Advanced Materials Technologies

ADVANCED MATERIALS TECHNOLOGIES
www.advmattechnol.com

RESEARCH ARTICLE OPEN ACCESS

Characterization of Droplet Formation in Ultrasonic Spray Coating: Influence of Ink Formulation Using Phase Doppler Anemometry and Machine Learning

Pieter Verding[1] | Danny E. P. Vanpoucke[1] | Yunus T. Aksoy[2] | Tobias Corthouts[1] | Maria R. Vetrano[2] | Wim Deferme[1]

[1]Institute for Materials Research (imo-imomec), Hasselt University, Belgium and Imec imo-imomec, Diepenbeek, Belgium | [2]Department of Mechanical Engineering, Division of Applied Mechanics and Energy, KU Leuven, Leuven, Belgium

Correspondence: Pieter Verding (pieter.verding@uhasselt.be) | Wim Deferme (wim.deferme@uhasselt.be)

Received: 30 September 2025 | **Revised:** 30 November 2025 | **Accepted:** 2 December 2025

Keywords: droplet formation | machine learning modeling | phase doppler anemometry | ultrasonic spray coating

ABSTRACT

This study examines droplet formation in ultrasonic spray coating (USSC) as a function of ink formulation (solvent, polymer, nanoparticles). First, acetone with polyvinylidene fluoride (PVDF) at concentrations from 0 to 4.5 wt.% is used to examine the effect of polymer additions. Additionally, acetone-based SiO_2 nanofluids (0–10 g/L), are explored. Finally, the combination of both polymer (PVDF) and nanoparticles (SiO_2) in acetone is studied. Droplet sizes are measured using Phase Doppler Anemometry under varying atomization power and flow rates. Machine Learning (ML) algorithms are employed to develop droplet size models from key spray parameters, including atomization power, flow rate, polymer concentration, and nanoparticle concentration. The model shows significantly higher accuracy than existing empirical models. The model is further validated on IPA-based inks with polyethylenimine (PEIE) or ZnO nanoparticles, and on acetone–cellulose acetate formulations, confirming its robustness across diverse ink systems. In addition to revealing the influence of coating parameters on the droplet formation and distribution, obtained both via experimental validation and ML, this study demonstrates that machine learning (ML) can be effectively applied to small experimental datasets, offering a robust framework for optimizing droplet formation and understanding key spray parameters in USSC for complex, unexplored inks enabling novel coating applications.

1 | Introduction

Thin film functional coatings are a rapidly evolving field, gaining interest due to their critical role in enhancing the efficiency and functionality of electronic devices such as OLEDs, solar panels, temperature, and strain sensors [1–4]. Together with the transition to flexible electronics, ultra-thin, homogeneous coatings are also beneficial for other fields of research, such as anti-microbial, UV-resistant, self-cleaning, and other functional coatings, where sustainability aspects of both the deposition method as well as the use of eco-friendly materials in reduced amounts (*i.e.*, ultrathin coatings) are required. Traditionally, these films have been deposited using spin coating, as well as vacuum techniques such as evaporation, sputtering, atomic layer deposition, and chemical vapor deposition [1, 5–7]. Although effective, these methods tend to be environmentally unfriendly, energy-intensive, and expensive.

As a more sustainable and green roll-to-roll compatible and cost-effective alternative, solution processing techniques are gaining traction [8–10]. Ultrasonic spray coating (USSC) stands out for its ability to coat large areas and complex 3D substrates with

minimal material waste. USSC operates by generating a fine mist of droplets (10–20 μm diameter) through ultrasound vibrations, allowing for the deposition of coatings with thicknesses as low as 30 nm [10–12]. The technique's ability to produce small droplets with a narrow size distribution is beneficial for homogeneous thin film deposition. The generated droplet size is essential for the thickness and quality of the deposited layers. However, numerous parameters influence the generated droplets and their size distribution, with key factors including atomization power, vibration frequency, and ink flow rate.

Additionally, the formulation of ink plays a critical role [13, 14]. The choice of solvent affects volatility, evaporation rate, and surface tension. Using polymers and/or nanoparticles introduces higher viscosities with elastic behavior. These complex inks can exhibit non-Newtonian behavior at the high shear rates induced by the ultrasonic vibrating nozzle tip. The parameters of complex inks are defined by their rheological properties, including viscosity, shear-thinning or thickening behavior, density, and viscoelasticity. These properties can be described by dimensionless numbers such as Weber (We), Ohnesorge (Oh), Capillary (Ca), and Reynolds (Re) numbers [15].

Existing correlations for droplet size prediction were primarily developed based on the capillary wave hypothesis, initially derived from the work of Lord Kelvin [16, 17]. This hypothesis posits that droplet formation is governed by the interplay of surface tension and inertial forces, leading to the formation of capillary waves on the liquid surface. However, cavitation can play a role, influencing droplet size and distribution [18, 19]. This introduces additional complexity, which the traditional capillary wave-based correlations do not fully account for. In the available models, parameters such as viscosity, liquid flow rate, and amplitude are not included, nor do they consider non-Newtonian or complex fluids. Several experimental correlations have been established using dimensionless numbers (e.g., In, We, Oh) in conjunction with physical parameters (e.g., frequency, density, surface tension), as demonstrated by researchers such as R.J. Lang, R. Rajan, and T.D. Donelly, B. Avvaru, Y. Zhang, K.A. Ramisetty, and S. Kooij [13, 14, 18–22]. These correlations provide insights into how droplet size is influenced by spraying parameters and ink characteristics. However, they are typically valid only within specific parameter ranges and often do not account for droplet size distribution or complex fluids.

Additionally, the measurement techniques used when these correlations were developed, though accurate for their time, are now considered outdated. Modern droplet characterization techniques (e.g., high-speed imaging, Phase Doppler Anemometry (PDA), laser diffraction, and time shift techniques) offer much higher precision [23, 24]. They can provide detailed information on droplet size and/or velocity, along with their distribution. This advancement in measurement technology opens up new possibilities for developing more accurate and universally applicable models, underscoring the need for a new model that is broadly applicable and tailored to the specific requirements of thin film deposition of functional films. Additionally, earlier studies did not typically experiment with complex fluids, such as those containing nanoparticles or polymers. These complex fluids are particularly relevant for functional coatings and their advanced applications. The behavior of such fluids under ultrasonic atomization can differ significantly from that of simpler fluids due to the high applied shear rates, requiring new correlations to accurately predict droplet formation under these conditions.

Our investigation includes a diverse range of inks, such as water, acetone, and acetone with varying concentrations of polyvinylidene fluoride (PVDF). Additionally, the incorporation of SiO_2 nanoparticles at varying concentrations introduces another level of complexity to the study. Beyond droplet size characterization, this research encompasses an analysis of rheological properties, such as surface tension and viscosity, which are fundamental to understanding droplet dynamics. By systematically exploring the influence of the spraying parameters and ink characteristics on droplet formation, we aim to establish a comprehensive understanding of the interplay between ink composition, rheological characteristics, and process parameters.

The outcomes of this investigation not only contribute to the fundamental understanding of ultrasonic spray coating but also provide practical insights for optimizing coating processes across diverse ink formulations. Machine learning performed exceptionally well with these small datasets, resulting in high-quality predictive models. These models were further evaluated using new ink formulations, including isopropanol (IPA)-based inks containing ZnO nanoparticles and/or PEIE as a polymer next to, alongside different concentrations of cellulose acetate in acetone. Therefore, this research lays the groundwork for optimizing the deposition of thin films in applications where precise control over coating properties is crucial, while also enabling the development of simple analytical models.

2 | Results

2.1 | Rheological Results

The rheological behavior of acetone-based solutions containing varying concentrations of PVDF and SiO_2 nanoparticles was characterized using shear-dependent viscosity measurements (Figure S1–S4). Figure 1 summarizes the influence of PVDF, SiO_2, and their combination (with 2.25 wt.% PVDF) on the dynamic viscosity at a shear rate of 100 s^{-1}.

An increasing concentration of PVDF in acetone results in a nearly linear increase in viscosity, attributed to the increased entanglement of polymer chains. Supplementary data (Figures S2 and S4) further reveal that solutions with higher PVDF concentrations exhibit clear shear-thinning behavior, where viscosity decreases with increasing shear rate. This non-Newtonian behavior is typical for polymer solutions due to the alignment and disentanglement of polymer chains under shear.

By contrast, solutions of pure acetone with dispersed SiO_2 nanoparticles (Figure S3) exhibit nearly Newtonian behavior, with minimal change in viscosity over a wide range of shear rates. When both SiO_2 and PVDF are present (Figure S4), the shear-thinning character is maintained and even enhanced, particularly at higher nanoparticle concentrations. This suggests that the combination of polymer and nanoparticles contributes to a more complex microstructure that responds more strongly to shear.

