosing only some of them (*feature selection*) or creating new derived ones (*feature creation*). Here, five different

feature sets were used (Table 3). For all five sets, the number of samples remained the same (n=614), however the input features varied for each feature set. More information on the structure of each dataset can be found in Tables S2, S3, S4, S5 and S6 (Supplementary Material).

Table 3. Summary of the five feature sets used

Feature Set	Input Feature	Input Value	Number of Features in the dataset
Material Name	Product grade name	Weight fraction	145
Material Type	Materials grouped together by their chemical structure	Weight fraction	36
Material Role	Materials grouped by their functional role	Weight fraction	12
Physical properties	Total weighted physical properties	Weighted physical properties	3
Physical properties per material type	Materials grouped together by their chemical structure.	Weighted physical Properties	108*

*3 Physical properties × 36 Material Type

The first feature configuration, named *Material Name*, corresponds to the inputs used in the original dataset, where the features were labelled by their manufacturer's tradename (Table S2, Supplementary Material). The weight fraction of each material used (ratio of the weight of each component to the total weight) for the drug-loaded formulation were used as values.

Using the individual manufacturer's tradename is an approach where the machine learning technique could classify the inputs according to the specific excipients and drug substances. A limitation could be that the resulting model may have been unable to decipher or comprehend a new material, whether it was produced by the same manufacturer with a different tradename or from a different manufacturer; despite it being chemically similar. For this reason, a second feature set was created to group chemically similar excipients together.

For example, Aqoat LG, MG and HG, as hypromellose acetate succinate formulations, were grouped into one feature labelled 'HPMCAS'. Another example are polyethylene glycols with molecular weights of 2000, 4000 and 8000 that were grouped together under the label 'PEG'. This feature configuration was named *Material Type*, where the features used were the weight fraction of each material type within the drug-loaded formulation (ratio of the weight of components of same type to the total weight). The active ingredients used were grouped together to produce a feature input labelled 'drug'. In this study, the materials were classified into 36 types (Table S3, Supplementary Material).

A third feature set, referred to as *Material Role*, grouped materials by their principal function (Table S4, Supplementary Material). For example, Soluplus and Eudragit were labelled as 'primary polymers' if they formed the main matrix of the admixture, as determined by the ratio of the weight of each component to the total weight. Polyethylene glycol and methyl paraben, both widely used plasticizers in pharmaceutical formulations, were labelled as 'plasticizers'. This feature set was less specific than the previous two, and thus new excipients could be classified in a straightforward manner and inputted. For example, a model built on this approach would only need the function of the material, without requiring the user to identify the exact trade name or chemical name. Again, the weight fractions for all material roles were used as values.

The fourth and fifth feature sets differ in that the weighted physical properties were used as values, rather than the ratio weight of each component to the total weight. For the fourth feature configuration, each material name in the drug-loaded formulation was replaced by three of its *Physical Properties*: the glass transition temperature T_g (ºC), the melting temperature T_m (ºC) and the molecular weight *mwt* (g/mol) (Table S5, Supplementary Material). Then, the whole drug-loaded formulation was represented as an average of each of

those three properties, but *weighted* by the proportion of the materials present in the mixture. For instance, the weighted average for T_g is (Equation 1):

$$Weighted\ T_g = \frac{(w_1 \times T_{g1}) + (w_2 \times T_{g2}) + \ldots (w_n \times T_{gn})}{(w_1 + w_2 + \ldots w_n)} \qquad \textbf{(Equation 1)}$$

where *w* is the weight fraction of the sample. As a result, under this configuration, each drug-loaded formulation is reduced to three input features: (the weighted averages for) T_g, T_m and *mwt*. The materials that did not have a T_g or a T_m, were not included in the equation. The physical properties for each material were obtained from either the Handbook of Pharmaceutical Excipient (Rowe et al., 2009), the respective manufacturer's datasheet, or from the literature. The rationale for using the physical properties as inputs was that physical properties offer a more objective, property-led classification than using the trade names.

For the fifth feature set, the materials' individual physical properties were used and were weighted by the proportion of the materials present in the mixture. For example, if 80 % w/w of a material was used in a formulation, and it possessed a T_g of 100 °C, then the T_g was weighted as 80 °C. Then the materials were grouped by their material type, whilst maintaining the individual material weighted physical property. The materials were grouped by their type for the same reason in the second feature set, as grouping chemically similar materials together could facilitate the model's ability to generalize to new data. Same as the second feature set, the active ingredients were grouped together in a feature input called 'drugs'. This led to a total of 108 features in the dataset, comprising of three physical properties for each of the 36 material type (Table S6, Supplementary Material).

2.7 Analysis of the data

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

2.7.1 Machine Learning Techniques (MLTs)

Six different machine learning techniques (MLTs) were used in this study to explore which MLTs best modelled the dataset. The tested MLTs, multivariate linear regression (MLR), *k*-nearest neighbors (*k*NN), support vector machines (SVM), random forests (RF), (traditional) neural networks (NN) and deep learning (DL) (using layered neural networks), are regularly used in the data science field with a strong evidence-base. Each MLT has its own learning characteristics. MLR was used for regression and *k*NN was used for classification whereas the other MLTs were used for both classification and regression.

The MLTs were developed using python 3.7 (Python Software Foundation). All of the models except for deep learning were installed using a machine learning library for the Python programming language (scikit-learn package, v0.21.3 (Pedregosa et al., 2011)). Neural networks were modelled using the *multi-layer perceptron* algorithm. Deep learning was performed using the neural-network library (Keras package, v2.3.1) and the machine learning platform TensorFlow (v2.0.0) as the backend (Géron, 2019). For the interested reader, an informal overview of MLTs can be found in reference (Domingos, 2015), containing the key bibliography references.

For the training of the models, the data were randomly split into training and testing data at a ratio of 75:25. The best hyper-parameters were determined for each model from a grid search during the 5-fold cross-validation stage, performed on the training set, and were then applied to the testing dataset (Schmidt et al., 2019; Wainberg et al., 2018). Figure S1

provides an illustration of the 5-fold cross-validation approach (Supplementary Material). Brief explanations of each MLT are described in the following sections and provided in the Supplementary Material Document 1.

Multi-Linear Regression (MLR)

A linear equation was developed, whereby the output variable is a linear combination and an approximation of all the input variables. No hyper-parameters were specified for MLR.

k-Nearest Neighbors (kNN)

The hyper-parameter *k* was defined based on the results of the grid search performed with *k* values ranging from 1 to 25.

Support Vector Machines (SVM)

The defined features were not linearly separable, so a *kernel* function was used to map their values into a higher dimensional space, whereby more hyperplanes were searched. For all experiments, a fixed kernel called the (*Gaussian*) *Radial Basis Function* (RBF) was used. The RBF kernel was adjusted by the parameter *gamma* (γ) and the cost *C* function, which was used to mitigate over-fitting. The grid search explored multiple combinations for the *C* and γ.

Random Forests (RF)

For classification, unseen data was evaluated against all of the ensembled trees and the classification was determined through a majority vote by the decision trees; hence, the mode of the classes was used. For regression, the target value was determined by the arithmetic

mode for the baseline. A valuable additional outcome of RF was that it provided a *rank of the importance of features,* used here to discriminate between classes (Belgiu and Drăguţ, 2016).

Neural Networks

Networks were organized in a layered structure with at least two layers: the first layer collected the values of all the input features and the last layer contained the target variables (Figure S2, Supplementary Material). The number of hidden layers and number of nodes were tuned by the grid search.

