

Article

A Machine Learning Approach for Predicting the Maximum Spreading Factor of Droplets upon Impact on Surfaces with Various Wettabilities

Moussa Tembely ^{1,*} , Damien C. Vadillo ² , Ali Dolatabadi ¹ and Arthur Soucemarianadin ³

¹ Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, ON M5S 2E8, Canada; dolat@mie.utoronto.ca

² 3M Corporate Research Analytical Laboratory, Saint Paul, MN 55144, USA; dvadillo@mmm.com

³ Univ. Grenoble Alpes, CNRS, Grenoble INP, LEGI, 38000 Grenoble, France; arthur.soucemarianadin@univ-grenoble-alpes.fr

* Correspondence: moussa.tembely@utoronto.ca or moussa.tembely@concordia.ca

Abstract: Drop impact on a dry substrate is ubiquitous in nature and industrial processes, including aircraft de-icing, ink-jet printing, microfluidics, and additive manufacturing. While the maximum spreading factor is crucial for controlling the efficiency of the majority of these processes, there is currently no comprehensive approach for predicting its value. In contrast to the traditional approach based on scaling laws and/or analytical models, this paper proposes a data-driven approach for estimating the maximum spreading factor using supervised machine learning (ML) algorithms such as linear regression, decision tree, random forest, and gradient boosting. For this purpose, a dataset of hundreds of experimental results from the literature and our own—spanning the last thirty years—is collected and analyzed. The dataset was divided into training and testing sets, each representing 70% and 30% of the input data, respectively. Subsequently, machine learning techniques were applied to relate the maximum spreading factor to relevant features such as flow controlling dimensionless numbers and substrate wettability. In the current study, the gradient boosting regression model, capable of handling structured high-dimensional data, is found to be the best-performing model, with an R²-score of more than 95%. Finally, the ML predictions agree well with the experimental data and are valid across a wide range of impact conditions. This work could pave the way for the development of a universal model for controlling droplet impact, enabling the optimization of a wide variety of industrial applications.

Keywords: drop impact; maximum spreading diameter; scaling laws; analytical models; machine learning



Citation: Tembely, M.; Vadillo, D.C.; Dolatabadi, A.; Soucemarianadin, A. A Machine Learning Approach for Predicting the Maximum Spreading Factor of Droplets upon Impact on Surfaces with Various Wettabilities. *Processes* **2022**, *10*, 1141. <https://doi.org/10.3390/pr10061141>

Academic Editor: Xiong Luo

Received: 16 April 2022

Accepted: 20 May 2022

Published: 7 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The impact and spreading of droplets on a solid substrate is of paramount importance from a scientific and practical standpoint. Ink-jet printing, metal solidification, spray cooling, microfluidic systems, ice accretion, combustion, coating processes, and pesticide deposition are just some of the myriad of industrial and environmental applications that have spurred decades of research on drop impacts [1–7]. In each of these applications, the maximum spreading diameter is critical, whether it relates to a higher dot resolution in ink-jet printing or better surface coverage efficiency aiming at maximum coverage of the target materials with the minimum amount of liquid. In other extreme cases, superhydrophobic surfaces have been developed to limit the spreading diameter due to their enormous potential applications, such as anti-icing, drag reduction, self-cleaning, anti-bacteria, and corrosion resistance, to mention a few [8–10].

From a scientific point of view, many studies have been devoted to understanding the different phenomena happening after impact and which are controlled by a subtle interplay between inertia, viscosity, and capillary forces for drops smaller than the capillary length. According to the literature [1,11,12], the droplet impacting process can be divided into four

phases: kinematic, spreading, relaxation, and wetting/equilibrium, with different outcomes. The outcomes are closely related to the kinematics and the interfacial properties of the liquid and the solid, the most important of the outcomes being deposition, receding contact line with break-up, splashing, and rebound, the latter being either partial or total [13–17].

Analytical models, which are further discussed in this paper, attempt to predict the maximum spreading diameter of droplet impingement using energy balance, momentum balance, and sometimes empirical considerations [18,19]. These models are crucial for understanding the underlying physics, as well as industrial applications. The kinetic energy, the surface energy, and the viscous dissipation of a droplet impacting on a solid surface must all be taken into account when analyzing the impact dynamics. It has been shown that drop impacts are best described and assessed when appropriate dimensionless groups are used. The Weber and Reynolds numbers are mostly employed in this regard. The Weber number ($We = \rho V_0^2 D_0 / \sigma$) relates the inertia forces to the forces resulting from surface tension; it represents the ratio of kinetic energy to surface energy and is a measure of the deformability of the droplet. The Reynolds number ($Re = \rho V_0 D_0 / \mu$) measures the ratio of inertia to viscosity, where V_0 and D_0 are the initial impact velocity and diameter of the liquid droplet, whilst ρ , σ , and μ are the density, surface tension, and viscosity of the fluid used for the experiments. Two other dimensionless numbers are also sometimes considered in the analysis, namely the Ohnesorge number ($Oh = We^{1/2} / Re$) and the capillary number ($Ca = We / Re$). It should be noted that these numbers are not independent of the first two (We , Re) but may be quite useful for scaling purposes and to better define regions of study and applications. Finally, another quantity of interest is β_{max} , the maximum spreading factor (or parameter). It serves to normalize the maximum deformation of the droplet and represents the ratio between the droplet's maximum spreading diameter (D_{max}) and its initial diameter (D_0).

Existing analytical models can be classified into two main categories: (i) β_{max} as a function of Re and We [17,20] and (ii) β_{max} as a function of Re , We , and θ [14,21], where θ could be the equilibrium contact angle or the advancing contact angle, with the latter being dynamic or not [22]. It is only when θ is introduced into the equations that the interactions between the solid and liquid are considered. Due to the difficulty in predicting the advancing dynamic contact angle theoretically, however, experimental data are often used instead. In this paper, some results from analytical models described above to those obtained using data-driven models are compared.

