Machine learning enhanced droplet microfluidics ⊕

Claire Barnes ⬤ ; Ashish R. Sonwane (आीशर सोनवने) ⬤ ; Eva C. Sonnenschein ⬤ ; Francesco Del Giudice ✉ ⬤

Check for updates

View Online   Export Citation

Articles You May Be Interested In

Confinement effect on the viscoelastic particle ordering in microfluidic flows: Numerical simulations and experiments

Physics of Fluids (April 2022)

Machine learning enhanced droplet microfluidics EP

View Online   Export Citation   CrossMark

Claire Barnes,[1] (ID) Ashish R. **Sonwane** (आशीष र. सोनवने),[2] (ID) Eva C. Sonnenschein,[3] (ID) and Francesco Del Giudice[4,a] (ID)

AFFILIATIONS

[1]Department of Biomedical Engineering, Faculty of Science and Engineering, School of Engineering and Applied Science, Swansea University, Fabian Way, Swansea SA1 8EN, United Kingdom

[2]Computational Foundry, Faculty of Science and Engineering, Swansea University, Fabian Way, Swansea SA1 8EN, United Kingdom

[3]Department of Biosciences, Faculty of Science and Engineering, School of Biosciences, Geography and Physiscs, Swansea University, Singleton Park, Swansea SA2 8PP, United Kingdom

[4]Department of Chemical Engineering, Faculty of Science and Engineering, School of Engineering and Applied Science, Swansea University, Fabian Way, Swansea SA1 8EN, United Kingdom

[a)]Author to whom correspondence should be addressed: francesco.delgiudice@swansea.ac.uk

ABSTRACT

Machine learning has recently been introduced in the context of droplet microfluidics to simplify the process of droplet formation, which is usually controlled by a variety of parameters. However, the studies introduced so far have mainly focused on droplet size control using water and mineral oil in microfluidic devices fabricated using soft lithography or rapid prototyping. This approach negated the applicability of machine learning results to other types of fluids more relevant to biomedical applications, while also preventing users that do not have access to microfluidic fabrication facilities to take advantage of previous findings. There are a number of different algorithms that could be used as part of a data driven approach, and no clear comparison has been previously offered among multiple machine learning architectures with respect to the predictions of flow rate values and generation rate. We here employed machine learning to predict the experimental parameters required for droplet generation in three commercialized microfluidic flow-focusing devices using phosphate buffer saline and biocompatible fluorinated oil as dispersed and continuous liquid phases, respectively. We compared three different machine learning architectures and established the one leading to more accurate predictions. We also compared the predictions with a new set of experiments performed at a different day to account for experimental variability. Finally, we provided a proof of concept related to algae encapsulation and designed a simple app that can be used to generate accurate predictions for a given droplet size and generation rate across the three commercial devices.

I. INTRODUCTION

The formation of droplets in microfluidic devices has been the subject of significant interest in the past few decades[1–3] for many applications such as single-cell analysis,[4] spheroids and organoids production,[5] and digital polymerase chain reaction (PCR).[6] The droplet formation mechanism in micrometre-scale devices is a complex phenomenon controlled by several parameters,[3,7–11] including surface tension, viscosity of the two non-miscible liquids, surface wetting properties, inertia, elasticity, addition of surfactant to either of the two liquid phases, and device geometry. Such complexity has so far prevented the establishment of a universal framework for the production of droplets in microfluidic devices, while the best approach is to empirically derive some scaling arguments that can only be applied within the experimental configurations presented in the study (see, for instance, the following articles[12,13] and references therein).

Machine learning (ML) and artificial intelligence (AI) have recently emerged as a very powerful tool to identify patterns across a complex set of data and to then deploy models to act on the data in a human-like manner but on a very short timescale.[14–16] ML and AI are sometimes used as synonyms, but they are different:[17,18] AI is capable of acting on a large set of data in order to trigger decisions on a short amount of time; ML is instead a class of AI-based methodologies that enable algorithms to learn without direct programming.[15] For instance, Nitta *et al.*[19] have employed ML to train the algorithm to recognize several populations of flowing cells in microfluidic devices; after recognition of the cell population type, the AI

activated a deflector in the microfluidic device to sort cells according to a specific target.

Considering its power, it is not a surprise that ML has been used in the attempt of tackling the problem of droplet formation in microfluidic devices.[17,20] Most studies[21–25] have combined ML with the dimensional analysis in order to predict the experimental parameters required to produce droplets of a given size. Chagot et al.[26] employed ML to establish the effect of surfactant concentration on the size of water-in-oil droplets formed in flow-focusing configuration. ML has also been used in the context of droplet characterization,[27,28] sorting,[29] microrheology with optical tweezers,[30] bubble motion,[31] droplet tracking,[27] computational fluid dynamics (CFD),[32] and experimental optimization for digital microfluidic applications.[33] However, the droplet microfluidic studies introduced earlier have been limited in their analysis, most often not allowing the user to depart from spherical droplets nor enabling selection of the droplet generation rate, which is an essential parameter, especially in the context of encapsulation.[3] Recently, Lashkaripour et al.[34] presented a detailed study where ML was employed to derive the dimensions of a flow-focusing microfluidic device and the experimental volumetric flow rate values required to produce droplets of a given size at a given generation rate, both pre-selected by the user. The authors also enabled the possibility to upload a new set of data to generate new predictions based on the requirements from the users, who may need to employ a combination of fluids different from the water and mineral oil presented in the study.

Despite the progresses made in the area of ML applied to droplet microfluidics, there are still some unaddressed problems. First, the majority of studies focused on the simple combination of water and mineral oil, which is not the preferred combination of liquids employed in biomedical applications, where fluorinated oil and phosphate buffer saline (PBS) are generally preferred.[35] Second, most of the studies relied on the possibility for the users to have access to microfluidic fabrication facilities, something that is not always true and that can hinder deployment of microfluidic solutions in remote areas. Finally, many of the papers published so far used and optimized a single architecture based on previous works or preference. A comparison between different approaches is essential, particularly for future work that may consider more parameters or more complex fluids.

In this work, we employed ML to predict the experimental parameters required for droplet generation in commercialized microfluidic devices using PBS and fluorinated oil as dispersed and continuous liquid phases, respectively. We compared the accuracy of three ML architectures in predicting the experimental conditions to obtain droplets of a given size and generation rate. We also provided a proof of concept related to the encapsulation of algae in both PBS and algae media. Finally, we developed a simple standalone app that can be used locally in conjunction with commercial microfluidic devices to predict the values of the volumetric flow rate of both the continuous and dispersed phases required to obtain droplets having a pre-determined size and generation rate.

II. MATERIALS AND METHODS
### A. Sample preparation and characterization

We employed phosphate buffer saline (PBS, Sigma Aldrich UK) as the dispersed phase and fluorinated oil FluoSurf 2 wt. % in HFE 7500 (Dolomite Microfluidics, UK) as the continuous phase. We used a force tensiometer (Biolin scientific) to measure the interfacial tension

γ between the FluoSurf and the PBS at room temperature ($T = 20\,°C$), obtaining a value of $\gamma = 3.39 \pm 1.09$ mN/m. The viscosity of the FluoSurf at 20 °C was evaluated using a Twin Drive rheometer working in a strain-controlled mode (Anton Paar MCR702) equipped with a 50 mm cone and plate metal configuration, obtaining a constant value of $\mu_c = 1.56$ mPa s.

B. Experimental setup

We employed three commercially available microfluidic flow-focusing devices (Dolomite Microfluidics, UK) having three etching depths, namely, 100 μm (part 3200512), 190 μm (part 3200506), and 275 μm (part 3200824), all having a fluorophilic coating on their surfaces, meaning that they could be used for the production of aqueous droplets when the continuous phase was fluorophillic (Fig. 1).