Deep Learning

The process was divided into three distinct stages. First, the model was defined whereby the number of hidden layers and their respective number of nodes, activation function and regularized terms were selected. Second, the model was compiled where the loss function, optimizer and metrics were selected. Lastly, the training and evaluation parameters were selected, in which the number of epochs, batch size and validation split were selected. A grid search was performed, factoring in the number of neurons, the number of hidden layers, the optimization algorithm, batch size and number of epochs.

Evaluation

Different metrics were used for scoring the accuracy of the MLTs, as no single metric conveys a complete picture of a model's performance. A brief explanation of each metric is provided in the Supplementary Material Document S1.

For classification analyses six classification metrics were used; *accuracy*, *Kappa-k*, *precision*, *recall*, F_1 and *area under a curve receiver operating characteristics* (AUROC). For the processing temperature analyses, four regression metrics were used: the *root mean*

square error (RMSE), the *mean absolute error* (MAE), the *mean absolute percentage error* (MAPE) and the *coefficient of determination* (R^2). Additionally, computational metrics were obtained for each MLT; the *computed time*, the *real time*, the *fit time* and the *predict time*.

3. Results

3.1 Exploratory Data Analysis

An exploratory data analysis was performed prior to the processing of the dataset using MLTs. The analysis provided an overview of the data to be analyzed and can aid in detecting anomalies, which can be detrimental to some MLTs. 49 of materials were used more than six times (Figure 2a). While 44 materials were only used once. The materials used as plasticizers or lubricants were frequently used (for example magnesium stearate (501 times), mannitol (341 times) or triethyl citrate (219 times)). The most used drug was paracetamol (373 times).

Figure 2. a) Distribution of the materials according to the number of times used (percentage), b) Histogram depicting the distribution of the glass transition temperature (T_g), melting temperature (T_m) and molecular weight (*mwt*). c) Pie charts classifying all the drug-loaded formulations based on the filament mechanical characteristics (percentage). d) The printability of the filaments as a function of filament mechanical characteristics. e) Histograms illustrating both the HME extrusion and FDM printing temperatures.

The physical properties of the materials were used as inputs for two out of the five of the feature configurations. The glass transition temperatures of the materials ranged from -60 to 1027 ºC, with the majority below 200 ºC (Figure 2b). For the melting temperature, the range was from -55 to 1855 ºC, with the majority of materials possessing values below 300 ºC. The high T_g and T_m values are likely to be due to the inclusion of inorganic fillers in the drug-loaded formulations, such as titanium dioxide and calcium phosphate. The molecular weight of the materials ranged from 10^1 to 10^7 g/mol. Therefore, a combination of both low and high molecular weight materials were used. From the analysis of the materials, it was evident that the inputs should be normalized. Normalization is a mathematical operation conventionally used in the pre-processing of the data. It ensures that the inputs occupy equivalent scales and prevents values that are considerably numerically larger from being assigned greater weights. Hence, normalization ensures that the MLT techniques are attributing equal weights to each input variable, and thereby avoiding errors when fitting the models to the test dataset (Nawi et al., 2013). Exploratory data analysis of the output parameter showed that 11.2% of the drug-loaded formulations could not be extruded via HME (Figure 2c). Although all the drug-loaded formulations were prepared with the aim of FDM printing, some of the tested drug-loaded formulations could not be extruded across a range of extrusion temperatures and were therefore considered unextrudable. Nevertheless, 88.8% were extrudable, with the majority found to produce good filaments.

The analysis revealed that using FDM good filaments were approximately twice as likely to be successfully printed than found to be unprintable, with 67.4% found to be printable (Figure 2d). Conversely, if a filament was either brittle or flexible, then there was an equal chance that it was printable. 52.4% of flexible filaments were printable, whereas 53.6% of brittle filaments were printable. The HME temperatures selected ranged from 30 to 190 ºC, with a mode of 100 ºC (Figure 2e). The temperatures used for FDM ranged from 80 to 240 ºC, with a mode value of 190 ºC. Thus, the dataset consisted of a wide processing temperature range for both HME and FDM. As expected, the FDM printing temperatures are higher than the HME temperatures. This culminated as 49.2% of the drug-loaded formulations being successfully printed via FDM. The value increased to 55.4% when the drug-loaded formulations were successfully extruded by HME.

3.2 Predictability Evaluation

3.2.1 Predicting the Filament mechanical characteristics

MLTs were able to predict the filament mechanical characteristics with high accuracy. To be precise, using the Material Name configuration, RF produced the highest test accuracy of 73%. A *kappa* value of 0.61 was obtained, indicating a significant improvement over the baseline (Figure 3a). RF was also found to outperform the other MLTs when using the feature sets of Material Type, Physical Properties and Physical Properties per Material Type; with respect to test accuracy. When analyzing Material Role, SVM was found to achieve the highest test accuracy of 59%. When using the Physical Properties configuration, a noticeable shrinkage in the coverage of the radar plot was observed, compared with the other configurations.

Figure 3. Metrics results for predicting a) the filament mechanical characteristics and b) printability using MLTs.

3.2.2 Predicting the Printability

The main objective of this study was to determine whether MLTs could predict the printability of drug-loaded formulations. Using the individual Material Name as inputs, SVM achieved the highest overall test accuracy of 76% (Figure 3b). The precision and recall were 80% and 73%, respectively. The *kappa* value was 0.52, which was within the 0.4-0.6 category classified as significantly better than random chance. RF and DL also achieved *kappa* values within this range. For predicting the printability when the drug-loaded formulations were grouped by their Material Types, NN was found to achieve the highest overall accuracy of 72%, which was one percentage point higher than RF. Both MLTs were found to possess precision and recall values of 72%. Using the Physical Properties per Material Type as inputs resulted in an overall test accuracy of 73%, which was by RF. The precision and recall values were also 73%, and the *kappa* value was 0.45. RF achieved the highest overall accuracy of 70% when the weighted Physical Properties were used as inputs. The precision and recall values were both 70%, and the *kappa* value was 0.4. When the materials were categorized by their Material Role, all MLTs were found to underperform in comparison to the Material Name dataset. The highest accuracy obtained was 63% for this dataset, which was obtained by both SVM and RF. It was concluded that four of the five datasets were able to achieve high prediction scores, with significant improvement over the baseline.

3.2.3 Predicting the Extrusion Temperature

The best performing MLT for predicting the extrusion temperature was RF. RF outperformed the other algorithms in four out of the five feature datasets, with SVM outperforming in the Material Role feature configuration (Figure S4, Supplementary Material). The lowest MAE recorded was 10.8 C, which was the Physical Properties per Material Name dataset. This was a very low value given the overall range for the extrusion temperature was 160 ºC. Furthermore, the baseline MAE value was 25.0 C, which was over 2-fold greater than RF. In general, SVM, NN and DL were found to produce MAE values below 15 ºC. Multivariate linear regression was found to perform poorly, with MAE above the baseline value.

RF produced the lowest error metric on the testing set and thus is the best technique tested here. Comparing the actual to predicted value produced an R^2 value of 0.56 on the testing set. In comparison, the R^2 value of the training set was 0.95. The discrepancy indicates that overfitting had occurred, whereby the technique captured specific information from the training data but is not able to fully transfer to this the testing set. As illustrated in Figure 4, the testing data had a greater range than the training dataset results.

Figure 4. Scatter plot illustrating the training and testing set results for (a) extrusion temperature and (b) printing temperature results. The metrics represent the testing dataset results.

3.2.4 Predicting the Printing Temperature

Once the filaments were extruded, they were loaded into the FDM printer. Here, assumptions were made for the ideal printing temperature. If premature extrusion was observed, then the temperature was decreased, typically in decrements of 5 ºC, until a stable and controlled flow was obtained. This approach is time-consuming and requires expertise in FDM 3DP. A similar process was followed if the filament could not be extruded through the nozzle, where the temperature would be gradually increased until extrusion was observed. However, the

maximum temperature was limited by the FDM printer, which is typically below 300 ºC. In addition, the maximum temperature was also limited in cases where thermal degradation of the filament occurred. Hence, identifying if a drug-loaded formulation can be printed within these constraints will minimize both time and material waste.

