Machine learning to explain printability induced by rheology additives

Ali Nadernezhad* and Jürgen Groll*

Chair for Functional Materials for Medicine and Dentistry at the Institute for Functional Materials and Biofabrication (IFB) and Bavarian Polymer Institute (BPI), University of Würzburg, Pleicherwall 2, 97070 Würzburg, Germany

* Correspondence to: ali.nadernezhad@fmz.uni-wuerzburg.de; juergen.groll@fmz.uni-wuerzburg.de

Keywords: machine learning, printability, rheology additives, explainable AI

Abstract

With the continuous growth of extrusion bioprinting techniques, ink formulations based on rheology modifiers are becoming increasingly popular, as they enable 3D printing of non-printable biologically-favored materials. However, benchmarking and characterization of such systems are inherently complicated due to the variety of rheology modifiers and differences in mechanisms of inducing printability. This study tries to explain induced printability in formulations by incorporating machine learning algorithms that describe the underlying basis for decision-making in classifying a printable formulation. For this purpose, a library of rheological data and printability scores for 180 different formulations of hyaluronic acid solutions with varying molecular weights and concentrations and three rheology modifiers were produced. A feature screening methodology was applied to collect and separate the impactful features, which consisted of physically interpretable and easily measurable properties of formulations. In the final step, all relevant features influencing the model's output were analyzed by advanced yet explainable statistical methods. The outcome provides a guideline for designing new formulations based on data-driven correlations from multiple systems.

1. Introduction

By introducing additive manufacturing (AM) technologies, particularly 3-dimensional (3D) bioprinting, to the field of tissue engineering (TE), a significant expansion in the scope and applicability of TE approaches was achieved.[1] The advancement of 3D bioprinting significantly depends on development in three critical frontiers, technological innovations,[2] the discovery of new functional biomaterials,[3] and deepening our understanding of regenerative biology.[4] While addressing all the requirements for a successful regenerative approach might seem out of reach for the moment, significant resources have been dedicated to approximating this process. In this respect, engineering biomaterial inks and bioinks includes a relatively large portion of research in the field, exclusively exploring the enabling possibilities through introducing new synthetic and natural biomaterials and formulations.[3]

Apart from meeting the strict biological requirements, the materials used in the 3D bioprinting approach, or in general terms, the 3D printing of soft biomaterials, need to fulfill some physical and mechanical criteria. The main class of materials used as the ink for this purpose includes hydrogels or polymer solutions. Traditionally, the natural biocompatible hydrogel inks failed to meet the 3D printing prerequisites, categorized primarily by lack of printability.[3] The phrase printability refers to the capability of the ink material in allowing the 3D printing process to form the designed structure with acceptable shape fidelity, mechanical stability, and structural integrity.[5] Despite the extent of these measures, the printability of a hydrogel ink or polymer solution is greatly influenced by its chemistry and mechanical properties.[3]

Due to the biological requirements, many attempts were made to enable or enhance the printability of promising known bioactive hydrogels and polymer solutions that formerly lacked physical and mechanical needs. These mainly included chemical modifications or blendings with a secondary material that could induce printability. The latter, generally referred to as formulations, is becoming increasingly popular for two main reasons. Firstly, several highly efficient additives are already available that meet the biological requirements and can significantly enhance the printability of the base hydrogel or polymer solution. Secondly, and more importantly, the low cost and the straightforward know-how of creating

new formulations constitute a significant advantage over developing sophisticated chemistries to induce comparable functionalities. Although this ease of processing does not replace the offerings of a sophisticated and application-tailored chemical modification, the literature shows an increasing trend in applying formulations in different domains related to the 3D printing of biomaterials.

Successful engineering of a new formulation for 3D printing needs a profound understanding of the material properties on micro and macro scales. However, in an interdisciplinary field of research such as biofabrication, there are tendencies in approximating the materials' related requirements, mostly towards more established biological measures. In this context, the performance of an additive is mainly weighted by the corresponding biological response rather than quantification and analysis of the material's properties. Nevertheless, the available literature shows increasing awareness and willingness of research groups to design the formulations based on traditional and state-of-the-art physical characterization methods.[3] These mainly included the rheological characterizations of inks and finding the correlations between the printing conditions and printability of the inks.

Analysis of such systems usually requires a degree of simplification, as increasing the number of parameters and variables could quickly deteriorate the interpretability of the readout by the conventional methods. A promising tool to overcome this limitation might be the data analysis techniques employing Machine Learning (ML) principles.[6] ML methods include statistical and mathematical tools which can reveal and exploit the relationships in data and deliver complex models to describe the system. Despite the visions and hopes for applying ML tools in the scope of 3D bioprinting,[7-9] there are a few reports on applying ML techniques to analyze correlations in 3D bioprinting of hydrogel inks,[10-13] which mainly focused on providing a metric on the predictability of the printability of the system, based on a few materials- and 3D printing process-related parameters.

A common approach in applied ML techniques is to provide metrics and models that enable a system's predictability based on some hidden correlations. By increasing the complexities of the models and their predictability power, the combinations and correlations of variables become more obscure, and interpretation of the process of making a particular decision by the model becomes inherently complicated. This problem gave rise to interpretable models mostly recognized as explainable artificial intelligence (AI), which provide a technically

equivalent but possibly more understandable approach than black-box models for data analysis and predictions.[14-16]

An interesting question in the context of 3D bioprinting is if the explainable AI could reveal and interpret the physical relationships in a practical problem such as inducing printability in a formulation. This question formed the rationale behind the current study, that whether it is possible to predict the printability of a formulation created by a rheology additive so that the process of making a decision by the model is understandable and interpretable by a user familiar with rheology and 3D printing. It should be noted that this study aims to provide insight into the complex correlations and relationships of different material-related properties that result in the creation of a printable formulation rather than providing a set of metrics for assessing the printability of different compositions. For this purpose, an extensive library of different formulations based on hyaluronic acid (HA) polymer solutions and three different rheology modifiers with distinct microstructural interactions were prepared. Different characteristics of these formulations were identified and labeled as the *features*; measurable properties shared between all the experimental conditions. Rheological data were the primary source of information. However, some combinations of printing conditions that were directly influenced by the material's properties were considered. After developing a predictive ML model with high precision, the main focus was to explain how the knowledge from the modeled data could interpret a particular model's decision towards the classification of a formulation.

2. Results and discussions

2.1. Production of data with minimized bias

Figure 1 shows an overview of the multiple steps taken in this study to identify and explain the contributing factors in enabling printability by adding rheology modifiers into a not-printable polymer solution. In the first stage, three different rheology modifiers with significantly different physical properties were selected to alter the rheology and printability of plain HA solutions. In addition to the interactions between HA molecules and the additives, each rheology modifier had unique interaction mechanisms to alter the viscoelasticity of final formulation: colloidal and granular interactions of Carbopol microgels, formation of a secondary network by electrostatic interactions between Laponite nanodiscs, and entanglement and network formation of one-dimensional self-assembled Fmoc-FF fibrils.

Several essential factors could potentially influence the extent of modification, notably the polymer-additive ratio and interactions, and the viscoelasticity of the initial polymer solution. This is apart from the fact that each additive, depending on the effectiveness of the reinforcing mechanism, might show a specific pattern in altering rheological properties. For example, colloidal-based additives could prominently induce yield behavior, while a secondary network formation might stabilize the frequency response over a broader range of frequencies. Showing such dependencies of the rheological behavior on the type of the additive might result in a strong bias throughout the analysis, potentially decreasing the precision and reliability of the outcome. For this purpose, the information regarding the type and content of the rheology modifiers in each formulation were anonymized. In this way, the outcome of rheology experiments only contained information on the viscoelasticity of the tested samples, which were linked with the printing performance of the same formulation in a later stage.