The flow was controlled using a syringe pump having two independent flow units and also a pressure sensor on each flow stream (Dolomite Quad Pumps), equipped with the Mitos Quad Pump Green
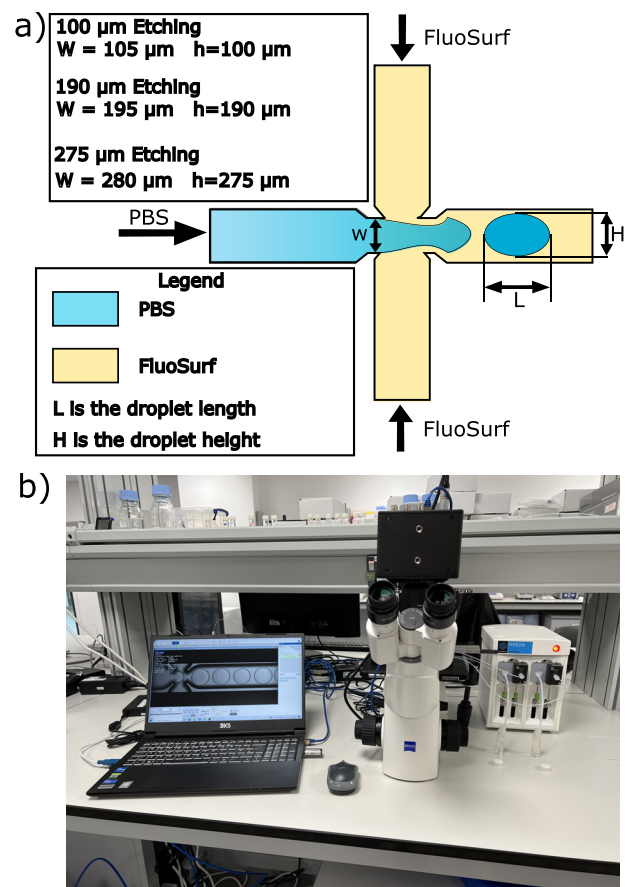
FIG. 1. (a) Schematic representation of the microfluidic flow-focusing devices employed here. Three different commercially available devices having different values of both orifice width W and channel etching height h are used. The dispersed phase is made of phosphate buffer saline (PBS), while the continuous phase is a FluoSurf 2 wt. % emulsion stabilizer. When the two phases meet at the junction, a droplet having length L and height H is formed. The flow direction is from left to right. Dimensions are not to scale. (b) Experimental apparatus in the laboratory.

syringe pairs (Dolomite Microfluidics), with a lower flow rate limit of $5\,\mu$l/min and an upper limit of $1250\,\mu$ l/min. The advantage of this system was the possibility of generating a continuous flow of liquids taken from a reservoir at a constant volumetric flow rate (at variance with more traditional syringe pumps that require the syringes to be filled in as soon as the liquid in the syringe ends), while also simultaneously enabling the measurement of the overall pressure drop in the system. This approach was essential to monitor pressure drop variations and accumulation points of pressure that could bias the accuracy and reliability of the experimental data set.[36,37] The lower end of the volumetric flow rate was $5\,\mu$l/min for both the dispersed or continuous phases (this was set by the type of syringes employed), while the upper flow rate limit was reached when the overall pressure drop for either of the two flow streams reached 10 bars which, in our experiments, roughly corresponded to a volumetric flow rate of $120\,\mu$l/min. According to the manufacturer, the maximum pressure tolerable by the microfluidic device was 25 bars, meaning that when the volumetric flow rates of the two streams were equal to $120\,\mu$l/min, the overall pressure drop was around 20 bars, close to the device limit. In the case of overpressure, the flow was automatically arrested by the syringe pumps.

The connections between the microfluidic device and the syringe pumps were made of fluorinated ethylene propylene (FEP) commercially available tubing with an internal diameter of $100\,\mu$m and an external diameter of 1.6 mm (Dolomite Microfluidics). The choice of the internal diameter of the tube was driven by the requirement to obtain a stable flow at flow rate values of $5\,\mu$l/min, which was the lower end investigated here for both flow streams. We observed that the droplet formation at the junction for flow rate values of $5\,\mu$l/min was unstable when employing the same FEP tube having an internal diameter of either 800 or $250\,\mu$m (data not showed). The reason for such a phenomenon was the fact that the overall pressure drop in the system (given by the sum of the pressure drops in the tubing and the device) was comparable with the pressure drop at the junction where droplet formation occurred, thus causing significant fluctuations in the flow, in line with well-established findings in the literature.[38]

The flow in the microfluidic channel was observed using an inverted microscope (Zeiss Primovert) connected to a high-speed camera (Photron Mini Ux-50) with videos acquired at frame rate values varying between 125 and 8000 fps. The recorded videos were manually analyzed with droplets annotated using Matlab to derive the droplet length L, the droplet height H, and the generation frequency f; a minimum of 20 droplets per case was analyzed: this was deemed sufficient as the droplet size and frequency remained constant for the whole duration of the experiment, as also expected in droplet microfluidic applications.[10]

C. Machine learning architectures

Three machine learning architectures, labeled as architecture 1, architecture 2, and architecture 3 were compared in this study (Fig. 2). Architecture 1 was made of a series of two neural networks with three hidden layers.

The layers consisted of a number of nodes and associated activation functions (in our case reLU) to determine the importance of a particular neuron when predicting the output. The layers were connected to each other via weights, which were subsequently updated during the learning process. The application of hidden layers enabled

the network to account for dependency between variables along with complex relationships between input and output parameters.

Our overall challenge was summarized by a multi-output regression task, and therefore, the number of target variables was set at the number of nodes in the output layer. Hereafter, we will refer to the input parameters as those required by the user (e.g., droplet size), while the output parameters will always refer to the predictions made by the machine learning (e.g., volumetric flow rate values).

The operating principle of architecture 1 is described in Fig. 2(a). First, the droplet length L and the droplet height H were used as input parameters to then predict the volumetric flow rate values of both continuous (Q_c) and dispersed (Q_d) phases. After this first step, the values of Q_c and Q_d estimated from the first neural network, together with the droplet length L, were fed as inputs into a second neural network to predict the value of the droplet generation frequency f. In practical terms, architecture 1 enabled the prediction of Q_d, Q_c, and f, given an initial set of input related to droplet length L and droplet height H. This meant that the generation frequency f could not be specified by the user *a priori*, but it could only be the result of the neural network.

Architecture 2 took a sequential modeling approach based on standard vector regression (SVR), which practically means an architecture based on the fitting of the training data [Fig. 2(b)]. Via the use of kernels, the algorithm could fit a hyperplane in high-dimensional space [green line in Fig. 2(b)], which could be used to estimate the relationship between the input and output parameters. Architecture 2 was easily able to deal with correlations between output variables; however, its success was dependent on the order of chained regressions. Here, L and H were employed as input parameters (similarly to architecture 1), but only Q_d was predicted via the first SVR. Furthermore, the predicted value of Q_d was employed together with the original input parameters to predict Q_c using a second SVR. Finally, L, Q_c, and Q_d were used to predict the value of f. This architecture was very rudimentary and did not rely on neural networks.

Architecture 3 was a single neural network presenting L, H, and f as input parameters and Q_c and Q_d as output parameters [Fig. 2(c)]. Architecture 3 was the most flexible among the three with respect to droplet microfluidic applications, as it enabled the user to set *a priori* droplet length, height, and generation frequency. The reliability of each architecture with respect to either the ground truth data (droplet size) or the mass balance (generation frequency) was assessed using the coefficient of determination R^2.

The choice of these architectures rested on some considerations made in relation to the existing state of the art. Machine learning algorithms similar to architectures 1 and 3 have been used separately in different studies but, to the best of our knowledge, they have never been directly compared. Such comparison is important to elucidate whether training the algorithm with the droplet generation frequency is desirable or not. Additionally, we wanted to see whether a neural network was required at all to enhance droplet microfluidic application: for this reason, we tested architecture 2, which is based on an advanced fitting, rather than a neural network.

D. Algae encapsulation experiments

Tetraselmis suecica CCAP 66 / 4 was obtained from the culture collection of algae and protozoa (CCAP) and cultivated in f/2 medium[39] prepared with seawater at 18 °C and 80 μmol m^{-2} s^{-1} constant light and shaken at 90 rpm. We used the force tensiometer
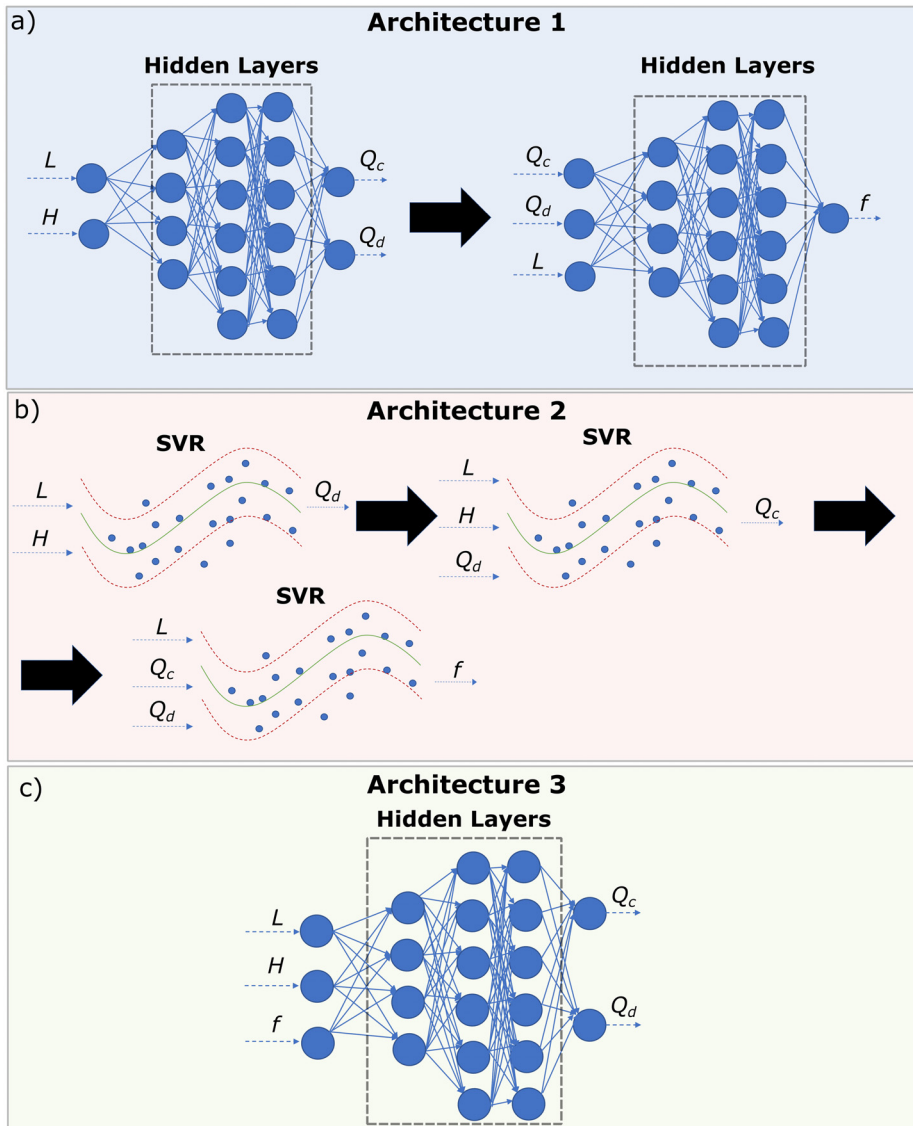
FIG. 2. Schematic representation of the three architectures employed here. (a) Architecture 1 is made of two neural networks in series, with L and H as inputs and Q_c and Q_d as outputs. Furthermore, L, Q_c, and Q_d are used again as input parameters to the second identical neural network to predict f. (b) Architecture 2 is based on a series of standard vector regressions (SVR) with workflow similar to (a). (c) Architecture 3 is a single neural network with L, H and f as inputs and Q_c and Q_d as outputs.

described earlier (Biolin scientific) to measure the interfacial tension γ between the FluoSurf and the algae medium at room temperature ($T = 20\,°C$), obtaining a value of $\gamma = 8.69 \pm 0.15$ mN/m.

The same experimental apparatus described before was used for the algae encapsulation experiments. For encapsulation in the algae media, the solution containing algae was employed as the disperse phase, and the experimental values of Q_c and Q_d were evaluated via architecture 3. For the experiments in phosphate buffer saline (PBS), the algae where first centrifuged using a micro-centrifuge, the media was removed using a pipette, and then the algae were re-suspended in the PBS, that was subsequently employed as the dispersed phase for the encapsulation experiments. The droplets generated were then collected on a Sedgewick rafter counting chamber, closed with a coverslip, and sealed within a petri-dish to prevent droplet evaporation.

E. App for droplet generation

To support this work, we designed a user-friendly graphical interface (GUI), enabling the user to easily predict the necessary flow conditions required to produce droplets of desired dimensions. Our app also enables the user to retrain our architectures using their own experimental data and use these to make predictions. The app was designed to tackle some of the common challenges associated with the task of droplet formation and, therefore, promote greater adoption within the microfluidic community. The GUI was developed using Rust, a multi-paradigm, general-purpose programing language that became prominent in the software industry in recent years due to the advantages of memory safety and better performance for certain tasks. All AI models were developed and trained using Tensorflow and Keras. The code for these steps was written in Python and called by

31 January 2025 10:26:27

our application using PyO3, a rust library that provides bindings for a Python interpreter.

III. RESULTS AND DISCUSSION

A. Generation of the ground truth

The set of data employed to train a machine learning algorithm is generally called "ground truth" to make a clear distinction between the experimental or simulation data employed for the training, which are considered to be the "truth," and the machine learning predictions. Therefore, it was critical that the ground truth was made of a highly reliable and reproducible dataset, or the predictions from the machine learning algorithm could be incorrect.[16,34] The formation of droplets in microfluidic devices is a very complex problem, which can be affected by several parameters, including pressure fluctuations, microfluidic fabrication, and surface wetting.[7–10] To minimize the impact of fabrication techniques and channel wetting on the droplet formation mechanism, we employed commercialized flow-focusing devices, whose reproducibility is ensured by industrial quality standards, while also reducing potential biases related to fabrication in different laboratory settings. Additionally, commercialized devices can be deployed in areas where microfluidic fabrication techniques, such as soft lithography,[40] milling,[41] or 3D printing,[42,43] are not accessible.

We performed experiments in three fluorophilic microfluidic devices with different etching height values h, all having a flow-focusing configuration (Fig. 1), and then compared the results with theoretical scaling laws available in the literature[12,44] to gain confidence in the reliability of the dataset as a whole. First, we identified the droplet shapes that we could achieve in each microfluidic device [Fig. 3(a)]. For devices having an etching height of either 100 or 190 μm, we mainly observed circular droplets (from the top view), while more elliptical-like droplets were observed in the channel with etching height of 275 μm. We then proceeded to verify the reliability of the experimental dataset to be used as ground truth [Fig. 3(b)]. It has recently been reported that, for either Newtonian[12] or non-Newtonian[44] liquid droplets, it is possible to identify a universal trend for the experimental data when plotting the droplet length L normalized by the orifice width W as a function of the ratio between the flow rate ratio q and the Capillary number of the continuous phase Ca_c, with $q = Q_d/Q_c$, where Q_d and Q_c are the volumetric flow rates of the dispersed phase and of the continuous phase, respectively, and Ca_c defined as[45] $Ca_c = \mu_c U_c/\gamma$, where μ_c is the viscosity of the continuous phase, γ is the interfacial tension between the two non-miscible liquids, and U_c is the velocity of the continuous phase, defined as $U_c = Q_c/(Wh)$. We observed that our data [Fig. 3(c)] qualitatively followed the relation introduced by Chen *et al.*,[12]

$$\frac{L}{W} = a + b\left(\frac{q^m}{Ca_c^n}\right), \tag{1}$$

with the two parameters $m = 0.45$ and $n = 0.35$ fixed using the same values recently reported in the literature,[46] while the other two parameters $a = 0.786 \pm 0.032$ and $b = 0.10585 \pm 0.0107$, obtained via a best fit of the data for the 100 μm etched device. The data for the 100 μm etched device followed the scaling closely, as expected by the fitting, while the other data obtained from the other two devices displayed some deviations, similarly to results available in the literature.[44] In practical terms, the deviation of some experimental data from the main fitting meant that a standard scaling curve could not be used to
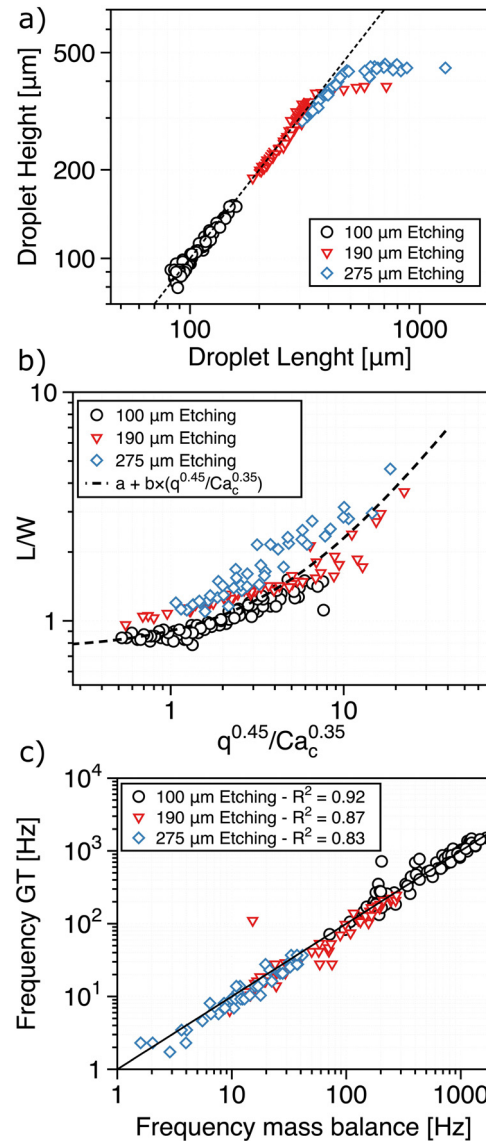
FIG. 3. (a) Droplet height H as a function of droplet length L for the three devices employed in this work. (b) L/W, where W is the channel width, as a function of $q^{0.45}/Ca_c^{0.35}$, where $q = Q_d/Q_c$, while Ca_c is the Capillary number of the continuous phase (see the main text for the definition). The dashed line is a fitting with $a = 0.786 \pm 0.032$ and $b = 0.0559 \pm 0.0056$. (c) Experimental droplet generation frequency f as a function of the f derived via the mass balance (see the text for more details). The lines in (a) and (c) have a unitary slope and a null intercept.

predict the experimental values of Q_c and Q_d to actually obtain droplets of a given length L, thus justifying the need of a machine learning algorithm capable of identifying trends across several experimental data. Moreover, the ability to generate ellipsoidal particles demonstrated the versatility of our work in terms of applications and justification for the use of L and H as opposed to a single droplet diameter or a droplet volume, often employed in the literature.[26,34]

We also tested the reliability of the experimental dataset with respect to the droplet generation frequency. Even though we could not identify a general scaling for our experimental data (not shown), we compared the experimental values of the droplet generation frequency with the one calculated from a mass balance, similarly to an approach adopted previously.[34] In general, the frequency of droplet formation f is linked to the flow rate of the dispersed phase Q_d by the following expression:

$$Q_d = V_d \times f, \tag{2}$$

where V_d is the volume of the droplet, being $V_d \approx L \times H \times h$. We found very good agreement between our experimental frequency data and those obtained via Eq. (2) [Fig. 3(c)], with a coefficient of determination $R^2 \geq 0.8$ for all the different devices.

Taken together, our results showed that the experimental dataset derived in the three microfluidic flow-focusing devices was robust enough to be employable as ground truth for the training of the machine learning algorithms. We were justified in employing a

machine learning algorithm over a standard fitting procedure, as we could not clearly identify a mastercurve describing all our experimental results with the data following the scaling only qualitatively.

B. Comparison among the different machine learning algorithms

In this section, we analyzed the prediction efficacy of the three architectures with respect to the ground truth data and the theoretical scaling introduced in Sec. III A (Fig. 4).

First, we focused on the microfluidic device with $100\,\mu m$ of etching depth and compared the flow rate ratio predicted by the machine learning algorithm against the ground truth data [Fig. 4(a)]. We employed $N_{GT} = 79$ experimental points to train the three architectures and observed that architecture 3 delivered the most accurate predictions, while architecture 2 provided the least accurate results. These results are not surprising considering that architecture 3 presented the droplet generation frequency f as input parameters, and f was linked to Q_d via Eq. (2). The addition of the experimental ground truth
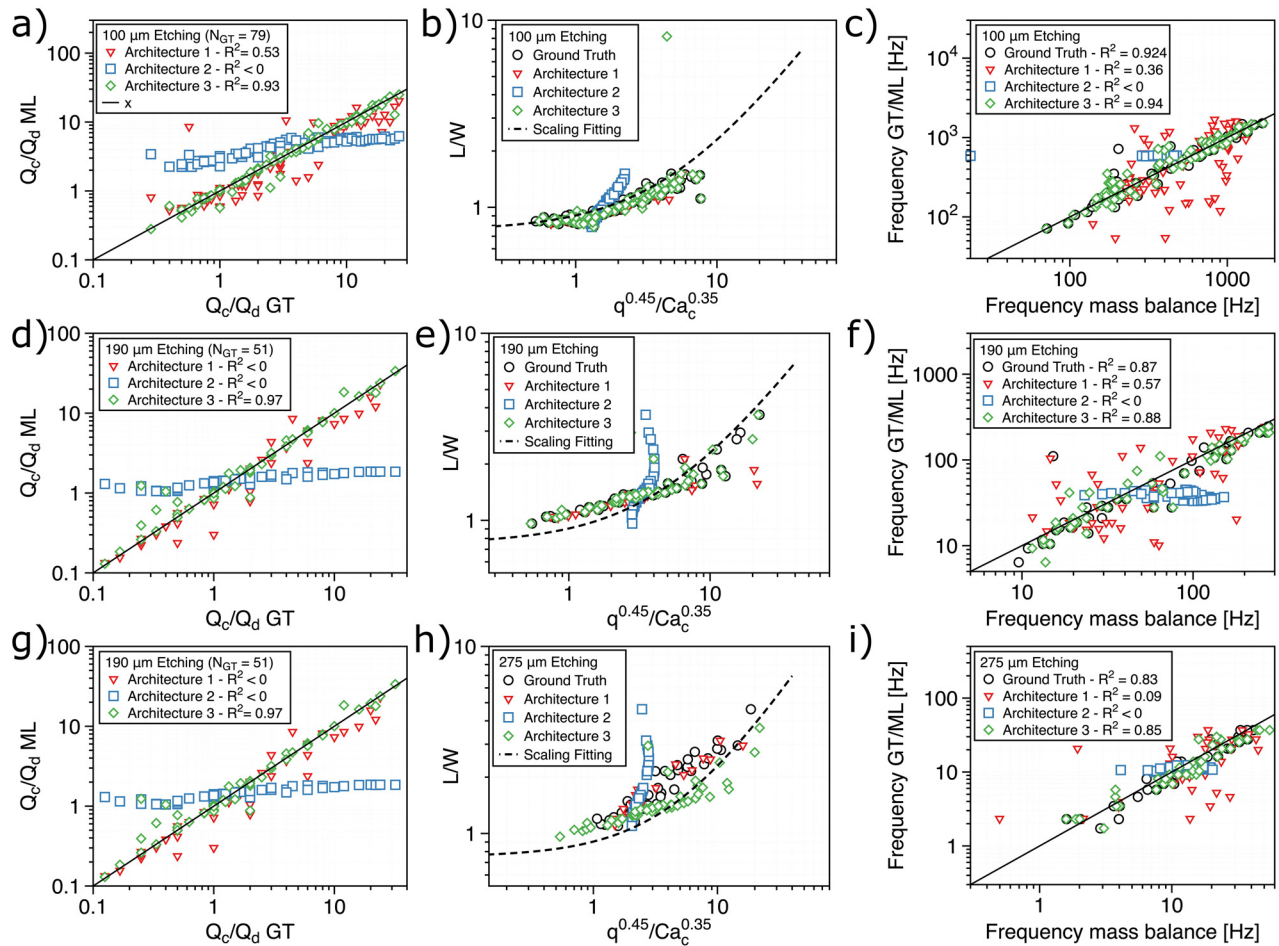
FIG. 4. Accuracy of the three different machine learning architectures in terms of flow rate ratio values Q_c/Q_d, droplet length over orifice width L/W and frequency of droplet generation f, for the three commercial microfluidic devices employed here, with etching depth of $100\,\mu m$ (a)–(c), $190\,\mu m$ (d)–(f), and $275\,\mu m$ (g)–(i). ML stands for machine learning and GT for ground truth.

frequency in the model training facilitated a more accurate predictions of the flow rate ratio required to achieve droplets of a given size.

Architecture 2 performed poorly, and this was also expected as this approach was based on a chained regression. Errors made by the algorithm when predicting Q_d would essentially accumulate, leading to a further deviation than typically expected when predicting Q_c. The number of ground truth data $N_{GT} = 79$ was low for the variety of phenomena experimentally observed: specifically, we could clearly observe a transition from droplets being arranged on two lines [Fig. 5(a)] to droplets being arranged on a single line [Fig. 5(b)], meaning that the overall number of dataset available for each scenario was low, thus causing architecture 2 to return incorrect predictions.

We also compared the machine learning predictions with the ground truth data and with the scaling in Eq. (1) [Fig. 4(b)]. The purpose of this numerical test was to both assess how well each of the models had trained along with comparing the data derived with the scaling previously reported in the literature.[12] We observed that the predicted data followed closely the scaling and the ground truth, which was also expected considering the fact that the scaling was found to describe better the ground truth data for the 100 μm etched device. Furthermore, we compared the data of droplet generation frequency obtained by three architecture with the mass balance in Eq. (2) [Fig. 4(c)], observing a behavior similar to our previous experiments, with architecture 3 being more accurate than the others, and with architecture 2 performing the worst among the three ($R^2 < 0$ for architecture 2).

We repeated the same investigation for the 190 [Figs. 4(d)–4(f)] and 275 μm [Figs. 4(g)–4(i)] etched devices. We again observed that architecture 3 was the best performing architecture. It is important to re-state here that the deviations observed between the empirical scaling and the machine learning predictions showed the need for a data driven approach, as the predictions as well varied from the theoretical scaling while being in close agreement with the ground truth data [Figs. 4(e) and 4(h)].

In summary, we observed that the predictions for architecture 3 were the most reliable with respect to the experimental ground truth data. This observation was rooted in the fact that architecture 3 employed the droplet generation frequency as an input parameter, thus somehow constraining the value of Q_d, and leading only to the prediction of Q_c. Architecture 2, based on chained regression, was found to perform the worst, and this was ascribed to the fact that this architecture was based on a chained approach with consequently larger errors. We were not too surprised by the supremacy of the neural network (architectures 1 and 3): due to the employment of a
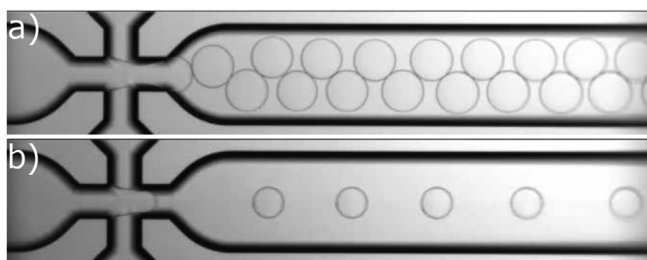
FIG. 5. (a) Droplet formation on two lines in the 100 μm etched device. (b) Droplet formation on a single line in the 100 μm etched device. Multimedia available online.

number of hidden layers, neural networks are often able to account for additional parameters that are not explicitly defined. Providing adequate steps are taken to prevent overfitting, the network is typically expected to perform well.

C. Avoiding overfitting problems: validation of the machine learning predictions using new experimental data

One of the key problems with supervised machine learning is overfitting,[47,48] where the experimental noise is computed within the architecture during the training, thus resulting in predictions that are only valid for specific experimental conditions and cannot be generalized to other systems, or even easily replicated across different laboratories because of experimental variability. This problem is particularly relevant in droplet microfluidics,[10] where the resulting droplet size, shape, and frequency are dependent among a large variety of parameters.[7,8] We already took some mitigation steps to address this problem, especially by choosing commercial devices, which are generally produced at scale in a highly controlled manner in order to guarantee homogeneity. However, it is important to see how the machine learning architectures would respond to experimental variability, in order to gain confidence that our predictions are not the result of overfitting. To this aim, we carried out the following new experimental campaign, carried out on different days from the original ground truth dataset. Depending on the architecture, we supplied the input parameters and then derived the values of Q_d and Q_c required to generate the droplets; then, we applied the resulting values to our experiments and then compared the predictions with the actual data [Fig. 6 and Figs. S1 and S2], thus simulating the real-life use of our algorithm.

We compared the predictions against the experimental data for the 190 μm etched device (Fig. 6). For each experiment (labeled with consecutive numbers on the x-axis), we compared the architecture value (white dashed bars) with the experimental observed value (red crossed bars), for the droplet length L [Fig. 6(a)], the droplet height H [Fig. 6(b)], and the droplet frequency f [Fig. 6(c)]. We did not compare the data with architecture 2, because it was the least performing, as already described in Sec. III B [Figs. 4(a)–4(c)]. For droplet length and height, we observed excellent agreement between predictions and observed data. For the frequency, instead, the minor disagreement observed for the frequency data was not uncommon in the literature, as other studies observed a similar trend.[34] Furthermore, it is well possible that not all the droplet sizes and frequency combinations were possible in our system, as we have constrained the microfluidic device geometry. It is, however, important to point out that the experimental parameters of Q_d and Q_c derived via architecture 3 were always returning value of the frequency f on the same order of magnitude compared to the observed values. We also repeated the same experimental campaign on the 100 μm etched device (Fig. S1) and on the 275 μm etched device (Fig. S2) and observed very good agreement. Note that we performed a larger number of experiments with architecture 3 to test the accuracy of the system in generating droplets having similar sizes but different generation frequencies.

Taken together, our results show that our training approach for three architecture does not suffer from overfitting, being the predictions reproducible on experiments performed on different days, thus accounting for some experimental variability. We also confirmed that
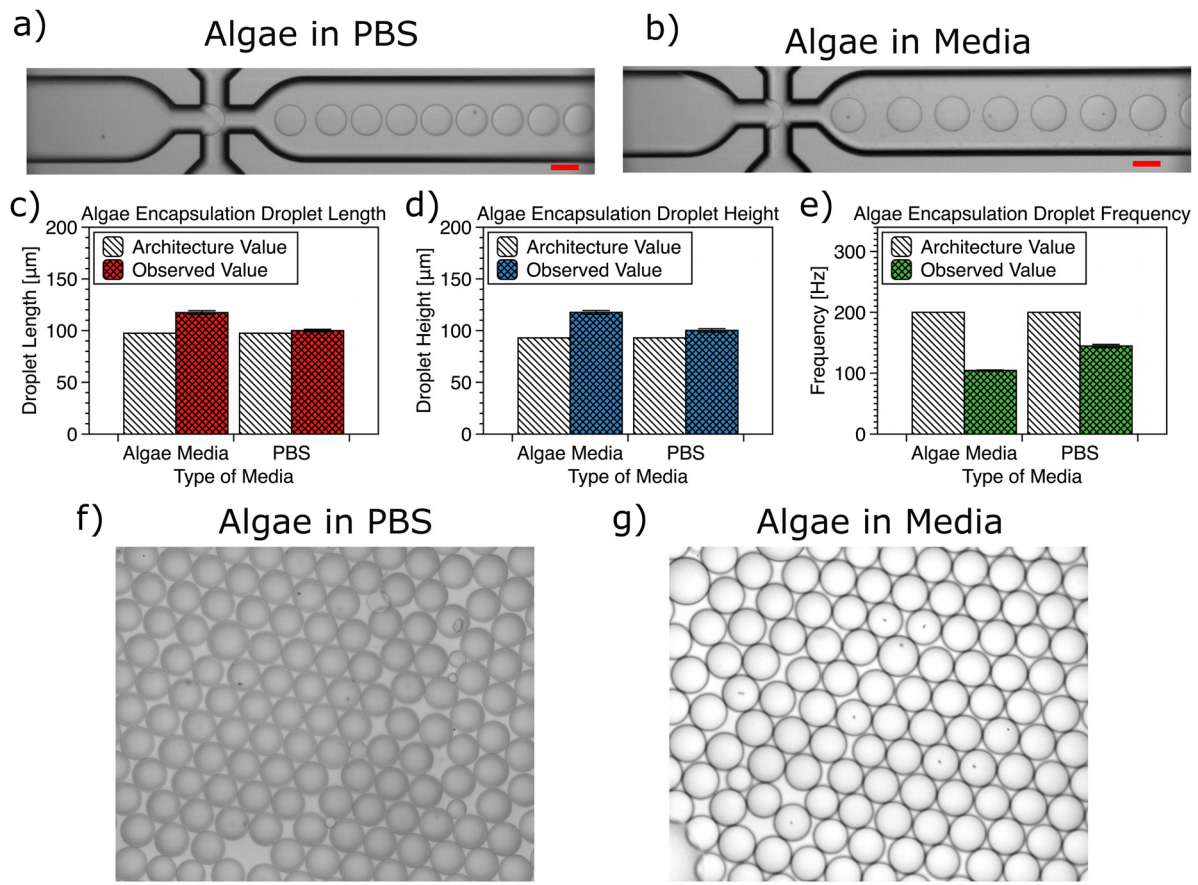
31 January 2025 10:26:27
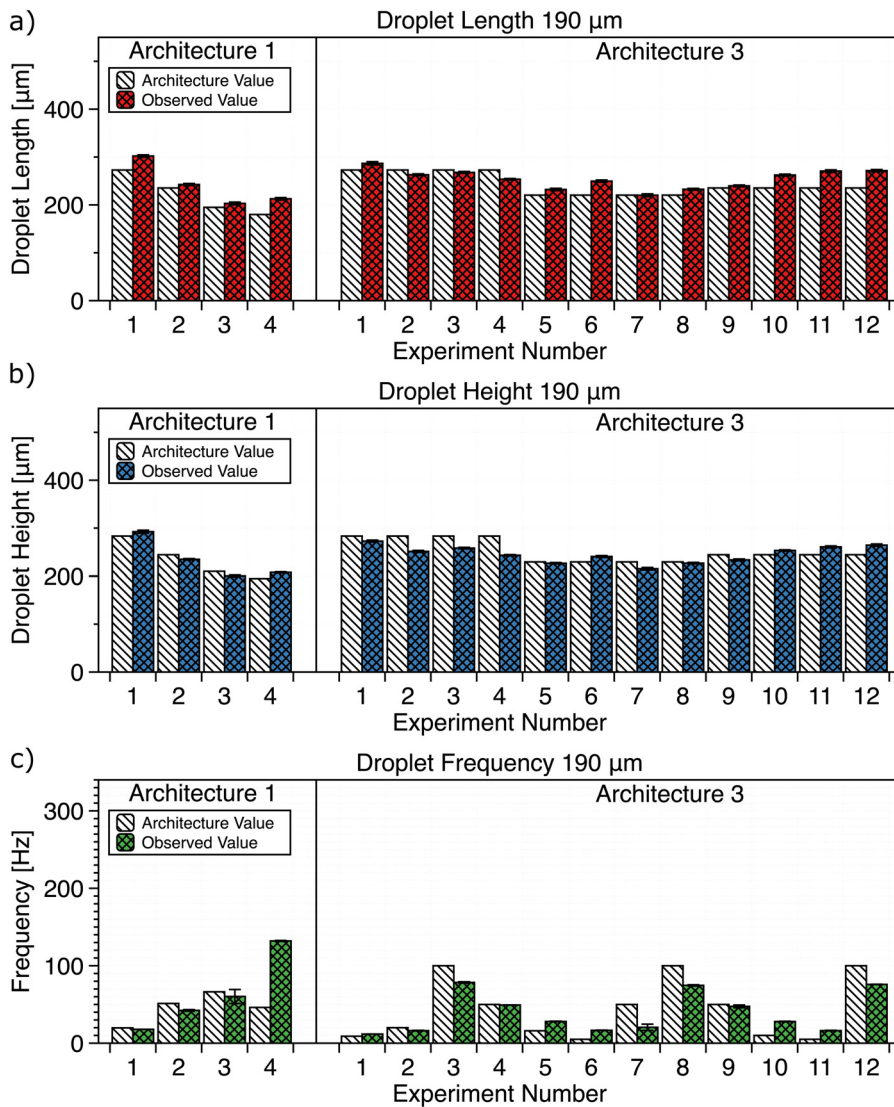
FIG. 6. Comparison between the specified values (dashed bars) and the experimental values (crossed bars) for droplet length L (a), droplet height H (b), and frequency f (c) for architectures 1–3, for the device with etching depth of 190 μ. The x-axis numbers refer to different experiments carried out on different days to test accuracy of the model against experimental variability.

architecture 3 is the best performing one for all the etched devices employed in this work.

D. Proof of concept for algae encapsulation

The formation of droplets in microfluidic devices has widely been used in the context of encapsulation applications.[3,49,50] In this context, the frequency of droplet formation has become an essential parameter to know before performing the experiment, so that such frequency could be synchronized to the frequency of equally-spaced objects approaching the encapsulation area to overcome the Poisson stochastic encapsulation limit.[44,46,51,52] The droplet size, instead, is a crucial parameters for microfluidic applications related to encapsulation and co-encapsulation of cells[53] together with other cells or functionalized particles.[54] As a proof of concept featuring the encapsulation of biological objects, we here employed architecture 3 to predict the droplet size and frequency in the 100 μm etched device

with the aim of encapsulating the *Tetraselmis suecica* algae (Fig. 7). We performed encapsulation experiments using both PBS [Fig. 7(a)] and Algae Media [Fig. 7(b)] as dispersed phase. The encapsulation of algae has been employed before to demonstrate intra-droplet growth,[35] as well as to elucidate inter-kingdom interactions in confined environments.[55] In our experiments, we clearly observed successful in-flow encapsulation at the flow-focusing microfluidic junction [Figs. 7(a) and 7(b)]. Notably, we compared architecture 3 values (the input parameters from which we obtained the values of Q_c and Q_d) and the ones observed in the experiments [Figs. 7(c)–7(e)]. For both PBS and Media, we observed good agreement on droplet sizes between the observed parameters and the input parameters in architecture 3. Most notably, we observed only a minor deviation when using the Algae media instead of PBS as the dispersed phase, despite the different value of interfacial tension measured for the two fluid combinations, in agreement with the literature.[56] Such observation can be justified in light of the weak dependency of the droplet length L with the Capillary

FIG. 7. Experimental snapshots for algae encapsulation in droplets made of phosphate buffer saline PBS (a), and algae media (b). Comparison between the specified values (dashed bars) and the experimental values (crossed bars) for droplet length L (c), droplet height H (d), and frequency f (e) using architecture 3, in both PBS and algae media. Experimental snapshots of algae encapsulated in droplets collected using a Sedgewick rafter counting chamber, for both PBS (f) and Media (g). Scale bars are 100 μm.

number Ca, as observed from the ground truth data [Fig. 3(b)], with $L \propto Ca^{-0.35}$. For the frequency, while the observed value fell within a reasonable order of magnitude of the given values, there were more significant discrepancies, in line with similar type of deviations previously observed in the literature.[34] In other words, our results are expected to be usable also for other type of dispersed aqueous Newtonian phases which not depart significantly from the PBS, for instance with similar viscosity and interfacial tension values. We collected downstream the droplets with the encapsulated algae [Figs. 7(f) and 7(g)] and observed that the droplets retained their size and shape for both PBS and algae media.

The take-home point from this section was the fact that we could use architecture 3 for the purpose of algae encapsulation, even when the dispersed phase presented some different properties compared to the training fluid (PBS).

E. Droplet microfluidic app to predict experimental parameters

We developed an application based on architecture 3 using the Rust Programing Language (Fig. 8). The user can either use our build

in model or insert their own trained model by activating the toggle on the top left of the panel: this is useful if the liquid combination departs significantly from the one employed in this work, especially in terms of viscosity and interfacial tension. On the top right, instead, the user is invited to define which etching depth the channel has, with choices among 100, 190, and 275 μm (which are the devices employed in this work). The users then insert the input data, namely, the droplet length L, the droplet height H, and the droplet generation frequency f. After inserting the input parameters, the app will return the predictions for the volumetric flow rate values of PBS (Q_d) and FluoSurf (Q_c). According to the results presented earlier in this manuscript, we expect the app to provide relatively accurate values even if the fluid combination is different from the one employed here, as long as the interfacial tensions between the two liquids is close to the 3 mN · m (measured here), and the viscosity values for the two phases are close to the ones reported in our work.