The results from the algorithms revealed that a MAE of 8.4 ºC was the best value obtained. This was achieved by RF using the Physical Properties per Material Type as input features. RF was also able to achieve an MAE of 8.9 ºC with the Material Type feature set. By using the arithmetic mean as a baseline, the MAE was found to be 23 ºC, which is approximately a 3-fold increase. Hence, a narrower prediction range was achieved by RF. Overall, the algorithms were able to achieve MAE values below 15 ºC, except for MLR. Again, MLR was unable to interpret the datasets, and performed poorly compared to the baseline. For the training dataset, the R^2 was 0.95, which decreased to 0.83 for the testing set (Figure 4) pointing out overfitting. To mitigate this effect, cross validation was performed on both the extrusion and printing temperature analyses. The results suggest that more measures were needed to further minimize overfitting.

3.3 Feature Importance

As aforementioned, RFs additionally provide a ranking function that reflects the importance of each feature in the obtained predictive model. In this way, RFs not only learn to make predictions, but also to provide some nuanced form of *knowledge*, highlighting the relevant factors involved in the prediction. This may be helpful when designing and fabricating a new drug-loaded formulation to identify which variable should be prioritized during formulation development. The results of applying RF feature importance for the Material Role feature set are presented in Figure 5. It was discovered that for predicting printability, filament mechanical characteristics and extrusion temperature, the key determinants were polymer and plasticizer. For predicting the FDM printing temperature, RF determined that the lubricant

was the key determinant, followed by plasticizer and polymer. The feature importance of colorants, inorganic fillers and disintegrants were comparatively minor. These materials were added to impart an additional functionality beyond the processing step, for example to enhance the aesthetic appearance. Thus, according to RF, the concentrations used for them had a relatively minor influence on the processing parameters.

Figure 5. Feature Importance plots for the different target variables determined by random forest for a) Extrusion temperature, b) Filament Mechanical Characteristics, c) Printing Temperature and d) Printability.

3.4 Results of Computational Time

The time it took to train the models and the time to predict the drug-loaded formulations was also measured whilst using the conventional PC computer. For these models, the training time, the time it took to complete the hyper-parameter tuning and train the models on the training set, ranged from a few seconds for *kNN* and *SVM*, to the order of days for *RF* and *NN* (Table S7, Supplementary Material). For all MLTs, the time taken to make a prediction for one drug-loaded formulation was a fraction of a second, ranging from 3 milliseconds to 129 milliseconds. The wide range in the training time was a reflection of the different hyper-parameters selected for tuning for each technique.

3.5 Online Web services

The tested MLTs were integrated in a web application service that can be accessed from any smart device connected to internet, using this link (http://m3diseen.com/predictions/). The application is hosted in an Elastic Compute Cloud (EC2) instance in Amazon Web Services (AWS), using the open-source software Apache HTTP Server for serving a web application written in Python3 using the Django web framework. The scikit-learn package was used to

integrate and make use of the machine learning models. This will allow any user to access the service from a laboratory, an office or a remote setting based on their requirements and situation. All the processing is done remotely, so the performance of the predictions does not depend on the user's device. This user-friendly software, named M3DISEEN, quickly predicts the printability, extrusion temperature, filament mechanical characteristics and printing temperature of any potential drug-loaded formulation before preparation, based on the selected composition.

4. Discussions

A family of AI machine learning tools were successfully built from a large dataset of drug-loaded formulations that were prepared *in-house*. The results demonstrated that using the Material Name inputs succeeded in achieving a high predictive score, with DL, RF and SVM able to achieve an overall accuracy above 70%. SVM in particular outperformed all MLTs using this feature set, achieving an overall accuracy of 76%, and achieving AUC (Islam et al., 2019), *kappa* and F_1 score metrics that are good predictive values. The *kappa* value verified that, in addition to high accuracies, MLTs were a significant improvement over the baseline. There is no one conclusive and superior MLT, and hence several were investigated. Different techniques perform better depending on the size of the dimensions, samples and the linearity of the data (Redkar et al., 2020; Wade et al., 2017).

For most, the optimization of the printing is performed manually using the experience generated from previous printing of drug-loaded formulations. This approach is time consuming and therefore not cost effective. In addition, one batch yields 1 to 5 m of filament that must be destroyed if found to be unprintable. The same happens if the raw material mixture is unextrudable. The complexity is further compounded by the numerous excipients available, increasing the possible combinations and making formulation optimization a high-dimensional problem. The results obtained from the feature importance ranking were also

consistent with the expectation that both the polymer carrier and plasticizer used were key determinants for drug-loaded formulations.

In addition to predicting the printability of filaments, the MLTs were taken a step further and employed to predict the mechanical properties of filaments. This is a key component of the fabrication process, and prior knowledge can allow the user to take the necessary measures to ensure a successful print. For example, if the filament is predicted to be brittle by the AI tool, the user can incorporate plasticizers to obtain a less brittle filament. The results demonstrate that MLTs were able to attain significant AUROC and *kappa* values. From this, it can be inferred that the AI models developed can predict the mechanical properties of the printed objects. This is of particular interest since 3D printing can address, for example, brittleness by altering the design. Thus, proactive measures can be taken by the operator to address these mechanical properties through design changes.

The capabilities of the MLTs were also tested to predict both the extrusion temperature of HME and the printing temperature using FDM. Identifying the processing temperature is also a time-consuming procedure, particularly if the initial temperature investigated is far from the ideal value. In addition to knowing whether a drug-loaded formulation is printable, there are instances when the user will need to know if desirable temperature conditions can be achieved, for example when working with thermally labile drugs. Thus, predicting the processing temperature can facilitate the fabrication process. The best prediction results were obtained with random forest, obtaining the lowest MAE of 8.8 ºC. Previous works have been guided by the *rule of thumb* of using a temperature range between 15 to 60 ºC above the glass transition temperature when processing via HME (Andrews et al., 2010). Thus, the current approach was advantageous because; (i) the MLT was able to predict a temperature range that was narrower than the *rule of thumb*, with a mean range of ± 8.8 ºC; and (ii) this was achieved without the need for determining the T_g of

a new drug-loaded formulation, which would require material wastage and time. FDM is a relatively new fabrication technique and a *rule of thumb* for the printing temperature has not been established yet. The AI models herein have provided a narrow range, and hence there is no need to establish a generic rule. Future work will seek to further minimize both MAE and RMSE values.

Another benefit of using machine learning was that it was directly applicable to an already existing formulation development dataset. Other predictive tools would require the addition of specific experiments to be performed to produce a model, which can be more expensive than a trial-and-error approach. The current work demonstrated that AI can be used effectively to predict the processing behavior of common pharmaceutical excipients with drug incorporated.