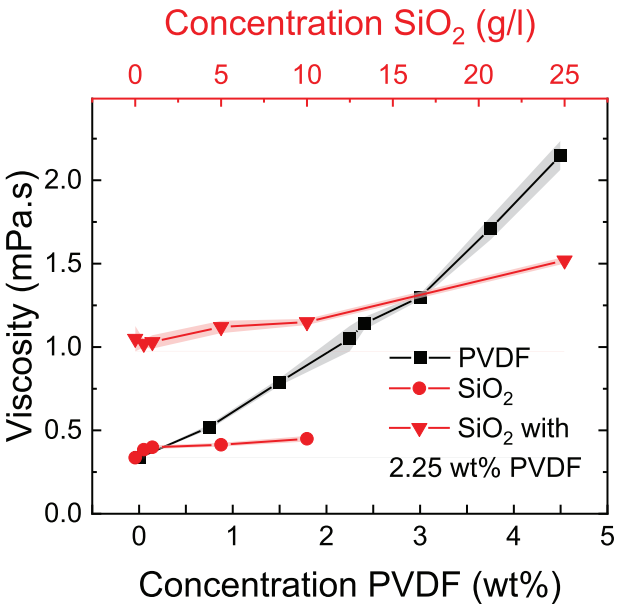

FIGURE 1 | Viscosity of acetone-based solutions as a function of PVDF concentration (wt.%) and SiO_2 concentration (g/L) at the shear rate of 500 s^{-1}. Black squares represent viscosity variations with increasing PVDF concentrations. Red circles indicate viscosity changes with increasing SiO_2 concentrations in pure acetone. Red triangles show viscosity changes with increasing SiO_2 concentrations in acetone containing 2.25 wt.% PVDF. The shaded areas represent the standard deviation ($\pm 1\sigma$) from three independent measurements.

Interestingly, the viscosity trends of SiO_2 in acetone and SiO_2 in 2.25 wt.% PVDF (Figure 1) are nearly parallel, indicating that the incremental viscosity increase due to SiO_2 is preserved even in the presence of PVDF. Solutions with PVDF concentrations above 4.5 wt.% could not be atomized at the maximum applied power (4 W) and were therefore excluded from further analysis.

In addition to viscosity measurements, surface tension measurements were performed to assess the interfacial behavior of the inks (Figure 2). For increasing PVDF concentrations (Figure 2a), surface tension remained relatively constant, indicating that PVDF has limited influence on interfacial tension. Similarly, dispersing SiO_2 nanoparticles in pure acetone (Figure 2b) resulted in only minor, non-systematic variations.

2.2 | Droplet Size and Size Distribution

In this section, droplet sizes and size distributions were measured using Phase Doppler Anemometry (PDA). While droplet velocity and its distribution were also analyzed, they are not discussed further, as measurements were performed directly at the nozzle exit, in the region immediately after atomization and before interaction with the shroud gas. As a result, shroud gas parameters such as type, pressure, temperature, and nozzle-substrate distance are not relevant to the measurer droplet characteristics. The focus of this study is therefore on the droplet size and distribution of five primary model inks: water, acetone, acetone with PVDF, acetone with SiO_2 nanoparticles, and acetone containing both PVDF and SiO_2 nanoparticles. In addition, acetone with cellulose acetate, IPA, IPA with ZnO nanoparticles, and IPA with PEIE were used for model validation.

2.2.1 | Water: Effect on Liquid Flow Rate and Atomization Power

The water droplet size distribution showed a narrow and well-defined peak around the characteristic droplet diameter, as presented in Figure 3c. All measurements in this figure were performed at a constant atomization power of 2.5 W. The corresponding Sauter mean diameter (D_{32}) increased systematically with the flow rate, rising from 22.1 µm at 0.2 mL/min to 30.3 µm at 4 mL/min (Figure 4a). In this increasing trend the presence of two distinct atomization regimes is evident: *Initial Atomization Regime (~0.2–0.6 mL/min)*: At flow rates below 0.2 mL/min, atomization is initially absent due to an insufficient liquid supply to form a stable film on the nozzle tip. Once the threshold is reached (at ~0.2 mL/min in this case), droplet formation begins, and D_{32} increases sharply. This rapid increase is associated with the gradual wetting of the nozzle surface. *Fully Wetted Film Regime (>~0.6 mL/min)*: Beyond approximately 0.6 mL/min, the nozzle becomes fully wetted. Further increase in flow rate leads to thicker liquid films and larger droplets, but the sensitivity of D_{32} to flow rate becomes less pronounced. The system transitions to a more stable atomization regime, where changes in D_{32} occur more gradually. This two-regime behavior is consistent with previous observations by Rajan and Pandit [14].

The effect of atomization power on droplet size distribution and Sauter mean diameter (D_{32}) is shown in Figure 3b,d. As the power increased from 1 to 4 W, while maintaining a 2.5 mL/min water flow rate, a shift toward larger droplet sizes is observed in the droplet size distribution (Figure 3c). This behavior is expected, as higher atomization power leads to increased ultrasonic vibration intensity.

2.2.2 | Acetone: Effect on Liquid Flow Rate and Atomization Power

The droplet size distribution and atomization behavior of pure acetone were analyzed under varying flow rates and atomization powers. In comparison to water, the acetone droplet size distribution exhibits a similarly narrow and unimodal profile, yet with a slightly smaller peak diameter than water (Figure S5c,d). This difference is consistent with the theoretical framework of Rajan and Pandit, which predicted smaller droplet sizes for fluids with lower surface tension [14]. Acetone has a significantly lower surface tension (~23 mN/m) than water (~72 mN/m), which reduces the capillary forces resisting droplet breakup, and thus facilitates the formation of finer droplets under ultrasonic excitation.

The evolution of the Sauter mean diameter (D_{32}) with increasing flow rate is shown in Figure 3e (at a fixed power of 3 W). A clear upward trend is observed, closely resembling the behavior of water. In contrast, the influence of atomization power on D_{32} is minimal for acetone, as shown in Figure 3f. Across a power range of 2.5 to 4 W at a fixed flow rate (1.5 mL/min), the

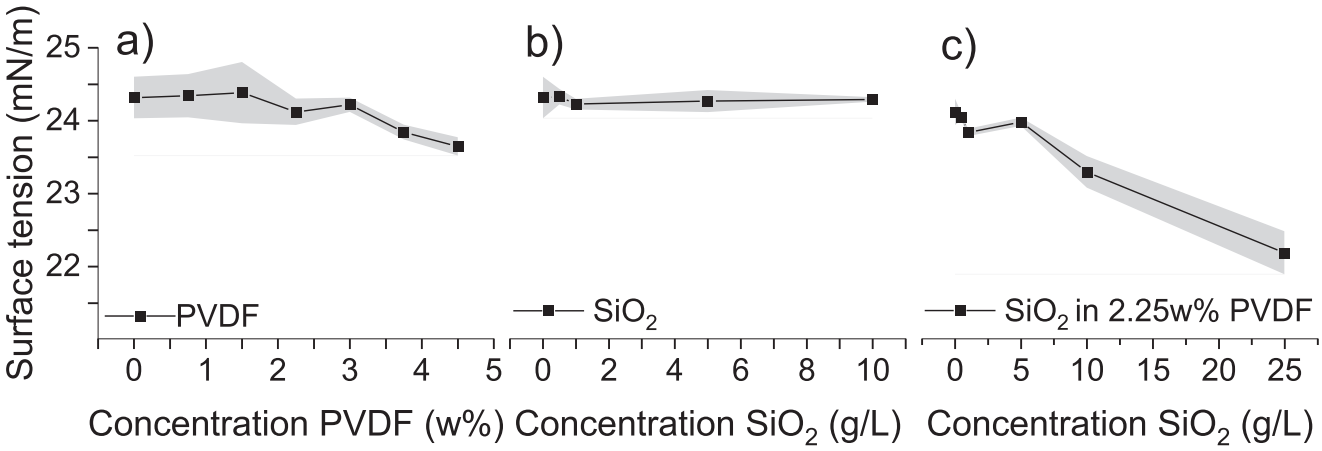

FIGURE 2 | Surface tension measurements of acetone-based solutions with varying concentrations of PVDF and SiO_2 nanoparticles. (a) Surface tension as a function of PVDF concentration in acetone. (b) Surface tension as a function of SiO_2 concentration in pure acetone. (c) Surface tension as a function of SiO_2 concentration in acetone containing 2.25 wt.% PVDF. The shaded regions represent the measurement spread (standard deviation) across repeated measurements. These results illustrate the influence of polymer and nanoparticle loading on the interfacial properties of the ink systems. The shaded areas represent the standard deviation ($\pm 1\sigma$) from three independent measurements.

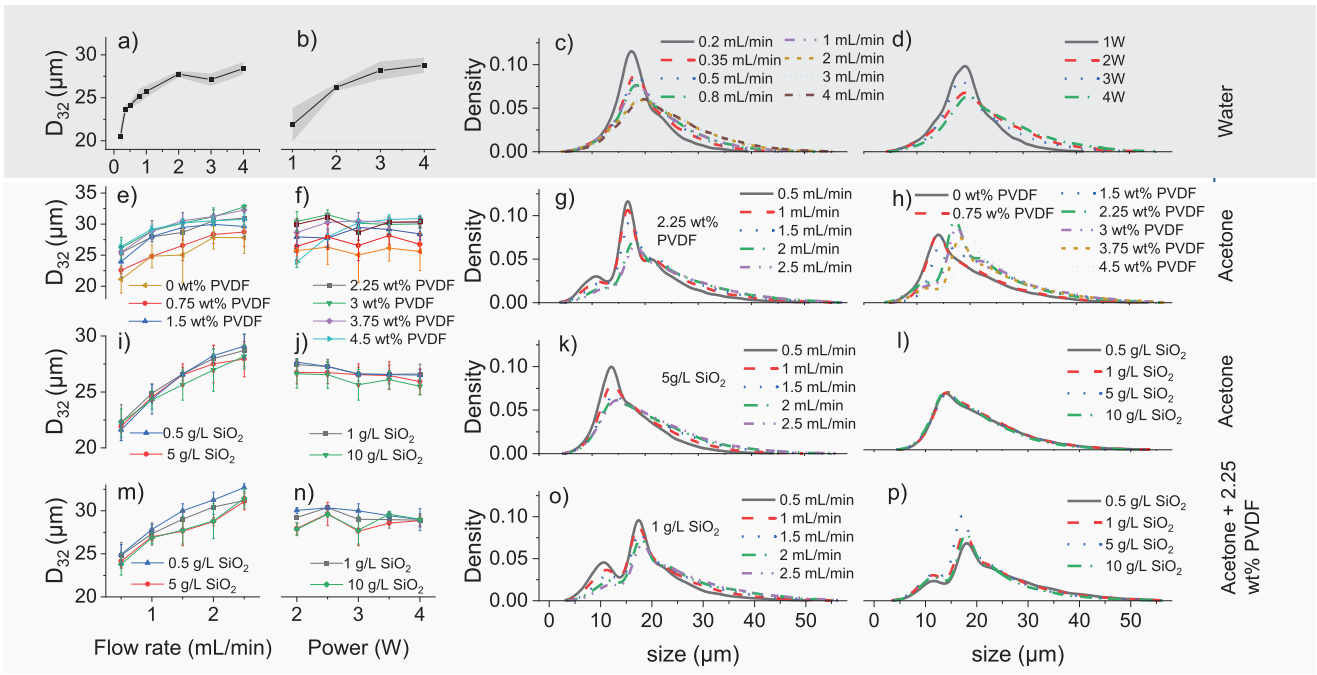

FIGURE 3 | Overview of droplet size distribution and Sauter mean diameter (D_{32}) evolution for different ink formulations in ultrasonic spray coating. (a,b) D_{32} as a function of flow rate and atomization power for water, showing an increase in droplet size with higher flow rates and a weaker correlation with power. (c,d) Droplet size distributions of water at different flow rates and atomization power levels, exhibiting a unimodal distribution with a peak around 10–20 μm. (e,f) Evolution of D_{32} with flow rate and atomization power for acetone with PVDF. D_{32} increased with flow rate, while atomization power introduces fluctuations at higher PVDF concentrations. (g,h) Droplet size distributions of acetone + PVDF, revealing bimodal and trimodal distributions at higher polymer concentrations. (i,j) Effect of flow rate and atomization power on D_{32} for acetone with SiO_2 nanoparticles, exhibiting a similar increasing trend with flow rate but with greater variability due to the presence of nanoparticles. (k,l) Droplet size distributions of acetone + SiO_2 nanoparticles, indicating a broader and more heterogeneous distribution compared to pure acetone. (m,n) D_{32} trends for acetone with both SiO_2 nanoparticles and 2.25 wt.% PVDF, showing increased droplet sizes with flow rate and an irregular response to atomization power due to complex interactions between nanoparticles and polymer viscosity. (o,p) Corresponding droplet size distributions, which exhibit strong multimodal behavior, highlighting the complex atomization dynamics of highly viscous and nanoparticle-laden formulations.

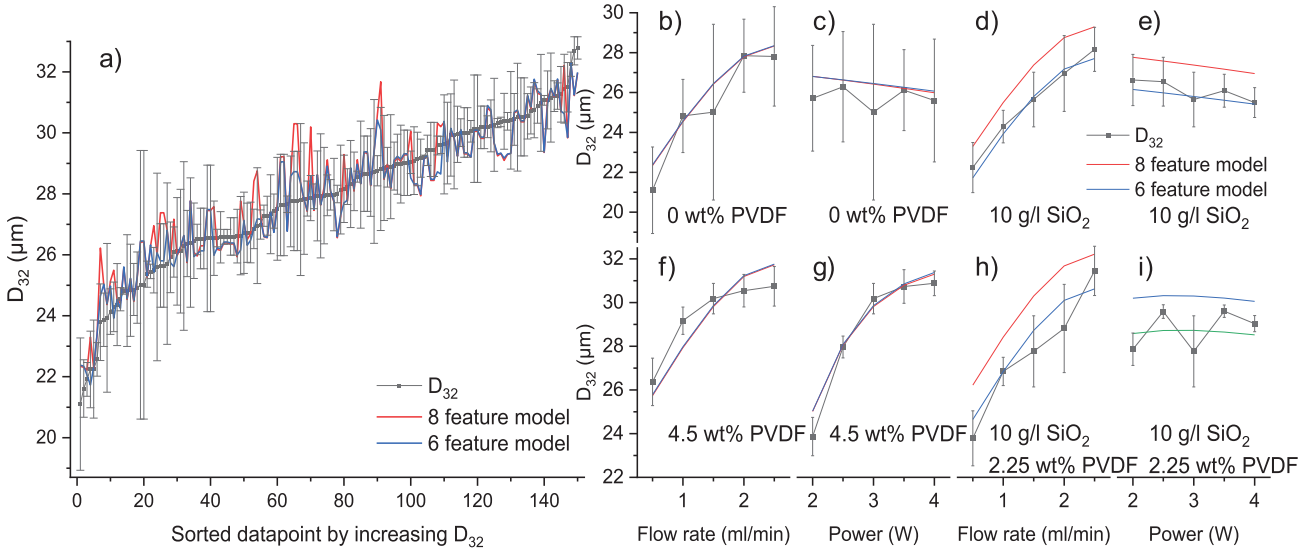

FIGURE 4 | Comparison of the 6-feature and 8-feature machine learning models with experimental data for predicting droplet size (D_{32}) across three different ink systems: pure SiO_2 in acetone, PVDF in acetone, and a combined system of SiO_2 + 2.25 wt.% PVDF in acetone. (a) Full dataset with all datapoints sorted by increasing D_{32}. Experimental data are shown with error bars, and model predictions are plotted for both 6-feature and 8-feature models. (b–i) Subsets of the data split by ink formulation and parameter: (b,c) 0 wt.% PVD, D_{32} as a function of flow rate and power; (d,e) 10 g/L SiO_2, D_{32} as a function of flow rate and power; (f,g) 4.5 wt.% PVDF, D_{32} as a function of flow rate and power; (h,i) 10 g/L SiO_2 + 2.25 wt.% PVDF, D_{32} as a function of flow rate and power. Both models show good agreement with experimental trends, with the 6 feature model performing comparably to the more complex 8-feature model across all formulations. These results highlight the generalization capacity of the models and their ability to capture the influence of polymer and nanoparticle concentrations on droplet formation behavior.

mean droplet size remains nearly constant within the margin of error. This stability indicates that for low-viscosity, low-surface-tension solvents, such as acetone, the primary breakup dynamics are already well established at moderate power levels. Further increases in ultrasonic intensity do not significantly alter the droplet formation process. This contrasts with water, where droplet sizes increase with power, likely due to enhanced capillary wave formation and thicker film dynamics under higher acoustic forces.

The droplet size distributions (Figure S5c,d) confirm this observation. Acetone distributions remain narrow and unimodal across all power levels, with the peak consistently centered around a slightly smaller size than water. The similarity in distribution shape – narrow, monotonic peaks with limited broadening. Acetone and water share the same fundamental atomization mechanism, although differing fluid properties, such as surface tension and viscosity, modulate this mechanism.

2.2.3 | Acetone + PVDF: Effect on Flow Rate, Atomization Power, and PVDF Concentration

The mean droplet size (D_{32}) increases with higher flow rates across all PVDF concentrations. This trend is consistent from 0.5 to 2.5 mL/min (Figure 3e, at a fixed power of 3 W), as higher flow rates results in a larger volume of liquid being atomized, leading to larger droplets. Regarding droplet size distribution density, at 0 wt.% PVDF (pure acetone) (Figure 3g), the distribution is narrow, with a peak around 10–15 μm, and broadens slightly with increased flow rate. With increasing PVDF concentration (Figure 3h), the peak shifts and broadens further.

Higher PVDF concentrations produce larger droplets due to increased solution viscosity (2.4E-4 to 2.2E-3 Pa.s for pure acetone and 4.5 wt.% PVDF, respectively, see Figure S2). The droplet size distributions show multiple peaks, indicating bimodal and trimodal distributions. For instance, at 4.5 wt.% PVDF and lower power settings (2 W appears to be the lower atomization limit, as observed in Figure S6f), a primary peak is observed at approximately 10 μm, and a secondary peak is observed around 20 μm. The observed bimodal/trimodal droplet size distributions in PVDF—acetone solutions (Figure 3h) may be partially attributed to the copolymeric nature and molecular heterogeneity of the PVDF used. Dyneon PVDF 11008/0001 is a copolymer of vinylidene fluoride and hexafluoropropylene, resulting in a molecular weight distribution. This characteristic can lead to variations in local viscosity and phase behavior, potentially causing multiple atomization mechanisms to occur simultaneously, resulting in distinct droplet size populations.

2.2.4 | Acetone + SiO_2 Nanoparticles: Effect on Flow Rate, Atomization Power, and SiO_2 Concentration

The trends observed for acetone with SiO_2 nanoparticles follow a similar pattern to those seen with pure acetone or water. As shown in Figure 3i, the Sauter mean diameter (D_{32}) increased with flow rate, which aligns with expectations based on liquid film formation and droplet detachment dynamics. The increase in droplet size with flow rate is consistent.

Unlike PVDF-containing formulations, the droplet size distribution remains unimodal, as seen in Figure (3 k-l). The absence

of bimodal or trimodal behavior suggests that the presence of nanoparticles alone at these concentrations does not significantly alter the fundamental droplet breakup mechanisms. Instead, the distribution profile is more comparable to that of pure acetone, with a well-defined peak and a gradual broadening as the flow rate increased.

A notable effect, however, is observed when examining the influence of atomization power (Figure 3j). Unlike the unclear trend observed in PVDF-containing formulations, an increase in atomization power leads to a more pronounced decrease in droplet size. This suggests that in low-viscosity, nanoparticle-loaden systems, higher power effectively enhances atomization efficiency, likely by increasing the intensity of capillary wave formation and secondary breakup processes (Movie S1). The decreased droplet size at higher power levels suggests that the presence of nanoparticles does not hinder the atomization process but, but instead follows similar trends to those base fluids.

2.2.5 | Acetone + PVDF+ SiO_2 Nanoparticles: Effect on Flow Rate, Atomization Power, and SiO_2 Concentration

For the formulation containing both PVDF and SiO_2 nanoparticles, the observed trends closely resemble those of acetone with PVDF alone. As seen in Figure 3m, D_{32} increased with the flow rate, following the same liquid film thickening mechanism previously described. Similarly, no distinct trend is observed with atomization power (Figure 3n), as viscosity-driven effects predominate over variations in ultrasonic energy.

However, a key difference emerges in the droplet size distribution (Figure 3o,p). The presence of both PVDF and SiO_2 nanoparticles results in a more pronounced bimodal or trimodal distribution, compared to PVDF-only formulations. This suggests that while PVDF increased viscosity and leads to the formation of larger primary droplets, the addition of nanoparticles further enhances secondary droplet formation, possibly through localized film instabilities and micro-scale interactions between the polymer matrix and dispersed SiO_2 particles, inducing shear thinning.

At higher PVDF concentrations (>2.25 wt.%), the bimodal and trimodal droplet size distributions become increasingly evident, suggesting a more complex breakup mechanism. This can be attributed to the pronounced shear-thinning behavior observed in these polymer-rich solutions (Figure S2), where viscosity decreases with increasing shear rate. Such non-Newtonian behavior alters the local flow dynamics within the jet, facilitating heterogeneous deformation and breakup.

Moreover, the presence of nanoparticles further increased the viscosity and introduced microstructural heterogeneities, especially when combined with PVDF. These heterogeneities may amplify local instabilities during atomization, promoting the formation of distinct droplet size populations. The synergistic effect of polymer-induced elasticity and particle-induced disruption supports the hypothesis that the interplay of viscoelastic and heterogeneous flow properties leads to multiple droplet breakup regimes within a single formulation.

2.3 | Results and Discussion: Machine Learning Models

Although empirical models have been presented in the literature, (Table S1) their applicability to the current spray coating examples seems to perform relatively poorly [13, 14, 18–22]. For that reason, we decided to approach the problem using a small dataset machine learning (ML) method, to determine simple analytical models which only depend on the machine parameters of the setup: (a) flow rate, Q, (b) power, P, and (c) concentration of nanoparticles/polymers, C_x [27]. Within the context of ML, these variables are referred to as features. The predicted properties include (I) the mean Sauter diameter, μ_{D32}, (II) the mean diameter, μ_{D10}, and (III) the mean droplet velocity, μ_v, which are referred to as targets within the ML context. Since μ_{D32} and μ_{D10} have a similar physical meaning and are formally related, we expect the models for these two targets to be similar in shape, providing a source of expert knowledge during feature selection. However, for the third target, μ_v, no meaningful model could be established due to poor predictive performance, in contrast to the diameter-related targets (Figure S12). Therefore, μ_v is not further discussed in this work

The datasets available in this work are relatively limited from the ML perspective, although they are of the typical size commonly found in standard laboratory investigations. The three data sets considered are used as obtained from experiments (*cf.*, Table 4 and *cf.*, Section 4.1. Ink Formulations). To efficiently handle the small size of the datasets, we will utilize small data ML methods. In the past, we have demonstrated that such ensemble methods consistently yield the highest quality models available, with the inference cost comparable to that of a single base model instance [25]. This approach has been successfully used to create simple analytical models for ink-formulations based on fewer than 30 datapoints [26]. In the current work, our approach was extended to include Dominance Importance, in addition to Ensemble Importance for feature selection (*cf.*, Section 4.1. Ink Formulations) [27, 28].

2.3.1 | Modeling the Mean Sauter Diameter μ_{D32} and the Mean Diameter μ_{D10}

For each of the three data sets (PVDF, SiO_2 NP, and SiO_2 NP+PVDF), specialized models are obtained as described in (*cf.*, Section 4.1. Ink Formulations). Starting from an 11-feature complex model, features are iteratively removed based on their ensemble and dominance importance (Tables S2–S19), in some cases extended with trends observed for top-performing sub-models generated for the APDI. While reducing the number of features in the models, the quality measures (MAE and RMSE for both in-bag and out-of-bag data) generally showed further slight improvements, as can be seen in Tables S28–S33. An improvement of the out-of-bag quality indicates an improved generalizability of the model and should be aimed for. Furthermore, if this coincides with a (slight) decrease in quality for in-bag measures, this may indicate that the reduced model also has reduced overfitting, which is another aspect to strive for when using ML. Finally, if the quality measure remains roughly constant upon removal of features, this removal is beneficial as it increases the quality of measures based on information theory. As a result of these

model reductions, three unique models are obtained, one for each data set, for μ_{D32} (Equations 1–3) and μ_{D10} (Equation S1–S3). The trained coefficients are presented in Tables S34–S39, and allow for the construction of analytical models. For μ_{D32} these are

$$\mu_{D32}^{PVDF} = 20.053 + 5.543Q + 2.474C_{PVDF} - 0.250PC_{PVDF} - 0.520e^{Q} - 0.861e^{C_{PVDF}-P} \quad (1)$$

$$\mu_{D32}^{SiO_2\ NP} = 21.242 + 5.359Q - 2.22 \times 10^{-4}C_{SiO_2}^3 - 0.345P - 0.438e^{Q} \quad (2)$$

$$\mu_{D32}^{SiO_2\ NP+PVDF} = 23.807 + 5.142Q + 1.60 \times 10^{-3}C_{SiO_2}^3 - 0.279C_{SiO_2} - 0.075P - 0.371e^{Q} \quad (3)$$

As can be seen from the three equations above, there are equivalent terms (e.g., Q and e^{Q}), as well as some distinct terms (e.g., $e^{C_{PVDF}-P}$). The origin of the latter partially stems from the different composition of the spray coating substance.

As the models for μ_{D32} were developed in lock-step, providing additional domain knowledge for each other, the selected features are the same. However, interestingly also the coefficients turn out to be comparable in size and trends, as one would expect for physically related parameters. The fact that no identical shape was obtained for the three different datasets indicates that these models may be too specialized. As the three data sets only contain four independent variables, a general model can be constructed by considering all unique features present in the three equations. This gives rise to a general 8-feature model, which when trained on the combination of the three data sets results in the following models for μ_{D32} (Equation 4 for μ_{D32} and Equation S4 for μ_{D10}).

$$\mu_{D32}^{gen,8} = 21.783 + 5.339Q - 0.448P + 1.429C_{PVDF} - 0.065C_{SiO_2} - 6.62 \times 10^{-4}PC_{PVDF} - 0.445e^{Q} - 0.682e^{C_{PVDF}-P} + 1.61 \times 10^{-3}C_{SiO_2}^3 \quad (4)$$

It is interesting to note that the coefficients of common features such as Q and e^{Q} as well as the intercepts are roughly the average of the coefficient values found for the specialized models. However, when considering that the general model is a linear combination of the specialized models, trained on the combined data sets, this becomes less unexpected. This stands in stark contrast with the coefficient for the P, C_{PVDF} feature, which reduced one to three orders of magnitude, indicating this feature may not be as essential expected from the specialized models. As our aim is to find a model as generally applicable as possible, we trained each of the four models (i.e., three specialized and one general) for each of four data sets (i.e., three individual data sets and the combined data set). As shown in Tables S40 and S41, the specialized model and its associated data set, which we refer to as native combinations, yield the best performance. The generalized models exhibits a slight loss in performance on the individual datasets, which is not entirely unexpected, as these models must be more general to fit all datasets well.

Considering the different models in detail again, we also note a significant variation in the size of the various coefficients. However, in parallel there is often a similar and opposite variation in the potential feature values. To gain a clearer understanding of the contribution of each feature to the overall prediction, we consider the product of the feature coefficients and the average feature values. These values can be compared as absolutes, or relative to the sum of all these terms in the model (Table S42). This highlights two features which only contribute a very small part of the overall prediction: PC_{PVDF} and $C_{SiO_2}^3$. Furthermore, the calculated Ensemble and Dominance Importance values indicate that these features are suitable for removal (Tables S43 and S44). From Table S42 it is also clear that the feature C_{SiO_2} provides a marginal contribution to the overall prediction. However, removing of this feature would eliminate all possible dependence on the SiO_2 nanoparticle concentration, which is at odds with the EI of this feature, which is 100% (in contrast to the previous two features). Furthermore, there is a significant difference in concentrations considered for PVDF and SiO_2 nanoparticles. Assuming the same model but with a SiO_2 nanoparticle concentration similar to PVDF, the contribution to the prediction would be similar to that of the PVDF concentration. This provides us with an important argument to retain this feature. Removal of the PC_{PVDF} and $C_{SiO_2}^3$ features gives rise to our final general models with only 6 features (Equation 5 for μ_{D32} and Equation S5 for μ_{D10}, Supporting Information).

$$\mu_{D32}^{gen,6} = 21.700 + 5.343Q - 0.408P + 1.427C_{PVDF} - 0.065C_{SiO_2} - 0.447e^{Q} - 0.680e^{C_{PVDF}-P} \quad (5)$$

A comparison of the different machine learning models is presented in Figure 4. Overall, all models perform very well, with predicted D_{32} values closely matching the experimental data and typically falling well within the experimental error margins. In the few cases where deviations occur particularly for datapoints with very narrow error bars, the models still provide reasonably accurate estimates. For the majority of data points, the predictions of the 6-feature, 8-feature, and native models nearly overlap, showing only minimal variation.

When sorting the datapoints by increasing experimental D_{32} values, it becomes clear that the models exhibit no systematic bias: the predictions oscillate around the experimental values while successfully capturing the overall trend (Figure S12). This same behavior is observed in the D_{10} predictions, further confirming the model's reliability. Notably, in Figure 4a, the 6-feature model appears to exhibit slightly smaller deviations from the experimental data compared to the 8-feature model, suggesting a more stable and generalizable performance across the full dataset.

These findings highlight that the model performance is not merely a function of dataset size or algorithm complexity, but also depends strongly on the nature of the data and the quality of the feature space. Even with fewer input features, the 6-feature model provides comparable, or in some cases superior predictive accuracy, supporting its use as a more efficient yet reliable approach for droplet size prediction in complex ink systems.

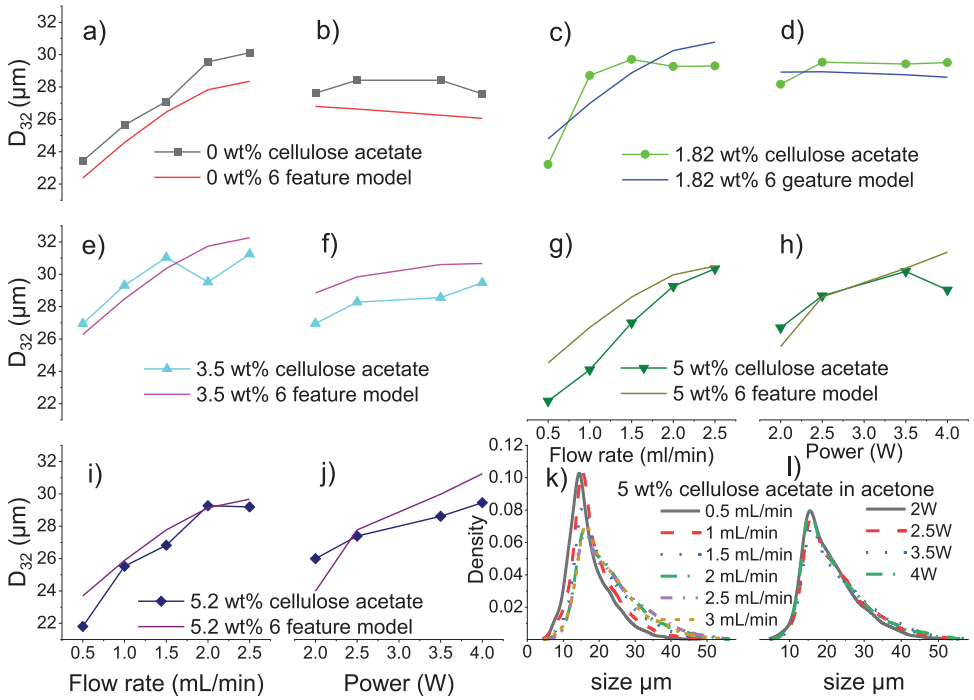

FIGURE 5 | Comparison between experimental results and the 6-feature machine learning model for droplet size (D_{32}) prediction across various concentrations of cellulose acetate in acetone. Panels show Sauter mean diameters at five different concentrations: 0 wt.% (a,b), 1.82 wt.% (c,d), 3.5 wt.% (e,f), 5 wt.% (g,h), and 5.2 wt.% (i,j). For each concentration, D_{32} is plotted as a function of flow rate (left) and atomization power (right). Panels (k–l) display droplet size distributions for 5 wt.% cellulose acetate in acetone, measured using Phase Doppler anemometry (PDA). (k) Increasing flow rate results in a clear shift toward larger droplet sizes. (l) Increasing power has only a minimal effect on the distribution shape and peak size.

2.3.2 | Evaluation of ML Model with New Complex Inks

To evaluate the robustness and generalizability of the developed machine learning model, additional experiments were conducted using new and untrained combinations of solvents, polymers, and nanoparticles. The aim was to assess the model's ability to accurately predict droplet size (D_{32}) under unfamiliar conditions.

2.3.2.1 | Solvent Validation – IPA and Water. 2-propanol (IPA) was tested as a new solvent, both as a pure fluid and in combination with ZnO nanoparticles and the polymer PEIE. In all IPA-based experiments, the model significantly deviated from the experimental results, as expected with a new, untrained solvent. Instead of the predicted decrease of D_{32} with increasing atomization power, as was observed with acetone (Figure S13), 2-propanol showed a clear increase in droplet size. A similar trend was seen with water (Figure S15), which was also not included in the training data. These findings indicate that the model can't be effectively extrapolated to untested solvents with different physicochemical properties, such as viscosity and surface tension. Adding ZnO nanoparticles and the polymer PEIE in IPA yielded in the same conclusion, as the solvent was not suitable for the model. This emphasizes that the model is sensitive to the solvent properties and would require training or feature extension for broader applicability. Importantly, in the case of water, the model failed particularly at the higher flow rate of 4 mL/min (Figure S15) because the model was only trained up to 2.5 mL/min with acetone. The use of PEIE in IPA as the polymer system resulted in a clear increasing trend in droplet size with increasing atomization power (Figure S14). Due to the different physicochemical nature of both the polymer and the solvent, an inverse and more pronounced effect was observed compared to acetone-based inks, with or without PVDF (Figure 3).

2.3.2.2 | Polymer Validation – Cellulose Acetate. Cellulose acetate in acetone was included as a validation experiments at multiple concentrations (1.82, 3.5, 5.0, and 5.2 wt.%) (Figure 5). Across all concentrations, the 6-feature model accurately predicted both the trend and magnitude of the Sauter mean diameter (D_{32}) as the flow rate and atomization power increased. The increase in droplet size with higher polymer concentration and flow rate is consistent with the expected rise in viscosity and film thickness at the nozzle tip. For power, the model correctly captured the relatively moderate influence compared to flow rate, especially at higher polymer concentrations.

Importantly, the results confirm the model's strong generalization capability to previously untrained polymers, as long as they are dissolved in known solvents and fall within the learned parameter space. This demonstrates the model's applicability to a broader class of ink systems with similar rheological behavior, even when the specific polymer chemistry differs from that used during training.

In addition to mean droplet sizes, the droplet size distributions were also examined at 5 wt.% cellulose acetate under varying flow rates and powers (Figure 5k,l). As the flow rate increased, a clear shift toward larger droplet sizes was observed, indicating more complete nozzle wetting and thicker liquid films, which promote the formation of larger droplets. In contrast, increasing the atomization power caused minimal change in distribution shape or peak location, reinforcing the observation that power

plays a secondary role in controlling droplet size for these ink systems.

2.4 | Conclusion and Outlook of Validation

The validation experiments conducted with various untrained ink formulations have revealed valuable insights into the capabilities and current limitations of the developed machine learning model.

The model performed poorly (Figure S14) on inks based on untrained solvents such as isopropanol (IPA) and water. In these cases, predicted trends for droplet size (D_{32}) diverged significantly from experimental observations, particularly at higher flow rates or varying power levels. These discrepancies highlight that the model is currently sensitive to solvent-specific properties and does not generalize well beyond the trained solvent (acetone). Similarly, the inclusion of ZnO nanoparticles, which were not represented in the training set, led to inaccurate predictions, indicating that nanoparticle effects must be explicitly incorporated into future models.

However, the model showed promising generalization to new polymer systems, most notably cellulose acetate in acetone. Across multiple concentrations, the model accurately predicted both the trend and absolute values of D_{32}. This is a significant result, demonstrating that the model is capable of extrapolating to unseen polymers when the solvent remains consistent and within the trained domain. These results suggest that rheological behavior, particularly viscosity, may be a more critical feature than the exact chemical identity of the polymer.

The findings underline the need to enhance the model's robustness by: expanding the training data to include a broader range of solvents, polymer types, and nanoparticle systems and incorporating rheological and physicochemical properties (e.g., viscosity, surface tension, density) directly into the feature set. Increasing the range of flow rates and atomization powers to prevent failure during extrapolation.

The 6/8 feature model demonstrates strong potential for practical use in droplet prediction for polymer-based inks in acetone, however, further refinement is required to reliably handle new solvents and additive types. The successful cellulose acetate results present an encouraging path forward for developing more universal models that are adaptable to diverse ink formulations in ultrasonic spray coating.

3 | Conclusion

This study presents a comprehensive investigation into droplet formation in ultrasonic spray coating by combining detailed experimental analysis using PDA with data-driven predictive modeling via machine learning. Through systematic variation of spray parameters (flow rate, atomization power) and ink formulations (solvent, polymer, nanoparticle concentration), we have provided new insights into the fundamental mechanisms that govern droplet size and distribution.

A key strength of this work lies in the use of PDA, which enabled not only the determination of average droplet sizes (D_{32}), but also the full droplet size distributions. As a result, we could distinguish between unimodal, bimodal, and even trimodal droplet populations. The number and shape of distribution peaks were found to depend strongly on the polymer concentration and the liquid flow rate, while atomization power and nanoparticle addition had a more secondary influence. Particularly at higher PVDF concentrations and moderate flow rates, multimodal distributions emerged, indicating complex atomization dynamics beyond simple capillary wave breakup.

The behavior of different solvents also shows clear distinctions. For instance, water and IPA displayed an increasing D_{32} trend with power, whereas acetone shows a decreasing trend, suggesting that the solvent's physicochemical properties strongly modulate atomization behavior. However, the effect of flow rate remained consistent across all formulations, with higher flow rates generally resulting in larger droplet sizes and distributions.

These results highlight that the ideal set of process parameters must be carefully selected for each ink formulation and application. For example, a narrow and uniform droplet size distribution is generally preferred to ensure homogeneous coatings. In water-based systems, excessive flow rate or power resulted in broad distributions. For acetone-based inks, droplet uniformity was compromised at low power or excessively high polymer concentrations.

The machine learning model developed in this study serves as a powerful tool to predict how spray parameters influence droplet size. Among the models tested, the 6-feature model proved to be consistently accurate and robust, often performing as well as the 8-feature version. Importantly, this work offers an experimental and data-driven study on droplet size distributions in ultrasonic spray coating, particularly for complex, multicomponent inks, an area that remains relatively underexplored in the literature. The integration of machine learning adds a new, valuable layer of understanding, offering a predictive framework to support both scientific exploration and industrial process optimization.

While current limitations exist, particularly for untrained solvents or nanoparticles the model's potential is clearly demonstrated. With further training on a broader range of fluids and rheological properties, it could evolve into a widely applicable tool for tailoring spray parameters to achieve optimal coating performance in advanced thin-film deposition.

4 | Experimental Section/Methods

4.1 | Ink Formulations

The ink formulations consisted of various combinations of solvents, polymers, and nanoparticles. A full overview of all materials and their properties is provided in Table 1, while the corresponding ink concentrations are summarized in Table 2. Polymers used included polyvinylidene fluoride (PVDF), polyethylenimine (PEIE), and cellulose acetate, dissolved in solvents such as acetone or isopropanol (IPA). Nanoparticles such as silicon

TABLE 1 | Materials used and their properties.

Component	Material Details
PVDF	3 M, Dyneon Fluoroplastics 11008/0001, Mw = 1.67×10^5 g/mol
PEIE	Polyethylenimine, 80% ethoxylated solution, 37 wt.% H_2O
Cellulose Acetate	ACROS, average Mw = 100 000 g/mol
SiO_2 Nanoparticles	SkySpring Nanomaterials, 99.8%, 10–20 nm, surface-modified (hydrophobic & oleophilic)
ZnO Nanoparticles	Sigma–Aldrich, 10–15 nm, dispersed in IPA, 99.99% purity
Solvents	Acetone (analytical grade), Milli-Q Water, 2-Propanol (IPA)

TABLE 2 | Ink formulations and concentrations.

Ink Type	Polymer Concentration (wt.%)	Nanoparticle Concentration (g/L)	Solvent
PVDF in Acetone	0.75, 1.5, 2.25, 3, 3.75, 4.5	—	Acetone
Cellulose Acetate	0, 1.82, 3.5, 5.0, 5.2	—	Acetone
SiO_2 in Acetone	—	0.5, 1, 5, 10	Acetone
SiO_2 + PVDF	2.25	0.5, 1, 5, 10	Acetone
PEIE in IPA	0, 2.34, 4.68	—	IPA
ZnO in IPA	—	2.08	IPA

dioxide (SiO_2) and zinc oxide (ZnO) were also included in specific ink formulations. Polymer-based inks were prepared by stirring the solutions in closed bottles at 50°C until complete dissolution. All inks were then characterized for droplet size and rheological behavior.

4.2 | Rheological Measurements

Viscosity measurements of the inks were performed using a Discovery HR-3 TA Instruments rheometer, configured with a gap of 1000 μm, at a temperature of 25°C, and the shear rate ranged from 5 to 500 1/s. Surface tension was determined with an OCA SCA20 instrument using a 2.08 mm needle and acetone as the solvent. Measurements were performed at 22.3°C, applying the Young-Laplace fitting method. The pendant drop method was employed, starting with an initial drop volume of approximately 15 μL. Several hundreds of data points were recorded, from which the average value was obtained.

4.3 | Ultrasonic Spray Coating Apparatus

Ultrasonic spray coating was performed using a Sonotek Impact nozzle operating at 120 kHz, chosen for its capability to provide uniform atomization of ink formulations. Different atomization powers and flow rates were explored, with flow rates ranging from 0.2 to 4 mL/min and power settings from 1 to 4 W.

4.4 | Droplet Size Measurement and Analysis

Droplet velocity and size were analyzed using a 1-D Phase Doppler Anemometry (PDA) system from Dantec Dynamics. The optical setup included a fiber-flow laser transmitter operating at a wavelength of 532 nm (green laser) with a maximum power of 300 mW, a 40 MHz frequency shift provided by the Bragg cell, and a Burst Spectrum Analyzer (BSA) configured with a photomultiplier sensitivity of 800 V, signal gain of 6 dB, a velocity center at 2.84 m/s, and a velocity span of 22.75 m/s. The focal length of the emission unit lens was set to 310 mm, allowing droplet size measurements to range from 1 to 56.2 μm when the angle between the emission and receiving units is set to 45 degrees [29]. The focal length of the receiver lens was a 112 mm.

Measurements were conducted in a controlled environment within a box equipped with float glass windows. Each measurement involved analyzing 10 000 droplets to ensure statistical accuracy, and each measurement was performed three times to ensure experimental repeatability. The probe volume of the PDA system was located 3 mm downstream from the atomization tip. At this position, the shroud gas had not yet influenced the droplets, as it is introduced approximately 3 mm further downstream. Therefore, it does not influence the measured droplet characteristics.

To better represent the droplet size distribution, including larger droplets that deposit significantly more ink locally, we calculated

characteristic droplet sizes, such as the Sauter mean diameter (D_{32}), from the sampled droplets (Equation (6)).

$$D_{32} = \frac{\sum_{i=1}^{N} n_i d_i^3}{\sum_{i=1}^{N} n_i d_i^2} \quad (6)$$

4.5 | Methodology: Machine Learning

The Machine Learning (ML) models are created using the scikit-learn-based AMADEUS framework, previously developed by one of the authors [25, 30, 31]. With this framework, pasting type ensemble models of 1000 base instances are generated [32]. Each instance is trained on a different random 80/20 split of the in-bag and out-of-bag parts of the entire dataset, resulting in 1000 different base instances. The ensemble model is then constructed as a single base instance equal to the average of the 1000 trained base model instances. The quality of the predictions on the out-of-bag data points determines the quality of the ensemble model. We consider linear and polynomial regression models as base model types, and use LASSO regularization [33]. The whole workflow is shown in Figure 6a. It consists of two parts: the first focuses on data preparation, and the second focuses on model generation. The data preparation phase involves selecting hyperparameters for Lasso regularization and generating features based on polynomial models. The model generation process begins with an 11 feature model, which is iteratively reduced in size to yield a simple analytical model.

4.5.1 | Hyperparameter Tuning

Due to the small size and the specific nature of the datasets studied, the regularization strength hyperparameter α could not be optimized using automated functionality, as the optimization process consistently triggered convergence warnings and pushed α toward zero, effectively eliminating regularization. This behavior, along with other signs of overfitting, justified manual tuning. Instead, it is selected based on the comparison of four values ($\alpha = 1.0,\ 0.1,\ 0.01$, and 0.001) while optimising polynomial models up to order ten for the μ_{D32} target of the most extensive dataset available (PVDF). As shown in Figure 6b, strong regularization ($\alpha = 1.0$) results a deterioration of the model compared to a simple linear regression model using the three basic features: flow rate (Q), power (P), and PVDF concentration (C_{PVDF}). Very weak regularization ($\alpha = 0.001$), on the other hand retains too many features in the polynomial model, and shows clear signs of overfitting. For models of polynomial order larger than three, the OOB quality stagnates, for all values of α. Considering the results of the OOB quality for Lasso regularized polynomial models up to order 3, the best balance between high quality (low RMSE) and regularisation strength (high α) is obtained for $\alpha = 0.01$. To achieve maximal transferability between the various models, we therefore choose to use the hyperparameter value $\alpha = 0.01$ for all datasets and targets.

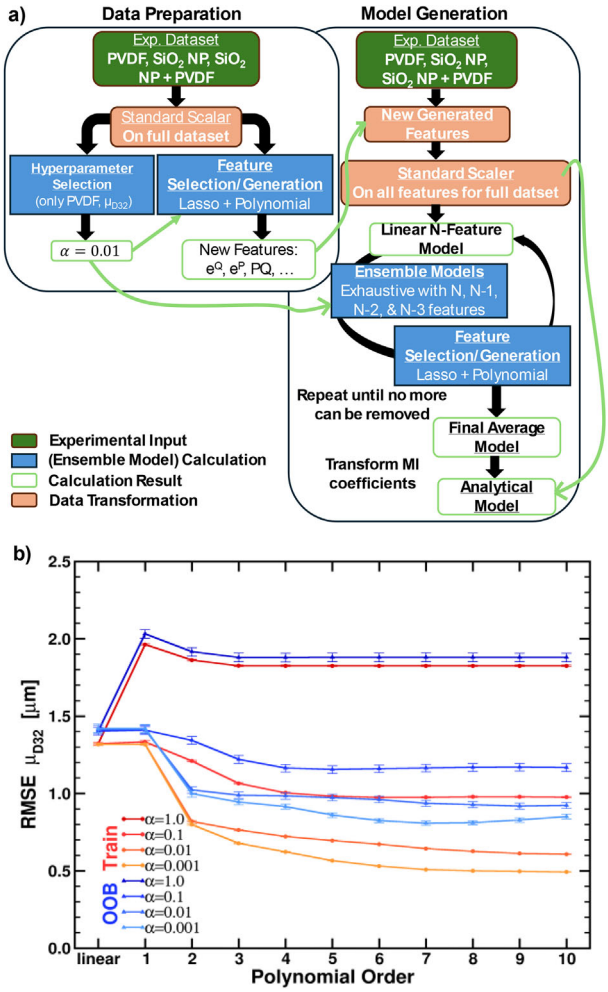

FIGURE 6 | (a) Schematic representation of the machine learning workflow used for droplet size prediction. The pipeline includes data preparation, feature generation, model selection, and ensemble averaging to generate interpretable analytical models. (b) Hyperparameter optimization for the prediction of the mean Sauter diameter (μD_{32}) for the PVDF dataset. Lasso-regularized polynomial models up to 10th order are compared to a non-regularized linear model using input features: flow rate (Q), power (P), and polymer concentration (C_{PVDF}). Root Mean Square Error (RMSE) is shown for training (red) and out-of-bag (OOB, blue) samples. Error bars indicate the 95% confidence interval.

4.5.2 | Feature Generation

Polynomial models evaluated during hyperparameter tuning demonstrated that higher predictive accuracy can be achieved compared to simple linear regression using the three main input features: flow rate (Q, in mL/min), atomization power (P, in W), and the concentration of either the polymer (C_{PVDF}, in wt.%) or added nanoparticles (C_{SiO2}, in g/L), denoted as (Q, P, C_x). However, polynomial models quickly grow in the number of features, leading to overfitting when dealing with small datasets. For example, starting from three base features, a third-order polynomial model already includes 20 fitting coefficients (one intercept and 19 feature terms), while a tenth-order model contains as many as 286 coefficients. The number of relevant

TABLE 3 | Initial features for the PVDF dataset.

		Features										
dataset	target	0	1	2	3	4	5	6	7	8	9	10
PVDF	μ_{D32}, μ_{D10}	Q	C_x	e^Q	e^P	e^{C_x}	QP	PC_x	Qe^{C_x}	$e^{(P-C_x)}$	$e^{(C_x-P)}$	C_x^3
	μ_v	Q	C_x	QP	e^{C_x}	$\frac{1}{1-C_x}$	QPC_x	Q^2PC_x	QC_x	Q^2C_x	PC_x	C_x^3

TABLE 4 | Description of datasets.

Name	Size	Features
PVDF	70	Q, P, C_{PVDF}
SiO_2 NP	38	Q, P, C_{NP}
SiO_2 NP + PVDF	50	Q, P, C_{NP}

features can be reduced through regularization, reducing the risk of overfitting. In the case of small datasets and the ensemble models constructed in this work, the regularization results in some features being relevant for some of the base model instances but not for others. Furthermore, some polynomial features will be relevant more often than others, indicating that these are more important for a global model. The Amadeus framework tracks the frequency at which a polynomial feature is retained within the ensemble, thereby providing a measure of importance for this specific polynomial feature. We refer to this as Ensemble Importance (EI). A feature with an EI of 100% will have a non-zero coefficient in each of the base model instances present in the ensemble. In comparison, a feature with an EI of 75% will have a coefficient of zero (due to regularization) in 25% of the base model instances in the ensemble. An EI of 0% means a polynomial feature is regularized away in each and every base model instance within the ensemble.

Investigation the features with the highest EI for polynomial models up to order 10 provides insight into possible more complex and relevant features to consider. This approach was successfully used in the past [26]. In the current work, such a study is performed for each of the three targets and, for each of the three datasets. The analysis for the μ_{D32} and μ_{D10} targets is performed in parallel for each dataset, resulting in a single set of generated features per dataset for both μ_{D32} and μ_{D10} models. In addition, knowledge obtained from the largest dataset (PVDF), which is studied first, is also considered during the feature generation for the models for the smaller datasets. The initial set of 11 features generated for PVDF model generation is shown Table 3.

Datasets: In the current work, three datasets are considered (*cf.*, Table 4. Description of datasets.). As the PVDF dataset is the largest, it was modelled first, providing additional insights for the feature selection and creation with the other datasets.

4.5.3 | Data Preparation

Although the data values can be used as they are in any ML approach, it is well known that more robust models are obtained when all features have values in similar ranges. Although this is roughly the case for the three base features, it is not the case for the newly generated features, where differences can be orders of magnitude (as will be seen when the analytical models are constructed). In this work, we constructed and applied a standard scaler to the entire dataset, as previous research has shown that for small datasets, a standard scaler constructed solely based on the training data is extremely volatile [25]. In a sense, we consider the application of the standard scaler to occur outside the ML context, with the ML training receiving a well-behaved multifeatured dataset by default.

4.5.4 | Model Quality

Within this work, model quality is evaluated using the Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE), both on the in-bag (or training) and out-of-bag (OOB) datasets. The latter provides a measure for the generalizability of the model (i.e. how well it is expected to perform for unknown data). The OOB quality measures is always lower than the in-bag quality measures, and too significant differences between the two could be a sign of overfitting. When creating new ML models, a new model with lower in-bag quality but the same OOB quality as before may be a superior model as it overfits less, and thus it may better represent the underlying global reality.

4.5.5 | Feature Importance

In this work, we consider four measures to determine feature importance. These measures can also be viewed in the context of ML model interpretability, where the goal is to decompose the prediction of a complex ML model into the contributions made by each feature.

Inter Actional Dominance Importance (IADI): The decrease in MAE when removing a feature f_i from a full N-feature model.

Average Partial Dominance Importance (APDI): The average decrease in MAE due to the removal of the feature f_i, considering a set of submodels differing only in the presence of feature f_i.In the current work, an exhaustive set of sub-models with up to three features removed was used.

Ensemble Importance (EI): The fraction of model instances that retain the feature f_i under regularisation (i.e. the associated coefficient is not reduced to zero).

Sub-model Ensemble Importance (SEI): The fraction of sub-models, generated for APDI, including feature f_i, for which feature f_i has an EI greater than 95%.

Dominance Importance measures are generally defined in terms of the R^2 quality measure, but they are equally applicable to other quality measures [27, 28]. The advantage of using the MAE (or RMSE) over R^2 is that the result also has a quantitative meaning. As MAE and RMSE have the same units as the target of interest, the IADI and APDI will as well. In addition, we chose the sign to reflect the direct impact of the feature f_i. For IADI and APDI, features with positive values improve the quality of the model's quality. In contrast, a negative value indicates the quality of the model improves if the feature is removed. Note that this sign convention is opposite to what is done by, for example, Ceder and co-workers [28]. Finally, we chose to use the MAE OOB quality measure to determine the IADI and APDI, which are used in a qualitative fashion. The feature importance values for the different models generated are presented in Table S2–S19.

4.5.6 | Model Generation

With the hyperparameter set, an experimental dataset available, and an initial set of complex features selected, ensemble models are generated using 1000 base model instances trained on random 80/20 in-bag out-of-bag splits of the dataset. One split per base model instance, means 1000 random splits of the dataset. The base model is a linear regression model that utilize the 11 previously selected complex features (*cf.*, Table 3). This is equivalent to having a complex non-linear regression model using the three base features, but computationally much more efficient. Although 11 features may be limited when considering complex ML tasks, this is still a lot for the small datasets considered. Thus, we aim to reduce this number as much as possible, with minimal loss in quality. We therefore iteratively modify the 11-feature set by removing the least relevant features (or introducing new potential candidate features). The evolution of the models and their quality measures are presented in Tables S27–S33. With regard to the μ_{D32} and μ_{D10} models it needs to be noted that these two are always considered as a pair for each dataset, based on the known relation. The relevance of the features is determined in each iteration cycle based on Dominance Importance and Ensemble Importance measures [27, 28]. When no more features can be removed without a significant loss in model quality, the iterative process ends, and the learned average coefficients of the final model are transformed back to create a simple analytical model, ready for use in lab conditions. Since each feature f_i is standard scaled from an experimental feature g_i via Equation 7

$$f_i = \frac{g_i - \mu_i}{\sigma_i} \tag{7}$$

with μ_i the mean and σ_i the standard deviation of the feature values. As such, the ML regression model

$$z = b + \sum_i a_i f_i \tag{8}$$

with intercept b and coefficients a_i, can be reformulated as

$$z = b' + \sum_i a_i' g_i \tag{9}$$

where $a_i' = \frac{a_i}{\sigma_i}$ and $b' = b - \sum_i \frac{a_i \mu_i}{\sigma_i}$. The values for μ_i, σ_i, a_i and b for the final models are listed in Tables S34–S39.

Acknowledgements

This work was partly funded by the Research Foundation-Flanders (FWO) via projects 1S99123N, 1SB2424N, and BOF B20INCEN17. The computational resources and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by the Research Foundation Flanders (FWO) and the Flemish Government – department EWI. The authors also gratefully acknowledge the in-kind support of the Department of Chemical Engineering, KU Leuven with the nanofluid characterization devices.

Funding

This work was partly funded by the Research Foundation-Flanders (FWO) via projects 1S99123N, 1SB2424N and BOF B20INCEN17.The computational resources and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by the Research Foundation Flanders (FWO) and the Flemish Government—department WEWIS.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. N. Ahmad, Y. Zhao, F. Ye, et al., "Cadmium-Free Kesterite Thin-Film Solar Cells with High Efficiency Approaching 12%," *Advanced Science* 10, no. 26 (2023): 2302869, https://doi.org/10.1002/advs.202302869.

2. C. Xu, Q. Li, Q. Song, et al., "Analyzing the Synergistic Effect of Ag and Ge co-incorporation on Cu2ZnSnSe4 Thin-film Solar Cells," *Materials Today Energy* 40 (2024): 101518, https://doi.org/10.1016/j.mtener.2024.101518.

3. M. Jose, A. Bronckaers, R. S. N. Kumar, et al., "Stretchable Printed Device for the Simultaneous Sensing of Temperature and Strain Validated in a Mouse Wound Healing Model," *Scientific Reports* 12, no. 1 (2022): 10138, https://doi.org/10.1038/s41598-022-13834-6.

4. R. S. Nithyananda Kumar, M. Eerdekens, Y. de Coene, et al., "Plasmon Enhanced Fluorescence from Meticulously Positioned Gold Nanoparticles, Deposited by Ultra Sonic Spray Coating on Organic Light Emitting Diodes," *Nanoscale Advances* 5, no. 6 (2023): 1750–1759, https://doi.org/10.1039/d2na00753c.

5. L. Y. Zhu, L. X. Ou, L. W. Mao, X. Y. Wu, Y. P. Liu, and H. L. Lu, "Advances in Noble Metal-Decorated Metal Oxide Nanomaterials for Chemiresistive Gas Sensors: Overview," *Nano-Micro Letters* 15, no. 1 (2023): 89, https://doi.org/10.1007/s40820-023-01047-z.

6. Z. Liang, Y. Zhang, H. Xu, et al., "Homogenizing out-of-plane Cation Composition in Perovskite Solar Cells," *Nature* 624, no. 7992 (2023): 557–563, https://doi.org/10.1038/s41586-023-06784-0.

7. Z. Ye, C. Tan, X. Huang, et al., "Emerging MoS_2Wafer-Scale Technique for Integrated Circuits," *Nano-Micro Letters* 15, no. 1 (2023): 38, https://doi.org/10.1007/s40820-022-01010-4.

8. S. M. Wang, H. Y. Zhang, B. H. Zhang, Z. Y. Xie, and W. Y. Wong, "Towards High-Power-Efficiency Solution-Processed OLEDs: Material and Device Perspectives," *Materials Science and Engineering: R: Reports* 140 (2020): 100547, https://doi.org/10.1016/j.mser.2020.100547.

9. N. Ikeda, S. Oda, R. Matsumoto, et al., "Solution-Processable Pure Green Thermally Activated Delayed Fluorescence Emitter Based on the Multiple Resonance Effect," *Advanced Materials* 32, no. 40 (2020): 2004072, https://doi.org/10.1002/adma.202004072.

10. R. S. N. Kumar, A. V. Ramirez, P. Verding, et al., "Deposition of Ultra-Thin Coatings by a Nature-Inspired Spray-On-Screen Technology," *Communications Engineering* 2, no. 1 (2023): 42, https://doi.org/10.1038/s44172-023-00093-0.

11. P. Verding, W. Deferme, and W. Steffen, "Velocity and Size Measurement of Droplets from an Ultrasonic Spray Coater Using Photon Correlation Spectroscopy and Turbidimetry," *Applied Optics* 59, no. 25 (2020): 7496–7503, https://doi.org/10.1364/ao.402579.

12. S.-I. Kwon, K.-H. Kyung, J.-Y. Park, et al., "Uniform Anti-reflective Films Fabricated by Layer-by-layer Ultrasonic Spray Method," *Colloids and Surfaces A: Physicochemical and Engineering Aspects* 580 (2019): 123785, https://doi.org/10.1016/j.colsurfa.2019.123785.

13. R. J. Lang, "Ultrasonic Atomization of Liquids," *The Journal of the Acoustical Society of America* 34, no. 1 (1962): 6–8, https://doi.org/10.1121/1.1909020.

14. R. Rajan and A. B. Pandit, "Correlations to Predict Droplet Size in Ultrasonic Atomisation," *Ultrasonics* 39, no. 4 (2001): 235–255, https://doi.org/10.1016/S0041-624X(01)00054-3.

15. C. Clasen, P. M. Phillips, and L. Palangetic, "Dispensing of Rheologically Complex Fluids: the Map of Misery," *AIChE Journal* 58, no. 10 (2012): 3242–3255, https://doi.org/10.1002/aic.13704.

16. J. W. S. Rayleigh and R. B. Lindsay, *The Theory of Sound, 2nd, Revised and Enlarged ed. (Dover Classics of Science and Mathematics)* (Dover, 1945).

17. B. W. Mwakikunga, "Progress in Ultrasonic Spray Pyrolysis for Condensed Matter Sciences Developed from Ultrasonic Nebulization Theories Since Michael Faraday," *Critical Reviews in Solid State and Materials Sciences* 39, no. 1 (2014): 46–80, https://doi.org/10.1080/10408436.2012.687359.

18. K. A. Ramisetty, A. B. Pandit, and P. R. Gogate, "Investigations into Ultrasound Induced Atomization," *Ultrasonics Sonochemistry* 20, no. 1 (2013): 254–264, https://doi.org/10.1016/j.ultsonch.2012.05.001.

19. B. Avvaru, M. N. Patil, P. R. Gogate, and A. B. Pandit, "Ultrasonic Atomization: Effect of Liquid Phase Properties," *Ultrasonics* 44, no. 2 (2006): 146–158, https://doi.org/10.1016/j.ultras.2005.09.003.

20. T. D. Donnelly, J. Hogan, A. Mugler, et al., "An Experimental Study of Micron-scale Droplet Aerosols Produced via Ultrasonic Atomization," *Physics of Fluids* 16, no. 8 (2004): 2843–2851, https://doi.org/10.1063/1.1759271.

21. S. Kooij, A. Astefanei, G. L. Corthals, and D. Bonn, "Size Distributions of Droplets Produced by Ultrasonic Nebulizers," *Scientific Reports* 9 (2019): 6128, https://doi.org/10.1038/s41598-019-42599-8.

22. Y. Zhang, S. M. Yuan, and L. Z. Wang, "Investigation of Capillary Wave, Cavitation and Droplet Diameter Distribution during Ultrasonic Atomization," *Experimental Thermal and Fluid Science* 120 (2021): 110219, https://doi.org/10.1016/j.expthermflusci.2020.110219.

23. W. Schäfer and C. Tropea, "Time-Shift Technique for Simultaneous Measurement of Size, Velocity, and Relative Refractive Index of Transparent Droplets or Particles in a Flow," *Applied Optics* 53, no. 4 (2014): 588–597, https://doi.org/10.1364/AO.53.000588.

24. R. Sijs, S. Kooij, H. J. Holterman, J. van de Zande, and D. Bonn, "Drop Size Measurement Techniques for Sprays: Comparison of Image Analysis, Phase Doppler Particle Analysis, and Laser Diffraction," *AIP Advances* 11, no. 1 (2021): 015315, https://doi.org/10.1063/5.0018667.

25. D. E. P. Vanpoucke, O. S. J. van Knippenberg, K. Hermans, K. V. Bernaerts, and S. Mehrkanoon, "Small Data Materials Design with Machine Learning: When the Average Model Knows Best," *Journal of Applied Physics* 128, no. 5 (2020): 054901, https://doi.org/10.1063/5.0012285.

26. D. E. Vanpoucke, M. A. Delgove, J. Stouten, et al., "A Machine Learning Approach for the Design of Hyperbranched Polymeric Dispersing Agents Based on Aliphatic Polyesters for Radiation-Curable Inks," *Polymer International* 71, no. 8 (2022): 966–975, https://doi.org/10.1002/pi.6378.

27. R. Azen and D. V. Budescu, "The Dominance Analysis Approach for Comparing Predictors in Multiple Regression," *Psychological Methods* 8, no. 2 (2003): 129–148, https://doi.org/10.1037/1082-989X.8.2.129.

28. H. Huo, C. J. Bartel, T. He, et al., "Machine-Learning Rationalization and Prediction of Solid-State Synthesis Conditions," *Chemistry of Materials* 34, no. 16 (2022): 7323–7336, https://doi.org/10.1021/acs.chemmater.2c01293.

29. Y. T. Aksoy, H. Cornelissen, P. Eneren, and M. R. Vetrano, "Spray Cooling Investigation of TiO_2–Water Nanofluids on a Hot Surface," *Energies* 16 (2023): 2938, https://doi.org/10.3390/en16072938.

30. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, and B. Thirion, "Scikit-Learn: Machine Learning in Python," *Journal of Machine Learning Research* 12 (2011): 2825–2830, https://doi.org/10.5555/1953048.2078195.

31. D. Vanpoucke, "Artificial Intelligence and Machine Learning Framework for Designing Useful Materials," https://github.com/DannyVanpoucke/Amadeus (accessed 2025).

32. L. Breiman, "Pasting Small Votes for Classification in Large Databases and On-Line," *Machine learning* 36, no. 1 (1999): 85–103, https://doi.org/10.1023/a:1007563306331.

33. R. Tibshirani, "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society Series B: Statistical Methodology* 58, no. 1 (1996): 267–288, https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.

Supporting file 1: admt70630-sup-0001-SuppMat.docx

Supporting file 2: admt70630-sup-0002-MovieS1.mp4