In view of the quantity of existing experimental results which have been generated over the years, of the various assumptions which are made to construct the models, and of the very large number of models, with some of them relatively specialized for certain fluids and operating conditions [23,24], it may be useful to consider whether one could not resort to some other technique to obtain the maximum spreading diameter. For instance, the lack of an analytical model can be circumvented if one applies machine learning (ML) or deep learning techniques able to unravel hidden interactions and to build a model directly from the experimental data [25]. The ML models in this work are applied in a supervised manner, which means that the data have to be prepared beforehand for a given target, which, in this paper, is the maximum spreading factor. As a result, the main aim of the present study is to investigate different tree-based machine learning models for the prediction of the maximum spreading factor of a single droplet onto a dry horizontal surface with various wettabilities. The simplicity of these models, relying on structured data, allows for better insight into the physics, unlike approaches such as deep learning applied on unstructured data [26,27]. The focus is essentially on the ensemble learning family of algorithms, which offer some of the most promising approaches in terms of prediction accuracy.

This paper's contribution is the development of a comprehensive ML approach for predicting the maximum spreading factors (β_{max}) over a wide range of impact conditions and machine learning models. Existing models do not address these issues and do not permit a direct comparison of β_{max} with empirical and energy models in the literature.

Current methods only predict the diameter after training on a subset of the diameter evolution or address the droplet impact classification problem [28,29].

This paper is organized as follows: Section 2 is devoted to the background on analytical models and scaling analyses which help to determine the maximum spreading diameter, and some of them will be used as benchmarks for comparisons later in the paper. Section 3 presents the machine learning methods used for regression, while Section 4 details the methodology and the results obtained, with special emphasis on the prediction of β_{max} for a large spectrum of dimensionless numbers and contact angles. Finally, Section 5 summarizes the main conclusions of this work and emphasizes the potentialities of machine learning methods in the analysis of heterogeneous data.

2. Prediction of the Maximum Spreading Diameter

2.1. Scaling Analyses and Analytical Models

Scaling laws between the maximum spreading factor (β_{max}) and the Weber (We) or Reynolds (Re) number have been proposed based on the type of force dominating the dynamics of drop impact. On one hand, for capillary-force-dominated spreading, one can obtain a scaling law of $\beta_{max} \propto We^{1/2}$ by energy conservation or as $\beta_{max} \propto We^{1/4}$ by using momentum conservation [17,30]. On the other hand, if the spreading is dominated by viscous force, then one obtains $\beta_{max} \propto Re^{1/5}$, whilst $\beta_{max} \propto Re^{1/4}$ when considering a balance of kinetic energy with viscous dissipation energy at initial and maximum spreading times [14]. It is noteworthy that Clanet et al. [17] proposed a model that fits well experimental data for impacts on superhydrophobic substrates and reads:

$$\beta_{max} = 0.9 We^{1/4} \quad (1)$$

The accuracy of the different estimations can be evaluated, in particular, on the basis of their ability to predict the maximum spreading factor, as performed by Clanet et al. [17], but only for a very specific type of substrate. In this context Laan et al. [31] explored the aforementioned scaling relations to enable universal scaling. The fact that no clear dependency on We or Re is found suggests that all three effects (inertia, capillarity, and viscosity) are important. By interpolating between $We^{1/2}$ and $Re^{1/5}$ using $\beta_{max} \propto Re^{1/5} f_{EC}(WeRe^{-2/5})$, where f_{EC} is a function of the impact parameter $P = WeRe^{-2/5}$, Laan et al. [31] reach the following relation for the maximum spreading factor:

$$\beta_{max} = Re^{1/5} \frac{P^{1/2}}{(A + P^{1/2})} \quad (2)$$

In the first category of models, which do not account for surface interactions, it is worth mentioning the semi-empirical correlation proposed by Roisman [20], based on a combination of scaling laws and experimental data fitting. This model has been extensively used in the literature for benchmarking but, however, fails for very low Weber impacts, whatever the wettability of the substrate, as well as for low-viscosity fluid drops landing on superhydrophobic substrates where the equilibrium contact angle θ_e is higher than 140° .

In the category of models where the surface wettability is taken into account, one of the best known and abundantly cited is probably the one due to Pasandideh-Fard et al. [14], for which the maximum spreading factor is given explicitly:

$$\beta_{max} = \sqrt{\frac{We + 12}{3(1 - \cos \theta) + 4\left(\frac{We}{\sqrt{Re}}\right)}} \quad (3)$$

In their work on carefully characterized substrates, Ukiwe and Kwok [18] concluded that the Pasandideh-Fard et al. model provided the best estimation, among a set of models, on their experimental data for β_{max} . This model was then modified to include different surface energy terms, resulting in new models [22].

2.2. Data-Driven Models

There are several traditional machine learning (ML) and deep learning methods, including decision trees [32], nearest neighbors, support vector machines, random forest, and many others [33]. Recently, neural networks [27] have shown significant potential in a wide range of applications [34–38]. Nevertheless, none of these models consistently outperform the other methods, and the performance of the model is data-dependent. Coming to the specific problem at hand, fluid dynamics presents challenges that differ from those commonly tackled in many applications of ML, such as speech and image recognition or even the control of manufacturing plants. The application of ML methods for characterizing drop impacts is rather limited, to the best of our knowledge. In 2018, Um et al. [39] proposed a new data-driven approach for modeling liquid splashes utilizing neural networks, which can learn to generate small-scale splash details using training data collected from high-fidelity numerical simulations. In the continuation of neural networks, Heidari et al. [28] present a method based on the NARX-ANN approach for predicting droplet spreading dynamics on a solid substrate. These types of neural networks seem to be particularly effective in predicting non-linear behavior [40], but, as usual with neural networks, the underlying phenomena are rather obfuscated. Moreover, Heidari et al. [28] do not conduct any comparisons with drop impact experiments performed on superhydrophobic substrates, making it difficult to determine if their neural network is equally effective regardless of the values of Weber, Reynolds, and contact angles, since it is found that superhydrophobic substrates have a peculiar behavior, as reported by [41]. Finally, Azimi et al. [29] utilized an approach based on machine learning to predict the behavior of impacting drops on hydrophobic and superhydrophobic surfaces. Their model, based on a random-forest approach, is used to predict drop behavior with high accuracy while aiming at finding those conditions favorable for producing a bouncing behavior upon drop impact for a wide range of Weber numbers and surfaces. Their results based on an ensemble method offer design criteria for superhydrophobic surfaces, so it is more of a classification problem than a regression one, as detailed in this paper.

3. Machine Learning Methods for Predicting Maximum Spreading Diameter

3.1. Multiple Linear Regression (MLR)

The first ML considered is a multiple linear regression (MLR), which involves predicting the target variable $F(x_i)$ (in our case, the maximum spreading factor) based on the p input features x_1, x_2, \dots, x_p so that the following linear relation is satisfied:

$$F(x) = \alpha_0 + \sum_{i=1}^p \alpha_i x_i + \epsilon \quad (4)$$

where α_i is the regression coefficient for the i th input feature x_i and ϵ represents the error. Using the set of observations compiled from our own experiments [22,42] and data from the literature (see below), the coefficients α_i can be computed using the least-squares approach by optimization.

3.2. Regression Tree (RT)

The regression tree (RT) method is a supervised nonparametric statistical method [43] that provides an alternative to parametric regression techniques, which typically need assumptions and simplifications to generate the described relationship. The main goal of a decision tree-based regression model is to predict the value of a target variable by learning simple decision rules based on the input data. To construct the regression tree, the entire dataset that represents the root node is recursively partitioned into more homogeneous groups, each of which is described by its own node. At the end of the splitting process, the resulting groups are known as leaves or terminal nodes. If the dataset is partitioned into M regions or homogeneous groups such as R_1, R_2, \dots, R_M , the system model can be expressed as:

$$F(x) = \text{average}(F(x_j)/x_j \in R_M) + \epsilon \quad (5)$$

where x_j corresponds to the input variables at j th observation while $F(x_j)$ represents target variables. Overall, Equation (5) illustrates that the predicted value can be computed based the average of $F(x_j)$ in region R_M with the associated error ϵ .

It should be mentioned that a small tree may not be able to capture a nonlinear relation, but a very large one can overfit with overly complex decision rules, which is more typical of the behaviour of a single tree. Consequently, tree size is a tuning parameter that can be adjusted in accordance with the data.

3.3. Random Forest (RF)

The random forest method comprises regression trees developed in a random manner. The purpose of the RF method is to build an ensemble of low-correlated regression trees and average them in order to address the weaknesses of decision trees and reduce variance, which is quite high in the case where a single tree is generated. The random forest method was first proposed by Breiman [44]. Randomization is applied in two processes to build low-correlated trees: each tree is formed using a random subset of observations and each node of a tree is split using a random subset of input features. RF can predict a new observation x by using a tree-based regression model in the forest to make new predictions $F_S(x)$ and then taking the average of the S prediction values as follows:

$$F(x) = \frac{1}{S} \sum_{s=1}^S F_s(x) + \epsilon \quad (6)$$

Thus, the random forest output is a meta-estimator that integrates the results of several predictions and many decision trees, with some adjustments that prevent overfitting and underfitting if sufficient trees are generated for a particular task.

3.4. Gradient Boost Regression Tree (GBRT)

Schapire [45] introduced the first boosting algorithm in 1990, followed by Freund and Schapire in 1996 with the AdaBoost method [46]. Boosting is a potent method for merging numerous base classifiers to create a group whose performance can be much superior to that of any base classifier. Boosting's central concept is to add new models to the ensemble sequentially but according to a predetermined pattern, as opposed to the random forest technique, in which trees are developed in parallel. At each successive iteration, a weak, base-learner model is trained relative to the error of the entire ensemble learned [47]. Gradient boosting, which was developed by Friedman [48], constructs the model in stages, such as in other boosting approaches, but generalizes weak learner models and optimizes the loss function based on gradient descent. As with the random forest method, it can be used for regression or classification problems and results in a lower variance than a single regression tree. The mathematical formulation for GBRT reads:

$$F(x) = \sum_{m=1}^M \gamma_m h_m(x) \quad (7)$$

where $h_m(x)$ are the basis functions, which, in the context of boosting, are typically referred to as weak learners. GBRT employs small, fixed-size decision trees for weak learners. Numerous abilities of decision trees make them advantageous for boosting. GBRT constructs the additive model in forward stages [48] as follows:

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x) \quad (8)$$

At each stage, the decision tree $h_m(x)$ is selected to minimize the loss function L , such as:

$$F_m(x) = F_{m-1}(x) + \operatorname{argmin} \sum_{i=1}^n L(y_i, F_{m-1}(x_i) - h(x)) \quad (9)$$

The initial model F_0 varies depending on the problem; for least squares regression, the mean of the target values is typically chosen, while the loss function is the mean squared error.

3.5. Data Description and Processing

The data that we use in this paper come from our own experiments [22,42], as well as quite a large number of experimental investigations which have been produced by different researchers over the past thirty years. Table 1 summarizes the provenance of the prominent data. The experimental results cover a large spectrum of Weber (0.2 to 802) and Reynolds numbers (8.7 to 14191), with drops impacting on substrates for which the equilibrium contact angle varies between 0 degree and 164 degrees.

Table 1. Selected experimental β_{max} and drop impact characteristics considered for ML models.

Investigation	Range of Weber Numbers	Range of Reynolds Numbers	Range of Equilibrium Contact Angles	Range of Values of Maximum Spreading Factors
Rioboo et al. [12]	36–614	20–10,394	0°–154°	1.4–5.4
Pasandideh-Fard et al. [14]	27–641	213–5833	20°–140°	2.15–4.4
Bayer et al. [15]	11.5	1078	10°–115°	2.1–2.5
Šikalo et al. [49]	50–802	27–12,300	0°–105°	1.7–5.2
Vadillo et al. and Vadillo [22,42]	0.21–12.4	39–2400	5°–160°	1.2 – 2.3
Roisman et al. [50]	0.88–7.9	400–1200	92°	1.2–1.5
Mao et al. [24]	11–511	1966–13,297	37°–97°	1.65–4.94
Kim and Chun [51]	30–582	3222–14,191	6.2°–87.5°	2.3–5.1
Antonini et al. [41]	33	3379	99°–164°	1.8–2.45

The dataset which has been constituted for this paper from various sources is probably one of the most extended reported in the literature on maximum spreading diameter. The detailed statistics on the dataset are shown in Table 2, with the number of data points(count), the minimum, mean, percentile, and the maximum of the different parameters describing the impact conditions. In addition, we represent various cross-plots between the dataset’s variables in Figure 1. Finally, to apply the ML models, the collected data are subdivided into two independent and non-overlapping datasets: a training set of 70% and a test set of 30%. While the training set is used for regression analysis to predict maximum diameters, the test set is used for assessing prediction accuracy.

Table 2. Statistics on the test dataset.

	We	Re	θ_{eq}	β
count	204	204	204	204
mean	101.7	2551.3	49.7	2.5
std	170.1	3315.0	36.5	1.1
min	0.2	8.7	0.1	1.0
25%	10.6	140.3	16.0	1.6
50%	33.0	1046.3	48.5	2.2
75%	117.1	3392.8	77.0	3.2
max	802.0	14,191.0	164.0	5.4

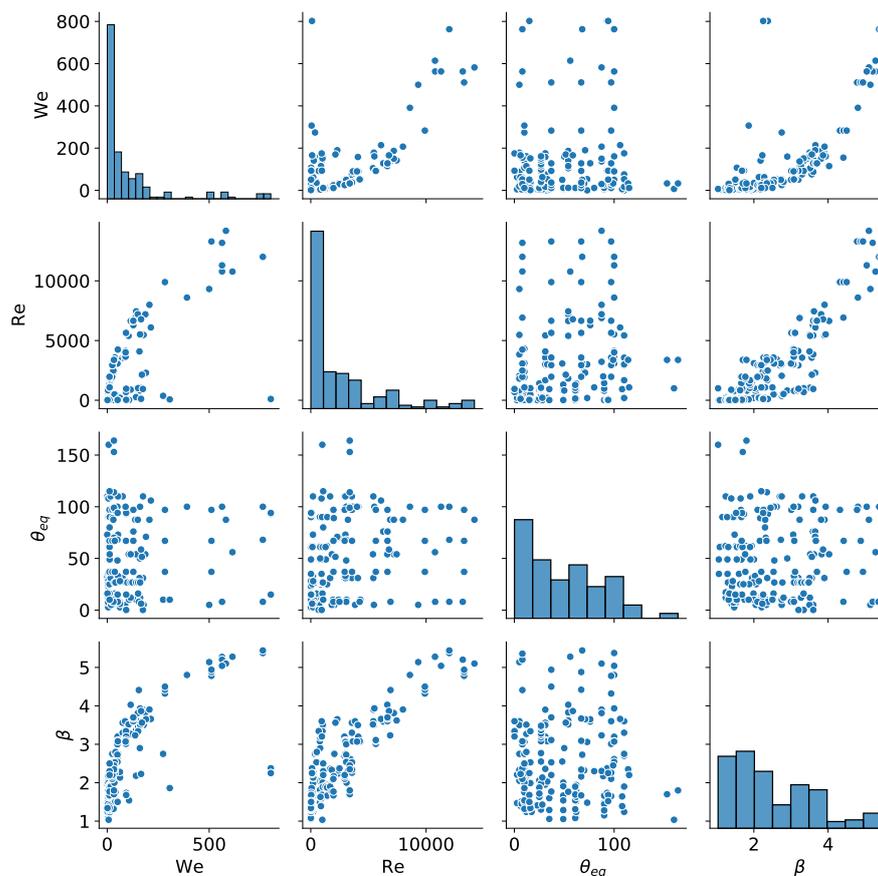


Figure 1. Exploratory analysis of the dataset based on the cross-plots of the different variables.

3.6. Feature Selection

As shown in Table 1, we have settled for the choice of Weber number (We), Reynolds number (Re), and equilibrium contact angle (θ_e) as features for the machine learning methods since these are the easiest to obtain when going through the different results reported in the literature. This does not mean, however, that other quantities such as surface tension, density of the fluid, and advancing and receding angles are unimportant, although some of them are present thanks to dimensionless numbers. To be practical, the point is that they are not mentioned in some papers, whilst all of them report the dimensionless flow numbers and the equilibrium contact angles of substrates used in the investigations.

3.7. ML Model Evaluation Metrics

To assess the performance of the different regression models, we use the R2-score between the actual (y_i) and predicted values (\tilde{y}), which measures the predictive capability of the ML model. This score can be defined as follows:

$$\text{R2-score} = 1 - \frac{\sum_{i=1}^N (y_i - \tilde{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (10)$$

where $\bar{y} = \sum_{i=1}^N y_i / N$, with N corresponding to the number of samples. In addition, the mean absolute error (MAE) is computed as follows:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - \tilde{y}_i| \quad (11)$$

Finally, we use randomized search and grid search to tune and set the hyperparameters of the different ML models.

4. Results and Discussion

4.1. Linear Regression Model (LRM)

After collecting and processing the data, we are in a position to test the accuracy of the data-driven models. A simple linear regression model constructs an optimized linear combination of the predictors variables, also known as features. The relatively strict assumption of a linear relationship makes it difficult to match with real-world problems, where many factors may contribute to the final result in a non-linear manner. One advantage of the multiple linear regression method is that it can be used to evaluate the contribution of each individual input variable to the target variable. The result of the model on estimating β_{max} is shown in Figure 2, with an R2-score of 78% and MAE of 0.41, highlighting the non-linearity, between the input features—chosen as Re , We , θ_e —and the target, β_{max} , which cannot be fully captured by the linear model.

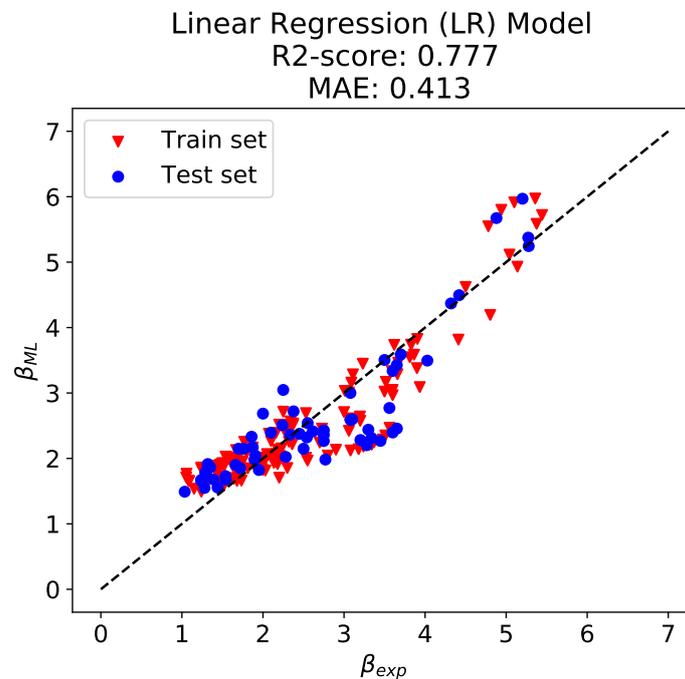


Figure 2. Linear regression model with three features.