Five different MLTs were employed herein for both classification and regression analyses, including the current *state-of-the-art* method, deep learning. The latter has been demonstrated to globally outperform RF and SVM (Qi et al., 2016), and hence, has gained recent popularity for a number of applications (Fadlullah et al., 2017). In our experiments, however, both RF and SVM were found to outperform DL, albeit marginally. The possible explanation for this is that DL's prowess lies in handling considerably larger datasets (from thousands to millions of examples), and hence, it would be expected to outperform both RF and SVM if the dataset was expanded from the current 614 drug-loaded formulations. *k*NN was found to yield the lowest predictability score when using the Material Names as input, where a decrease of 10 percentage points was observed, relative to RF, when predicting the printability and filament mechanical properties. However, when using the weighted physical properties as inputs, *k*NN was found to be either equal or marginally lower in predicting the outputs. *k*NN is a technique that is regarded as the easiest to operate, largely owing to the single hyperparameter that needs to be tuned. However, the technique is susceptible to the

curse of dimensionality. This means that datasets with high number of features (e.g. Material Name and Material Type) can cause learning problems for *k*NN (He et al., 2003; Stommel and Herzog, 2009). For regression analyses, multivariate linear regression was found to produce poor prediction scores, even when compared with the baseline value determined by the arithmetic mode. This suggests that the relationship between the input variables and the target variables is non-linear. In summary, different MLTs have different advantages and, thus, a reasonable AI prediction tool should incorporate a number of MLTs in order to attain the highest predictability performance.

Suppliers often introduce new excipients into the pharmaceutical market. At the same time, existing grades are discontinued. Furthermore, our dataset does not include all potential excipients available. Thus, it is important to ensure that an implemented AI tool is amenable to the introduction of new materials. With this approach, the software developed offers users flexibility when inputting their data.

The materials were categorized by grouping similar grades together. The results demonstrated that the MLTs were robust, with RF having a prediction value of 73%, when the materials were classed by their relative grades. Therefore, the MLTs were able to process datasets that converged similar material grades together, illustrating their versatility to recognize similar compounds.

RF was also capable of attaining a high predictability score when the total weighted physical properties of the drug-loaded formulations were examined. It assigned equal feature importance to both molecular weight and melting points (data not included). The Physical Properties selected for this study were based on readily available information. There are other properties to consider, such as viscosity and dynamic mechanical properties, which have the potential to improve the current predictive scores. Adding more properties would be more complex but would provide a more holistic representation of the formulation development

data. From this study, it was determined that the MLTs can handle both large datasets and non-linear relationships. Using the Physical Properties allows the AI to learn from the dataset by using the fundamental behavior of materials, rather than learning in terms of their brand names. Moreover, some excipients can be incorporated for different processing functions in different drug products. For example, polyethylene glycol can be used as the primary polymer or as a plasticizer. Thus, using the Physical Properties is the preferred option as it removes subjectivity by not labelling the excipients by their Material Name, Material Type or Material Role. The Physical Properties dataset were found to produce the lowest predictability metrics from the datasets examined herein, however the *kappa* values confirmed that the prediction scores were an improvement over the baseline. The results suggest that MLTs were able to capture aspects of the underlying variation in physical properties between the different drug-loaded formulations.

The computational times reported herein were discovered to be ideal for providing a web-based service. In real terms, the AI models are able to provide predictive results for all of the key parameters studied in a matter of seconds. Even though the time to adjust the models could take some days for some MLTs, the time it takes to make a prediction for one drug-loaded formulation is less than a second. The training time could be also reduced using better computer processor or cloud computing. Future work will seek to increase the data size, which is known to improve some MLTs predicting performance. In addition, alternative ML approaches can be explored, centering on techniques that will explain their decision making process. Lastly, a focal aspect of the machine learning pipeline is the pre-processing stages and feature engineering. Hence, different pre-processing strategies will also be explored. The tested MLTs integrated in the web services offer to any researcher around the world suitable predictions for the printability, extrusion temperature, filament mechanical characteristics and printing temperature of any potential drug-loaded formulation by selecting

the materials from the database and their proportions. Since the prediction of a single drug-loaded formulation takes less time than the connection of the computer to the webpage by the internet, this indicates that this approach can be suitable for rapid prediction of drug-loaded formulations saving valuable time and resources, even in remote locations without the need of big computational resources. M3DISEEN allows the user to prove the trained models offsite and then compare the predicted results with FDM studies in the pharmaceutical laboratory, expediting the 3DP process. Users of the service are afforded the efficiency of a high-throughput screening tool, combined with high predictability scores.

5. Conclusions

Here, a software tool named M3DISEEN was designed to accelerate the FDM formulation development process using AI machine learning techniques. Importantly, the AI models after training and testing can predict key fabrication parameters, namely the printability and filament characteristics, with high accuracies and HME and FDM temperatures within a relatively accurate narrow range. Our approach demonstrates that inputting in the composition of the drug-loaded formulation achieved high printing accuracies, and thus obviating the need for in-depth knowledge of the properties of pharmaceutical materials. M3DISEEN can be used as a tool to guide formulation development researchers, advancing the 3DP fabrication process.

Acknowledgements

The authors would like to thank Dr. Atheer Awad for her support with the graphical abstract. This research was funded by the Engineering and Physical Sciences Research Council (EPSRC) UK, grant number EP/L01646X. CITIC is a Research Center of the University

System of Galicia, funded by Consellería de Educación, Universidade e Formación Profesional of Xunta de Galicia and co-financed 80% by ERDF (Ref. ED431G 2019/01).

References

Acherjee, B., Kuar, A.S., Mitra, S., Misra, D., 2011. Application of grey-based Taguchi method for simultaneous optimization of multiple quality characteristics in laser transmission welding process of thermoplastics. The International Journal of Advanced Manufacturing Technology 56, 995-1006.

Alhijjaj, M., Belton, P., Qi, S., 2016. An investigation into the use of polymer blends to improve the printability of and regulate drug release from pharmaceutical solid dispersions prepared via fused deposition modeling (FDM) 3D printing. European Journal of Pharmaceutics and Biopharmaceutics 108, 111-125.

Alhnan, M.A., Okwuosa, T.C., Sadia, M., Wan, K.W., Ahmed, W., Arafat, B., 2016. Emergence of 3D Printed Dosage Forms: Opportunities and Challenges. Pharmaceutical Reseach 33, 1817-1832.

Allahham, N., Fina, F., Marcuta, C., Kraschew, L., Mohr, W., Gaisford, S., Basit, A.W., Goyanes, A., 2020. Selective Laser Sintering 3D Printing of Orally Disintegrating Printlets Containing Ondansetron. Pharmaceutics 12, 110.

Andrews, G.P., Abu-Diak, O., Kusmanto, F., Hornsby, P., Hui, Z., Jones, D.S., 2010. Physicochemical characterization and drug-release properties of celecoxib hot-melt extruded glass solutions. Journal of Pharmacy and Pharmacology 62, 1580-1590.

Aprecia_Pharmaceuticals, 2015. FDA approves the first 3D printed drug product. https://www.aprecia.com/pdf/2015_08_03_Spritam_FDA_Approval_Press_Release.pdf, last accessed 09-2018.

Awad, A., Fina, F., Goyanes, A., Gaisford, S., Basit, A.W., 2020. 3D printing: Principles and pharmaceutical applications of selective laser sintering. International Journal of Pharmaceutics 586, 119594.

Awad, A., Trenfield, S.J., Gaisford, S., Basit, A.W., 2018. 3D printed medicines: A new branch of digital healthcare. International Journal of Pharmaceutics 548, 586-596.

Baker, D.J., 2018. Chapter 11 - Artificial Intelligence: The Future Landscape of Genomic Medical Diagnosis: Dataset, In Silico Artificial Intelligent Clinical Information, and Machine Learning Systems, in: Lambert, C.G., Baker, D.J., Patrinos, G.P. (Eds.), Human Genome Informatics. Academic Press, pp. 223-267.

Balducci, B., Marinova, D., 2018. Unstructured data in marketing. Journal of the Academy of Marketing Science 46, 557-590.

Belgiu, M., Drăguţ, L., 2016. Random forest in remote sensing: A review of applications and future directions. ISPRS Journal of Photogrammetry and Remote Sensing 114, 24-31.

