Data-Driven Real-Time Prediction for Interfacial Fluid Mechanics: Droplet Evaporation

Sahar Andalib (✉ sandalib@ucla.edu)

University of California, Los Angeles

Kunihiko Taira

University of California, Los Angeles

H. Pirouz Kavehpour

University of California, Los Angeles

Data-Driven Real-Time Prediction for Interfacial Fluid Mechanics: Droplet Evaporation

Sahar Andalib[a,*]**, Kunihiko Taira**[a]**, and H. Pirouz Kavehpour**[a]

[a]Department of Mechanical and Aerospace Engineering, University of California, Los Angeles, CA 90095
[*]sandalib@ucla.edu

ABSTRACT

Droplet evaporation plays crucial roles in biodiagnostics, microfabrication, and inkjet printing. Experimentally studying the evolution of a sessile droplet consisting of two or more components needs sophisticated equipment to control the vast parameter space affecting the physical process. On the other hand, non-axisymmetric nature of the problem, attributed to compositional perturbations, introduces challenges to numerical methods. In this work, droplet evaporation problem is studied from a new perspective. We analyze evolution of a sessile methanol droplet through data-driven classification and regression techniques. The models are trained using experimental data of methanol droplet evolution under various environmental humidity levels and substrate temperatures. At higher humidity levels, the interfacial tension and subsequently contact angle increase due to higher water uptake into droplet. Therefore, different regimes of evolution are observed due to adsorption-absorption and possibly condensation of water which turns the droplet into a binary system. We use classification algorithms to predict the regime of droplet with point-by-point analysis of droplet profile. Decision tree demonstrates a better performance compared to Naïve Bayes (NB) classifier. Furthermore, through utilizing regression techniques, we predict the humidity level surrounding droplet as well as time evolution of macroscopic parameter (diameter or contact angle) of droplet. The prediction results show promising performance for four cases of methanol droplet evolution under conditions that are unseen by the model which demonstrates the capability of the model to capture the complex physics underlying binary droplet evolution.

Introduction

Droplet evaporation is omnipresent in nature and engineering technologies such as biodiagnostics, inkjet printing, microfabrication, spray cooling, and agricultural irrigation[1–10]. Over the last two decades, evaporation of a liquid droplet has been studied extensively through theoretical, experimental, and numerical studies[11–22]. Although evolution of a single component droplet is mainly understood[23–27], there is limited knowledge in the complex physics for a multi-component or even a binary droplet involving absorption/adsorption and possibly condensation of a second component, Marangoni effect, multi-component diffusion, nucleation of microdroplets, and phase transition[28–39]. Such physics with numerous parameters in play make experimental studies sophisticated and time-consuming while requiring advanced equipment to finely control the environmental condition. On the other hand, highly non-asisymmetric nature of the problem due to compositional inhomogeneities brings up significant challenges to numerical models. In this study, we introduce a new perspective to analyze binary droplet evaporation through data-driven techniques. Our method shows a great potential in opening up new paths to analyze more complicated multi-component droplet evolution and interfacial fluid mechanics in general.

Machine learning methods have emerged as powerful tools for analyzing various fluid mechanics problems ranging from turbulence modeling to phase transition[40–54]. For example, image processing and pattern recognition techniques have been employed to indirectly measure the target components in blood droplets after evaporation[55–57]. In this work, we use data-driven classification and regression algorithms to analyze the real-time behavior of a methanol droplet at different levels of environmental humidity and temperature of substrate. Water uptake into droplet through adsorption-absorption and possibly condensation turns methanol droplet into a binary system. Based on environmental condition, droplet evolves in different regimes: evaporation-dominated, transition, or condensation-dominated.

The capability of the proposed model is evaluated in three ways. First, a classification algorithm is trained to predict the regime of droplet evaporation through analysis of diameter and contact angle evolution over time. The objective for the model is to detect the regime of droplet evolution with even a single data point at a specific time. Predicting the regime of droplet evolution is crucial for various industries. In many instances, the occurrence of one regime or the other should be avoided. For example, a droplet sitting on a surface forever is not desired for printing technologies or droplet-based biosensors. Second, a regression algorithm is utilized to detect the humidity of surrounding by analyzing droplet evolution. High hygroscopic nature of methanol, allows greater amount of water uptake into droplet in humid environments. The higher content of water in droplet increases the contact angle and alters the rate of change of volume. The regression model analyzes these changes and reversely
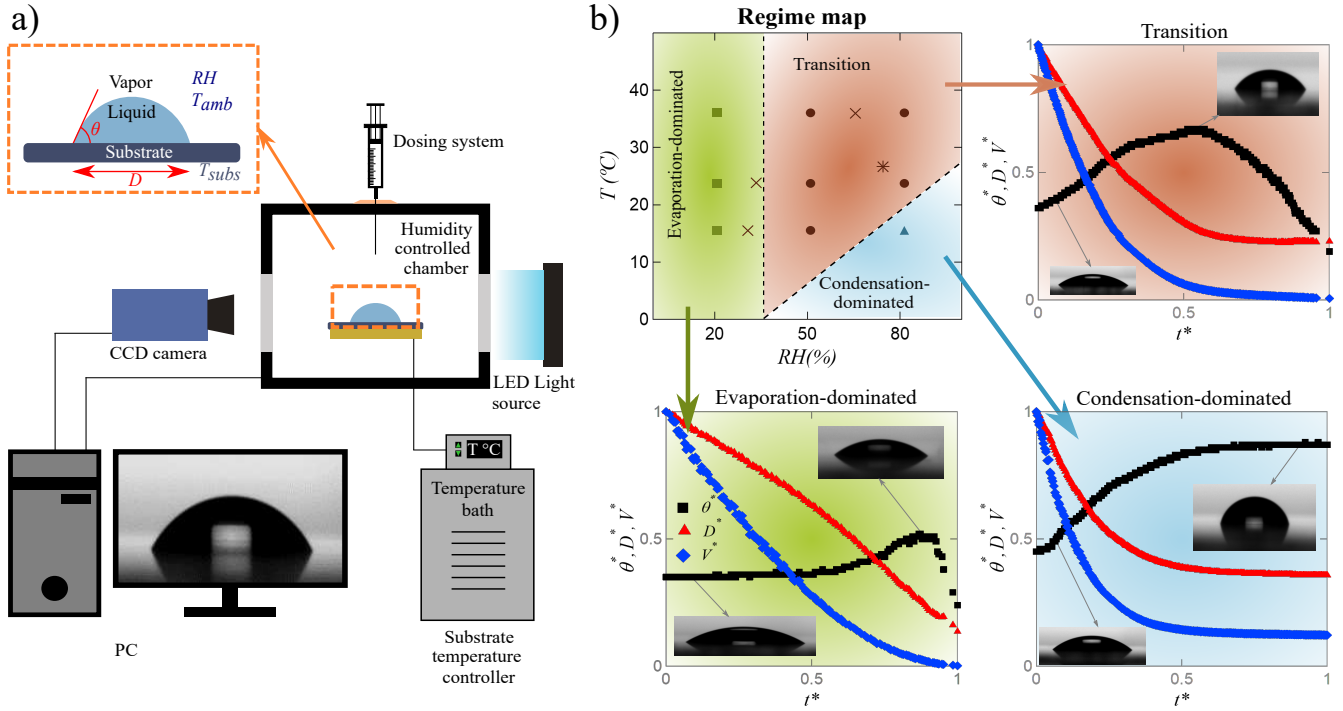
Figure 1. a) Schematic of the experimental setup with macroscopic parameters of droplet shown in the inset; b) droplet evolution regime map (top-left), evolutions of nondimensional contact angle, volume, and diameter versus time for evaporation-dominated (bottom-left), transition (top-right), and condensation-dominated (bottom-right) regimes.

predicts the humidity. Furthermore, given the condition of surrounding, the continuous evolution of macroscopic parameters of droplet, i.e., diameter and contact angle, is predicted.

Results

Physics of Droplet Evaporation

Droplet evaporation is influenced by numerous factors including liquid/substrate properties as well as environmental conditions[14,18,19,21,27,35,58]. We analyze the evolution of a sessile methanol droplet through macroscopic parameters: volume, V, diameter, D, contact angle, θ, and time, t, under controlled relative humidity of surrounding, RH, and substrate temperature, T (see Fig. 1a inset). The variables are nondimensionalized as: $t^* = t/t_f$, $V^* = V/V_0$, $D^* = D/D_0$, $\theta^* = \theta/\theta_0$, $T^* = T/T_0$, $RH^* = RH/RH_0$, where $\theta_0 = 90°, T_0 = 35°C, RH_0 = 100$, and t_f, V_0, D_0, that are experimentally measured, stand for total evaporation time, initial volume, and initial diameter, respectively. The experiments are conducted in a chamber with controlled humidity and on a substrate with controlled temperature (Fig. 1a). Fig. 1a Details of experimental procedure are given in Materials and Methods Sect. Three regimes of droplet evolution are observed under various relative humidity of the surrounding and substrate temperature (Fig. 1b-top left) namely: evaporation-dominated, transition, and condensation-dominated. Three sub figures represent the evolutions of θ^*, V^*, and D^* over t^*. Nondimensional plots are reported to better visualize different evolution patterns.

At low relative humidity (green-shaded region in Fig.1b), change in substrate temperature does not alter the qualitative evolution of droplet. In this regime, the contact angle stays constant for most of droplet lifetime followed by a slight increase and a sharp decrease towards the end (1b-bottom left). The modest rise in contact angle is attributed to the interplay of high evaporation rate of methanol and receding speed at the triple line[30]. Diameter and volume monotonically decrease during droplet lifespan.

Due to high hygroscopic nature of methanol, at higher relative humidity, water vapor transfers into the droplet at liquid-gas interface. Water adsorbing-absorbing and possibly condensing on the interface is reported in previous studies[28,32,35,58–60]. The growth in the concentration of water content changes the interfacial tensions and results in higher contact angle[32,35,60]. Unlike low relative humidity, substrate temperature plays a determining role in regime of droplet evolution at high relative humidity of surrounding. In transition regime (red-shaded region), contact angle rises to a maximum value before gradually decreasing towards the end of droplet lifetime. Increasing contact angle demonstrates water uptake into droplet while methanol
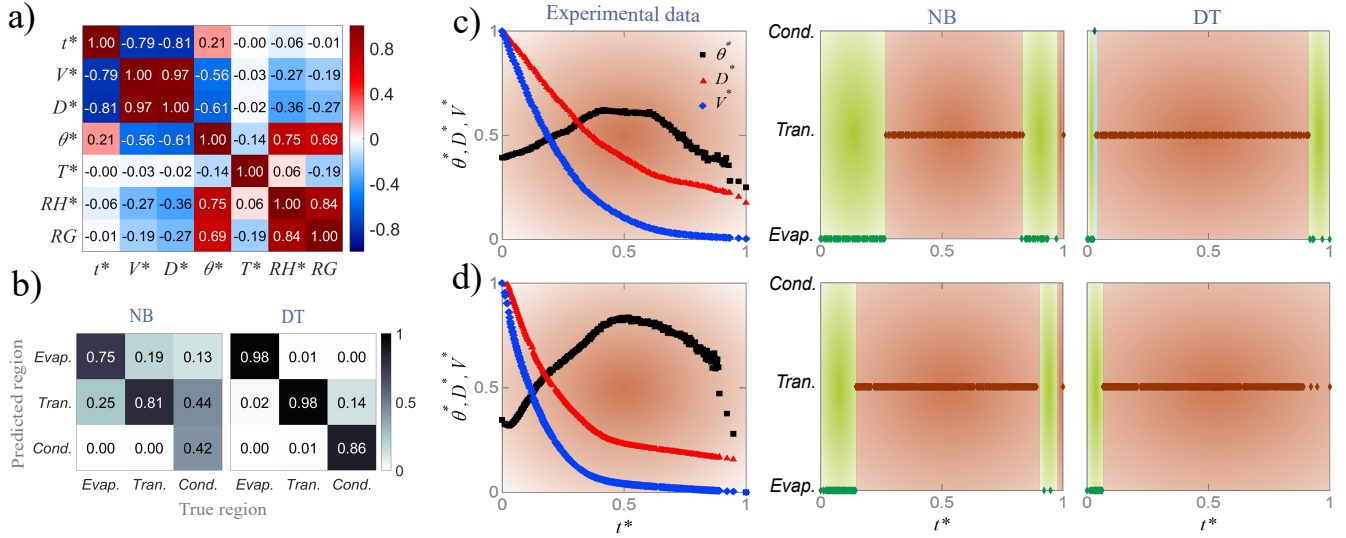
Figure 2. a) Correlation matrix for parameters in droplet evaporation; b) confusion matrices for the test set with Naïve Bayes (NB) and decision tree (DT) algorithms; c) experimental data of droplet evolution in validation set (RH = 80% and T = 35°C) and point-by-point validation results with NB and DT classifiers; d) experimental data for experiment 4 in prediction set (RH = 75% and T = 25°C) and point-by-point prediction results with NB and DT classifiers.

is evaporating. At the point of maximum contact angle, most of methanol has already evaporated and droplet consists mainly of water. However, some studies revealed that small amount of residual methanol remains until the end of droplet lifetime[30,31]. Even though diameter and volume decrease monotonically, two obvious slopes are observed in their evolutions (1b-top right). The two slopes correspond to two stages: the initial stage when merely methanol evaporates and the second stage when water mainly evaporates at a slower rate.

When the humidity of the environment is high and the substrate temperature is sufficiently low, another regime is observed. In condensation-dominated (blue-shaded) regime, contact angle monotonically increases until it reaches a plateau. Both diameter and volume converge to a non-zero value. Lower substrate temperature enhances water uptake through condensation by dropping the liquid-gas interface temperature below that of dew point[60]. In this regime, droplet comes to a quasi-steady state with a remaining droplet consisting mainly of water[30–32,35].

Regime Classification

A classification algorithm is trained to detect the droplet regime given the evolutions of contact angle and diameter at each specific point in time. Dependence of variables is shown by the correlation matrix in Fig. 2a where RG stands for the regime. Diameter and volume are coupled for a spherical cap sessile droplet through the relation $V = (\pi/3)(D/2)^3(2+\cos\theta)(1-\cos\theta)^2$ which assumes slow quasi-static evaporation. t^*, D^*, and θ^* are used as input variables and RG is the target variable. It is observed that the contact angle is highly proportional to humidity because the higher the humidity, the higher the amount of water uptake into drop. Higher water content increases the interfacial tension at the triple line which results in higher contact angles.

The framework characterizes the behavioral pattern of droplet by learning the values of contact angle and diameter at each specific time and classifying them to each regime. The model then labels the test and validation sets based on similar evolution observed previously during training. The ratio of training to test set is 80 to 20%. Two classifiers of Naïve Bayes (NB) and decision tree (DT) are trained and confusion matrices are used to compare the performance of classifiers on the test set (Fig. 2b). Precision, recall, F-score and overall accuracy values (shown in Table 1) provide a comprehensive evaluation on performance of each classifier on the test set. Based on the results shown in the Fig. 2b and Table 1, DT outperforms NB for all regimes of the test set. It is also observed that detection of condensation-dominated regime is challenging for both classifiers. For NB, around half (43%) of the points in condensation-dominated regime are classified as transition regime (see Fig. 2b). Replacing diameter with volume slightly improves the results for both classifiers (6% on average) for this regime. This is due to a more discernible evolution of V^* compared to D^* towards the end of droplet lifetime for this regime. However, since measuring diameter is a more direct approach and also more convenient for the user, the model is trained with diameter.

Performance of classifiers is evaluated for each specific condition through validation set. The model is validated using the data from a single experiment under specific condition that is held out during training/testing. Validation results, averaged for each condition, are presented in Table 2. It should be noted that the accuracy and recall are the same for validation because

Table 1. Performance of two different algorithms for regime classification: test set

Classifier	Accuracy	RG	Precision	Recall	F-score
		Evap.	0.74	0.65	0.6
NB	0.75	Trans.	0.81	0.78	0.79
		Cond.	0.46	0.93	0.61
		Evap.	0.99	0.97	0.97
DT	0.96	Trans.	0.97	0.96	0.96
		Cond.	0.83	0.89	0.85

Table 2. Performance of two different algorithms for regime classification: validation set

Classifier	$RH(\%)$	$15^\circ C$	$23^\circ C$	$35^\circ C$
	20	0.72($\pm$0.04)	0.73($\pm$0.08)	0.78($\pm$0.06)
NB	50	0.80($\pm$0.06)	0.62($\pm$0.11)	0.63($\pm$0.02)
	80	0.41($\pm$0.10)	0.88($\pm$0.06)	0.83($\pm$0.08)
	20	0.98($\pm$0.03)	0.96($\pm$0.08)	0.96($\pm$0.08)
DT	50	0.94($\pm$0.10)	0.94($\pm$0.03)	0.98($\pm$0.03)
	80	0.74($\pm$0.23)	0.97($\pm$0.05)	0.94($\pm$0.07)

at each condition there is only one true regime. Fig. 2c illustrates a sample of validation for $RH = 80\%$ and $T = 35^\circ C$. The classifier assigns a region for each point in time based on the value of contact angle and diameter. Both classifiers correctly detect the region of the majority of the data points although DT demonstrates a better performance. NB seems to struggle at the beginning and end of droplet lifetime. This issue is less pronounced for DT.

Once the model is trained, tested, and validated, its performance should be evaluated on a prediction data set from conditions that are unseen by the model and do not contribute to the model training, testing, and validation. The values of RH and T for these conditions are randomly selected in the range of 20%<RH<80% and $15^\circ C$<T<$35^\circ C$ (shown with cross and star marks on the regime map). The model classifies each experiment under each regime with different ratios as shown in Table 3. For example, experiment 1 with RH of 30% and T of $15^\circ C$ is close to the boundary of evaporation-dominated and transition regimes. With NB, 54% of the data points in this experiment are classified as evaporation-dominated regime and 46% as transition regime. This is expected due to the location of experiment 1 on the regime map. It should be noted that the lines on the regime map are approximate boundaries. Fig. 2d demonstrates regime prediction of all data points of experiment 4 with both classifiers. The shown evolution of contact angle is not similar to any of the evolutions shown in Fig. 1b subfigures. In fact, contact angle decreases at the beginning and then starts rising. NB and DT correctly classify 84 and 93% of the data points in experiment 4 to transition regime.

Relative Humidity Prediction

In this section, we show the ability of the model to detect the environmental humidity by analyzing the evolution of contact angle and diameter through regression algorithms. Polynomial regression with four different orders (linear, quadratic, third order, and fourth order) and regression tree are used for training. The coefficient of determination (R^2) increases from 0.66 for linear up to 0.93 for fourth order polynomial regression. The test results for all five regression methods are shown in Fig. 3a. The horizontal axis shows the true value of RH^* and the vertical axis shows the predicted values averaged over all the points for each RH^*. As it can be seen, the higher the order of polynomial regression, the closer the average prediction to the ground truth value and the smaller the error bar. Furthermore, regression tree performs more accurately compared to all polynomial regression methods. Model performance under each specific condition through validation set is shown in Fig. 2b. The consistent

Table 3. Regime prediction for four experiments under new condition with NB and DT classifiers

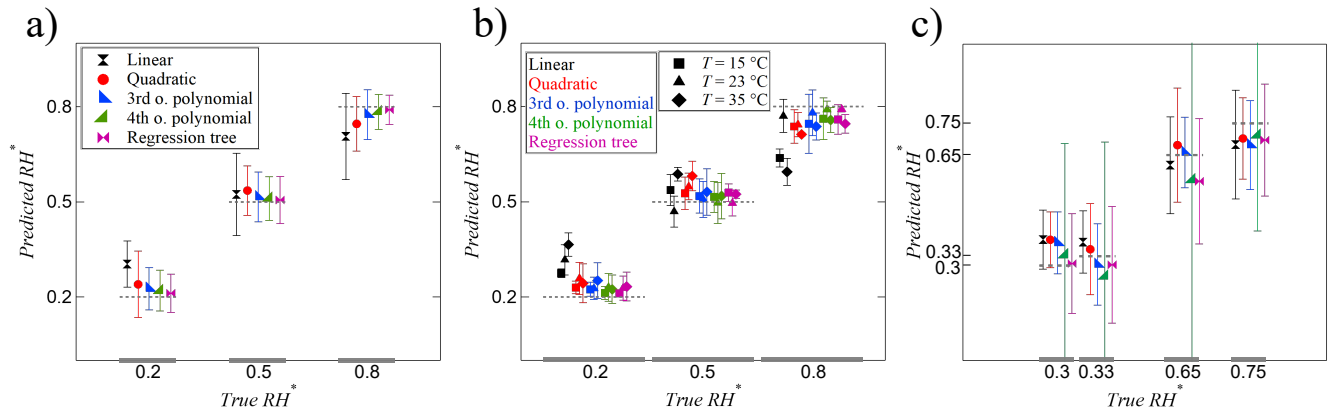| Experiment	$RH(\%)$	$T(^\circ C)$	NB Evap.	NB Trans.	DT Evap.	DT Trans.
1(X)	30	15	0.54	0.46	0.67	0.33
2(X)	33	23	0.61	0.39	0.79	0.21
3(X)	65	35	0.32	0.68	0.16	0.84
4(*)	75	25	0.18	0.82	0.07	0.93

Figure 3. Relative humidity regression results: a) test set; b) validation; c) prediction. Markers for all colors in (b) represent different temperatures as shown in legend. Markers and colors in (c) are the same as in (a).

colors throughout Fig. 3a-c represent different regression methods and different markers are used for each substrate temperature in Fig. 3b. Based on the results shown in Fig. 3b, all methods except linear regression produce reasonably accurate results. The validation results get closer to actual values as the order of the polynomial regression increases. Nonetheless, it must be noted that higher order polynomial regression increases the computational cost as well as the chance of over-fitting. Performance of regression tree is comparable to third and fourth order polynomial. The capability of different regression methods to predict new humidity values by analyzing the time evolution of contact angle and diameter is presented in Fig. 3c. The new relative humidity values (30, 33, 65, and 75%) are randomly selected in the range of 20-80%. It is noteworthy that the model has not seen any data of droplet evolution under these *RH* values during training, testing, or validation. It is seen that unlike testing and validation where increasing the order of polynomial or complexity of the model (i.e., regression tree) produces more accurate results, higher order polynomials do not result in better prediction of unseen conditions. As a matter of fact, linear, quadratic, and third order polynomials predict more accurately. This is a common issue when the model fits the training data very well and it negatively affects the model performance on the new data set. Fig. 3c clearly indicates over-fitting with fourth order polynomial.

Diameter and Contact Angle Prediction

In this section, the capability of the model to predict the continuous evolution of contact angle and diameter over time is evaluated through regression algorithms. The input variables include T^*, RH^*, t^*, and θ^* (or D^*) and the target variable is D^* (or θ^*). By increasing the order of polynomial, the coefficient of determination, R^2, for training improves from 0.87 to 0.99, and from 0.78 to 0.96 for diameter and contact angle prediction, respectively. The performance of five different regression methods on the test set is presented in Fig. 4a. First row represents the results when D^* is the target variable and the second row illustrates the results for θ^* as the target variable. The closer the distribution of data to the diagonal line in these plots, the better the performance of the model on the test set. Based on the results shown in Fig. 4a., the diameter test results saturate after third order polynomial while for contact angle the performance keeps improving when increasing the degree of polynomial from third to fourth.

The validation results are summarized in Fig. 4b. for D^* (left) and θ^* (right) in terms of R^2 and root mean square error *rmse*. With D^* being the target variable, an average R^2 of 0.8 or higher and average *rmse* less than 0.1 are achieved for all nine conditions. Going from linear to quadratic to third order polynomial increases and decreases the value of R^2 and *rmse*, respectively. The profiles of R^2 and *rmse* exhibit saturation, and further increase in the order of the polynomial does not improve model performance on validation data. This is consistent with the test results where model performance saturate at third order. Furthermore, regression tree demonstrates accuracy comparable to third and fourth order polynomials. By comparing the range of axes in Fig. 4b- right with left, it is obvious that R^2 values are generally lower (hardly reaching 0.7) and error is higher when θ^* is the target variable. In fact, there are a few instances where the average R^2 turns negative, suggesting that the overall prediction of the model is worse than a prediction with constant average value.

The performance of the model on predicting evolution of θ^* and D^* versus time under four new conditions that did not contribute to model training, testing, or validation as shown in Fig. 4c. One value of R^2 and *rmse* is reported for each condition (or experiment) which shows the overall quality of the fit. Higher coefficients of determination and lower *rmse* values demonstrate the better performance of the model in predicting D^* then θ^* evolution. Based on the results shown in Fig. 4c, third order polynomial regression has the best performance in predicting diameter. The results become less accurate with

fourth order polynomial which suggests over-fitting. It is interesting to note that regression tree, which had higher accuracy during testing and validation, is outperformed even by linear regression during prediction. Evolution of diameter versus time predicted by quadratic regression for Experiment 3 is depicted in Fig. 4d. As it can be seen, even with a quadratic regression, the model predicts the evolution of D^* quiet accurately for an unseen condition. Considering the range of values on axes of Fig. 4c-left and right, predicting the evolution of contact angle is more challenging for the model. Unlike predicting diameter, increasing the order of polynomials has negligible effect. The accuracy of the model stays almost constant for linear, quadratic, and third order. However, it worsens drastically for fourth order polynomial due to over-fitting the data. The R^2 and *rmse* values for fourth order polynomial fall outside the range shown in the plot. Since prediction of θ^* is generally more challenging for the model, the effect of over-fitting is more noticeable compared to D^* prediction. The overall better performance of the model for diameter prediction compared to contact angle prediction is due to the fact that diameter evolution is relatively smooth and therefore more predictable where contact angle evolution changes substantially under different conditions. Fig. 4d-right illustrates contact angle evolution over time predicted with third order polynomial for Experiment 1. There is a good agreement during most of droplet lifetime except for the region of steep rise in θ^* where the model underestimates the value of θ^*. Nevertheless, it is important to note that the corresponding time of the maximum θ^* is predicted accurately.

Discussion

In this study, we have analyzed the complex evolution of a binary sessile droplet through machine learning, classification and regression algorithms. Point-by-point analysis of droplet profile enables real-time predictions given a limited number of data points. This means that the model does not need the entire evolution profile of a droplet to make a prediction. Instead, only a few (or even a single) data points are (is) sufficient for prediction, although more data points result in more accurate prediction. The model prediction capability is then assessed on the data that do not contribute to training, testing, or validation.

Knowledge on droplet evolution regime is necessary for compatible designs in numerous industries such as droplet-based biosensors or ink-jet printing. We have shown the ability of the model to detect the regime of droplet evolution by a simple easy-to-interpret algorithm, i.e., NB, as well as using a more robust algorithm, i.e., DT. As expected, DT outperforms NB due to a more robust internal structure in the expense of computational cost and transparency. Both classifiers showed impressive performance (minimum 54% accuracy) on predicting the regime of droplet.

Furthermore, regression techniques are used to predict the humidity level surrounding droplet as well as time evolution of diameter (or contact angle) of droplet. Polynomial regressors as well as regression tree were trained through point-by-point analysis of droplet evolution. The model performance improved by increasing the order of the polynomial and using regression tree for training, test, and validation sets. However, when predicting the new conditions unseen by the model, fourth order polynomial and regression tree suffered from data over-fitting. The best performance of the model is achieved by third order polynomial. In general, the model prediction results are more accurate when predicting diameter evolution compared to contact angle prediction. This is due to smoother hence more predictable evolution of diameter with time. The sharp changes in θ^* under different conditions makes the prediction of its evolution challenging for the model. Information of this type is of great importance for technologies such as ink-jet printing, or droplet-based biodiagnostics where the predictions can provide critical information on the base diameter or contact angle of the droplet at a specific time.

In the current work, binary droplet evolution is studied through data-driven techniques. The model demonstrated promising performance detecting the regime of droplet evolution, the humidity level surrounding droplet, and time evolution of diameter and contact angle. The results of the current study demonstrate the potential of utilizing machine learning algorithms to better analyze interfacial fluid mechanics. Our preliminary study opens up new ways to study binary or multi-component droplet evolution which might lead into better analyzing the complex physics of the problem.

Methods

Experimental Setup

The experimental data is generated by methanol (Fisher Scientific, 99.8% purity) droplet evaporating in a chamber ($127 \times 127 \times 76\ mm^3$) under controlled humidity. The temperature of the substrate is controlled via an electrical system and a temperature bath (Fig. 1a). For training, testing, and validation set, nine conditions are created by a combination of relative humidity of the chamber and substrate temperature. The relative humidity values are: 20,50, and 80% and the values of substrate temperature are: 15, 23 (room temperature), and 35°C. Methanol droplet is deposited on a glass slide, coated with a thin layer of polydimethylsiloxane (PDMS), by using a dosing system which is fitted in the chamber through a small hole at the top. The transparent windows in the sides of the chamber allow for illumination with LED light source and recording of the droplet evolution by a CCD camera. The setup is placed on an anti-vibration table to eliminate environmental disturbances. The prediction set includes data of methanol droplet evolution under conditions that are unseen by the model during training,
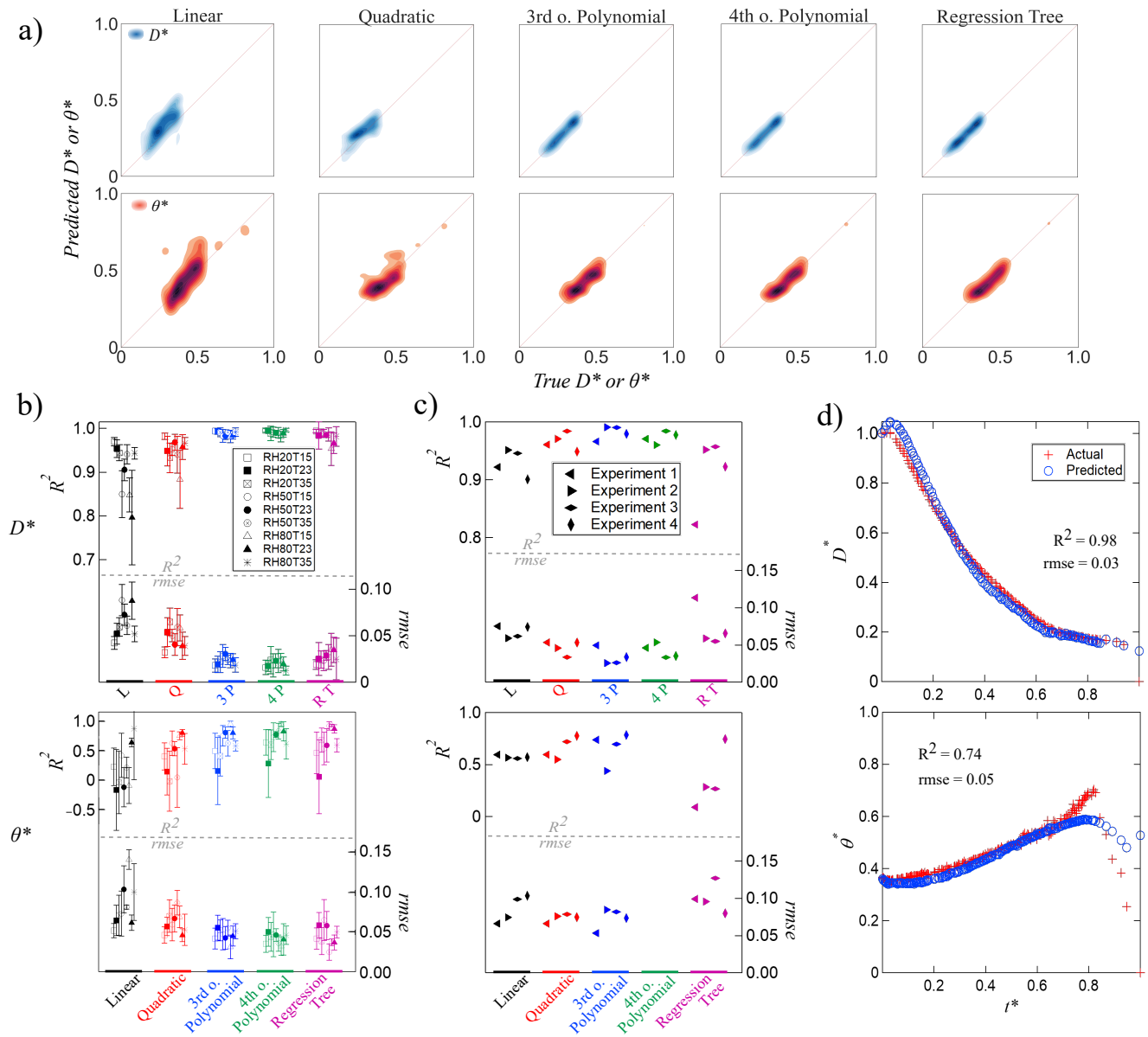
Figure 4. Diameter (D*) and contact angle (θ^*) regression results: a) test set; b) validation; c) prediction set; d) diameter prediction with quadratic regression for E3 (left) and contact angle prediction with 3rd order polynomial regression for E1 (right)

testing, and validation. These data are generated under new conditions with *RH* and *T* values randomly selected in the range of 20%<*RH*<80% and 15°*C*<*T*<35°*C*.

Data Acquisition

Once the relative humidity in the chamber and substrate temperature is set to desired values and enough time is passed to create quasi-steady state, the recording begins and the droplet is deposited on the substrate. The substrate is coated with hydrophobic layer in order to achieve measurable contact angle and reproducible results. In order to confirm that methanol did not interact with the coating on the substrate, multiple droplets are repeatedly deposited at the same location. No change is observed in their initial contact angle or evolution during evaporation. The deposited droplet volumes are all smaller than 10 μl to keep the droplet size lower than the capillary length. The corresponding volume for the capillary length for methanol is around 45 μl. A CCD camera is used to record the evolution of droplet at 50 frames per second. Post processing characterizes the behavioral pattern of droplet through the time evolution of macroscopic parameters i.e. contact angle (θ) and base diameter (D) (see Fig.1a inset).

Data Partitioning and Processing

The data for training, testing, and validation includes total number of 10,850 data points generated by 60 experiments of methanol droplet evolution under nine conditions. The nine conditions are created by a combination of three levels of relative humidity: low (20%), medium (50%), and high (80%) in combination with three substrate temperatures: 15, 23 (room temperature), and 35 °*C*. For each model run, one experiment (validation set) is excluded from training and testing data in order to specifically evaluate the model performance on a particular RH and T condition. The remaining data (i.e. 59 experiments) are partitioned into a training and a test set with 80-20% ratio using a 10-fold cross validation. Once the model is trained and tested, it is then examined on the validation set. This process is repeated 60 times until each experiment is kept out once and validated. The validation results are then averaged for each of *RH* and *T* condition. Validation demonstrates the performance of the model for each specific condition of humidity and substrate temperature. After this step, the predicting capability of the model is evaluated on the prediction set. The prediction set includes the data from droplet evaporation under conditions that are unseen by the model and do not contribute to model training, testing, or validation. The prediction set consists of a total of 761 data points generated by four experiments under conditions that are new to the model. The values of *RH* and *T* for these conditions are randomly selected.

Performance Criteria

The performance criteria of the model is reported by standard metrics. For classification, the confusion matrix of the model is used as well as the precision, recall, F-score, and overall accuracy values for each regime. For validation, the accuracy values are the same as recall values for each combination of *RH* and *T* because there is only one true regime for each validation set.

For regression methods, determination of coefficient (R^2) is reported to show how good the model fits the data in the training set. For *RH* testing, validation, and prediction, the actual *RH* of the environment is compared against the predicted value of *RH*. When D^* (or θ^*) is the target variable, the test results are demonstrated as predicted values versus actual values. The validation and prediction results are reported by R^2 that shows the quality of the fit; the proportion of variance in the target variable which is predictable from the input variables; and root mean squared error (*rmse*) which represents how much the prediction is off on average when predicting the average target variable.

Classifiers

We have used Naïve Bayes Classifier[61–64] as a simple and easy-to-interpret algorithm. Since the algorithm is simple, there is less chance for over-fitting the data, it is faster and needs smaller memory footprint. However, the restrictive underlying assumptions compromises its accuracy for real case scenarios when the variables are not fully independent of each other.

Bagged Decision Tree[65–67] with 250 trees is also used. It is a powerful classifier with built-in support for cross validation and a specialized function to measure feature importance. However it results in complex models that are not very transparent. It is often hard to understand how it makes predictions.

References

1. De Gans, B.-J., Duineveld, P. C. & Schubert, U. S. Inkjet printing of polymers: state of the art and future developments. *Adv. Mater.* **16**, 203–213 (2004).

2. Kawase, T., Sirringhaus, H., Friend, R. H. & Shimoda, T. Inkjet printed via-hole interconnections and resistors for all-polymer transistor circuits. *Adv. Mater.* **13**, 1601–1605 (2001).

3. Park, J. & Moon, J. Control of colloidal particle deposit patterns within picoliter droplets ejected by ink-jet printing. *Langmuir* **22**, 3506–3513 (2006).

4. Smith, P. J. & Shin, D. H. *Inkjet-based micromanufacturing* (John Wiley & Sons, 2012).

5. Shedd, T. & Pautsch, A. Spray impingement cooling with single-and multiple-nozzle arrays. part ii: Visualization and empirical models. *Int. J. Heat Mass Transf.* **48**, 3176–3184 (2005).

6. Sodtke, C. & Stephan, P. Spray cooling on micro structured surfaces. *Int. J. Heat Mass Transf.* **50**, 4089–4097 (2007).

7. Molle, B., Tomas, S., Hendawi, M. & Granier, J. Evaporation and wind drift losses during sprinkler irrigation influenced by droplet size distribution. *Irrigation Drainage* **61**, 240–250 (2012).

8. Hurth, C. *et al.* Biomolecular interactions control the shape of stains from drying droplets of complex fluids. *Chem. Eng. Sci.* **137**, 398–403 (2015).

9. Trantum, J. R., Wright, D. W. & Haselton, F. R. Biomarker-mediated disruption of coffee-ring formation as a low resource diagnostic indicator. *Langmuir* **28**, 2187–2193 (2012).

10. Gulka, C. P. *et al.* Coffee rings as low-resource diagnostics: detection of the malaria biomarker plasmodium falciparum histidine-rich protein-ii using a surface-coupled ring of ni (ii) nta gold-plated polystyrene particles. *ACS Appl. Mater. Interfaces* **6**, 6257–6263 (2014).

11. Deegan, R. D. *et al.* Capillary flow as the cause of ring stains from dried liquid drops. *Nature* **389**, 827–829 (1997).

12. Hu, H. & Larson, R. G. Evaporation of a sessile droplet on a substrate. *J. Phys. Chem. B* **106**, 1334–1344 (2002).

13. Popov, Y. O. Evaporative deposition patterns: spatial dimensions of the deposit. *Phys. Rev. E* **71**, 036313 (2005).

14. Ristenpart, W., Kim, P., Domingues, C., Wan, J. & Stone, H. A. Influence of substrate conductivity on circulation reversal in evaporating drops. *Phys. Rev. Lett.* **99**, 234502 (2007).

15. Ledesma-Aguilar, R., Vella, D. & Yeomans, J. M. Lattice-boltzmann simulations of droplet evaporation. *Soft Matter* **10**, 8267–8275 (2014).

16. Yang, K., Hong, F. & Cheng, P. A fully coupled numerical simulation of sessile droplet evaporation using arbitrary lagrangian–eulerian formulation. *Int. J. Heat Mass Transf.* **70**, 409–420 (2014).

17. Mollaret, R., Sefiane, K., Christy, J. & Veyret, D. Experimental and numerical investigation of the evaporation into air of a drop on a heated surface. *Chem. Eng. Res. Des.* **82**, 471–480 (2004).

18. Bhardwaj, R., Longtin, J. P. & Attinger, D. A numerical investigation on the influence of liquid properties and interfacial heat transfer during microdroplet deposition onto a glass substrate. *Int. J. Heat Mass Transf.* **50**, 2912–2923 (2007).

19. David, S., Sefiane, K. & Tadrist, L. Experimental investigation of the effect of thermal properties of the substrate in the wetting and evaporation of sessile drops. *Colloids Surfaces A: Physicochem. Eng. Aspects* **298**, 108–114 (2007).

20. Bonn, D., Eggers, J., Indekeu, J., Meunier, J. & Rolley, E. Wetting and spreading. *Rev. Mod. Phys.* **81**, 739 (2009).

21. Sobac, B. & Brutin, D. Thermal effects of the substrate on water droplet evaporation. *Phys. Rev. E* **86**, 021602 (2012).

22. Lohse, D. & Zhang, X. Physicochemical hydrodynamics of droplets out of equilibrium. *Nat. Rev. Phys.* **2**, 426–443 (2020).

23. Cazabat, A.-M. & Guena, G. Evaporation of macroscopic sessile droplets. *Soft Matter* **6**, 2591–2612 (2010).

24. Shahidzadeh-Bonn, N., Rafai, S., Azouni, A. & Bonn, D. Evaporating droplets. *J. Fluid Mech.* **549**, 307 (2006).

25. Gelderblom, H. *et al.* How water droplets evaporate on a superhydrophobic substrate. *Phys. Rev. E* **83**, 026306 (2011).

26. Sefiane, K. & Tadrist, L. Experimental investigation of the de-pinning phenomenon on rough surfaces of volatile drops. *Int. Commun. Heat Mass Transf.* **33**, 482–490 (2006).

27. Sefiane, K., Wilson, S. K., David, S., Dunn, G. J. & Duffy, B. R. On the effect of the atmosphere on the evaporation of sessile droplets of water. *Phys. Fluids* **21**, 062101 (2009).

28. Sefiane, K., Tadrist, L. & Douglas, M. Experimental study of evaporating water–ethanol mixture sessile drop: influence of concentration. *Int. J. Heat Mass Transf.* **46**, 4527–4534 (2003).

29. Christy, J. R., Hamamoto, Y. & Sefiane, K. Flow transition within an evaporating binary mixture sessile drop. *Phys. Rev. Lett.* **106**, 205701 (2011).

30. Sefiane, K., David, S. & Shanahan, M. E. Wetting and evaporation of binary mixture drops. *J. Phys. Chem. B* **112**, 11317–11323 (2008).

31. Liu, C., Bonaccurso, E. & Butt, H.-J. Evaporation of sessile water/ethanol drops in a controlled environment. *Phys. Chem. Chem. Phys.* **10**, 7150–7157 (2008).

32. Chen, P. *et al.* Evaporation of binary sessile drops: infrared and acoustic methods to track alcohol concentration at the interface and on the surface. *Langmuir* **32**, 9836–9845 (2016).

33. Tan, H. *et al.* Evaporation-triggered microdroplet nucleation and the four life phases of an evaporating ouzo drop. *Proc. Natl. Acad. Sci.* **113**, 8642–8647 (2016).

34. Diddens, C. *et al.* Evaporating pure, binary and ternary droplets: thermal effects and axial symmetry breaking. *J. Fluid Mech.* **823**, 470–497 (2017).

35. Kita, Y. *et al.* Quantifying vapor transfer into evaporating ethanol drops in a humid atmosphere. *Phys. Chem. Chem. Phys.* **20**, 19430–19440 (2018).

36. Li, Y. *et al.* Evaporation-triggered segregation of sessile binary droplets. *Phys. review letters* **120**, 224501 (2018).

37. Li, Y. *et al.* Gravitational effect in evaporating binary microdroplets. *Phys. review letters* **122**, 114501 (2019).

38. Li, Y. *et al.* Evaporating droplets on oil-wetted surfaces: Suppression of the coffee-stain effect. *Proc. Natl. Acad. Sci.* **117**, 16756–16763 (2020).

39. Diddens, C., Li, Y. & Lohse, D. Competing marangoni and rayleigh convection in evaporating binary droplets. *J. Fluid Mech.* **xx**, xxx (2020).

40. Duriez, T., Brunton, S. L. & Noack, B. R. *Machine learning control-taming nonlinear dynamics and turbulence*, vol. 116 (Springer, 2017).

41. Brenner, M., Eldredge, J. & Freund, J. Perspective on machine learning for advancing fluid mechanics. *Phys. Rev. Fluids* **4**, 100501 (2019).

42. Brunton, S. L., Noack, B. R. & Koumoutsakos, P. Machine learning for fluid mechanics. *Annu. Rev. Fluid Mech.* **52**, 477–508 (2020).

43. Taira, K. *et al.* Modal analysis of fluid flows: An overview. *AIAA J.* **55**, 4013–4041 (2017).

44. Duraisamy, K., Iaccarino, G. & Xiao, H. Turbulence modeling in the age of data. *Annu. Rev. Fluid Mech.* **51**, 357–377 (2019).

45. Fukami, K., Fukagata, K. & Taira, K. Super-resolution reconstruction of turbulent flows with machine learning. *J. Fluid Mech.* **870**, 106–120 (2019).

46. Fukami, K., Fukagata, K. & Taira, K. Assessment of supervised machine learning methods for fluid flows. *Theor. Comput. Fluid Dyn.* **34**, 497–519 (2020).

47. Ribeiro, J. H. M., Yeh, C.-A. & Taira, K. Randomized resolvent analysis. *Phys. Rev. Fluids* **5**, 033902 (2020).

48. Singh, A. P., Medida, S. & Duraisamy, K. Machine-learning-augmented predictive modeling of turbulent separated flows over airfoils. *AIAA J.* **55**, 2215–2227 (2017).

49. Baydin, A. G., Pearlmutter, B. A., Radul, A. A. & Siskind, J. M. Automatic differentiation in machine learning: a survey. *The J. Mach. Learn. Res.* **18**, 5595–5637 (2017).

50. Jadrich, R., Lindquist, B. & Truskett, T. Unsupervised machine learning for detection of phase transitions in off-lattice systems. i. foundations. *The J. chemical physics* **149**, 194109 (2018).

51. Verma, S., Novati, G. & Koumoutsakos, P. Efficient collective swimming by harnessing vortices through deep reinforcement learning. *Proc. Natl. Acad. Sci.* **115**, 5849–5854 (2018).

52. Popov, P. P. *et al.* Machine learning-assisted early ignition prediction in a complex flow. *Combust. Flame* **206**, 451–466 (2019).

53. Raissi, M., Wang, Z., Triantafyllou, M. S. & Karniadakis, G. E. Deep learning of vortex-induced vibrations. *J. Fluid Mech.* **861**, 119–137 (2019).

54. Hou, W., Darakananda, D. & Eldredge, J. D. Machine-learning-based detection of aerodynamic disturbances using surface pressure measurements. *AIAA J.* **57**, 5079–5093 (2019).

55. Gorr, H. M., Xiong, Z. & Barnard, J. A. Pattern recognition for identification of lysozyme droplet solution chemistry. *Colloids Surfaces B: Biointerfaces* **115**, 170–175 (2014).

56. Kim, N. *et al.* Identification of fluid and substrate chemistry based on automatic pattern recognition of stains. *Anal. Methods* **4**, 50–57 (2012).

57. Hamadeh, L. *et al.* Machine learning analysis for quantitative discrimination of dried blood droplets. *Sci. Reports* **10**, 1–13 (2020).

58. Fukatani, Y. *et al.* Effect of ambient temperature and relative humidity on interfacial temperature during early stages of drop evaporation. *Phys. Rev. E* **93**, 043103 (2016).

59. Innocenzi, P. *et al.* Evaporation of ethanol and ethanol- water mixtures studied by time-resolved infrared spectroscopy. *The J. Phys. Chem. A* **112**, 6512–6516 (2008).

60. Andalib, S., Alshehri, A. & Kavehpour, P. Combined effect of relative humidity and substrate temperature on evaporation of methanol droplet. *J. Coatings Technol. Res.* **16**, 1691–1698 (2019).

61. Buntine, W. L. *A theory of learning classification rules*. Ph.D. thesis, university of technology, sydney (1990).

62. Kononenko, I. Comparison of inductive and naive bayesian learning approaches to automatic knowledge adquisition. b. wielinga editors, current trends in knowledge adquisition (1990).

63. Langley, P. Induction of recursive bayesian classifiers. In *European Conference on Machine Learning*, 153–164 (Springer, 1993).

64. Langley, P. & Sage, S. Induction of selective bayesian classifiers. In *Uncertainty Proceedings 1994*, 399–406 (Elsevier, 1994).

65. Swain, P. H. & Hauska, H. The decision tree classifier: Design and potential. *IEEE Transactions on Geosci. Electron.* **15**, 142–147 (1977).

66. Safavian, S. R. & Landgrebe, D. A survey of decision tree classifier methodology. *IEEE Transactions on Syst. Man, Cybern.* **21**, 660–674 (1991).

67. Rao, J. S. & Potts, W. J. Visualizing bagged decision trees. In *AAAI Knowledge Discovery in Databases Proc.*, 243–246 (1997).

Acknowledgements

Author contributions statement

S.A., K.T., and H.P.K. designed research, K.T. and H.P.K. supervised research, S.A. performed research, analyzed data, and wrote the paper.

Competing interests

The authors declare no competing interests.
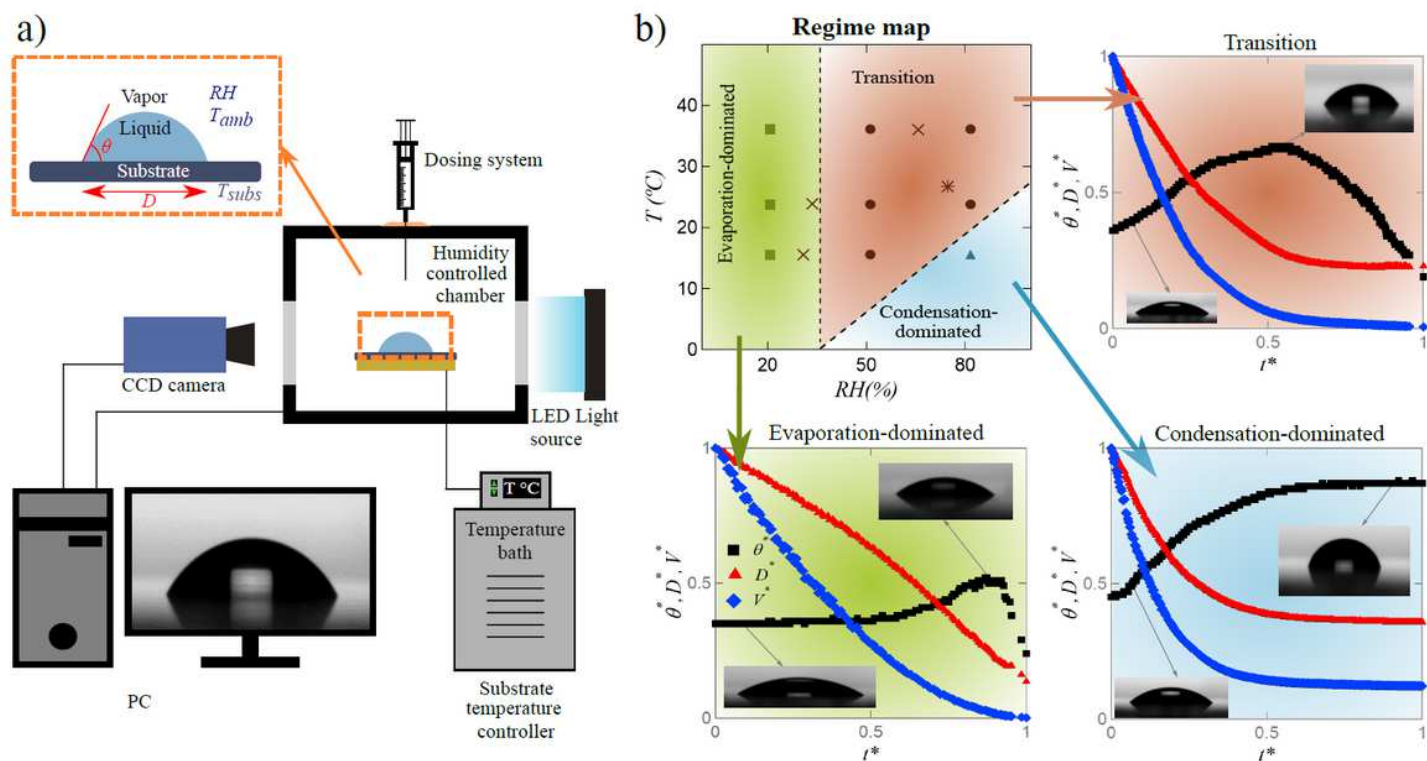
Figures

Figure 1

a) Schematic of the experimental setup with macroscopic parameters of droplet shown in the inset; b) droplet evolution regime map (top-left), evolutions of nondimensional contact angle, volume, and diameter versus time for evaporation-dominated (bottom-left), transition (top-right), and condensation-dominated (bottom-right) regimes.
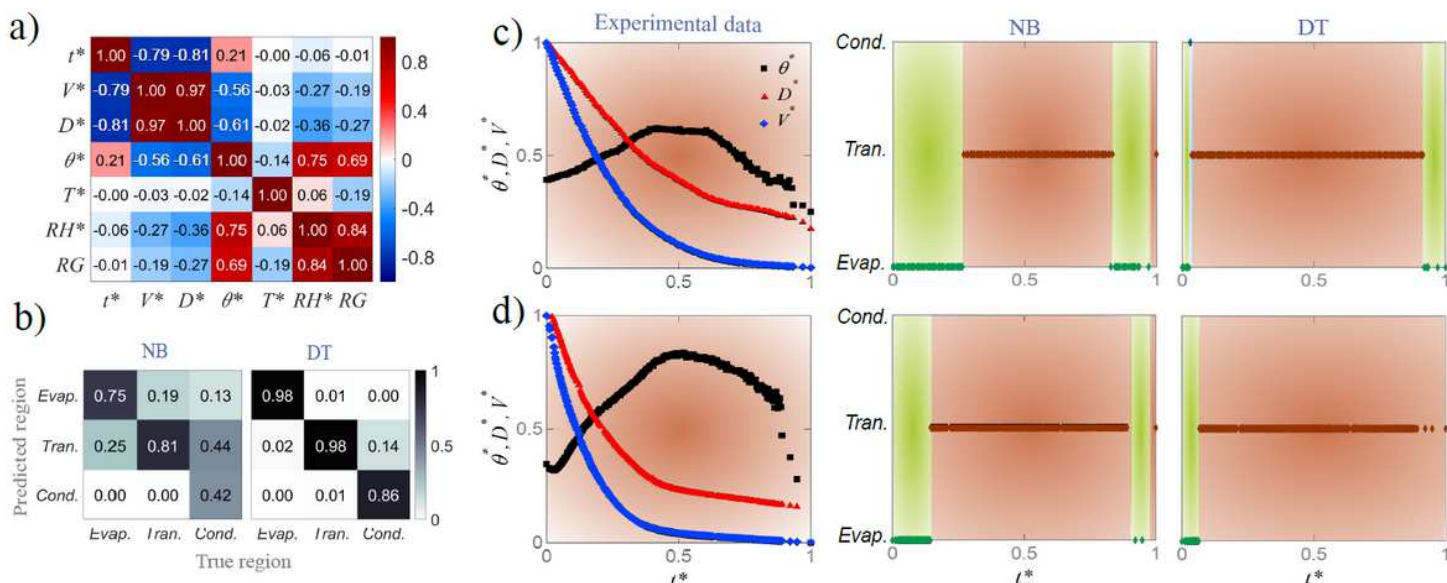
Figure 2

a) Correlation matrix for parameters in droplet evaporation; b) confusion matrices for the test set with Naïve Bayes (NB) and decision tree (DT) algorithms; c) experimental data of droplet evolution in validation set (RH = 80% and T = 35⬚C) and point-by-point validation results with NB and DT classifiers; d) experimental data for experiment 4 in prediction set (RH = 75% and T = 25⬚C) and point-by-point prediction results with NB and DT classifiers.
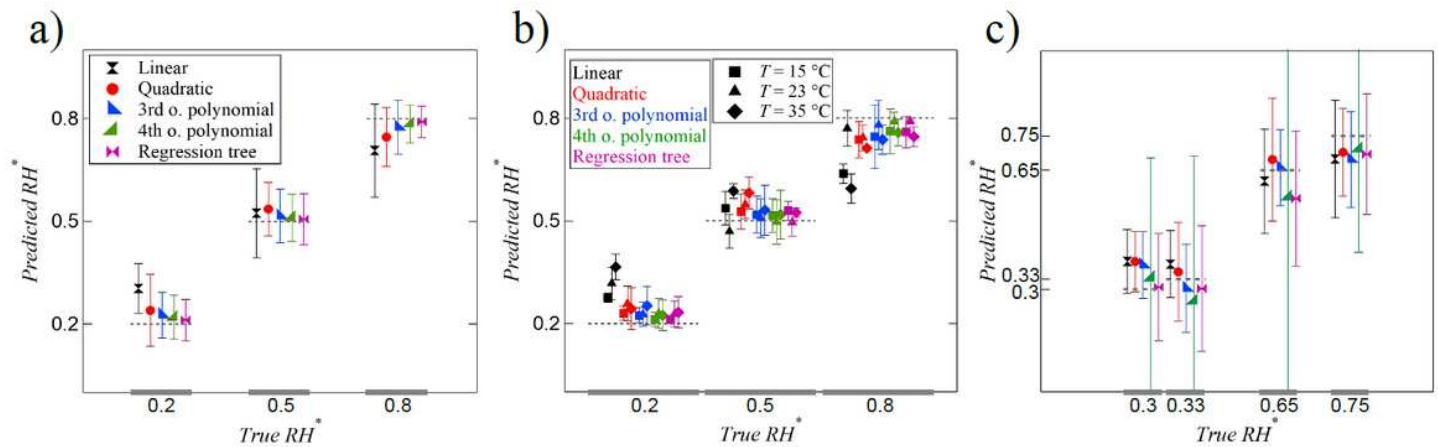

Figure 3

Relative humidity regression results: a) test set; b) validation; c) prediction. Markers for all colors in (b) represent different temperatures as shown in legend. Markers and colors in (c) are the same as in (a).
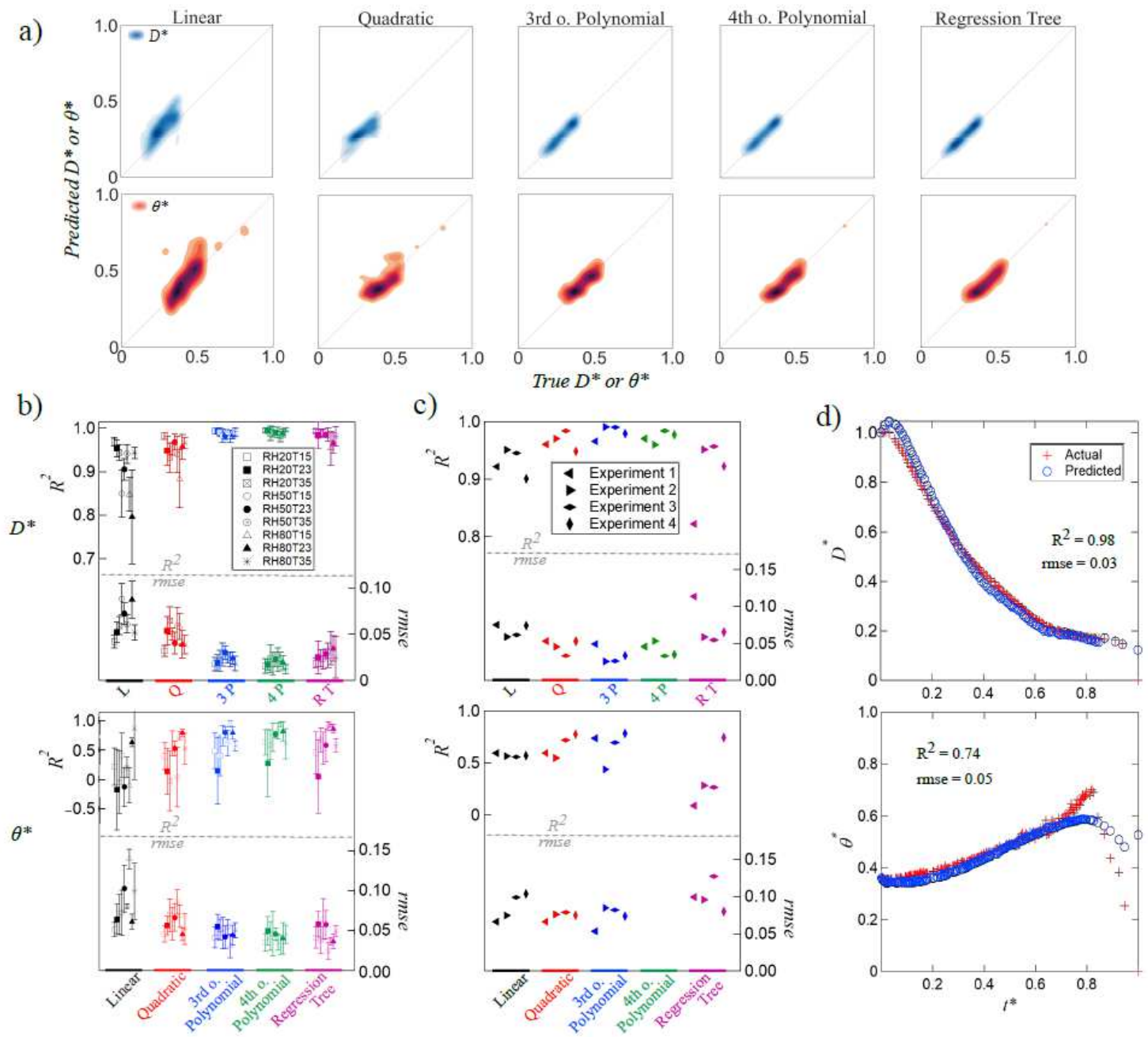
Figure 4

Diameter (D*) and contact angle (θ ⬚) regression results: a) test set; b) validation; c) prediction set; d) diameter prediction with quadratic regression for E3 (left) and contact angle prediction with 3rd order polynomial regression for E1 (right)