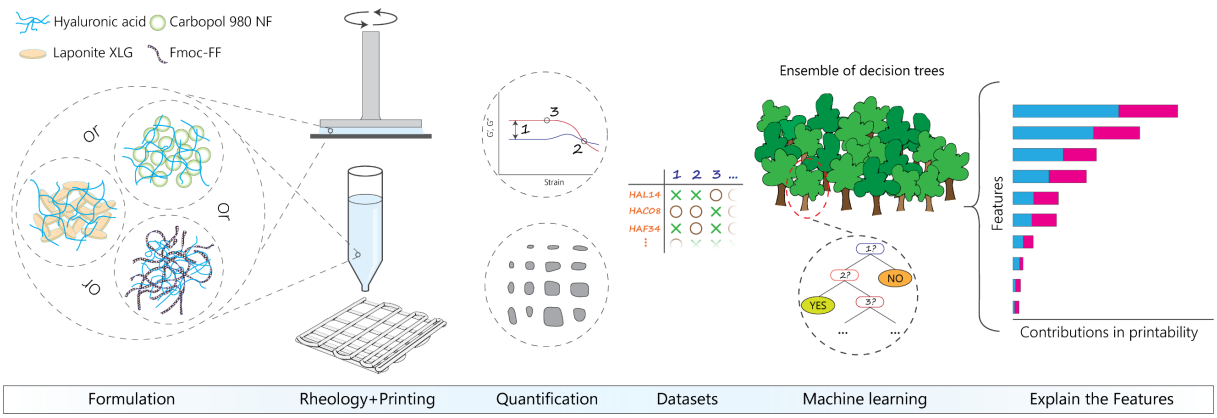


Figure 1. A demonstration of the multi-steps taken in this study to explain the printability. The flow of this study is depicted from left to right of the figure. First, the formulations based on different additives and starting HA solutions were prepared. At the second step, rheological characterizations and printing experiments were performed. The quantification step involved the extraction of different rheological features and quantitative analysis of the printability of formulations. In the next step, datasets based on a combination of all acquired features and printability scores were fabricated. By using the generated datasets, a random forest ML algorithm was trained. In the final step, a post-analysis of the obtained model revealed the correlations between data and the influence on making a decision by the model.

The multi-step rheological testing protocol was designed to acquire information on viscoelasticity and flow properties of samples holistically. The viscoelastic properties of formulations were directly extracted from the frequency and amplitude sweeps, including the values of storage modulus, damping factor, and the observed crossovers. A previous study on collagen derivatives showed that the elasticity of viscoelastic hydrogel ink during deformation could significantly influence the prediction of printability.[13] The 3-Interval Thixotropy and Oscillatory Tests (3ITT and 3IOT) returned the time dependant percentage of recovered viscosity (in 3ITT) and storage modulus (in 3IOT) of the formulations after a significant shear deformation. These two methods could serve as alternatives to more sophisticated creep and stress relaxation studies, as they can provide a rough approximation of the recovery of microstructural interactions of the material after deformation. This is especially important in a process like extrusion 3D printing, as the hydrogel ink experiences a significant amount of deformation during extrusion, and the ability to retain the shape afterward could be correlated with its recoverability. The apparent yield behavior of formulations was identified by whether a peak in the viscosity-shear stress plot was observed, and if so, the corresponding value was recorded. Similarly, estimation of printability of collagen-based bioinks showed significant dependency on the value of yield stress.[13] However, since the screening tests (data not shown) showed difficulties in acquiring yield stress in a hybrid system containing different proportions of additives, it was instead decided to capture the occurrence of a peak in viscosity during the sweep and record the corresponding value. From the physical point of view, this peak and its value corresponds to the buildup of resistance against the flow on a macroscopic scale. In the last step, the viscosity of formulations subjected to increasing shear rates was fitted by the Carreau-Yasuda model, and the values of zero-shear viscosity and the consistency index of the model were extracted. In particular, the consistency index (λ) was chosen to provide an empirical estimate of the shear rate at which a transition from Newtonian to power-law behavior was evident (λ^{-1}).

The testing protocol was aimed to maximize the obtained information with minimum complexities. Especially, creep-related tests were avoided since the preliminary experiments (data not included) for optimization of protocol showed the sensitivity of creep tests towards not-printable formulations, potentially resulting in a bias in the analysis. Moreover, sophisticated methods such as generation of the relaxation spectra and studying the

advanced rheological features of the samples were avoided since these methods generally depend on sophisticated post-processing steps by a knowledgeable user.

The printing experiments were conducted by considering that no HA solution was printable by itself. Even at high molecular weights and high concentrations, HA solutions could not retain the shape after extrusion, and the shape fidelity due to the spreading of the extrudate was quickly diminished. Moreover, the printing experiments were performed considering the volumetric flow of each formulation during printing. A recent study by Fisch et al.[17] demonstrated the sensitivity and susceptibility of pneumatic driven extrusion systems to over-or under-extrusion if the volumetric flow and the cartesian translation feedrate are mismatched. For this reason, each formulation was printed at a unique combination of applied pressure and feedrate; the latter was derived automatically based on the extruded mass of the formulation prior to printing for a given time. Two features of the printing process, the ability to form a filament and the proportionality of the volumetric flow to printing pressure, were recorded as the characteristics of the printing process per formulation.

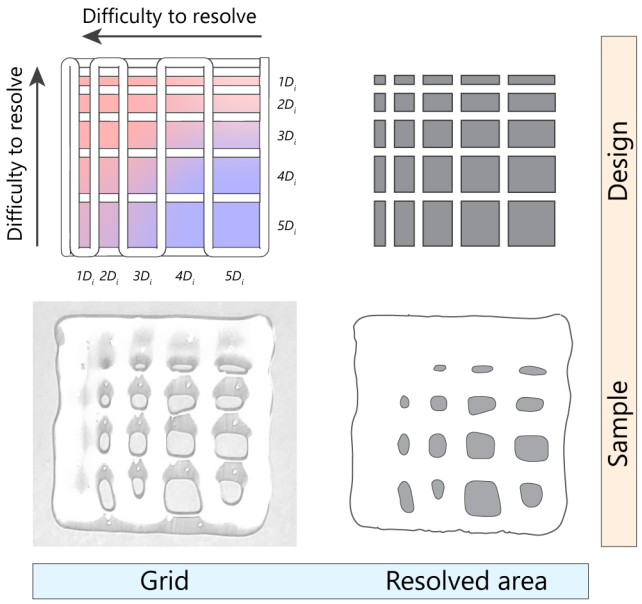


Figure 2. A schematic and experimental representation of the metric used in this study to quantify printability. Top-left) a 2-layer grid design with varying fiber spacing was used to assess the printability of the formulations. The fiber spacing was increased by increments of inner nozzle diameter (D_i). Bottom-left) an image of an actual sample printed according to the grid design. Top-right) The resolved area of design is used for benchmarking printability.

Bottom-right) The segmented image of a printed formulation is used to calculate the percentage of resolved area.

The designed path for the printability assessment included a 2-layered rectangular mesh with a directional increase in fiber spacing based on increments of the inner diameter of the nozzle (D_i) (**Figure 2**). The printing pattern with changing 2-dimensional (2D) spacing allowed rigorous benchmarking of the printability of formulations. The 3D assessments of printability were deliberately avoided since criteria such as general 3D shape fidelity and fiber sagging supposably demand a certain extent of viscoelasticity, which could interfere with the objectives of this study in the unbiased evaluation of induced printability by rheology additives. Based on our preliminary screening experiments, a weighting approach to penalizing the easy to resolve areas was employed (Supporting Information, Figure S1). The weighted quantified areas of each print were normalized to the designed pore area to produce the printability index of each formulation. In general, the induced printability in different formulations was not significant, as only about 14% of the formulations could resolve more than 33% of the designed area. We speculate that several factors contributed to such behavior. Among them, the concentration restrictions (maximum concentration of additives was 2.5 wt./v% in a 1:1 additive:polymer ratio) imposed by the experimental design and the wide range of physical properties of initial polymer solutions (resulted by variation of concentration and molecular weight) would play the critical role.

The printing parameters could significantly influence printability.[18] This influence is more evident in resolving geometrical features for which an abrupt change in the printing process, such as a change in direction, is expected.[17] In addition, other factors such as extrusion rate and substrate interactions potentially influence the spreading and fidelity of extruded filaments. As described earlier, the printing speed adjustment based on the volumetric flow was employed to eliminate the factor of over-and under-extrusion, while the same substrate was used to print all the formulations. However, to avoid complexities caused by pressure-dependent extrusion delay, no changes in printing speed at turning points were implemented. This phenomenon resulted in the underscoring of weakly printable formulations, as revealed by the review of the raw quantitative analysis of individual printing experiments (data not included). However, in an equally conditioned set of experiments, this

could be considered a result of the lack of meeting the viscoelasticity requirements for extrusion printing.