D'Souza, S., Prema, K.V., Balaji, S., 2020. Machine learning models for drug-target interactions: current knowledge and future directions. Drug Discovery Today 25, 748-756.

Dehnad, K., 2012. Quality control, robust design, and the Taguchi method. Springer Science & Business Media.

Domingos, P., 2015. The Master Algorithm: How the Quest for the Ultimate Learning Machine will Remake our World. Basic Books.

Ekins, S., 2016. The Next Era: Deep Learning in Pharmaceutical Research. Pharmaceutical Research 33, 2594-2603.

Fadlullah, Z.M., Tang, F., Mao, B., Kato, N., Akashi, O., Inoue, T., Mizutani, K., 2017. State-of-the-Art Deep Learning: Evolving Machine Intelligence Toward Tomorrow's Intelligent Network Traffic Control Systems. IEEE Communications Surveys & Tutorials 19, 2432-2455.

Genina, N., Boetker, J.P., Colombo, S., Harmankaya, N., Rantanen, J., Bohr, A., 2017. Anti-tuberculosis drug combination for controlled oral delivery using 3D printed compartmental dosage forms: From drug product design to in vivo testing. Journal of Controlled Release 268, 40-48.

Géron, A., 2019. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. Sebastopol: O'Reilly Media, Incorporated, Sebastopol.

Gioumouxouzis, C.I., Baklavaridis, A., Katsamenis, O.L., Markopoulou, C.K., Bouropoulos, N., Tzetzis, D., Fatouros, D.G., 2018. A 3D printed bilayer oral solid dosage form combining metformin for prolonged and glimepiride for immediate drug delivery. European Journal of Pharmaceutical Sciences 120, 40-52.

Gioumouxouzis, C.I., Katsamenis, O.L., Bouropoulos, N., Fatouros, D.G., 2017. 3D printed oral solid dosage forms containing hydrochlorothiazide for controlled drug delivery. Journal of Drug Delivery Science and Technology 40, 164-171.

Goyanes, A., Buanz, A.B.M., Basit, A.W., Gaisford, S., 2014. Fused-filament 3D printing (3DP) for fabrication of tablets. International Journal of Pharmaceutics 476, 88-92.

Goyanes, A., Buanz, A.B.M., Hatton, G.B., Gaisford, S., Basit, A.W., 2015a. 3D printing of modified-release aminosalicylate (4-ASA and 5-ASA) tablets. European Journal of Pharmaceutics and Biopharmaceutics 89, 157-162.

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

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

Goyanes, A., Wang, J., Buanz, A., Martínez-Pacheco, R., Telford, R., Gaisford, S., Basit, A.W., 2015c. 3D Printing of Medicines: Engineering Novel Oral Devices with Unique Design and Drug Release Characteristics. Molecular Pharmaceutics 12, 4077-4084.

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

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

Hatamlou, A., 2013. Black hole: A new heuristic optimization approach for data clustering. Information Sciences 222, 175-184.

He, J., Tan, A.-H., Tan, C.-L., 2003. On Machine Learning Methods for Chinese Document Categorization. Applied Intelligence 18, 311-322.

Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L.H., Aerts, H.J.W.L., 2018. Artificial intelligence in radiology. Nature Reviews Cancer 18, 500-510.

Hussain, A.S., Yu, X., Johnson, R.D., 1991. Application of Neural Computing in Pharmaceutical Product Development. Pharmaceutical Research 8, 1248-1252.

Islam, M.M., Nasrin, T., Walther, B.A., Wu, C.-C., Yang, H.-C., Li, Y.-C., 2019. Prediction of sepsis patients using machine learning approach: A meta-analysis. Computer Methods and Programs in Biomedicine 170, 1-9.

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

Jamróz, W., Kurek, M., Łyszczarz, E., Szafraniec, J., Knapik-Kowalczuk, J., Syrek, K., Paluch, M., Jachowicz, R., 2017. 3D printed orodispersible films with Aripiprazole. International Journal of Pharmaceutics.

Kempin, W., Domsta, V., Grathoff, G., Brecht, I., Semmling, B., Tillmann, S., Weitschies, W., Seidlitz, A., 2018. Immediate Release 3D-Printed Tablets Produced Via Fused Deposition Modeling of a Thermo-Sensitive Drug. Pharmaceutical Research 35, 124.

Landin, M., Rowe, R.C., 2013. Artificial neural networks technology to model, understand, and optimize drug formulations, Formulation Tools for Pharmaceutical Development, pp. 7-37.

Leuenberger, H., Leuenberger, M.N., 2016. Impact of the digital revolution on the future of pharmaceutical formulation science. European Journal of Pharmaceutical Sciences 87, 100-111.

Li, Z., Zhang, Z., Shi, J., Wu, D., 2019. Prediction of surface roughness in extrusion-based additive manufacturing with machine learning. Robotics and Computer-Integrated Manufacturing 57, 488-495.

Maroni, A., Melocchi, A., Parietti, F., Foppoli, A., Zema, L., Gazzaniga, A., 2017. 3D printed multi-compartment capsular devices for two-pulse oral drug delivery. Journal of Controlled Release 268, 10-18.

Nam, J., Jo, N., Kim, J.S., Lee, S.W., 2020. Development of a health monitoring and diagnosis framework for fused deposition modeling process based on a machine learning algorithm. Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture 234, 324-332.

Nasereddin, J.M., Wellner, N., Alhijjaj, M., Belton, P., Qi, S., 2018. Development of a Simple Mechanical Screening Method for Predicting the Feedability of a Pharmaceutical FDM 3D Printing Filament. Pharmaceutical Research 35, 151.

Nawi, N.M., Atomi, W.H., Rehman, M.Z., 2013. The Effect of Data Pre-processing on Optimized Training of Artificial Neural Networks. Procedia Technology 11, 32-39.

Nikolaev, P., Hooper, D., Webber, F., Rao, R., Decker, K., Krein, M., Poleski, J., Barto, R., Maruyama, B., 2016. Autonomy in materials research: a case study in carbon nanotube growth. npj Computational Materials 2, 16031.

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

Paulo, F., Santos, L., 2017. Design of experiments for microencapsulation applications: A review. Materials Science and Engineering: C 77, 1327-1340.

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. Journal of Machine Learning Research 12, 2825-2830.

Pereira, B.C., Isreb, A., Isreb, M., Forbes, R.T., Oga, E.F., Alhnan, M.A., 2020. Additive Manufacturing of a Point-of-Care "Polypill:" Fabrication of Concept Capsules of Complex Geometry with Bespoke Release against Cardiovascular Disease. Advanced Healthcare Materials, e2000236.

Popova, M., Isayev, O., Tropsha, A., 2018. Deep reinforcement learning for de novo drug design. Science Advances 4, eaap7885.

Qi, Z., Wang, B., Tian, Y., Zhang, P., 2016. When Ensemble Learning Meets Deep Learning: a New Deep Support Vector Machine for Classification. Knowledge-Based Systems 107, 54-60.

Rantanen, J., Khinast, J., 2015. The Future of Pharmaceutical Manufacturing Sciences. Journal of Pharmaceutical Sciences 104, 3612-3638.

Redkar, S., Mondal, S., Joseph, A., Hareesha, K.S., 2020. A Machine Learning Approach for Drug-target Interaction Prediction using Wrapper Feature Selection and Class Balancing. Molecular Informatics 39, 1900062.

Rowe, R.C., Sheskey, P., Quinn, M., 2009. Handbook of pharmaceutical excipients. Libros Digitales-Pharmaceutical Press.

Schmidt, J., Marques, M.R.G., Botti, S., Marques, M.A.L., 2019. Recent advances and applications of machine learning in solid-state materials science. npj Computational Materials 5, 83.