4.2. Decision Tree (DT)

Due to their simplicity, practicality, and interpretability, decision trees have gained in popularity as a ML technique in recent years. In Figure 3, we find that the first split, at the level of the root node, is performed for a Reynolds number smaller than or equal to 3613. The mean maximum spreading factor value for the 142 samples, which concerns only the train set, in the root node is 2.44 and the mean squared error (MSE) is 1.16. The resulting tree is of quite small depth with only six leaves and is representative of the trees that will be used in the methods discussed later on. It should be emphasized that at such a low depth level, the highest MSE for a leaf is 0.13, so more than 10 times lower than the initial value, and the MSE could be as small as 0.04, i.e., more than 30 times smaller than the initial value. The regression tree seems to accommodate quite well for non-linearities that exist in the drop spreading phenomenon.

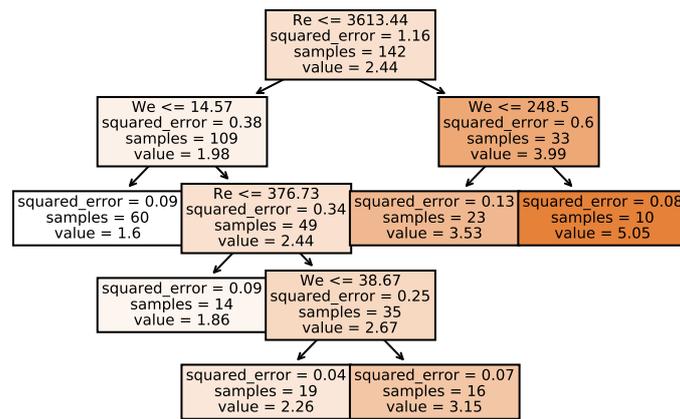


Figure 3. Decision tree graph for the maximum spreading factor.

4.3. Random Forest (RF)

As already explained, a single classification or regression tree is easy to obtain and its interpretability is straightforward, but one is confronted with the problem of high variance. As with the random forest approach, we plot β_{exp} vs. β_{ML} in Figure 4 below. It can be noted that there is a quantitative difference between the coefficients of determination found for the multiple linear regression and decision tree methods and the one for the random forest method (variation from 0.78 to almost 0.95), demonstrating the excellent non-linear mapping generalization ability of the latter method. We lose somewhat in terms of interpretability because of the number of trees which are grown, but we gain quite significantly in terms of prediction accuracy.

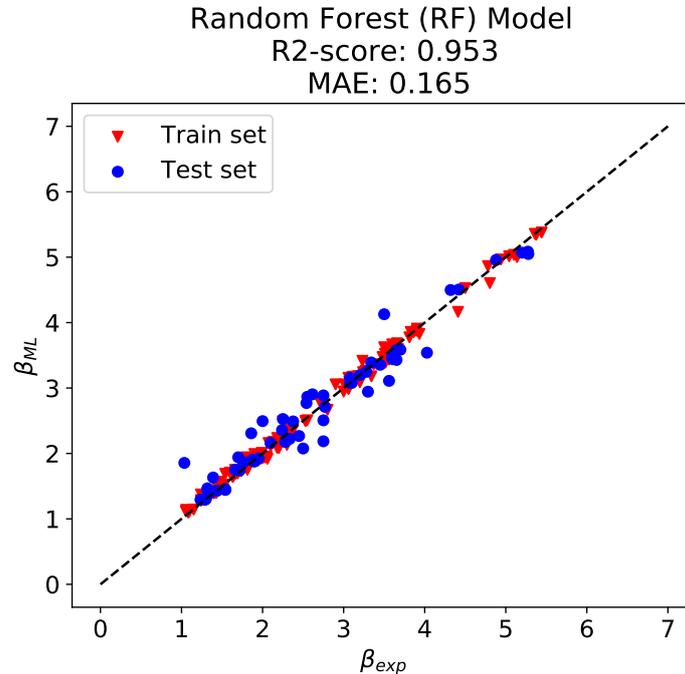


Figure 4. Experimental maximum spreading factor plotted against β_{max} given by the random forest method.

4.4. Gradient Boost Regression Tree (GBRT)

As a final method, we have chosen the gradient boost regression tree technique, which performs well, if not the best, on a variety of datasets. Instead of trees grown parallel, the trees are grown sequentially, correcting each time. As for the random forest method, there are a number of parameters which need to be tuned: (i) the number and depth of

trees used in the ensemble, (ii) the learning rate, which should be a small number, used to limit how much each tree contributes to the ensemble, and (iii) random sampling such as fitting trees on random subsets of features and samples. One may note that the comparison (Figure 5) is slightly better than for the random forest method, which is a testimonial to the adequacy of this method to tackle highly non-linear problems, in line with results reported in the literature.

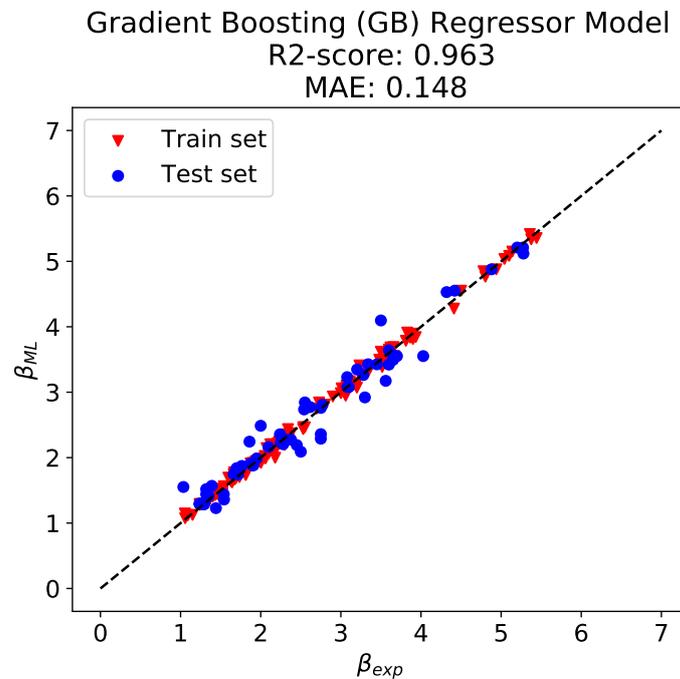


Figure 5. β_{max} experimental plotted against β_{max} given by the gradient boosting method.

Finally, we have added Table 3 summarizing the performance of the different ML models. We can observe that GBRT outperforms the other models in terms of both R2-score and MAE.

Table 3. Summary of the performance of the different models.

Evaluation Metrics	LR	RT	RF	GBRT
R2-score	0.777	0.885	0.953	0.963
MAE	0.413	0.308	0.165	0.148

4.5. Importance of Features

Since GBRT outperforms the different ML models tested, this technique has been used to determine the relevance of the input features controlling droplet impact on a flat surface. The result is shown in Figure 6, where the relevance of Re and We compared to θ_e is retrieved.

Finally, we provide the results of various models on the test set, hitherto unseen data (Table 4), including machine learning-based GBRT, the models of Pasandideh-Fard et al. [14], and Laan et al. [31]. The results, as shown in Figure 7, demonstrate that machine learning models can outperform existing models, paving the way for a data-driven approach to a better understanding and control of droplet formation and impact.

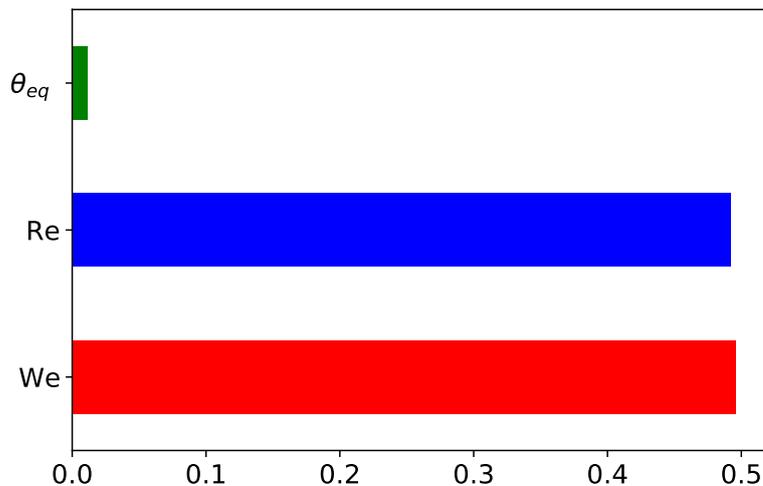


Figure 6. Feature importance as given by the GBRT.

Table 4. Statistics on the test dataset.

-	We	Re	θ_e
count	62	62	62
mean	122.7	2722.2	47.8
std	191.2	3394.8	39.1
min	0.2	8.7	0.1
max	802.0	13,297.2	160.0

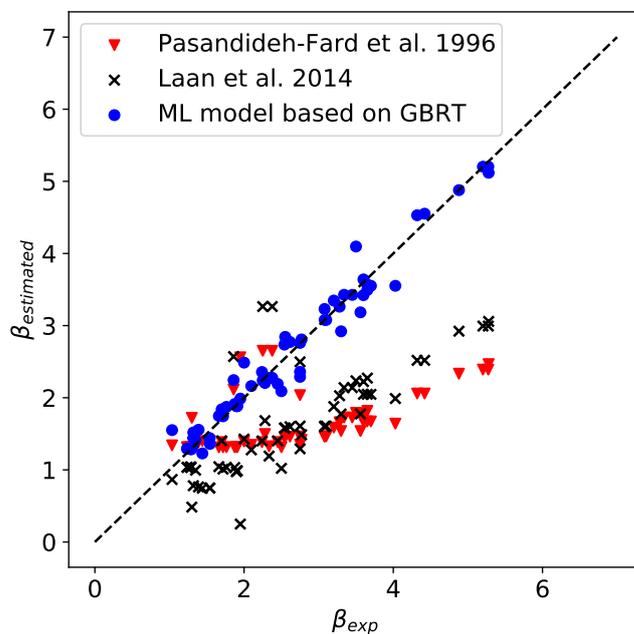


Figure 7. A comparative estimation of the maximum spreading factor based on machine learning, scaling, and analytical models. Pasandideh-Fard et al. [14], and Laan et al. [31].

5. Conclusions and Perspectives

The maximum spreading diameter of impacting drops plays a significant role in a number of industrial applications. Currently, there are three main methods for estimating the maximum spreading factor: (i) scaling law analyses, (ii) energy balance approach, and (iii) numerical simulation. In contrast, this paper offers a different approach to the issue. Indeed, recent advances in computing technologies have led to the development of an array

of accurate and sophisticated machine learning techniques. In addition, the best machine learning technique cannot be determined beforehand, and numerous algorithms must be examined to determine their applicability for a particular task. This study compared the performance of commonly used scaling laws and analytical models to ensemble-based machine learning approaches, such as random forest (RF) and gradient boost (GB), which are both widely employed for prediction. We found that machine learning algorithms (RF and GB) that use decision trees as their base learners can predict experimental results of the maximum spreading factor more accurately than conventional methods. We posit that some of this improved estimation is due to hidden interactions (fluid properties fluctuations, substrate heterogeneities, experimental artifacts, and so on) that are not always completely controlled in experiments and are not accounted for in analytical models. Finally, the proposed machine learning models are highly versatile; they do not require any adjustable parameters and may be used to improve the interpretation of heterogeneous data, the detection of trends, and the performance of industrial processes involving droplet impact and spreading.

Author Contributions: Experimental Investigation, D.C.V.; Modeling, M.T.; Methodology, Conceptualization, Supervision, Writing and Editing M.T., D.C.V., A.D. and A.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

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

Acknowledgments: D.C.V. thanks 3M Corporate Research for permission to publish this paper

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Yarin, A. Drop Impact Dynamics: Splashing, Spreading, Receding, Bouncing ... *Annu. Rev. Fluid Mech.* **2006**, *38*, 159–192. [[CrossRef](#)]
2. Josserand, C.; Thoroddsen, S. Drop Impact on a Solid Surface. *Annu. Rev. Fluid Mech.* **2016**, *48*, 365–391. [[CrossRef](#)]
3. Danzebrink, R.; Aegerter, M.A. Deposition of micropatterned coating using an ink-jet technique. *Thin Solid Film.* **1999**, *351*, 115–118. [[CrossRef](#)]
4. De Gans, B.J.; Duineveld, P.C.; Schubert, U.S. Inkjet printing of polymers: State of the art and future developments. *Adv. Mater.* **2004**, *16*, 203–213. [[CrossRef](#)]
5. Fedorchenko, A.I.; Wang, A.B.; Wang, Y.H. Effect of capillary and viscous forces on spreading of a liquid drop impinging on a solid surface. *Phys. Fluids* **2005**, *17*, 093104. [[CrossRef](#)]
6. Tembely, M.; Vadillo, D.; Soucemarianadin, A.; Dolatabadi, A. Numerical Simulations of Polymer Solution Droplet Impact on Surfaces of Different Wettabilities. *Processes* **2019**, *7*, 798. [[CrossRef](#)]
7. Gomaa, H.; Tembely, M.; Esmail, N.; Dolatabadi, A. Bouncing of cloud-sized microdroplets on superhydrophobic surfaces. *Phys. Fluids* **2020**, *32*, 122118. [[CrossRef](#)]
8. Liao, R.; Zuo, Z.; Guo, C.; Yuan, Y.; Zhuang, A. Fabrication of superhydrophobic surface on aluminum by continuous chemical etching and its anti-icing property. *Appl. Surf. Sci.* **2014**, *317*, 701–709. [[CrossRef](#)]
9. Bhushan, B.; Jung, Y.C. Natural and biomimetic artificial surfaces for superhydrophobicity, self-cleaning, low adhesion, and drag reduction. *Prog. Mater. Sci.* **2011**, *56*, 1–108. [[CrossRef](#)]
10. Yang, Z.; Liu, X.; Tian, Y. Hybrid Laser Ablation and Chemical Modification for Fast Fabrication of Bio-inspired Super-hydrophobic Surface with Excellent Self-cleaning, Stability and Corrosion Resistance. *J. Bionic Eng.* **2019**, *16*, 13–26. [[CrossRef](#)]
11. Rioboo, R.; Tropea, C.; Marengo, M. Outcomes from a drop impact on solid surfaces. *At. Sprays* **2001**, *11*, 155–165. [[CrossRef](#)]
12. Rioboo, R.; Marengo, M.; Tropea, C. Time evolution of liquid drop impact onto solid, dry surfaces. *Exp. Fluids* **2002**, *33*, 112–124. [[CrossRef](#)]
13. Chandra, S.; Avedisian, C.T. On the collision of a droplet with a solid surface. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **1991**, *432*, 13–41. [[CrossRef](#)]
14. Pasandideh-Fard, M.; Qiao, Y.M.; Chandra, S.; Mostaghimi, J. Capillary effects during droplet impact on a solid surface. *Phys. Fluids* **1996**, *8*, 650–659. [[CrossRef](#)]