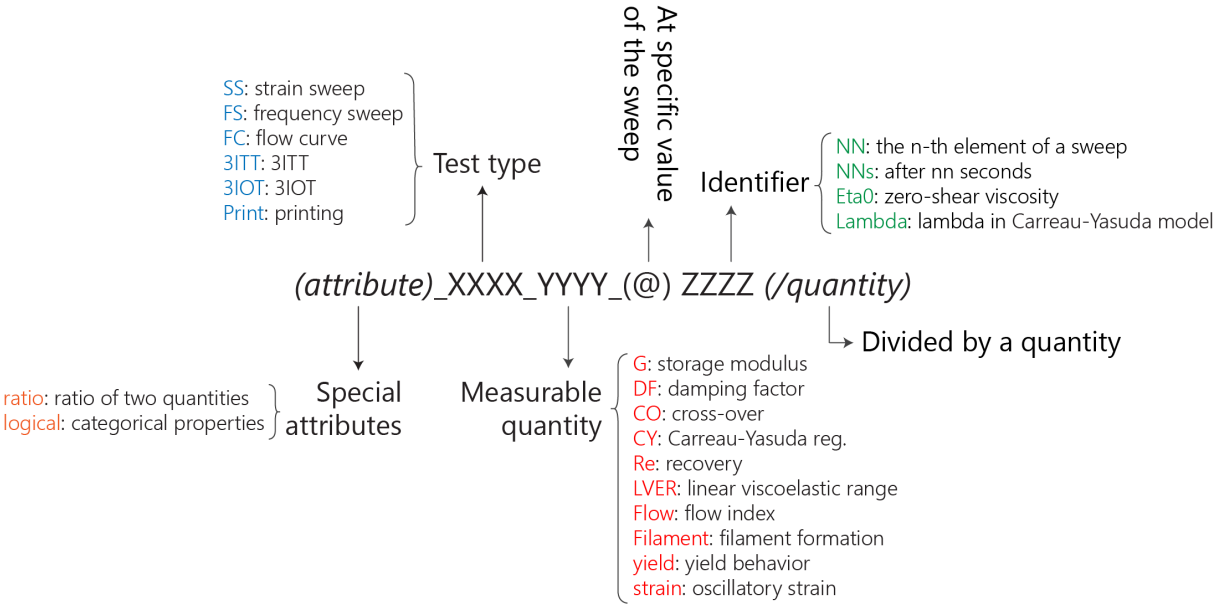
After acquiring the rheological and printing data, the generation of the dataset for further analysis was performed by considering two feature types. Concretely, some aspects of the raw data from rheological experiments were used as the *primary* features. These included, but were not limited to, the storage modulus and damping factor in amplitude and frequency sweeps, the crossover points, the filament-forming ability, or the results of regression analysis for recovery and flow experiments. In order to include the additive nature of the rheology modifiers in the analysis, the majority of the extracted primary features of formulations were individually divided by the corresponding values of the respective HA solutions of different concentrations and molecular weights. In this way, instead of an absolute value of a given data point for a formulation, the proportion of change to the respective HA solution will be analyzed. The details of the complete list of features are provided in Supporting Information (Table S2).

The second type included the *combinatory* features. The combinatory features were generated based on combining some aspects of the data collected during experiments. A few examples of such features include the proportionality of the volumetric flow to printing pressure, the ratio between damping factors at the two extremes of frequency sweeps, or the ratio between the storage modulus of high and low ends of the sweep regions.

As a principle, all the primary and combinatory features were selected based on physically explainable factors rather than abstract quantities. This would guarantee that this study's outcome is physically interpretable and can shed light on the complexities of printability induced by the addition of rheology modifiers.

2.2. Selection of relevant features influencing printability

The data from rheological and printability experiments of all the formulations were prepared and consolidated into a single structured data frame. An intuitive nomenclature system to identify the different features of the data was devised. The complete list of nomenclature used to address different elements in the dataset is provided in Supporting Information (Table S2). To increase the legibility of the manuscript, **Scheme 1** provides the guideline for interpretation of the coded nomenclatures.



Scheme 1. Reference for the nomenclature used to address different features in this study.

If evaluated by an experienced user familiar with rheology and extrusion printing, some correlations, patterns, and trends in data might be recognized. Analytically, this was demonstrated by performing Spearman's correlation analysis on the raw data (Supporting Information, Figure S2). In practice, rheological data obtained from printable formulations might show some patterns in different aspects. For example, as revealed by the correlation analysis, a monotonic correlation between the values of damping factor and elastic modulus is expected, as increased elasticity of the formulation usually corresponds to a decreased loss of energy during deformation, or a slow initiation of recovery in a formulation subjected to 3ITT tests is followed by a slow continuation. Considering these trends, a refinement and selection of features are necessary. This refinement could potentially increase the model's usefulness and interpretability by including only relevant information, while the possibility of overfitting during the training would be minimized. This process results in collections of features that directly and significantly influence the printability of the formulation, and at the same time, they can be explained by the physics of the problem.

2.2.1. All relevant features to explain printability

Implementing the statistically robust algorithm of Boruta feature selection[19] refined the features to those that significantly contributed to printability prediction with a high F-score of the RF model. Boruta algorithm identifies the importance of the features for constructing

the ML model based on the performance of a randomized version of the features through many iterations. Eventually, the statistically impactful features are selected from the top 0.5% of the binomial distribution of the iterations. This procedure resulted that, among the initial 65 features, only 15 were identified as having a significant influence on the predictability of the RF model. To explore and understand the existing correlations between these essential features, the Spearman rank-order correlation matrix and the linkage based on hierarchical cluster analysis of Spearman's correlations are demonstrated in **Figure 3**. This analysis shows three main clusters of features (as demonstrated by different colors of the dendrogram) with distinct linkage distances from each other.

To physically interpret the results, these three clusters included features which: 1) describe the degree of plasticity of the flow (orange leaves in Figure 3), 2) describe the ease of inducing flow and its homogeneity (green leaves in Figure 3), and 3) describe the viscoelasticity of the formulation prior to and after the flow (red leaves in Figure 3).

The first cluster included the ratio between storage modulus at the flow point and the limit of the linear viscoelastic range, and the proportionality index of the flow during printing. The former describes the plasticity of the formulation prior to the flow, while the latter is a measure of how the formulation's flow can be maintained during the extrusion through a fine nozzle. Although obtained from two different methods, the information from these two features complements each other, shown by their correlation.

The second cluster also combines the information gathered from rheological experiments and printing, correlating the quality of the flow, indexed as the ability to form filament after extrusion, with the flow characteristics of the formulations determined by the Carreau-Yasuda model. The analysis showed that the yield viscosity is closely correlated with whether or not a filament would be formed during the extrusion, both to a lower extent, with the transition from Newtonian to non-Newtonian flow during shear deformation. The latter is quantified by the consistency index Lambda in the Carreau-Yasuda model.