Schneider, P., Walters, W.P., Plowright, A.T., Sieroka, N., Listgarten, J., Goodnow, R.A., Fisher, J., Jansen, J.M., Duca, J.S., Rush, T.S., Zentgraf, M., Hill, J.E., Krutoholow, E., Kohler, M., Blaney, J., Funatsu, K., Luebkemann, C., Schneider, G., 2019. Rethinking drug design in the artificial intelligence era. Nature Reviews Drug Discovery.

Shen, H., Du, W., Sun, W., Xu, Y., Fu, J., 2020. Visual Detection of Surface Defects Based on Self-Feature Comparison in Robot 3-D Printing. Applied Sciences 10, 235.

Singh, B., Kapil, R., Nandi, M., Ahuja, N., 2011. Developing oral drug delivery systems using formulation by design: vital precepts, retrospect and prospects. Expert Opinion on Drug Delivery 8, 1341-1360.

Stommel, M., Herzog, O., 2009. Binarising SIFT-Descriptors to Reduce the Curse of Dimensionality in Histogram-Based Object Recognition, in: Ślęzak, D., Pal, S.K., Kang, B.-H., Gu, J., Kuroda, H., Kim, T.-h. (Eds.), Signal Processing, Image Processing and Pattern Recognition. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 320-327.

Sun, W., Lee, J., Zhang, S., Benyshek, C., Dokmeci, M.R., Khademhosseini, A., 2019. Engineering Precision Medicine. Advanced Science 6, 1801039.

Tiwari, R.V., Patil, H., Repka, M.A., 2016. Contribution of hot-melt extrusion technology to advance drug delivery in the 21st century. Expert Opinion on Drug Delivery 13, 451-464.

Trenfield, S.J., Awad, A., Goyanes, A., Gaisford, S., Basit, A.W., 2018. 3D Printing Pharmaceuticals: Drug Development to Frontline Care. Trends in Pharmacological Sciences 39, 440-451.

Vithani, K., Goyanes, A., Jannin, V., Basit, A.W., Gaisford, S., Boyd, B.J., 2019. 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, 102.

Wade, B.S.C., Joshi, S.H., Gutman, B.A., Thompson, P.M., 2017. Machine learning on high dimensional shape data from subcortical brain surfaces: A comparison of feature selection and classification methods. Pattern Recognition 63, 731-739.

Wainberg, M., Merico, D., Delong, A., Frey, B.J., 2018. Deep learning in biomedicine. Nature Biotechnology 36, 829-838.

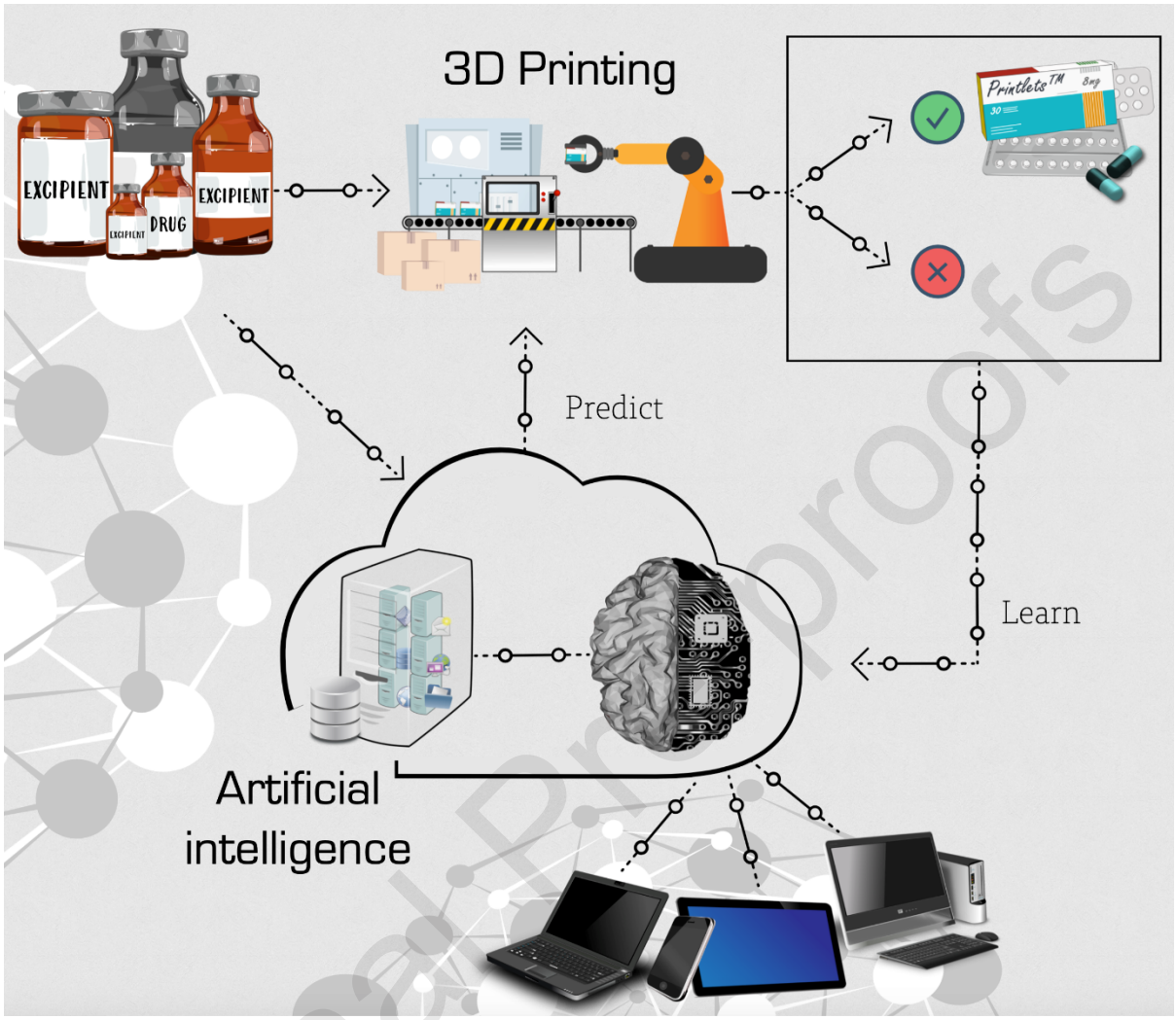
Xianyu, Y., Wang, Q., Chen, Y., 2018. Magnetic particles-enabled biosensors for point-of-care testing. TrAC Trends in Analytical Chemistry 106, 213-224.

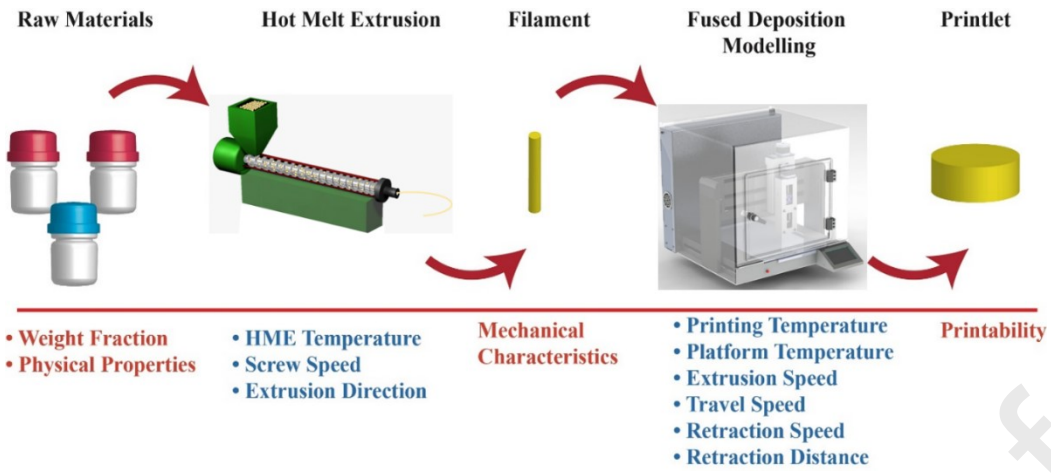
Xu, J., Yang, P., Xue, S., Sharma, B., Sanchez-Martin, M., Wang, F., Beaty, K.A., Dehan, E., Parikh, B., 2019. Translating cancer genomics into precision medicine with artificial intelligence: applications, challenges and future perspectives. Human Genetics 138, 109-124.