IV. CONCLUSIONS

In this work, we employed ML to predict the experimental parameters required for droplet generation in commercialized microfluidic devices using phosphate buffer saline (PBS) and fluorinated oil
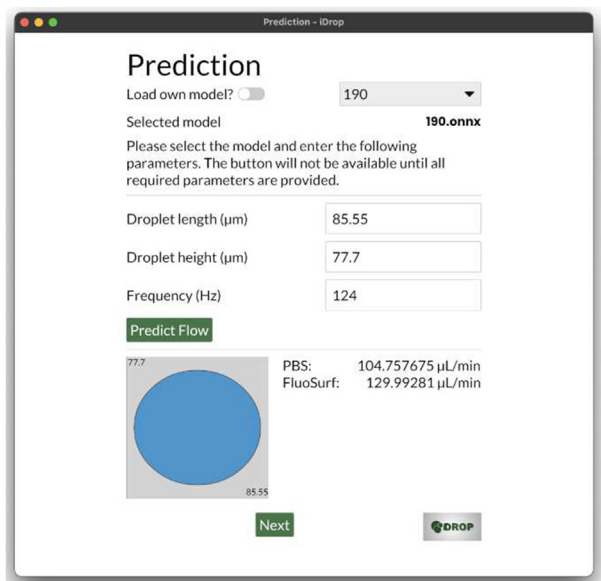
FIG. 8. User interface of the app developed in this study. The user is prompted to select which of the 3 channels they want to employ (top right rectangle). The user then selects a value for droplet length, droplet height and frequency. By clicking the "Predict Flow" button, a prediction for the volumetric flow rate values of PBS and FluoSurf can be obtained.

as dispersed and continuous liquid phases, respectively. We showed that the experimental dataset derived in the three microfluidic flow-focusing devices was robust enough to be employable as ground truth for the training of the machine learning algorithms. We were justified in employing a machine learning algorithm over a standard fitting procedure, as we could not clearly identify a mastercurve describing all our experimental results, with the data following the scaling only qualitatively. We observed that the predictions for architecture 3 were the most reliable with respect to the experimental ground truth data. This observation was rooted in the fact that architecture 3 employed droplet generation frequency as an input parameter, thus somehow constraining the value of Q_d, and leading only to the prediction of Q_c. Architecture 2 based on chained regression was found to perform the worst, and this was ascribed to the limited number of data available for the training. We performed a further experimental campaign demonstrate the validity of the predictions made by architecture 3 on different days, where experimental variation is a potential issue. We also demonstrated the use of architecture 3 predictions in the context of Algae encapsulation. Finally, we introduced an app where the users could select the device type, the droplet size, and generation rate to derive the predicted values of the liquid flow rates required.

We expect our results to be useful to the wider microfluidic community, being droplet microfluidic employed across a variety of different fields.[1–3] Future directions should include study of non-Newtonian droplet formation,[12,57,58] which has currently gained further attention in the context of controlled encapsulation.[44,46] Our results could also be employed to help tackling the problem of stochastic encapsulation, which prevent wide-spread adoption of such technologies within the medical setting.

SUPPLEMENTARY MATERIAL

See the supplementary material for Fig. S1 (predictions and experiments in the $100\,\mu$m channel) and Fig. S2 (predictions and experiments in the $275\,\mu$m channel).

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Claire Barnes: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Software (equal); Supervision (equal); Writing – review & editing (supporting). **Ashish Sonwane:** Software (lead). **Eva C. Sonnenschein:** Data curation (supporting); Resources (supporting); Writing – review & editing (supporting). **Francesco Del Giudice:** Conceptualization (equal); Data curation (equal); Formal analysis (equal); Funding acquisition (lead); Investigation (equal); Methodology (equal); Project administration (equal); Resources (equal); Supervision (equal); Validation (equal); Writing – original draft (lead); Writing – review & editing (lead).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request. The application reported in Fig. 8 is available in GitHub at Ref. 59.

REFERENCES

[1] Y. Geng, S. Ling, J. Huang, and J. Xu, "Multiphase microfluidics: Fundamentals, fabrication, and functions," Small **16**, 1906357 (2020).

[2] Y. Ding, P. D. Howes, and A. J. deMello, "Recent advances in droplet microfluidics," Anal. Chem. **92**, 132–149 (2020).

[3] F. Del Giudice, G. D'Avino, and P. L. Maffettone, "Microfluidic formation of crystal-like structures," Lab Chip **21**, 2069–2094 (2021).

[4] Z. Jiang, H. Shi, X. Tang, and J. Qin, "Recent advances in droplet microfluidics for single-cell analysis," TrAC Trends Anal. Chem. **159**, 116932 (2023).

[5] Y. Wang, M. Liu, Y. Zhang, H. Liu, and L. Han, "Recent methods of droplet microfluidics and their applications in spheroids and organoids," Lab Chip **23**, 1080 (2023).

[6] Y. Hou, S. Chen, Y. Zheng, X. Zheng, and J.-M. Lin, "Droplet-based digital PCR (ddPCR) and its applications," TrAC Trends Anal. Chem. **158**, 116897 (2023).

[7]J. Nunes, S. Tsai, J. Wan, and H. A. Stone, "Dripping and jetting in microfluidic multiphase flows applied to particle and fibre synthesis," J. Phys. D: Appl. Phys. **46**, 114002 (2013).

[8]P. Zhu and L. Wang, "Passive and active droplet generation with microfluidics: A review," Lab Chip **17**, 34–75 (2017).

[9]S. Battat, D. A. Weitz, and G. M. Whitesides, "Nonlinear phenomena in microfluidics," Chem. Rev. **122**, 6921–6937 (2022).

[10]K. S. Elvira, F. Gielen, S. S. Tsai, and A. M. Nightingale, "Materials and methods for droplet microfluidic device fabrication," Lab Chip **22**, 859–875 (2022).

[11]S. L. Anna, "Droplets and bubbles in microfluidic devices," Annu. Rev. Fluid Mech. **48**, 285–309 (2016).

[12]Q. Chen, J. Li, Y. Song, B. Chen, D. M. Christopher, and X. Li, "Pressure-driven microfluidic droplet formation in Newtonian and shear-thinning fluids in glass flow-focusing microchannels," Int. J. Multiphase Flow **140**, 103648 (2021).

[13]N. Paiboon, S. Surassmo, U. R. Ruktanonchai, and A. Soottitantawat, "Hydrodynamic control of droplet formation in narrowing jet and tip streaming regime using microfluidic flow-focusing," Int. J. Multiphase Flow **150**, 104013 (2022).

[14]Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature **521**, 436–444 (2015).

[15]J. Riordon, D. Sovilj, S. Sanner, D. Sinton, and E. W. Young, "Deep learning with microfluidics for biotechnology," Trends Biotechnol. **37**, 310–324 (2019).

[16]A. Isozaki, J. Harmon, Y. Zhou, S. Li, Y. Nakagawa, M. Hayashi, H. Mikami, C. Lei, and K. Goda, "AI on a chip," Lab Chip **20**, 3074–3090 (2020).

[17]S. Srikanth, S. K. Dubey, A. Javed, and S. Goel, "Droplet based microfluidics integrated with machine learning," Sens. Actuators, A **332**, 113096 (2021).

[18]A. V. Joshi, *Machine Learning and Artificial Intelligence* (Springer, Cham 2020).

[19]N. Nitta, T. Sugimura, A. Isozaki, H. Mikami, K. Hiraki, S. Sakuma, T. Iino, F. Arai, T. Endo, Y. Fujiwaki *et al.*, "Intelligent image-activated cell sorting," Cell **175**, 266–276 (2018).

[20]D. McIntyre, A. Lashkaripour, P. Fordyce, and D. Densmore, "Machine learning for microfluidic design and control," Lab Chip **22**, 2925 (2022).

[21]J. Zhang, S. Zhang, J. Zhang, and Z. Wang, "Machine learning model of dimensionless numbers to predict flow patterns and droplet characteristics for two-phase digital flows," Appl. Sci. **11**, 4251 (2021).

[22]M. Nazari, S. M. Varedi-Koulaei, and M. Nazari, "Flow characteristics prediction in a flow-focusing microchannel for a desired droplet size using an inverse model: Experimental and numerical study," Microfluid. Nanofluid. **26**, 26 (2022).

[23]S. Solanki, S. Lee, A. Jebakumar, J. Lum, M. Hamidi-Haines, C. Denison, M. Sundheim, K. Schauer, P. Stevenson, J. Hintzman *et al.*, "Machine learning for predicting microfluidic droplet generation properties," Comput. Fluids **247**, 105651 (2022).

[24]A. E. Siemenn, E. Shaulsky, M. Beveridge, T. Buonassisi, S. M. Hashmi, and I. Drori, "A machine learning and computer vision approach to rapidly optimize multiscale droplet generation," ACS Appl. Mater. Interfaces **14**, 4668–4679 (2022).