15. Bayer, I.S.; Megaridis, C.M. Contact angle dynamics in droplets impacting on flat surfaces with different wetting characteristics. *J. Fluid Mech.* **2006**, *558*, 415–449. [[CrossRef](#)]
16. Marengo, M.; Antonini, C.; Roisman, I.V.; Tropea, C. Drop collisions with simple and complex surfaces. *Curr. Opin. Colloid Interface Sci.* **2011**, *16*, 292–302. [[CrossRef](#)]
17. Clanet, C.; Béguin, C.; Richard, D.; Quéré, D. Maximal deformation of an impacting drop. *J. Fluid Mech.* **2004**, *517*, 199–208. [[CrossRef](#)]
18. Ukiwe, C.; Kwok, D.Y. On the maximum spreading diameter of impacting droplets on well-prepared solid surfaces. *Langmuir* **2005**, *21*, 666–673. [[CrossRef](#)]
19. Du, J.; Wang, X.; Li, Y.; Min, Q.; Wu, X. Analytical Consideration for the Maximum Spreading Factor of Liquid Droplet Impact on a Smooth Solid Surface. *Langmuir* **2021**, *37*, 7582–7590. [[CrossRef](#)]
20. Roisman, I.V.; Rioboo, R.; Tropea, C. Normal impact of a liquid drop on a dry surface: Model for spreading and receding. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **2002**, *458*, 1411–1430. [[CrossRef](#)]
21. Choudhury, R.; Choi, J.; Yang, S.; Kim, Y.J.; Lee, D. Maximum spreading of liquid drop on various substrates with different wettabilities. *Appl. Surf. Sci.* **2017**, *415*, 149–154. [[CrossRef](#)]
22. Vadillo, D. Caractérisation des phénomènes Hydrodynamiques lors de l'impact de Gouttes sur Différents Types de Substrats. 2007. Available online: https://tel.archives-ouvertes.fr/tel-00178665/file/These_Vadillo.pdf (accessed on 15 April 2022).
23. Seo, J.; Lee, J.S.; Kim, H.Y.; Yoon, S.S. Empirical model for the maximum spreading diameter of low-viscosity droplets on a dry wall. *Exp. Therm. Fluid Sci.* **2015**, *61*, 121–129. [[CrossRef](#)]
24. Mao, T.; Kuhn, D.C.S.; Tran, H. Spread and rebound of liquid droplets upon impact on flat surfaces. *AIChE J.* **1997**, *43*, 2169–2179. [[CrossRef](#)]
25. Brunton, S.L.; Noack, B.R.; Koumoutsakos, P. Machine Learning for Fluid Mechanics. *Annu. Rev. Fluid Mech.* **2020**, *52*, 477–508. [[CrossRef](#)]
26. Tembely, M.; AlSumaiti, A.M.; Alameri, W. A deep learning perspective on predicting permeability in porous media from network modeling to direct simulation. *Comput. Geosci.* **2020**, *24*, 1541–1556. [[CrossRef](#)]
27. Hassantabar, S.; Wang, Z.; Jha, N.K. SCANN: Synthesis of Compact and Accurate Neural Networks. *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.* **2021**. Available online: <https://arxiv.org/abs/1904.09090v2> (accessed on 15 April 2022). [[CrossRef](#)]
28. Heidari, E.; Daeichian, A.; Sobati, M.A.; Movahedirad, S. Prediction of the droplet spreading dynamics on a solid substrate at irregular sampling intervals: Nonlinear Auto-Regressive exogenous Artificial Neural Network approach (NARX-ANN). *Chem. Eng. Res. Des.* **2020**, *156*, 263–272. [[CrossRef](#)]
29. Azimi Yancheshme, A.; Hassantabar, S.; Maghsoudi, K.; Keshavarzi, S.; Jafari, R.; Momen, G. Integration of experimental analysis and machine learning to predict drop behavior on superhydrophobic surfaces. *Chem. Eng. J.* **2021**, *417*, 127898. [[CrossRef](#)]
30. Bartolo, D.; Josserand, C.; Bonn, D. Retraction dynamics of aqueous drops upon impact on non-wetting surfaces. *J. Fluid Mech.* **2005**, *545*, 329–338. [[CrossRef](#)]
31. Laan, N.; De Bruin, K.G.; Bartolo, D.; Josserand, C.; Bonn, D. Maximum diameter of impacting liquid droplets. *Phys. Rev. Appl.* **2014**, *2*, 044018. [[CrossRef](#)]
32. Salehi, H.; Burgueño, R. Emerging artificial intelligence methods in structural engineering. *Eng. Struct.* **2018**, *171*, 170–189. [[CrossRef](#)]
33. Touzani, S.; Granderson, J.; Fernandes, S. Gradient boosting machine for modeling the energy consumption of commercial buildings. *Energy Build.* **2018**, *158*, 1533–1543. [[CrossRef](#)]
34. Ha, S.; Jeong, H. Unraveling Hidden Interactions in Complex Systems With Deep Learning. *Sci. Rep. Nat.* **2021**, *11*, 12804. [[CrossRef](#)] [[PubMed](#)]
35. Yuan, X.; Li, L.; Shardt, Y.A.W.; Wang, Y.; Yang, C. Deep Learning With Spatiotemporal Attention-Based LSTM for Industrial Soft Sensor Model Development. *IEEE Trans. Ind. Electron.* **2021**, *68*, 4404–4414. [[CrossRef](#)]
36. Ou, C.; Zhu, H.; Shardt, Y.A.W.; Ye, L.; Yuan, X.; Wang, Y.; Yang, C. Quality-Driven Regularization for Deep Learning Networks and Its Application to Industrial Soft Sensors. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, 1–11. [[CrossRef](#)]
37. Zhu, H.H.; Zou, J.; Zhang, H.; Shi, Y.Z.; Luo, S.B.; Wang, N.; Cai, H.; Wan, L.X.; Wang, B.; Jiang, X.D.; et al. Space-efficient optical computing with an integrated chip diffractive neural network. *Nat. Commun.* **2022**, *13*, 1044. [[CrossRef](#)]
38. Luo, S.; Zhang, Y.; Nguyen, K.T.; Feng, S.; Shi, Y.; Liu, Y.; Hutchinson, P.; Chierchia, G.; Talbot, H.; Bourouina, T.; et al. Machine Learning-Based Pipeline for High Accuracy Bioparticle Sizing. *Micromachines* **2020**, *11*, 1084. [[CrossRef](#)]
39. Um, K.; Hu, X.; Thurey, N. Liquid Splash Modeling with Neural Networks. *Comput. Graph. Forum* **2018**, *37*, 171–182. [[CrossRef](#)]
40. Lee, W.J.; Na, J.; Kim, K.; Lee, C.J.; Lee, Y.; Lee, J.M. NARX modeling for real-time optimization of air and gas compression systems in chemical processes. *Comput. Chem. Eng.* **2018**, *115*, 262–274. [[CrossRef](#)]
41. Antonini, C.; Amirfazli, A.; Marengo, M. Drop impact and wettability: From hydrophilic to superhydrophobic surfaces. *Phys. Fluids* **2012**, *24*, 102104. [[CrossRef](#)]
42. Vadillo, D.C.; Soucemarianadin, A.; Delattre, C.; Roux, D.C. Dynamic contact angle effects onto the maximum drop impact spreading on solid surfaces. *Phys. Fluids* **2009**, *21*, 122002. [[CrossRef](#)]
43. Eedi, H.; Kolla, M. Machine learning approaches for healthcare data analysis. *J. Crit. Rev.* **2020**, *7*, 806–811. [[CrossRef](#)]
44. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
45. Schapire, R.E. The Strength of Weak Learnability. *Mach. Learn.* **1990**, *5*, 197–227. [[CrossRef](#)]

46. Freund, Y.; Schapire, R.E. Experiments with a New Boosting Algorithm. In Proceedings of the 13th International Conference on Machine Learning, Bari, Italy, 3–6 July 1996; pp. 148–156. [[CrossRef](#)]
47. Natekin, A.; Knoll, A. Gradient boosting machines, a tutorial. *Front. Neurobot.* **2013**, *7*, 1–21. [[CrossRef](#)]
48. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [[CrossRef](#)]
49. Šikalo, Š.; Marengo, M.; Tropea, C.; Ganić, E.N. Analysis of impact of droplets on horizontal surfaces. *Exp. Therm. Fluid Sci.* **2002**, *25*, 503–510. [[CrossRef](#)]
50. Roisman, I.V.; Opfer, L.; Tropea, C.; Raessi, M.; Mostaghimi, J.; Chandra, S. Drop impact onto a dry surface: Role of the dynamic contact angle. *Colloids Surfaces A Physicochem. Eng. Asp.* **2008**, *322*, 183–191. [[CrossRef](#)]
51. Kim, H.Y.; Chun, J.H. The recoiling of liquid droplets upon collision with solid surfaces. *Phys. Fluids* **2001**, *13*, 643–659. [[CrossRef](#)]