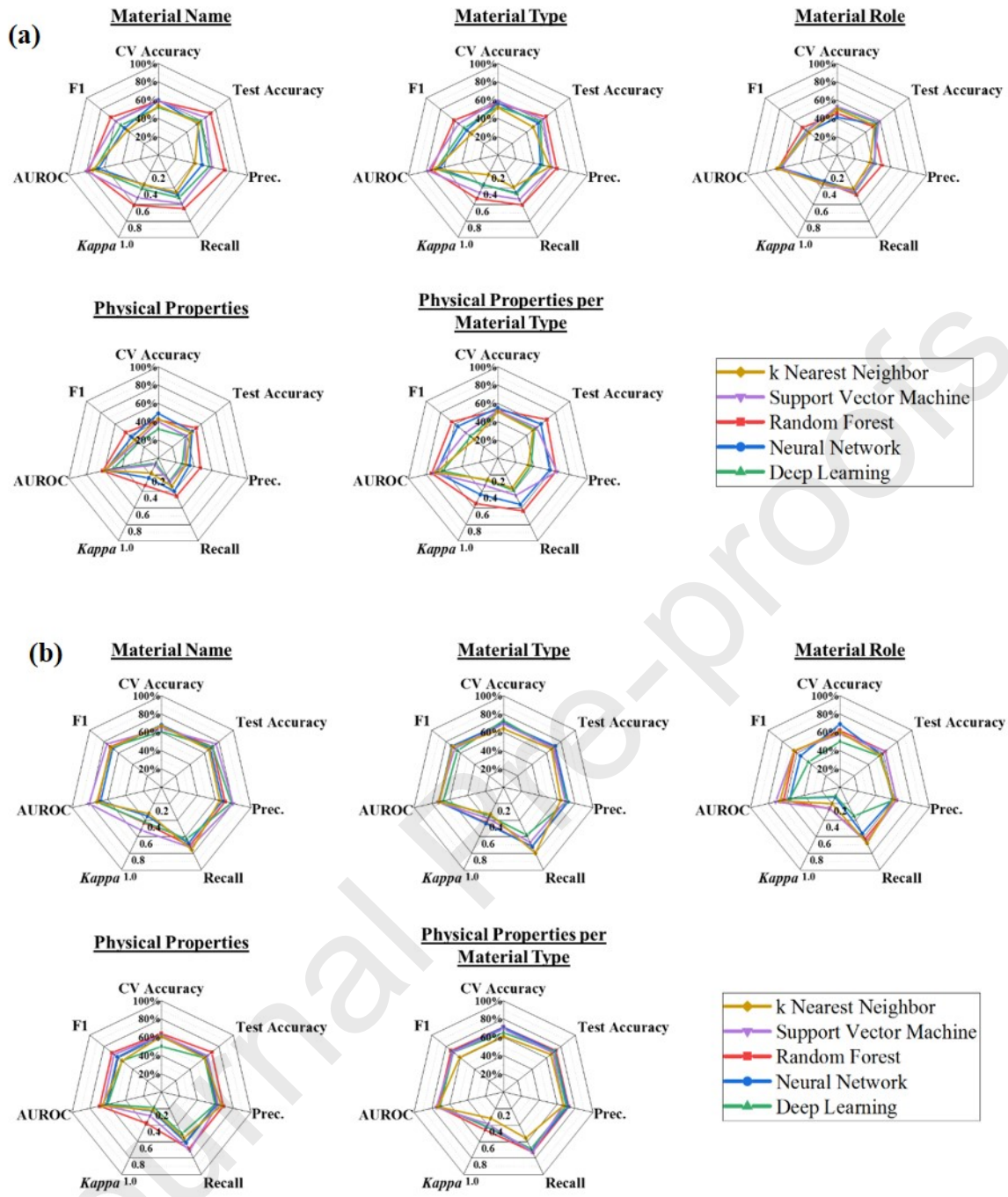
Xu, X., Robles-Martinez, P., Madla, C.M., Joubert, F., Goyanes, A., Basit, A.W., Gaisford, S., 2020. Stereolithography (SLA) 3D printing of an antihypertensive polyprintlet: Case study of an unexpected photopolymer-drug reaction. Additive Manufacturing 33, 101071.

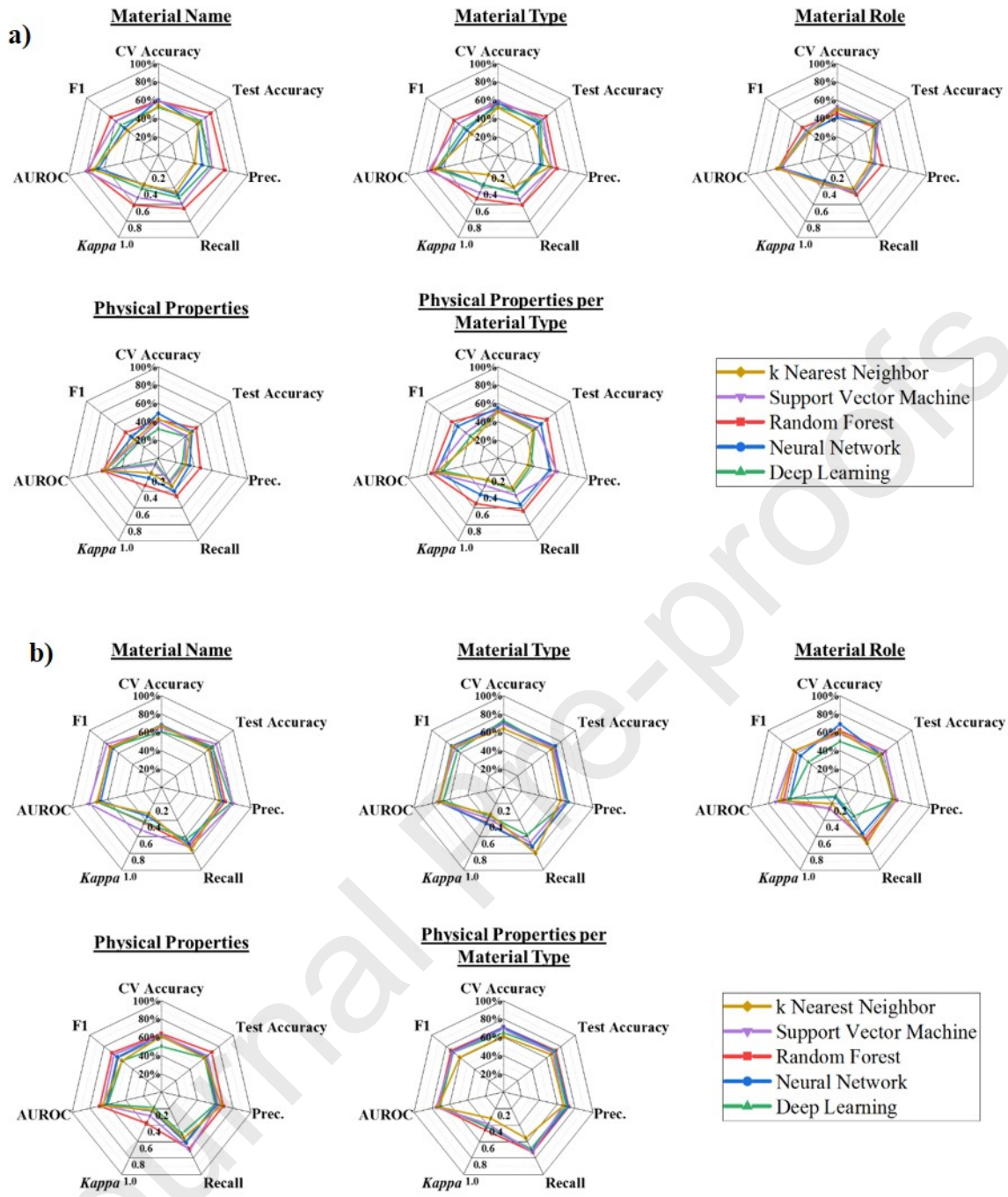
Zema, L., Melocchi, A., Maroni, A., Gazzaniga, A., 2017. Three-Dimensional Printing of Medicinal Products and the Challenge of Personalized Therapy. Journal of Pharmaceutical Sciences 106, 1697-1705.