[25]J.-X. Wang, W. Yu, Z. Wu, X. Liu, and Y. Chen, "Physics-based statistical learning perspectives on droplet formation characteristics in microfluidic cross-junctions," Appl. Phys. Lett. **120**, 204101 (2022).

[26]L. Chagot, C. Quilodrán-Casas, M. Kalli, N. M. Kovalchuk, M. J. Simmons, O. K. Matar, R. Arcucci, and P. Angeli, "Surfactant-laden droplet size prediction in a flow-focusing microchannel: A data-driven approach," Lab Chip **22**, 3848–3859 (2022).

[27]M. Durve, A. Tiribocchi, F. Bonaccorso, A. Montessori, M. Lauricella, M. Bogdan, J. Guzowski, and S. Succi, "Droptrack-automatic droplet tracking with yolov5 and deepsort for microfluidic applications," Phys. Fluids **34**, 082003 (2022).

[28]O. O. Kartashov, S. V. Chapek, D. S. Polyanichenko, G. I. Belyavsky, A. A. Alexandrov, M. A. Butakova, and A. V. Soldatov, "Online microfluidic droplets characterization using microscope data intelligent analysis," Big Data Cognit. Comput. **7**, 7 (2023).

[29]M. Shahab and R. Rengaswamy, "Reinforcement-learning designs droplet microfluidic networks," Comput. Chem. Eng. **161**, 107787 (2022).

[30]M. G. Smith, J. Radford, E. Febrianto, J. Ramírez, H. O'Mahony, A. B. Matheson, G. M. Gibson, D. Faccio, and M. Tassieri, "Machine learning opens a doorway for microrheology with optical tweezers in living systems," AIP Adv. **13**, 075315-1–075315-12 (2023).

[31]W.-X. She, Z.-Y. Zuo, H. Zhao, Q. Gao, L.-X. Zhang, and X.-M. Shao, "Novel models for predicting the shape and motion of an ascending bubble in Newtonian liquids using machine learning," Phys. Fluids **34**, 043313-1–043313-12 (2022).

[32]D. J. S. Aulakh, S. B. Beale, and J. G. Pharoah, "A generalized framework for unsupervised learning and data recovery in computational fluid dynamics using discretized loss functions," Phys. Fluids **34**, 077111-1–077111-23 (2022).

[33]F. Ahmadi, M. Simchi, J. M. Perry, S. Frenette, H. Benali, J.-P. Soucy, G. Massarweh, and S. C. Shih, "Integrating machine learning and digital microfluidics for screening experimental conditions," Lab Chip **23**, 81–91 (2023).

[34]A. Lashkaripour, C. Rodriguez, N. Mehdipour, R. Mardian, D. McIntyre, L. Ortiz, J. Campbell, and D. Densmore, "Machine learning enables design automation of microfluidic flow-focusing droplet generation," Nat. Commun. **12**, 1–14 (2021).

[35]J. Pan, A. L. Stephenson, E. Kazamia, W. T. Huck, J. S. Dennis, A. G. Smith, and C. Abell, "Quantitative tracking of the growth of individual algal cells in microdroplet compartments," Integr. Biol. **3**, 1043–1051 (2011).

[36]A. Kalantarifard, E. A. Haghighi, and C. Elbuken, "Damping hydrodynamic fluctuations in microfluidic systems," Chem. Eng. Sci. **178**, 238–247 (2018).

[37]A. Kalantarifard, E. Alizadeh-Haghighi, and C. Elbuken, "A microfluidic droplet system for ultra-monodisperse droplet generation: A universal approach," Chem. Eng. Sci. **261**, 117947 (2022).

[38]C. N. Baroud, F. Gallaire, and R. Dangla, "Dynamics of microfluidic droplets," Lab Chip **10**, 2032–2045 (2010).

[39]R. R. Guillard, "Culture of phytoplankton for feeding marine invertebrates," in *Culture of Marine Invertebrate Animals: Proceedings-1st Conference on Culture of Marine Invertebrate Animals Greenport* (Springer, 1975), pp. 29–60.

[40]Y. Xia and G. M. Whitesides, "Soft lithography," Annu. Rev. Mater. Sci. **28**, 153–184 (1998).

[41]D. J. Guckenberger, T. E. De Groot, A. M. Wan, D. J. Beebe, and E. W. Young, "Micromilling: A method for ultra-rapid prototyping of plastic microfluidic devices," Lab Chip **15**, 2364–2378 (2015).

[42]S. Razavi Bazaz, O. Rouhi, M. A. Raoufi, F. Ejeian, M. Asadnia, D. Jin, and M. Ebrahimi Warkiani, "3D printing of inertial microfluidic devices," Sci. Rep. **10**, 5929 (2020).

[43]R. Su, F. Wang, and M. C. McAlpine, "3D printed microfluidics: Advances in strategies, integration, and applications," Lab Chip **23**, 1279 (2023).

[44]K. Shahrivar and F. Del Giudice, "Beating Poisson stochastic particle encapsulation in flow-focusing microfluidic devices using viscoelastic liquids," Soft Matter **18**, 5928 (2022).

[45]T. M. Squires and S. R. Quake, "Microfluidics: Fluid physics at the nanoliter scale," Rev. Mod. Phys. **77**, 977 (2005).

[46]K. Shahrivar and F. Del Giudice, "Controlled viscoelastic particle encapsulation in microfluidic devices," Soft Matter **17**, 8068–8077 (2021).

[47]T. Dietterich, "Overfitting and undercomputing in machine learning," ACM Comput. Surv. **27**, 326–327 (1995).

[48]X. Ying, "An overview of overfitting and its solutions," J. Phys.: Conf. Ser. **1168**, 022022 (2019).

[49]G. Aubry, H. J. Lee, and H. Lu, "Advances in microfluidics: Technical innovations and applications in diagnostics and therapeutics," Anal. Chem. **95**, 444–467 (2023).

[50]W. Zong, X. Shao, J. Li, Y. Chai, X. Hu, and X. Zhang, "Synthetic intracellular environments: From basic science to applications," Anal. Chem. **95**, 535–549 (2023).

[51]J. F. Edd, D. Di Carlo, K. J. Humphry, S. Köster, D. Irimia, D. A. Weitz, and M. Toner, "Controlled encapsulation of single-cells into monodisperse picolitre drops," Lab Chip **8**, 1262–1264 (2008).

[52]L. Li, P. Wu, Z. Luo, L. Wang, W. Ding, T. Wu, J. Chen, J. He, Y. He, H. Wang *et al.*, "Dean flow assisted single cell and bead encapsulation for high performance single cell expression profiling," ACS Sens. **4**, 1299–1305 (2019).

[53]J. Y. Tan, T. E. Saleski, and X. N. Lin, "The effect of droplet size on syntrophic dynamics in droplet-enabled microbial co-cultivation," Plos one **17**, e0266282 (2022).

[54]E. Z. Macosko, A. Basu, R. Satija, J. Nemesh, K. Shekhar, M. Goldman, I. Tirosh, A. R. Bialas, N. Kamitaki, E. M. Martersteck *et al.*, "Highly parallel

31 January 2025 10:26:27

genome-wide expression profiling of individual cells using nanoliter droplets," Cell **161**, 1202–1214 (2015).

[55] S. S. Terekhov, I. V. Smirnov, M. V. Malakhova, A. E. Samoilov, A. I. Manolov, A. S. Nazarov, D. V. Danilov, S. A. Dubiley, I. A. Osterman, M. P. Rubtsova *et al.*, "Ultrahigh-throughput functional profiling of microbiota communities," Proc. Nat. Acad. Sci. U. S. A. **115**, 9551–9556 (2018).

[56] H. A. Abdulbari and M. Zahera, "Investigating the effect of a monovalent ion on the droplet's size and distribution in a surfactant-free droplet generation microfluidic chip," J. Flow Chem. **12**, 31–39 (2022).

[57] Y. Ren, Z. Liu, and H. C. Shum, "Breakup dynamics and dripping-to-jetting transition in a Newtonian/shear-thinning multiphase microsystem," Lab Chip **15**, 121–134 (2015).

[58] W. Du, T. Fu, Q. Zhang, C. Zhu, Y. Ma, and H. Z. Li, "Breakup dynamics for droplet formation in a flow-focusing device: Rupture position of viscoelastic thread from matrix," Chem. Eng. Sci. **153**, 255–269 (2016).

[59] A. R. Sonwane and C. Barnes (2023), "iDrop," GitHub. https://github.com/asher-gh/iDrop/

31 January 2025 10:26:27