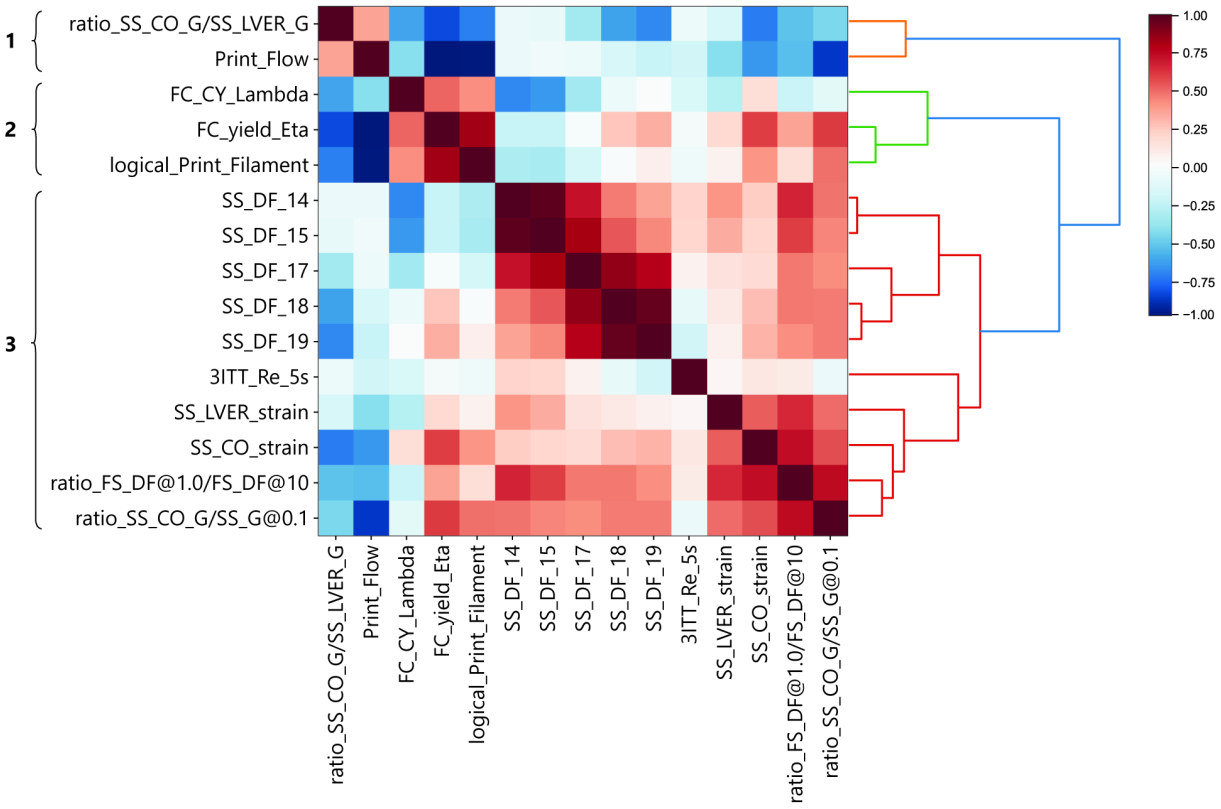


Figure 3. The Spearman rank-order correlation matrix and the linkage according to hierarchical cluster analysis of Spearman's correlations shown by the dendrogram. Numbers on the left identified three clusters. Value of 1 or -1 in the correlation matrix denotes high correlations, and the direction of association depends on the sign of the correlation coefficient.

The third cluster represents more complex correlations between the viscoelastic properties of the formulations. On the one hand, the closely linked damping factors at moderate shear strains (14.8 and 21.7%) are, with a considerable linkage distance, correlated with the values of the damping factors at high shear strains (46.6, 68.5, and 100%). This indicates that while the extent of the viscous portion of deformation plays a significant role in predicting printability, the flow pattern by itself is also of great importance. On the other hand, five different aspects of the viscoelasticity of the formulations are correlated with each other, including the energy required to induce the flow of formulation with consideration of the elasticity of the starting HA solution (ratio between storage modulus at low strain and flow point multiplied by the storage modulus of corresponding HA at 0.1% strain), the extent of stability of the interactions by changing the frequency of deformation, the required strain to

induce the flow, the extent of linear viscoelasticity, and the rate of recovery of interactions after the flow of the formulations.

Although the hierarchical clustering analysis reveals the correlations between different features, it does not provide information on how these features interact and contribute to predicting the printability of the formulations. As mentioned earlier, it should be noted that the purpose of this study is not to provide a metric for the prediction of printability based on specific thresholds of the features; rather, the subject of the forthcoming discussions is to understand the contributions and interaction of features, which induced printability.

2.3 Explanation of induced printability

Despite identifying the important features that contributed to printability prediction, the extent of the contributions is unknown. This is a typical characteristic of models generated with most ML methods, as interpreting the predictive model's output is a tedious task, especially by increasing the complexities in non-linear models. Several measures and methods have been developed in recent years to enhance the interpretability of the ML outcome. The details and review of such methods are out of the scope of this study. However, in the context of this study, it is crucial to understand how interpreting the decision-making process of the ML model could result in understanding the contribution of different factors in inducing the printability in a polymer solution ink.

The ML model identifies printable formulations in a pool of many observations based on the prior knowledge of correlations and interactions gained during training. Apart from metrics to describe the accuracy of such a model, the conditions that resulted in recognizing the printable formulations with high accuracy are the key subjects in explaining how a new formulation with a given set of attributes would be classified. Analogically, the rationale behind making a particular prediction of a model is comparable with tunning some principal and dependant physical properties of the base polymer solution by adding additives that would enhance or worsen printability from an experimental point of view. In this scope, prior knowledge of the full extent of the modification is not necessarily available, although some understandings about simple correlations were obtained through training.

Shapley additive explanations (SHAP) is a powerful tool for interpreting the prediction by ML models.[16] SHAP has a solid theoretical foundation in game theory and can provide contrastive explanations and analyze the model's output locally and globally. The extent of

the contribution of different features in the prediction of printability of formulations is demonstrated in **Figure 4**A. The average SHAP value quantifies the impact of each feature on the model output by interpreting the average expected contribution of the feature after all the possible combinations of other features are considered.

The SHAP feature importance plot shows that the most important features in the global scale for printability prediction included the yield viscosity, ratio of the storage modulus at flow point to that of the limit of the viscoelastic range, and the consistency index in the Carreau-Yasuda model. This is an interesting finding since it shows that the combination of resistance to flow, the plasticity of the formulation prior to flow, and the transition state from Newtonian to non-Newtonian flow critically determine the model's output. Although the impact of the other features on the model's output is significantly less than the main three features, their contributions to the model's accuracy can not be disregarded since the SHAP analysis was performed on all the relevant features selected by the Boruta algorithm.

The average SHAP value provides information about the contribution of the features on a global scale. However, on a local scale, evaluation of individual observations and the corresponding SHAP value demonstrates how each feature contributed to predicting a printable formulation experimentally (Figure 4B). Careful analysis of Figure 4B demonstrates that different features contributed in different directions to enabling printability. For instance, a high yield viscosity of the formulation resulted in a higher printability score, while a low ratio between the storage modulus at the flow point and that of the limit of viscoelastic range resulted in a printable formulation. On the contrary, moderate to high values of the consistency index in the Carreau-Yasuda model correspond to printable formulations, translating to transition to a shear thinning behavior at moderate to low shear rate values. Additionally, formulations with a lower flow proportionality index were statistically more susceptible to be identified as printable, meaning that either a lesser extrudate mass at constant pressure or a larger force to extrude the same mass of the ink determines a printable formulation. This feature directly corresponds to the microstructural interactions of the formulation; the stronger the interaction, the better resistance to deformation and better printability.

The formulations that resulted in filament formation during printing had a higher probability of being recognized as printable; especially, the formation of broken and fragmented

filaments did not positively contribute to printability prediction. A higher ratio between the damping factor at low and high frequencies of deformation contributed more to the printability of the formulation. From the physical point of view, this meant that lesser variation of the damping factor in the frequency range results in a higher chance of printability. In other words, more stability of the interactions at the moderate to the low portion of the frequency spectrum results in better printability. Similarly, a higher ratio of storage modulus at the flow point to the one at 0.1% strain (proportional to the starting HA's storage modulus) resulted in a higher probability of printability, meaning that the higher elasticity of starting HA solution could contribute to more printability of the formulation. The higher flow strain also resulted in higher printability of the formulations, and similarly, formulations with a moderate to high level of the limit of linear viscoelastic range showed more printability.

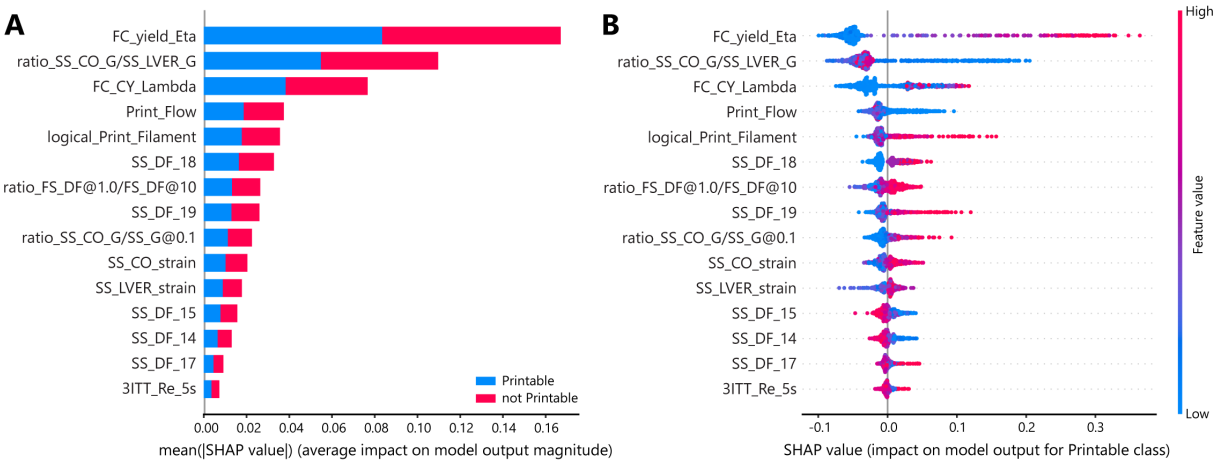


Figure 4. SHAP values of different features show their contributions to the model output in two scales. A) on a global scale, the mean SHAP value represents the feature's average impact on the predictions made by the model. B) on a local scale, the rank-ordered features explain the margin output of the model, which is the change in printability of formulations. The plot also shows the range of influence over the dataset. The color shows how the change in the value of a feature affects the change in the prediction of printability.

In contrast to the monotonic behavior of the features mentioned so far, the impact of the ratio between the viscous and elastic characteristics of the formulations on the printability strongly depended on the state of the deformation. The formulations with viscous dominant behavior at high shear strain (68.5 and 100%) showed more printability, meaning they

already started to flow at such strain values. On the other hand, the dominant elastic behavior at lower shear strains (14.8 and 21.7%) contributed positively to printability, while a mixed impact of the value of damping factor at 46.6% strain on printability was observed. In other words, the dominancy of elastic behavior at low strain values and prominent viscous behavior at high strain values contributed to the formulations' printability.

The information obtained from the SHAP values for each observation reveals many dependencies of printability on individual factors. Nevertheless, a significant amount of mutual contributions of different factors could occur in a multi-variable system. Figure 5 shows some notable dependencies observed between all the combinations of the features. Every dot in plots of Figure 5 corresponds to one observation, and a SHAP value above zero meant a positive contribution of the feature in that observation towards being classified as a printable formulation. In contrast, a value below zero corresponded to a negative influence on the outcome of the observation.

The combination of increasing the elasticity of starting HA solution and decreasing the plasticity of formulation prior flow led to a positive impact on the model's prediction, if the flow's proportionality index during printing was low (Figure 5A). Additionally, the flow's proportionality index positively impacted the predictability when extruding the formulation formed a filament at relatively high pressures (Figure 5B). The damping factor at the high end of the amplitude sweep profile of the formulations started to significantly impact the model's output when it was larger than 1 (a viscous behavior of the formulation) and the plasticity prior to the flow was small (Figure 5C). From a physical point of view, this could be interpreted as a printable formulation had probably started to flow with minimized force at lower strain values. However, the formulation with a low damping factor at low strain values positively impacted the model's predictability, meaning that a high degree of elasticity prior to flow critically impacts the prediction of printability of the formulation (Figure 5D).

The analysis showed that the observations, in which higher stability in microstructural interactions was combined with a higher elasticity imposed by the starting HA's solution, had a more positive impact on the model's predictability towards identifying printability (Figure 5E). Ultimately, the observations in which the degree of plasticity prior to flow was low significantly impacted the model's predictability, while they mostly showed high values of yield viscosity (Figure 5F).

These were a few examples of the many combinations of different properties that could enhance or worsen the model's predictability power. The key aspect in explaining the printability induced in a formulation is to consider the many contributing factors which could individually play a minor or major role in imposing the desired effect. Figure 5G-I show three different example observations in which the mutual contributions of different features respectively resulted in a minor negative, significantly positive, and significantly negative offset in the final prediction of the model ($f(x)$) compared with a random guess ($E(f(x))$, the baseline value of the model).

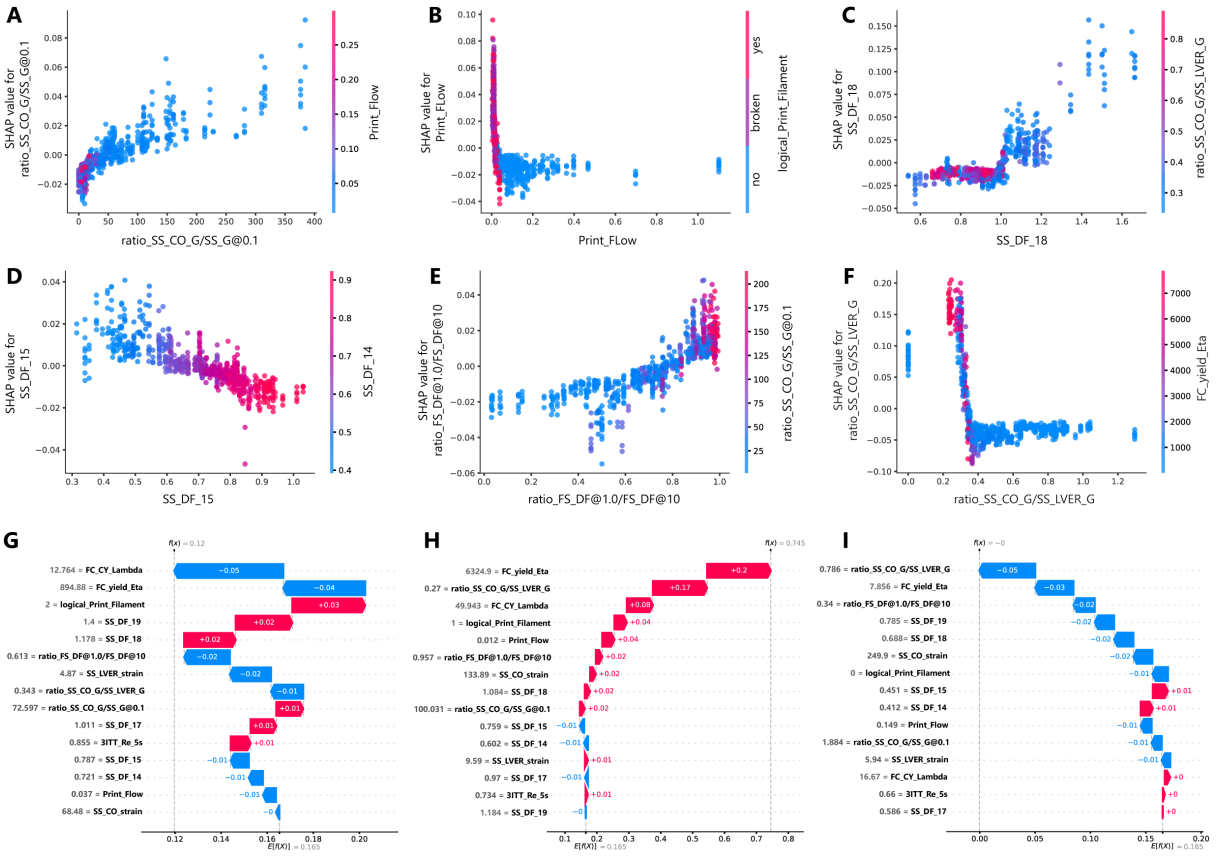


Figure 5. SHAP dependence plots of some notable combinations of features to show how the estimation of printability varies by feature value. These include the combinations of A) HA-dependent plasticity of formulation prior flow and the flow proportionality index, B) flow proportionality index and filament formation ability, C) damping factor at high strain, and the plasticity of formulation prior to the flow, D) damping factor at low ends of the strain sweep, E) stability of frequency response and the HA-dependent plasticity of formulation prior to flow, and F) plasticity of formulation prior to the flow and yield viscosity. Waterfall

plots showing the explanations for individual predictions where positive (red) and negative(blue) contributions of each feature offset the model output (dashed vertical lines) by a G) minor negative, H) significantly positive, and I) significantly negative value from a random guess by the model (the baseline value of the model, $E(f(x))$. Grey numbers indicate the actual value of the features per observation.

Printability of a formulation is a result of the complex interactions between different features of additives and the base ink, together with the requirements of the process. The formulations investigated in this study involved a range of characteristic behaviors and included three rheology additives with essentially different governing physics. From a general point of view, the mechanisms driving the reinforcement and variation of viscoelasticity in these formulations could vary. Additionally, some factors such as chemistry and affinity of the additives and base polymer solutions could significantly change the mechanisms of interactions. These fundamental aspects were not considered in this study, since as demonstrated, the multitude of influencing factors could easily complicate the process of engineering a formulation ink. Nevertheless, some rules of thumb or general guidelines could be identified, especially if we consider that the features investigated in this study were the observed effects caused by the material-dependent properties.

We speculate that these inherent complexities resulted that the few available studies on the application of ML in the 3D (bio)printing domain focused on either investigating a group of inks with similar characteristics or a set of process parameters with minimum alterations in ink properties.[10-13] The rationale behind limiting the factors is very reasonable, considering the interpretability of the outcome and the required resources. However, it should be noted that the results of such studies, including the current one, should be interpreted cautiously, as introducing new elements to the system might form new correlations and contributions, which could potentially influence the outcome. The current study's findings are almost in the same direction as the previous literature. However, a curious reader is courageously welcomed to challenge and explore the different dimensions of similar systems, as the results of these studies might not replace an expert's knowledge of the data in a particular problem.

3. Conclusions

The current study attempted to investigate and analyze the complex process of inducing printability in a non-printable polymer solution by adding rheology modifiers. By applying a classification machine learning algorithm, it was possible to build a model to predict the printability of formulations based on three different rheology modifiers in a matrix of HA with varying concentrations and molecular weights. Instead of providing a metric, this study's focus was to establish a correlation between different rheological aspects of the formulation and their flow behavior with the observed trends in printability. Advanced data analysis tools were employed to unveil the correlations between the essential properties of formulations.

In summary, the obtained model predicts that from a statistical point of view, a formulation becomes printable when it starts to show a high yield viscosity and a low degree of plasticity prior to flow, while the transition from Newtonian to non-Newtonian behavior of the flow occurs at relatively low shear rates. The formulation tends to flow at higher pressures during printing, and extruding through a small nozzle forms a filament rather than a droplet. The formulation based on polymer solutions with higher elasticity tends to be more printable, and a higher degree of stability of microstructural interaction over a range of frequencies is favored. While generally, a more extended range of linear viscoelasticity is desired, the damping factor of the formulations at low and high strain values should follow a pattern, as a formulation with more elasticity at the lower range and high viscous nature at the higher range is more desired. The kinetics of the recovery seems to have no directional influence on the prediction of printability, and a case-specific study approach could be of more usefulness.

Rather than providing a metric for printability, the objective of this study was to establish correlations between the multitude of different viscoelastic and printing properties of formulations. While not intended to construct conclusive metrics to predict printability, this study attempted to draw a more generic map for designing new formulations based on data-driven correlations from inherently different systems.

4. Experimental section

Materials

Hyaluronic acid sodium salt with three different molecular weights was purchased from Carbosynth (Mw 0.6-1.0 MDa, 1.0-2.0 MDa, and 2.0-2.5 MDa; Biosynth Carbosynth, Compton, UK). Carbopol 980 NF was purchased from Lubrizol (Lubrizol Pharmaceuticals, OH, USA). Laponite XLG was purchased from BYK Additives (BYK-Chemie GmbH, Germany). N-fluorenylmethoxycarbonyl diphenylalanine (Fmoc-FF) was purchased from Bachem (Bachem, Switzerland). Dimethylsulfoxide (DMSO) was purchased from Sigma-Aldrich (Sigma-Aldrich, USA). Sodium hydroxide was purchased from Merck (Merck KGaA, Germany).

Formulations

HA-Carbopol (HAC), HA-Lapointe XLG (HAL), and HA-Fmoc-FF (HAF) formulations were prepared with different concentrations and starting HA molecular weights by thoroughly mixing the required amounts of HA and the respective additive's stock solutions, as listed in **Table 1**. All the formulations were incubated at 4 °C for 24h after mixing. One hour of equilibrium time at room temperature was administered before each measurement.

For HAC formulations, the stock solution of Carbopole at 30 $mg.mL^{-1}$ was prepared in MilliQ water, and the pH of the solution was neutralized with dropwise addition of 10N sodium hydroxide solution. The 100$mg.mL^{-1}$ stock solution of Fmoc-FF in anhydrous DMSO was used to prepare HAF formulations. The Fmoc-FF stock solutions were prepared freshly. To prepare HAL formulations, 55 $mg.mL^{-1}$ stock solutions of Laponite-XLG in MilliQ water were prepared. Stock solutions of HA with different concentrations and molecular weights in MilliQ water were prepared by vigorous shaking at 250 rpm at 40 °C overnight, using New Brunswick Innova 40 incubator shaker (Eppendorf, Germany).

Table 1. Different formulations based on additives used in this study

Formulation	**HA: Additive**	**Total concentration ($mg.mL^{-1}$)**	**HA Mw (MDa)**
HAC	10:10…10:01	15	0.6-1.0
			1.0-2.0

			2.0-2.5
		30	0.6-1.0
			1.0-2.0
			2.0-2.5
HAL	10:10…10:01	15	0.6-1.0
			1.0-2.0
			2.0-2.5
		30	0.6-1.0
			1.0-2.0
			2.0-2.5
		45	1.0-2.0
			2.0-2.5
		50	0.6-1.0
HAF	10:10…10:01	5	0.6-1.0
			1.0-2.0
			2.0-2.5

Rheology

An Anton Paar MCR702 rheometer with a 25 mm parallel plate geometry at 500 µm gap was used to analyze the formulations. A general protocol for rheological measurements was designed and followed for each experiment. The protocol included the following steps: 1) homogenizing the sample by constant rotation at 1.0 s^{-1} for 60 s 2) frequency sweep between 0.1-100 rad.s^{-1} at 0.1% strain 3) amplitude sweep at 10 rad.s^{-1} in logarithmic scale between 0.01-500% strain 4) 3-Interval Thixotropy Test (3ITT) at 1.0, 100, and 1.0 s^{-1} shear rates. The recovery viscosity was calculated as the percentage of the rest viscosity at 5, 10, and 30 seconds 5) 3-Interval Oscillatory Test (3IOT) at 10 rad.s^{-1} with 0.5, 50, and 0.5% strain. The recovery storage modulus was calculated as the percentage of the rest storage modulus at 5, 10, and 30 seconds. 6) shear stress sweep in rotation in linear scale from 1.0 to 100 Pa with 0.5 Pa increments 7) transient shear steps with shear rates in logarithmic scale from 0.1 to 100 s^{-1}, using a dynamic data acquisition method. The viscosity at each discrete shear rate value was monitored every 100 ms, and the corresponding viscosity value was reported if a 0.5%

tolerance threshold over 10 observed values was met. A viscosity vs. shear rate curve was generated using the acquired data, followed by fitting with the Carreau-Yasuda model.

A rest period between each measurement step was implemented to recover the sample after deformation. Samples were gently loaded from standard 5 mL syringes on the lower plate without any needles attached for each experiment.

In order to benchmark the change in rheological properties induced by the additives, the equivalent HA solution of each formulation with the relevant concentration and molecular weight was tested with the same general protocol.

Printability assessment

The induced printability of each formulation was quantified based on the ability to resolve a 2D mesh patterns with varying intra-fiber distances (Figure 2). Printing was performed using RegenHU Discovery bioprinter (RegenHU, Switzerland). The hydrogels were loaded in 3 mL Luer lock syringe barrels (Nordson EFD, USA), equipped with a blunt G22 general-purpose dispensing tip (Nordson EFD, USA). For each formulation, the minimum air pressure of the pneumatic dispensing unit, which resulted in constant flow, was used. The path-plans and the printing speed for each formulation were created using an in-house developed VisualBasic program created in VisualStudio (Microsoft, USA). The printing speed for each sample was automatically calculated based on the extruded mass during 20 seconds of extrusion with the set pressure. The images of 3 printed grids were acquired 3 minutes after printing and were further quantified using Fiji.[20] The printability index was calculated as the ratio between the resolved and expected area of the grid, weighed by the difficulty index of resolving a specific mesh area (Supporting Information, Figure S1). Larger grid areas of the designed path plan were penalized by a lower weight. A ratio above 0.33 was considered printable.

Machine learning algorithm

Data generation

Rheological data and printing conditions related to each formulation were processed according to the template provided in Supporting Information (Table S1). To avoid skewness of dataset due to possible measurement errors at high and low ends of the frequency and strain sweeps, a clipped range of data was used by limiting the angular frequency and oscillatory strain values between 1.0-10 rad.sec^{-1} and 0.1-100%, respectively. The generated

tabulated data were consolidated into a randomly distributed dataset with 65 features per formulation using a MATLAB script (MathWorks, USA).

ML model and selection and evaluation of relevant features

A classification ML model based on Random Forest (RF) algorithm was implemented in Python using the scikit-learn package.[21] The RF classifier is an ensemble non-parametric model based on many decision trees. In order to build an ML model including the features with the relevant and statistically meaningful contributions, a Python implementation of the Boruta all-relevant feature selection method[19] was used (BorutaPy). Iterative randomization steps were employed to compensate for the imbalance in the dataset (positive sample population ~14%). Initially, a subset of the dataset with a balance of 75:25 between the not-Printable: Printable classes was randomly chosen from the original dataset. An RF classifier was trained with the subset, and if the F-score (Equation 1) of the trained model on the test portion of the subset was above 0.80, a Boruta feature screening was subsequently applied. This process was iterated 10000 times, and the most-occurring relevant features were selected from the accumulated responses of the Boruta feature screening algorithm. The F-score was calculated by the following equation:

$$\mathrm{F-score} = \frac{TP}{TP+\frac{1}{2}(FP+FN)} \qquad \text{(Equation 1)}$$

where TP, FP, and FN are the classifier's true-positive, false-positive, and false-negative predictions, respectively.

Shapley values were used as the principal way of describing features' contributions in the implemented ML model. For this purpose, Shapley Additive Explanations (SHAP) were employed to explain features' correlations, interactions, and contributions in predictions.[22] SHAP values were obtained and accumulated through a 20-fold cross-validation of the trained ML model on random splits of the dataset.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the authors.

Acknowledgements

The authors would like to thank Peyman Beyranvand for the insightful discussions during this study.

[1] W. Sun, B. Starly, A. C. Daly, J. A. Burdick, J. Groll, G. Skeldon, W. Shu, Y. Sakai, M. Shinohara, M. Nishikawa, Biofabrication 2020, 12, 022002.
[2] L. Moroni, T. Boland, J. A. Burdick, C. De Maria, B. Derby, G. Forgacs, J. Groll, Q. Li, J. Malda, V. A. Mironov, Trends in biotechnology 2018, 36, 384.
[3] G. Decante, J. B. Costa, J. Silva-Correia, M. N. Collins, R. L. Reis, J. M. Oliveira, Biofabrication 2021, 13, 032001.
[4] L. Moroni, J. A. Burdick, C. Highley, S. J. Lee, Y. Morimoto, S. Takeuchi, J. J. Yoo, Nature Reviews Materials 2018, 3, 21.
[5] S. Naghieh, D. Chen, Journal of Pharmaceutical Analysis 2021.
[6] G. D. Goh, S. L. Sing, W. Y. Yeong, Artificial Intelligence Review 2021, 54, 63.
[7] J. Kim, J. A. McKee, J. J. Fontenot, J. P. Jung, Frontiers in bioengineering and biotechnology 2020, 7, 443.
[8] C. Yu, J. Jiang, International Journal of Bioprinting 2020, 6.
[9] J. An, C. K. Chua, V. Mironov, International Journal of Bioprinting 2021, 7.
[10] A. Conev, E. E. Litsa, M. R. Perez, M. Diba, A. G. Mikos, L. E. Kavraki, Tissue Engineering Part A 2020, 26, 1359.
[11] J. M. Bone, C. M. Childs, A. Menon, B. Poczos, A. W. Feinberg, P. R. LeDuc, N. R. Washburn, ACS Biomaterials Science & Engineering 2020, 6, 7021.
[12] K. Ruberu, M. Senadeera, S. Rana, S. Gupta, J. Chung, Z. Yue, S. Venkatesh, G. Wallace, Applied Materials Today 2021, 22, 100914.
[13] J. Lee, S. J. Oh, S. H. An, W.-D. Kim, S.-H. Kim, Biofabrication 2020, 12, 035018.
[14] A. Shrikumar, P. Greenside, A. Kundaje, "Learning important features through propagating activation differences", presented at *International Conference on Machine Learning*, 2017.
[15] M. T. Ribeiro, S. Singh, C. Guestrin, "" Why should i trust you?" Explaining the predictions of any classifier", presented at *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 2016.
[16] S. M. Lundberg, S.-I. Lee, "A unified approach to interpreting model predictions", presented at *Proceedings of the 31st international conference on neural information processing systems*, 2017.
[17] P. Fisch, M. Holub, M. Zenobi-Wong, Biofabrication 2020, 13, 015012.
[18] G. Gillispie, P. Prim, J. Copus, J. Fisher, A. G. Mikos, J. J. Yoo, A. Atala, S. J. Lee, Biofabrication 2020, 12, 022003.
[19] M. B. Kursa, W. R. Rudnicki, J Stat Softw 2010, 36, 1.
[20] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, Nature methods 2012, 9, 676.
[21] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, the Journal of machine Learning research 2011, 12, 2825.
[22] S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, S.-I. Lee, Nature machine intelligence 2020, 2, 56.

Supporting Information

Machine learning to explain printability induced by rheology additives

Ali Nadernezhad* and Jürgen Groll*

Chair for Functional Materials for Medicine and Dentistry at the Institute for Functional Materials and Biofabrication (IFB) and Bavarian Polymer Institute (BPI), University of Würzburg, Pleicherwall 2, 97070 Würzburg, Germany

* Correspondence to: ali.nadernezhad@fmz.uni-wuerzburg.de; juergen.groll@fmz.uni-wuerzburg.de

Template for collection of features from rheology experiments

The rheological data from the testing protocol were collected according to Table S1. The data obtained from rheological experiments were further processed in later steps to generate features.

Table S1. List of rheological data used for further processing and generation of features

Test type	Measured quantities	The range used for further processing
Frequency sweep	Storage modulus (G)	Angular frequency: 1.0-10 rad/s
	Damping factor (DF)	Angular frequency: 1.0-10 rad/s
Amplitude sweep	Storage modulus (G)	Oscillatory strain: 0.1-100%
	Damping factor (DF)	Oscillatory strain: 0.1-100%
3-Interval Thixotropy Test (3ITT)	Recovery of viscosity during the 3rd interval	Time: 5, 10, and 30 seconds
3-Interval Oscillatory Test (TIOT)	Recovery of storage modulus during the 3rd interval	Time: 5, 10, and 30 seconds
Shear stress sweep	Viscosity	Rotational shear stress: 1-100 Pa
Transient shear steps	Viscosity	Shear rate steps: 0.1-100 s^{-1}

List of features used for training the ML model

Table S2 lists all the features which were used to train the ML model in this study. The description of each coded feature is provided for a better understanding of each feature.

Table S2. List of full features and their corresponding descriptions

Feature	Description	Proportioned[1]
logical_FS_CO	If a crossover in frequency sweep was observed	NA
FS_DF_6	Damping factor at 1.0 rad/s angular frequency	Yes
FS_DF_5	Damping factor at 1.58 rad/s angular frequency	Yes
FS_DF_4	Damping factor at 2.51 rad/s angular frequency	Yes
FS_DF_3	Damping factor at 3.98 rad/s angular frequency	Yes
FS_DF_2	Damping factor at 6.31 rad/s angular frequency	Yes
FS_DF_1	Damping factor at 10.0 rad/s angular frequency	Yes
ratio_FS_DF@1.0/FS_DF@10	Proportioned ratio between Damping factor at 1.0 rad/s to DF at 10.0 rad/s	NA
logical_SS_CO	If a crossover in Amplitude sweep was observed	NA

SS_G_1	Storage modulus at 0.1% strain	Yes
SS_G_2	Storage modulus at 0.148% strain	Yes
SS_G_3	Storage modulus at 0.217% strain	Yes
SS_G_4	Storage modulus at 0.318% strain	Yes
SS_G_5	Storage modulus at 0.467% strain	Yes
SS_G_6	Storage modulus at 0.685% strain	Yes
SS_G_7	Storage modulus at 1.0% strain	Yes
SS_G_8	Storage modulus at 1.48% strain	Yes
SS_G_9	Storage modulus at 2.17% strain	Yes
SS_G_10	Storage modulus at 3.18% strain	Yes
SS_G_11	Storage modulus at 4.67% strain	Yes
SS_G_12	Storage modulus at 6.85% strain	Yes
SS_G_13	Storage modulus at 10.0% strain	Yes
SS_G_14	Storage modulus at 14.8% strain	Yes
SS_G_15	Storage modulus at 21.7% strain	Yes
SS_G_16	Storage modulus at 31.8% strain	Yes
SS_G_17	Storage modulus at 46.6% strain	Yes
SS_G_18	Storage modulus at 68.5% strain	Yes
SS_G_19	Storage modulus at 100.0% strain	Yes
SS_DF_1	Damping factor at 0.1% strain	Yes
SS_DF_2	Damping factor at 0.148% strain	Yes
SS_DF_3	Damping factor at 0.217% strain	Yes
SS_DF_4	Damping factor at 0.318% strain	Yes
SS_DF_5	Damping factor at 0.467% strain	Yes
SS_DF_6	Damping factor at 0.685% strain	Yes
SS_DF_7	Damping factor at 1.0% strain	Yes
SS_DF_8	Damping factor at 1.48% strain	Yes
SS_DF_9	Damping factor at 2.17% strain	Yes
SS_DF_10	Damping factor at 3.18% strain	Yes
SS_DF_11	Damping factor at 4.67% strain	Yes
SS_DF_12	Damping factor at 6.85% strain	Yes
SS_DF_13	Damping factor at 10.0% strain	Yes
SS_DF_14	Damping factor at 14.8% strain	Yes
SS_DF_15	Damping factor at 21.7% strain	Yes
SS_DF_16	Damping factor at 31.8% strain	Yes
SS_DF_17	Damping factor at 46.6% strain	Yes
SS_DF_18	Damping factor at 68.5% strain	Yes

SS_DF_19	Damping factor at 100.0% strain	Yes
SS_LVER_strain	Strain at the limit of linear viscoelastic range	No
SS_LVER_G	Sorage modulus at the limit of viscoelastic range	Yes
SS_CO_strain	Crossover strain in Amplitude sweep	No
ratio_SS_CO_strain/SS_LVER_strain	Ratio between crossover strain and the strain at the limit of viscoelastic	NA
ratio_SS_CO_G/SS_G@0.1	Ratio between storage modulus at crossover and proportioned storage modulus at 0.1% strain	NA
ratio_SS_CO_G/SS_LVER_G	Proportioned ratio between storage modulus at crossover and storage modulus at the limit of the viscoelastic range	NA
logical_FC_yield	If a peak viscosity in the stress sweep test was observed	NA
FC_yield_Eta	Peak viscosity in the stress sweep test	No
FC_CY_Eta0	Zero-shear viscosity from Carreau-Yasuda model	Yes
FC_CY_Lambda	Consistency index (λ) from Carreau-Yasuda model	No
3ITT_Re_5s	Recovery of viscosity in 3ITT after 5s	Yes
3ITT_Re_10s	Recovery of viscosity in 3ITT after 10s	Yes
3ITT_Re_30s	Recovery of viscosity in 3ITT after 30s	Yes
3IOT_Re_5s	Recovery of storage modulus in 3IOT after 5s	Yes
3IOT_Re_10s	Recovery of storage modulus in 3IOT after 10s	Yes
3IOT_Re_30s	Recovery of storage modulus in 3IOT after 30s	Yes
Print_Flow	Proportionality index of extruded volume to the applied pressure	No
logical_Print_Filament	If a filament was formed during extrusion (no filament, broken filament, yes)	No

[1] Proportioned ratio was calculated by dividing the respected value by that of the corresponding HA solution with the same molecular weight and concentration

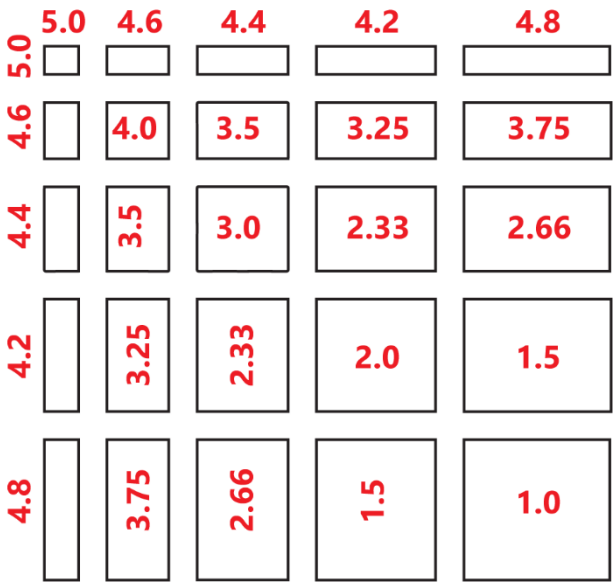
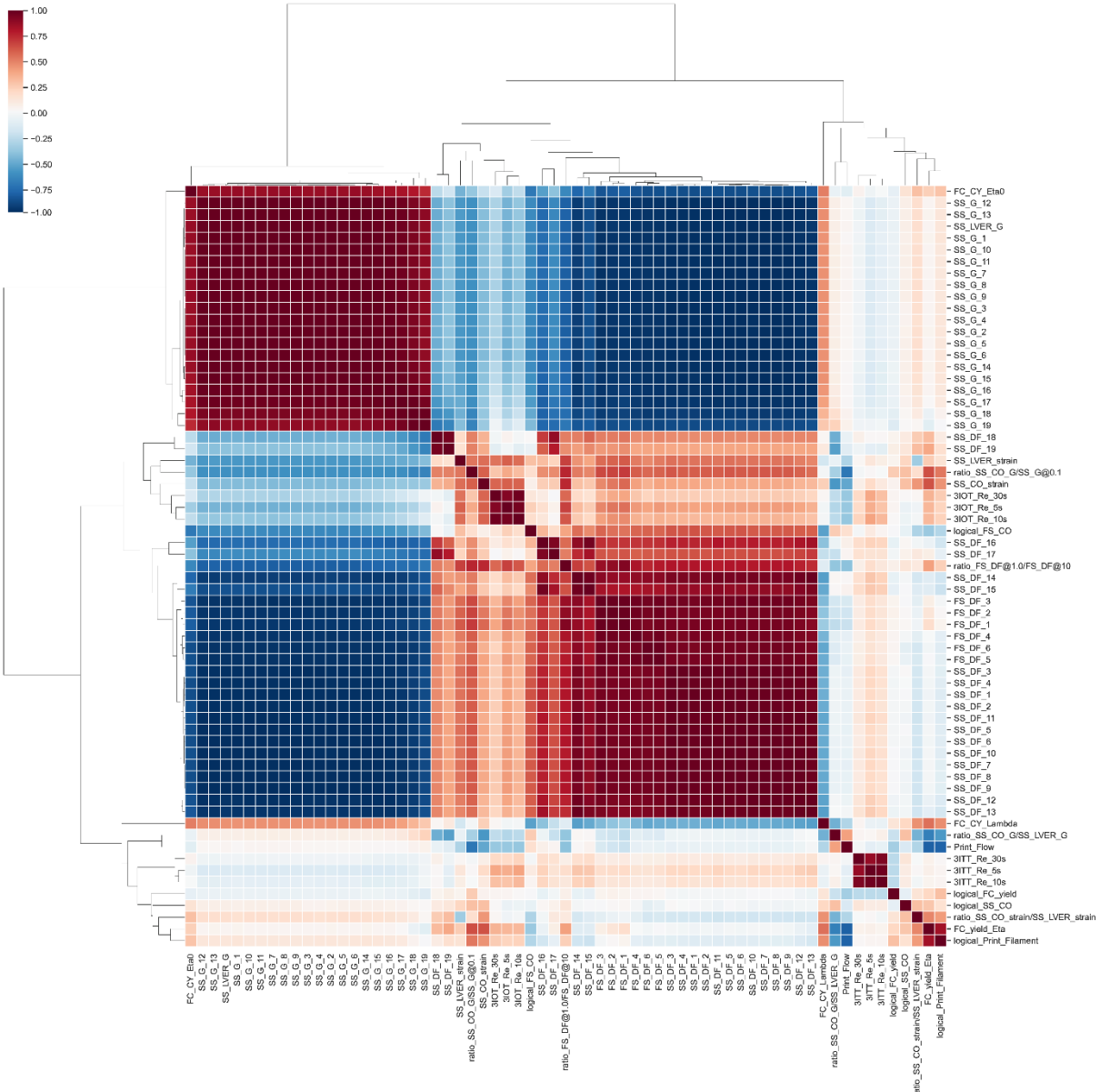


Figure S1. The weights utilized to penalize the easy to resolve areas of the grid. The numbers next to or inside each grid show the multiplication factor.


1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
FC_CY_Eta0
SS_G_12
SS_G_13
SS_LVER_G
SS_G_1
SS_G_10
SS_G_11
SS_G_7
SS_G_8
SS_G_9
SS_G_3
SS_G_4
SS_G_2
SS_G_5
SS_G_6
SS_G_14
SS_G_15
SS_G_16
SS_G_17
SS_G_18
SS_G_19
SS_DF_18
SS_DF_19
SS_LVER_strain
ratio_SS_CO_G/SS_G@0.1
SS_CO_strain
3IOT_Re_30s
3IOT_Re_5s
3IOT_Re_10s
logical_FS_CO
SS_DF_16
SS_DF_17
ratio_FS_DF@1.0/FS_DF@10
SS_DF_14
SS_DF_15
FS_DF_3
FS_DF_2
FS_DF_1
FS_DF_4
FS_DF_6
FS_DF_5
SS_DF_3
SS_DF_4
SS_DF_1
SS_DF_2
SS_DF_11
SS_DF_5
SS_DF_6
SS_DF_10
SS_DF_7
SS_DF_8
SS_DF_9
SS_DF_12
SS_DF_13
FC_CY_Lambda
ratio_SS_CO_G/SS_LVER_G
Print_Flow
3ITT_Re_30s
3ITT_Re_5s
3ITT_Re_10s
logical_FC_yield
logical_SS_CO
ratio_SS_CO_strain/SS_LVER_strain
FC_yield_Eta
logical_Print_Filament