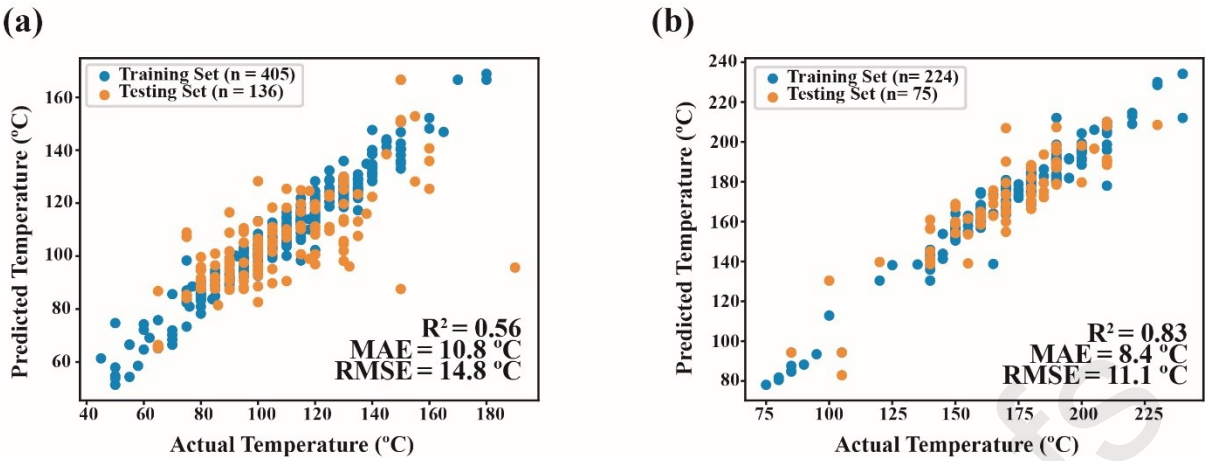
Zhang, J., Feng, X., Patil, H., Tiwari, R.V., Repka, M.A., 2017. Coupling 3D printing with hot-melt extrusion to produce controlled-release tablets. International Journal of Pharmaceutics 519, 186-197.

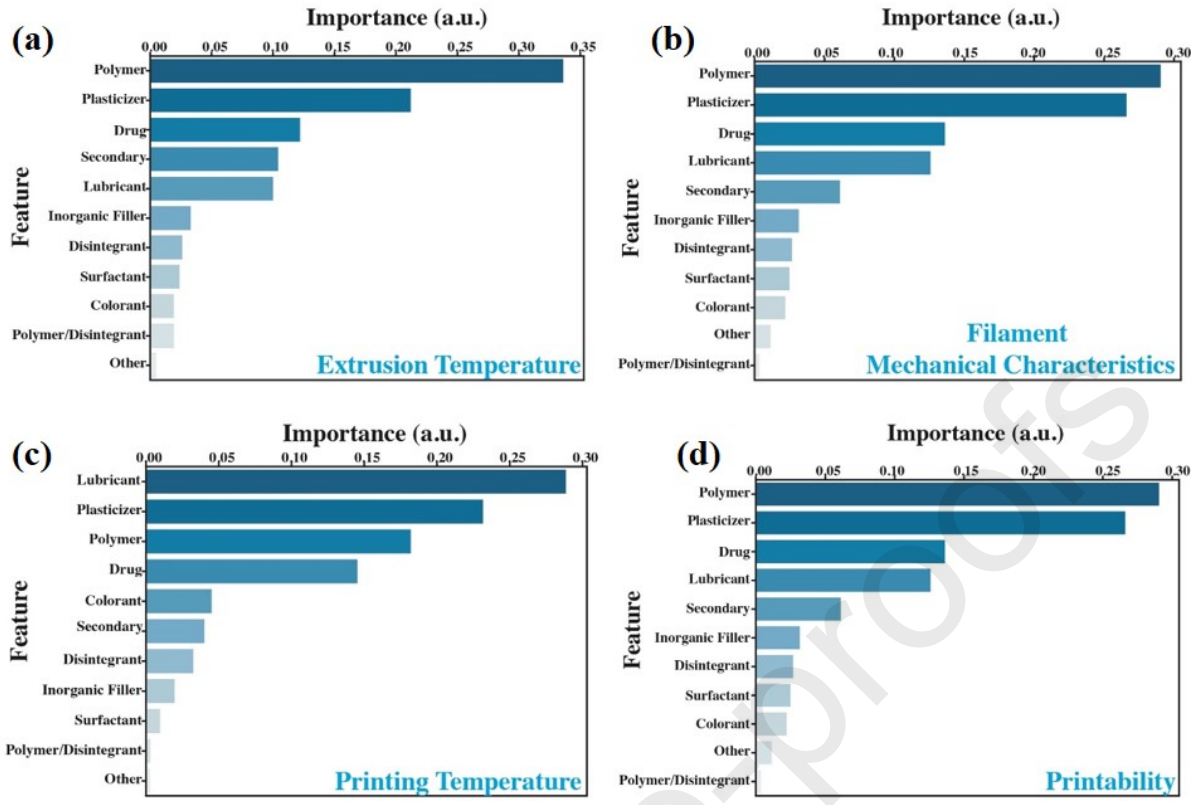












CRediT roles: Conceptualization; Data curation; Formal analysis; Funding acquisition; Investigation; Methodology; Project administration; Resources; Software; Supervision; Validation; Visualization; Roles/Writing - original draft; Writing - review & editing

Moe Elbadawi: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing - original draft; Writing - review & editing
Brais Muñiz Castro: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Roles/Writing - original draft; Writing - review & editing
Francesca K.H. Gavins: Investigation, Roles/Writing - original draft; Writing - review & editing
Jun Jie Ong: Data curation, Formal analysis, Roles/Writing - original draft; Writing - review & editing
Simon Gaisford: Supervision, Funding acquisition, Validation, Writing - review & editing
Gilberto Pérez: Supervision, Project administration, Validation, Resources, Writing - review & editing
Abdul W. Basit: Supervision, Project administration, Validation, Funding acquisition, Resources, Writing - review & editing
Pedro Cabalar: Supervision, Project administration, Funding acquisition, Resources, Validation, Writing - review & editing
Álvaro Goyanes: Conceptualization, Supervision, Validation, Project administration, Funding acquisition, Resources, Writing - review & editing

Declaration of interests

☒ 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.

☒The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: